NORTH ATLANTIC TREATY
ORGANIZATION

SCIENCE AND TECHNOLOGY
ORGANIZATION

AC/323(IST-085)TP/561

www.sto.nato.int

STO TECHNICAL REPORT

TR-IST-085

Interactive Visualization of Network Dynamics

(Visualisation interactive de la dynamique des réseaux)

Final Report of Task Group IST-085.

Published June 2014

Distribution and Availability on Back Cover

NORTH ATLANTIC TREATY
ORGANIZATION

SCIENCE AND TECHNOLOGY
ORGANIZATION

AC/323(IST-085)TP/561

www.sto.nato.int

STO TECHNICAL REPORT **TR-IST-085**

Interactive Visualization of Network Dynamics

(Visualisation interactive de la dynamique des réseaux)

Final Report of Task Group IST-085.

The NATO Science and Technology Organization

Science & Technology (S&T) in the NATO context is defined as the selective and rigorous generation and application of state-of-the-art, validated knowledge for defence and security purposes. S&T activities embrace scientific research, technology development, transition, application and field-testing, experimentation and a range of related scientific activities that include systems engineering, operational research and analysis, synthesis, integration and validation of knowledge derived through the scientific method.

In NATO, S&T is addressed using different business models, namely a collaborative business model where NATO provides a forum where NATO Nations and partner Nations elect to use their national resources to define, conduct and promote cooperative research and information exchange, and secondly an in-house delivery business model where S&T activities are conducted in a NATO dedicated executive body, having its own personnel, capabilities and infrastructure.

The mission of the NATO Science & Technology Organization (STO) is to help position the Nations' and NATO's S&T investments as a strategic enabler of the knowledge and technology advantage for the defence and security posture of NATO Nations and partner Nations, by conducting and promoting S&T activities that augment and leverage the capabilities and programmes of the Alliance, of the NATO Nations and the partner Nations, in support of NATO's objectives, and contributing to NATO's ability to enable and influence security and defence related capability development and threat mitigation in NATO Nations and partner Nations, in accordance with NATO policies.

The total spectrum of this collaborative effort is addressed by six Technical Panels who manage a wide range of scientific research activities, a Group specialising in modelling and simulation, plus a Committee dedicated to supporting the information management needs of the organization.

AVT	Applied Vehicle Technology Panel
HFM	Human Factors and Medicine Panel
IST	Information Systems Technology Panel
NMSG	NATO Modelling and Simulation Group
SAS	System Analysis and Studies Panel
SCI	Systems Concepts and Integration Panel
SET	Sensors and Electronics Technology Panel

These Panels and Group are the power-house of the collaborative model and are made up of national representatives as well as recognised world-class scientists, engineers and information specialists. In addition to providing critical technical oversight, they also provide a communication link to military users and other NATO bodies.

The scientific and technological work is carried out by Technical Teams, created under one or more of these eight bodies, for specific research activities which have a defined duration. These research activities can take a variety of forms, including Task Groups, Workshops, Symposia, Specialists' Meetings, Lecture Series and Technical Courses.

The content of this publication has been reproduced directly from material supplied by STO or the authors.

Published June 2014

ISBN 978-92-837-0211-5

Table of Contents

List of Figures

List of Tables

IST-085 Membership List

CANADA

Dr. Zachary JACOBSON
Health Canada Applied Research and Analysis Directorate
Jeanne Mance Building, Floor 5
Tunney's Pasture
Postal Locator 1905C
Ottawa, Ontario K1A 0K9
Email: zack@sigmaxi.org

Dr. Marcus LEM
First Nations and Inuit Health British Columbia
Region, Health Canada
406 – 1138 Melville Street
Vancouver, BC V6E 4S3
Email: marcus_lem@hc-sc.gc.ca

Dr. Maurice Martin TAYLOR
Martin Taylor Consulting
369 Castlefield Avenue
Toronto, Ontario M5N 1L4
Email: mmt@mmtaylor.net

DENMARK

Ms. Lisbeth M. RASMUSSEN
Danish Defence Acquisition and Logistics Organization (DALO)
FMT-PA-NBO05 DALO
Lautrupbjerg 1
DK-2750 Ballerup
Email: lr@mil.dk

GERMANY

Mr. Michael KLEIBER
Fraunhofer-FKIE
Ergonimics and Human Systems Integration
Neuenahrer Str. 20
D-53343 Wachtberg-Werthhoven
Email: michael.kleiber@fkie.fraunhofer.de

SWEDEN

Dr. Pontus SVENSON
FOI (Swedish Defence Research Agency)
SE-16490 Stockholm
Email: pontus.svenson@foi.se

UNITED KINGDOM

Dr. Margaret J. VARGA
QinetiQ Malvern
Spectrum Solutions Division
Malvern Technology Park
St Andrew's Road
Malvern, Worcestershire WR14 3PS
Email: mjvarga@qinetiq.com

UNITED STATES

Dr. David L. HALL
The Pennsylvania State University College of Information Sciences and Technology
University Park, PA 16802-6823
Email: dhall@ist.psu.edu

Mr. Richard MAY
National Visualization and Analytic Center/PNNL
902 Battelle Boulevard
PO Box 999 MSIN J4 – 32
Richland, WA 99532
Email: richard.may@pnl.gov

Dr. Amy K.C.S. VANDERBILT (Chair)
Chief Strategist, Trend POV
141 S Nursery Avenue
Purcellville, VA 20132
Email: Amy@TrendPOV.com

Interactive Visualization of Network Dynamics

(STO-TR-IST-085)

Executive Summary

IST-085/RTG-041 follows a series of predecessor study groups that considered the visualisation of massive military datasets of different kinds and from different viewpoints. One of these predecessors, IST-013/RTG-002 produced a major report (Visualisation of Massive Military Datasets: Human Factors, Applications, and Technologies) that guided the subsequent work of IST-021/RTG-007 (Multimedia Visualisation of Massive Military Datasets) and IST-059/RTG-025 (Visualisation Technology for Network Analysis). In all the earlier works, the problem of how to present networks to aid visualisation frequently arose.

IST-085/RTG-041 and its associated Workshops and Visualisation Network of Experts have noted that interactive visualisation of network dynamics continues to be a major topic of interest and a concern in many problem domains. For decision-making in the areas of:

a) Network discovery;

b) Simulation and prediction supporting adaptive operations, political effects, public health and safety, and security issues; and

c) Uncertain environments and abstract concepts, one needs to understand how to visualise the changes taking place within a network (dynamics) and the trends within that change.

Interactive visualisation, in this context, includes the human in a feedback loop using a variety of presentation and control devices and refers to the interactive nature of the visualisation itself. The Framework and Survey developed by IST-059/RTG-025 appears to provide a good foundation for investigation of the problem space. Anticipated security benefits include a better understanding of how interactive visualisation should be used to discover, simulate and predict network dynamics, and how such interactive visualisation may aid military command decision-making, public health and security operations as well as intelligence network analysis tasks.

The objectives of the TG have been to:

- Compare the utility of various interactive visualisation styles for providing the user knowledge of the dynamics of a network and subsequent trends.
- Develop the required experiments to provide insight into what characteristics of interactive visualisations are most likely to aid the military user in determining and predicting the types of change happening within a network, given various influence factors.
- Produce the present report highlighting interactive visualisation methods that facilitate and make more effective the analysis of network dynamics in applications such as netcentric warfare, counterterrorism including bioterrorism, peacekeeping, public security, and peace support operations.

Collaboration has been an important part of the IST-085 efforts to advance fundamental research and to further its interaction with the global technical community. In partnership with IST-085, different organizations have assisted in the instigation of technology creation, evaluation of developed technology, and engaging in technical exchanges to further NATO goals and objectives.

By conducting these research collaborations and publishing the results, the IST-085 committee is providing a beacon to help focus other researchers onto problems that require additional effort. Through research collaborations it makes sure the issues important to NATO objectives and military missions are being addressed.

Visualisation interactive de la dynamique des réseaux

(STO-TR-IST-085)

Synthèse

L'IST-085/RTG-041 vient à la suite de groupes de travail qui ont étudié la visualisation d'ensembles massifs de données militaires de différentes sortes et sous différents angles. L'un de ces groupes, l'IST-013/RTG-002, a rédigé un important rapport (Visualisation des ensembles massifs de données militaires : facteurs humains, applications et technologies) qui a orienté le travail consécutif de l'IST-021/RTG-007 (Visualisation multimédia des ensembles massifs de données militaires) et de l'IST-059/RTG-025 (Technologie de visualisation pour l'analyse des réseaux). Au cours des premiers travaux, le problème de la présentation des réseaux pour faciliter la visualisation s'est fréquemment posé.

L'IST-085/RTG-041, les séminaires associés et le réseau d'experts de la visualisation ont remarqué que la visualisation interactive de la dynamique des réseaux demeure un sujet d'intérêt majeur et une préoccupation dans beaucoup de domaines problématiques. En vue de la prise de décision dans les domaines de :

a) La découverte des réseaux ;

b) La simulation et la prédiction appuyant des opérations de transformation et concernant les effets politiques, la santé et la sécurité publiques, ainsi que la sûreté ; et

c) Les environnements incertains et les concepts abstraits ; il faut comprendre comment visualiser les changements qui ont lieu au sein d'un réseau (dynamique) et les tendances qui sous-tendent ce changement.

Dans ce contexte, la visualisation interactive inclut l'homme dans une boucle de rétroaction à l'aide de divers dispositifs de présentation et de commande et se réfère à la nature interactive de la visualisation en soi. Le cadre et l'étude développés par l'IST-059/RTG-025 semblent fournir de bons fondements pour l'étude du problème. En matière de sûreté, les avantages seraient une meilleure compréhension de la façon d'utiliser la visualisation interactive pour découvrir, simuler et prédire la dynamique des réseaux et de l'aide que pourrait apporter cette visualisation interactive à la prise de décision du commandement militaire, aux opérations de santé et de sûreté publique et aux tâches d'analyse des réseaux du renseignement.

Les objectifs du groupe de travail étaient les suivants :

- Comparer l'utilité de différents styles de visualisation interactive pour informer l'utilisateur de la dynamique d'un réseau et des tendances qui s'ensuivent.
- Elaborer les expérimentations nécessaires pour cerner quelles caractéristiques des visualisations interactives sont les plus susceptibles d'aider l'utilisateur militaire à déterminer et prédire les types de changements à l'œuvre au sein d'un réseau, étant donné divers facteurs d'influence.
- Rédiger le présent rapport en soulignant les méthodes de visualisation interactive qui facilitent et renforcent l'efficacité de l'analyse de la dynamique des réseaux dans des applications telles que la guerre réseau-centrée, le contre-terrorisme incluant le bioterrorisme, le maintien de la paix, la sûreté publique et les opérations de soutien de la paix.

La collaboration a constitué une part importante des travaux de l'IST-085 visant à faire progresser la recherche fondamentale et faciliter son interaction avec la communauté mondiale des techniciens.

En partenariat avec l'IST-085, différentes organisations ont contribué à inciter à la création de technologies, évaluer la technologie développée et établir des échanges techniques pour favoriser les objectifs de l'OTAN.

En menant ces collaborations de recherche et en en publiant les résultats, le comité IST-085 est un phare qui aide les autres chercheurs à se concentrer sur les problèmes qui nécessitent plus de travail. Par le biais des collaborations de recherche, l'IST-085 veille à ce que les questions importantes pour les objectifs et missions militaires de l'OTAN soient traitées.

Chapter 1 – INTRODUCTION

1.1 BACKGROUND

Visualisation, a means by which people make sense of complex data, can be seen as a human activity supported by technology. A key element of visualisation is the interface through which the human interacts with the data. It includes both the "how" as well as the "what, when, where, and why" of information presentation and control. Visualisation technologies include search engines, algorithmic processes, display and control devices, but what matters is how these technologies enhance and allow people to do their tasks in a timely and effective manner.

During the course of the work of the visualisation Technical Teams, IST RTG-002, RTG-007 and RTG-025, interactive visualisation of network dynamics appeared to be a major topic of interest in numerous problem domains. This was reinforced by the deliberations and recommendations of the Québec, Halden, Toronto and Copenhagen NATO Workshops and was also supported by the observations of the 2001, 2003 and 2005 meetings of the visualisation Network of Experts (N/X). For successful network discovery, simulation and prediction supporting adaptive operations, political effects, public health and other security issues, uncertain environments and abstract concepts one needs to understand how to visualise the changes taking place within a network (dynamics) and the trends within that change. Interactive visualisation, in this context, includes using hardware platforms ranging from PDAs, to laptop or desktop computers to larger platforms such as immersive environments; and refers to the interactive nature of the visualisation itself. The human is part of the process producing and refining the presentations that support decision-making. The Framework and Survey developed by IST-059/RTG-025 provides a foundation for investigation of the problem space.

1.2 MILITARY APPLICATIONS

Military applications of interactive visualisation of network dynamics include:

- **Counterterrorism:** Interactive visualisation of the threat network and network of relationships are important tools in counterterrorism. It is possible to analyse who might be a threat, and to see the effect of removing that threat.
- **Network Common Operational Picture (NetCOP):** NetCOP is the Common Operating Picture (COP) with an additional layer of network information, and is essential in Network-Based Operations (NBO) and cyber warfare.
- **Special Operations and Urban Warfare:** Special operations and urban warfare are by nature adaptive. Interactive visualisation of the network in use and the Common Operating Picture (COP) facilitates simulation and prediction.
- **Threat Network:** A network visualization of a threat network is only of use if the analyst is able to extract the encoded information quickly and precisely. This can be assessed by applying the method of cognitive analysis and modelling either in the design process or in a later evaluation of an existing network visualisation.
- **Display of Networks of Relationships:** Networks of relationships include causal or probabilistic networks that affect planning of military and civil protection operations. These relationships are often not made evident in current planning systems but are created in a commander's mind through his experience and interpretation of his map displays and related data presentations. Improvements in the display of such relationships should promote common understanding across roles as well as improving the speed and robustness of operational planning. (See factsheet [Chapter 2] and later annexes).

- **Network Interdependencies:** The operational context of any mission involves a number of networks that may be inter-related. When planning military operations, one has to assess the influence it may have in the operational environment. Although the direct effects on a network may be clear, side effects on related networks may not be obvious. Using the junction model for visualising network interdependencies is beneficial for predicting the potential impact of side effects in inter-related networks. (See factsheet [Chapter 2] and later annexes).

Chapter 2 – FACT SHEETS

All the factsheets in this chapter start on the new, right-hand page to facilitate their extraction as separate documents.

2.1 COLLABORATION IN NATO IST-085

Collaboration has been an important part of the IST-085 efforts to advance fundamental research and to further our interaction with the global technical community. In partnership with IST-085, different organizations have assisted in the instigation of technology creation, evaluation of developed technology, and engaging in technical exchanges to further NATO goals and objectives.

IST-085 has been involved in a series of direct research collaborations over its existence. The Framework, Embedding Fields, Hypernode, and Visual Interfaces for Text Analytics (VITA) technologies are examples of the vital collaboration conducted between committee members and with partners. The Framework and associated Embedding Fields research represent the core of the committee's research and collaborations to address complex network challenges. The original work on the Framework was derived in collaborative discussions and has been evaluated by groups within the UK Ministry of Defence and Health Canada. The Hypernode work is a key example of how committee members have collaborated on fundamental research. The VITA work was conducted outside of the committee but we helped provide direction and feedback to further the research and resulting technology. All these technologies derived from collaborative efforts and addressing fundamental challenges to better represent large and complex networks are discussed in factsheets and annexes.

The committee's research efforts in these and other areas will not solve the huge challenges we face but it has served to help bring together the larger community of researchers. By conducting these research collaborations and publishing the results, the IST committee is providing a beacon to help focus other researchers onto problems that require additional effort. Through research collaborations we help make sure the issues important to NATO objectives and military missions are being addressed. In some cases the collaborations are in areas not receiving sufficient attention and our work can help open up new research areas. In other cases, we can help to redirect ongoing research. For instance the Hypernode work has provided opportunities for discussions and future collaboration with others doing similar research. This will result in more research opportunities for follow on committees.

We have also worked to raise awareness of the committee and its work through involvement in and collaboration with professional, industrial, and government organizations. For instance we have been involved in the VisWeek conference for the several years serving many different roles. IEEE VisWeek is the premier forum for advances in scientific and information visualization. The event-packed week brings together researchers and practitioners from academia, government, and industry to explore their shared interests in tools, techniques, and technology. In 2012, an IST-085 committee member chaired the conference providing an opportunity for greater exposure and interaction. IST-085 maintains the Network of Experts (N/X) to assist in the development and promotion of collaborations. The N/X represents a new concept in NATO research discussions and activities. It offers an unofficial forum for researchers to exchange information, data and expertise. It carries the advantages that the NATO umbrella can offer, notably in protecting members' intellectual property, while avoiding some of the problems with more formal arrangements, including Governments' current reluctance to join in official arrangements. The N/X supports NATO in its mandate to study and develop methods to aid understanding and use of the contents of massive datasets.

N/X organizes its own set of workshops that are not NATO sponsored and operate under quite different procedures. N/X Workshops are a voluntary association of interested persons attended by any N/X member who cares to come. An N/X Workshop typically emphasizes discussion of issues rather than presentation of results, and even though many members of the N/X are from commercial organizations, attempts are made to make all presentations more relevant to visualization issues than simple "show and tell" to advertise the successes of some particular system or approach.

In 2010 the committee was invited to present at the Visual Analytic Community (VAC) Consortium meeting. The VAC serves as a home for visual analytics discussions and emerging communities of research and practice. VAC is a coordination point for the breadth of disciplines that all play important roles in the continued growth of visual analytics. The National Visualization and Analytics Centre leads the VAC consortium and provides leadership and coordination among the academic community, industry, national laboratories and government to create and deploy visual analytics technologies to help counter future terrorist attacks in the United States and around the globe.

Connection with NVAC has proven to be a vital connection to the broader visual analytics research community. Through NVAC connections have been made with VIVA, UKVAC and VACCINE.

Vancouver Institute for Visual Analytics (VIVA) was launched as a joint SFU/UBC research institute to leverage these institutions' significant expertise in the field of Visual Analytics (VA). VIVA seeks to develop VA expertise in industry, government, and university organizations in Canada, as well as to promote collaborative VA research.

The UK Visual Analytics Consortium (UKVAC) is a partnership of UK universities with a shared interest in establishing a multi-disciplinary scientific community in the UK dedicated to promoting and contributing to the visual analytics research and development agenda in the UK. UKVAC is coordinated from Middlesex University and includes Imperial College London, Oxford University, Bangor University and University College London. UKVAC is working with the support of the US Department of Homeland Security via the US National Visualization and Analytics Centre at Pacific Northwest National Laboratory, Washington State, and in close collaboration with the UK Home Office.

The Visual Analytics for Command, Control and Interoperability Environments (VACCINE) Centre is lead by Purdue University and funded by the U.S. Department of Homeland Security. VACCINE was created to develop interactive visualization and analytical tools to help our country's 2.3 million+ extended homeland security personnel understand an ever-growing sea of data. VACCINE focuses on the education, research, development, and deployment of new tools and technologies that communicate and disseminate information to first responders and decision makers.

Through shared research, outreach, workshops, and coordination with other research organizations around the world, IST-085 has established an important collaboration network. We have and will continue to use this resource to generate new ideas and develop technology vital to the NATO mission.

2.2 VISTG REFERENCE MODEL

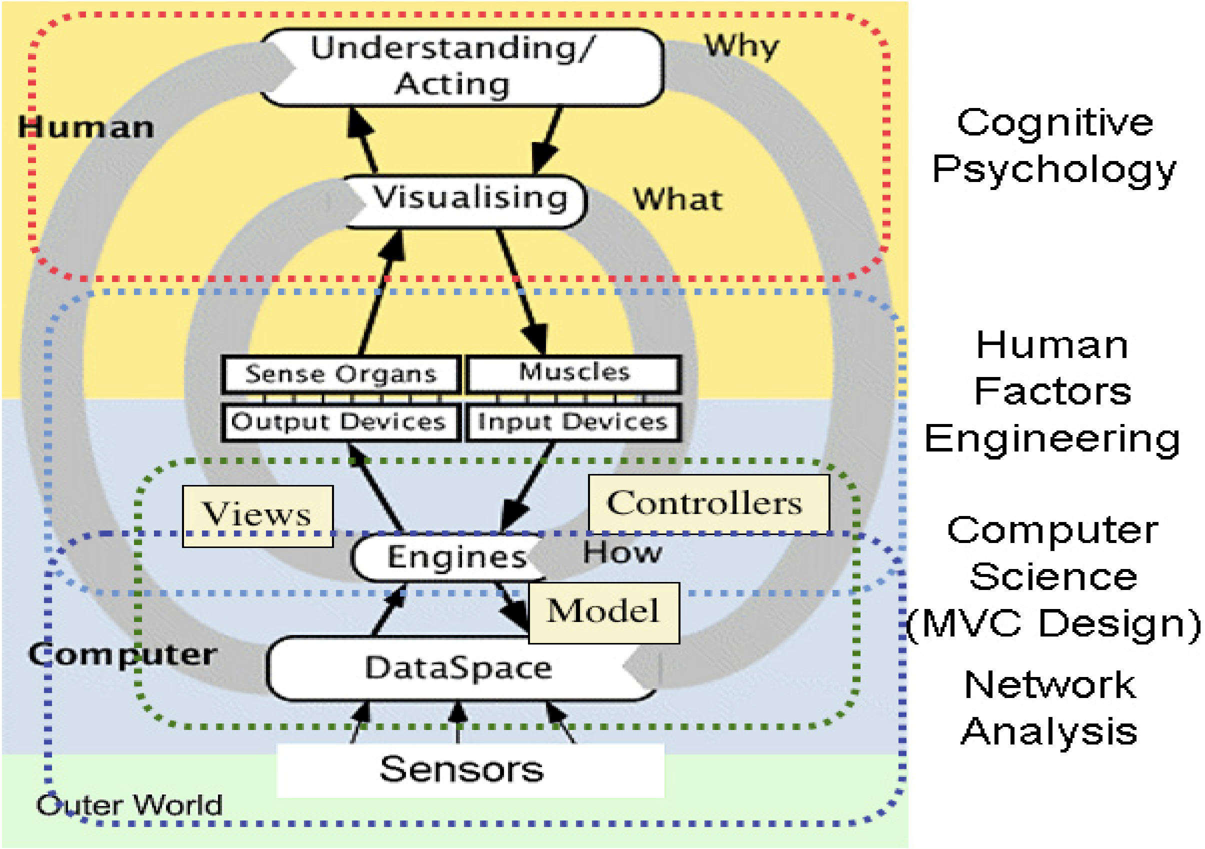


Figure 2-1: VisTG Reference Model and Associated Disciplines.

The VisTG Reference model is the foundation of the Framework for Network Visualization. It consists of three levels of feedback loop:

- *User's task to the data space;*
- *User's mental visualization process to the computational engines that manipulate the data; and*
- *The physical interfaces for communication with the computer.*

Military Utility

The VisTG Reference Model would not be used directly by military users, as it is a conceptual basis for developing Frameworks suitable for work in different specific areas.

Visualisation Framework

The VisTG Reference Model is the foundation for the Framework for Network Visualisation. It specifies six questions, of which the first asks what the user wants to perceive. A Framework developed from the Reference

Model will be likely to specify what kinds of thing are available to be perceived, and what kinds of actions can be taken with respect to them.

Technical Details

The VisTG Reference Model was first developed by IST-05, and is described in the Final Report of IST-013 (TR-IST-030). The Framework for Network Visualisation is an elaboration of the Reference model that develops the outer loop in cases where the data space represents a network.

Each loop of the VisTG Reference model represents many individual “elementary” feedback loops, each characterized by some condition of the computer side of the loop that the user wants to perceive. If the state is not in the desired condition, the user may act to bring its perception closer to the desired condition. This holds whether the state in question is the key to be struck by a finger at a keyboard in the innermost loop or in the outer loop some fundamental relationship among thousands of nodes in a network. In each case, if the state is not as desired, the user does something to correct the discrepancy.

To instantiate the VisTG reference model in any particular environment such as a complex network, it must be elaborated. The figure shows some of the scientific disciplines that may be involved.

For More Information, Contact:

Dr. Maurice Martin TAYLOR
Email: mmt@mmtaylor.net

2.3 MEASUREMENT AND CHARACTERIZATION OF DYNAMIC-LAYERED NETWORKS

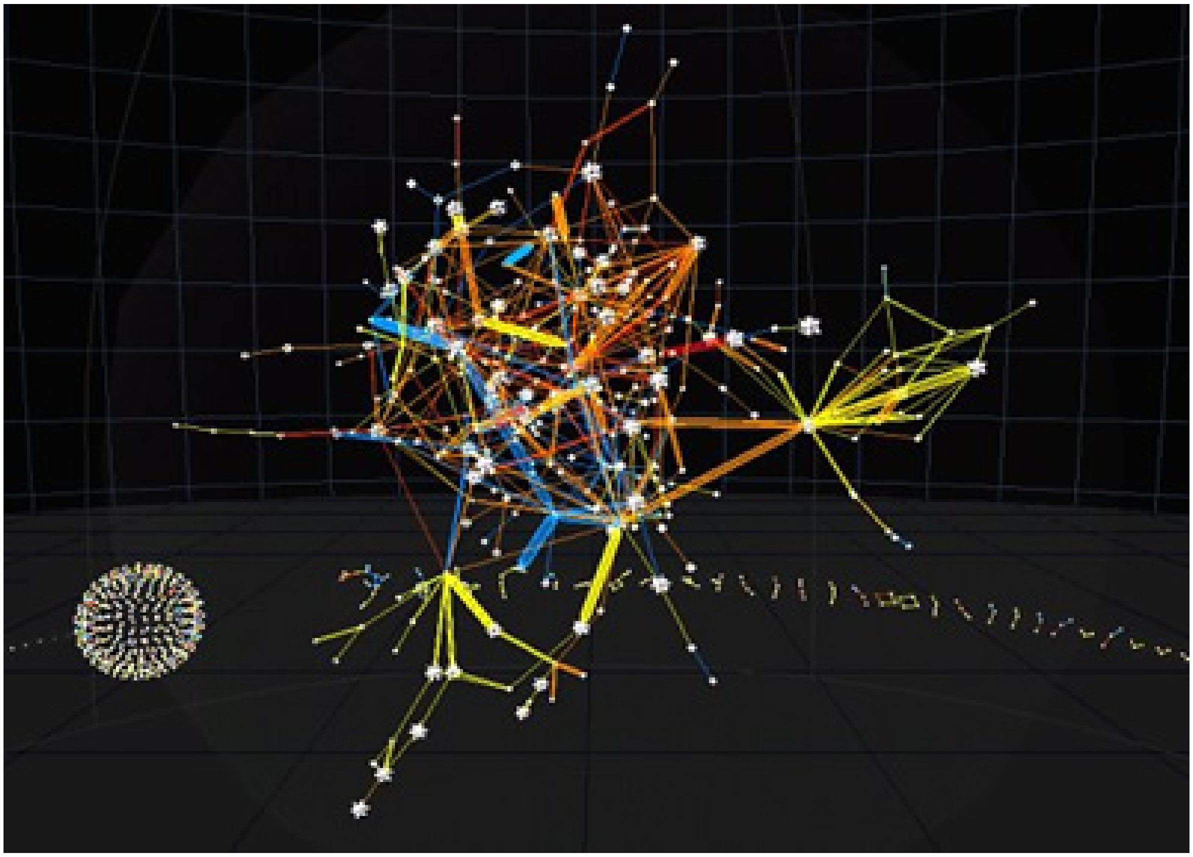


Figure 2-2: IST-085/RTG-041 are applying information theory in a new way for more accurate dynamic network analysis.

IST-085/RTG-041 are applying information theory in a new way. This effort is producing a means to measure and characterize multiple dynamic network types. Analyzing multiple networks as they change is a great way to increase accuracy and prediction; making the most of the data deluge.

Military Utility

To carry out the mission, you often need to understand the networks inherent to your area of operation. To understand the network environment, we must analyze interconnected stacks of many different and dynamic network types simultaneously as they change. Simultaneous multi-layered dynamic network analysis will not be possible until we understand the underlying "physics of networks" at the mathematical level.

Visualisation Framework

Bridging the gap between the problems a user wants to solve, and how to analyze and visualize the network data to solve that problem, requires a bit of translation in the middle, etc., a unified way of measuring and characterizing any kind of network. The framework provides the first step by helping the user to define the problem and the networks associated with it. These measures and characterizations are then the second step towards more effective visual analysis.

Technical Details

Similarity

How can we measure the similarity between two networks? You can choose a variable for any two networks and look at the difference between the two distributions. BUT there is no general "similarity". There is only similarity in regards to a variable. That requires a context of the user's goal or task.

Attributes

Do the attributes of a node mean anything when its nearest neighbours are in other layers? Maybe, but you may need to re-define the meaning based on your application.

How can we measure the attributes of a network? First identify how many variables you want to take with you, then sample each variable, then you have a distribution and you can apply Shannon to compute information-theoretic measures for each variable and for aggregations.

For More Information, Contact:

Dr. Amy Vanderbilt
Vanderbilt Consulting
DrV@DrAmyVanderbilt.com

2.4 IST-085 FRAMEWORK FOR NETWORK VISUALISATION

Figure 2-3: The Border Between East and West Germany During the Cold War. This bridge was a syntactic link, a semantic "bridge-type" link, but not a pragmatic link.

The IST-085 Framework is based on the principle that the results of the user's actions are more important than the actions themselves. It uses atomic-level descriptions of task needs, data sources, display and interaction capabilities, user training and the static and dynamic properties of networks to help users decide on the most appropriate techniques for solving their problems of the moment, and for researchers and developers to identify areas of potential future value.

Military Utility

Effective use of the Framework should allow the use of displays appropriate to the training status of operators performing different tasks. The different kinds of displays required for discovery tasks, "what-if" planning tasks, briefing tasks, etc., should be discoverable by using a combination of the Framework and knowledge of the currently available display and interaction techniques that fit the Framework specification.

Novel Concepts in the Framework

- The concept of embedding fields that support and the properties of real-world networks.
- Separation of syntactic, semantic, and pragmatic network properties.
- Fuzzy network properties, including negative fuzzy membership values dependent on the pragmatic embedding fields.
- Information-based dynamics and limitations.

Technical Details of the Framework

Visualization is one of two mutually supporting routes to understanding, the other being analysis. The Framework helps in developing displays useful for task-relevant visualization by providing a set of descriptive elements for user tasks and abilities, display and interaction techniques, and most particularly, network properties.

Visualization is taken to happen in the user’s head, not on a display surface. It can be supported through any sensory system, and by any method of representation, whether linguistic, symbolic, or pictorial.

For More Information, Contact:

Dr. Maurice Martin TAYLOR
Email: mmt@mmtaylor.net

2.5 EMBEDDING FIELDS

Figure 2-4: The regional geography is a pragmatic embedding field for the road network.

An embedding field places a network in its real-world context. Supporting networks, such as the Internet that supports the network of http links existing at any moment are syntactic or semantic embedding fields. Pragmatic embedding fields are not usually networks, but often are distributed environmental properties that affect the network.

Military Utility

Embedding fields both constrain and potentiate a network of interest. In analyzing a terrorist network, the pragmatic embedding fields of local political trends, geographic and temporal event patterns, the semantic embedding field of social relationships, and the syntactic embedding fields of contact possibilities and media all affect the implications of analytical aspects of the network itself.

Visualisation Framework

The concept of embedding fields is one of the novel aspects of networks introduced by the Framework. Embedding fields should be considered when developing or using visualisations of networks.

Technical Details

Networks are often analyzed as though they were mathematical abstractions, complete unto themselves. Networks in the real world, however, are supported by and constrained by the properties of the world in which they exist.

The network of messages between people is supported by transportation media, such as vehicles that carry paper letters, electronic links, or vibrations in the air. Such support structures can be abstracted as networks, whose properties determine the limits of possible properties for the components of the message network.

If only their interconnection patterns and weights are considered, the supporting networks are syntactic embedding fields for the supported network. If semantic properties of the supporting networks are considered, they are semantic embedding fields.

Pragmatic embedding fields are (usually) not themselves networks, but states of the world, such as the topography and ecology of the region containing a road network, or the political relationships among the entities taken to be network nodes.

For More Information, Contact:

Dr. Maurice Martin TAYLOR
Email: mmt@mmtaylor.net

2.6 FUZZY NETWORKS

Figure 2-5: Two Bridges with Different Fuzzy Membership in Class "Link" of a Road Network or a Hiking Network.

In the real world it may not be clear whether a connection between two entities is a link, or whether the entities are nodes. The degree to which a connection is perceived to be a link is its fuzzy membership value. Classical networks are fuzzy networks in which all the membership values of entities in classes "link" or "node" are 1.0 or zero.

Military Utility

The assignment of fuzzy membership values to network nodes and links paradoxically allows for increased precision in the description of networks and network behaviour. Fuzzy membership values depend on the uses of the network in the real world, and so require the analyst to incorporate pragmatic situations into the analysis. In many cases, the results will be more reliably related to the real-world situation than for a crisp network.

Visualisation Framework

Fuzzy network description is a part of the general descriptive framework for fuzzy networks, and is an inherent aspect of the Framework for Dynamic Network Visualisation.

Technical Details

Consider the two bridges in the picture. The bridge on the left (over the French River in Ontario, Canada) has gravel paths through forest at its ends, whereas the multi-lane toll bridge on the right (the Forth Road Bridge in Scotland) is connected to a superhighway network. Each is capable of carrying a car across a river, but the bridge on the left is much less a part of a road network than is the Forth Bridge.

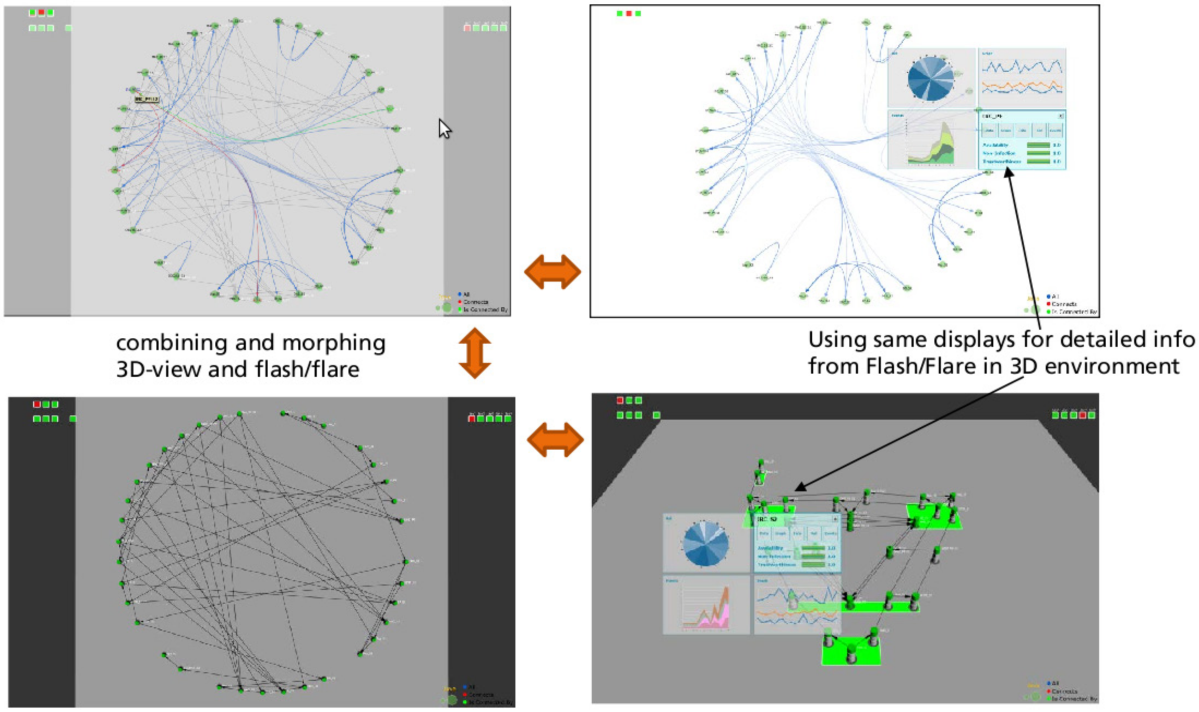
Consider the bridges from the point of view of a hiker. Each is equally capable of carrying pedestrians across the river, but the Forth Bridge crossing with its connecting roads is much less pleasant than is the French River bridge. From the viewpoint of the hiker, the French River bridge is much more of a link in the transportation network than is the Forth Bridge.

Each bridge could potentially perform the function of a link in the car-driver's network or the hiker's network, but they do so to different degrees, represented by their fuzzy membership values in the class "link" of the relevant networks.

For More Information, Contact:

Dr. Maurice Martin TAYLOR
Email: mmt@mmtaylor.net

2.7 COGNITIVE ANALYSIS OF DISPLAY ELEMENTS FOR NETWORK VISUALISATIONS



combining and morphing
3D-view and flash/flare

Using same displays for detailed info
from Flash/Flare in 3D environment

Figure 2-6: The perceptive process of using such visualization can be analyzed to create a computable cognitive model.

An analysis and a subsequent computational model of cognitive processes in the usage of different display elements for network visualization allows the understanding, verification and prediction of their utility within different tasks. By modelling the perceptive process of the user visual load, reaction times and other measures can be predicted.

Military Utility

Applying the method of cognitive analysis and modelling either in the design process of network visualization or in a later evaluation of an existing visualization allows an assessment of its utility. For example: a network visualization of a threat network is only of utility if the analyst is able to extract the encoded information quickly and precisely.

Visualization Framework

The framework allows the selection of different visualization techniques and tools based on a questionnaire. The cognitive analysis can be seen as an extension of the framework in that it employs analysis techniques to further gain an understanding of the impact of using a specific visualization technique.

Technical Details

Before a computable cognitive model can be created a task analysis, e.g. by using the GOMS (Goals, Operators, Methods, and Selections) method or constructing a sequence diagram, must be performed. The detail of such an analysis determines the detail of the later model. Times, resources or HCI principles can then be assigned to the identified sub-tasks.

A very simple computational cognitive model can be based on the perceptive and cognitive steps carried out by the user during the analysis of one specific diagram. Earlier research has determined certain HCI principles. One example is the uncertainty principle, with which the required time of choosing one of multiple options can be modelled. Another principle is the limitation of humans in the transfer and comparison of spatial relations within a diagram. By assigning probability distributions to steps or choices in such a sequence diagram the complete perceptive process can be described in a computable analytical model which can either be analytically solved or used in a simulation for prediction and evaluation.

For More Information, Contact:

Mr. Michael Kleiber, Mr. Carsten Winkelholz
Fraunhofer FKIE, Germany
michael.kleiber@fkie.fraunhofer.de / carsten.winkelholz@fkie.fraunhofer.de

2.8 INFORMATION THEORY OF NETWORKS

Figure 2-7: Networks are structured. Information recovers structure. Some of the structure may be easily seen, some may not. (Picture of the Eiffel Tower, Paris, France)

Information Theory is a quantitative and dynamic theory of uncertainty and its complement, information. Uncertainty is always that of a potential observer at some point about something, and depends on the prior knowledge of the observer. Information gain is reduction of uncertainty as a consequence of continued observation.

Military Utility

Information theory in itself is not of great utility to the operator. Its value is in the organization of techniques for analysis and display under different conditions. On the other hand, an understanding of the conceptual background of information theory can be helpful in practical situations, by helping the operator to recognize possibilities and limitations inherent in tasks when particular tools are available.

Visualisation Framework

Information Theory is one of the formal underpinnings of the Framework, most significantly in analyzing the dynamics of networks and the display of temporally variable network properties and traffic.

Technical Details

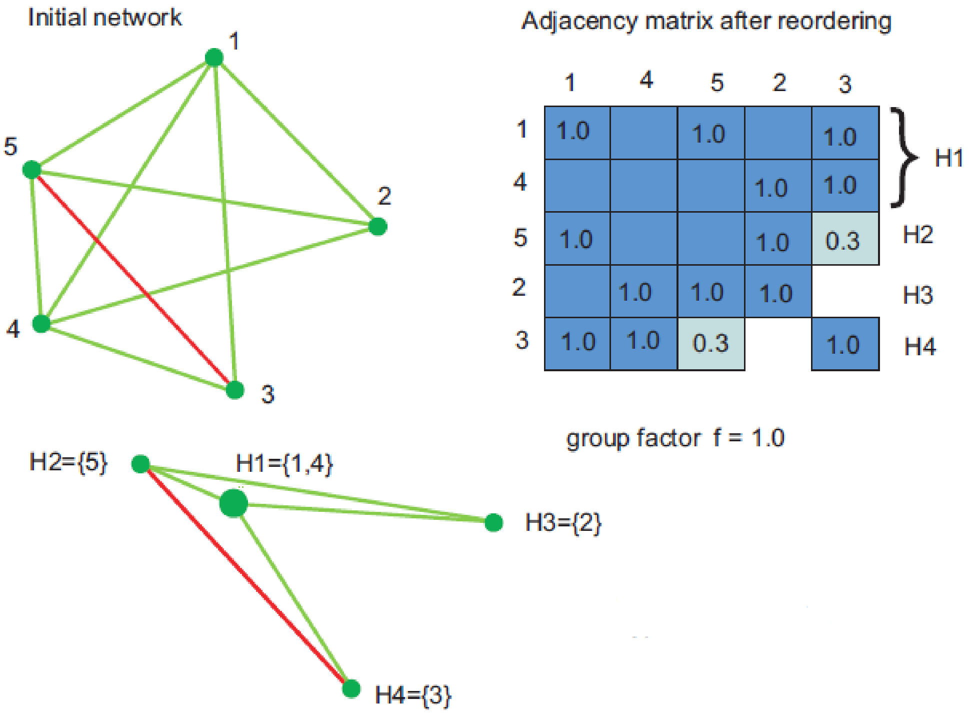
The central concept of Information Theory (IT) is uncertainty. A potential observer is uncertain about the actual state of a “target”; the degree of uncertainty is a function of the probabilities that the potential observer assigns to the possible states of the target. The “observer” might be as simple as a node in a network, and the target a node somewhere else in the network, or as complex as a battlefield commander with the complex dispositions and movements of friendly and enemy forces as the target.

“Information” is reduction of the observer’s uncertainty about the target. Information can flow through “channels” that have a capacity measured in bits/second and have some transport lag. The transport lag and channel capacity determine the degree to which changes at one point in the network can be recognized at another point, and therefore the precision with which a controller at one place in the network can control something at another place.

For More Information, Contact:

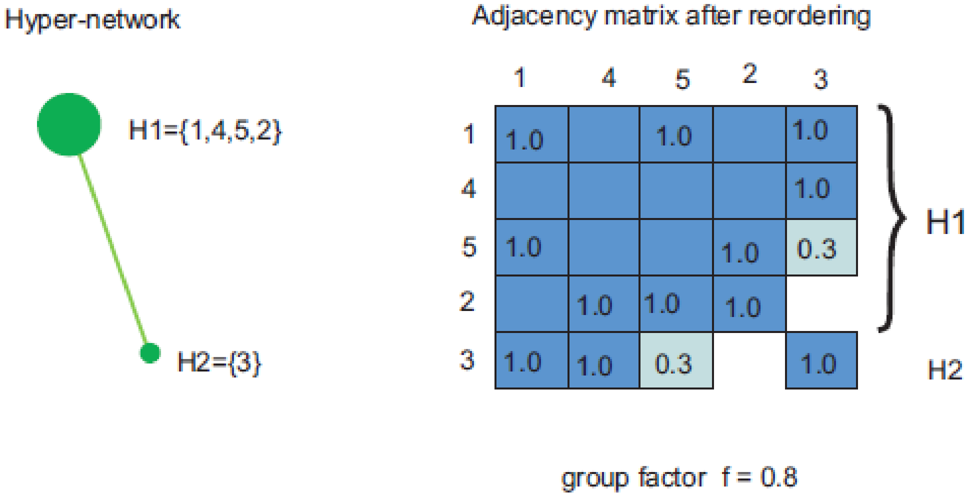
Dr. Maurice Martin TAYLOR
Email: mmt@mmtaylor.net

2.9 UNCERTAINTY



	1	4	5	2	3	
1	1.0		1.0		1.0	H1
4				1.0	1.0	
5	1.0			1.0	0.3	H2
2		1.0	1.0	1.0		H3
3	1.0	1.0	0.3		1.0	H4

Figure 2-8: A Weighted Network and its Adjacency Matrix and its Network of Hypernodes – the Group Factor *f* is Set to 1.0. The green nodes and links have membership value $\mu(r_{i,j}) = 1.0$, yellow when $\mu(r_{i,j}) < 0.8 > 0.4$ and the red link $\mu(r_{i,j}) < 0.4$.



	1	4	5	2	3	
1	1.0		1.0		1.0	H1
4					1.0	
5	1.0			1.0	0.3	
2		1.0	1.0	1.0		
3	1.0	1.0	0.3		1.0	H2

Figure 2-9: A Weighted Network and its Adjacency Matrix and its Network of Hypernodes – the Group Factor *f* is Set to 0.8. The green nodes and links have membership value $\mu(r_{i,j}) > 0.8$, yellow when $\mu(r_{i,j}) < 0.8 > 0.4$, and the red link $\mu(r_{i,j}) < 0.4$.

Graphs are often used in the modelling of real-world networks, the representation of abstract concepts, etc. One of the key issues for the visualization of real networks is that uncertainty is an aspect that must be considered. Uncertainty is inherent in much data/information and this poses a significant challenge in the analysis and visualization of networks. As the number of nodes and links in a network increases, compounded with uncertainty, the representation of the network needs simplification in order to provide clarity/tractability of representation for the user. The simplification must also taking into account the propagation of, and the degree of, uncertainty. Uncertainty will have a significant impact on the effective topological structure of the network. The certainty-based Hypernode algorithm has thus extended to handle uncertain relationships.

Military Utility

There is a need to:

- Simplify to trackable levels the complexity of large and complex network without losing the visual clarity of the network structure and underlying characteristics;
- Detect/identify the uncertainties in a network;
- Measure and represent uncertainties;
- Know how uncertainties propagate and interact in a network; and
  - Represent, visualise and manage uncertainties.

Visualisation Framework

The framework provides the selection of different visualization techniques and tools based on the users' needs. When the users are faced with large and complex network with uncertainties, Hypernode provides a possible means to analyze and understand the complexity of the network at different granularity. It also provides a framework in which to deal with uncertainties within the network.

Technical Details

There are two types of uncertainties in a network, namely uncertainty in the nodes and uncertainties in the links/edges.

Uncertainties in network data can, for example, be represented by a weighted network see Figure 2-8 a traffic light scheme is used here. The green nodes and links have membership value $\mu(r_{i,j}) = 1$, yellow when $\mu(r_{i,j}) < 1$ and > 0.4 and red when $\mu(r_{i,j}) < 0.4$, i.e. representing respectively reducing certainty.

In the Hypernode algorithm a 'group factor', *f,* is used to define the allowed/required degree of clustering together of the nodes in the hypernode creation process; in this example *f* is set to 1. The resulting hypernodes show that due to the weak link between nodes 3 and 5, node 5 was not aggregated to hypernode H1, see Figure 2-8.

In Figure 2-9, the group factor is set to 0.8. The role of H1 changes due to the 'uncertain' link between Nodes 3 and 5, i.e. it is no longer the most connected node. It can thus be seen that the uncertainty has a clear impact on the effective topological structure of the network.

The topological structure and relationships are depicted very clearly in this hypernode representation; indeed H2 is the most connected hypernode in this example.

The examples show how ‘uncertainty’ in network data can be visualized by the application of colour coding. The group factor *f* determines the structure of the hypernode and must be set with careful consideration in respect to the application in question. Figure 2-9 shows how radically the hypernode structure can be changed by reducing the group factor from 1.0 to 0.8.

For More Information, Contact:

Prof. Jan Terje Bjorke (†)
Dr. Margaret J. Varga
Email: m.varga@signal.QinetiQ.com

2.10 HYPERNODE

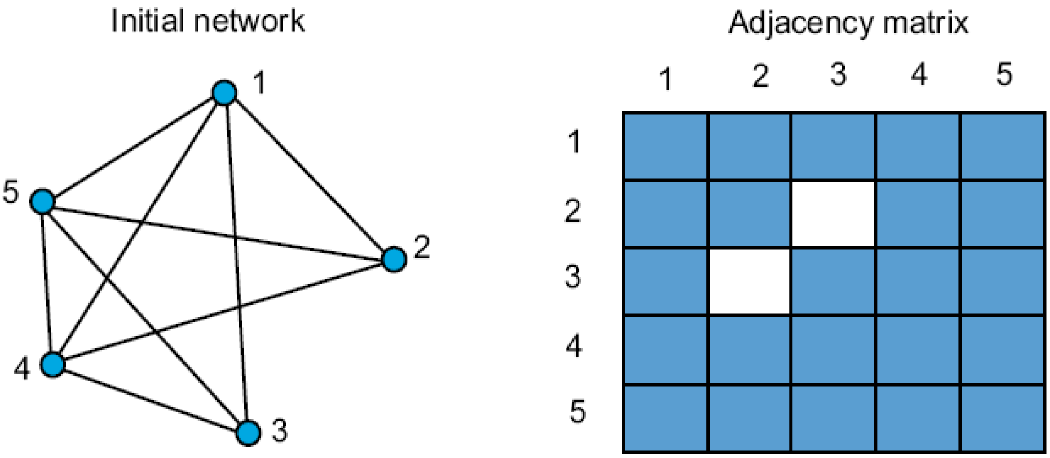


Figure 2-10: A Network and its Adjacency Matrix.

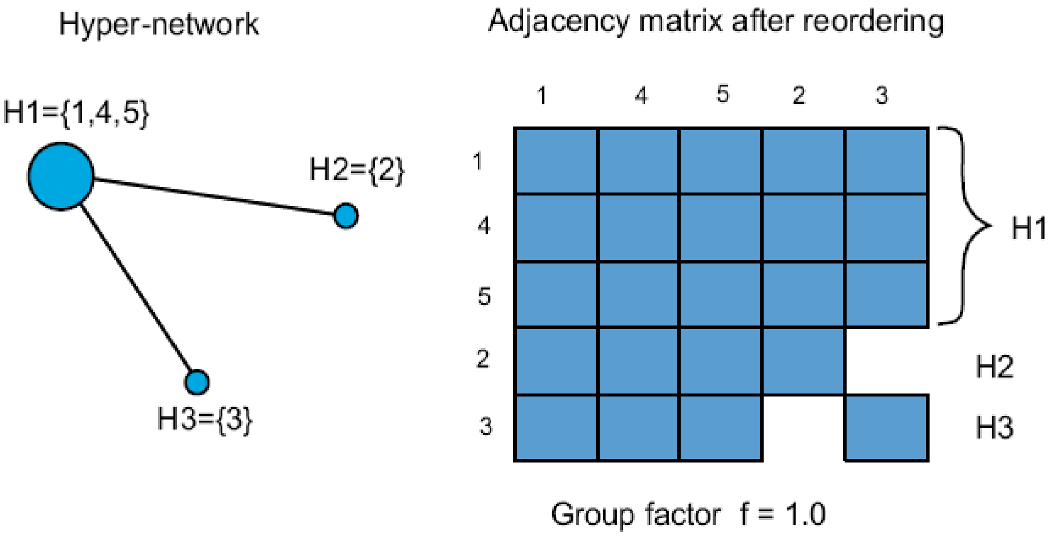


Figure 2-11: Hypernode and Re-Ordered Adjacency Matrix.

Much intelligence and military data are in the form of networks. The reduction of complexity is an important element in the visualization, understanding and management of networks with large numbers of nodes and links. The Hypernode technique was developed to abstract complex networks. It is based on Information theory to aggregate nodes and links into 'hyper-nodes' and 'hyper-links' by reordering the adjacency matrix to generate hierarchies of hyper-networks. It provides an effective means to reduce visual overload and thus simplify and enhance the understanding and analysis of complex networks.

Military Utility

Visualisation is a mental process, for example:

- **Military Commander:** Visualises relationships among his forces, between his forces and enemy, between forces and landscape, between events at one time and events at another.
- **Intelligence Analysts:** Assess potential threats (e.g. terrorist) by visualising the development and structure of networks of malicious groups; defenders can visualise the likely effects of specific interventions and non-intervention.

Hypernode provides the capability to:

- Abstract a large and complex weighted network into a significantly smaller and simpler network;
- Enable networks to be visualised at different levels of abstraction;
- Enable the user to zoom in and out to different levels of detail/abstraction;
- Transform a flat network into a hierarchical structure; and
  - Reveal and highlight the underlying structure/pattern of the network.

Technical Details

The adjacency matrix of a network can be mapped to a grid so that similar rows, or similar columns, are clustered. Figure 2-10 and Figure 2-11 demonstrate how re-ordering can be used to generate a view of the adjacency matrices from where groups of nodes can be derived. From the re-ordered matrix, three groups of nodes can be derived, i.e. the hypernodes H1, H2 and H3, as shown in Figure 2-11.

A similarity measure is used to group the nodes, i.e. a group factor. An index on the scale between [0,1] is used to measure the minimum similarity for rows in a group. If the group factor is 1, the rows must be completely the same in order to be mapped to the same group. Figure 2-11 shows the result of using group factor1. As the group factor determines the aggregation of the nodes, it is user defined to enable flexible exploration of the network.

From the network of Hypernodes, it can be observed that the initial network in Figure 2-10 can be regarded as a tree structure, with one mother node and two leaf nodes. The mother node H1 is composed of three sub-nodes (1, 4 and 5) which have strong connection, i.e. one node is connected to the other two. Since there is no direct connection between nodes 2 and 3, hypernodes H2 and H3 have no direct link. The network in Figure 2-11 is constructed by clustering. The size of the node is proportional to its connectivity. The thickness of the link is proportional to the number of sub-links of a hyper-link.

Visualization Framework

The framework provides the selection of different visualisation techniques and tools based on the users' needs. When the users are faced with large and complex network, Hypernode provides a possible and effective means to support the analysis and understanding of large and complex networks at different granularities.

For More Information, Contact:

Prof. Jan Terje Bjorke (†)
Dr. Margaret J. Varga
Email: m.varga@signal.QinetiQ.com

2.11 INTERACTIVE HYPOTHESIS VISUALISATION

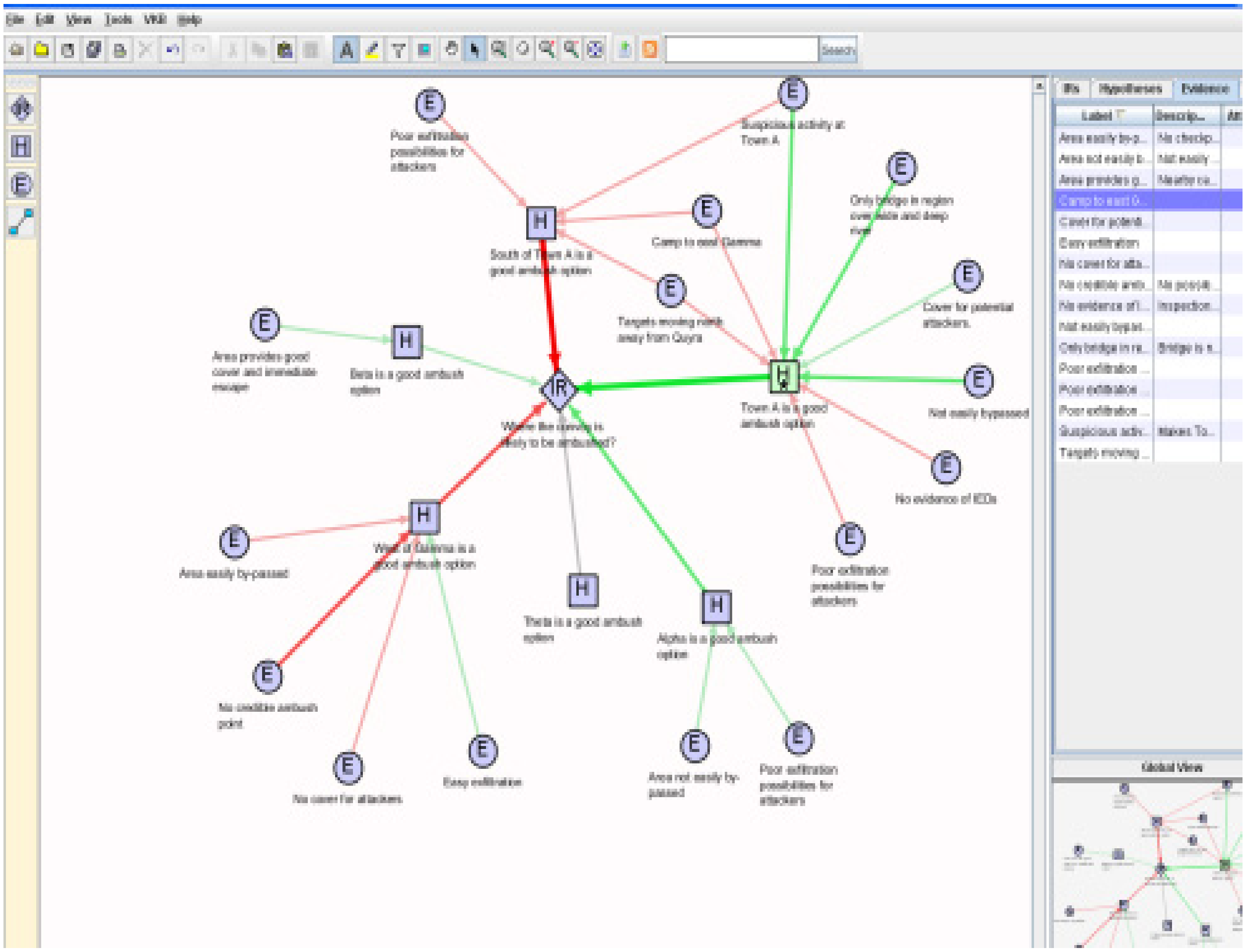


Figure 2-12: Hypothesis Visualization.

Hypothesis visualization is concerned with supporting the construction of an argument; one of its key strengths is easily showing the balance of argument. It is concerned with showing and managing information items of relevance and significance, the temporal development of facts and arguments, and how items of information support or refute alternative deductions. It is important to manage and represent information uncertainty and/or credibility, as well as duplicate, redundant and implausible items of data/information.

Military Utility

There are two main scenarios of hypothesis generation. In the first scenario, a hypothesis can be proposed and evidence is then gathered, in the second scenario a hypothesis is generated on the basis of gathered evidence.

The formation of hypotheses is supported by different sources of evidence and the strength of evidence influences the outcome of the hypotheses.

Analysts therefore need to have tools that enable interactive dynamic hypothesis formulation, evaluation and visualisation for:

- Handling multiple hypotheses;
- Assessing data/information sources; and
- Integrating uncertainty.

Visualization Framework

The framework provides the selection of different visualization techniques and tools based on the users' needs. When the users are faced with large and complex network of evidence with uncertainties, hypothesis visualisation provides a possible means to generate dynamically possible hypotheses with the evidence to support or refute them accordingly

Technical Details

The user can interactively analyse and visualise the formation of hypotheses and their corresponding evidence, see Figure 2-12.

The Information Request (IR) poses a question to the analysts. The Analysts then generate Hypotheses (H) and their associated Evidence (E) to support or refute the hypothesis in question.

The system supports the management of the uncertainties inherent in the evidence and their role by assessing the:

- Reliability of the source;
- Value and relevance of the information from the source; and
    - Degree the evidence supports or refutes a hypothesis.

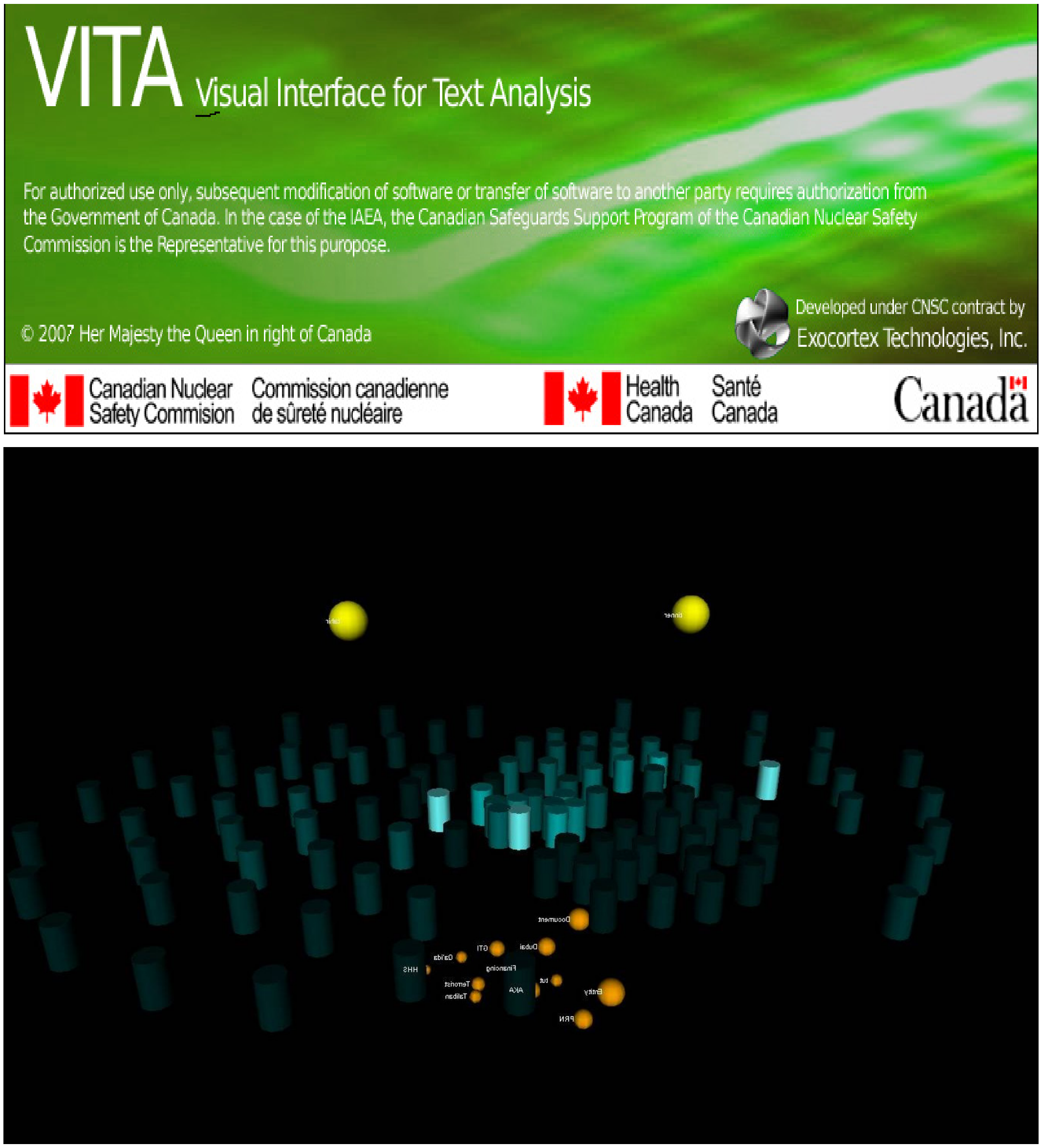
The system automatically identifies the most likely hypothesis among all plausible hypotheses based on the available evidence and in the presence of uncertainty, i.e. the one with the thickest green line in Figure 2-12.

The system updates the hypothesis or hypotheses either when new evidence becomes available and/or new hypotheses are generated so that the *current* most likely hypothesis is identified.

For More Information, Contact:

Dr. Maurice Martin TAYLOR
Email: mmt@mmtaylor.net

2.12 VISUAL INTERFACE FOR TEXT ANALYSIS (VITA)



VITA Visual Interface for Text Analysis

For authorized use only, subsequent modification of software or transfer of software to another party requires authorization from the Government of Canada. In the case of the IAEA, the Canadian Safeguards Support Program of the Canadian Nuclear Safety Commission is the Representative for this puropose.

© 2007 Her Majesty the Queen in right of Canada

Developed under CNSC contract by
Exocortex Technologies, Inc.

Canadian Nuclear Safety Commision
Commission canadienne de sûreté nucléaire
Health Canada
Santé Canada
Canadä

Figure 2-13: VITA Screen Taken from a Counter Terror-Related Search in an Intelligence Database.

VITA is a visual analytic tool fielded to assist an analyst or executive to understand and manipulate the contents in a large text database, as they relate to his/her immediate interest. It aims to create an intuitive grasp in the user, of the concepts in the text dataset, as well as their contexts and internal relations. The user enters his/her search terms, chooses among several available search engines and specifies one or more target text corpora. A final search through the returns is generated by the software as soon as the initial searches are complete; this serves to discover the new significant concepts, terms and topics of likely interest that also appear in the hits. This will suggest important new avenues of enquiry.

Military Example – Nuclear Intelligence

The search shown above might begin with a technology of target interest, e.g. nuclear centrifuge development; those terms would appear as yellow balls. The hits [cylinders] are then searched for clusters of informative terms [orange balls] inferring the related topics related in the text corpora.

For example, if the topic of target interest were nuclear centrifuge, the inferred terms likely would include the names of scientists and engineers working on the technology whose names appear in the documents discovered.

Technical Details

VITA is a meta-search engine, programmed in $C^{\#}$, using the users' input to drive external or internal search engines – Yahoo and Bing for the Web – with dtSearch and Google Desktop currently available for private or secure text bases.

Discussion

VITA [Visual Interface for Text Analysis] or VisualTA] is a meta search engine fielded for use in intelligence and health-related applications. The interface joins the analyst to the corpora through a series of standard search engines. Analysts find this tool highly useful in allowing them to see the relations and content of large text corpora without a need to read each article within the corpus.

A VITA user can target a search to an internal or external or mixed data corpus for keywords related to a query, and the documents discovered are rendered in a 3D space, grouped by the search terms. The user is then able to relate and select groups of the returns and manipulate those in a variety of ways. That utility alone makes it a useful means to organize, understand and interact with text corpora. However and beyond this, a search engine internal to VITA scans the returns and extracts other discovered informative terms. This last facility shows the user the related concepts and terms in the literature; as such it helps users to discover fruitful avenues of researcher he might have missed otherwise:

- Supports popular search engines;
- Ability to search the Internet, a specific website or private data bases;
- Examine views and explore text relationships not apparent with other tools;
- Automatic clustering;
- Ability to save and print searches;
- Provides a visual insight to how documents in a dataset are related; and
    - Quick to set up and very easy to use.

Trial Download: Visual TA Evaluation Installer

For More Information, Contact:

Dr. Zachary Jacobson, CA
Zachary.Jacobson@hc-sc.gc.ca / jzj3@rogers.com

2.13 JUNCTION MODEL FOR VISUALISING NETWORK INTERDEPENDENCIES

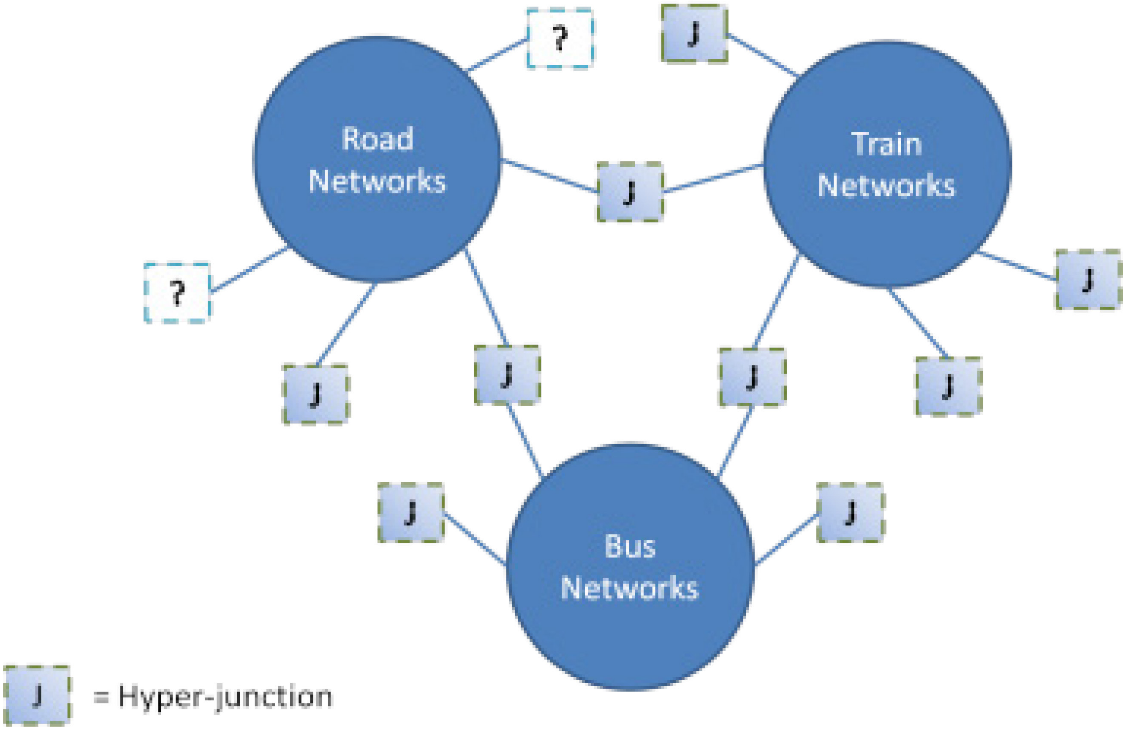


Figure 2-14: High-Level Network Interdependencies are Represented by Hyper-Junctions.

Being able to represent network interdependencies is beneficial for predicting non-obvious side effects on inter-related networks. A junction model was developed to isolate the interdependencies from the network model. Potential visualizations of the junctions' nature and effects are also proposed.

Military Utility

The operational context of any mission involves a number of networks that may be inter-related. When planning military operations, one has to assess the influence it may have in the operational environment. Although the direct effects on a network may be clear, side effects on related networks may not be obvious. Being able to visualize the interdependencies between networks would be beneficial for predicting the impact of potential operations.

Visualisation Framework

The Junction Model provides a mechanism to display complex interrelationships providing a construct for facilitating visualization. Through this model the Visualisation Framework has a new tool to support analysis through customizations of interconnected parameters.

Technical Details

Network Interdependencies Challenges

What kind of tasks involved analyzing network interdependencies? Finding cascading effects such as the implications of the perturbations across networks as compared to steady state, prediction / "what if" analysis, alerting, exploring connections between disparate networks, inferring links / discovering connections.

What is the main visualization challenges associated with network interdependencies? Four important challenges related to interdependencies visualization were identified:

- Incomplete information about interdependencies;
- Complexity of interdependencies;
- Uncertainty of interdependencies; and
  - Temporal aspects of interdependencies.

For More Information, Contact:

Dr. Valérie Lavigne
Email: valelavi@gmail.com

Chapter 3 – WORKSHOPS

3.1 PENN STATE UNIVERSITY

Report: Visualization Network-of-Experts

10th Workshop: Understanding Dynamic Networks – Visualization Research, Theory and Practice

6-8 October 2009

The Nittany Lion Inn, 200 West Park Avenue, State College, PA 16803-3598, USA

Dynamically evolving networks are ever-present in a wide variety of complex systems and there are pressing interests in understanding the evolution of such complexes, e.g. small world networks and scale-free networks.

With massive empirical data sets from strategic and military intelligence sources, the public safety domain, criminal investigations, economics and social sciences, there comes the challenge of how to display key graph/ network measures, attributes and uncertainties. The Workshop focused on discussing and exploring a wide range of visualization and visual analytic techniques that are applied to detecting, tracking and predicting change in evolving networks – tasks which are critical to defence, intelligence, health, law enforcement and public safety. It brought together international experts to share their work and experience. Dr. Vanderbilt's keynote speech on the future challenges in visualization and analysis set the direction of the Workshop. This was followed by presentations on multi-future, multi-perspective network simulations in support of operations analysis and planning, role of visualization in human centred information fusion as well as evaluation of network visualization tools for intelligence applications. The workshop also discussed entropy, uncertainties, sonification and stereoscopic displays. Different application domains ranging from defence, forensic science as well as legal and medical were included. Working groups were formed for in-depth discussions on Information theory and experimentation and research strategy for visualizing networks dynamics. It was a very successful Workshop which stimulated fruitful discussions and the gaining of some insight in the understanding of dynamic networks.

Dr. Margaret Varga
UK National Leader

3.2 VANCOUVER, CANADA

Report: Visualization Network of Experts

11th Workshop on Visual Analytics for Network Health

29 November – 1 December 2011

Health Canada, Vancouver, Canada

Dynamically evolving networks are ever-present in a wide variety of applications, such as those in the defence, security, biological, medical, healthcare and social domains. There are pressing interests and needs in understanding and maintaining the health of these networks. The health of a network encompasses its functionality, capability, stability, and robustness to intrusions, as well as its ability to restructure due to contraction or expansion. Visual analytics provides a means to understand and analyse the network's health and the implications when it is undergoing expected or unexpected changes.

This Workshop brought together operational users, developers and researchers to explore how visual analytics technologies support network health assessment and management for military and civil applications. Application domains include information assurance, healthcare, infectious disease management, cyber warfare, terrorism, along with peace-keeping and peace-support operations.

A core objective of the Workshop was to have users interact with developers and researchers. The Workshop stressed the importance of a multi-disciplinary approach as both human factors and technological innovation must collaborate to improve visual analytics systems.

The Workshop had three inter-related sub thrusts:

- Social Networks;
- Physical Networks; and
- Logical Networks.

Of particular interest were military and civil applications and problems whose dynamic nature poses threats to the health of the networks.

Prof. Brian Fisher opened the workshop with a very stimulating presentation on a cognitive system approach to visual analytics. This was followed by a session on challenges and strategies for counter-insurgency visualization as well as issues on data sources and healthcare applications. One of the areas of interest was sense making, i.e. how to make sense of the complex data so as to explore, exploit, understand, analyse and visualize the data.

The Workshop also had a session on the work of the international visual analytics groups which included the presentation from the:

1) Canadian Visual Analytics Consortium (CANVAC);
2) The US National Visualization and Analytics Centre (NVAC);
3) The UK Visual Analytics Consortium (UKVAC) as well as; and
4) The NATO Visualization Research Task Group (IST85/RTG41).

It was very informative and showed that the various groups share common interests in some areas and complement each other in other areas. It provided an invaluable opportunity to discuss how the various groups could share information and collaborate.

The Workshop also had demonstrations to show the state-of-the-art visual analytics technologies in various application domains. The demonstrations provided an opportunity for the attendees to not only see how the systems/prototypes/tools work but also the opportunity to use them. The demonstrations included:

- CZsaw 2.0;
- Scalable Reasoning System;
- IN-SPIRE;
- Analyst Notebook;

- Collaborative Analytics Environment;
- GreenGraph Suite;
- 3D Cough simulator;
- Starlight; and
- Symptom tracker.

It was a very successful Workshop addressing the challenges in visualising and analysing the health of dynamic networks.

Dr. Margaret Varga
UK National Leader

Annex A – VITA

We had a meeting in 1993 in Ottawa featuring the best in that year's data visualizations, presenting them to the clients of the Government's internal consulting office. Prior to that, my own clients in DND had petitioned NATO for an Exploratory Group to meet on the same topic. The techniques were billed as visualizing non-visual data. It was clear to me that numeric data visualization – letting an analyst see at a glance what a complex set of quantitative variants was doing – was a set of technical challenges but less challenging conceptually; the big enchilada of a problem was going to be text visualization – answering the same question for text. I threw that down as a challenge to the attendees at that meeting. No one took it up. A few years later and I was still waiting.

There's not a lot to write, VITA speaks for itself.

VITA emerged from my request, asked of Els Goyette and Ben Houston, for "something that will show the relations among documents in the user's private database at the level of ideas." This was in 1997. It was to be an intelligence tool, able to deal with the text flood. Other specs included that it would need to be transparent to whether or not the database or the user had appended metadata, and it would need to be able to handle huge datasets of the petabyte size that we forecasted would become commonplace in Intel operations.

Els spotted that it needed to use a search engine. A search engine would already have indexed the texts according to its own imposed metadata and would even then be ready to handle the extreme scalability issues that were to come. Els wanted a cognitive search engine but after a few false starts on commercially available cognitive search engines I chose a 3D display that Ben [then my summer intern] was working on. "Is this one okay?" "Yes! That's the one we've been here all summer trying to get."

We worked with the best of the available cognitive search engines but they proved to be too expensive and not sufficiently scalable for our budget and use.

Ben built his own search engine driver for the then-current most powerful available search engines – Google and Alta Vista. He adapted the display from one of the two most successful displays he had devised the previous summer and the gravity from another. The prototype worked better than hoped but funding to move on did not develop. But there the project languished for several years.

Ultimately, in 2004, CNSC took me to show the then current version of VITA [adopted from the original by Ollie Dagenais] to IAEA as an Intel tool. They encouraged CNSC to commission further development. We recruited Ben back into the project. He added a secondary search, to extract the informative frequent terms [the orange balls in the display] in the hits. This greatly extended VITA's power and in 2007 became the tool I had first asked for – to show the relations in a client's dataset that showed the important common threads at the level of ideas.

IAEA have adopted it for to monitor nuclear traffic. Exocortex markets it under the name Visual-TA.

Dr. Zachary JACOBSON
Health Canada Applied Research and Analysis Directorate
Jeanne Mance Building, Floor 5
Tunney's Pasture, Postal Locator 1905C
Ottawa, Ontario K1A 0K9
Canada
Email: zack@sigmaxi.org

Annex B – APPROACH FOR A COGNITIVE ANALYSIS OF DISPLAY ELEMENTS IN NETWORK VISUALIZATIONS

Dr. Carsten Winkelholz, Mr. Michael Kleiber, Prof. Dr. Christopher Schlick
Fraunhofer-FKIE
Neuenahrer Str. 20
53343 Wachtberg
GERMANY

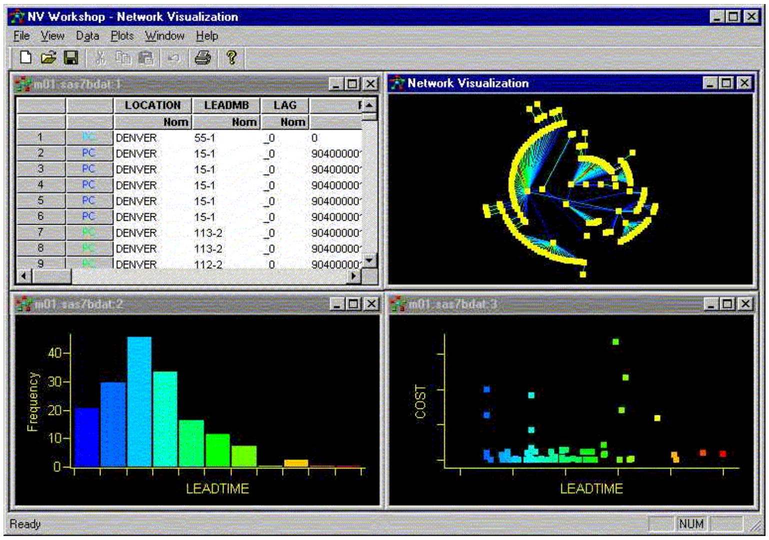
carsten.winkelholz@fkie.fraunhofer.de / michael.kleiber@fkie.fraunhofer.de /
christopher.schlick@fkie.fraunhofer.de

ABSTRACT

The proposed contribution will present a framework we developed to model human spatial information processing within a symbolic architecture of cognition. Within a sub-symbolic layer, spatial relations are represented as probability density functions modelling the uncertainty of their cognitive representation. Reasoning about spatial relations perceived from an external graphical display and their relation to internal cognitive spatial relations is reflected within the framework by the principle of Bayesian inference. In this way the cognitive processes of the human operator inspecting a graphical layout or diagram can be modelled and possible potential bottlenecks be identified.

B.1 INTRODUCTION

The application shown in Figure B-1 employs different kinds of display elements to graphically depict the underlying data.



		LOCATION	LEADMB	LAG	
		Nom	Nom	Nom	
1	PC	DENVER	55-1	_0	0
2	PC	DENVER	15-1	_0	90400000
3	PC	DENVER	15-1	_0	90400000
4	PC	DENVER	15-1	_0	90400000
5	PC	DENVER	15-1	_0	90400000
6	PC	DENVER	15-1	_0	90400000
7	PC	DENVER	113-2	_0	90400000
8	PC	DENVER	113-2	_0	90400000
9	PC	DENVER	112-2	0	90400000

Figure B-1: SAS Network Visualization which Employs Two Diagrams and One Graph (From http://www.sas.com/index.html).

All the used diagrams differ but they do depict the same or related data. Some diagrams require the data to be in a certain format but nonetheless there are still many possibilities to depict the same underlying data. The visualization should support the user to build his own internal model/representation of the data on which he does his reasoning process. To achieve this, the visualisation should be optimized in a way that all dimensions can not only be discriminated perceptually but also cognitively. Often the user wants to compare features by their dimension, their hue or saturation, their angle or relation to each other. When the user is exploring data it might be more appropriate to quickly flip through different kinds of visualizations in order to see a common trend or find an interesting outlier.

When flipping through different representations of the data the user needs to recall memorized content from other visualizations reliably and quickly to make comparisons with a currently perceived view of the data. We propose a perceptive and cognitive model in which visual features are perceived and encoded for later retrieval. The common visual features used in diagrams to encode numerical data are lengths, angles, areas or even volume. The cognitive model which will be introduced later allows modelling the perceptive and cognitive processes involved when analyzing diagrams by breaking this process down into basic steps. Breaking down the process into basic steps allows using a minimal set of characterizing constants when formulating the cognitive model.

In Figure B-2 different layouts for a tree are shown. The layout influences the spatial relation of how objects are related to each other. The graphical layouts are mostly used to present a user a map of items he is interested in. The hierarchical structure is determined by the semantic context and the underlying ontology of the data. The layout should highlight this hierarchical structure. However, there are different ways to do this and it is not clear what layout should be the best. For a map the layout should facilitate the user to remember fast and reliable locations of items. Humans encode locations in local reference systems, which mean that a location is encoded in relation to salient features in its environment. These allocentric reference systems are constituted by virtual connecting lines between features. The attributes that are memorized within these allocentric reference systems are angles and distances. In general the same is true for memorizing features constituted by data points within diagrams as shown in Figure B-3. Also many connecting lines are present within the diagram; these connecting lines only make specific allocentric reference systems salient. The theory proposed in the following is based on the idea that angles and distances are basic/atomic attributes memorized to internalize an external visualization.

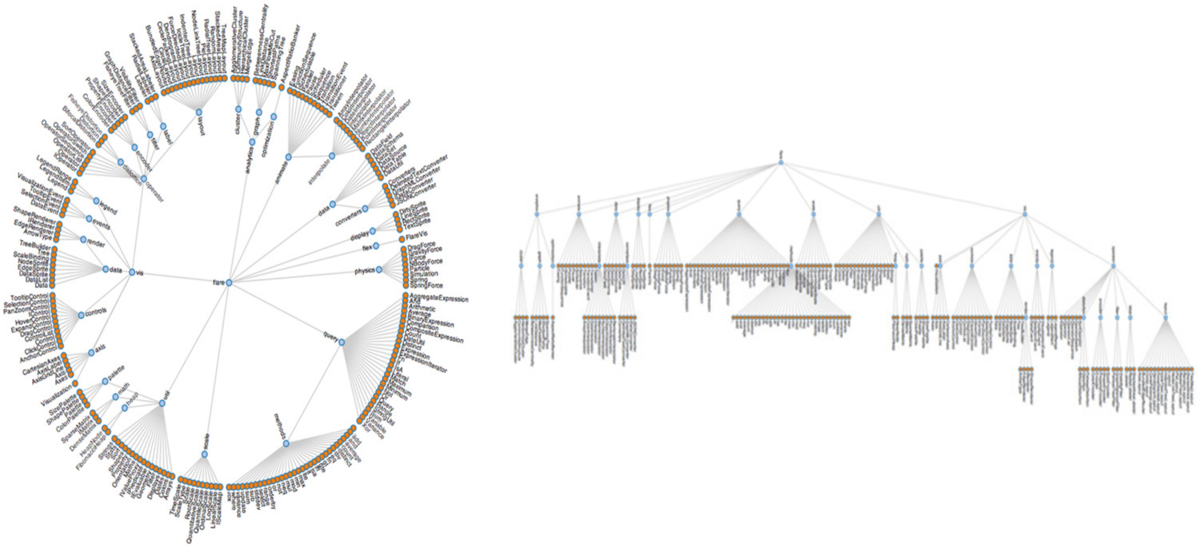


Figure B-2: Example for Different Graph Layouts (From http://hci.stanford.edu/jheer/files/zoo/).

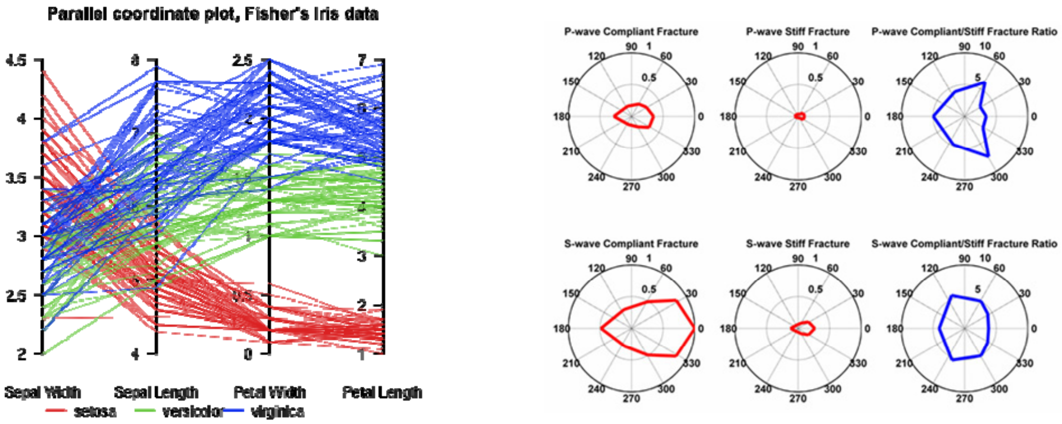


Figure B-3: Examples for Different Diagrams.

B.2 MODELING HUMAN SPATIAL COGNITION

Spatial perception and the geometry of visual space have been studied for more than a century. In the early days, scientists were interested in identifying a single geometry that maps Euclidean physical space into perceived visual space [1],[2]. Later, the focus shifted to investigating the memorial space [3]-[6]. Whereas detailed mathematical descriptions of the visual geometry have been developed for visual perception [2],[7]-[9], mostly normative models exist for cognitive maps. For instance, [6] proposed that locations are memorized in egocentric and allocentric coordinate systems. Allocentric coordinate systems define locations with respect to objects in the environment, which is consistent with the view that humans process spatial information in a mostly sequential fashion. The visual system scans and encodes a scene using a number of sequentially performed attention shifts, and the locations are recalled by sequentially retrieving spatial relations from memory that are related to the location in question. Therefore, it is reasonable to assume that spatial relations that shape the mental model are acquired during attention shifts. We believe that shifts of attention between several locations in space define the reference axes and planes of local allocentric coordinate systems within which the spatial relations are encoded. This assumption is consistent with the idea that locations are encoded by intrinsic frames of references [10]-[13]. These intrinsic reference frames would result naturally from salient landmarks of the scene that attract attention. It should be possible to obtain a mathematical description of these spatial relations within the allocentric reference systems. In general, locations are described mathematically by scalar values. Since the neural system can only represent noisy values of scalar dimensions, it is reasonable for the cognitive system to memorize a location in the dimensions of a coordinate system by storing the values that are most suitable for an expected task. If locations in the vicinity of a landmark location need to be encoded with greater precision, the dimensions of a spherical coordinate system are a rational choice. This assumption is consistent with the fact that humans think about visual space in terms of distances and angles. In the following we describe a framework that is able to integrate the numerical mathematical description of spatial relations into a symbolic architecture of cognition. We used this framework already to model phenomena related to memorizing object locations in graphical structures [14] and for symmetry detection [15].

B.2.1 Symbolic Architectures of Cognition

Symbolic architectures of cognition describe cognitive processes by distinct steps. The current state of the environment and the cognitive system activates a rule that describes the actions and the state transitions of the cognitive system. ACT-R [16] is a popular architecture of cognition. Its essential feature is a sub-symbolic layer which influences the activation processes of rules and the state transitions of the cognitive system. The values of the parameters of the sub-symbolic process are variable and might adapt during the cognitive process, which enables the cognitive system to learn. The most elaborated module of ACT-R is the memory module. The memory is assumed to be a collection of symbolic entities $d^{(i)}$ called chunks. The probability of a successful retrieval and the retrieval time are determined by the activation of the memory chunks, which is calculated by a formula taking into account the time spans in which a chunk has already been retrieved, the strength of its association with the current goal and the similarity of the attributes in a request to the values in the attributes of a memory chunk. This central equation of ACT-R is given by:

$$a_i(t) = b_i(t) + \sum_j w_j s_{ji} + \sum_k u_k m_{ki} \tag{1}$$

The base activation $b_i(t)$ decays logarithmically over time and increases each time the memory chunk is retrieved. The parameters s_{ji} reflect the frequency of how often chunk $d^{(i)}$ has been retrieved if the symbolic value of attribute ν_{jg} of the goal was identical to the current value. The parameter m_{ki} is the similarity parameter, and we think it could best be interpreted as the log-probability $m_{ki} = \ln(P(\nu_{kx} = \nu_{ki}))$ that the value in attribute ν_{kx} in the request x is identical to the value in the attribute ν_{ki} of the chunk $d^{(i)}$. Because in this case the ACT-R equation for the probability that a memory chunk will be retrieved factorizes to:

$$P_i \propto e^{b_i + \sum_k u_k \ln(P(\nu_{ki} = \nu_{kx}))} = e^{b_i} \prod_k e^{u_k} P(\nu_{ki} = \nu_{kx}) \tag{2}$$

The product in the factorized version expresses the probability that all attributes are equal to the attributes in the request. The parameters w_j and u_k are weighting factors reflecting the importance of a single attribute. To decide during a simulation which chunk will be retrieved from memory, noise is added to (1), and the random variable $A_i = a_i(t) + X + Y$ (2) is considered. The random variables X and Y are independent normal distributed with a mean of zero and a standard deviation σ_X; σ_Y. The value of the first random variable X is added when the chunk has been created. And the second one Y is added when $a_i(t)$ is re-evaluated. The memory chunk with the highest activation will be retrieved. If the activation of no memory chunk exceeds a threshold τ_a a failure will be retrieved. Because of the decaying of the base activation a memory chunk will be forgotten unless it is not frequently retrieved. The time needed for a successful recall also depends on the activation by the relation $t \sim e^{-A}$. The higher the activation, the faster a memory chunk can be recalled.

To model the variability of reproduced spatial relations from human memory such a sub-symbolic layer is also needed for a visual spatial module. Visual information is acquired by single shifts of attentions.

B.2.2 Reference Systems

The location of an object can only be identified within a frame of reference. In experimental psychology it is a well accepted procedure to divide the frames of references into two categories: An egocentric reference system, which specifies the location of an object with respect to the observer and an environmental (allocentric) reference system, which specifies the location of an object with respect to elements and features of the environment. According to Mou and McNamara [10] humans also use reference systems concerning the intrinsic axis of the object configuration. For example, two salient objects on a plane create an axis that is used to specify the location of other objects. The most natural way to integrate this into the concept of attention is to consider the last two attended objects as an axis of reference. However, creating object-location memory chunks in this "semi-allocentric" reference system is less effort to the visual module because it only needs to keep track of two objects, whereas in the case of the pure allocentric reference system three objects are needed (Figure B-4).

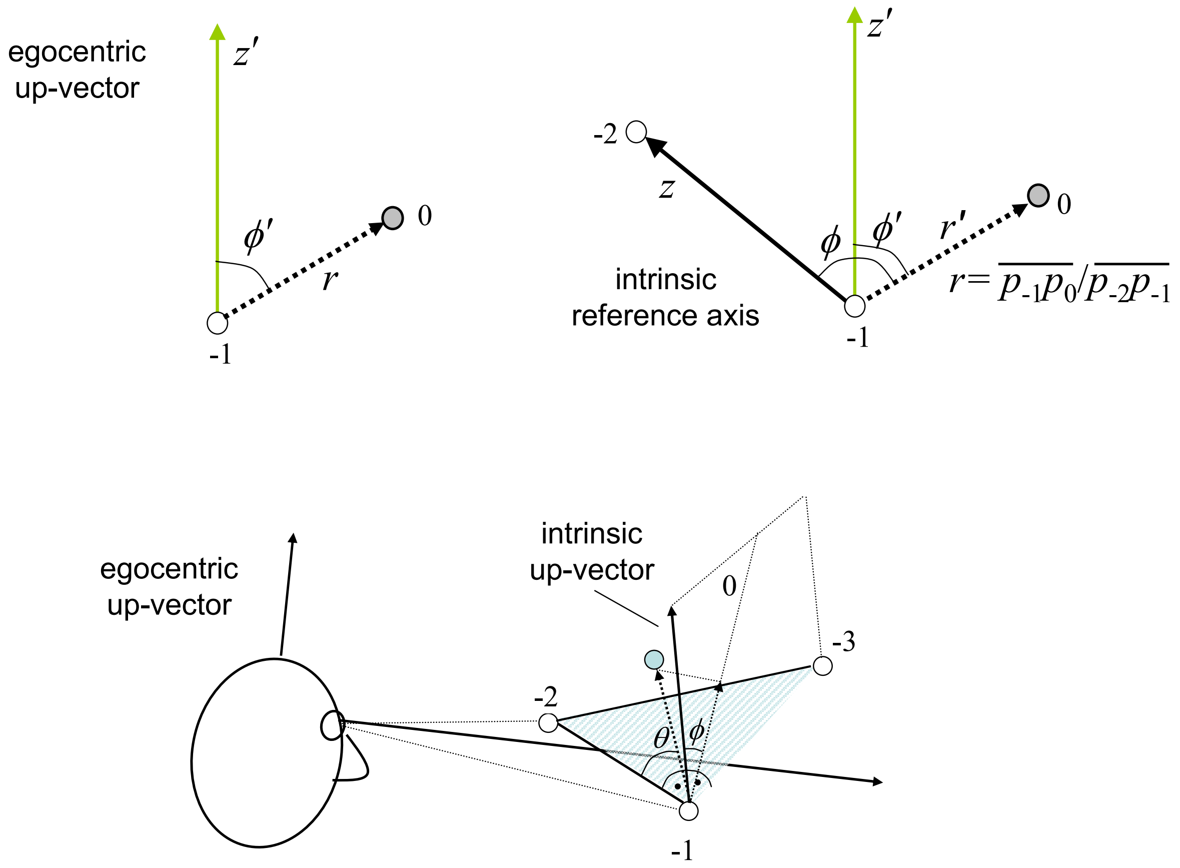


Figure B-4: Different Reference Systems. The objects are attended in the order (p_{-3}, p_{-2}, p_{-1}, p_0).

Therefore in some situations the production system might be forced to use spatial memory chunks in the semi-allocentric system. In three dimensions three locations are needed to define a pure allocentric reference system. Three locations define a reference plane, whereby the cross-product of the two linear independent vectors that can be defined by the three locations determines a third axis for a right handed reference system. In general these three axes can be used to define the spatial relation of a fourth object to one of the three objects. Semi-allocentric reference systems for a three-dimensional spatial relation can be defined by replacing one of or both of the intrinsic reference axes with the viewing axis or the up-vector of the egocentric reference system. We think that humans represent, in their mental model, a spatial relation by its angle and the fraction of its length to the reference axes. This can be deduced by measuring the structure of the variability of reproduced object locations [16]-[18]. The introduction of object-relations based on three (four) objects is important for three reasons: first, it fits well with the concept of intrinsic axis in the object configuration as reported by Mou and McNamara [10]. Second the concept of angles is essential to most cognitive operations in geometric tasks. Third, it is the simplest percept for spatial memory chunks that allows reconstructing object locations, also if the whole configuration is rotated. The distance dimension is crucial, since there is no universal scale like for angles where any angle can be seen in relation to a complete rotation around 360°. The distance can be represented in respect to the length of the reference axis or in respect to the field of view.

The representation in respect to the length of the reference axis is scale invariant and can later also be applied to a scaled environment (e.g. the screen is farther away). Representations of distances in respect to the field of view are considered in a semi-allocentric reference system. A semi-allocentric reference system is considered to use one location as the origin but to interpret direction and distance of a location relative to this point in an egocentric reference system. In our understanding, a distance is not perceived directly, but as a dimension of a location in one of these allocentric reference systems. For the model described in the following section it needs to be emphasized that if the distance is considered in the pure-allocentric reference system, the visual system needs to shift attention three times to assess a location – and therefore a relative inter-object distance. However, if the distance is represented in the semi-allocentric reference system only two attention shifts are needed (taking the attention shift to the first reference object into account).

The variances in recalled object-locations require the scalar values of a spatial relation to be noisy. To integrate noise into the location information of memory chunks the first question is how object-locations in different reference systems are represented in memory. [16] Showed among other things, that the distribution of recalled locations supports the assumption that subjects imagine object-locations on a plane relative to a centre in polar coordinates. We generalized this to use spherical coordinates in respect to model spatial cognition in three dimensions. This assumption has also some interesting implications on the representation of locations on a screen. Spherical coordinates are a system of curvilinear coordinates that are natural for describing positions on a sphere or spheroid. Generally ϕ is defined to be the azimuth angle around the polar axis, which is normally the up-vector, θ to be the zenith angle from the polar axis and r to be distance (radius) from a point to the origin. In the case of an allocentric reference system on a two dimensional screen this means, that if the three points p_{-2}, p_{-1}, p_0 were attended and p_0 has to be represented in a local allocentric reference system, the point p_{-1} defines the origin, the polar axis is given by (p_{-1}, p_{-2}), and the local spherical y-axis points orthogonal into the screen. For the semi-allocentric reference system on a screen, again p_{-1} is the origin, but the polar axis is parallel to the vertical axis of the screen and the x-axis is parallel to its horizontal axis. The next question is, if θ, ϕ, and r should be considered as single, independent memory chunks. Because it is impossible to imagine a distance without a direction and an angle without corresponding lines, it is reasonable to combine distance and angle as one percept in one memory chunk. Because of this argument, also in the case of the actual allocentric reference system the egocentric orientation of the reference system should be stored into the same memory chunk. This does not imply that the angle or the different dimensions of one chunk cannot be separated later. In spatial reasoning often two angles have to be compared. But this can be handled as commands to the visual module of the cognitive system. Then, timing issues can also be considered, e.g. for the mental rotation of an actual allocentric reference system. In principle the spatial information of the semi-allocentric reference system is now also present in the chunk of an actual allocentric reference system. This might suggest discarding memory chunks of the semi-allocentric reference system. But as mentioned above, creating object-location memory chunks in this semi-allocentric reference system is less effort to the visual module and therefore in some situations needful.

B.2.3 Noise

Finally a spatial location is represented by $D(r,\theta,\phi, r',\theta',\phi',e_{rs})$, where r, θ, ϕ are the spherical coordinates as described above, e_{rs} indicates in which reference system r, θ, ϕ have to be interpreted, and r', ϕ', θ' are additional attributes for the actual allocentric reference system and additionally hold the polar and azimuth angle in the semi-allocentric reference system. The values of the spherical coordinates in the memory chunk are interpreted as random numbers distributed according to a truncated normal distribution $N(x, x_0, \sigma_x)$, with standard deviations (σ_r, σ_θ, σ_ϕ) corresponding to each dimension. The scalar value in the slot of the memory chunk indicates the mean x_0 of the distribution. Independent of whether the distance is measured compared to the field of view or to the length of the reference axis the ability of humans to discriminate them should obey the Weber-Fechner Law. Therefore, the noise σ_r should scale linearly with r_0. For later calculations it is easier to consider the quantity

$f_r = r/r_0$ to specify the probability distribution. The noise model of f_r is again simply normally distributed with mean at 1 and a specific standard deviation σ_{fr}. In this way the noise σ_r is scale invariant. Therefore, if no scale transformations are to be considered the physical measurements of a distance can be used in the calculations. This is true regardless of whether the distance is represented as a fraction of the reference axis or as a fraction of the field of view.

Every time a location is to be encoded, it is decided if the perceived values for the location correspond to an already existing memory chunk. The posterior probability $P_{Di}=P(D_i=d^{(i)}|F_x)$ that the location of a feature F_x belongs to a memory chunk D_i and the probability P_0 that no appropriate memory chunk already exists, are given by:

$$P_{D_i} = \frac{P(F_x \mid D_i)}{V^{-1} + \sum_i P(F_x \mid D_i)}, \quad P_0 = \frac{V^{-1}}{V^{-1} + \sum_i P(F_x \mid D_i)} \tag{3}$$

The parameter V^{1} describes the weight of a noisy background and it is assumed that single dimensions of the spatial relation are not correlated.

$$\begin{aligned} P(F_x = (r_x,\theta_x,\phi_x,r'_x,\theta'_x,\phi'_x) \mid D = d(r,\theta,\phi,r',\theta',\phi')) = \\ N(r_x,r,\sigma_r)N(\theta_x,\theta,\sigma_\theta)N(\phi_x,\phi,\sigma_\phi)N(r'_x,r',\sigma_{r'})N(\theta'_x,\theta',\sigma_{\theta'})N(\phi_x,\phi,\sigma_{\phi'}) \end{aligned} \tag{4}$$

On the other hand, if an object-location is requested based on a memory chunk $d(r,\theta,\phi, r',\theta',\phi',e_{rs})$, the values are set to random values x according to (5). After the noise has been added to the location request, it is decided if the values are latched on possible features in the display. Therefore, the object-locations of all features $F_i(r_i, \theta_i,\phi_i,r_i',\theta_i',\phi_i')$ in question are calculated in the current local reference system corresponding to the reference system in the request. The probability P_{Fi}, that the location request is caught by feature F_i and the probability P_0 that it is not, are given similarly to (3) by:

$$P_{F_i} = \frac{P(x \mid F_i)}{V^{-1} + \sum_i P(x \mid F_i)}, \quad P_0 = \frac{V^{-1}}{V^{-1} + \sum_i P(x \mid F_i)}. \tag{5}$$

These equations express the posterior probability $P_{Fi}=P(F_i|x)$ that if a noisy location x from memory is given it results from the feature F_i. The likelihood probability functions $P(x|F_i)$ are the probability density according to if the feature F_i would have been the stimulus and are similar to (4). This process of encoding and reconstructing a location into a random number in memory is illustrated in Figure B-5.

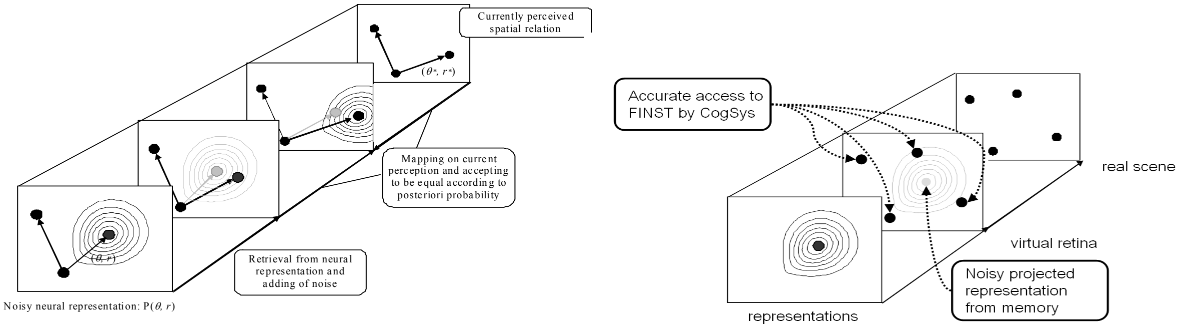


Figure B-5: Perception, Representation and Reconstruction of a Location.

The presented noise model has two interesting properties. First, because the truncated logistic distribution is asymmetric, the expected report of an object-location is biased away from the reference axis. This is the same effect as has been reported about categorical boundaries [16]. Second, for object-locations on a flat screen the values of ϕ, ϕ' are discrete $\phi,\phi'=(\pi/2,0,-\pi/2)$ and encode whether the object-location in question is on the left side, on the right side, or aligned, when facing into the direction of the reference axis. This is consistent with the assumption of interpreting the reference axis as a categorical boundary, where ϕ encodes the category.

From the results of our experiments we calculated for the noise parameter $\sigma_{fr} = 0.11$, which means that the length of a spatial relation can be reproduced by a human with a standard deviation of 11%. For the angle we got $\sigma_{\phi} = 6.3°$, which is 1.75% of 360° and $\sigma_{\theta} = 11.3°$ which is 3.1% of 360°. In general we try to use the same values for both allocentric reference systems. These values were measured for one specific view in 3D, where the viewing axis was along the reference axis. The experiments suggested that the values in 3D strongly depend on the view. In compromise with other data we suggest to use in the 2D case $\sigma_{fr} = 0.1$ and $\sigma_{\{\theta,\phi\}} = 8°$. However, all values show that directions can be memorized more efficiently than length. This is an aspect which could be an important design criterion of graphical displays. One example will be discussed below.

B.2.4 Visual Indexing

It is evident that subjects browsing a graphical layout structure encode environmental characteristics of object-locations, e.g. if an object is located on the border of a matrix. To encode such environmental features the cognitive system needs to attend objects nearby. The crucial point is that after some objects in the environment have been attended, attention needs to return to the object in question. If this return would depend on noisy spatial memory chunks, the strategy to encode environmental features might be highly counterproductive. At this point the concept of visual indexing, or FINST-FINger INSTantiation, [20] is needed. According to this theory the cognitive system has "access to places in the visual field at which some visual features are located, without assuming an explicit encoding of the location within some coordinate system, or an encoding of the feature type". Experiments suggest that the number of FINSTs in the visual system is limited to 4 to 5. In the visual module of ACT-R the concept of FINST is used to decide if an object has already been attended. Whenever an object is attended, a FINST is created. Because the number of simultaneously existing FINST is limited, any time a new visual object is attended the oldest FINST is removed to create a new FINST for the currently attended object. To implement environmental scan patterns, FINST needs to provide, additionally to the information that an object has already been attended, also information for accessing its location without, or at least with only minimal noise. In the visual module interface described below this has been accomplished by determining a visual index through the sequential position in the chain of attended locations. This index can be used in visual module commands to return (or avoid to return) attention to a particular location in the chain of attended locations.

B.3 MODELING EXAMPLES

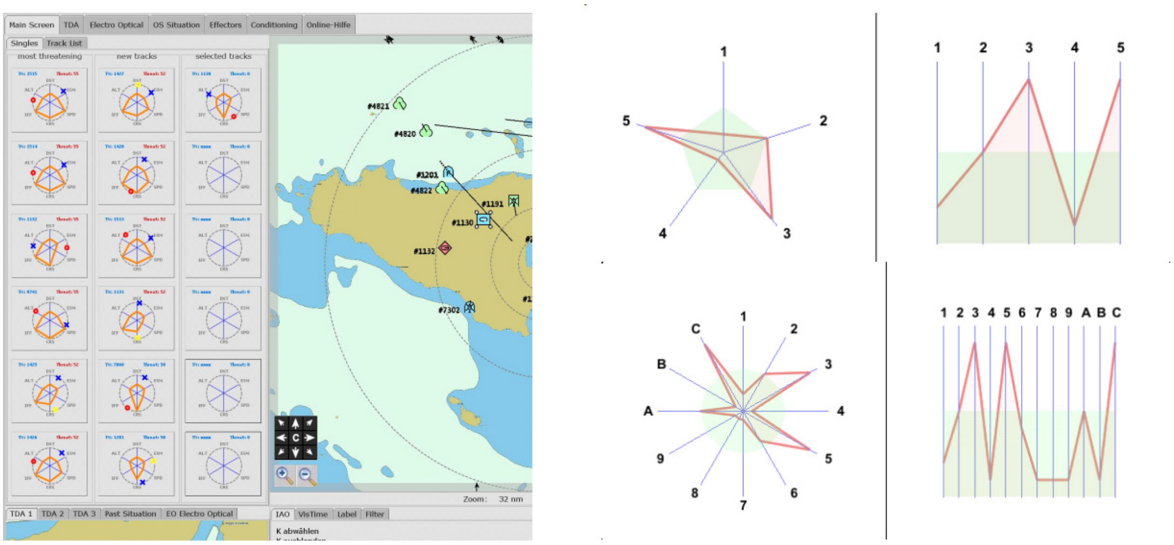
In the following we will give some examples on how human spatial information processing can be modelled within the proposed framework.

B.3.1 Parallel Plot vs. Polar Plot

The Fraunhofer FKIE developed a Tactical Situation Display (TDS) for Anti-Air Warfare (AAW) scenarios with some innovative display elements (Witt et al. 2009). One of these new unconventional display elements is the polar display for fast survey over the track attributes. The polar displays were used to constitute in integrated form the track attributes that are crucial for threat classification, such as Distance to the Ownship (DST), Altitude (ALT), Speed (SPD), Course (CRS), IFF-Information (IFF) and ESM-Emissions (ESM). For a single track attribute the polar display generates display proximity between the current attribute value and – taking into account predefined identification criteria – an un-critical attribute value on the attribute's parameter beam.

A symmetric figure represents the a priori defined scenario knowledge. For instance, it is known in advance which friendly, neutral (civil) and hostile radar emissions (ESM emissions) are to be expected. Similar friendly and neutral IFF (identification friend foe) codes are defined a priori. If potentially threatening attribute values are detected the respective value indicator breaks the symmetry within the figure. The normal range of kinematic attributes like speed or altitude is defined in advance as a tolerance area. By connecting the current indicator values of single attributes in a polar diagram a figure is generated which integrates the single pieces of information on a higher level of abstraction. This figure by means of symmetry or a-symmetry forms a so-called emergent feature which helps to transfer the interpretation of information content to the perception phase of human information processing, i.e. direct perception: A symmetric figure indicates an un-critical airborne contact. In contrast, an easy to perceive a-symmetry emphasizes the criticality of the contact. The most relevant advantage of polar displays, in spite of the graphical aggregation of individual attributes, is that they assure that the single pieces of information are better noticeable and perceivable. Thus, in contrast to classical alarm displays polar displays are alarming and diagnostic at the same time, because the possibly symptomatic characteristic of a single parameter value can be noticed easily. Furthermore, under different parameter constellations the figure-forming aggregation of single attributes allows for the direct derivation of higher-level task-related manifestations. In contrast, the notification about several pieces of information on separated displays as mentioned above requires multiple mental transformations and comparisons.

However, also alternative designs of displays exist that integrate the values of attributes into a figure. Figure B-6 von the right shows the comparison of the same attributes displayed in two different designs for a graphical display. Both designs constitute figures that can directly be perceived and interpreted by the user. However, considering equally distributed values for the attributes, the characteristics of the figures that are constituted differ in both designs. The question is which design produces the more salient features that can be recognized more easily from figures stored in memory. This question can be discussed within the proposed framework, by examining the emergent angles and fractions of distances within the figures. These angles and fractions of distances are restricted by the design of the display (see Figure B-7).



(a) **(b)**

Figure B-6: (a) The Usage of Polar Display in a Tactical Situation Picture of an AAW Interface; (b) Comparison of the Polar Display to a Linear Display.

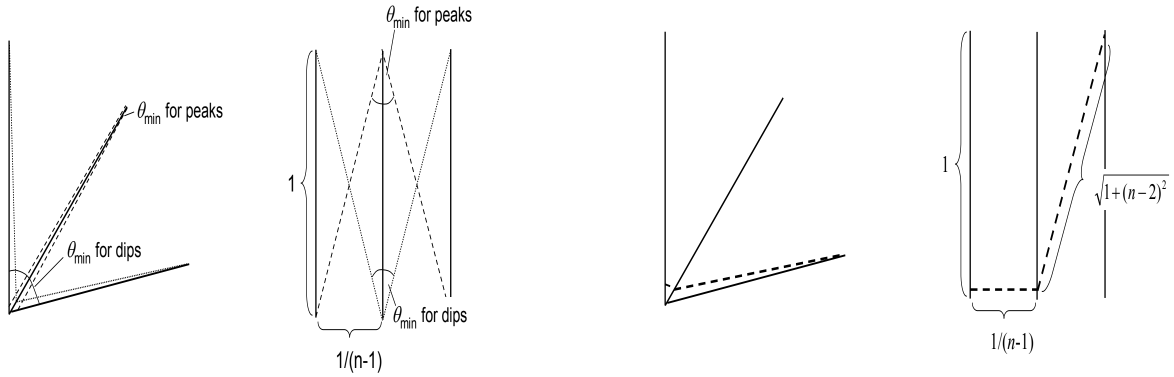


Figure B-7: Illustration of the Restrictions the Design of the Displays Put on Values of Attributes of Spatial Relations.

In case of the linear display the angle ranges $[2\cdot\tan^{-1}(1/(n\text{-}1)),180°]$ within peaks and dips, where n is the number of axes in the diagram. In a polar display the ranges differ between peaks and dips. The angle within a peak can theoretically become zero as a minimum and $180° - 360°/n$ as a maximum. For peaks pointing inward (dips) the smallest angle that can emerge is $2\cdot 360°/n$. Therefore the ranges of angles for peaks and dips are $[0,180° - 360°/n]$ and $[720°/n, 180°]$. For the fraction of distances of one arm to another in a linear display the range is $[(n\text{-}1)\sqrt{(1+(n\text{-}1)^{-2})}, 1/((n\text{-}1)\sqrt{(1+(n\text{-}1)^{-2})})]$. In case of the polar display theoretically the fraction of length can range from 0 to infinity. However, we assume here that it is bounded by the $r = 10$, and we think that also $\sigma_{fr} = 0.1$ only holds for r smaller than 10 or larger than 1/10.

Figure B-8 visualizes the ranges of angles and fractions of lengths. As can be seen from Figure B-8 the linear display provides a larger range of angles, especially if the number of axes in both displays is low. For the fraction of length the polar display provides a larger range of fractions that could emerge and could be discriminated by an operator.

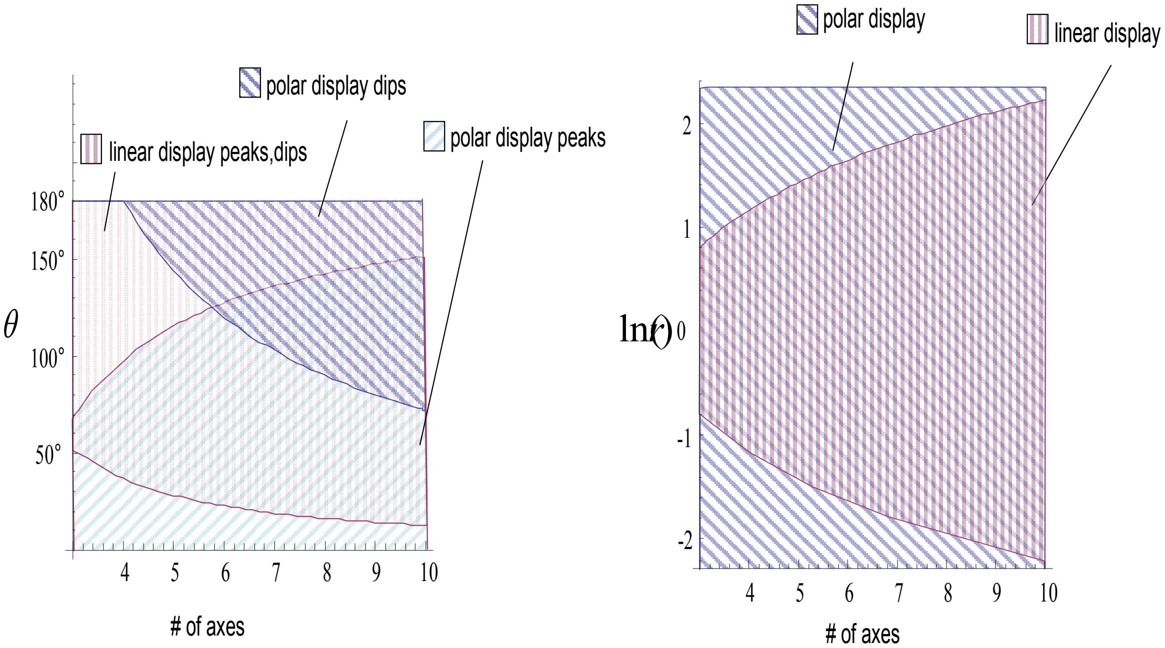


Figure B-8: Comparison of the Ranges for the Values of the Spatial Attributes Between Polar and Linear Display.

In a common task an operator needs to classify an alert that is signalled by the constituted figure within a display. For this he needs to compare the figure perceived with the figures from memory. The operator will extract salient features from the figure and will try to recall a figure possessing this salient feature from memory. These salient features are spatial relations in nature. As discussed above humans are limited in recalling spatial relations and this process is noisy. To identify a figure he will compare different features to come to a decision, about the meaning of the displayed alert. If one salient feature is very important for the characteristics of one figure he might have stored this by more than one single spatial relation. We will not discuss a complete model here. The differences in performance between both display designs result from the different ranges of spatial relations that can occur in the figures displayed. As shown above, these ranges depend on the number of axes, and the linear display seems to be advantageous for less number of axes. However, the noise from memory is different for different attributes of the spatial relations and these needs to be considered within the assessment of both display types.

To get a first model based assessment for both display types we consider the step in recalling the first salient feature of a displayed figure. For this two spatial relations are needed: one for the shape of the salient feature and one for its position within in the display. Both spatial relations are shown for both display types in Figure B-9.

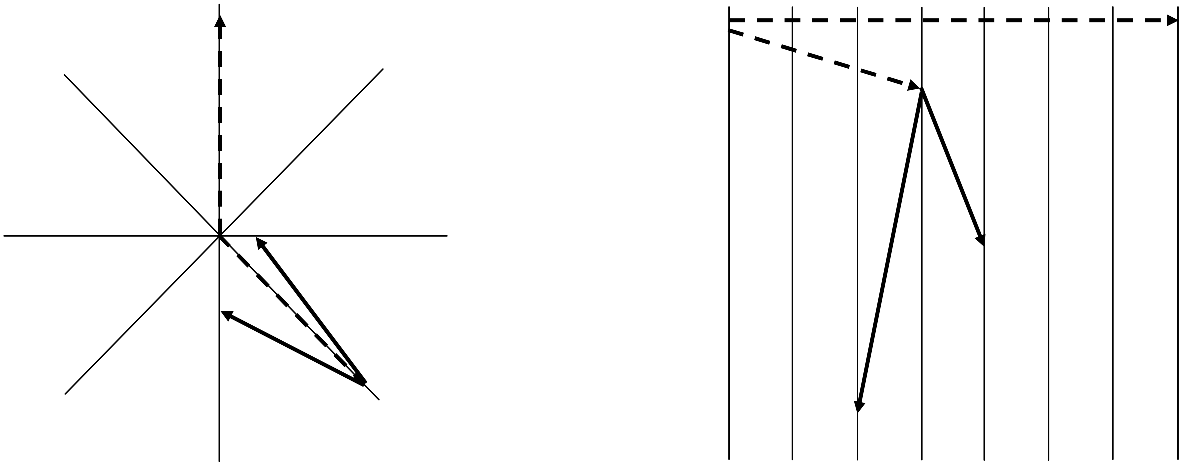


Figure B-9: Encoded Spatial Relation for the Relative Location for One Salient Feature in the Two Different Displays.

As a measure of performance we calculate the mean probability that the shape of the salient feature can be recalled from memory successfully and if the associated position which is recalled from memory will correctly be recognized. Actually the performance in both steps depends on how well the spatial relations can be discriminated from other spatial relations that are associated with figures in the display. The overall performance will be a weighted sum of the performance in both steps. Both steps could also be merged into one single recall for one memory chunk holding both relations, for the shape and its location. In this case, again the overall performance would be a weighted sum of the ability to discriminate the spatial relations from others. In fact we are not sure on how the weights have to be chosen or if they could be concluded from the ACT-R theory. However, the separation into these two steps clarifies the different contributions of the shape of the feature and its location.

For the first step we consider that only this salient feature is in memory and it only needs to be distinguished from the background noise. The expected probability $P(D_{Fx} = d^{(Fx)}|F_x)$ that the framework will propose the salient feature F_x perceived to be equal to the same noisy salient feature $d^{(Fx)}$ in memory can be calculated by integrating eq. (3) over all possible attributes of the spatial relations that constitute the feature within the display.

$$\left\langle P_{D_{Fx}} \right\rangle = \int_{\Omega(Fx)} \frac{P(F_x \mid D_{Fx})}{V^{-1} + P(F_x \mid D_{Fx})} P(F_x)\, dF_x\,, \tag{6}$$

To simplify calculations we consider that the displayed attributes on the axes are distributed in a way, that the probability density functions is uniform $P(F_x)=1/\Omega(F_x)$, which can also be interpreted as the background noise $V^{1} = P(F_x)$. Furthermore, it is assumed that exactly the same spatial relation has to be retrieved from memory that has been stored previously. Now the conditional probability density functions within the integral only depends on the standard deviation as a parameter.

$$P(F_x \mid D_{F_x}) = N(x_0, x_0, \sigma_x) = g(\sigma_x) \tag{7}$$

The variable ϕ, which is in the 2D case discrete, determines if the feature is a peak or a dip. We assume that there is no noise in the cognitive system about the categorization of a feature into a dip or a peak. Therefore, the consideration of ϕ causes (7) to be separated into a weighted sum of two integrals: one for the peaks and one for the dips. For the linear display the ranges of spatial attributes are the same for peaks and dips, as is also the probability of their occurrence. For the polar display it must be considered, that there are only dips for $n > 4$ and that the ranges of the attributes differ between dips and peaks. In the polar display the probability of the occurrence of peaks and dips differ and correlate with the ranges of the angles. In the same way should the probability mass of the fractions of distances be equally distributed on the interval $[1/n, 1]$ and $[1, n]$. This is achieved by taking the logarithm of r, which has also already been done in Figure B-8 for illustration purposes. Putting it all together results in Figure B-10. It shows the mean probability that the shape of one salient feature can be discriminated from the background noise and can successfully be recalled as a function of the number of axes.

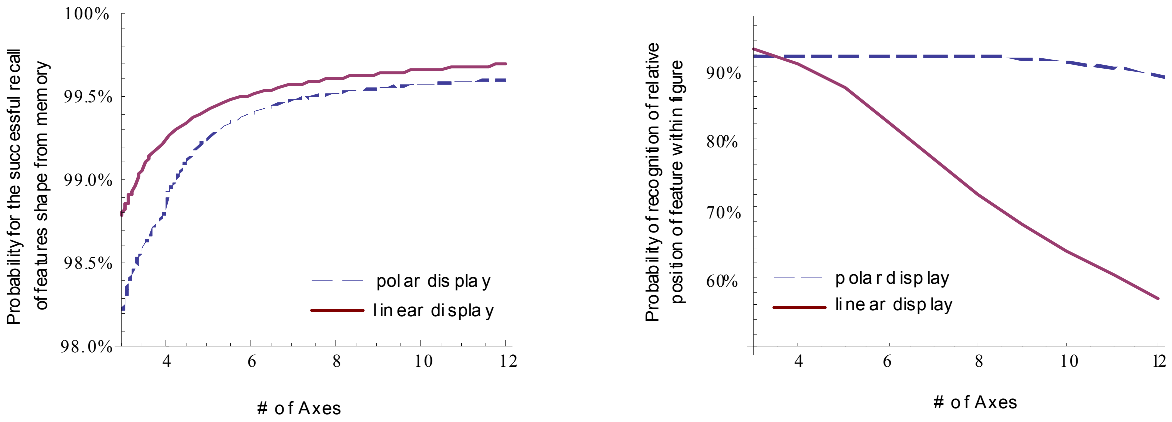


Figure B-10: Predicted Performances for the Recall of the Feature's Shape and for the Recognition of its Relative Position Within the Figure.

As can be seen an increasing number of axes improves the recall probability of the shape of the salient feature. The linear display shows, in this aspect, a better performance than the polar display which reflects the larger range of angles. However, it should be stressed, that in a real ACT-R simulation this probability also depends on the base activation level for the memory chunk. Furthermore, we omitted by this calculation, that in a polar display the direction of the peak might inherently be encoded within one memory chunk, because the allocentric reference system have different orientations depending on which axis the peak is. This is not the case for the linear display, where the orientations of the allocentric reference systems are always the same and aligned with the egocentric up-vector.

After the memory chunk of one salient feature has been retrieved successfully it is plausible to assume that the cognitive system has stored an associated memory chunk for the position of the salient feature within the figure. The cognitive system will than validate this position of the salient feature with the perceived position. The probability that the correct position of the salient feature will be accepted can be calculated within the framework by:

$$P(F_{x_0} \mid D_0 = d^{(0)}) = \int_{\Omega(x)} P(F_{x_0} \mid x) P(x \mid D_0 = d^{(0)}) dx = \int_{\Omega(x)} \frac{P(x \mid F_{x_0})}{V^{-1} + \sum_{i=1}^{n} P(x \mid F_{x_i})} P(x \mid D_0 = d^{(0)}) dx$$

The integral 'simulates' the noise that is added to the spatial attributes stored in the memory chunk. The possible features on which the location from the memory chunk could be latched are the discrete locations on the axes and therefore are given by: $P(x_0 \mid F_{x_i}) = N(x_0, x_i, \sigma_x)$. Again it is assumed that the correct feature spatial relation has been recalled from memory: $P(x \mid D_0 = d^{(0)}) = N(x, x_0, \sigma_x)$. The resulting predicted performance is shown in Figure B-10 on the right side. For the recognition of the position of the feature within the display the polar display clearly outperforms the linear display. This can be explained by the fact that the standard deviation of the noise for angles is smaller than for distances. The linear display strongly decreases with a higher number of axes, while in the polar display it seems to stay constant up to twelve axes.

B.3.2 3D-Vektorfields

The following example was originally designed for illustrating the theory for the assessment of a three dimensional operational picture. But the same modelling approach might be useful in applications of visualizing three dimensional vector fields.

To make tactical decision the operator needs an operational picture. In general he needs to know where different entities are located in relation to the ranges of effectors and sensors and how these locations will change in time. The operator needs answers to questions like: which entity will be next within reach of the effectors; which sensors and effectors will I be able to use; where is the best location to hit the enemy aircraft without endanger civilian entities in the proximity. All these questions can be answered by algorithms implemented in the combat computing system with high accuracy.

However, as long as the operator needs to make these queries sequentially or has to read the answers of the system in form of lists, performance in making the right decisions will be hampered. An external visualisation of the spatial layout of the operational picture should enable the operator to get answers to some of his queries that refer to spatial relations and time evolution intuitively by perception. To handle all spatial relations numerically in a list is impossible, because the number of possible spatial relations increase strongly (~ n^2) with the number of entities in the operational picture. However, also the derivation of spatial relations from a spatial layout can be biased and can only be derived by humans up to certain level of accuracy.

The physical parameters of sensors, effectors and tracks are three dimensional in nature. The choice of the views on the operational picture strongly influences performance of different cognitive tasks the operator needs to perform to extract the information he needs for his decision-making process. By modelling the process of information acquisition from the display, not only different visualization approaches can be compared prospectively, but the model can also give clues about how a single interface design can be improved. One short example on how a certain view can be assessed by the proposed framework is illustrated in Figure B-11 and will be discussed in the following.

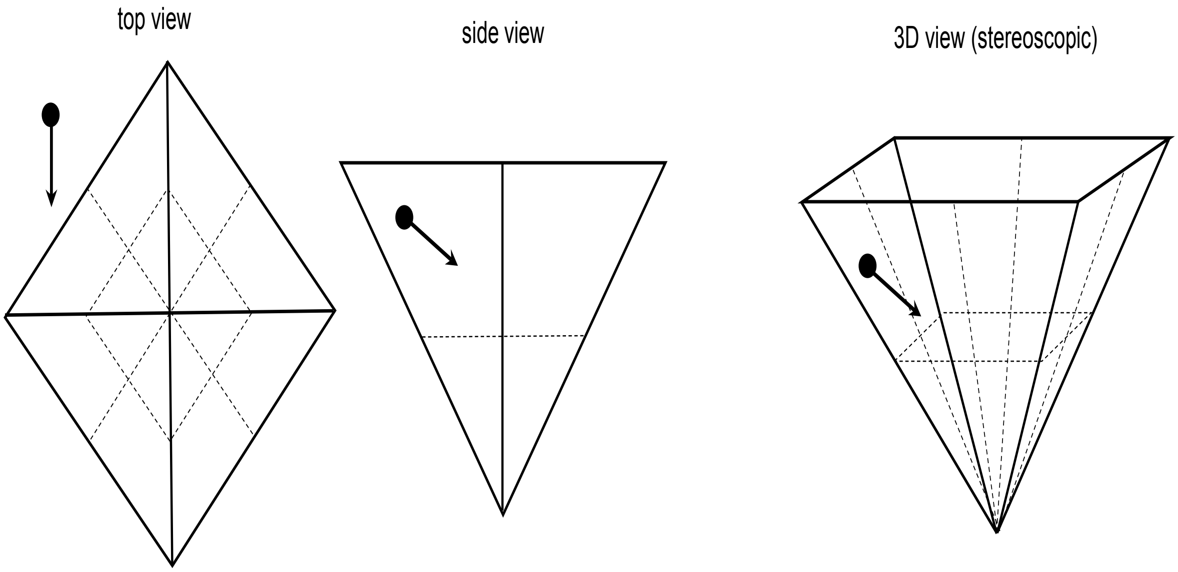


Figure B-11: Scenario – Decide if the continuation of a vector intersects a volume.

One basic question an operator is interested in to query from the operational picture is, if an aircraft will come into the range of one specific effectors. This information can only be retrieved if the spatial relations are imagined in three dimensions. To simplify argumentation we assume in the example of Figure B-8 that the shape of the range-volume is a pyramid standing on its top. To infer if the aircraft will enter or pass the volume, the edge in proximity to the projected flying path is critical. One solution to decide if the flight path crosses the volume is to derive if the flight path runs in front or behind the edge, when the scene is looked at from a viewpoint where the edge in question points towards the observer. To do so, the speed vector of the aircraft is projected in its length beyond the edge. Then, an arbitrary location on the edge is attended and the attention is directed back to the location of the aircraft afterwards. In this way an allocentric three dimensional reference system has been created by two axes. As explained in the theory section a third “up”-axis is inherently defined by both these axis. Now the cognitive system simply needs to decide if this inherently created “up”-axis of the local reference system has an up or a down component. More generally the operator could also compare the “up”-axes of this local reference system with the “up”-axes of the reference system that can be constituted by the edges at the base of the pyramid. This is illustrated in Figure B-12.

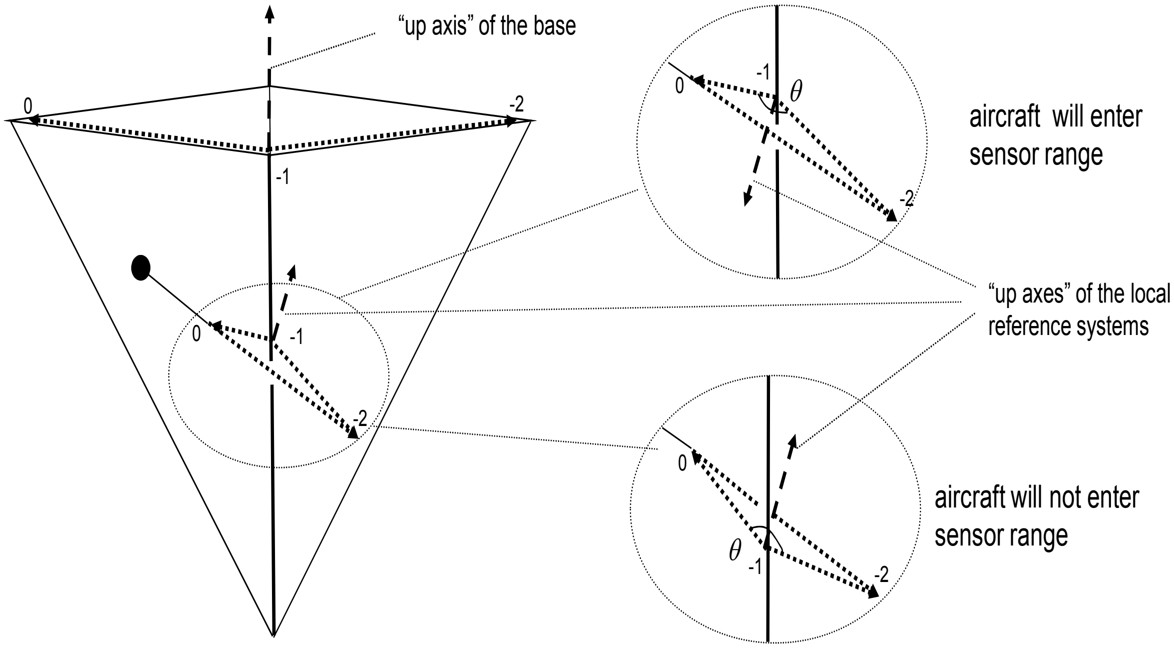


Figure B-12: Cognitive Steps to Decide if the Continuation of a Vector Intersects a Volume.

The imaginary picture might be constituted from a stereoscopic 3D-visualization or multiple 2D views. In a stereoscopic 3D view the operator only needs to project the flight path by extending the speed vector in his imaginary picture. All other locations needed to constitute the reference system can be retrieved by perception. If the operator uses multiple 2D views he needs to memorize one spatial relation and to transform it into a second view. There are two drawbacks if a spatial relation needs to be reconstructed from memory. Beside that its reconstruction is noisy it possibly needs to be retrieved multiple times, if the number of FINSTs are not sufficient to keep one of it constantly on the constructed location. Whenever this FINST is needed, the spatial relation needs to be retrieved from memory again, which is time-consuming. In the example above the operator could memorize from the top view the relative location of the aircraft to the centre axis of the volume. To construct this location of the aircraft in the side view, he than attends in the side view one location on the centre axes in the height of the aircraft and add the rotated spatial relation from memory to this location. This imaginary location he needs to memorize, since he needs to perform another rotational task and he will not be able to keep attention constantly on this imaginary location. The other rotation task is needed to reconstruct the imaginary edge in the three dimensional space of the side view. Only if both imaginary spatial relations have been created by transforming them from memory, they can be tested as described above. As can be seen by this brief model description one query to multiple 2D views needs a chain of multiple noisy steps to perform the task. It is even arguable, if the number of FINSTs is sufficient to complete the task. However, a detailed analysis by a detailed model is ongoing work. But this short discourse shows how the framework might help to assess the limits of an operator to perform such geometric tasks. In a stereoscopic 3D view the task can be completed, because for some of the locations that are needed to constitute the reference systems, attention is caught by the real stimulus and variability of the attended locations are lower. But some steps are still noisy and this produces inaccurate responses. The probability of inaccurate responses in dependence on the operational picture and the viewpoint could be predicted by the model.

B.3.3 D-Graphs

The display of graphs using a 3D layout can be beneficial when the third dimension has a meaning in the embedding space. In earlier work it was shown that 3D visualizations of networks can increase the number of links which can be followed, the number of groups of nodes discernible or the time required to follow a network link [23],[24]. When stereoscopic 3D visualizations are used disparity and vergence provide additional depth cues to the user to understand the three dimensional structure of the network [22]. However, what is true for two dimensional display elements is also true for 3D elements: it is not obvious which display elements are the most appropriate for a given task. It is therefore advantageous to quantitatively model the perception of stereoscopically displayed 3D elements.

Graphs consist of nodes and links. When graphs are displayed as projected two dimensional images a user perceives the nodes' positions on the two dimensional display but integrates the depth information in these two dimensional images into a three dimensional mental model. He is therefore able to judge their relative distances. When designing three dimensional graph visualizations it is advantageous to know the limits of the user perceiving these relations. The limits and mechanisms involved when perceiving three dimensional visualizations will be similar. This means that a model of the cognitive procedure involved when perceiving one kind of three-dimensional visualization will be useful in judging other kinds of three dimensional visualizations.

An example of a three dimensional visualization is depicted in Figure B-13. The isosurface depicted in the figure is a common display element of three dimensional visualizations. Knowing how well the user is able to perceive the distance between the isosurface depicted by the isolines will help in designing adequate visualizations. As in the two dimensional case there are undoubtedly distinctive features in the three dimensional visual scene which are considered by the perceptual system as key elements. They attract attention and the relative distances in-between them are the basics of the spatial mental model. The perception of a purely transparent surface such as the isosurface in Figure B-13 is difficult even if perceived stereoscopically because the entire process of object recognition is disturbed by the reduction of the elementary features on the surface e.g. there is almost no colour contrast. For this reason the curvature and feature edges of a surface are harder to recognize and a matching between the stereoscopic half images is very hard.

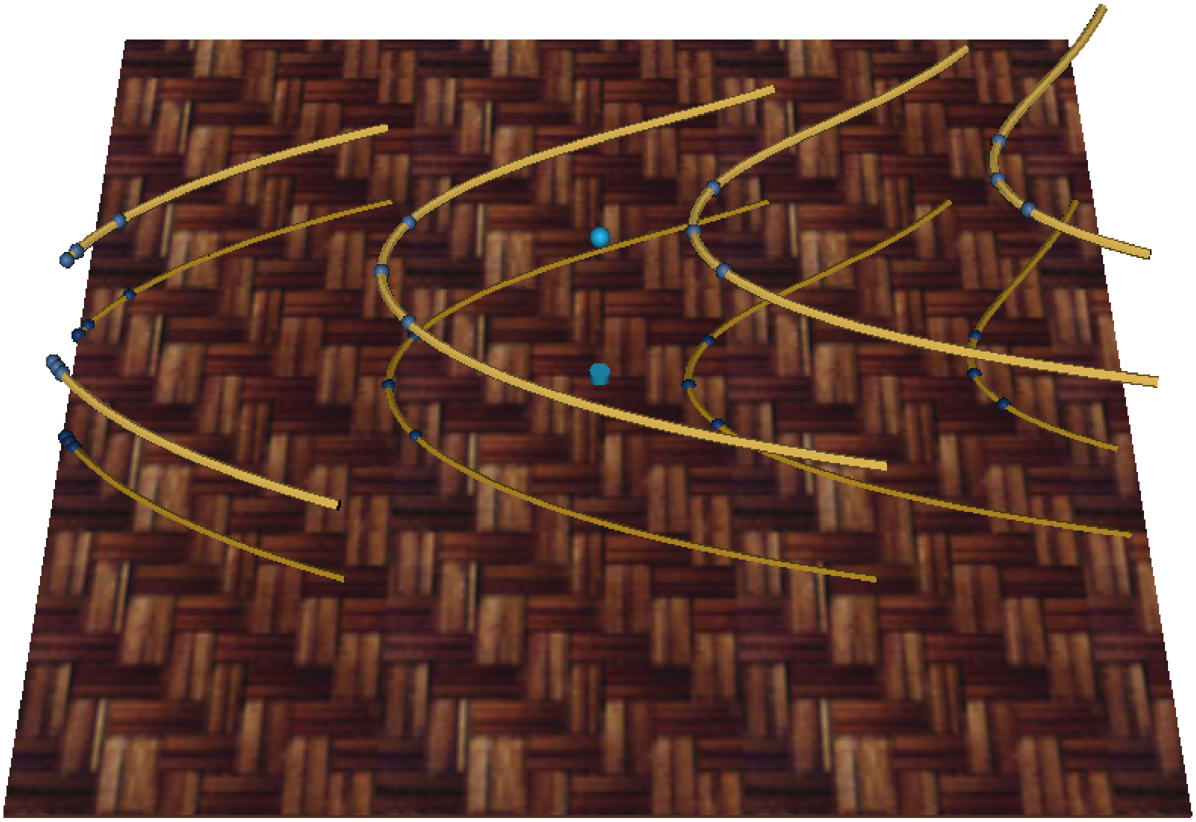


Figure B-13: Image of Isosurface from Experiment.

B.3.4 Display of Transparent 3D Surfaces

Gibson [25] proposed a theory of depth perception from texture gradients. Single texture elements appear more closely graduated with increasing distance. This invariant depth reference makes it possible to perceive sizes, distances and forms exactly. Applying small opaque features regularly onto a 3D surface will in effect be similar.

Interrante et al. [26],[27] examined the effects of using different kinds of semi-opaque features for transparent surfaces regarding the perception of shape. Their results showed significant differences between the purely transparent and the semi-opaque representation. Yet, no differences between individual semi-opaque methods i.e. curvature based versus isolines were proven.

B.3.5 Response Time for Distance Judgement

In order to investigate which semi-opaque visualization allows for the quickest estimation of depth we conducted an experiment where curvature oriented strokes were compared with isolines. To measure the accuracy of distance estimation we used the method of adjustment which means the subjects were asked to control a stimulus and adjust it to the test stimulus. The advantage of this method is that less experimental runs are needed to calculate the variance of the estimation.

The result of the experiment stimulated the research presented here. The curvature oriented strokes allowed for a quicker estimation of the distance than the isolines. In conclusion, the experiment yields that the semi-opaque methods show similar performance in accuracy. Yet, the temporal factor does show significant differences between the isoline and curvature-oriented method. The subjects were able to judge the distance more quickly in the curvature oriented method. There are two possible explanations for this. Firstly the number of individual elements, i.e. there were more strokes than isolines, and the size of the elements, i.e. the strokes are more compact, were not controlled properly within the experiment. Secondly, processing isolines is more time consuming for the visual system because to understand the semi-opaque surface prominent points on the isolines have to be determined. This process might be based on judging the curvature of the isolines which is time consuming. This paper investigates the second hypothesis in more detail.

B.3.6 Model of Distance Judgement

Before presenting the actual hypotheses the task of judging the distance is broken down into sub-steps. The hypotheses concern these sub-steps and the visual process involved. The process will be described based on the model introduced earlier. The model presented in this section was formulated based on earlier work [6] which suggests that the perceptual process consists of several attention shifts during which inter-object relations are assessed and represented as noisy values with a specific variance. The model will be formulated in regard to the results of the experiment presented in the previous section.

Judging the distance from a point on the ground to an overlaying semi-opaque surface can be decomposed into two main steps:

- Forming a vector orthogonal to the ground (base) at the point of interest; and
- Finding the intersection of the formed vector (from the point of interest) with the overlaid semi-opaque surface.

In order to evaluate the proficiency with which step one can be carried out a test point can be imagined which must be judged for orthogonality. The sub-steps will then look like the following:

a) User looks at central point;
b) User looks at another point on the base;
c) User looks at another point on the base;
d) An allocentric coordinate system is formed with the allocentric up-vector orthogonal to the base;
e) User looks at central point;
f) User looks at test point; and
g) Judgment concerning the angle is passed if unsure repeat e-f (or maybe even a-f) with a certain probability.

The formation of an orthogonal vector to the base must be assumed to be a basic task which the user is able to carry out using a template or built in mechanism. The template will work with noisy representations of points which mean that even if the template does work exactly the results will not be. When recalling the orthogonal angle the user will therefore recall a noisy representation.

The accuracy of the orthogonality judgment will certainly depend on the ability to form an allocentric up-vector. This again, will depend on the accuracy with which the points of the allocentric coordinate system are perceived in the three dimensional space. Building upon our earlier work6 we therefore hypothesize:

> H_1: The accuracy of the orthogonality judgment will be flawed or systematically be wrong when no visible perpendicular is present.

B.3.7 Experimental Validation

In order to provide an exact specification for the model based on the principles introduced earlier we devised a set of simple experiments. These experiments were supposed to allow us to find the exact parameters for the model and also to test our hypotheses.

B.3.7.1 Subjects and Apparatus

We had 13 volunteer subjects which were recruited from the staff of our institution. Four of these subjects were women. Before the experiments all participants' ability for binocular vision was tested. All subjects passed the requirement of 100" for stereoscopic vision. Subjects were also shown a virtual scene on the auto-stereoscopic display to get used to the artificial stereoscopy.

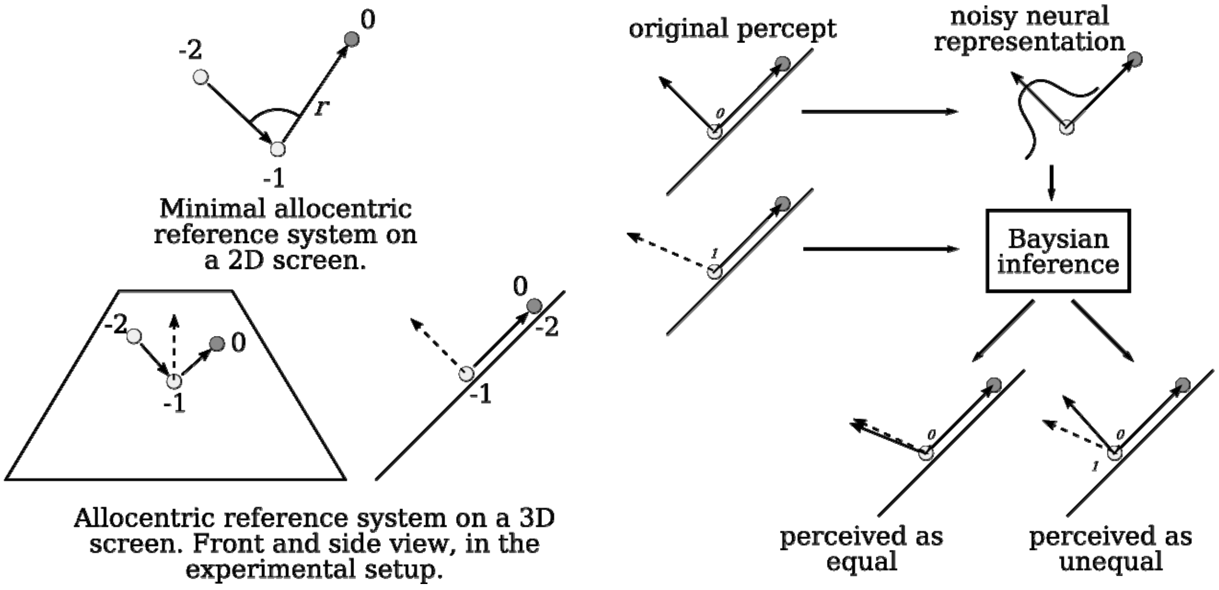
Experiments were conducted using a PC with a quad buffered graphics card. As a stereoscopic display system we used a tracked auto-stereoscopic display from SeeReal which uses movable prismatic lenses. The actual screen size of the auto-stereoscopic display was 35.8 cm by 28.7 cm. The screen resolution was 1280 x 1024. We also used super-sampling (16x) to facilitate good distance judgements. For rendering the virtual scene we used the Open Active World 3D toolkit which employs asymmetric view frustums and models the perspective distortion through the user's distance to the screen and the display size. We used an eye-tracking system from Tobii (X120) to record the user's eye movements. As the tracker allows the recording of both eyes separately we were able to calculate the three dimensional gaze position. However, the accuracy of this three dimensional measurement is not very good.

Users were placed so that their viewing distance was 60 cm. They were restrained in their head movement by a chin rest. We did not measure the inter-pupillary distance as earlier experiments have not shown its significance [28].

B.3.7.2 Task

The aim of the experiment was to investigate the quality with which the subjects are able to assess an orthogonal vector. We used a forced choice experimental design. The users had to decide whether a given test stimulus was in front of the orthogonal vector or behind. As the base a plane textured with a tiled image was displayed. Therefore, there were perspective and stereoscopic depth cues to judge positions on the plane. The centre point was a small cylinder in blue which had a good contrast to the base. The test point to be judged was coloured with the same colour and was placed approximately 3 cm above the centre point. A sample screenshot of the general experimental setup can be seen in Figure B-10 – in experiment 1 we did not show the semi-opaque surface.

We varied the position of the test point in two dimensions – vertically and in depth. Horizontally it was always placed at the same position as the centre point. In essence this means that the user only had to judge one angle. This can also be seen in the depiction of the allocentric coordinate system in Figure B-14(a) and Figure B-14(b). To control the change of the position of the test point we used three interleaved groups from the transposed up-down method described by Levitt [29]. The groups chosen were group 8 (P (X) = 0.5), group 2 (P (X) = 0.707) and group 3 (P (X) = 0.293). Choosing these three groups allowed a relatively precise measure for the calculation of the variance without too many trials per run.



(a) (b)

Figure B-14: (a) Minimal Allocentric Reference System in the General Case and in the Experimental Setup; (b) Perceptual Representation of an Angle.

The experimental procedure started with a blank screen. After the start the test scene was shown consisting of the base, the cylinder in the centre and the test point. The base was positioned with a uniformly distributed random horizontal offset of 2 cm – the rest of the objects had this offset, too. The cylinder also had a uniformly distributed random offset which was 2 mm in both directions on the plane. The test point was adjusted accordingly. The test point was placed approximately 3 cm above the cylinder with a depth offset determined by the up-down method. The up-down method used a step size of 1 mm. The subjects had to decide whether the test point was too far to the front (left mouse button) or to the back (right mouse button). After the click of the mouse the screen was blanked. After one second blank time the next test configuration was shown. The test terminated after all three groups of the transformed up-down method had at least five runs. Randomizing the horizontal position of the whole scene and the position of the cylinder was done in order to force the user to reassess the position of the objects in every trial.

B.3.7.3 Results

Overall the responses from the subjects were that they found passing judgment to be very hard. They had the feeling they were guessing most of the time. Also, none of the subjects said to have noticed the up-down method – interleaving the three groups was therefore a successful procedure. Since the up-down method will oscillate around the 50% point where the subject has to guess as he is not able to pass judgment, the feeling of the subjects that they were guessing most of the time proves that the method converged successfully.

We used the transformed up-down method of Levitt [29] with three groups. We therefore had as a result the mean and the variance of the supposed psycho physiological function. We also recorded response times. The accuracy results of the experiment can be seen in Table B-1. What can clearly be seen on first sight is that the results of the individual subjects differ. Yet, what is also clear is that there is a preference for a negative offset. All but two subjects wanted the test sphere to be farther in the back. This means that subjects either saw the sphere to be too far to the front or that the slant of the base was judged to be smaller. The standard error of the $X_{0.5}$ offset measure (o) is $SE_o = 8$.

Table B-1: The Values from the Three Up-Down Groups are Shown as well as the Calculated Variance (($X_{0.707}$ – $X_{0.293}$)/1.09).

Subject	Offset $X_{0.5}$	Offset $X_{0.707}$	Offset $X_{0.293}$	Variance
1	-3	-1	-5	3
2	-3	-1	-2	0
3	-8	-7	-7	0
4	0	1	1	0
5	-5	-2	-7	4
6	-2	-3	-4	2
7	-9	-8	-11	3
8	-2	-1	-3	2
9	-5	-3	-4	1
10	-6	-4	-10	5
11	-7	-2	-6	4
12	0	2	-1	2
Mean	-4	-2	-5	2

The response times of the subjects varied considerably. This can clearly be seen in Figure B-15.

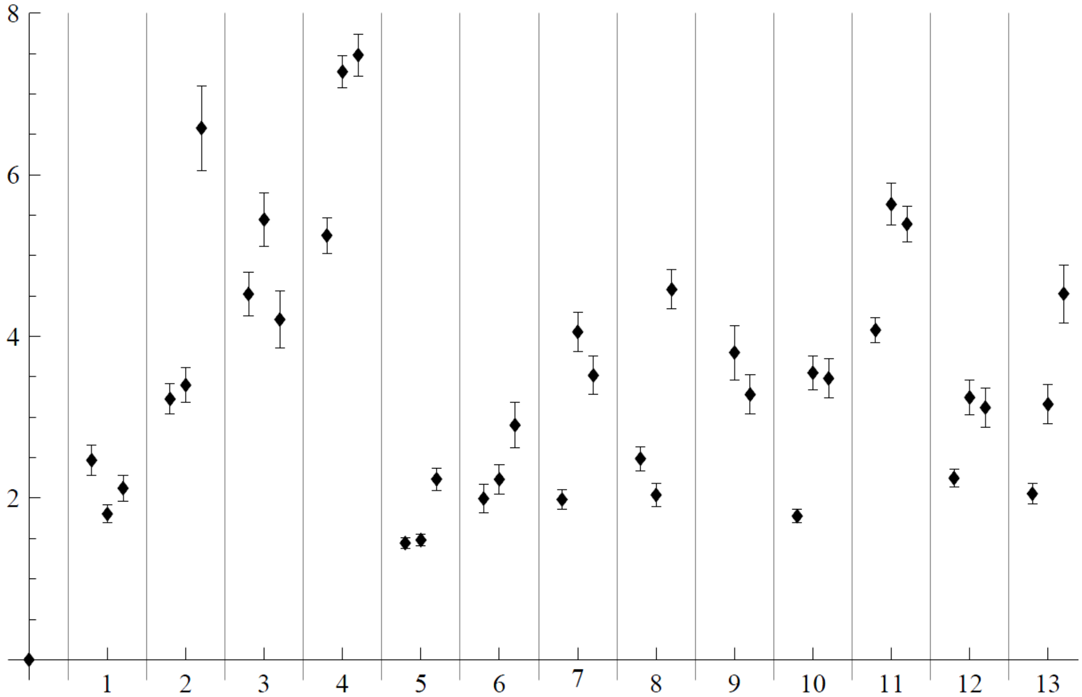


Figure B-15: The Results of the Experiments. The response time ± SE are depicted per subject.

B.3.7.4 Discussion

The experiment was designed to be minimalistic in order to be able to measure the hypothesized effects. Yet, the recorded data shows an enormous difference between subjects' response times and accuracy. We therefore decided to model the perceptual process for the experiment with normalized data to show the general applicability of the method to model the basic three dimensional perceptual processes.

B.3.8 Specification of the Model

B.3.8.1 Decision Process

A single assessment of a distance or angle is noisy. The cognitive system may improve the quality of a final assessment by making repeated assessments. The touchstone for the theory is whether or not the characteristics of response probabilities and response times can be explained by the same parameters. One critical aspect is how the cognitive system integrates the outcome of the multiple assessment steps. In the following, we assume a very restricted paradigm where one single step of the process only depends on local properties and involves no long-term correlations, such as accumulating the outcome of the assessments.

The most simple decision process consisting of more than one assessment step would be that the cognitive system is sceptical if, in the first assessment, the angle formed by the centre point and the test point is perceived to be actually 90 degrees and therefore orthogonal to the base. With probability p, the cognitive system makes a new assessment and only accepts the assessment with probability $(1 - p)$. The description of the single steps read:

- S1: Form an allocentric coordinate system by attending the central point and then attending two points on the base. Remember the angle φ0 of the up-vector which will be represented with noise.
- S2: Asses the angle (φ1) of the vector (v1) formed by the central point and the test point in the allocentric coordinate system i.e. the angle to the up-vector.
- S3: If the angle (φ1) is perceived by Bayesian inference to be the angle (φ0) of the remembered up- vector of the allocentric coordinate system from step S1 then proceed with a probability p with step S1 or if not proceed with probability (1 − p) with step S4.
- S4: If the angle (φ1) is perceived to be larger than (φ0) then proceed with step Sf otherwise with step Sb.
- Sf: Test point lies in front of the perpendicular.
- Sb: Test point lies behind the perpendicular.

Example Sessions:

- $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4 \rightarrow S_f$
- $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4 \rightarrow S_b$

The transition probability of $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_3$ is one. The transition probability of $S_3 \rightarrow S_1$ is given by the probability of an "equal" judgment and p according to the mentally projected angle. The transition probability $S_3 \rightarrow S_4$ is given by the probability of an "unequal" judgment. The latency time of the subject's answer can be calculated from the expected number of steps after which the decision process stops. The probability for the event R_{eq} of an equal or unequal response after n steps is the following:

$$P(n, R_{eq} = 0 | \varphi_0, \varphi_1) = P(B = 0 | \varphi_0, \varphi_1) P(B = 1 | \varphi_0, \varphi_1)^{n-1} p^{n-1} \tag{8}$$

$$P(n, R_{eq} = 1 | \varphi_0, \varphi_1) = P(B = 1 | \varphi_0, \varphi_1)^{n} p^{n-1} (1 - p) \tag{9}$$

This means the subject has made $n - 1$ equal judgments and has therefore repeated steps S1-S3 with a probability of p^{n-1}. In step n he stopped because he made an unequal judgment. The probability for the event $R_{<>}$ of a front/back response after n steps is then the following sum:

$$\begin{aligned} P(n, R_{<>} | \varphi_0, \varphi_1) = \; & P(R_{<>} | n, R_{eq} = 0, \varphi_0, \varphi_1) P(n, R_{eq} = 0 | \varphi_0, \varphi_1) + \\ & P(R_{<>} | n, R_{eq} = 1, \varphi_0, \varphi_1) P(n, R_{eq} = 1 | \varphi_0, \varphi_1) \end{aligned} \tag{10}$$

The term $P(R_{<>} |n, R_{eq} = 0, \varphi_0, \varphi_1)$ describes the probability that the user will come to a front/back conclusion with the parameters $n, R_{eq} = 0, \varphi_0$, φ_1 set. The next term $P(n, R_{eq} = 0|\varphi_0, \varphi_1)$ describes the probability that after n steps the event that the angle is perceived to be different than 90 degrees occurs. In the second part of the sum the term $P(R_{<>} |n, R_{eq} = 1, \varphi_0, \varphi_1)$ describes the probability that the user will come to a front/back conclusion with the parameters $n, R_{eq} = 1, \varphi_0, \varphi_1$ set. We can assume here that the user will randomly choose one direction as he came to the conclusion that the angle is actually 90 degrees. As a simplification the parameter can be set to 0.5 – this can also be adapted later e.g. to model the favouring of one direction. The last term $P(n, R_{eq} = 1|\varphi_0, \varphi_1)$ describes the probability that after n steps the event that the angle is perceived to be equal occurs. When deciding whether the angle φ_1 is larger or smaller than the memorized angle φ_0 the term $P(R_{<>} |Req = 0, \varphi_0, \varphi_1)$ describes the probability which can be formulated using the integral or cumulative distribution function:

$$\begin{aligned} P(R_{<>} = f | R_{eq} = 0, \varphi_0, \varphi_1) = \; & \int_0^{\varphi_1} P_\sigma(\varphi \,|\, \varphi_0 = \pi / 2) d\varphi \\ P(R_{<>} = b | R_{eq} = 0, \varphi_0, \varphi_1) = \; & 1 - P(R_{<>} => | \, R_{eq} = 0, \varphi_0, \varphi_1) \end{aligned} \tag{11}$$

When considering the simplifications mentioned above the probability that the decision process ends after n steps, if the two angles φ_0, φ_1 are given, is:

$$\begin{aligned} P(n \,|\, \varphi_0 \varphi_1) \;\; = \;\; & P(R_{<>} = f, n \,|\, \varphi_0 \varphi_1) + P(R_{<>} = b, n \,|\, \varphi_0 \varphi_1) \\ = \;\; & P(R_{<>} = f \,|\, n, R_{eq} = 0, \varphi_0 \varphi_1) P(n, R_{eq} = 0 \,|\, \varphi_0 \varphi_1) + 0.5 P(n, R_{eq} = 1 \,|\, \varphi_0 \varphi_1) + \\ & (1 - P(R_{<>} = f \,|\, R_{eq} = 0, \varphi_0 \varphi_1)) P(n, R_{eq} = 0 \,|\, \varphi_0 \varphi_1) + 0.5 P(n, R_{eq} = 1 \,|\, \varphi_0 \varphi_1) \\ = \;\; & P(n, R_{eq} = 0 \,|\, \varphi_0 \varphi_1) + P(n, R_{eq} = 1 \,|\, \varphi_0 \varphi_1) \end{aligned} \tag{12}$$

The expected number of steps can therefore be calculated using:

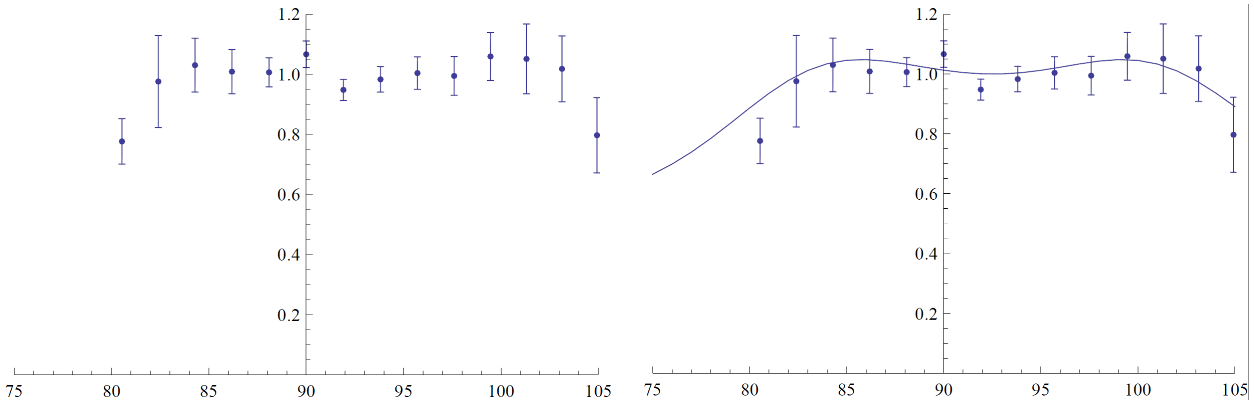
$$\langle n \rangle = \sum_{n=1}^{\infty} n P(n \,|\, \varphi_0 \varphi_1) \tag{13}$$

B.3.8.2 Response Time and Latency

One step means $S_4 \rightarrow S_f$ or $S_4 \rightarrow S_b$. One must also consider the basic latency of first reacting to a new display as well as the fixed latency of answering by clicking the mouse. This results in a fixed offset (700 ms) for the latency. To obtain the latency on a time scale, the duration of each step needs to be determined. In most architectures of cognition (e.g. ACT-R) the firing of one production rule or information processing step costs 50 ms. Therefore, the three shifts of attention in step S_1 cost 50 ms each. So for steps $S_1 \rightarrow S_3$ the overall time needed is 0.3 seconds. This means that when the user does three repetitions of the assessment process he will spend 900 ms for the repetitions, 100 ms to go from step S_3 to either S_f or S_b and additionally the basic offset latency of 700 ms.

Looking at the results from all the experiments yields that the fastest response time (for obvious configurations) is about 950 ms. With these assumptions it is now possible to completely specify the model and set parameters for V^{-1}, σ and p. However, in the results of the experiments there is an enormous difference between the subjects. It is therefore not feasible to just calculate the mean of all subjects and decide how to fit the model. The differences can clearly be seen in Figure B-12. One option is then the normalization of the results. But even the normalized results still differ greatly. The problem arises because of the different convergence points of the subjects. The mean offset $X_{0.5}$ of the subjects is spread from − 7 mm to 0 mm (76 – 90 degrees). Besides normalizing the data of the experiment the results also need to be shifted to account for individual differences in perception.

When joining the corrected and shifted results a trend does seem to emerge (see Figure B-16(a)). It can clearly be seen that users need less time for more obvious configurations than for the ones which are closer to the perpendicular. Fitting the model to the data can be done by using the variance of the accuracy for the probability density function σ = 3.8 – the model seems to fit nicely as can be seen in Figure B-16(b). However, the response time at 90 degrees is clearly not on the curve. This could be an outlier but the error bar only slightly touches the curve. The curve also seems to suggest that it took slightly less time to come to a conclusion when the test stimulus was very close to the perpendicular. This behaviour can also be found in the model – the chance of choosing to guess is higher for angles close to 90 degrees.



(a) **(b)**

Figure B-16: (a) The Normalized and Offset Corrected Results; (b) The Fitted Model Using *V-1 = 0, σ = 3.8, p = 0.96* and an Offset of *o = -3°*.

B.3.9 Discussion

We have presented our approach for stereoscopically visualizing radar ranges by using semi-opaque textures. To be able to choose the best semi-opaque texturing method we carried out an experiment which has shown that curvature oriented strokes allow for quicker distance judgements. In order to explain the results of this experiment we tried to describe the perceptual process using a Bayesian model which we derived from our earlier work. Additional experiments were carried out to prove our hypotheses. We formulated the hypothesized perceptual process of one of the experiments using our Bayesian approach. As the results of the experiments were very diverse between the individual subjects we normalized and shifted them. We were then able to fit the model to the experimental results. What remains to be done is to compare the accuracy of the results with the model using the same parameters as for the response times. Additionally the model needs to be able to account for the individual differences more directly.

The Bayesian approach seems to be very promising as it allows a compact and sensible description of the perceptual process. However, to see if it is truly applicable for the description of the stereoscopic perceptual process further experiments with a limited scope need to be carried out. The number of parameters which arise because of stereoscopy and three dimensional visualization are complicated to control. When the experiments lead to very diverse experimental results it is almost impossible to formulate a general model encompassing all these parameters.

B.4 OUTLOOK

This article proposes a framework to model human performance for spatial reasoning tasks. However, even for simple tasks the modeller needs to make some assumptions about how operators guide their attention to encode the visual space into memory. However, if it is true that operators are equally restricted as it is described within the framework, special expertise in spatial reasoning tasks can be trained by learning to choose the appropriate reference system for one specific environment. The framework could be used in future work to analyse how experts solve specific problems in spatial reasoning and transfer the knowledge to novices.

Overall we feel that more basic research is needed to confirm the parameters for the noisy representation of basic spatial relations in memory. The measurement of these basic variability parameters is difficult, because it is difficult to control which reference locations participants will use to encode one spatial location within a laboratory environment. For example any dot represented within a display could be encoded by using any of the corners of the screen. Also an unsolved question is, if or how the noise of basic spatial relations increases in time. Actually, performance in recalling locations or an operational picture decreases in time. But this does not necessarily mean, that noise of one spatial relation in a single memory chunk increases. It could also be explained by a model in which the accuracy of a single spatial relation is relatively stable in time, but some of multiple memory chunks of spatial relations that were used to encode the location of one single object could not be retrieved anymore. Because some spatial relations to refine the spatial location of one object are missing, accuracy in reconstructing the object location decreases.

Finally, we think that the proposed framework will be useful to understand limitations of humans in processing spatial information. In the first stage this will bring attention to possible problems in visualizing such information. In a second stage it could also be used for a model based assessment of different visualization designs.

B.5 REFERENCES

[1] Helmholtz, H.L.F. (1867). Handbuch der physiologischen Optik (1st ed.). Leipzig: Voss.

[2] Luneburg, R.K. (1947). Mathematical analysis of binocular vision. New York: Oxford University Press.

[3] Stevens, A. and Coupe, P. (1978). Distortions in judged spatial relations. Cognitive Psychology, 10, 422-437.

[4] Byrne, R.W. (1979). Memory for urban geography. Quarterly Journal of Experimental Psychology, 31, 147-154.

[5] Tversky, B. (1981). Distortions in memory for maps. Cognitive Psychology, 13, 407-433.

[6] McNamara, T.P. (2003). How are the locations of objects in the environment represented in memory? In Freksa, C., Brauer, W., Habel, C., and Wender, K. (Eds.), Spatial cognition III: Routes and navigation, human memory and learning, spatial representation and spatial reasoning. (174-191). Berlin: Springer.

[7] Blank, A.A. (1957). The geometry of vision. British Journal of Physiological Optics, 14, 1-30.

[8] Indow, T. (1974). On geometry of frameless binocular space. Psychologia, 17, 50-63.

[9] Heller, J. (1997). On the psychophysics of binocular space perception. Journal of Mathematical Psychology, 41, 29-43.

[10] Mou, W. and McNamara, T.P. (2002). Intrinsic frames of reference in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 162-170.

[11] Schmidt, T., Werner, S. and Diedrichsen, J. (2003). Spatial distortions induced by multiple visual landmarks: How local distortions combine to produce complex distortion patterns. Attention, Perception & Psychophysics, 65, 861-873.

[12] Schmidt, T. (2004). Spatial distortions in visual short-term memory: Interplay of intrinsic and extrinsic reference systems. Spatial Cognition & Computation, 4(4), 313-336.

[13] Schmidt, T. and Lee, E.Y. (2006). Spatial memory organized by environmental geometry. Spatial Cognition & Computation, 6(4), 347-369.

[14] Winkelholz, C. and Schlick, C. (2007a). Modelling human spatial memory within a symbolic architecture of cognition. In Barkowsky, Th., Knauff, M., Ligozat, G. and Montello, D.R. (Eds.), Spatial cognition V: Reasoning, action, interaction. (229-248), Berlin: Springer.

[15] Winkelholz, C. and Schlick, C. (2007b). Bridging psychophysics and cognitive engineering in visual perception, Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2520-2527.

[16] Huttenlocher, J., Hedges, L.V. and Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. Psychological Review, 98(3), 352-376.

[17] Winkelholz, C., Kleiber, M. and Schlick, C.M. (2010a). Analysis of the Variability of Three-Dimensional Spatial Relations in Visual Short-Term Memory. Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society 2010. Mahweh, NJ, USA: Lawrence Erlbaum Associates.

[18] Winkelholz, C., Kleiber, M. and Schlick, C.M. (2010b). Modelling human performance in processing basic three-dimensional spatial relations. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC 2010).

[19] Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C. and Qin, Y. (2004). An integrated theory of the mind. Psychological Review 111(4), 1036-1060.

[20] Pylyshyn, Z.W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition 32, 65-97.

[21] Witt, O., Grandt, M. and Küttelwesch, H. (2009). Direct Perception Displays for Military Radar-based Air Surveillance. In: Engin. Psychol. And Cog. Ergonomics, HCII 2009, LNAI 5639 proceedings (S.606-615). Springer.

[22] Ware, C. and Franck, G. (1996). Evaluating stereo and motion cues for visualizing information nets in three dimensions, *ACM Trans. Graph.*, Bd. 15, Nr. 2, S. 121-140, April.

[23] Baier, H., Buss, M., Freyberger, F. and Schmidt, G. (2000). Benefits of combined active stereo vision and haptic telepresence, in *Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on*, Bd. 1, S. 702 -707, Vol.1.

[24] Kleiber, M., Weltjen, B. and Förster, J. (2012). Stereoscopic desktop VR system for telemaintenance, S. 828803-828803-8.

[25] Gibson, J.J. (1950). The Perception of the Visual World, Houghton Mifflin, Boston, MA, USA.

[26] Interrante, V., Fuchs, H. and Pizer, S.M. (1997). Conveying the 3d shape of smoothly curving transparent surfaces via texture, IEEE Transactions on Visualization and Computer Graphics 3(2), 98-117.

[27] Interrante, V. (1997). Illustrating surface shape in volume data via principal direction-driven 3d line integral convolution, in SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques, 109-116, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA.

[28] Kleiber, M. and Winkelholz, C. (2008). Distortion of depth perception in virtual environments using stereoscopic displays: quantitative assessment and corrective measures, Stereoscopic Displays and Applications XIX 6803(1), 68030C, SPIE.

[29] Levitt, H. (1971). Transformed up-down methods in psychoacoustics, The Journal of the Acoustical Society of America 49, February.

Annex C – THE VISTG FRAMEWORK FOR DYNAMIC NETWORK VISUALISATION

C.1 INTRODUCTION

A Framework is a strong structure erected on some firm base to guide and support the detailed construction of a particular edifice. When the edifice is complete, the Framework may or may not remain visible; it may even be removed entirely. The VisTG Framework for Dynamic Network Visualisation (hereafter called "the Framework") is no different. The "edifices" it supports and guides are sets and choices of tools and techniques for performing tasks that involve networks changing over time, tools and techniques that are aided by displays designed to enhance the user's ability to visualise what is happening in the network.

The VisTG Framework has been developed over the lifetimes of several Task Groups and their associated NX and official NATO Workshops. The reports of these bodies have largely discussed the developments within those groups, but nowhere has an overall description of the Framework been presented. This Annex endeavours to rectify that, and therefore contains extensive material quoted or paraphrased from the various earlier Final Reports and Workshop Proceedings. In most cases, the original report expands on the parts quoted or referenced here, though the newer developments during the life of IST-085 are integrated with the older material without specific note.

The "firm base" for the Framework is the VisTG Reference Model first described in the Final Report of IST-013, known colloquially as "the HAT Report". The VisTG Reference Model starts with the presumption that a user wants to know the state of some property of the world that is represented in a computer data space. The reason the user wants to know the state is that if the state differs from the way the user would like it to be, the user may be able to act to correct the discrepancy, or that knowledge of this state contributes to the user's understanding of some situation in the world inside or outside the computer. Acting to correct discrepancies between the observed and desired state of a variable is called "control" [7].

C.1.1 Uses of Perception

"Control" is just one way of using real-time observations of a changing system. Instead, the user may want to determine the state in order to keep it within tolerable bounds, leaving it alone when it remains within those bounds. This use of real-time observation is called "maintaining situation awareness", or more simply "monitoring". Monitoring and Controlling are usually treated together in the VisTG Reference Model, and hence in the Framework, under the name "Monitoring/Controlling", since both monitoring and controlling have similar requirements for presentation and display in support of the visualization that happens in the user's head.

In order to determine the value of a property of interest that is to be controlled or monitored, it may be necessary to discover something as yet unknown. The user searches for it, and when it is found, the search terminates. This use of perception is simply labelled "Search".

A third use of perception involves actions very much like those of Search. Various aspects of the data space are examined, but the examination does not terminate when the value of a particular factor is discovered. Instead, the results are held in memory, so that they may be readily available if they are later needed as part of a Controlling/Monitoring process. This process is called "Exploring". In contrast to Search, the values obtained by Exploration are useful only if they are likely still to be close to their discovered value when they are later needed. Exploration discovers slowly changing and stable aspects of the environment, whereas the values found during Search are needed at the time of Search and can change freely thereafter.

A fourth use of perception, called “Alerting”, differs in character from the other three. Whereas the other three demand ongoing observation of the changing value of some property or factor of the data space, “Alerting” depends only on whether the value of a property matches some criterion (e.g. Does the acoustic babble at a party contain the sound of my name?) If it does, then the user may want to attend to something related to the alerting condition, and observe the value of the alerting property or of something related to it.

Alerting can be considered as a filtering process with many filters that can run in parallel in the background. It is valuable when the data space is too large for the user to continuously observe all of it. Humans have such filters, as, for example, the unexpected movement in the unattended visual periphery that allows one to glance at something that might turn out to be an approaching threat. In automated systems, an arbitrary number of such filters can be running at any moment, scanning large data spaces or monitoring incoming observations of real-time events. These four uses of perception were first presented by Taylor [10].

C.1.2 A Framework for a Particular Kind of Data Space

The VisTG Reference Model and the uses of perception apply no matter what kind of data are of interest. A Framework for a particular kind of task must, however, is concerned with the nature of the data. The IST-085 mandate was to consider the visualisation of dynamic networks, which means that a large part of the Framework for Visualisation of Dynamic Networks must be concerned with what kinds of network properties might possibly be the subject of the different modes of perception, and with the ways these properties might be represented in the different pathways of the VisTG reference Model. Accordingly, a large part of the Framework development has been the gradual evolution of what some Task Group members have called “a universal theory of networks”.

Networks consist of entities, in this chapter called “nodes” that are connected in some way by “links” sometimes called “edges” in discussions of network properties. A link may consist of a description of a relationship between two nodes, such as that A is the brother of B, or it may represent some traffic-carrying connection such as a road or an electronic link between computers. Networks may be physically embodied, or they may be entirely conceptual.

Networks as analyzed represent abstractions of structures perceived to exist in the real world. The analyzed network is “clean” in the same way as are the integer numbers used to count biological populations without taking note of differences in the size and health of the individuals. Clean networks are amenable to mathematical analysis techniques, but the user of such techniques always has to be aware of the possibility that the results may miss something important in the real-world environment that affects the behaviour of the real-world structure represented by the analytic network. This report uses the concept of “embedding fields” introduced in the Final Report of IST-059 as a first step in addressing the issue.

C.1.3 Complementary Routes to Understanding: Visualisation and Analysis

Albert Einstein is said to have claimed that while he could easily visualise the physics involved in his theories, putting those visualizations into mathematical analytic form was much more difficult. To visualize is to see a picture in one’s mind, whereas to analyze is to determine the relationships among isolated identifiable objects and properties in the picture. Displays suitable for effective visualisation differ from displays appropriate for effective analysis. Analysis may suffer from clutter, whereas Visualisation often benefits from a high density of presentation elements, analogous to the high density of visual elements we see in the natural world.

Good understanding often depends on an interaction between visualisation and analysis. In popular culture, the relationship is often called “right brain” and “left brain”, nomenclature that carries a germ of truth despite its oversimplification. For analysis, individual display items should be easily identified and distinctive, whereas for visualisation patterns over large numbers of individually indistinct items should be visually apparent. To understand the distinction, imagine a turbulent wind over a wheat field. If the wheat field is dense, the bending of the thousands of stalks makes very clear the twisting flow of the wind as it changes from moment to moment, but if there is only one stalk per square meter, no flow pattern can be seen, although the visible stalks still bend to the wind in the same way, and the flow patterns could be analysed, perhaps more easily than by using all the stalks. Substituting a wind gauge for the wheat stalks would allow more precise data to be recorded, and these data could later be turned into a map of flow properties by analysis and re-representation, but the dense representation offered by the full wheat field is much easier to visualise directly.

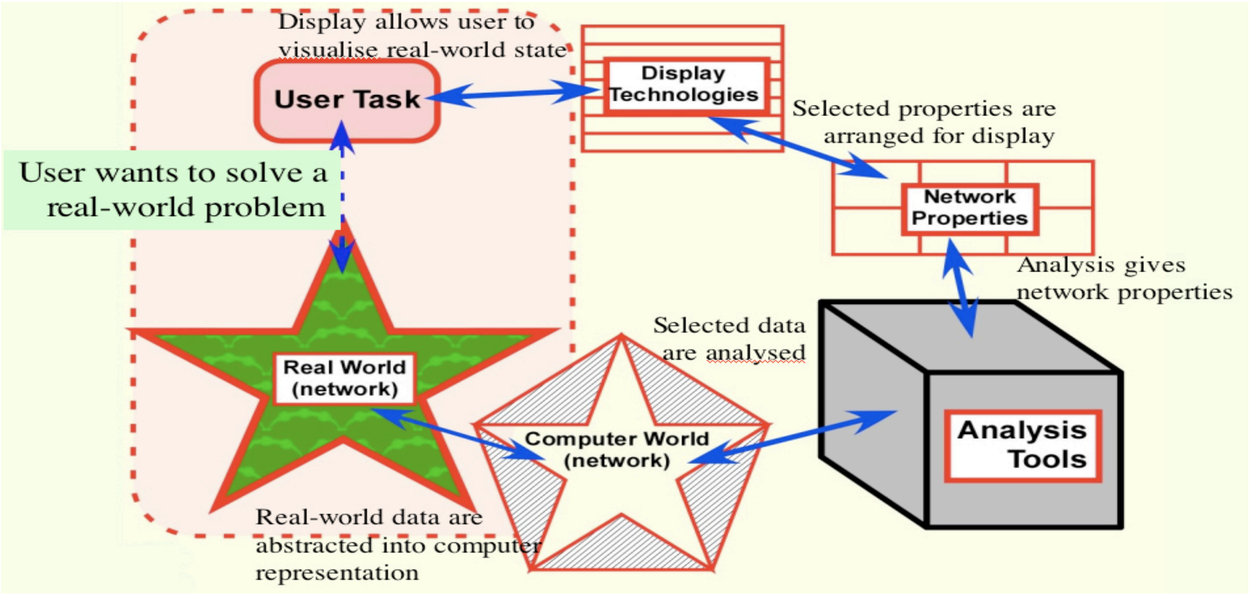
Analysis and visualisation work together. Some people habitually emphasise one pathway rather than the other, and these differences appear also in the way they learn new things [5] and [6]. Analytically oriented people usually learn better when new facts cohere with their existing body of understanding, whereas visually oriented people often can pick up isolated bits of information and eventually place them all into a “big picture”. These interpersonal differences may well be important in designing displays for different users and for different purposes.

C.1.4 The Nature of the Framework

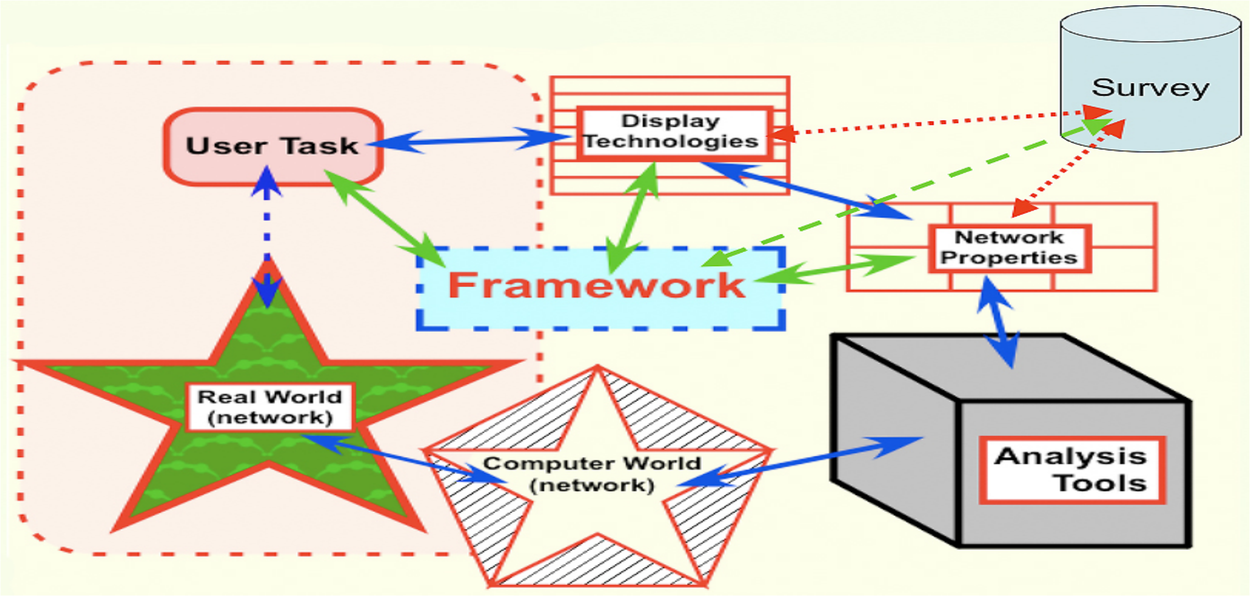
The Framework can be seen from several different viewpoints. It could be seen as analogous to the periodic table of the elements in chemistry. One cannot do chemistry using only the periodic table, but to know the table makes doing chemistry much easier. In that sense, it can be a conceptual aid to thinking about problems involving networks. From another viewpoint, the Framework is a process for determining what kinds of presentation might be suitable for a given task in a particular domain.

The Framework for any visualisation domain includes issues such as how the data are collected, which affects the dynamics of a presentation, questions about the task and the context of the presentation (a briefing, an analysis by a single interpreter, etc.), and so forth. For Network Visualisation, the Framework incorporates a comprehensive description of the dimensions over which a network may vary, and for networks that are dynamic or uncertain it includes some information-theoretic considerations, discussed in Annex D.

The Framework concept is intended as an aid to a user, rather than as a prescriptive procedure. The user is assumed to have some kind of real-world task. The Framework offers ways of describing tasks that may help the user to identify other tasks that have a similar underlying structure thought they may look different on the surface. Solutions for comparable tasks might help with the current one. The data for a task are accessible to the user only insofar as they are represented in some computerized data space, and are seen by the user only after they have passed through a variety of analytic processes that output individual or global properties of the data space (network properties in Figure C-1, since networks are the domain of interest to IST-085. The Framework offers a rational set of descriptions of possible network properties. The analyzed properties must be displayed in a way that the user can relate to the task at hand. The Framework offers a rational way of describing display methods, so that they can be related to techniques used in other situations (labelled as ‘survey” in Figure C-1).



(a)



(b)

Figure C-1: The Framework Concept. (a, Above) A typical workflow for an analyst who wants to solve a real-world problem through understanding the surrogate data held in a computerized data space; (b, Below) The framework is intended to help the user define the task and the properties of the data space so that effective use can be made of existing experience in choosing or designing appropriate methods of selecting, analyzing, and displaying the data in ways useful for the particular task.

C.2 THE VISTG REFERENCE MODEL

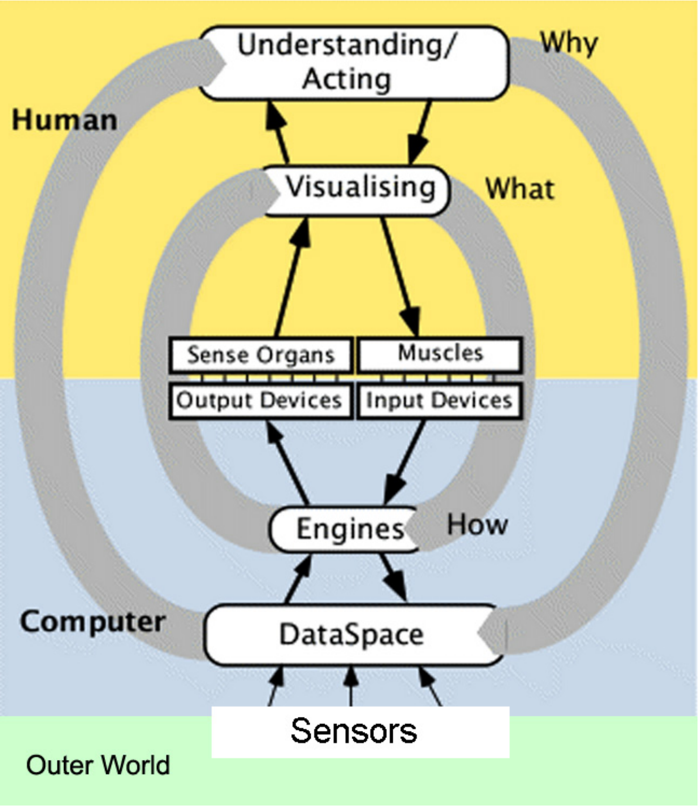


Figure C-2: The VisTG Reference Model.

The VisTG Reference Model starts from the proposition that the human user is trying to understand something about the real world that is represented in a data space, and perhaps act upon that world. It is conceptually derived from the "Perceptual Control Theory" (PCT) of W. T. Powers [7], and was initially described in the Final Report of IST-013 (RTO-TR-030, 2000).

The user may be in any of the attentive modes of perception described above, namely Controlling/Monitoring, Searching, or Exploring. What the user can perceive of the data space is represented in the figure by the leftmost broad grey arrow. The rightmost grey arrow represents the user's influences on the data space if the user is controlling. The two outer grey arrows together form a complete feedback loop.

Since the user has no telepathic connection with the contents of the data space, the actual selection, organization, and presentation of the data is done by processes labelled "engines" in the figure, and the interpretation by the process labelled "visualising" in the human. As discussed above, a parallel process called "Analyzing" would exist in a more complete model, but IST-085 is concerned only with visualising.

The passage of data between the engines and the visualizing process is represented in the figure by the inner pair of grey arrows that form an inner loop. However, the visualising process can communicate with the engines only through physical interfaces such as display surfaces seen by eyes, keyboards touched by fingers, sounds heard by ears, and so forth. The actual signal paths are represented in the figure by solid black arrows.

The structure of the VisTG Reference model is thus a set of nested feedback loops, the inner ones supporting the outer. Each arrow represents possibly many parallel paths. When analyzing or designing a presentation to support visualisation it is often necessary to consider these "atomic" paths individually while remaining alert for possible "molecular" interactions among them.

When one is interacting with anything such as an individual property of some complex structure, the feedback loop is complete, as shown in Figure C-3. One perceives the current state of the property, compares it with a desired state, and acts so as to influence the state until the revised state matches the desired value. This is controlling, and it applies universally to the interactions between the user and the Engines through the physical interface. Control can be analyzed as an engineering discipline; the effects of time lags, resolution and bandwidth, noise and external disturbance apply equally whether the components are mechanical, electronic, biological, or a mixture of them all. The approach to Control used in the VisTG Reference model is that of Perceptual Control Theory [7]. In PCT, control is exercised over both simple and complex variables by structures composed of Elementary or atomic Control Units (ECUs), each of which is structured as shown in Figure C-3.

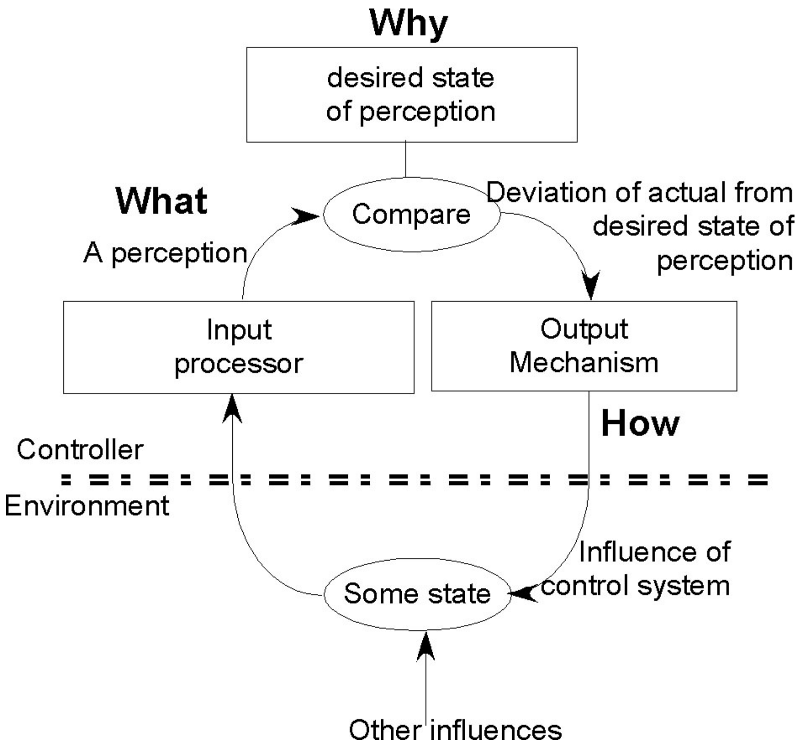


Figure C-3: A Generic Elementary Control Unit (ECU).

The ECU of Figure C-3 controls a "perception", which is an internal representation of some state in its environment, an environment that may well include many other ECUs. The perception is "*What*" the ECU controls. The perception is created by some input processor usually called the "perceptual input function", which works on possibly many inputs from the environment of the ECU to produce a single value of the perception. The ECU acts to bring the value of the perception to a reference value provided from an outside source. This reference value is the "*Why*" of control. The difference between the reference value and the actual value of the perception is called the "error". The error is the input to an output mechanism, often called the "output function" of the ECU. The output function provides a signal that acts on the environment of the ECU. In using the Framework, the important question is often the clarification of Why and what, which jointly determine what needs to be done.

The environment of an ECU may include other ECUs, and its output function may contribute to the reference signal of some of them, as suggested in Figure C-4, which shows how ECUs at two control levels work together to allow a person to ring a doorbell. The output function and its distribution to the reference points of other ECUs is the "How" of perceptual control. According to Perceptual Control Theory, the entire complex for controlling and monitoring consists of similar connections among ECUs of the same configuration, but with different perceptual input functions and output functions, analogous to the way that all complex chemistry is made by linkages among atoms similarly configured but with different nuclei and electron valence shells.

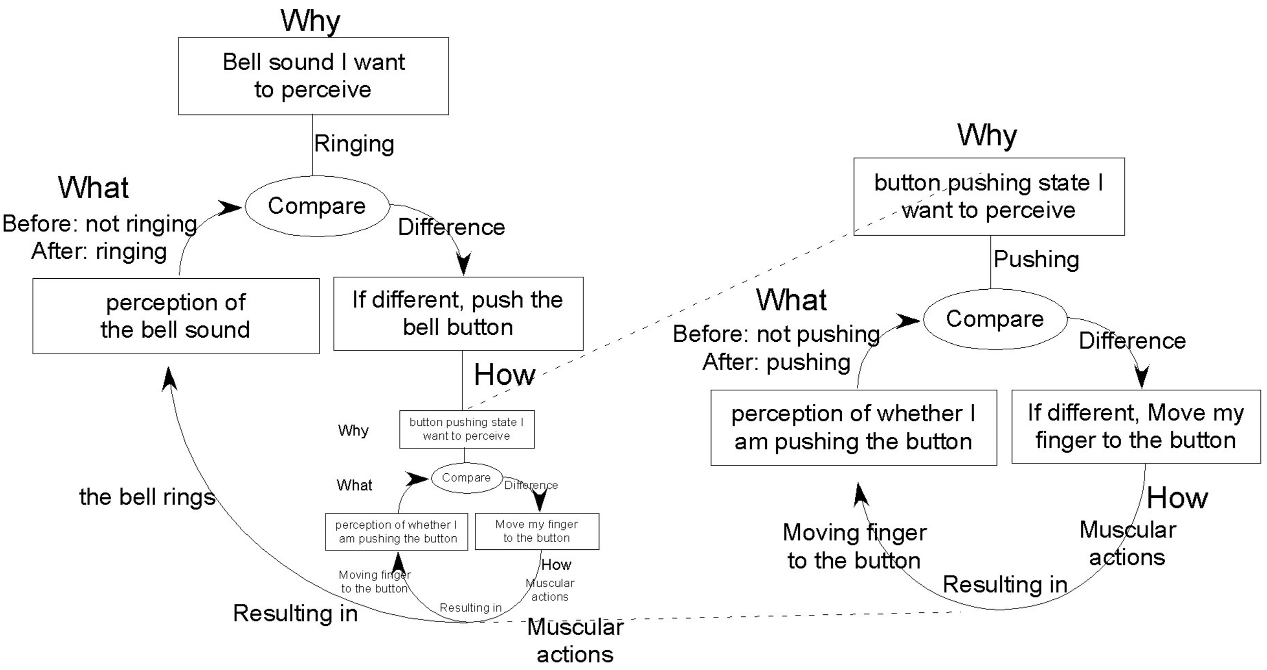


Figure C-4: Two Levels of Perceptual Control. The "How" of ringing a bell is a control system for perceiving one's finger to be pushing a button. The "why" of pushing the button is that the upper control unit wants to hear a bell ringing.

The perceptual input function of an ECU determines what property of the environment it perceives. In the entire complex of nested control units, no variation in the environment has an effect unless there is a perceptual input function that is influenced by that variation. As far as the control complex is concerned there is nothing else in the environment at all.

J.G. Taylor [9] argued and demonstrated that the ability to perceive a property of the environment depended almost entirely on whether the observer acted with reference to the thing perceived. PCT ascribes the development of perceptual input functions to a process known as "reorganization" that leads to effective control. Effective control happens when the thing perceived corresponds to some property of the environment that is preferentially influenced by control output actions. PCT therefore suggests that it is not at all assured that the complex control systems in different people will see the same properties of the environment. One of the roles of training and practice is to ensure that trainees reorganize so that they become able to see the same properties of their environment – in this case their display presentations – as are seen by the instructor.

Figure C-5 illustrates what may be called the "perceptual mirror" principle: if a person perceives something in the environment, for that person the thing *is* a part of the environment, whether or not anyone else sees it.

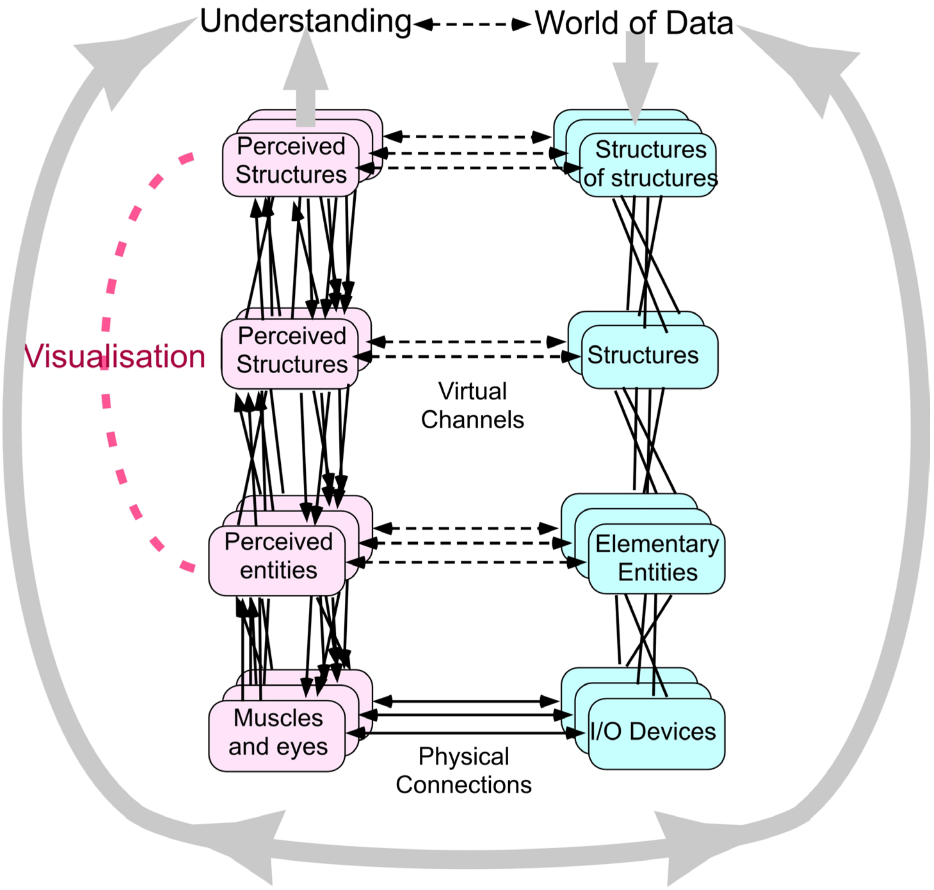


Figure C-5: Representing the Perceptual Side of the Inner "Grey Loop" and the I/O Loop of the VisTG Reference Model. Perceptual functions in the human (left column) create properties that can be perceived in the data space (right column). The perceived structures may or may not represent something useful in the data space, but they are what the user perceives. Acting on them tests their reality.

As Figure C-5 suggests, each of the "grey loops" in the VisTG Reference model may consist of several nested loops or control layers. The figure illustrates the loop that connects the human visualisation with the data space abstractions created by the computer's engines. What the figure does not show is the important contribution made by memory in completing the interpretations of the data at the different levels of perception.

The VisTG Reference Model, and more particularly its breakdown into "atomic" control loops, suggests a set of questions to be asked. Treated generically, these questions lead too much of the Framework considered as process, and to the descriptive aspects of the Framework considered as structure. The questions (from TR-IST-021) are:

- Q1. What user purpose is being considered?
  - Q1a. What higher-level purpose does this one support?
- Q2. What information does the user need to get from the computer to achieve the purpose?
- Q3. What does the user need to tell the computer to allow it to provide the needed information?
- Q4. What impediments might inhibit the user from taking advantage of the information provided?
- Q5. What impediments might inhibit the user from providing the computer the information it needs?
- Q6. Is there any mechanism to alert the user to information that might be important for the purpose, but that is not currently evident in the display?

These questions provide the basis for the development of a Framework for any domain. Q1 asks what kinds of tasks might be wanted. Q2 asks about the nature of the data space, which in the case of the Framework for Dynamic Network Visualisation constitutes much of the Framework development. Q3 ties in with the answers to Q2, and provides a start to defining an interface. Q4 deals with the limitations of human processing and of presentation media, as well as with possible interferences among different requirements imposed by the answers to Q1 and Q2. The same applies to Q5, though with a different emphasis. Q6 refers to Alerting, which introduces an independent set of issues.

C.3 THE FRAMEWORK FOR DYNAMIC NETWORK VISUALISATION

C.3.1 What is in a Framework, and Why

A Framework for visualisation contains three major components, which are principled descriptions of:

- The data domain;
- User tasks likely to be performed in the data domain; and
- Presentation methods suited to different possible tasks and user expertise levels.

The data domain of interest to IST-085 is that of dynamic networks, and by far the largest part of the Framework development has been devoted to the properties of dynamic networks in the real world. The user tasks depend on who the user might be, and since many different kinds of users may be interested in working with dynamic networks, the generic descriptions must be at a fairly basic level, ready to be expanded if the Framework is to be used for particular purposes as disparate as Intelligence visualisation, Civil infrastructure vulnerability visualisation, textual report visualisation, and so forth. The same applies to the presentation methods. The best that can be done within the remit of IST-085 is to link certain kinds of presentation methods to certain kinds of tasks appropriate to different basic properties of dynamic networks.

The VisTG Framework can serve several purposes, for end users and for researchers and developers. One of its uses is to help end users to define their problems, by asking the right questions. To ask a question properly often turns out to provide the answer. To this end, the descriptive components of the framework can be cast in the form of queries for the user to answer, such as “Is your data stable, or is it coming in sporadically, continuously, or repetitively?” or “Will you use the presentation to develop your own understanding or to brief others?” Ideally, the queries should be developed into an interactive medium, but this has not yet been done.

For researchers and developers, the Framework might help them to examine the multi-dimensional descriptive space to determine where suitable methods have not been developed for important problems. Research into such lacunae should offer higher payoffs than research into random issues of network visualisation. Similarly, presentation techniques that have proved successful under specific condition may be slotted into the multi-dimensional descriptive space in such a way that the underlying effective aspects of effective techniques can be abstracted and used in other situations. Again, this has not been done by IST-085, but the Framework should make it possible for others to begin to do it.

C.3.2 Task Description

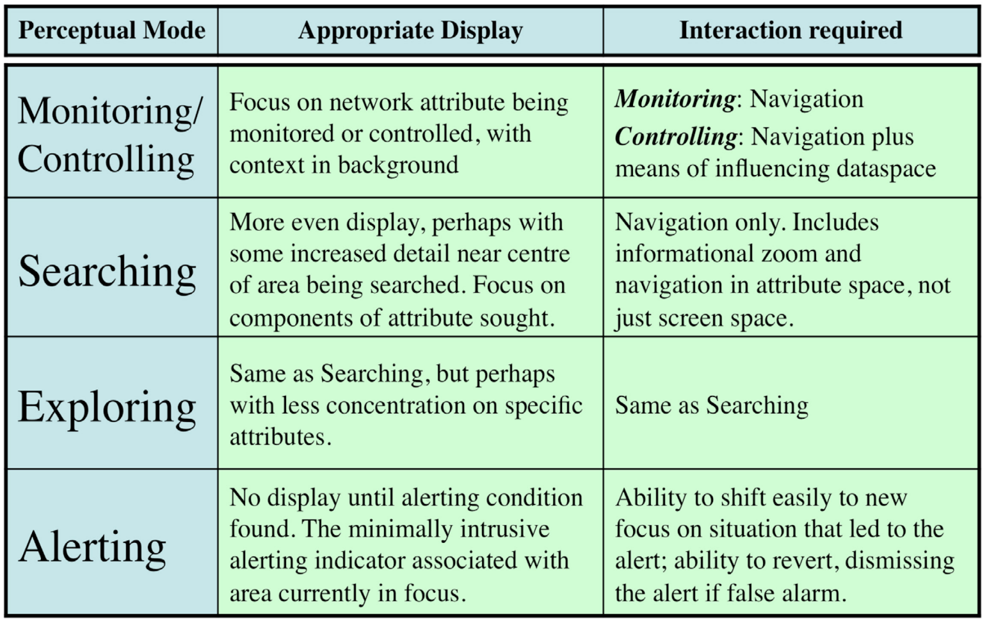
The complexity of possible tasks in different domains is probably beyond any easy description, but by analogy with chemistry, in which all the complex compounds are made of a few simple atoms joined in simple ways, we may be able to describe at least some of the atoms that combine to make any task of arbitrary complexity in any domain.

In the introduction, four "modes of perception" were introduced. These are:

- **Controlling/Monitoring**: The perceived condition may be changing; in monitoring mode, the perceiver maintains situation awareness of the changing situation, whereas in controlling mode the perceiver acts to influence the situation to bring the monitored condition toward a desired reference condition, or to keep it there by correcting deviations caused by known or unknown external influences.
- **Searching**: The Searcher cannot determine the state of some condition being monitored, and discovery of some contributing condition will improve its definition or permit effective action to influence that state. The Searcher examines a small portion of the data space, changing the portion observed until the sought data is discovered. The search is for the contributing condition, and terminates when the definition of the primary monitored condition is sufficient for the user's purpose of the moment. Search may involve waiting until appropriate data arrives, rather than waiting until it is found in a static data space. Since the sought data is used immediately in the primary monitoring or controlling process, it can be used even if it may change rapidly.
- **Exploring**: Controlling or monitoring is more efficient if some of the data relevant to the monitored condition is available in memory, rather than having to be sought when it is needed. Exploring provides these long-lasting data. Although, like Searching, Exploring involves roving examination of small portions of the data space, the discovered data must change very slowly if it is to be useful at some unknown later time. Exploring discovers the relatively invariant background structure and content of the data space, whereas Searching discovers a momentary state. Exploring is thus an important aspect of learning and training.
- **Alerting**: Alerting differs from the other three modes in that it is normally a background activity, and in that it is not involved with any ongoing controlling/monitoring, nor in preparing for future monitoring or controlling. To initiate an alerting process, criteria must be defined such that if the process detects some part of the data space or of incoming data that satisfies the criteria, an alerting signal occurs. Humans have such processes. For example, if one is expecting a phone call during a party, one is not continuously listening for the call, but if the phone rings, one hears it through the babble when one might not if one were not expecting the call. Some alerts are built-in; an unexpected touch on the back of the neck will usually draw one's attention, as will an unexpected movement in the visual periphery. In computational systems, an unlimited number of background alerting processes may be initiated with no visible display, their function being to draw the user's attention by some probably visual or auditory means to a situation that might warrant control or monitoring. Effective alerting signals use the human's built-in or trained alerting processes, clearly indicate to the human user what property of the data led to the alert, and make it easy for the user either to take on the new controlling or monitoring condition or to dismiss the alert and return to whatever the user was doing before the alert signal occurred.

The different modes of perception have implications for the design of displays, as suggested in Table C-1.

Table C-1: Implications of the Modes of Perception for Display Design.



Perceptual Mode	Appropriate Display	Interaction required
Monitoring/ Controlling	Focus on network attribute being monitored or controlled, with context in background	***Monitoring***: Navigation ***Controlling***: Navigation plus means of influencing dataspace
Searching	More even display, perhaps with some increased detail near centre of area being searched. Focus on components of attribute sought.	Navigation only. Includes informational zoom and navigation in attribute space, not just screen space.
Exploring	Same as Searching, but perhaps with less concentration on specific attributes.	Same as Searching
Alerting	No display until alerting condition found. The minimally intrusive alerting indicator associated with area currently in focus.	Ability to shift easily to new focus on situation that led to the alert; ability to revert, dismissing the alert if false alarm.

As Table C-1 suggests, a display that has as its main purpose the user's ability to monitor or control some specific aspect of the data space will tend to be more focused and probably more sparsely populated than a display intended for search, with an Exploration display being even less focused and more densely populated.

These four modes of perception are properties of what we may call atomic tasks. What a user might describe as a "task", such as "determine the areas of likely vulnerability in the electrical infrastructure" is a complex amalgam of many atomic tasks. Most complex, or "molecular", tasks involve all the modes of perception, but an atomic task involves only one mode at any one moment.

Interactions with the Engines of the VisTG Reference Model are all in the Controlling mode. These tasks mostly can be considered atomic. They are performed because the user wants to see something in the data space that is not currently available in the presentation (or in the user's memory). However, the task that requires some interface interaction may be any mode except Alerting. The initiation of an Alerting process is not itself an Alerting mode function, but is a Controlling mode function; the user wants to see that a particular Alert has been set up, and seeing that it has not, acts to rectify the situation. Analysis of the task types concerned with engine interaction are of interest in themselves, but are independent of the tasks relating to visualising interesting aspects of the data space.

Questions Q1, Q2, and Q3 of the canonical questions of the VisTG Reference Model offer another approach to categorizing tasks in general, but it is probably more profitable to consider them within a particular domain than as generic questions. Most of this Annex C and the following Annex D are devoted to the major component of the Framework for Visualisation of Dynamic Networks, the description of the dynamic network domain. But first, we consider aspects of the user's interactions with the display, which remain valid no matter what the subject domain.

C.3.3 Interaction and Display Syntax

Human-Computer Interaction is a vast domain, all of which is implicitly incorporated within the Framework. The aspect of interaction important in the present context is its effect on the ability of a person to make sense of a complex display.

The construction of understanding of a computer's data space observation of its display has much in common with the development of understanding a person from observing them. One can understand much by simply watching the other person, but one can understand a great deal more by interacting with them. One can ask questions and consider the answers; one can request certain actions and determine whether they occur. Unless one can interact with the other person, one can judge only by observing what they do, making assumptions as to how those actions conform to convention to create a coherent hypothesis about the other person.

The process of scientific discovery follows a similar pattern. One can observe the changing world and make hypotheses about what might be happening behind the scenes to generate one's observations. Indeed, astronomers can do little more than that. But if one has the possibility of manipulating some aspect of the world and observing the results of the manipulation, one can compare and contrast different hypotheses that would equally well have explained observations if they had been made passively. The situation is similar when one tries to understand the implications of the contents of a complicated data space. If one is passively presented with views into the data space, one may develop hypotheses about it, but if one can manipulate those views, the nature of the content becomes much clearer. As an example, consider the display of a network as a set of nodes (displayed as blobs) connected by links (lines). When the display is static and flat, one can make sense of the network if it is not too large and not too many of the link lines cross one another. If the display moves as it would if the viewer were to change viewpoint, ambiguities resolve themselves, and the viewer can make sense of a larger network. The same is true if the network is displayed in stereoscopic 3D. But a further large improvement occurs if the user can manipulate the viewpoint in a 3D display, whether it is shown in projection on a flat surface or in stereo 3D.

One of the problems with the network representation is the lack of syntax. Syntax is the name for a convention about the way things fit together. In English, "Window open air in allows an" is intelligible, but "An open window allows air in" is much easier to understand. The latter uses the English syntax for word ordering, whereas the former does not, requiring the reader to use the semantic meanings of the words in order to puzzle out how they fit together. Displays similarly can be made easy or hard to understand according to whether their elements fit together in ways understood by the viewer – whether they conform to a display syntax known to the viewer.

Just as in English text the syntax of a document has many levels, such as the syntax of strokes within a letter, of letters within words, of words within phrases and sentences, of sentences within a discourse, so also does the syntax of a display have many levels, as suggested in Figure C-6. Within the 2D syntax level, for example, one element is the convention of displaying mutually relevant collections of content in windows that have frames whose components have certain interactive properties.

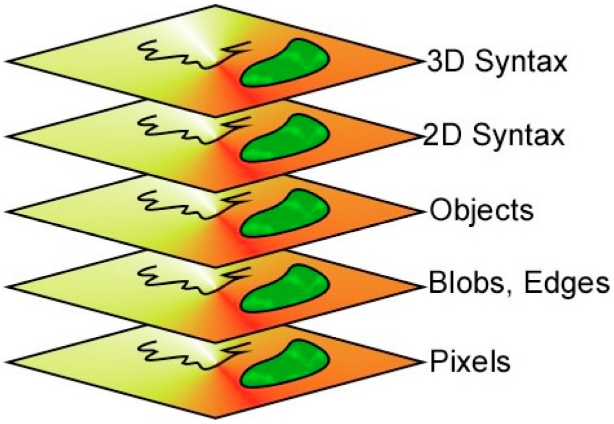


Figure C-6: Levels of Syntax for a Display on a 2D Screen.

The syntax of a display is nearly irrelevant when the display is constructed piece-by-piece by the end-user of the display – in other words, interactively. The process of construction allows the user to understand how the elements fit together. But if the resulting display is to be seen by another who did not witness the construction, whether its syntax is well formed matters a great deal. An individual who can communicate with the creator of the display can ask how this bit relates to that bit, but when the viewers are the audience at a briefing or lecture, the display syntax has to perform that function.

C.3.4 Presentation Context

The VisTG Reference Model is based around the twin concepts of perceiving what is, and the purpose for which that percept is required. Both affect the manner of presentation of any particular data.

The kind of display depends greatly on the use to which it will be put. One dimension of usage varies along the perceptual dimensions described above. Another dimension of usage depends on the relationship of the controller of the display to the viewing audience, who might not be the same person. If the user is working alone and building a display interactively as understanding increases, the kind of display required is quite different from the situation in which the user is using the presentation to brief an expert audience, and different again from the case in which the user is briefing a naïve audience, or is working collaboratively to solve a problem in concert with others who see the same or related presentations.

The viewing audience for the display affects the nature of the display because it affects the user-display interaction. If the user of the display is one person who controls the display, the user can build a very complex display bit by bit, while maintaining an understanding of how the elements fit together, even though the display becomes a tangle that no outside viewer could make head or tail of it. Such a constructed display often has no syntax, whereas a display intended for others to view must have an internal syntax that clarifies the relationships among the display elements.

The perceptual modes likely to be prominent for a display depend strongly on the audience. The audience at a briefing or a lecture cannot Search, except within the data being displayed by the presenter. They may be able to Explore, and often should, since what is presented early in a briefing is likely to be help the audience understand what is presented later. The better the audience can find linkages within and among successive displays, the easier should be their understanding of later material. This audience cannot Control at all, though they may be able to monitor if the presenter is showing streaming data in real time.

Consider now a different relation between the display controller and display viewer; a senior officer is instructing an expert to show something important for the officer's consideration of an ongoing situation. The situation is almost interactive. The officer does not actually manipulate the display, but does instruct the operator to act on the display depending on what is and has been shown. The difference between this and true interaction is bandwidth. The officer's communications with the display through the operator are much slower and more error-prone than the operator's interactions with the display on his own behalf.

Yet a third kind of interaction occurs when several people cooperatively work with a common data space. If the control of data selection and display method is allowed to just one of them, the situation is like a blend of the two previous examples – the officer with the operator and the briefing. If several of them can simultaneously influence the display, their actions are likely to lead to conflict, a condition in which the actions of two or more actors cannot simultaneously achieve what all of them want. In such a situation, conflict can be avoided if no two participants control the same aspect of the display at any one moment, or if they view the same data space on individual displays, possibly linked, and do not simultaneously act to change the same elements of the data space.

Table C-2 suggests the likely perceptual modes that might be used in eight different operator-audience-display relationships. The column headings of the table refer to the different ways in which one or more people viewing the display might interact with each other and with the display. "Coordinated" means that several users could manipulate the display but use a protocol that prevents them from interfering with each other's work; "mediated" means that the person controlling the display is not the actual user of the display. "Passive" means that the viewer has no control over what is displayed or how it is displayed.

Table C-2: Different Perceptual Modes Likely to be Used in Different Viewing Conditions.

	Interactive	Coordinated	Mediated	Passive
Single End-User	All Modes	N/A	(e.g. Officer Directing Operator) Explore, Search	(e.g. Reading) Explore
Multiple Simultaneous Viewers	N/A	(e.g. Planning) Monitor, Explore, Search, Alert	(e.g. Questions at a Briefing) Explore	(e.g. Lecture) Explore
Multiple Users Viewing Separately	N/A	Monitor, Explore, Search, Alert	N/A	Explore

The kind of display most suited for the occasion depends also on the kind of data to be displayed. IST-013 identified six dimensions on which data might differ, as shown in Table C-3.

Table C-3: Dimensions of Variation of Data (From Final Report of IST-013).

Acquisition	Streamed	Sporadic	
		Regular	
	Static		
Sources	Single		
	Multiple		
Choice	User-Selected Interactive		
	Externally Imposed		
Identification	Located		
	Labelled		
Values	Analogue	Scalar	
		Vector	
	Categoric (Crisp)	Symbolic	Linguistic
			Non-Linguistic
		Non-symbolic	Linguistic
			Non-Linguistic
	Categoric (Fuzzy)	Symbolic Non-linguistic	
		Non-Symbolic Non-Linguistic	
Relations	User-Structured		
	Source-Structured		

The details of these 19 different aspects of data are discussed in the Final Report of IST-013, and are only sketched here.

The cells of Table C-3 refer to elementary units of data. Most real datasets are heterogeneous, containing elements of different kinds. For example, a traffic flow dataset contains (relatively) static data describing the individual road junctions and the roads that connect them, together with streamed data about the traffic flow in different parts of the network. Of the traffic flow data, some may be located by the geographic coordinates of points of congestion, while others may be labelled, using identifiers such as "on the approaches to the Skyway". Some data may be externally selected, as when members of the public report conditions at particular places, while others may be user-selected, as when a central traffic reporter asks an observer in the air for "conditions on the approaches to the Skyway".

Within any dataset, whether it is static or dynamically varying, the requirements for presentation depend on how the presentation is to be used, and what about the data the user needs to understand. The Final Report of IST-013 (The "HAT Report") identified four dimensions of description for displays, dimensions that work with the descriptive dimensions of data to suggest broad outlines for effective display, as shown in Table C-4.

Table C-4: Descriptive Dimensions of Data Display.

Display Timing	Static picture	
	Dynamic Variation	
Data Selection	User-Selected Interactively	
	Algorithmically Selected	
Data Placement	Located	
	Labelled	
Data Values	Analogue	Scalar
		Vector
	Categoric	Linguistic
		Symbolic

As with the descriptive dimensions for data, the cells of Table C-4 describe elementary components of displays. Any complete display is likely to incorporate a heterogeneous variety of elements, unified by a display syntax that assists the user to relate those elements.

C.4 THE NETWORK DOMAIN

The data domain is one of the major variables in the development of a Framework from the VisTG Reference Model. It determines what the user could possibly perceive in the world, and thus limits the range of possible tasks. Once the parameters of the data domain are established, it becomes possible to refine the task description beyond the generic approach dictated by the modes of perception. Much of the following was discussed in the Final Reports of IST-013, IST-025, and IST-059, and is abstracted here. For more detail, refer to those reports.

Networks are often considered as abstract mathematical objects, stripped of some or their entire real-world context. The Framework restores that context as a component of the description of a real-world network, in two ways. Firstly, it introduces the concept of an “Embedding Field” (not to be confused with the mathematical concept of embedding). Secondly, it introduces the concepts, borrowed from linguistics, of syntactic, semantic, and pragmatic levels of description of networks.

C.4.1 Embedding Fields

An embedding field for a network may be a physical support structure that enables the network to operate. An example is the network of wired and wireless signal connections among computers that enables the TCP protocol network that connects some of the computers on the signal network, but not all of them. Likewise, the TCP/IP protocol structure enables the connected network of pages of the World Wide Web, and independently the network of accounts on computers defined by the recent passage of e-mail messages. As these three examples illustrate, embedding fields can be hierarchically related, and can form a tree structure, as shown in Figure C-7.

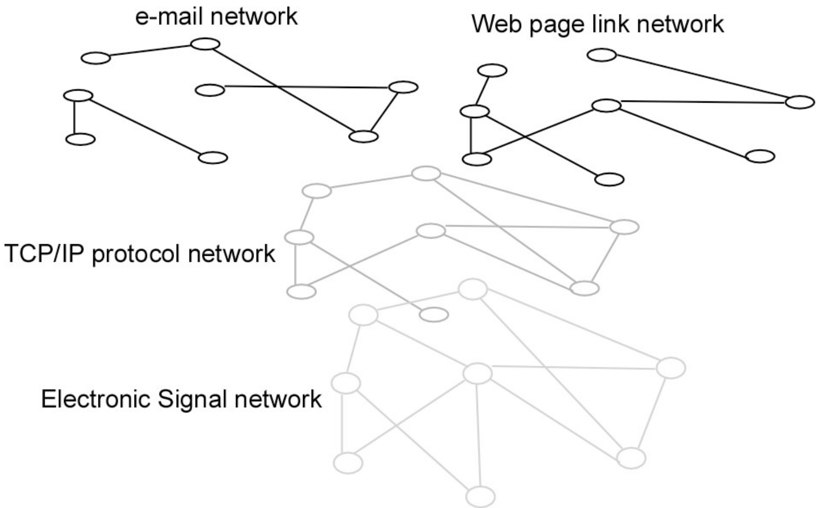


Figure C-7: A Schematic Example of Levels of Embedding Fields. Nodes represent computers or accounts within computers, while links represent network connections. Note that every link in a higher-level network also exists in its embedding fields, but the reverse is not true.

Another way and embedding field may support a real-world network is through influences from the world outside the network. For example, a road network exists in a landscape of towns, fields, hills and valleys. The character of these external factors determines where the roads should be built and how they evolve over time. The formal network consists of nodes that represent road junctions or other significant places, connected by links that represent the roads. These links may have different weights (a concept to be defined below), but they do not reflect the character of what is seen by a driver along the road. Such an embedding field is defined as the set of influences that matter from the viewpoint of the user of the network presentation software, not as the total set of influences that might be important to any arbitrary user.

C.4.2 Syntactic, Semantic, and Pragmatic Properties

In analyzing natural language, one discriminates among properties called “Syntactic”, “Semantic” and “Pragmatic”:

- Syntactic properties are quite generic. A word belongs to some grammatical category, but has no other properties. Adjectives modify nouns, verbs relate nouns by actions, and so forth. A word string such as the famous “Colourless green ideas sleep furiously” is syntactically well formed, whereas “sleep furiously green ideas colourless” is not.
- Semantic properties are more specific. Certain adjectives do not fit well with certain nouns, for example. Neither “Colourless” nor “green” is a semantically appropriate modifier of “ideas”, and if something is colourless it cannot be green. Semantic properties are timeless, in the sense that their appropriateness is determined within the language rather than within the context in which the language is used. “President Abraham Lincoln yesterday tweeted about his success in begging on the street” is semantically well formed. It seems wrong because we know that President Lincoln is long dead, Presidents, do not usually beg on the street, and Presidents are not expected to tweet. The wrongness is pragmatic.

- Pragmatic properties are those that link the language to the actual state of the world beyond the words themselves. It is pragmatically wrong that President Lincoln tweeted and that he begged on the street while President. A word string such as "Down there?" may be pragmatically meaningful without necessarily being syntactically or semantically well formed, because the pragmatic meaning depends on the real-world context of the moment, including the context of an ongoing conversation.

Consider now the analogous categories in the description of networks:

- Syntactic properties again are quite generic. A network consists of two kinds of entity, nodes and links that connect nodes. Nodes are junction points where links meet. Node properties apply to all nodes, and link properties to all links. A node has at least the property of a count of how many links originate and how many terminate at it. A link may be directed or undirected, and may have "weight" or "strength". Graph-theoretical analyses are based largely on the syntactic properties of the network.
- Semantic properties differentiate among classes of nodes and of links. Different kinds of links may connect specific types of nodes. For example a link of type "owns" may be directed from a node of type "human" to a node of type "house" but not in the reverse direction. Semantic properties are timeless in the sense that they do not depend on the changing circumstances of the world outside the network. They exist totally within the network. Social Network Analysis (SNA) incorporates semantic as well as syntactic properties of networks.
- Pragmatic properties are concerned with the way circumstances outside the network influence the network. Consider a road network. Syntactically, it consists of connections among junction points and potential stopping points (nodes with links to only one or two other nodes, meaning that paths do not branch at a stopping point). Semantically it consists of a variety of road types: gravel paths, six-lane expressways, toll roads, roads with different speed limits, and so forth. Pragmatically the roads lie on a landscape that may be scenic, hilly or flat, urban with the risk of pedestrian crossing traffic, the day may be foggy or there may be an expected major sporting event or concert in the area. All of these possibilities affect the way the traffic on the road network behaves, even though they do not form part of the network itself. No well-established system of analysis uses the pragmatic properties of networks.

Embedding fields may be categorized using the syntactic, semantic, and pragmatic categories:

- Syntactic properties are again generic. A Syntactic embedding field is a network that "supports" the network in question ("support" is not used with any standard mathematical meaning, but implies only that without the supporting network, the one of interest could not exist). A node in the network must correspond to a node in, or a sub-net of, the embedding field, and a link in the network must correspond to a path that consists of one or more links in the embedding field network.
- A Semantic embedding field is ordinarily a network. The semantic properties of the embedding field constrain the possible semantic properties of the network in question. If the Semantic embedding field is not a network, its properties nevertheless constrain the semantic properties of the network. A network may have more than one semantic embedding field, since many different circumstances may constrain its semantic properties. Which of these form part of "the" semantic embedding field for a user of a network analysis tool depends entirely on the user's needs (in the VisTG Reference Model, what the user needs to perceive for the task at hand).
- A Pragmatic embedding field consists of those influences from outside the network that affect the pragmatic properties of the network. Which of the indefinitely many possible influences are included within "the" pragmatic embedding field depend entirely on the user's needs of the moment. The examples of influences on the road network used when discussing the pragmatic properties of networks, above, would be elements of the pragmatic embedding field of the road network.

The separation of network properties and of embedding fields into syntactic, semantic, and pragmatic categories is useful, because it suggests what kinds of tools might be useful for analysing how the networks relate to the user's task. Syntactic properties and the effects of syntactic embedding fields are readily addressed mathematically, but it is normally the case that mathematical approaches fail when applied to pragmatic embedding fields and network properties. Semantic properties and embedding fields fall in the middle. Mathematical approaches are often useful (as SNA illustrates) but may be supplemented by less rigorous and more task-oriented approaches.

The categorization into syntactic, semantic, and pragmatic also may influence the character of useful presentations. If syntactic properties commend themselves to mathematical analysis, it makes sense that these properties would not be individually displayed, but instead, displays should represent the results of the analyses, and should assist the user to judge which kinds of analysis might be most suited to the task at hand. On the other hand, pragmatic properties, and the influence of the pragmatic embedding field, are more likely to be amenable to human visualisation, suggesting that the appropriate displays should be densely populated rather than oriented to the analysis of isolated objects.

C.4.3 Link Properties

What makes a network into a network rather than a set of isolated objects is the fact that the objects are linked in some way. The links may be conceptual – this is similar to that, or this person is that person's sister – or embodied in some tangible way such as an electrical wire or a pathway. Links may be capable of carrying "traffic", such as cars on a road, signal packets on a wire, or messages on a protocol network. "Traffic" has rather a broad definition. If two nodes are linked, and changes at the origin of the link cause some later effect at the termination of the link, the link carries traffic. Barring accidents, if a car enters a road that goes only from A to B, it will later arrive at B. If a signal pulse enters one end of a wire, it will later appear at the other end, perhaps modified. The transmission may be at nearly the speed of light, but it is not instantaneous. Some consequences of traffic transit time are considered in Annex D.

Instantaneous effects are not traffic. If A "owns" B, there is a conceptual link from A to B, but if A sells B to someone else, the link A®B instantaneously vanishes, both ends at the same time.

Whether a link carries traffic is only one of its possible properties. A frequently considered property is link strength or weight. But what might "weight" or "strength" mean? If the link does not carry traffic, weight is just a number at the syntactic level. At the semantic level, weight indicates that there is more or less of whatever the link signifies, whether it is degree of similarity between the entities represented by the nodes, the degree of determination of one node by the properties of the other, or some other semantic property. Again, it is just a number.

If the link carries traffic, "weight" can have a variety of meanings, several of them concurrently. Here are three:

- Link capacity: How much traffic can the link sustain?
- Link utilization: How much traffic does the link carry?
- Link availability: How likely is the link to be open to traffic?

Each of these three can be specified for any traffic-carrying link. They are therefore syntactic properties of the link, as is "weight". But which of them, if any, should be considered the "weight" of the link? It depends on the user's interests, and on whether the user even is concerned with specifying a single number to represent something that might be called "weight".

Links may have competing effects on the nodes they connect. For example, inhibitory input to a neuron influences the likelihood that the neuron will fire when excitatory input arrives. Information from a trusted source might reduce the likelihood that an analyst would believe contradictory material from a less well-known source. Such links might be considered to have negative weight, or to have positive weight but a negative connection to the receiving node.

In a dynamic network with varying traffic load, other link properties may specify aspects of the traffic variability. For example:

- Is the link closed to traffic regularly, irregularly, or not at all?
- Does the traffic consist of a continuous flow or of discrete items such as cars or signal packets?
- To what degree is the traffic flow predictable?
- What is the time delay between traffic entering the input of the link and exiting at the output?
- Is this delay fixed or variable, and if variable, to what extent is it predictable?

All these examples refer to the traffic on the link rather than to the link itself. The network exists only to support the traffic, and a link exists even if the link utilization is zero. But one can look at it from the other side, imagining that it is impossible to detect the links other than by observing the traffic. The network of e-mail connections is of this kind. Unless one has access to the address books of all users on the internet, one could not, in principle, know whether there exists a possible e-mail link between two randomly chosen users. The only way to define the social network of e-mail connections is to observe emails that are actually sent.

If the links in the network can, by some other criterion, be segregated into different classes, and those classes affect the values of the properties mentioned, then the properties have a semantic character. For example, a road network can usually be segregated into classes such as “expressway”, “arterial road”, “paved road”, “and gravel road “,” forest path”. The classes have different capacities for different kinds of traffic. An expressway is not expected to have pedestrian traffic, and a forest path is not expected to have motorized traffic.

The possibility, indeed the necessity, of delay between traffic entering the link at one node and exiting the link at its terminus, means that the concept of the link being open or closed to traffic is ill-defined. If the link is a road with a traffic light at its terminus, traffic can enter the link at any time, but can leave it only in pulses. Some entering traffic encounters no closure at the terminus, while other traffic finds itself stopped temporarily at the terminus. Is the link open or closed for this latter kind of traffic?

Semantic properties are widely variable, whether we are dealing with traffic-carrying or traffic-free links. Links in the so-called “Semantic Web”, which is concerned with the meanings of things, have quite a variety of different characters. Some varieties of networks in which semantic properties are defined are called “coloured”.

The pragmatic properties of links include anything that is affected by conditions outside the network. For example, a road may semantically be a six-lane highway, but pragmatically may not exist as a link because it is closed on account of extensive wildfires in the surrounding terrain. Pragmatic properties are the properties that are subject to dynamic variation, and therefore must be of primary interest when considering the visualisation of dynamic networks.

C.4.4 Node Properties

Syntactically, nodes are places where links meet. The basic syntactic properties of a node are the counts of how many links originate and how many terminate at that node (an undirected link does both). When analyzing networks or parts of networks, properties relating to the connections made by the links to and from a node, such as any of the many varieties of “centrality” become important, but such properties belong to the node as a component of the network rather than to the node in and of itself.

In a traffic-carrying dynamic network, nodes can have other syntactic properties. For example, a node may introduce delay between traffic arriving on one link and departing on another. There may be contingencies among the traffic on the different links. For example, in a Petrie Net, a node may not emit an outgoing element of traffic (a token) until it has received, say, two incoming tokens since the last one emitted. Such a node might also illustrate inhibition, if the arrival of a token from one specific link increased the required total of tokens arriving from the node’s other links. As another example, consider a traffic light in a road network. If traffic is accepted from one incoming link, it may inhibit the acceptance of traffic from a different incoming link.

If all nodes in the network have the same set of properties, these properties are syntactic. Semantic properties include all the varied properties that pertain to distinct classes of nodes or only to idiosyncratic individual nodes. The semantic properties of a node can have any level of complexity. If one considers a node in a dynamic network to be a place where the conditions and event arriving from in-links are transformed into conditions and events at the inputs to the nodes out-links, classes of nodes are distinguished by the algorithms that generate these transformations. In particular, the internal structure of a node could be an entire network with its own dynamic properties.

In a traffic-free network, the semantic properties of nodes are limited to the kinds of links that may initiate or terminate at a node. For example, a node representing a human may have an out-link of type “owns”, but a node representing an inanimate object cannot.

The dynamics of a traffic-free network are reflected in node properties only as changes in those properties. A link may change its status, the node may change its potential out links, as, for example might happen if a link of type “provided licence” changes from “false” to “true”, which would then allow the node to permit the existence of a link of type “drives” to a node of type “car”. The node type itself might change, as in the foregoing example, in which the node might change from class “non-driver” to class “driver”.

As before, the pragmatic properties of a node depend on conditions outside the network. If the person who acquired a licence happens to be in hospital, the corresponding node would semantically be of class “driver”, but pragmatically that class would be over-ridden, and the node would pragmatically be of class “non-driver”.

Pragmatic considerations can often over-ride the semantic class of a node, and indeed of any other network property, whether of links or of network structure, very much in the way that in object-oriented programming sub-classes may not only introduce methods with no counterpart in the parent class, but may also override methods that exist in the parent class and that would otherwise be inherited unchanged.

C.4.5 Regional and Global Properties

The properties of a network are defined not by the properties of its individual nodes and links alone, but by those properties in conjunction with the structural shape of the network. These emergent structural properties also have syntactic, semantic, and pragmatic characters. To a large degree, the character of a structural property depends on that of its component elements. If the structural property has at least one semantic component, the network property cannot be syntactic. Likewise, if at least one component property is pragmatic, the structural property must be pragmatic.

Even though the nature of a structural property is determined by that of its most specialized component (pragmatic is more specialized than semantic, for example), most structural properties studied in Social Network

Analysis (SNA) are syntactic. The various kinds of "centrality" and "between-ness" depend only on the existence or lack of a link between the members of all pairs of nodes in the network, or perhaps on the weight of a link that exists. Such properties are syntactic, belonging to every type of link or node.

Semantic analogues of such properties can be defined. Suppose, for example, that nodes come in two classes A and B and all links are directed, such that nodes of one class are linked only to nodes of the other class. An example might be infection through a mosquito vector. Humans can infect mosquitoes, and mosquitoes can infect humans, but humans cannot infect humans and mosquitoes cannot infect mosquitoes. Semantically, the nodes "mosquito" and "human" may be labelled differently, but they differ in no other properties except which of the two kinds of link they may have as in-links or out-links. The two classes of link are both semantically defined as "can infect", their difference being defined by their input and output possibilities. The inputs of one class are always from human and the output to mosquito, while those of the other class are the reverse.

Other semantically different kinds of link can connect the nodes in a universe of humans and mosquitoes. For example, a bidirectional link of type "friend of" can connect only human nodes, whereas a unidirectional link of type "has stung" can go only from a mosquito to a human. Whereas syntactically, there either is or is not a link from node A to node B (this statement will be modified when we deal with fuzziness), semantically, there can be many links of different kinds between any two nodes.

A link that exists semantically (and therefore also syntactically) may not exist pragmatically. For example, a link from mosquito to human of class "can infect" pragmatically does not exist if the human and mosquito are kilometres apart. Figure C-8 illustrates two road links on the East-West German border during the Cold War in which the pragmatic external influence is political. In each case a road link exists syntactically and semantically but not pragmatically, because a person attempting to use the link without official authorization was liable to be shot.

Figure C-8: Two Roads that Crossed the East-West German Border During the Cold War. In each case the link in the road network exists syntactically and semantically, but not pragmatically.

As with all pragmatic properties, this pragmatic effect cannot be analysed from data about the network, but has to be discovered by separately examining the environment of the network. Sometimes, a pragmatic effect may be due to influence from a different network that is ordinarily analyzed separately. For example, when considering the road network, one rarely considers the electric power network, but if the power goes out, the traffic lights and street lighting go out, drastically changing the traffic flow pattern on the road. This kind of interaction is considered below under “Junction Nodes”.

C.4.6 Broadcast and Point-to-Point Networks

When we think of a network, we usually think of a collection of entities that are pair wise connected. Node 23 might be connected to node 48, node 174, and node 61. The node could, in principle, identify the destination of every item of traffic it emits, and treat traffic destined for node 61 differently from its treatment of traffic for node 174. Likewise, receiving nodes could, in principle, know whence each element of traffic came, and treat it appropriately to its source.

Not all networks have this point-to-point character. Some nodes may broadcast their traffic, to be collected by whichever nodes might be listening to it at the moment. The receiving nodes may be able to know the source of the traffic, as we know where the video stream we are watching came from, but the source may not be able to know who is listening or being influenced. The reverse situation is also possible, though less common. The receiver may not know the source of a message, while the source knows very well where the message was destined. Many wartime deceptions us the strategy of making false information seem to come from a reliable source.

Broadcast networks, especially those in which the receiving nodes choose from which sources to accept traffic, have a very labile structure. Consider the network of which homes are listening to which TV stations at any one moment, and then contrast it to the same network five minutes later, or worse, after the next even hour. Whereas a link in a point to point network must be constructed, with some deliberation and choice if not real effort, a link in some broadcast networks is so readily made or broken that it is hard to say it is constructed at all. On the other hand, there are some broadcast networks in which reception is obligatory for all nodes within range of a source. One cannot decide to be unaffected by smoke from a fire that envelopes one.

Much of what we say below about network dynamics assumes that the networks are point-to-point, or at least that if they are not, at least the links have certain stability within the time-scale of interest to the analyst.

C.4.7 Traffic Dynamics and Networks of Influence

Much of the preceding discussion has been about the links of a network and the traffic on those links. However, an important class of networks has no overt traffic and no changes in link structure, yet has dynamic variation. These are networks in which the states of the nodes may change, and the state of one node may influence the state of another. Such networks may be called “networks of influence”.

Influence networks are not truly traffic-free. The influence may be carried by chemical messengers, electric messengers, word-of-mouth messages, casual observation, or anything else, but there must be some way in which a node can detect the state of some neighbour. Indeed, the very structure of the network is defined by the existence of means by which the state of one node affects the state of another. The difference between networks of influence and the kind of traffic-carrying networks discussed earlier is that the earlier discussion treated the nodes as static elements, the interest being concentrated on the traffic flow and the link structure. In considering networks of influence, traffic flow cannot be ignored, but the emphasis is on changes in node states.

Provided that the network is not acyclic, and has at least one place where the effects of changes at node A later influence the state of node A, the network or sub-network has three possible dynamics: dissipative, oscillatory, and chaotic. If the dynamic is dissipative, then in the absence of input from some external source, the states of all the nodes will eventually stabilize to some invariable condition. If it is oscillatory, then in the absence of external input each node will cycle endlessly through a pattern of states. Finally, if it is chaotic, each node will go through a never-ending pattern of states that might seem periodic for a while but that would never repeat exactly.

Without any change in the network parameters, networks can have several different dynamic states. Which one is manifest may depend on the initial conditions. The same network could at one time show dissipative behaviour, several different oscillatory patterns, or different chaotic behaviours, and might move from one regime to another on receipt of particular kinds of input. Depending on the extent of external input, the actual pattern of node states might be dominated by either the external input or by the intrinsic dynamic of the network or sub-net.

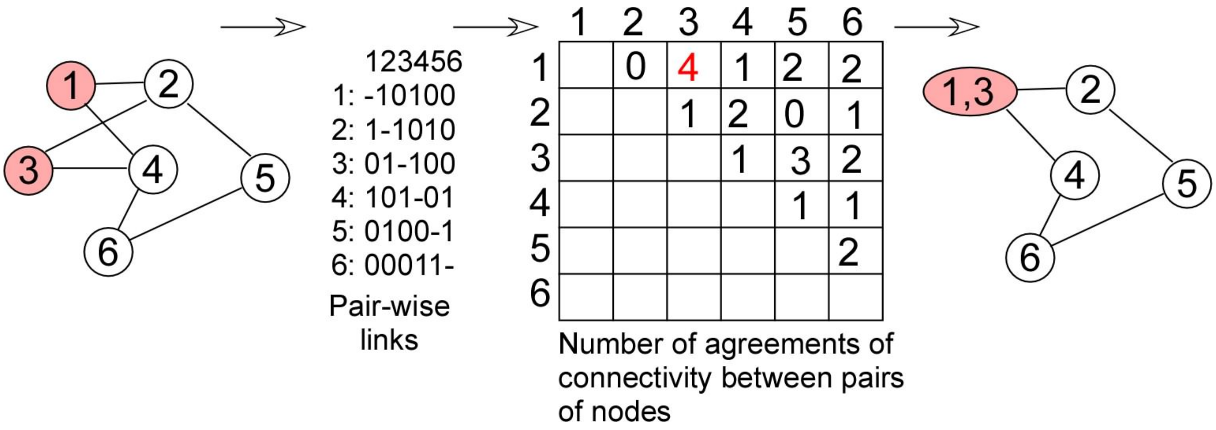
Regimes of behaviour are defined by what are known as “basins of attraction”. If the initial condition is with a basin of attraction, and no further external input occurs, the behaviour will eventually arrive at an attractor, which may be a single point (a dissipative system), an oscillation, or a chaotic “strange attractor”. The transition between basins of attraction may be caused by external inputs or by changes in some external condition that influences the properties of the network elements (the pragmatic embedding field).

Sometimes changes of regime may be signalled by changes in the behaviour of observable states of some node or link. Many of these possibilities have been described by Scheffer et al. [8], who point out that it is often impossible to analyze the dynamic possibilities of any but the simplest net. Nevertheless, no matter how massive or complicated the network, there are certain characteristic behaviours that provide clues that changes of regime might be in progress. Their discussion applies equally to networks that overtly carry traffic and to networks of influence, in which the traffic is less apparent.

C.4.8 Hypernodes

The concept of Hypernode was introduced to IST-059 by Bjørke, [3]. A simple way to describe a Hypernode is to call it a collection of nodes that have more in common with each other than with other nodes in their network. The index of similarity is information-theoretic: to what degree the descriptions of two nodes within a Hypernode are mutually redundant.

At the syntactic level, one characteristic by which the similarity of nodes can be measured is the pattern of weights of the links that connect a node to all the other nodes in the network (a non-existent link has a weight of zero). Figure C-9 shows a trivial example in which the linkage patterns of nodes 1 and 3 have more in common than do the linkage patters of any other pair.



123456
1: -10100
2: 1-1010
3: 01-100
4: 101-01
5: 0100-1
6: 00011-

	1	2	3	4	5	6
1		0	4	1	2	2
2			1	2	0	1
3				1	3	2
4					1	1
5						2
6						

Figure C-9: Construction of a Simple Hypernode. (a) A trivial network; (b) The connectivity matrix, ignoring self-connections; (c) The number of agreements as to whether two nodes are or are not linked to each other node, showing that nodes 1 and 3 have the greatest number of connections and non-connections in common; (d) The reduced network, showing a hypernode consisting of nodes 1 and 3.

What do the nodes constituting such a syntactic hypernode have in common? Their means of influencing the rest of the network, in this case the fact that both nodes in the hypernode connect to nodes 2 and 4, and to no other point in the rest of the network. In more complicated networks, the reduced network can be subjected recursively to the same process to create hypernodes of progressively lower commonality, always ending with the entire network becoming one giant hypernode. The benefit of this particular construction is that a sequence of connected hypernodes identifies a path in a large network more readily than in the raw network.

Even at the syntactic level, reduction of a network by hypernodes can suggest issues that an analyst might want to investigate. For example, in the network of Figure C-10, two hypernodes have been abstracted. The contributing nodes are keyed by colour in the figure only for convenience in discussing them. The red nodes in the original network diagram are similar in that each is linked to all the blue nodes and to the yellow node. The blue nodes are similar in that each is linked to both of the red nodes, and one of the blue nodes is linked to the yellow node. At the first stage of hypernode construction, using the strictest criterion of similarity, the blue node connected to the yellow node would be outside the blue hypernodes, but with a slight relaxation of the similarity criterion, the hypernode structure would be as shown, with a red and a blue hypernode strongly linked, and connected to the rest of the network through the yellow node.

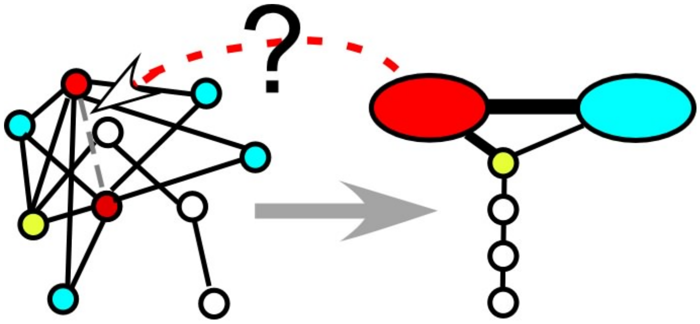


Figure C-10: A Hypernode Reduction. The analyst may want to know whether a hidden connection exists between the red nodes.

In the hypernode reduction, the two red nodes are so similar that an analyst might wonder whether there might be a hidden link between them, and if so, why the link is hidden. In any case if the links represent influence in some way, there is the potential for the two red nodes either to assist one another, to conflict with one another, or if they pragmatically do not know of each other, to disturb each other's influence erratically. The Hypernode concept becomes more interesting and useful at the semantic level. Whereas at the syntactic level, two nodes increase their degree of similarity if they are connected to the same node, at the semantic level this is true only if their links are of the same semantic kind. In the previous example of Figure C-9, if the link between 1 and 4 is a directed "owns" link such that 1 owns 4, and the link between 3 and 4 is a directed link such that 3 is "part of" 4, the fact that both are linked to 4 has very little influence on their measure of similarity. However, it can be useful to consider the syntactic and semantic properties together. Even though semantically 1 and 3 have no common relationship to 4, yet the semantics indicate that 1 probably "owns" 3 even though there is no indicated syntactic link between them. A syntactic Hypernode structure may suggest the possibility of common semantic properties among the nodes within a hypernode. In Figure C-10, are the blue nodes of the same semantic class and the red nodes of another? The syntactic hypernode structure cannot answer that question, but it may lead the analyst to ask it.

C.4.9 Temporal Variation and "Network Objects"

At the pragmatic level, the hypernode structure becomes dynamic. External conditions may influence some parts of the network differently from other parts. For example, a power outage may affect only part of a city, changing the properties of some of the nodes in the road network, thereby changing the structure of similarities across the nodes. Display of the changing hypernode structure might be useful in deciding strategies to take advantage of the changing situation. Temporal variation has another, very important consequence. In our everyday world, things that consistently move together usually are perceived as belonging to a single object, even though they may be physically separate. The legs and arms of a chair may not connect (though they may), but because they move together, along with the seat and the back, the coordinated set of movements allows us to see a new object, the chair. Several chairs set around a table are a configuration often seen, but it is a looser connection than that among the components of a chair. Such a configuration is sufficiently common to have a name such as a "dining suite". And on the other side, the legs and arms can become detached from the rest of the chair under circumstances usually unfortunate, leading them to be seen as individual objects as well as parts of the chair.

The creation of a sequence of tighter or looser hypernodes has the same character as that of the dining suite (loose connection among the components), chair (tight connection), and the legs, arms, seat, and back (very tight connection). Temporal variation allows for, and indeed determines, this hierarchy of "objectness". Within a network, likewise corresponding variations of properties over time can be used as the criterion for the construction of tighter or looser hypernodes that could be said to represent "network objects". The division of networks into objects allows the network world to be perceived more coherently than is possible considering only the individual network elements, just as it is easier to perceive a dining room suite than to perceive each individual chair leg and arm.

C.4.10 Layered Hypernodes

At the semantic level, a new concept comes into play, that of "layered" hypernodes. The hypernode construction procedure depends on comparison of the similarity of properties across nodes. However, if by "node" is meant some identifiable entity in the physical world, the number of its properties is undefined. At any one moment only a few of its possible properties are known to the user, and even fewer are of interest.

The hypernode construction procedure can be used with a single property, as at the syntactic level, or with many properties at once to provide a more stringent measure of pair wise similarity. If a single property is used, its values must have a sufficient range of variation to allow meaningful differences in the measure of similarity. In other words, the uncertainty associated with the property must be sufficient to allow for a range of mutual information measures (Information Theoretic concepts are discussed in Annex D). Figure C-11 suggests how real-world entities that constitute nodes of a network may exist in different hypernodes if the hypernode structure is based on different properties.

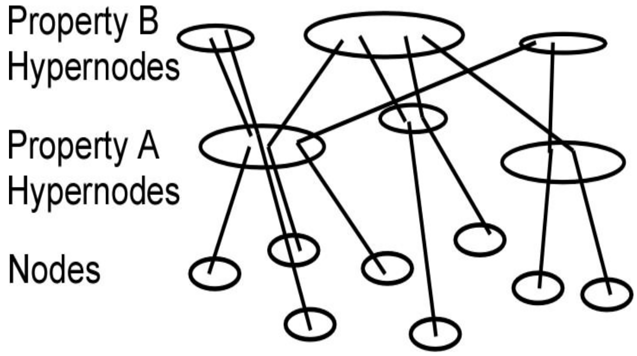


Figure C-11: Hypernodes can be based on different properties, and the nodes may partition differently over the structures based on different properties.

Figure C-12 suggests how a multi-level hypernode structure might be analytically useful. If it is used with several different properties individually, and some of the hypernodes based on the different properties contain the same set of nodes, it is reasonable to assume that these nodes have something in common that might later be used to infer unknown property values of one from the known property values of others that share hypernodes at different levels, as do the nodes marked by red, blue and green traces through the levels of Figure C-12.

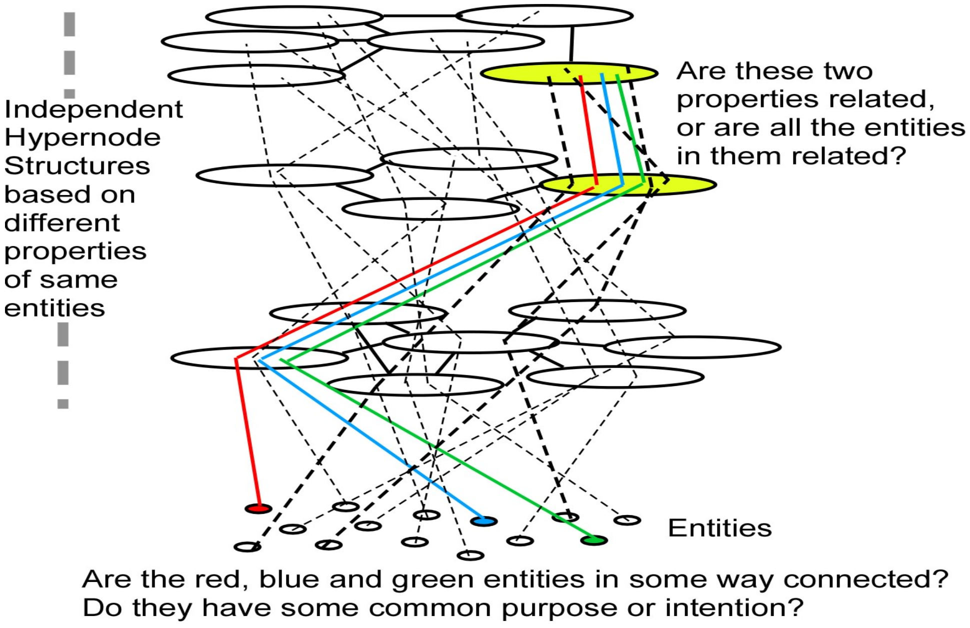


Figure C-12: Hypernode structures based on different individual properties may suggest covert relationships among the entities that are considered nodes.

Of course, the different properties might be combined to create a single measure of mutual information over the entity pairs, leading to a flattened hypernode structure, but to do this would obscure difference among the structures based on particular properties, differences that might be important to the analyst.

C.4.11 Hypernodes and Network Additions

The different levels of multi-level hypernodes are defined by different properties. Entities that agree with each other on several levels may well have other properties on which they are similar, and this includes entities that are not presently nodes in the network. For example, consider the growth of a local gang. A child who lives in the neighbourhood (similar geographic property), has parents of similar backgrounds to those of several gang members, has similar academic achievement, and so forth, is more likely to join that gang than to join a group whose members have different geographic, social, and academic properties. The concept of Hypernode can therefore be extended into the pragmatic embedding field of a network, to influence the likelihood of future changes to the network.

Many naturally grown networks are scale-free in structure. At the syntactic level, such a structure occurs if a new node attaches to each old node with a probability proportional to the number of links already connected to that old node [2]. Such a mechanism, called “Preferential Attachment”, works, but has no obvious underlying reason other than that nodes with many links are more exposed to the outer world than are nodes with fewer links.

Why should one node have more links than another? A simple reason is spontaneous symmetry-breaking. If we assume an initial node in which every node has the same number of links to the others and use any probability-based method of adding new nodes, then after the first new node is added, some of the old nodes have more links than do the others. Using the Barabsi and Albert Preferential Attachment (PA) mechanism, the next new node will be more likely to link to the nodes to which the first new node linked, increasing the disparity. After a number of new nodes have been linked in, the original symmetry of the network will have been completely obscured.

The PA mechanism asserts that nodes with many links are more exposed to the external environment than are nodes with few links, without offering a reason why this might be so. Such reasons are to be found at the semantic and pragmatic levels. Papadopoulos et. al. [4] showed that incorporating similarity between nodes in the network and newly joining nodes can also result in the development of scale-free networks.

Similarity of properties of nodes within a network is the criterion for assigning nodes to the same hypernode, and similarity of hypernode assignment across different properties defines similarity of entities in terms of those properties of interest to the analyst/user. Entities not already members of the network have no syntactic properties in common with entities already connected, but they may have semantic and pragmatic properties in common with network members. If the analyst knows enough of those properties for entities within and outside the network, the known properties can be used to construct multi-level hypernode structures distinct from the hypernode structures constructed from network structural properties.

The analyst may wish to judge whether a particular entity is actually in the network, linked in unknown ways, or may wish to judge how the entity is likely to link into the network if and when it joins. If the multi-level hypernode structure has incorporated the new entity into the same entity set on several levels (as with the coloured lines in Figure C-12), the analyst may legitimately conclude that if the entity is part of the network, its linkage structure will be like that of the other entities in the same hypernodes at the different levels, or if the entity is not yet part of the network, its linkage structure will evolve to be like that of the others when it does join.

Papadopoulos et. al. [4] has offered a specific mechanism for determining to which old nodes a new network node is likely to link. Their mechanism seems to complement the hypernode construction mechanism, inasmuch as they leave open the choice of similarity measure. The multi-level hypernode procedure offers a similarity measure, and at the same time proposes the link structure toward which the new node will evolve rather than just asserting that it will instantly attach to specific network nodes.

C.4.12 Stigmergic Properties

A stigmergic system is one in which the environment retains some change consequent on an earlier event, and that change affects subsequent behaviour of elements in that environment. A network on a stigmergic embedding field is one in which the traffic to or from a node leaves something behind that influences the behaviour of some other node an indeterminate time later. The long-term potentiating of synapses in a network of neurons is of this kind. Giving someone a piece of information that influences the interpretation of later information is also stigmergic. (Quoted from Annex D of the Final Report of IST-059).

Traffic on a link can alter the properties of the link for future traffic. For example, the passage of a vehicle over a muddy road creates ruts and may increase the muddiness of the track. The ruts make it easier to steer along the road, sometimes making it unnecessary to steer at all, while the enhanced muddiness may make it harder to make progress along the road, changing the delay characteristics of the link, and perhaps severing the link entirely.

The existence of stigmergic properties is important when considering network dynamics, since they create the possibility of a feedback loop between network traffic and network structure. The feedback loop affects both the semantic and the syntactic properties of the set work viewed at any single moment, while the loop itself depends on the pragmatic properties. The initial muddiness of the road is influenced by the pragmatic embedding field (recent rainfall, drainage properties of the environment, for example), and its actual pragmatic muddiness changes over time. Semantically, the link may change its nature from country road to difficult track, and syntactically it may change weight or even cease to exist. When stigmergy enters the analysis, all levels are affected.

C.4.13 Junction Nodes

Much of this section is based on the report of the Junction Node Working Group at the 2011 Vancouver NX Workshop (and presented in abstract in Chapter 2). When the behaviour of one network may influence conditions on another, the possible effects may be analyzed by treating the two networks as a single network in which the semantic properties of the nodes of one sub-network include either that they belong only to that sub-network or that they also link to nodes of the other sub-network. Since the two sub-networks might well be analyzed separately, with their mutual influences being considered pragmatic, only a small proportion of the nodes of either sub-network will have links to the other.

When analyzing the mutual influences between networks, it may be convenient to consider the roles of a cross-linked node within its primary sub-network separately from its role as a connector, by conceptually splitting the node into two parts. The abstract part that communicates with the other network is called a “Junction Node”. The Junction Node abstraction may or may not carry traffic. For example, consider the following example that involves three networks.

The three networks in the Working Group’s example are train, bus (as public road transport) and road (as private vehicle transport). The nodes in the train network are train stations; in the bus network they are bus stops and bus stations; in the road network they are any place where car drivers and passengers might want to get in or out of the car. The three networks are traffic-carrying, where the unit of traffic is a single person.

An integrated bus-train terminal is a physical entity that connects the three networks, as suggested in Figure C-13. At the terminal, a person may transfer between networks.

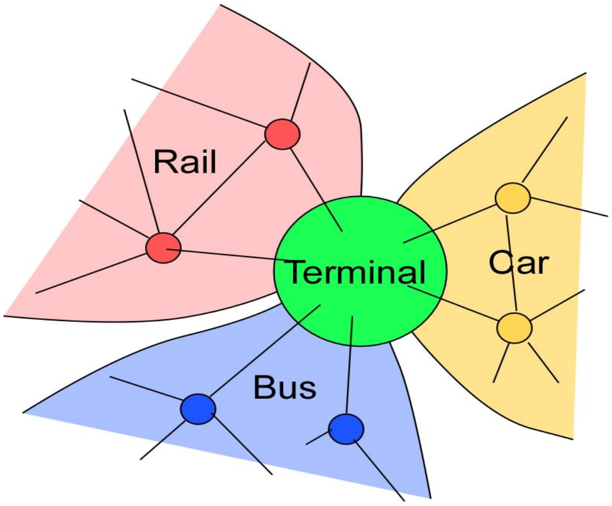


Figure C-13: Three Networks that Share a Node in Common.

Syntactically, the terminal is a single node that has links to nodes in each of the three sub-networks. In each of the sub-networks, the nodes are semantically “train station”, “bus stop”, or “point of interest”. The integrated terminal is none, or all, of these. It is of its own semantic class.

Although for traffic analysis at a syntactic level, the terminal has the same property as all the other nodes, namely that the sum of people entering the node must equal the number of people leaving over any sufficiently long interval, yet it has separate properties that are semantically and pragmatically better shown by separating out three abstract Junction nodes, as shown in Figure C-14(a) and Figure C-14(b). Figure C-14(b) illustrates that the junction nodes are not properties of the node labelled “Terminal” in Figure C-13 and Figure C-14(a), but describe the ways the behaviours at the railway station, the bus stop, and the car parking place influence one another. If a train is late, does the connecting bus wait for it? How reliable is the bus schedule, and how does that affect the timing of car arrivals at the parking lot? These are the kinds of questions whose answers may be properties of the junction nodes. Rather than attributing the possible effects of states at a node and link on what happens to a different network to the node or link, the influences are abstracted into the junction nodes.

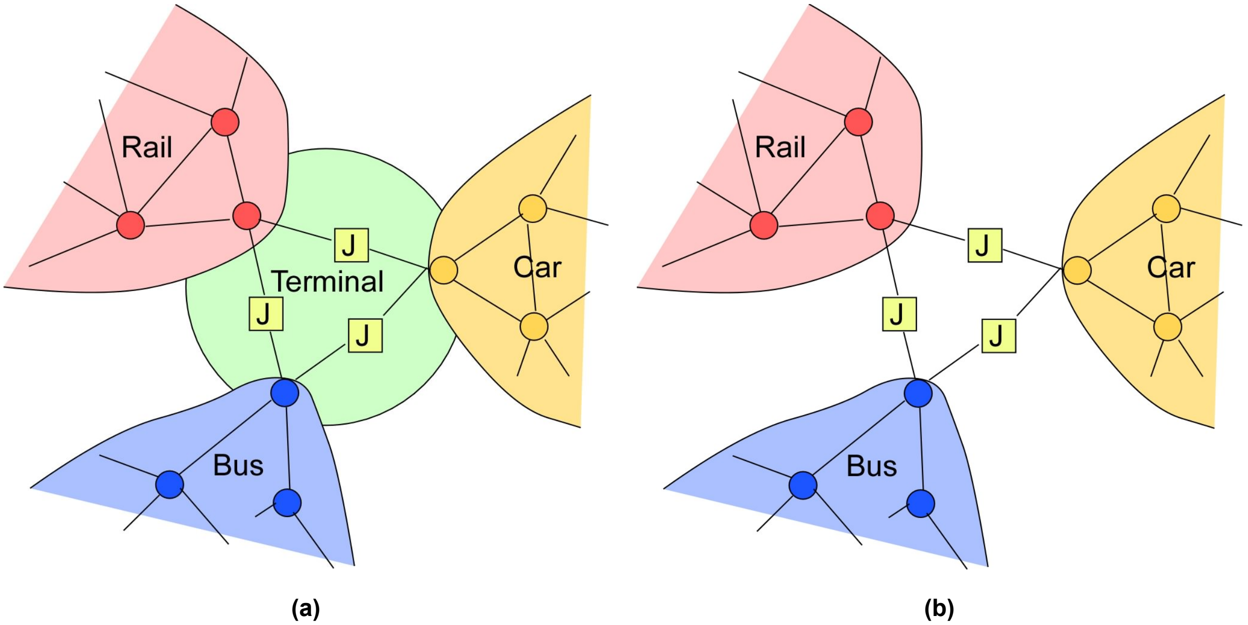


Figure C-14: (a, left) Separating zone of mutual influence into junction nodes; (b, right) There may pragmatically be no physical transport terminal, but only nearby train station, bus stop, and car parking lot. The Junction Node properties are the same.

Finally, junction nodes need not be associated with individual nodes within the interacting networks. If snow generates widespread traffic delays on the bus and car networks, no single node on either network is truly responsible, even though the direct influence is through the specific bus stop, car park, and train station. The same effects will be found at every place these networks interact. Accordingly, the junction nodes may be shown as existing between networks, rather than between individual nodes of networks, as in Figure C-15.

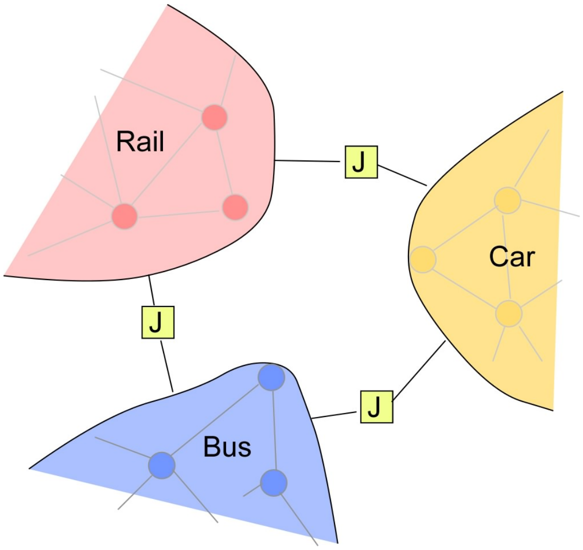


Figure C-15: Junction nodes may connect whole networks, rather than simple nodes of the different networks.

Another example in which the influence is between networks rather than between network nodes might be the influence of the electrical power network on the city road network. A power outage blacks out the traffic lights, snarling the traffic city-wide. In turn, the widespread traffic tie-up inhibits the repair crews from fixing the outage.

Although it is legitimate to show junction nodes as connecting one network to another, the actual influence obviously is between individual elements of the two networks. Failure of the electrical supply to one traffic light is logically, though not necessarily in practice, independent of power failure to another traffic light. By showing the junction node as connecting two networks, we combine several (possibly many) similar node-to-node connections into one channel of influence. Considered microscopically, each individual traffic light is affected by the power outage, but to treat them individually would obscure the global nature of the effect.

C.4.14 Network Junctions as Hypernodes

Although it is legitimate to show junction nodes as connecting one network to another, the actual influence obviously is between individual elements of the two networks. Failure of the electrical supply to one traffic light is logically, though not necessarily in practice, independent of power failure to another traffic light. Properly speaking, the effects should be symbolized by myriads of junction nodes, each with the property of connecting one traffic light to its own road network node. By showing the junction node as connecting two networks, we combine several (possibly many) similar node-to-node connections into one channel of influence.

The solution to this opposition between accurate representation and understandable representation is to recognize that a network junction node is very like an ordinary hypernode. In fact, if the large set of individual junction nodes is subjected to the hypernode construction procedure, the result would be a single hypernode that included all the junction nodes affected by the power outage. For this construction, only two properties of the junction nodes would be used:

1) The property that the node connects the power supply to a traffic light; and
2) The property that the power is out.

The resulting hypernode is the network junction node. By adding other semantic properties to the similarity measure used in the hypernode construction process, such as semantic distinctions among traffic lights on major and minor roads, progressive loosening of the similarity criterion in the hypernode construction might produce a sequence of junction hypernodes between the all-or-none categorization of element-to-element and network-to-network junction nodes.

C.4.15 Syntactic, Semantic, and Pragmatic Properties of Junction Nodes

At a syntactic level, networks are defined simply by the connectivity among undifferentiated nodes. Some sub-net may be defined by being more strongly connected among its member nodes than with nodes outside itself, but that definition is often ill-defined. At this level, the abstraction of a junction node has little relevance. It is quite probable that semantically distinct sub-nets such as the electrical supply system and the road network, or the rail network and the bus network, are much more linked within themselves than between nodes of one and nodes of the other, but this is not assured.

At the syntactic level, the insertion of a junction node simply breaks a link to insert a node between its two terminals. The inserted junction node has no effect on syntactic link properties. However, if the sub-nets between which the junction nodes are inserted are more tightly interconnected than their connections to the other net, the inserted junction nodes will have high "between-ness", no matter which of the many measures is used, and are also likely to have high centrality. As such, they represent points of considerable interest when an analyst considers matter such as network vulnerability or susceptibility to external influence. For example, to act on either sub-net, one might attempt to turn the abstract junction node into a concrete node by inserting a physical agent. The bus-train junction might be disrupted by installing a fence across the physical path from train station to bus-stop, or enhanced by installing a moving sidewalk between them and instituting a scheduling policy that facilitates rapid transfers.

At the semantic level, the place of junction nodes is much clearer. They connect two sub-nets that consist of nodes and links that are distinct in at least one semantic property, such as having links made of copper wire in one sub-net and of concrete pavement in the other, or of carrying traffic in buses, or in trains, or in private cars. Junction nodes therefore have at least one property not shared with nodes in either of the connected nets, and that is the paired semantically differentiating properties of the connected sub-nets. A junction node that connects a bus stop with a railway station is semantically distinct from one that connects a railway station with a car parking lot.

Pragmatically, a junction node has properties that depend on the current state of the networks it connects. If the transit authorities have arranged a schedule that is intended to allow for easy transfer between arriving and departing buses and trains, the junction node is pragmatically different from what it would be if it had no such scheduling property. If there is snow that disrupts bus service but not train service, the junction node is again pragmatically altered.

Semantically and pragmatically, the definition of a junction node includes a description of how events and conditions in one of the sub-nets influence events and conditions in the other. Junction nodes are therefore as directed as the links into which the "junction" abstraction is inserted. However, whereas it is often proper to substitute an undirected link for two opposed directed links, this is unlikely to be a proper substitution for the corresponding junction nodes. The kind of influence the electrical power network has on the road network is quite different from the influence in the other direction. Junction nodes are therefore almost always intrinsically directed.

C.4.16 Link-to-Link Influences

The cross-influence between two networks of different kinds need not be from node to node. It could be from link to link. For a particularly clear instance, consider a swing or lift bridge that carries a road over a river. Tower Bridge in London is a well-known example. When the bridge is available for road traffic, it is unavailable for river traffic except for small boats without tall masts. Dynamically, the road link changes from existing to not existing, and the same is true for the river link. The cross-influence happens because when one of the links exists, the other does not. Of course, such link-to-link influences need not be all-or-none.

How can link-to-link influences be represented by junction nodes? One way is to insert into each of the two links a virtual node that has the function of passing or impeding traffic. Such a virtual control node eliminates the requirement to consider the links as passing into and out of existence. A junction node, which is always a virtual construct, can then be connected between the two virtual control nodes, as shown in Figure C-16.

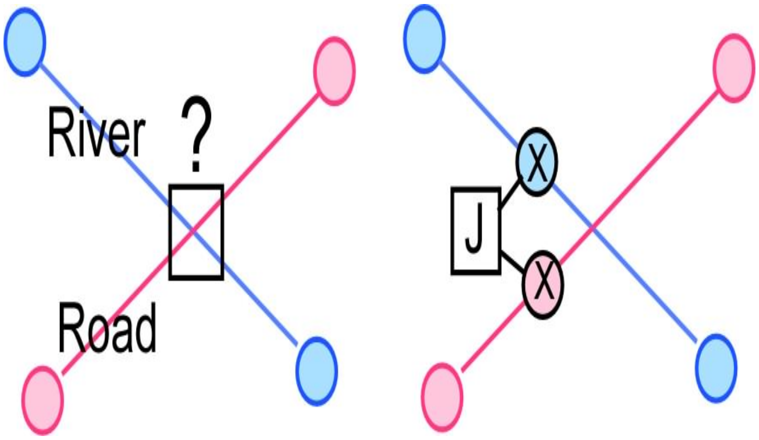


Figure C-16: A road crosses a river by a lift bridge, so that if road traffic can pass, river traffic cannot, and vice versa. The cross-network influence is between links, not between nodes. However, mutual influence between link states can be represented by a junction node connecting two virtual nodes that implement the state changes of the links.

Even though the individual networks may have been described purely syntactically, the fact that the mutually influencing links behave differently from all the other links in either network introduces semantics into the description of the combined super-network. The virtual nodes take on this semantic load, as they behave differently from all the other simple nodes. There is, therefore, no conceptual difference if the individual networks are considered semantically, with a rich variety or properties and behaviours for their nodes and links. The introduced virtual nodes take from the semantically rich links only those properties affected by the inter-network influence, and only those two-network properties are represented at the junction node.

C.4.17 Fuzzy Nodes and Links

In most considerations of network properties, a pair of nodes either is or is not connected by a link. In real life, the situation is not always so clear. Consider the two scenes in Figure C-17.

Figure C-17: Two Links in a Road Network. But are they equally to be considered links? It depends on the user. For a nature-loving hiker, the road in the left picture is a much better link than the one on the right, whereas the reverse is true for a logistics officer. (Photos by the author).

Figure C-17 shows two roads, each of which could be travelled by motorized vehicles, one more easily than the other, and both of which could serve as hiking routes, one more pleasantly than the other. A logistics officer (not serving under Hannibal) would probably not consider the left-hand road to be a link in the road network, whereas a hiker who enjoyed the peace and quiet of nature would not consider the right-hand road to be much of a link. Both might think a two-lane paved road through a forest to be an acceptable, but not perfect, link.

The example suggests two things. Firstly, it suggests that the existence of a link is a graded phenomenon, not all-or-none, and secondly it suggests that the grade assigned to the link is not absolute, but depends on who is doing the grading and the grader's purpose for the link.

Graded existence is a signal that it might be useful to consider fuzzy mathematics. Fuzziness is often confused with probability; indeed, a Google Scholar search for "fuzzy networks" during the life of IST-059 produced one paper, and that paper (not referenced here) improperly treated probabilistic variables as fuzzy.

C.4.18 Distinguishing Between Fuzziness and Probability

To clarify the distinction between fuzziness and probability, consider the height of a man. He is measured as being 185.3 cm (about 6 ft 1 in) tall. His height is known to the millimetre. Is he "tall", "normal", "very tall", and "quite short"? The answers (for this writer) are respectively "yes", "pretty much", "sort of", and "no". There are many descriptors for a man's height, of which these four are a sample. Each descriptor defines a class of tallness to which the man might belong. Once given his exact height, it is possible to say whether the man is a member of the class. However, as the example illustrates, membership may not be all-or-none. The 185.3 cm man clearly belongs in the class "tall" (unless he is a professional basketball player). But he only partly belongs in the classes "normal" and "very tall". If we added a class "moderately tall" he would clearly belong in that class. Membership in one class does not preclude membership in another.

The "fuzzy" approach assigns class membership a value between zero and unity. This man would have a fuzzy membership value of near 1.0 in classes "tall" and "moderately tall", near zero in the class "quite short", and say 0.7 in "normal" and 0.3 in "very tall". None of this is probabilistic, though different people might assign different numeric values to the man's membership in the different classes. Figure C-18 suggests the way men of different heights might be assigned membership in the different classes.

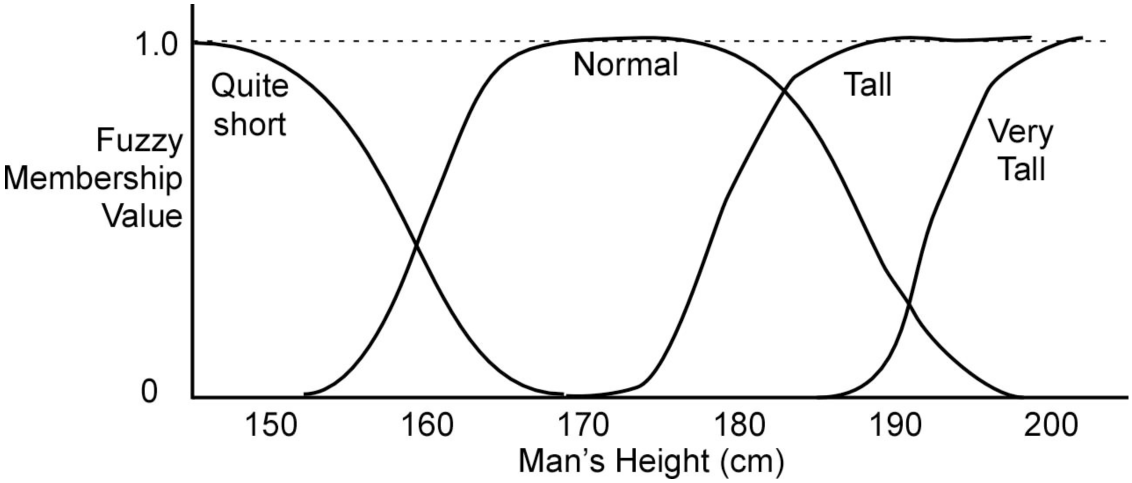


Figure C-18: Fuzzy Membership in Different Classes for a Man as a Function of His Height.

Contrast fuzzy membership with the concept of probability, again using the man's height as an example. If you are asked how tall the man is, and you answer "about six feet" (182.9 cm), then using the graph of Figure C-18, you can say precisely that his membership in class "Tall" is about 0.7 , the same as his membership in the class "Normal", and that his membership in classes "Very Tall" and "Quite Short" is about zero.

What would the man's height be if he were to be measured? That is a probabilistic question. He will have only one measured height, and that measurement will exclude all other possibilities. He may be measured to be 180.9 cm, and if so, his measurement is not 182.1 cm. When he is specified as "about six feet", this specification defines a range of possibilities for what the final result will be. Each 1 mm interval has a distinct probability that the final measure will be in that interval, and the sum of all those probabilities is 1.0. Fuzzy membership and probability behave quite differently.

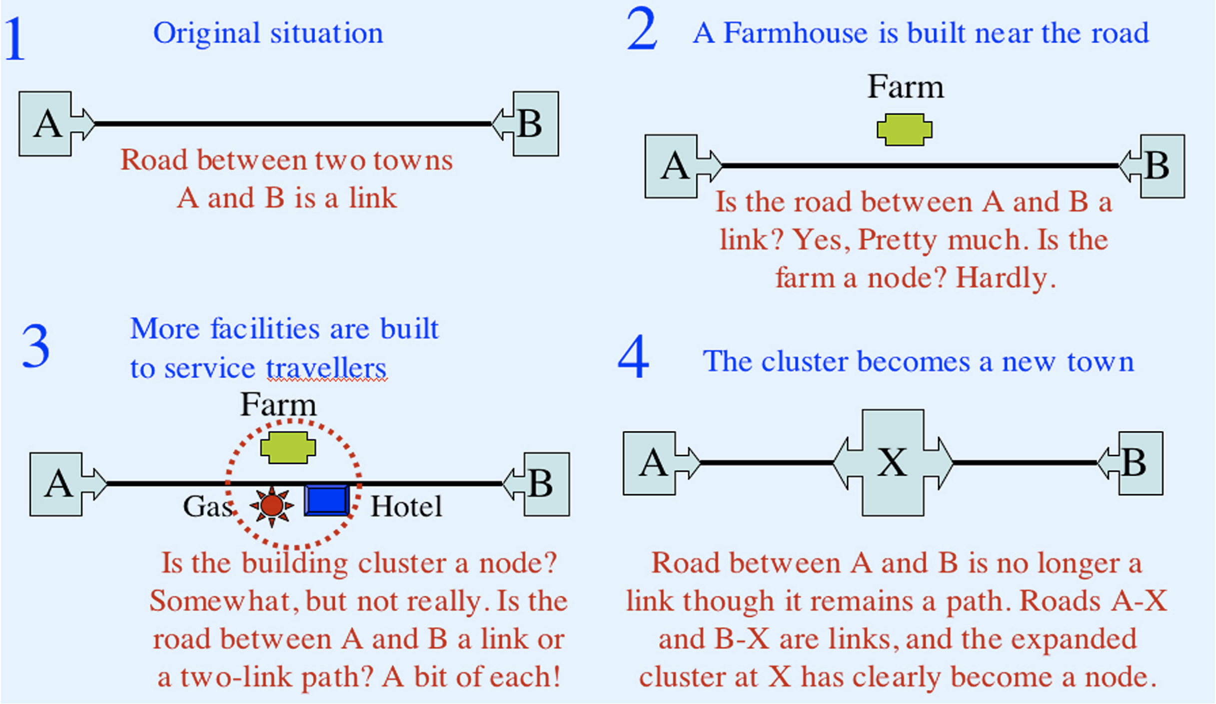
As indicated above, an entity can belong fuzzily to any number of classes. In networks, at the syntactic level there are only two possible classes, node and link, but when we come to the semantic level the number of possible classes becomes indefinite, limited only by the properties of interest to the user at any moment. Consider the two bridges in Figure C-19 as members of the class "river crossing".

Figure C-19: Two Contrasting Suspension Bridges. (a, left) The Forth Road Bridge that carries expressway traffic across the Forth River, Scotland; (b, right) A bridge across the French River, Ontario, Canada, designed for snowmobiles and pedestrians, though capable of supporting a car (Photos by the Author).

Each of these bridges carries traffic across a river, so using only that criterion, both have full membership in the class “river crossing”. However, different users may see them differently. For a logistics officer wanting to take heavy military vehicle over the river, the French River Bridge is not a river crossing at all. For that user, it has a membership of 0.0 in the class “river crossing”. In contrast, for a hiker, the Forth Road Bridge, which can carry pedestrian traffic, is not a very pleasant crossing, and for that user the Forth Road Bridge would have a membership value of perhaps 0.5 in the class “river crossing”, while the French River Bridge would have a membership of near unity. Fuzzy membership values often, perhaps always, depend on the person assigning the values.

C.4.19 Temporal Variability of Fuzzy Nodes and Links

Fuzzy membership, unlike crisp “all-or-none” values, can change continuously over time. Consider the temporal sequence suggested by Figure C-20. A link is defined as a connection between two towns, and a path as a route between towns that consists of one or more links.



1 Original situation

Road between two towns A and B is a link

2 A Farmhouse is built near the road

Is the road between A and B a link? Yes, Pretty much. Is the farm a node? Hardly.

3 More facilities are built to service travellers

Is the building cluster a node? Somewhat, but not really. Is the road between A and B a link or a two-link path? A bit of each!

4 The cluster becomes a new town

Road between A and B is no longer a link though it remains a path. Roads A-X and B-X are links, and the expanded cluster at X has clearly become a node.

Figure C-20: Stages in the Development of a Node and the Partitioning of a Link into a Path of Two Links. Art stage 1, the path AB is a link. By stage 4, a town has grown up between A and B, and the path between them has become two links. How should the intermediate stages be characterized?

Figure C-20 shows two towns, A and B, connected by a road that is both a path and a link (Panel 1). Someone builds a farmhouse midway between A and B, unnoticed by most travellers along the road, though some may stop at the farm for welcome refreshment offered by the owners (Panel 2). Around this welcoming farmhouse, others build facilities that attract travellers to stop and spend some money (Panel 3). The facilities need staff, some of whom may choose to build houses nearby. Eventually, the cluster of staff housing and buildings serving the associated needs of the residents result in the formation of a real town around the original site of the farmhouse (Panel 4).

It is clear that in Panel 1 there is a one-link path between A and B (the connection has a membership of 1.0 in the class "link"), and in Panel 4 the path between A and B consists of two links (the AB connection now has a membership of 0.0 in the class "link"). But what of the intermediate stages shown in Panels 2 and 3? Does the connection shift abruptly between being a one-link path to being a two-link path? On the ground, there is no such moment of transition. Instead, the connection becomes gradually less like a single link and more like a two-link path. Its fuzzy membership in the class "link" decreases continuously as the cluster of buildings around the farm grows into a town.

The fuzziness of link membership is, in this specific case but not ordinarily, anti-correlated with the fuzziness of node membership. The path between A and B splits into two links only as the cluster of buildings between them turns into a true town, which is by definition a node. The location of the cluster started with a fuzzy membership of zero in the class "node", and finished with a membership value of 1.0. It is tempting to suggest that the membership of the AB connection in the class link should be directly related to the membership of the building cluster in the class node, and even that $L = 1 - N$ where L is the link membership of the connection and N is the node membership of the building cluster. Whether such a suggestion can be supported, or if it is supported, whether it can be generalized to other configurations, remains to be seen.

C.4.20 Negative Fuzzy Membership Values

In fuzzy mathematics, the membership of something in a class is limited to the range 0.0 to 1.0. However, in addressing network questions, it has proved useful to consider the possibility of negative fuzzy membership values. To render this concept more plausible, consider the case of negative counting numbers, a concept that is now taken for granted, but that once may have seemed impossible.

Imagine Arthur, a poor farmer with two cows in his field. He counts them "one, two". His cottage roof leaks, and his neighbour Bruce are good at fixing roofs. He offers Bruce a cow in exchange for fixing the roof. Now he has one cow he can see and count in his field. Soon afterwards, to get some money, Arthur takes his remaining cow to market. How does he now count cows?

At one time, the answer might have been that you cannot count what is not there, so Arthur can't count cows in an empty field. Eventually, however, the concept of zero as a count became understood, and Arthur could say that he had zero cows in his empty field, even though he cannot see those zero cows. But what if he now needed Bruce to fix a crumbling wall? He could do it on credit by offering to give Bruce a cow when next he would be able to acquire one. How many cows would Arthur have at that point? Nowadays, it is easy to say that he would have minus one cow, because if he were to add one he would have zero cows. But at the time, the answer might have been that the whole idea was ridiculous that one could count cows that did not exist and might never exist. Imagined in this way, the concept of a negative counting number is seen to depend on circumstances other than direct vision of the things being counted. From zero upward, the answer to the count is based on looking at the field, but to know that Arthur has minus one cow, one has to know of his arrangement with Bruce. Looking at the field does not help.

Using this analogy, it is possible to argue for the usefulness of negative fuzzy membership values in network analysis, even though on the surface it seems non-sensical to say that something has less than no membership in some class. As with negative counting numbers, the concept becomes less non-sensical when factors outside the network are taken into consideration. In other words, the pragmatic embedding field can induce what cannot be deduced from within the network, just knowledge of Arthur's arrangement with Bruce can induce what cannot be counted by looking at the empty field.

Think of the two bridges of Figure C-19 above, the French River Bridge for pedestrians and snowmobiles and the Forth Road Bridge for heavy expressway traffic. For a logistics officer needing to move heavy military vehicles, the French River Bridge has a membership value of 0.0 in the class “link”, whereas the Forth Road Bridge has a membership value of 1.0.

What would it take to increase from 0.0 to, say, 0.5 the membership value to the logistics officer of the French River Bridge? Physically, the bridge would need to be strengthened and widened, and the approaches greatly improved. But beyond the physical requirements are other considerations, both political and financial. The bridge is in a park setting, and is much used by pedestrians as a viewpoint. It is part of a large snowmobile network the users of which would presumably be rather unhappy with such a modification to the bridge. The political opposition of park users and snowmobilers would have to be overcome. Then funding would have to be found, which might require more effort. Overall, to make such an increase in the membership value of the bridge in the class “link” would take some effort. The circumstances requiring such effort are part of the pragmatic embedding field for that part of the road network.

Now imagine the situation before the French River canyon was bridged. To create the bridge as it now stands, with its membership value of zero for the logistics officer, took both political and financial resources, just as raising the membership value of the existing bridge would do. It seems entirely reasonable to consider that the membership value of the unbridged canyon would be negative by an amount commensurate with the effort that would be required to bring it up to a membership value of zero. The case seems entirely analogous to that of Arthur having minus one cow because it would take the effort required for him to acquire a cow in order to bring him to a count of zero cows.

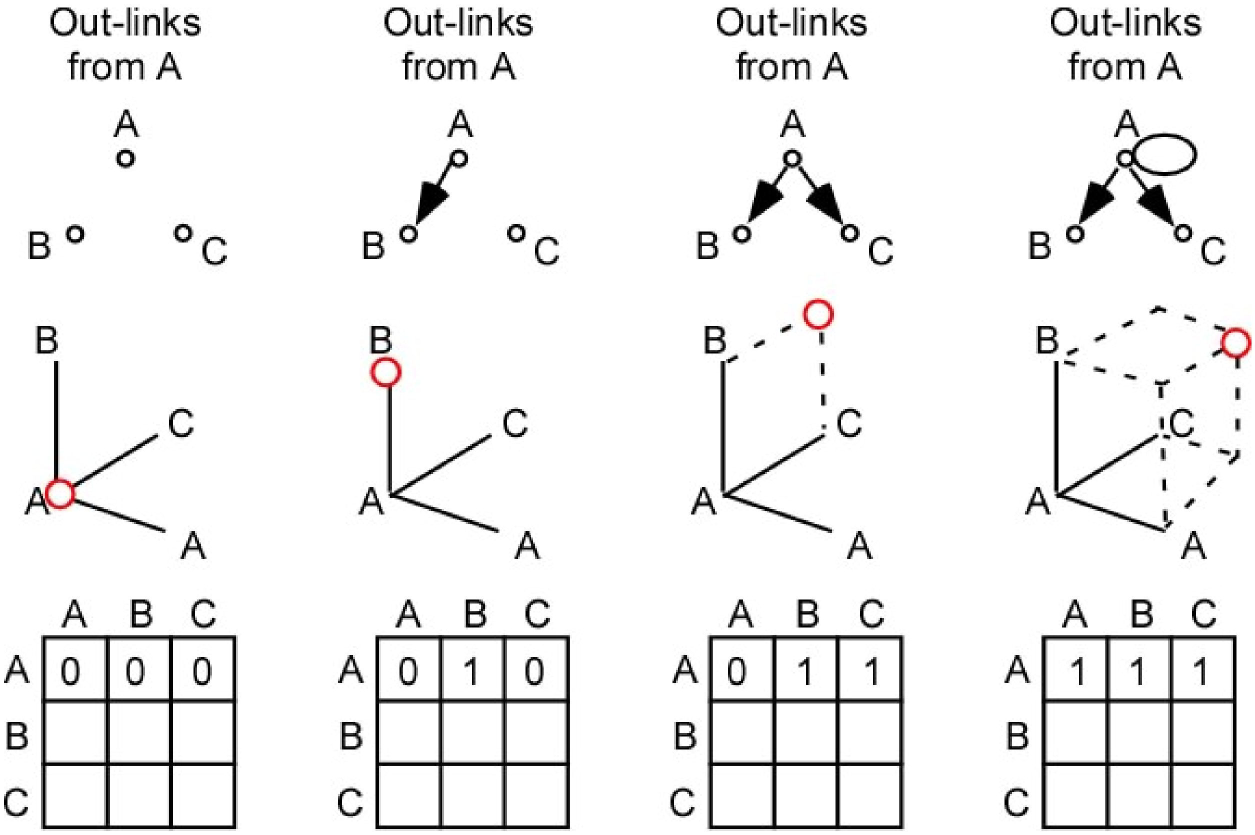
In terms of effort required to change fuzzy membership values, the range from zero to unity seems to be substantially non-linear. It would probably take much less effort to bring the French River bridge to a membership value of 0.5, at which state it could support one heavy vehicle at a time, than it would take to further “improve” it to six-lane expressway standard. If this non-linearity continues into negative membership values, the scale of admissible negative membership values seems open-ended.

Not much work has been done to follow the implications and the mathematics of negative fuzzy membership values, but they do seem to be a necessary consequence of incorporating the concept of pragmatic embedding field in the analysis (and visualisation) of networks.

C.4.21 Structural Dynamics of Networks

IST-085 is concerned not just with static networks, but also with network dynamics. Considering fuzzy membership provides a tool to investigate the dynamics of network structure.

At the syntactic level, a network can be represented as an NxN matrix, where N is the number of nodes, and a cell of the matrix represents the weight of a directed link between the corresponding nodes (self-links are often permissible). If weight is ignored or irrelevant, the cell entries are 1 or 0, depending on whether a link exists, as shown in Figure C-21.



	A	B	C
A	0	0	0
B			
C			

	A	B	C
A	0	1	0
B			
C			

	A	B	C
A	0	1	1
B			
C			

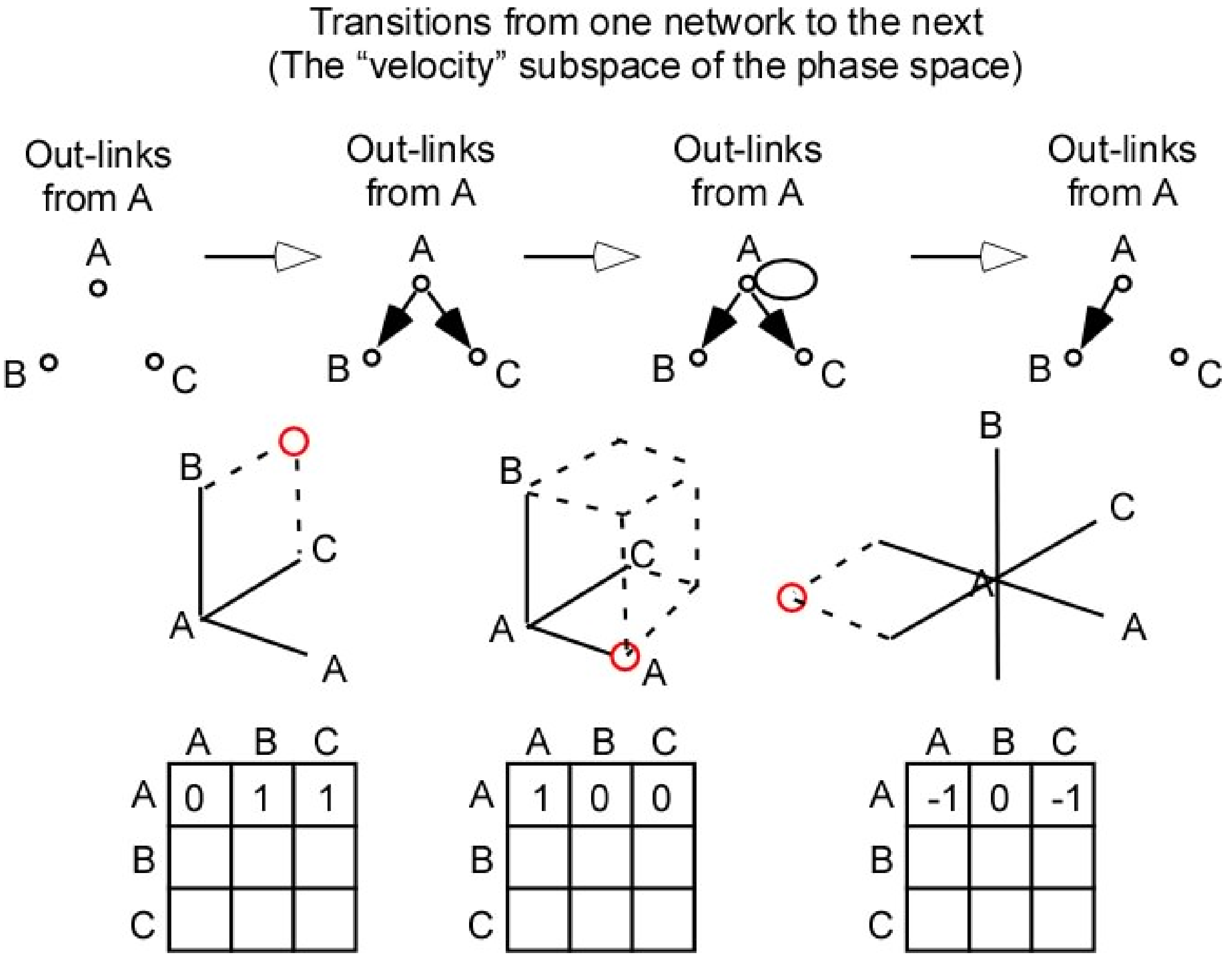
	A	B	C
A	1	1	1
B			
C			

Figure C-21: Three Different Ways of Representing the Outlinks from One of the Nodes of a Three-Node Network. The same representations may be used for any number of nodes, but the graphical form requires 6 N dimensions to represent in-links as well as outlinks, where N is the number of nodes. The geometric analyses work in such high dimensional spaces, but they are hard to show on the page.

Figure C-21 shows three equivalent ways of representing just the outlinks from only one node of a three node network. The top row shows the conventional node-and link diagram. The bottom row shows a matrix representation in which a “1” represents a link that exists and a “0” represents a non-existent link. The important aspect of this figure is to show how the structure of the network can also be represented as a location in a multi-dimensional space.

The cells of an NxN matrix can be arranged as a vector of N^2 elements. The elements of any vector define a location in a space of as many dimensions as the vector has elements, changes in the values of the elements define a directed motion of the location, and the velocity of movement if time is taken into account. Any network structure can therefore be defined by its location in a space of very high dimensionality that cannot be directly visualised, but that can be treated geometrically.

After a structural change such as the addition or removal of a link, the location that represents the structure has moved in the multi-dimensional representation space. The difference between the two locations is itself a vector in that space. Examples are illustrated in Figure C-22. In this figure, the matrix and the graphic represent the transitions, not the structural locations.



	A	B	C
A	0	1	1
B			
C			

	A	B	C
A	1	0	0
B			
C			

	A	B	C
A	-1	0	-1
B			
C			

Figure C-22: Changes of network structure can be represented as movements in space. The individual matrices and 3D graphics illustrate the vector of the movement, not the initial and final locations of the network structure.

Links are not restricted to simply existing or not existing. They may have weights, and they may have fuzzy memberships in the class "link". Either of these possibilities allow for a continuum of possible values for each cell of the matrix or each dimension of the graphic, as suggested in Figure C-23.

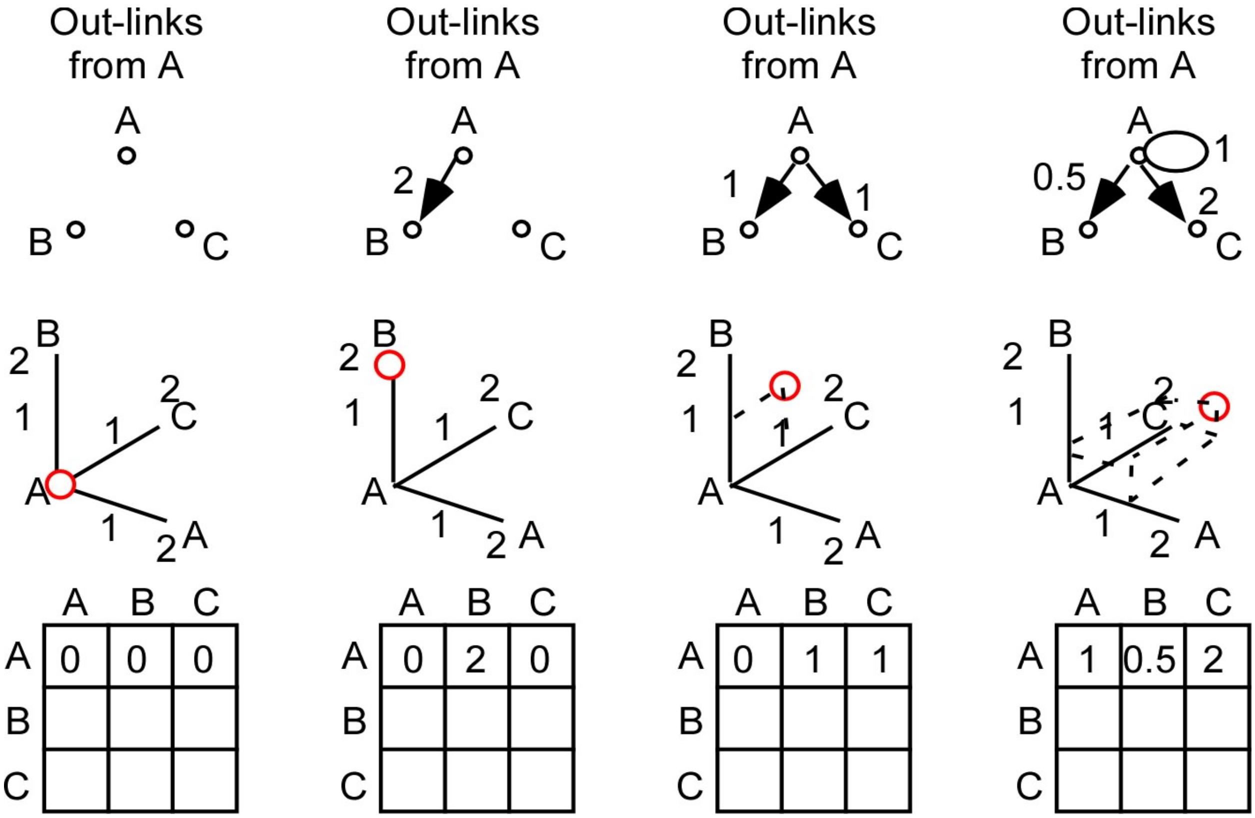


Figure C-23: Links with Graded Values, Such as Weights, that Range Upward from 0.0, or Fuzzy Memberships that Range Downward from 1.0.

The location representations in Figure C-23 correspond to network structures, this time including continuous variation of weights. As the locations change, they may change continuously. If they do, it becomes reasonable to talk about dynamic structural momentum for the network, which is not true if the links are described only by their existence or non-existence.

The same argument applies if the link and node membership values are fuzzy. Changes in the fuzzy membership values can define a directed velocity of structural change in the network, in the same way as changes of link weight. The two measures can be used together, defining a vector of 2 x N^2 elements whose changes in fuzzy membership and link weight together define the dynamic structural changes of the network.

To reinforce the idea that link weight and fuzzy membership are independent variables, consider two examples of road networks. The first example is a city street network, the second the connection between towns A and B used above to illustrate the development of fuzzy membership values. The fuzzy membership values of the elements of the city street network remain unchanged through the day and night, but the weight (defined as the traffic load on each stretch between road junctions) varies over time, both overall and in the distribution of weight over different regions of the city. In contrast, as the cluster of buildings between towns A and B grow into a town, the membership of the AB path in class “link” declines from 1.0 to 0.0, while quite possibly the traffic flow (weight) stays the same or increases.

C.4.22 Macrostates and User's Purpose

At the semantic and pragmatic levels, a network has an indefinitely large number of possibilities, since its real-world elements have an indefinitely large number of properties. However, only a small number of properties are of interest to a user at any one moment, the states of the others being irrelevant. In Annex Y, the information-theoretic implications of this fact are considered at length. Here, we consider the implications for analysis.

Two words are important in the following discussion: microstate and macrostate. The microstate of a network is described by representing **all** of its properties and external influences, whether or not they are known to or of any interest to the user. A macrostate contains possibly very many microstates. All microstates whose differences are not observed by the user or are of no interest to the user at the particular moment belong to the same macrostate. Figure C-24 offers an example.

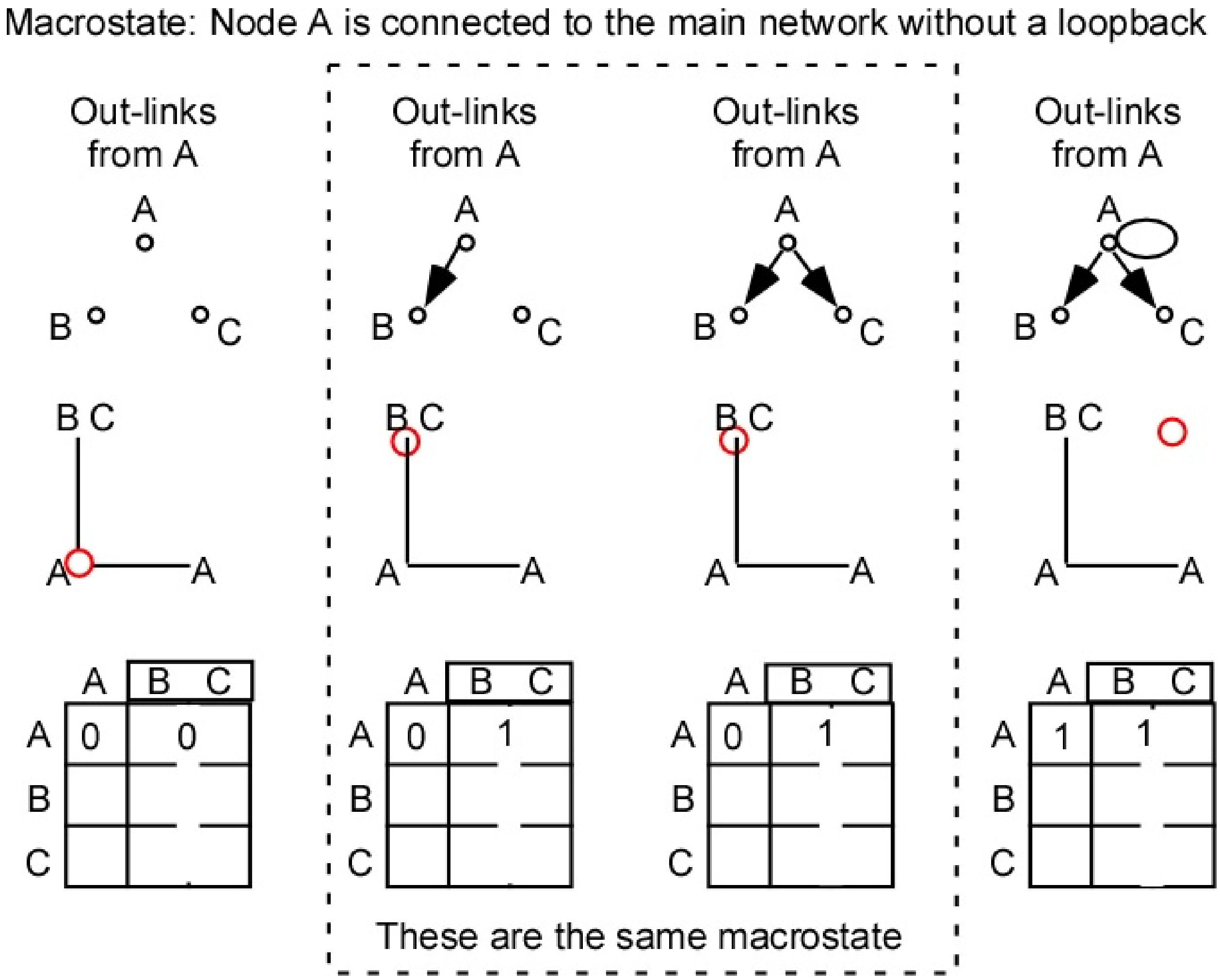
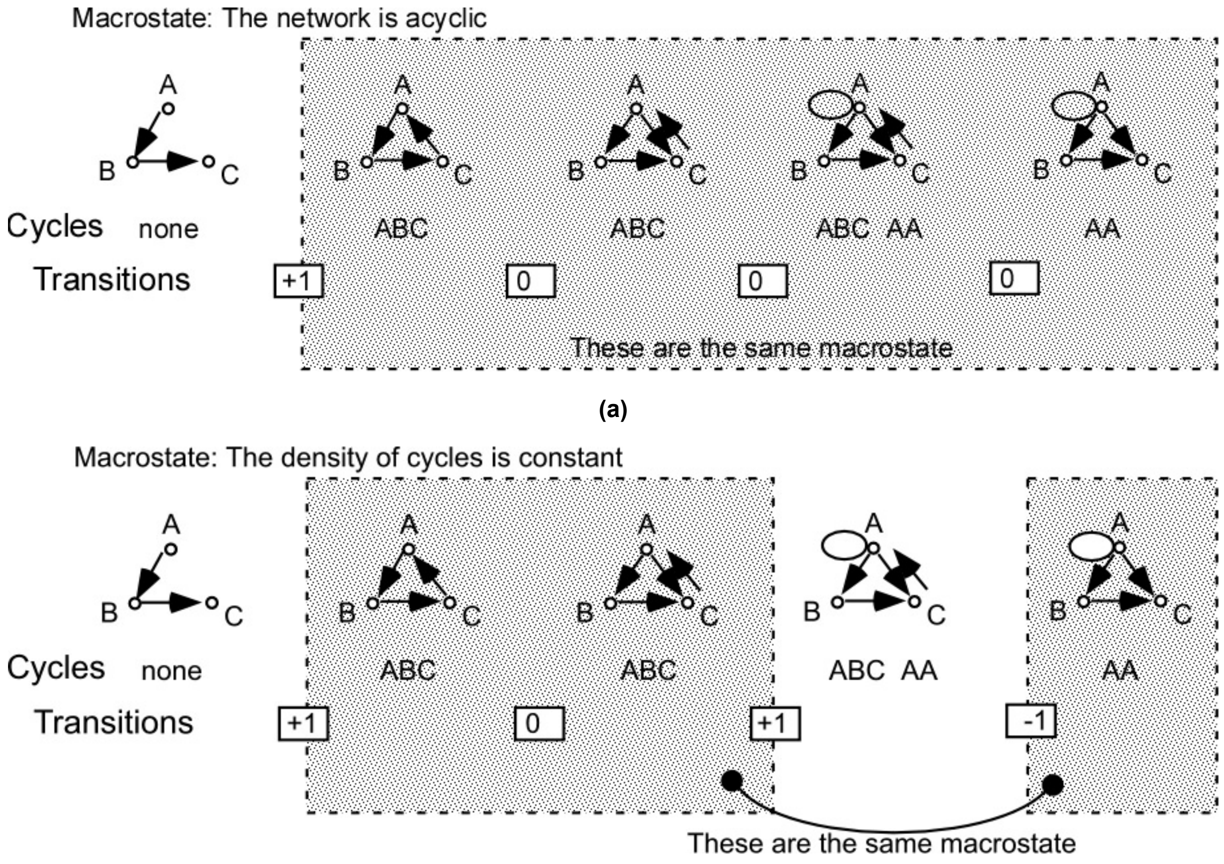


Figure C-24: If the value of some property of the network is irrelevant to the user, variants of that property do not change the macrostate. (Top) Example network structures. The user does not care whether A is connected to C; (Middle) Graphically, the addition of the link AC does not affect the location of the structure in the network space; (Bottom) Matrix representation of the location component of the phase space for outlinks of A.

In the velocity component of the phase space of the network, the transition between the second and third configuration of Figure C-24 does not affect the user's view of the structure, since the user is not interested in whether A is linked to C. The addition of that link in panel 3 changes the microstate but not the macrostate.

Figure C-25 illustrates the fact that the macrostate depends on the user's interests of the moment.



(a)

(b)

Figure C-25: The boundaries of macrostates depend on the user's interests of the moment. (a, top) The user is interested in whether any cycles exist in the network. Acyclic networks belong in one macrostate, networks with cycles belong in another; (b, bottom) The user is interested in the density or number of different cycles in the network. Networks with zero, one, and two cycles belong in different macrostates.

In Figure C-25, the sequence of microstates is the same for both panels, but the sequence of macrostates differs because the user is interested in different things about the network. In the upper panel, the network initially has no cycles. Through all the subsequent changes of microstate, the network has at least one cycle, and therefore remains in the same macrostate for an observer who is interested only in whether the network contains cycles. The existence of a cycle means that the network might be able to sustain autonomous traffic activity, whereas a network without cycles has only sources and sinks. Traffic in an acyclic network must be sustained by insertions form outside the network.

In the lower panel of Figure C-25, the user is concerned with the density of cycles in the network. The higher the cycle density, the more likely it is that the self-sustained traffic dynamics becomes complex and even chaotic. Accordingly, even though the sequence of microstates is the same as it is in the upper panel, the sequence of macrostates is quite different.

The concept of microstates and macrostates based on the user's momentary interest becomes important when we consider the information dynamics of networks and their analysis. The concept is not restricted to consideration of network structure; a macrostate consists of all microstates that do not differ in any way that matters for the user's purpose. For road traffic analysis it may not matter at all whether a particular car is red or blue, but for police looking for a suspect after a robbery the difference could be crucial. The policeman has a set of macrostates quite different from that of the traffic analyst.

C.4.23 Network Dynamic Phase Space

Macrostate boundaries may depend on whether certain dimensions of variation among the microstates are uninteresting to the user, as in the case of the red versus blue car, or whether the boundary depends on some variable measure within a dimension, as is the case if children under a certain height ride free on city buses. It does not matter how much taller or shorter they are than the criterion; for the bus driver, one macrostate contains shorter children, the other contains taller children. For the parent, the height dimension may be separated into many more macrostates, or not, depending on the circumstances. When getting on a bus, the same two macrostates apply as are used by the bus driver, but when buying clothes, the macrostates are defined by whether the child can wear the clothes to be bought.

The implications for display differ between macrostates defined by ignored dimensions and macrostates defined by variation within a dimension. For analysis, all that need be displayed is the identity of the macrostate. Display of any aspects of the microstate within the macrostate would constitute "clutter". On the other hand, for visualization it may well be useful to display more detail about the microstate, as well as about the embedding fields that could influence the macrostate if they were to move the microstate in predictable ways.

Macrostates defined by the dimensions used in their definition differ from macrostates defined by ranges of variation within a dimension in another important way. For the latter, but not the former, dynamical changes can be predicted by noting variations of the microstate. A homely example is afforded by a growing child. One possible macrostate is defined by the child's shoe size, and when the mother notes that it is getting harder to get the foot into the shoe, she may say "We will soon have to get you new shoes". The actual system dynamics is determined by variations in the microstate, which may be continuous, whereas changes of macrostate are ordinarily discontinuous. The concept of "velocity" makes sense only is systems with continuous variation.

Where there is directed velocity and a location, there is a *phase space*. Phase spaces are much used in the study of mechanical systems, and some of those techniques are likely to prove useful in the analysis of real networks for which some history is available.

Networks present a particular problem as compared to mechanical systems. The phase space of a mechanical system usually involves momentum rather than simply velocity. Momentum is the product of mass and velocity, not just velocity alone. A system with momentum cannot change its velocity magnitude or direction instantly, whereas a mass less system can. If the velocity can change instantly, the future is inherently unpredictable even for the smallest distance into the future. Is there any equivalent to mass in the network context? Possibly there is, in the pragmatic embedding field of the network.

C.4.24 The Pragmatic Embedding Field and the Shadow Network

The pragmatic embedding field of a network consists of all the influences outside the network that affect anything within the network. These external influences constrain the possible changes in the network structure. To continue the bridge example, it takes time to assemble the political approval, the funding, the physical machinery, the architectural and engineering design skills needed to transform a lightweight narrow bridge into an expressway link. The change in the network structure cannot happen instantaneously, though the overt change of link character does happen in the instant that the ceremonial ribbon is cut on the new and improved bridge. Behind the scenes, the new bridge develops slowly, but as a component of the network, it does not exist until that ribbon is cut. Hence an analyst wanting to predict the future of a network must often analyze not just the network, but also its pragmatic embedding field.

If a bridge is known to be under construction or deteriorating from neglect, or a social relationship is being frayed by external influences on the participants, the analyst may well be able to say with reasonable assurance that at some point the network structure will change abruptly, when the ribbon is cut, or the partners decide on divorce. In the phase space, although the overt location of the network is static, yet the analyst can trace covert movement – a shadow network changes under the surface of the existing network, until the existing network snaps to match the shadow one or the shadow one vanishes because the analyst learns of a stop-work decision for the bridge or a reconciliation of the partners.

There is a relationship between fuzzy membership and the hidden shadow network, but they are not the same. Fuzzy membership changes depend on the effort that would be required to bring the link or node to its new value, without that effort being actually expended. Changes in the virtual network depend on the effort being actually expended to bring about those changes in fuzzy membership values.

C.4.25 Uncertain and Unknown Portions of Networks

It is rare that a network is known entirely, and it is rare that for a user's purpose the network needs to be known in every detail. Certainly, for a network of more than trivial size, the human mind is unable to maintain an understanding of the existence of every link and node, let alone the full description of the properties of each one. Nevertheless, there are times when data are missing as to whether a particular pair of nodes is connected by a link of a certain character. Are these two people of interest friendly with each other, unfriendly to each other, or unknown to each other? The answer could be crucial in a counter-espionage operation or the assessment of a possible terrorist threat.

No tools or techniques could, even in principle, determine that a link does or does not exist when the relevant data are missing. However, various factors may influence the user's probability that the link does or does not exist. For example, if the entities belong to the same hypernodes at several levels (based on several different properties) and other members of those hypernodes belong also to common hypernodes at other levels where data are missing for the entities of interest, it is not improbable that the entities would also belong to the same hypernodes at these other levels. "*Birds of a feather flock together*."

The situation is a little different when what is known can be considered to be a local sub-net and what is unknown is a possibly much larger network outside the locality. This situation is akin to knowing the street plan of London, but not of Manchester, New York, Delhi, and Tokyo. What happens to the traffic in those places has little bearing on the traffic in London, but there are roads leading to the edge of the "known world" that act as sources and sinks of traffic into and out of London. Even though the network as a whole may not have significant sources and sinks, by segmenting the network the broken links change the situation / if the user/analyst knows nothing about the network outside the broken ends, they have to be considered as potential sources of substantial variation in the dynamics of the local network.

Similar problems arise if the network is traffic-free, the links representing relationships of different kinds among the nodes. In this case, the existence of links from known nodes to entities outside the network may well be unknown, so there is no possibility of inferring their existence, at least at the syntactic level. Consideration of the pragmatic embedding field might suggest their existence, and correlations of properties across nodes that have no obvious relationship might also suggest that some external factor might be in play. The main question arising from uncertainties that seem to matter is whether the data exist in the data space, and if not, whether it is worthwhile trying to obtain the data in the real world.

C.4.26 Training and Expertise

The VisTG Reference Model (Figure C-2, repeated as Figure C-26) is based on the assertion that in order to act usefully, a person must be able to perceive the thing acted upon. To perceive something does not mean simply to be aware of the entire environment, which necessarily includes the thing in question. To see a number of brown, red, gold blobs is not the same as perceiving a wooden chair with richly brocaded upholstery. Seeing such a chair is not the same as perceiving a valuable antique dated around 1740 by an important French furniture maker. To create the perception, the perceiver must have specific processing abilities and some background knowledge. “Chair” is a macrostate. Many different arrangements of colour blobs in the visual field can be perceived as being a chair, and if one is interested only in seeing a place designed for sitting, all of these different arrangements of blobs belong to the same macrostate – the same perception.

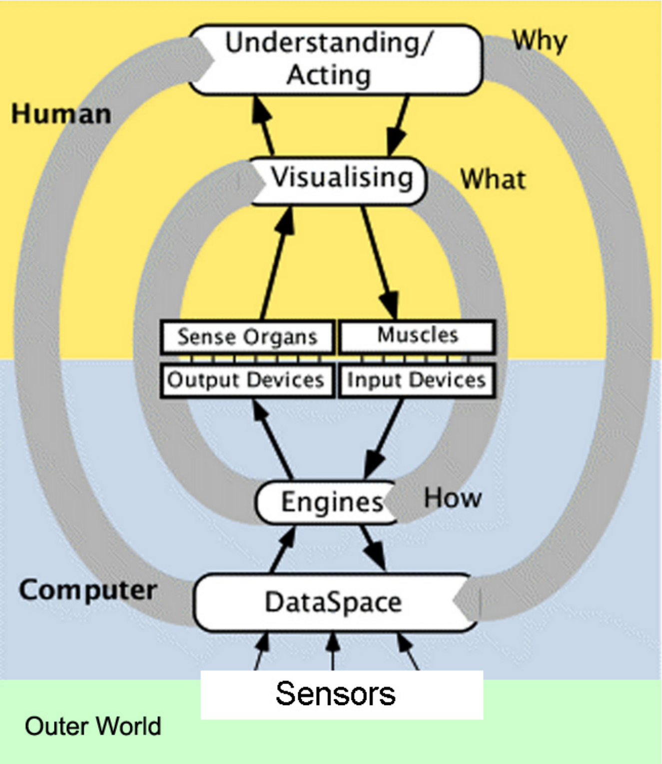


Figure C-26: The VisTG Reference Model (Figure C-2 Repeated).

For the expert in antique furniture, to perceive an arrangement of blobs as simply "a chair" may or may not define the macrostate, depending on whether he is simply weary or is conducting an evaluation. For the evaluation, the "chair" macrostate may become a myriad of smaller macrostates, distinguished by the expert's perceptions of the date of manufacture, the maker, the rarity of the type of chair, the condition of the wood and of the upholstery, and so forth. Many of these perceptions are unavailable to the inexpert observer of the colour blobs. If one cannot process incoming data appropriately to determine whether the microstate belongs in a particular macrostate, the effect is equivalent to functional blindness. The person may know that a chair by a particular designer would be very valuable, but on seeing a chair, cannot perceive whether it is by that designer.

After receiving the expert's evaluation, the inexpert user may perceive the chair to belong to one of several macrostates such as "chair to keep", "chair to sell" or "chair to discard". Nevertheless, the expert has identified several different possible macrostates that might be valuable to the inexpert customer. Suppose the owner's primary macrostate based on the expert's evaluation is "chair to sell". The question for the owner then becomes how and where to sell. The answers would be quite different for an expert's identification of the chair as "1740 French Rococo" as opposed to "1820 Philadelphia" or "School of Chippendale". Such names identify microstates that the owner knows to exist, but cannot perceive, in the same way that a colour-blind person knows that there are colour macrostates such as "green" and "red" but cannot perceive any difference between blobs other people label with one or the other name.

Any set of microstates could, in principle, define a macrostate. At any one moment, however, the functional macrostates depend on the user's shifting purposes. The user may act differently if the system is in one macrostate as opposed to another, and that difference may not depend on whether the observer can directly perceive the macrostate boundaries. The colour-blind person must stop if the traffic light is red and go when it is green, even though he cannot see the colour difference. The chair owner takes the "chair to sell" to a different auction or saleroom depending on what the expert says about the chair's background. And both naïve and expert military observers of a display tailored to suit them should be able to determine which macrostate is implied by a particular pattern of data.

Here we begin to deal with the question of information analysis, training, the partitioning of work between computer and human, and the construction of displays for novice and expert users. A novice user, naïve to the patterns that signify militarily important differences, will be confused by a display that allows the expert to visualise the positions of microstates within his current macrostates in a complex dataset. An expert, on the other hand, may be hampered by a display that simply shows the results of an algorithmic analysis in terms of the probable current macrostate, even if that analysis is based on the expert having efficiently identified to the computer his purposes for defining the macrostate boundaries.

Referring to the display type's definitions (Table C-3) the macrostate analysis suggests that for display the naïve user may be well served by algorithmically selected data, with labelled placement, whereas the expert may work better with data user-selected interactively, with located placement. The "labelled placement" dimension suggests that the computer presents labelled macrostates in preference to the microstate detail within the macrostate, while the "algorithmically selected" dimension indicates that the computer makes the choice as to how much of the microstate detail to present. As the user gains expertise, this selection may become progressively more interactive. The novice is led by the hand through an analytic exercise that shows the possibilities in broad outline, while the expert uses the detail to visualize the possible outcomes of different unknown states and different actions upon the microstate properties.

The transition from novice to expert is largely a transition from analyzing the implications of different factors "according to the book" to visualizing the situation or "seeing the big picture" from the patterns of detail in the data. A similar transition occurs as a child learns to read, from laboriously putting together the sounds of the letters to seeing the meaning when exposed to the word form. It is a very general form of learning to interpret any complex data input.

The VisTG Reference Model suggests that the primary effect of learning is the development of the ability to perceive the macrostates that suggest the possibilities of different actions. The novice needs to become able to "see" at least some parts of the "big picture". J.G.-Taylor demonstrated that the development of the ability to perceive novel feature of the environment depended crucially on a person's interactions with the factors of which the feature is composed. One would therefore argue that a novice should be allowed to manipulate a display of some of the microstates that contribute to a macrostate she is learning to perceive, while also being aware of changes in the macrostate as defined by an expert or by an algorithm.

C.4.27 Expertise and Network Dynamics

A display that shows only the macrostates and not the underlying microstates cannot allow the user to anticipate changes of macrostate. Figure C-27 illustrates the problem. If the user is expert enough to be able to visualize the macrostate when the microstate is displayed (usually in a dense and complex display), then the likely future changes of macrostate can be anticipated. The middle panel of the figure shows a geometric representation of the macrostate without showing the macrostate. The geometric representation makes clear what the "neighbour" macrostates might be, but in a multi-dimensional structure where the macrostate depends on several properties, such a presentation is difficult to design, and the macrostate is likely to be presented only as a label, suggested by the "2, 2" identifier in the right-hand panel of the figure.

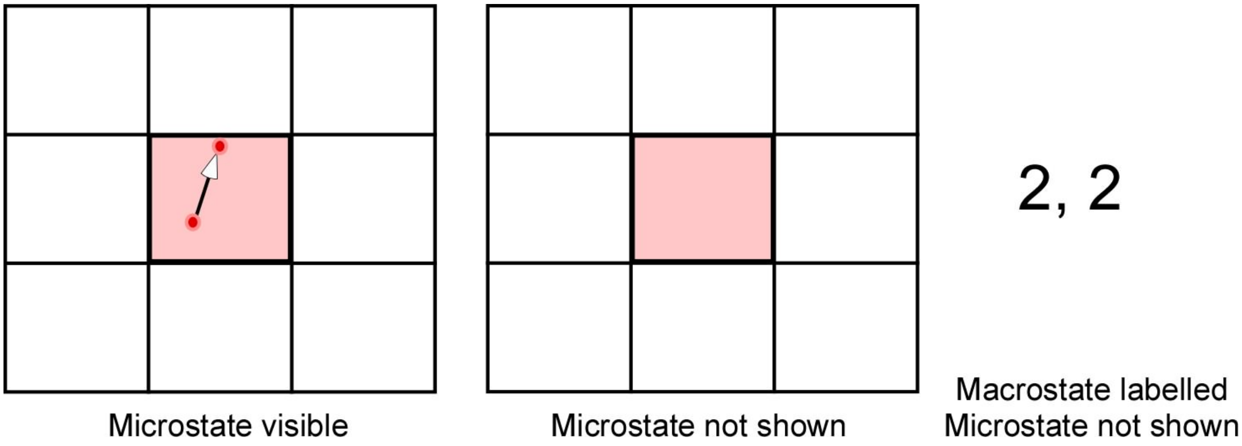


Figure C-27: If the microstate is not shown, the user cannot anticipate when or how the system is most likely to change macrostate. If the macrostate is simply labelled, it is hard for the user to judge which macrostates are available as the next state.

The presumption is that an expert would be able to identify the macrostate from seeing the microstate, whereas the novice would not, and would rely on the algorithm to identify the current macrostate. Of course, an algorithm could also identify likely changes of macrostate, if the user supplied all the possibilities of interest, along with precise definitions of the macrostate boundaries.

Above, network structure was represented as a point in a high-dimensional space, and structural changes in the network were represented as changes in the location of the point (Figure C-21, Figure C-22). Changes could, of course, be in any direction in this high-dimensional space, but certain directions may be preferred. For example, a change that affects only one link is represented by a move parallel to an axis of the space. A change caused by something happening in a node may be reflected in the links connected to that node. Such correlated changes occur within the sub-space defined by the axes that represent the links in question, and often the direction of the change will be parallel to a main diagonal of that sub-space.

It is difficult to display a space of more than three dimensions on a two-dimensional screen, and when we are talking about network structure, the space could have millions of dimensions. How, then could be helpful to talk about directions nearly parallel to a main diagonal of a sub-space? One way is simply to point up the inherent problem.

More importantly, geometric thinking can feed back to other ways of representing network properties. "Directions nearly parallel to a main diagonal of a sub-space" indicate correlations among events in different parts of the network, and it is only correlations among events that indicate to our human perceptions the existence of objects and complex structures with real-world importance. The states of such complexes are perceivable variables, though they may be perceptible only to an expert. Nevertheless, they are the only variables available to define the macrostates that might be important to a user, and must be learned if a novice is ever to become an expert.

C.5 USING THE FRAMEWORK

The VisTG Framework is just that, a framework. As such, it is of no value until the framework is fleshed out with the specifics of the user's or designer's problem. Why is the user or designer interested in visualizing something, and what is to be visualised? These questions must be answered differently for each problem, and in most cases for several mutually dependent phases of the problem.

To help the user or designer to approach each unique problem, IST-059, the predecessor group to IST-085 created a spreadsheet that listed questions based on the state of the Framework at that time. This spreadsheet was tested against four widely different problems in three different domains – the spread of infection, anti-terrorism, and computer security. In its first draft form, some of the questions were not easily answered, and it was obvious that revisions were required, as well as an upgrade from a static spreadsheet to an interactive web-based system. This necessary upgrade was not performed during the life of IST-085, and remains as unfinished business.

Despite the failure to address the upgrade of the interface for using the Framework, nevertheless, the extension of the Framework in several aspects, and its implicit use in different areas, allows a reconsideration of the questions to be asked in either a static or an interactive form. These questions are designed to help the user to understand the problem and thereby to find a solution more easily than if the problem were to be addressed as a unique novelty.

C.5.1 Starting Out

The first step in using the Framework is to identify what aspects of the problem involve networks. Any part of the problem that does not involve a network can be treated in the same manner, but using the generic elements common to all Frameworks of this type rather than the network-specific aspects of the IST-085 Framework for dynamic network visualisation.

The questions the user should ask begin quite generically, and become refined as the answers to the generic questions come available. The first question has been suggested above:

Question 1: Is the real-world problem primarily concerned with the properties of one or more networks, or is the network a matter of peripheral interest?

In what follows, we will assume that the network is of primary interest in solving the real-world problem, because if it is not, the results of working on the other parts of the problem are likely to have considerable influence on what will be asked of the network(s) involved. Under that assumption, the next few questions are concerned with the kind of network and the kind of information about the network that the user might want to get from the computerized version of the real-world network. Which questions are applicable, and in which order they should be considered depends greatly on the answers to the earlier questions. The question set is itself a fairly complex directed network (as are many flow charts).

Question 2: How expert are you on the subject matter of this network? (e.g. Expert, some knowledge, novice).

The answer to this question should be a guide as to the kinds of displays that might be best suited to the problem. An expert may be better able than a novice to take advantage of a complex cluttered display suited to visualisation. On the other hand, the novice must be shown material that the expert will retrieve from memory, such as the implications of complex analyses, which the expert may request but which the novice may not realize to be useful. On the other hand, visualization can often help a novice to understand relatively simple properties such as whether a large network has obvious divisions into relatively coherent sub-nets.

Question 3: Why are you observing the network?

This is the starting point of the VisTG Reference model, and implies the five related question listed at the end of Section 1.2, which are not repeated here other than to say that they deal with how to get the required information and what might stand in the way of getting it. The answer to question 3 will be related to one of the following:

- To learn something about the network that may be useful later (Exploring);
- To learn something I need to know right now to make sense of something I have in memory or am now observing (Search); and
- To maintain awareness of some property of the network (and perhaps to act to influence this property) (Monitoring/Controlling).

Alerting will not explicitly occur as the answer to this question, at least initially, since Alerting is a background activity that does not involve observing the network. However, "To set up an alerting condition that may be invoked if some state is found in the network" is an aspect of Controlling, but at a lower level relating to the Engines of the computerized system rather than to the main purpose.

Question 4: Who will use the display?

Are you a subject-matter expert producing a briefing for operational personnel, an analyst trying to find out something by interacting with the display, a planner trying to develop a plan for the approval of executive authority – who is the display intended for? You, interacting with it, you interacting with it under instruction, or passive observers? The implications for the display and for interaction techniques are quite different depending on the answer.

C.5.2 Network-Related Questions

Question 5: Are you mainly interested in structural aspects of the network or in the traffic flow over the network?

Clearly, if the network is traffic-carrying, its traffic flow properties will be affected by its structural properties, and if it is traffic free, the answer to the question must be that you are interested in something related to the network structure. If you are interested in how the structure affects the traffic flow, the answer would be "Both", but it is quite possible you are interested only in the traffic, particularly if the answer to Question 3 is related to the real-time activities of Monitoring and Controlling.

Question 6: Does the supporting environment of the network matter?

This question related to the constraints implied by the syntactic and semantic embedding fields of the network. Dot you care about the physical support of the network, such as physical wires, broadcast bandwidths, social meeting places, and so forth, or are you interested only in what happens within the network itself? This question does not concern the pragmatic embedding field, which changes by the moment, but with the embedding fields that are necessary for the network to exist. If you answered Question 2 with a Controlling/Monitoring kind of answer, the answer to Question 5 will be necessary, but not sufficient to describe what you require.

Question 7: Are you concerned with the situation of the moment in the environment of the network?

This question addresses the pragmatic embedding field. Would it matter to you if the status of two of the people represented by nodes in your network changed, as for example if they had a quarrel? Would the existence of a forest fire near a road in your network affect your purpose?

The pragmatic embedding field may be at the same time the most important and the most difficult aspect of the problem to pin down. Unlike the situation with the syntactic structure, or even the pragmatic stricture, of the network, the range of possibilities for the pragmatic embedding field is unbounded. The bounds must be imposed by the analyst/user. The macrostates must be described, perhaps not explicitly, but at least in the mind of the user. What changes to the environment might affect the purpose of the moment, and what infinite variety of possible changes would not?

Question 8: Are you concerned with global or local properties of the network?

Do you expect to zoom in on a small proportion of the network, or are you interested in the larger patterns of structure over a wide region of the network? An example of the former might be an investigation of the vulnerable points of your own or an adversary's network, whereas the latter might be concerned with the partitioning of the network into viable interacting sub-nets.

Question 9: Are you concerned with the interactions of networks that have different characters?

An example of when one might answer "Yes" to the question is in consideration of infrastructure vulnerabilities. Militarily, the supply lines affect the abilities to wage a normal war across battle lines; the intelligence network affects the ability to treat non-conventional aggression. In civil space, the functioning of the water supply depends on the electricity supply, as do the traffic lights that strongly influence traffic flow over the urban street system. The interaction of different kinds of networks concerns Junction Nodes. Junction nodes have the singular property of defining how the elements of one network influence the elements of the other. Such being the case, both the character and the distribution of junction nodes must be available for display.

Question 10: Are you concerned with uncertain data about the network?

This question assumes that you know the network fairly well, but lack data to know whether some links or nodes exist, or if they do, what their properties are within your domain of interest. The difference between this and Question 5 is that Question 5 asks about things that are in the computer's data space, whereas Question 10 asks about things that are not, and that must be inferred from existing properties of the network or its embedding fields as they exist in the data space. The implications for display are quite different, a "Yes" answer to question 10 implying the use of concepts such as hypernodes, whereas a "Yes" answer to Question 5 implies the ability to Search within the data space using algorithmic filters to highlight the interesting aspects.

Question 11: Are you concerned with real-time dynamic changes in the network traffic?

Question 12: Are you concerned with real-time dynamic changes in the network structure?

Questions 11 and 12 are listed separately to emphasise that there are different aspects of "network dynamics", but they can be discussed together.

The words "real-time" imply that you are observing an on-going situation, possibly with a view to intervention. The operators at a power station do this continually. In a military environment that is constantly in flux, "real-time" operation is central to effective operation. When one is dealing with dynamics, "real-time" often is concerned with projection into the future, and projection implies the likelihood of "what-if" scenarios. If you are concerned with these aspects, your software must allow you to manipulate a virtual network that differs from the real one without affecting your access to the effects of incoming data from the real network.

As we noted when dealing with the phase space of the network, the pragmatic embedding field often implies the existence of a virtual network. For example, the building of a bridge starts with an idea, permissions must be obtained, political, financial, and technical resources must be brought to bear, physical construction must proceed, and finally a ribbon is cut, at which point the actual network suddenly acquires a new link. All the time, a virtual network implied by the pragmatic embedding field of the actual network that contains the new link was coming closer to reality. The structural change to the real network comes as no surprise when it occurs, but its effect on the traffic dynamics might be – which leads to Question 12.

Question 13: Are you concerned with the effects of network structure on the dynamics of network traffic?

This question implies that you are Exploring, though it is possible you might be searching for a change in network structure that might create a particular change in the traffic dynamics. The implications for display and interaction capabilities are similar in the two cases. Network structure determines whether the dynamics are dominated by feedback processes, but the character of the traffic determines whether these feedback processes lead toward stability, oscillation, or chaos. The appropriate software probably will need considerable analytic capability, as it is often difficult to visualize the effects of non-linear dynamics, no matter how expert you may be.

In considering the effects of structural manipulation on projection into the future, we noted that the software must permit the user to treat virtual networks without losing track of the actual network. The same is true when considering the effect of structural changes on the character of the traffic dynamics, except that the virtual network to be manipulated is not determined by the pragmatic embedding field of the network.

C.5.3 Using the Answers

Every aspect of the Framework implies a question that is worth asking. The dozen questions above are but a small selection chosen to illustrate the kinds of question that might be asked. The first four would need to be asked with any domain of interest substituting for "network", whereas the other nine are specific to networks. Similar questions will be implied for any other domain.

The answers to the questions posed above and others implied by the Framework should guide the user toward choice of appropriate interaction and display techniques. The key point to remember is that the VisTG Reference

Model is recursive. The answer to the Question 1, about the purpose of observing the network, leads to subsidiary questions based on what the user needs to know in order to achieve that purpose. Obtaining the different elements of what the user needs to know creates a set of new purposes, each of which may well lead to asking the rest of the questions in the new context.

C.6 REFERENCES

[1] References for the HAT REPORT and other Final Reports and Workshops since IST-013 (Various dates).

[2] Barabasi, A.-L., and Albert, R. (1999). Emergence of scaling in random networks, *Science*, 286, 509-512.

[3] Bjørke, J.T. (†), Nilsen, S. and Varga, M. (2010). Visualization of network structure by the application of hypernodes, *Int. J. Approximate Reasoning*, 51, 275-293.

[4] Papadopoulos, F., et al. (2012). Popularity versus similarity in growing networks, *Science*, 489, 537-540.

[5] Pask, G. and Scott, B.C.E. (1972). Learning Strategies and Individual Competence, *Int. J. Man-Machine Studies*, 4, 217-253, doi:/10.1016/S0020-7373(72)80004-X.

[6] Pask, G. (1976). Styles and Strategies of Learning. *Brit. J. Ed. Psych.*, 46: 128-148. doi: 10.1111/j.2044-8279.1976.tb02305.x.

[7] Powers, W.T. (1973, 2nd. Ed. 2005). *Behaviour: The control of perception*, 2nd Ed. NewCanaan: Benchmark Publications.

[8] Scheffer, M., et al. (11 others) (2012). Anticipating Critical Transitions, *Science*, 338, 344-348.

[9] Taylor, J.G. (1962). *The Behavioural Basis of Perception*, New Haven and London: Yale U.P.

[10] Taylor, M.M. (1972). Perceptual Principles relating to Remote Sensing. *Proceedings of the First Canadian Remote Sensing Symposium*, 2, 497-503.

Annex D – INFORMATION ANALYSIS OF NETWORKS

Martin Taylor
Martin Taylor Consulting, Toronto
CANADA

D.1 INFORMATION THEORY CONCEPTS

D.1.1 Uncertainty, Entropy, and Information

Networks represent relationships among different entities. A relationship implies that there is something in common between related entities, those changes in the state of one of the entities affects the state of another, or that some traffic passes from one entity to another. No matter which of these possibilities is represented in a particular network, the implication is that the related entities are related by information. Information obtained about one of the entities also can provide information about the related entity.

Historically, the three main conceptual approaches to the quantitative study of information and entropy may be labelled with the names of their major proponents, namely Boltzmann-Gibbs, Shannon, and Kolmogorov-Chaitin. All three starts from quite different basic considerations, but all three arrive at very similar conceptual results. In this annex, we will use Shannon for the most part, but that is a matter of choice, not of necessity.

D.1.2 Boltzmann-Gibbs

Boltzmann [3] recognized that the state of an ideal gas was determined completely by the positions and momenta of the individual molecules of the gas, while at the same time the observable properties of the gas were limited to measures such as temperature, volume, pressure, and the like. A myriad of different configurations of positions and momenta could give rise to any given set of values of the observable variables. The values of the observable variables constitute a "macrostate", while the actual positions and momenta of the individual particles constitute the "microstate" of the system. He discovered that under very reasonable assumptions, the number of different configurations consistent with a particular set of observable values was proportional to its entropy. The entropy therefore represents what you do not know about some World in which you can specify a few general measures. Gibbs [8] refined Boltzmann's ideas by taking the combined probabilities of the microstates in a macrostate, rather than their number, to define the probability of the macrostate. The difference is important, but the basic idea was due to Boltzmann.

Boltzmann assumed that for a given total energy of the gas no point in the microstate space that had that energy (sum of squares of molecular velocities) would be more probable *a priori* than any other. Accordingly, the hyper-volume of a macrostate within the microstate descriptive space would be proportional to the probability that the microstate would be in that macrostate. The negative logarithm of that probability was the entropy of the macrostate, the measure of what you do not know about microstate of the gas when you have specified its macrostate. Gibbs's refinement allowed for the mixture of different constituents and for the interactions among constituents, but both Boltzmann and Gibbs proposed the $-K^*\Sigma p(i)\log p(i)$ formulation that later suggested to Shannon that the word "entropy" might be used for his "lack of information" measure.

D.1.3 Shannon

Shannon [10] was concerned with telephone communication. His basic question was how much you can learn of what a transmitter intended to send when you receive a message through a channel. Before receiving the message, you know something of what messages the transmitter might intend to send, and after receiving the possibly garbled message you know something more about what the transmitter did intend to send. The difference between these two states is the information about the transmitter's intent obtained from the message.

Being concerned with the properties of the telephone network, Shannon was concerned with the rate at which a noisy, distorting channel can allow the receiver to become more precise about the transmitter's intent, not about what that intent might be for different users of the channel. As he pointed out in the introduction to his monograph, the meaning of the messages that might be transmitted through the communication channel is irrelevant when considering the properties of the channel itself. As recently as Denning and Bell [5], this comment has almost universally been misinterpreted as saying that quantitative information has nothing to do with meaning. In fact, information *always* has to do with meaning, since it depends entirely on what the receiver believed about something related to the transmitter before and after receiving the message.

Shannon took the view of an engineer observing both ends of the channel, who could know the set of messages that the transmitter could have sent and their probabilities, as well as the set of messages compatible with whatever was received by the receiver; based on the properties of the message received, the receiver could assign probabilities to the various possibilities for the transmitter's intent. Before the message was received, the receiver had a certain set of probabilities for the different possibilities, and this set of probabilities limited the amount of information that the receiver could possibly get from receipt of the message. Setting the initial probabilities to be equal for all possible messages defined the maximum amount of information an initially ignorant receiver could possibly get from the message ensemble. This ignorance, however, did not extend to ignorance of the set of possibilities.

The receiver's prior knowledge of both the message coding and the set of possible intended messages were key to the effect of the actual received message on the receiver. In other words, although Shannon did not say this explicitly, the inference from what he did say is that a message that is meaningless to the receiver is a null message that conveys no information. Meaning is irrelevant to designing and measuring a communication channel, but meaning is critical to the interpretation of "information".

Shannon gave the name "Entropy" to the amount that the receiver does not know about the transmitter's intent, because the expression for its measurement is exactly the same as the Boltzmann-Gibbs expression for the entropy of a gas. The entropy of a gas was described by Boltzmann and Gibbs as what you don't know about the microstate of the gas after you observe its macrostate. For Shannon, a "message" defines a macrostate, while the set of possible variants that could specify the same message are the microstates. If the message is perfectly received, there remains only one microstate in the macrostate determined by the reception of the message.

It is important to recognize that for Shannon, what constituted a "message" can be treated at several different levels of perception. Suppose that the transmitter intended to send a message inviting the recipient to meet at a certain place and time. This message had to be translated into words, the words into a letter or sound stream, and the stream into a pattern of variation in an electric current. At the receiving end, the receiver might interpret the electrical pattern into a sound or letter stream. At one level of perception, the message would have been correctly received if the stream matched the transmitted stream. However, if that were all that happened at the receiving end, the message would not have been received. From the stream, the message still must be converted into words, and the words into an understanding that the transmitter wanted a meeting, that the meeting should be in a particular location, and that it should be at a particular time. Suppose the transmitter had identified the location as "at Jake's" and the receiver knew of no place that would be so identified. Would the message have been correctly received at that level? It would have been correctly received in that all the words in it had been exactly as transmitted, but the *meaning* would not have been conveyed. The message would have conveyed information about when the meeting was requested, but it would not have conveyed information about where. Without considering the meaning of a message to the recipient, the concept of information makes no sense.

Garner [6] treated observation of the world as the equivalent of receiving a message. The key difference is that the world had no intention of being observed, so the observation does not result in the attribution of intent to a transmitter. It does, however, have the same property of changing the observer's distribution of probabilities over the observable states of the world, and of creating a meaning in the observer's head. Before looking out of the window, you might say there was a 50 – 50 chance that it is raining. After looking out of the window, your probability distribution is likely to be 100 – 0. You now know whether or not it is raining, and if the result is "raining", that means you would get wet if you went out without protection. You have gained one bit of information. Of course, if beforehand you thought the odds were 4:1 that it was raining, you got less information from your observation.

Shannon initially used "uncertainty" for his "missing information" measure, until von Neumann suggested "Entropy" because of the identity of Shannon's and Gibb's equations. Garner [6] used "Uncertainty" exclusively, because "Uncertainty" takes the viewpoint of the receiver/observer, whereas "Entropy" seems to imply a value computed by some omniscient being able to know correctly all the possibilities inherent in a situation. Such omniscience may be available to the engineer designing a communication system, but it is seldom available either to the user of the communication system or to the observer of the world. We use "Uncertainty" in this Annex. Uncertainty is the opposite of precision. If Uncertainty is reduced by one bit, precision is doubled, or increased by one bit.

The concept of uncertainty implies that we take the viewpoint of Shannon's receiver, for it is at the receiver that the uncertainty of the message is reduced by what is transmitted through the communication channel. The receiver observes the transmitter through the channel in order to determine what the transmitter means by sending the message. The transmitter may not even be intending to send a message, but by observing, the receiver can learn something meaningful about the transmitter, regardless.

The word "observe" is important. For Garner, the receiver of a message is simply observing its environment through the communication channel. Whether there is an active transmitter at the other end is irrelevant and possibly unknowable. What matters is how the observation changes the receiver's ("observer's) understanding of the thing observed. If the observer initially believes a measure is between 4 and 6 mm, and after an observation believes it to be between 4.5 and 5.5 mm, the observation has provided one bit of information about the thing measured (has halved the range of uncertainty about the length in question).

The relation between Shannon's transmitter-receiver language and Garner's observer-observed language may be encapsulated in Table D-1.

Table D-1: Mapping Between Shannon's Communication Terminology and the Language of Observation.

Shannon Communication	Natural Observation
Transmitter	Observable World
Receiver	Observer
Channel	Sensors, Displays, and Processing
Transmitted Message	Sensor Data
Received Message	Observer's Perception and Understanding
Possible Distinct Messages	Possible Distinct States of the Observed World
Entropy	Uncertainty
Information	Information

"Information" is a term widely used, and, as a technical construct, almost as widely misunderstood. The problem is that the same word is used to mean several different but related concepts, three in particular within the Shannon conceptual structure. Information gained from a message, which is the change in uncertainty about the transmitter's intent as a consequence of receipt of the message; one bit of information equates to a doubling of precision. Information about the state of the transmitter, which is obtained from consideration of all the messages so far received from that transmitter:

> Information that could yet be obtained about the state of the transmitter, which is the remaining uncertainty about the transmitter's state. Sometimes this is misleadingly called "Information in" the transmitter. This usage is the one that causes most misunderstanding, as there is no sense in which "information" resides in anything waiting to get out. It will not be used in this chapter. However, a legitimate form of statement is "information potentially available about" the transmitter, and occasionally a form such as this may be used in what follows.

Although we will avoid mathematical descriptions where possible, one crucial formula should be kept in mind in any discussion of information and information. Because of interests of the moment, the receiver/observer may divide the messages or observations that might be received into classes. For the receiver/observer, at any moment "t" the ith possibility has a probability $p_t(i)$, where "i" ranges over all the possibilities. Given this set of classes and probabilities the receiver/observer's uncertainty at that moment about the transmitter's intent or the state of the thing observed is $U(t) = -\Sigma p_t(i) \log p_t(i)$, and *the information gained by receipt of a message or making an observation is the change in the value of U(t).*

Shannon's concept and Boltzmann's coincide at one point. Shannon's entropy and Boltzmann's both depend on the idea that a macrostate contains many possible microstates, and that only the macrostate enters into a measurement or an observation. For Shannon, the question is which macrostate contains the current microstate. For Boltzmann, the macrostate is known, and it is the microstate that is unknown.

The boundaries of a Shannon macrostate depend on the receiver's interest at the moment of receiving a message. The symbols A, a, 𝒜, a might be all in the same macrostate, depending on whether or not the receiver is interested in font and capitalization, but A and Q would usually not be in the same macrostate. Macrostate boundaries define the possible meanings to the receiver/observer of the possible messages or observations.

The receiver may be completely sure of having received a message correctly (uncertainty zero about the message) while remaining ignorant of other information that might have been obtained from the message had the receiver's interest been different. For example, in the days of paper mail the recipient of a typed letter might note only the sender and the content, whereas the detective, after the murder of the recipient, finds information in the paper quality, the envelope postmark, and DNA from the licked stamp, all of which could have been available but were of no interest to the original recipient.

It is worth re-emphasising the fact that information is always *about* something. The nature of that something, the possible meaning of the information, is determined only by the observer/recipient of the information. The receiver's initial uncertainty is determined not by the actual set of possible messages from which the transmitter might select the actual message, but by what the receiver believes to be the set of possibilities and the probabilities associated with the different messages in that set.

Likewise, although the possibilities are unlimited for what an arbitrary observer might observe of some part of the real world, any particular observer will observe only in which macrostate the observation belongs. Again, the set of possible macrostates and their initial probabilities is determined only by the observer, and it varies over time. Although the potential information obtainable from observation may be unlimited, the actual information obtainable is limited by the number of macrostates that make a difference to the observer – the limit is $\log_2 N$ bits (each of N macrostates having a probability 1/N). That limit is reduced if the observer knows anything about the observed part of the world prior to making the observation.

D.1.4 Kolmogorov-Chaitin

The Kolmogorov-Chaitin approach is sometimes called "Algorithmic Information Theory" [4]. The basic idea is that using a particular language and symbol set, the state of any system can be described in many ways. Of these, one is the shortest. This description can be considered as the maximally compressed representation of the system, and the uncertainty associated with it, treated as a message, is the "information in" (the uncertainty of) the system.

Using the Kolmogorov-Chaitin approach once again brings to the fore the importance of the observer's prior knowledge and processing capability, since in order to determine the state of the system from the compressed description, the observer must be able to decode the description. The more the observer knows about the system from previous experience, the less need be included in the compressed description, and the more the user knows about how to decode descriptions, the shorter the description need be in order to describe the system against a given background knowledge.

This annex will not refer very much to the Kolmogorov-Chaitin approach, because it is not practical to determine the shortest possible description associated with a world of which something is known to the observer, in a language incompletely understood by the observer. Furthermore, in principle the approach points up much the same set of considerations as does the Shannon approach: the importance of the observer's prior understanding, and the restricted range of distinctions that make a difference. This Annex will instead depend primarily on the more intuitive Shannon-Garner concepts of uncertainty and information.

No matter which interpretation of "information" is used, the values obtained by calculation are often limiting values and not the values that will occur in practice, which are usually difficult or impossible to compute, since they depend on the varying capabilities and interests of people. In particular, "channel capacity" indicates how fast information *can* get from A to B, but not how fast it *does* get from A to B, just as a pipe that can sustain a flow of 30 litres/sec may have an actual flow of only 1 litre/sec at some particular moment. In most of what follows, the consideration is more on the direction and relative magnitude of changes in uncertainty than on actual values. Hence we will have little or no recourse to mathematical derivations.

D.1.5 Fuzziness and Information

Fuzziness is discussed in Annex C, but because the mathematical concept of fuzziness is easily (and often) confused with uncertainty or probability, we mention it again here. The confusion may arise from the fact that fuzzy membership values, like probabilities, range between zero and unity (though in Annex C we argue that negative fuzzy membership values may legitimately occur). Fuzziness is actually quite distinct from any construct based on probability, as the following example illustrates.

How might one describe the height of a man? One might say he is of “average height”, “shortish”, “very tall”, etc. Whichever descriptor one uses, it fuzzily describes the man’s height. Imagine a man of 183 cm (6 ft). He might be described as being of “average height”, or perhaps as being “tallish” or “tall”, but one would probably not say he was “short” (unless he was being described in the context of his membership in a basketball team). Such a man might have a membership value of about 0.4 in the class “average”, near 1.0 in the class “tallish”, around 0.9 in the class “tall” and a membership near zero in any other class. An entity can have high membership values in several fuzzy classes without contradiction, whereas for probability if the entity belongs in one class it does not belong in any other, and summed over the possible classes, the probabilities that it belongs to each must total 1.0.

Contrast the overlapping classes of fuzzy membership with the concept of probability. If you are uncertain about the man’s height, it is not because he can be at the same time 183, 182, and 184 cm tall, but because you do not know which of those values is closer to the truth. Only one can be the closest, and the probabilities summed over all possible choices of which one it is must add to exactly 1.0. But for membership in fuzzy classes, no matter which of these heights is nearest the truth, and even if you know precisely that the man is 182.763 cm tall, the man may be described as “average”, “tallish” and “tall” without contradiction. He has significant membership in all those classes, and the sum of his membership values has no specified maximum other than that implied by the fact that the maximum membership in any single class is 1.0.

Fuzzy membership may sometimes be related to likelihood, which is itself related to probability. Given the datum that a man is described as being of average height, the likelihood of his being 145 cm tall is very low, as would be his membership in the class “average height” if he did happen to be that short. For any suggested height “X”, the likelihood that his height is exactly X varies in much the same way as would the fuzzy membership of a man of height X in the class “average height”. The difference between them may be encapsulated in the difference between the conversational phrases “can reasonably be called” (fuzzy membership) and “is likely to be” (likelihood). Likelihoods are relative to each other, and have no set range of values. What matters with likelihood values is the ratio between the likelihood of one value to that of another.

Since men come in a wide range of heights, one has a certain expectation about what heights are more probable than others. Full grown men are seldom under 1.5 m or over 2.5 m, for example, though some few are outside these limits. Knowing no more than that someone is an adult man, one has some kind of mental probability distribution for the likely height of the person. However, if you are told that he is of average height, your uncertainty about his height has been substantially diminished. Neither your initial nor your final uncertainty is well defined, because the actual uncertainty depends on the size of your macrostates. Are they a millimetre or an inch wide? The answer determines the probability that the man’s height lies within any particular macrostate for you. But the reduction in uncertainty consequent on finding that he is of “average height” is defined by the change in the width of your probability distribution over heights. If your range is reduced by a factor of four, you have gained two bits, no matter how fine or coarse your macrostates might be.

D.1.6 Observation and Mutual Information

We will consider an observation to be a time-limited period during which information about the observed part of the world passes through an observation channel to the observer. In the later part of this Annex, “the observer” is often a node in a network, not necessarily an observer of the network. The viewpoint of the observer must always be kept in mind. Sometimes it is that of the external analyst observing the network, and at other times it is that of a point in the network. The analyst-observer can know things about both ends of a network link in the same way that Shannon’s engineer can know the probability distributions at both ends of a communication channel before and after a message is sent.

The information in an observation is ordinarily not selected by the thing observed, in contrast to the information in a message, which depends on an active selection by the transmitter from the ensemble of possible messages. The receiver of a message learns something of what the transmitter may have meant by sending the message, whereas an observation has meaning only to the observer.

D.1.7 Uncertainty and Belief

An observation provides information about the world observed. But what of the information gathered from previous observations? Some, at least, may still be valid. One could not act effectively in a world that changed so radically from observation to observation that the probabilities of the possible states of the world reverted to their baseline values for each observation. What the observer knows about the world grows, on average, with successive observation. In other words, the observer’s uncertainty about the observed part of the world is usually reduced by an observation. We deal later with situations in which a new observation contradicts an older one, a situation that might well occur when one has unreliable or untrustworthy informants who form part of a commander’s or a politician’s observation channel onto the world. The observer must develop a degree of belief in something about the world as a consequence of the observation and its assessed reliability.

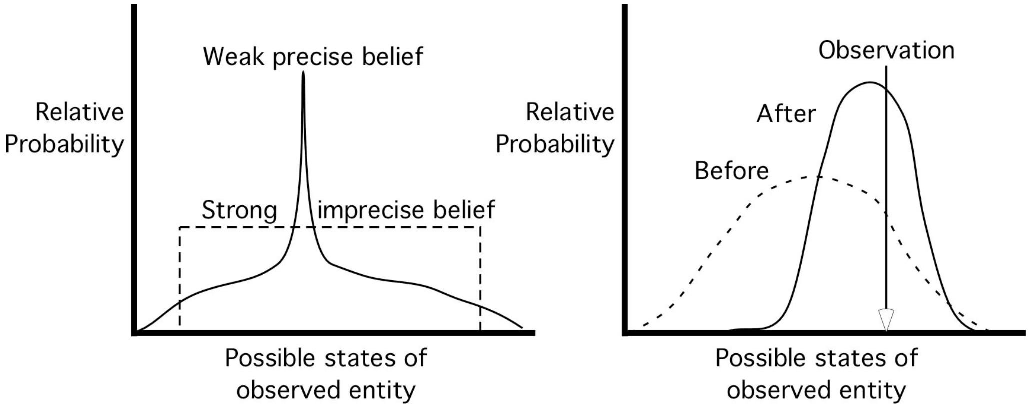


Figure D-1: Probability Distributions, Belief, and Observation: (Left) a Precise Belief has a Small Region of High Probability for the Possible States of the Entity, Whereas a Strong Belief has a Sharp Distinction Between States Believed Reasonably Probable and States Given Very Low Probability; (Right) the Effect of an Observation is Usually to Increase Both the Strength and Precision of Belief.

The concept of belief occurs throughout this Annex. As used here, it has a technical meaning based on, but not identical to, the everyday meaning of the word. As we use it, “belief” always refers in some way to the distribution of probabilities over the possible states of some entity. The everyday meaning might be encapsulated in statements such as “I believe he said that the river is flooded” which might lead to “I believe the river is flooded”. In our technical usage, such statements specify a peak of the probability distribution over possible states such as <he said the silver is corroded>, <he said the middle is muddy>, <he said the river is flooded>, and <the river is dry>, <the river is ankle-deep>, <the river is flowing well>, <the river is flooded>.

Shannon’s entropy [6] measure is one parameter of the probability distribution over the states of the entity. In the first set of examples, the entity is “what he said”, whereas in the second set it is “the state of the river”. In either case, the uncertainty is measured by $U = \Sigma\ p(i) \log p(i)$, where “i” represents either the possible things he might have said or the possible states of the river. The everyday concept of belief is that one believes the “i” for which p(i) is greatest, or those “i” for which p(i) is near its maximum.

If the observer’s interest is in the state of the river, the probability distribution over believed flood states depends both on the probabilities of “what he said” and the distribution conditional on each of the possibilities for “what he said”. For the ith flood state, $p(i) = \Sigma\ p(\text{statement})*p_{\text{statement}}(i)$ where the sum is over the possible statements. In passing, it should be noted that all probabilities have this kind of conditionality. A probability has a particular value only in the context of certain conditions being true. This fact is often ignored, and probabilities taken as absolute measures, but underlying all such “absolute” measure is a tacit understanding that they apply only under a presumed set of conditions.

We extend the everyday concept of belief to define “strong belief” to be an uncertainty structure for which there is a clear break between a set of states of near equal, near maximum probability and all other possible states, for any of which the probability is very low. Weak belief is the contrasting condition in which there is no clear break between the higher and the lower probability states. “Strong” and “weak” are fuzzy descriptors of the shape of the probability distribution.

The other dimension in which we extend the concept of belief is that of precision. A precise belief has a sharp maximum that extends over a small range of the possible states. In contrast, an imprecise belief describes a probability distribution in which a wide range of states have near maximum probabilities. “Precise” and “imprecise” are two more fuzzy descriptors of the probability distribution.

It should be clear that the concept of “state” has an important bearing on the apparent precision of a belief. A “state” is clearly a macrostate for the person defining it. If one person’s macrostate is seen by another person as being a collection of smaller macrostates, the two will have different ideas about the precision of the first person’s belief. On hearing that the river is flooded, one hearer may consider that to be a precise description, whereas another hearing the same thing may want to know whether the flood has reached his mother’s house. The first person has a strong precise belief: “The river is flooded”, whereas the second has an imprecise belief that encompasses two equally probable macrostates: “the flood reaches my mother’s house” and “the flood does not reach my mother’s house”. The first may gain no more information from hearing “The flood has reached Front Street”, while for the second, hearing this may provide as much as 1 bit of information, and for a third person interested in mapping the precise boundaries of the flood, it might provide several bits of information.

D.1.8 Information from Disparate Sources

Suppose an observer gets information from two sources about the value of some property. One says “the height is around 180”, while the other says “the height isn’t far from 185”. Both are somewhat imprecise, but they are compatible with both having been reports of the height of the same thing. The observer might be justified in thinking that the height is likely to be between 180 and 185, with an uncertainty distribution rather narrower that that justified by either report alone, as suggested in Figure D-2a.

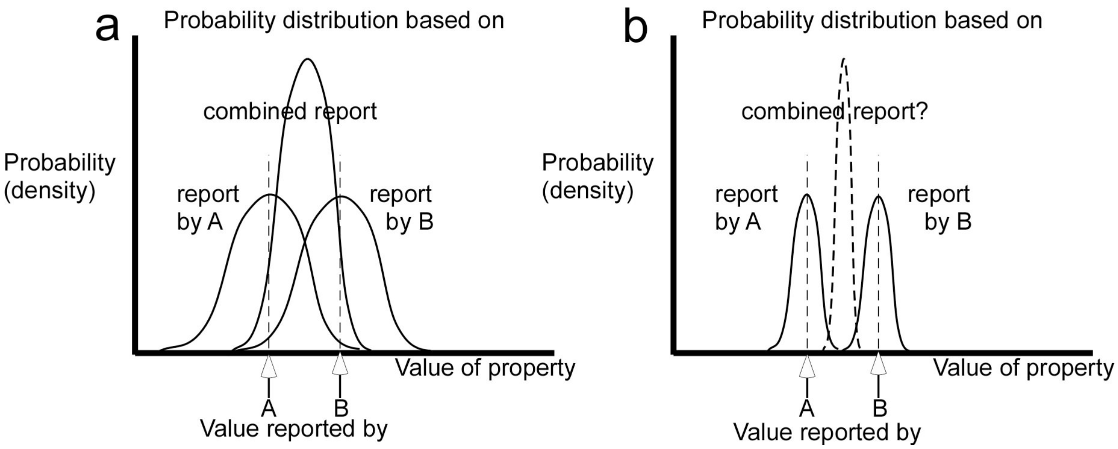


Figure D-2: Combination of Information from Disparate Reports: (a, Left) the Reports are Compatible with a Value Between Them; (b, Right) the Reports are not Compatible with Both Being Valid Reports of the Same Thing.

But what of a situation like that of Figure D-2b, in which the two reports, ostensibly of the same thing, are incompatible. A has said “the height is about 180.2” while B has said “the height isn’t far from184.8”. They cannot both be reliable reports of the same thing, and the observer is clearly not justified in an uncertainty distribution centred on a value between the two. One or the other must be in error, or else the height changed between the two measurements and neither represents the current height. The observer must assign some probability to the various possibilities and to other possibilities that might present themselves (e.g. B always overestimates by about 4.5), and use the appropriate combination of probabilities to produce a joint probability distribution for the height. Whatever way the observer combines probabilities, the resulting distribution determines the uncertainty measure for the height.

It is quite possible that before receiving the report from B, the observer had a fairly sharp probability distribution, whereas after receiving the contradictory report the probability distribution is flatter, and the observer’s uncertainty has increased. In such a case, the information gained from the observation of B’s report would be quantitatively negative. From B’s viewpoint, the transmitted information was positive, so how could the received information be negative? The answer is that information is not a conserved quantity like energy. Like physical entropy, uncertainty usually trends in one direction, but again like entropy, reversals can occur. The observer did receive a positive quantity of information about what B said, but a negative quantity of information about the state on which both A and B reported. The meaning of the message determined the quantity of information received; had B’s report meant something different from the meaning of A’s report, there would have been no local increase of uncertainty about either, and no reception of a less than zero quantity of information.

Prior probabilities become quite important in determining the post-observation probability distribution. If the observer strongly believes that the entity does not change in height, the situation is quite different from that if the observer believes it is continually changing and that B made the measurement some time later than A. In the latter case, the reports mean different things. They provide information about two separate topics, which could be treated as the height and the rate of change of height. With reports such as those of Figure D-2a, the observer gains little information about rate and much about height, whereas the reverse is true with reports such as those of Figure D-2b. We will explicitly or implicitly consider observations of dynamic systems in much of what follows.

D.1.9 Belief About a Changing World

If the observed world was static, the observer could determine all its properties with indefinitely fine precision, simply by making enough observations. But worlds do change, and the information gained about any state from an observation becomes less reliable over time. Like entropy, uncertainty usually increases.

At this point we must introduce the concept of "Mutual Information", without (yet) getting into its mathematical expression. Consider an observer O (Figure D-3) who is able to see two systems, A and B, with perfect precision and understands precisely their relationship. Before observing, O has uncertainty U(A) about A and U(B) about B. After observing A with absolute precision in the sense that O now knows the macrostate of A exactly, the observer has gained information equal to U(A). But what does O then know about B? O knows the relationship between A and B, such as their degree of correlation. A and B may be highly correlated, so that all of A is included in B, or all of B is included in A, or may have some less rigid, but still significant, relationship. If so, then by learning all about A, this unnaturally knowing observer has reduced the uncertainty about B. The difference between the observer's uncertainty about B before and after observing A is the mutual information between A and B, U(A:B). Its value is symmetric; U(A:B) = U(B:A), and has its maximum value at the lesser of U(A) and U(B).

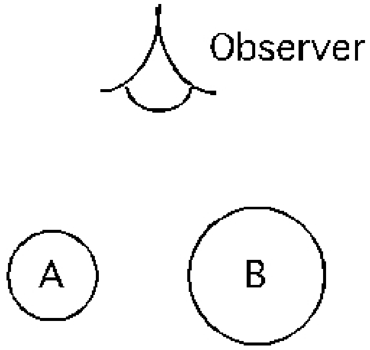


Figure D-3: An Observer Observing Two Possibly Related Systems.

Now imagine that B is observing A, and the observer O can see both. Figure D-3 schematizes some of what happens. In Figure D-4, some entity A (left column) is observed by some other entity B (right column). Entity A is illustrated as having ten properties, and five of an indefinite number of properties of the observer B are shown. Two of these five represent B's understanding of two of A's properties. Supposing that each of the properties could be described using five bits, the uncertainty of entity A to O is 50 bits, and the uncertainty of the illustrated part of observer B is 25 bits, making 75 bits in all. But if O knows that B has made the observation and has learned all 10 bits that describe two properties of A, then O will need only 65 bits to describe all the properties.

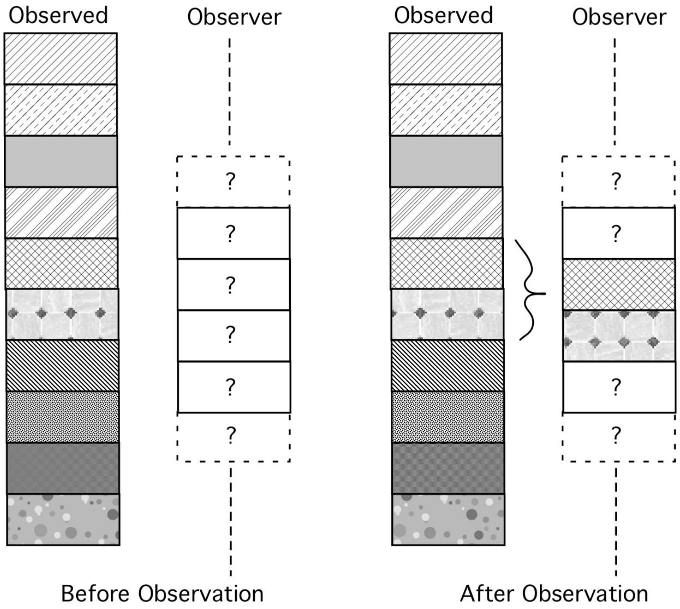


Figure D-4: An observed entity has several possible properties. Before making an observation, the observer does not know their values. After making the observation the observer knows some of those values. The information available to an external viewer of the joint system of observer and observed is less after the observation than before, because some of the properties of the observed influence the properties of the observer.

This is an obviously crude illustration, and in practice many more bits would be required to define the relationships between observer and observed, but it suffices to illustrate the point that mutual information between observer and observed reduces the uncertainty of the joint observer-observed system.

When we talk about B observing A later in this chapter, B need not be an active player. "Observing" implies no more than that because of information about A reaching B, the mutual information between A and B is increasing.

D.1.10 Information Loss Rate

Now suppose A is a system observed at time t_0, and B is the same system observed at a later time, t_1. Now the mutual information between A and B is a measure of how rapidly the system is changing. If it is static, then the mutual information is just the uncertainty of A. Having obtained all the available information from A, the observer need never look at it again, provided the observer knows it will not change. On the other hand, if A is changing unpredictably but slowly, the mutual information between the original observation and the current state of the system will slowly decline to zero.

Though information may not be conserved, uncertainty can be partitioned. In fact, Garner and McGill [7] showed that there is an exact parallel between the analysis of variance and the analysis of uncertainty, with uncertainty taking the place of variance at every stage. At this point, we need only to know that the sum of the mutual information between A and B and the uncertainty of A when B is known are constant and equal to the uncertainty of A before any observation. So, as time since an observation increases and the mutual information between the observed state and the current state declines, the uncertainty about the current state increases in complementary fashion, as shown in Figure D-5. This rate of increase is often known as the "information generation rate" of the observed part of the world, whereas it is better described as an "information loss rate" or "uncertainty generation rate" for the observer. What is lost is the mutual information between the observer and the observed. What is generated is the observer's uncertainty about the current state of the observed entity, which is the same as the information still potentially available from it.

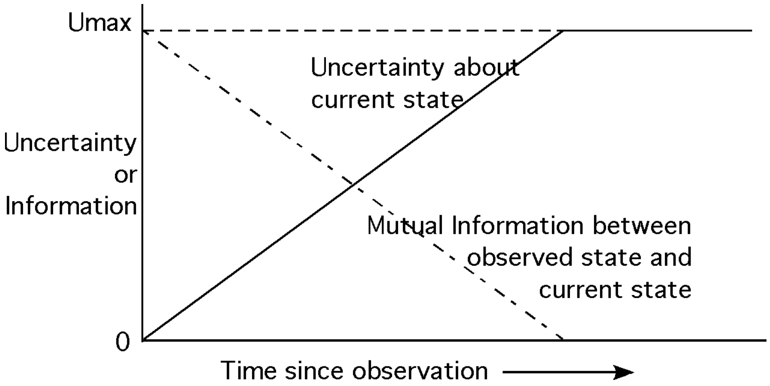


Figure D-5: Decreasing Mutual Information and Increasing Uncertainty Since an Observation.

In linear systems there is a direct parallel to the information loss rate, the autocorrelation function. As time passes, the average correlation between the current value of a variable and its value at the time of a measurement decreases from 1.0 to 0.0. The time course of this decrease may not be monotonic. Indeed, for some signals the autocorrelation function may oscillate between +1.0 and –1.0 as time passes, but eventually it will subside to zero. Likewise, the mutual information loss function will decline from Umax at the time of the observation to zero and while the loss may not be monotonic, eventually the mutual information will subside to zero.

D.1.11 Channel and Channel Capacity

As Shannon described a communication system, a message can get from transmitter to receiver by way of encoding hardware and software at the source, which results in signals sent over the air or through wires or optical fibre, and then through decoding hardware and software in the receiver where the meaning is interpreted. All of that apparatus constitutes a "channel", as suggested in Figure D-6.

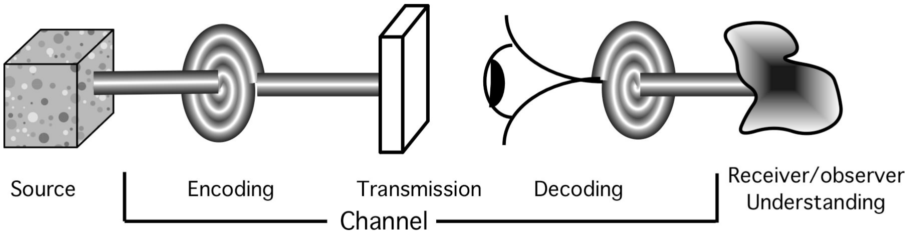


Figure D-6: Schematic of a Channel. A source has much potential for observation. Some aspect of the source is observed. The result is encoded for transmission through various media, which might include human senses, and is decoded to the receiver, which could be human understanding.

When we talk about an observation, we again must consider a channel between the observed system and the observation made by the observer. For unaided viewing, the channel may consist of the space between the object and the eye, the sensitive retina, the optic nerve, and a lot of processing in the brain. When we observe some data in a computer, the channel includes the processing needed to construct the image on a display screen, the screen itself, and all the apparatus used in direct viewing. To put it shortly, a channel is whatever allows the observer to get information from the thing observed.

Channels have limits. For purely physical reasons, every channel has some latency, if only that due to the speed of light. It takes time for any change at the source to have its first effect at the observer. In any realistic case, most of the latency is likely to be in the processing required to create the observation rather than in the physical transmission across space. Channels also are limited in how fast they can transmit information after the latency period is finished. The two physical limits are the speed with which the channel can accommodate changes in the input value (its "bandwidth") and how precisely the instantaneous input value can be specified (its "Signal-to-Noise Ratio", or SNR). Bandwidth determines how many statistically independent values can be transmitted per second (for a bandwidth of W Hz, 2 W independent samples can be transmitted per second). Signal-to-Noise ratio determines how many bits each of those samples can transmit. Overall, any channel can transmit at most some calculable number of bits/sec, its "Channel Capacity" or simply "Capacity".

D.1.12 Channel Distortion and Noise

Shannon's main concern was that channels are imperfect, and that by observing what arrived at the receiver, the receiver might not be able to determine what was sent by the transmitter. Shannon treated all forms of distortion and interference as "noise", something that altered what had been sent into a possibly different thing that was received. His major contribution was to show that regardless of the noise in a channel, by incorporating redundancy in the encoding, the channel could communicate information at a rate up to a limit that could be specified in bits/sec, but at no more than that rate. The limiting rate is known as the "channel capacity", symbolized as "C".

In a classical linear system, noise could be treated as a random value added to the transmitted value to create the received value. The range of possible values for the transmitted signal could be treated as signal power, and the range of added values as noise power, the ratio of which was Signal-to-Noise Ratio (SNR). We deliberately do not offer the standard electrical engineering formulae here, because we wish to treat the concept as more generic. The noise must transmute what is received from one possibility to another within the set or range of allowable possibilities. The wider the range over which the transmutation can occur, the lower the SNR. Whereas Shannon measured signal and noise in volts, we can measure it in terms of the probability distributions over the different possibilities. Noise and distortion both alter these probabilities.

D.1.13 Information, Channel Authenticity, and Meaning

It is usually asserted that Shannon denied a connection between information and meaning. He did not. What he did say was that when designing or measuring a communication channel, the information measures were irrelevant to the meanings of whatever the users of the channel might want to communicate. But at the receiver, the meaning becomes all-important, since the receiver's ability to determine what the transmitter sent depends crucially on the receiver's prior probabilities for what the transmitter might have sent. If the transmitter had been talking about financial matters and the receiver heard "bank", it would be unlikely for the receiver to interpret that as meaning the boundary of a river, or as a French word for a bench.

The important point is that the observer or receiver is always getting information *about* something through the channel. "Information" does not arrive in grand isolation. If it does not connect with something already available to the observer, it is meaningless in the sense that it does not change the observer's uncertainty about anything.

That the information is about something leads into the question of whether the channel is in fact transmitting information about what the observer/receiver believes it is about. Is the microscope pointing at the right piece of tissue? Does the document actually come from the source it purports to come from or is it a fake meant to mislead? If the observer assigns a probability appreciably less than 1.0 to the answer that the source is what it should be, that probability must enter into all the calculations of uncertainty *about* the thing supposedly observed. The calculations are similar to those required in situations in which the actual words of a message might be misinterpreted, as in the "river flood" example above.

Uncertainty as to the quality of the source can be treated as an irreducible noise that cannot be corrected by reference to redundancies in the transmitted material. The receiver may be able to receive the message with perfect fidelity, but remain quite uncertain about the state of the world to which the message seems to refer. Maybe the channel has been subverted by an enemy and is reporting the wrong things about the right part of the world. Maybe the supposed 17th century document is an 18th century forgery. How do these doubts affect the observer's uncertainty about the aspect of the world that is supposedly being observed?

There are two possibilities, "T" (the channel is authentic and has not been subverted) and "F" (either the channel is reporting data about the wrong part of the world or has been subverted and is reporting deliberately wrong data about the right part of the world). The "F" possibility has many variants, but here we assume that the observer cannot distinguish between them. If the observer believes "F", then nothing received through the channel will affect the observer's probabilities p_i for the different possible states of the thing observed. On the other hand, if the observer believes "T", then the normal change of probabilities as a result of receiving the information will occur, changing the observer's uncertainty about the state of the thing observed.

Label the observer's probability that the observed world is in state "i" before the observation as p(i) and as $p_O(i)$ after the observation. If the observer neither believes nor disbelieves T and F, T and F become possible states of the world, to which the observer can assign probabilities p(T) and p(F). The observer's probability for state "i" after the observation then becomes $p_O(i) = p(T)p_{T,O}(i)+p(F)p(i)$, and the observer's uncertainty about the state purportedly observed is $U = -\Sigma p_O(i)\log p_O(i)$.

D.1.14 The Effect of Channel Latency on Observation Precision

If the system is observed through a channel of latency L, the observed state is L seconds different from the current state of the system (current in a synoptic, or "Newtonian", view of the world). If the system has an information loss rate of G bits/sec, the observer's minimum uncertainty about its current state is G x L bits.

Ordinarily, an observer has some prior notion about the possible range of variation of the observed entity. Suppose the uncertainty associated with that range of variation is Umax. If the information loss rate of the entity is G, and the time since the last observation is L, the uncertainty is G x L if G x L < Umax, but Umax otherwise, as shown in Figure D-7. If the channel latency is too long, the observer can never learn anything about changes in the observed entity. All the observer can do is compile statistics about what the entity tends to do, so as to look for patterns that might allow the observer to improve predictions of the changes, and thereby reduce G.

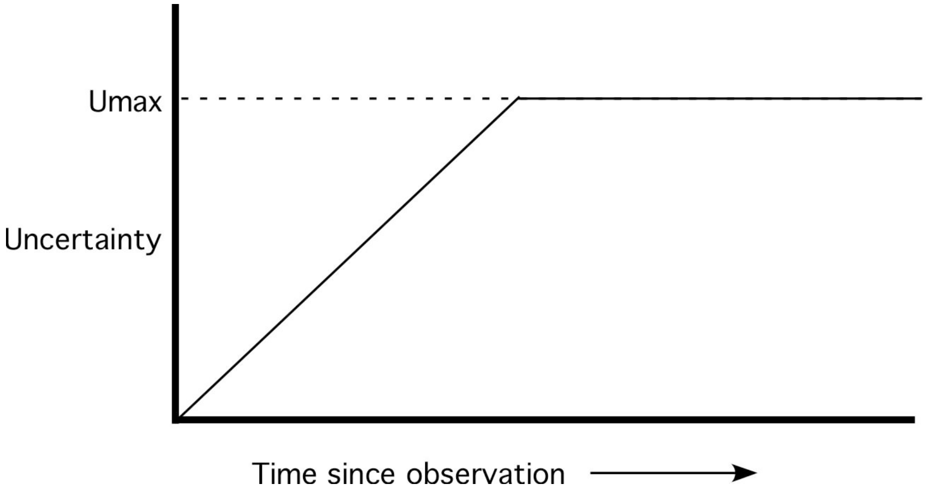


Figure D-7: The observer's uncertainty about the current state of the thing observed increases over time since the observation at the "uncertainty generation" or "information loss" rate G bits/second.

So far, we have not considered the effect of channel capacity C. On average, provided C > G, C does not matter, but it does matter when the source changes abruptly. It takes time for the observer to acquire precision about the change, and C determines the rate at which precision can be acquired. At the moment of the abrupt change, the observer has an uncertainty $U(t_0)$ about the new state. Each bit of information acquired by further observation reduces this uncertainty by one bit and increases the precision by a factor of two. But at the same time, the thing observed has an uncertainty generation rate G, so the rate at which precision is increased is not C, but C – G.

Some methods of compression for moving images take advantage of this fact, recognizing that the channel capacity of the visual processing system is limited. If the motion of a scene object reveals something that was hidden, the hidden object is drawn initially only crudely, and with increasing exactness over the next few frames. The "potentially available information" from the revealed part of the scene grows only at a rate compatible with the channel capacity of the visual process, allowing more transmitted bits to be devoted to increasing the displayed precision of the stable parts of the scene.

In many cases, the observer has limited attainable resolution, which determines a lower limit to the size of the Boltzmannian macrostates within which differences of microstate cannot be distinguished by the observer. If a human is only able to discriminate two dots separated by one minute of arc, it makes no difference whether they are separated by 10 seconds of arc or 5, even if the ongoing flow of information from the dots continued through a channel that could provide hundreds of bits per second. Resolution limits of this kind put an upper bound on the information that could be obtained from an observation, or, equivalently, to the maximum uncertainty of the thing observed. When we deal with displaying data, this becomes an important consideration.

D.1.15 Information Flow to the Analyst

The Analyst, the one who has an interest in understanding something about the world for some purpose, is no different, informationally, from any other observer. We will call the Analyst "Me" for easy reference. I have a current topic of interest, and have some level of uncertainty about aspects of that topic. By observing, I can reduce that uncertainty, gaining information from the observation channel, a channel that may be conceived as being like that of Figure D-6, above.

Since I am interested in only a very small part of the world I might be able to observe, the information available is only a small part of the information that could be available from the observable world. The information I have already gained is a smaller amount yet. These relationships are symbolized in Figure D-8 in terms of the remaining uncertainty I have about the world, the observable world, and the topic of current interest to me. All of those have been reduced only by the amount symbolized by the darkest, smallest patch in the figure. And all is changing in parts of the world that I am not watching.

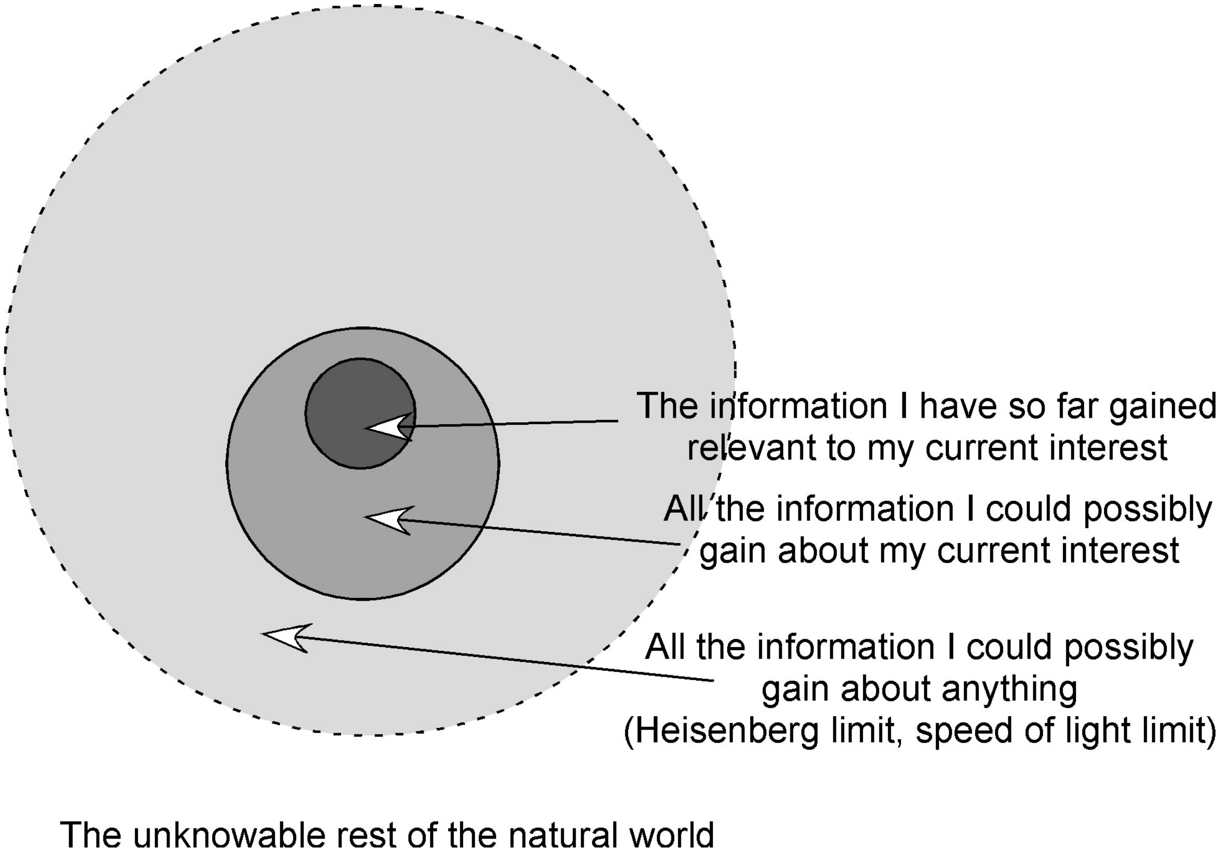


Figure D-8: Uncertainty Regions. From the middle outward, the smallest disk represents the reduction in uncertainty (information gained) about the topic due to prior observation, the next disk represents the initial uncertainty I had about the topic, which is the information that could be gained by observation, and the outer disk represents all the information I could possibly gain about all possible topics in the world by observation. Outside the disks is everything that I cannot learn about the world by observation. The boundary of the large disk is shown dotted, as it changes with observation technology.

The same idea can be represented in more detail by referring to the channel between the Analyst's target of interest and the Analyst's understanding of the target by way of a representation in a computer of the world of interest, as suggested in Figure D-9.

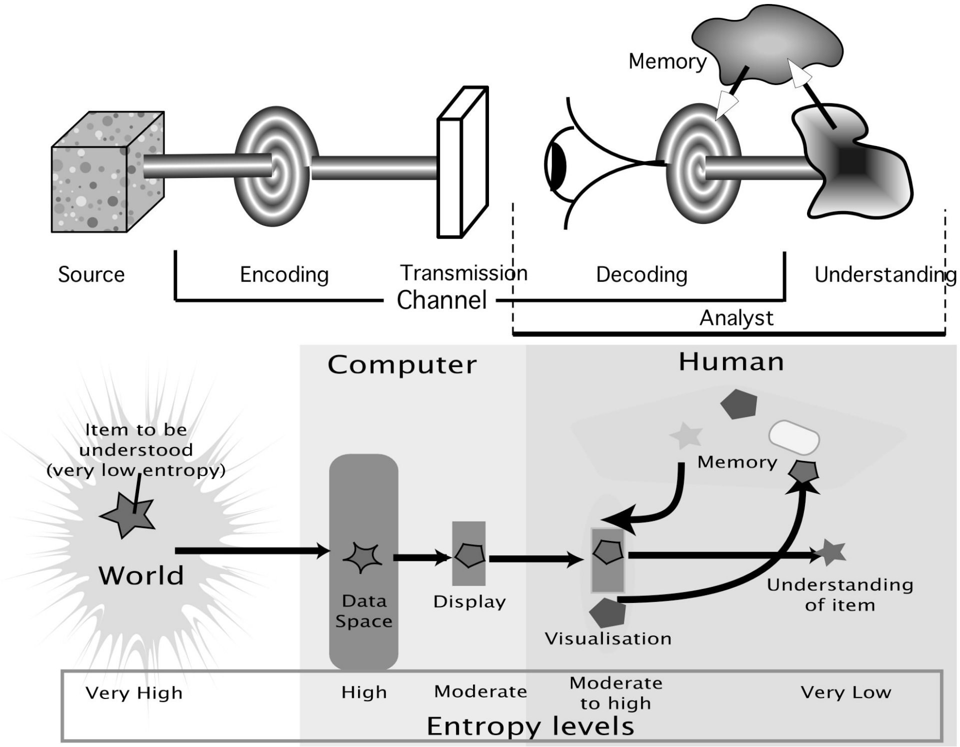


Figure D-9: Schematic of Entropy (Before Observation Uncertainty) Levels at Different Stages in a Multi-Stage Channel Between the Real World and the Analyst Who Uses a Computer Display to Help in Understanding a Small Aspect of the World.

The Analyst is interested in only a very small part (symbolized by the star) of the very high entropy world. The question at issue may require only one bit of information “Is there an X in the area – Yes/No?”. However, the data required to answer the question may be widely distributed across observable elements of the world, some of which might not be represented in the data space of the computer (represented by a slight change in the shape of the star). To illustrate the dimensions of that problem, suppose that the question was “Is there a Higgs Boson in the universe?” and consider the enormous effort that has been expended to get that one bit of information from observations of the Universe.

The possible contents of the data space have a very high uncertainty, but although it is very high, the uncertainty of what is in the data space is orders of magnitude less than that in the world it represents. The reduction in available information is dramatically greater at the display surface, and it is likely that more has been lost of what might be useful in perceiving the one bit the Analyst wants.

Despite these dramatic reductions in available information in the passage from real world to data in the computer to display surface, the human Analyst has a trump card – memory. If the Analyst is skilled and experienced, memory may well be able to fill in elements that are missing from what could have been available from a direct view of the real world, and the results of the Analyst’s visualisation of the current situation will be in memory, ready to be used in some later task.

We now consider the implications of informational analysis within networks, recognizing that every link in a network represents a place where the mutual information between two nodes is likely to be non-zero.

D.2 NETWORK ANALYSIS PART 1: TRAFFIC IN A STABLE NETWORK

D.2.1 Nodes, Links, Channels, and Observation

In talking about complex channels, we begin to invade the realm of networks, and it is in the network domain that this Annex continues. In this section, we consider the information-theoretic analysis of traffic in a structurally static network – one in which the properties of its nodes and links do not change over time. “Traffic” in such a network goes from one node to another, and the arrival of traffic at a node constitutes an observation by the arrival node of the node that originated the traffic. In most of this chapter, we use the more general term “observation” in place of the word “traffic”, since “traffic” often connotes the movement of physical objects, whereas “observation” includes the movement of photons, electron packets, acoustic pulses, and all kinds of other things, not excluding cars, planes, and trains.

Not all networks carry traffic or have nodes that observe other nodes. Networks such as family kinship, the set of all http links on Web pages, or the set of syntactic relationships among the words in a text do not carry traffic, though they may enable traffic. For example, the set of all http links enables the passage of http requests to remote pages, and the set of links used over a defined span of time is a traffic-carrying network. In the same sense, the road network enables the passage of motorized traffic, but is itself simply a set of relationships among the nodes that are the junction points of different roads – albeit a set of relationships embedded in a spatial (geographic) pragmatic embedding field. Traffic-free networks are discussed in a later section of this chapter.

Remember that in an “observation” there is no need for active participation of the observer. “To observe” means only that the mutual information between the observer and the observed is increased by virtue of the traffic between them. In network terminology, a node observes another node if there is a path from the observed node to the observer node. If the path consists of a single link, the observation is direct (as when one sees something with one’s own eyes), but if the path is longer, passing through at least one intermediate node, the observation is indirect (as when one is told of something by another person).

D.2.2 Node Types

Nodes come in three classes, as shown in Figure D-10. Pure source and pure sink nodes are more of an analytic convenience than representative of anything in the real world symbolized by the network. A source node can represent any undefined or undetermined source of observable variation from outside the network, while a sink node is an indication that something unspecified may be observing, but not contributing the results of its observations back into the network. Almost all nodes are of the “common” or “transceiver” type that accepts inputs and produces outputs that depend at least in part upon those inputs.

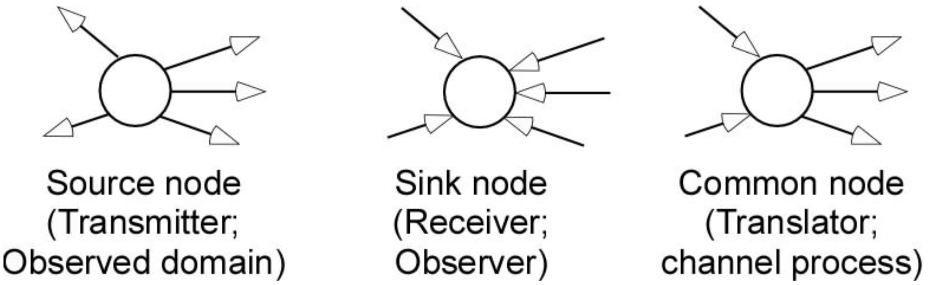


Figure D-10: Types of Node. A Source is an information generator, a sink is an observer, and a common node is both. Almost all nodes in a network are likely to be of the common type.

D.2.2.1 Nodes May Process Their Inputs to Produce Their Outputs

Common nodes are assumed to have memory, in the sense that their outputs may depend in part on past values of their inputs. One implication is that although an output may be totally dependent on a single input, its values over time may follow an entirely different course from those of the input. To make this concrete, consider a node with one input and one output, both of which are simple variations of a value over time. The node's output is the integral of its input. The input and output are uncorrelated according to standard statistical measures, but the output contains all the information about the input, as can be seen by the fact that the derivative of the output will reconstitute the input waveform exactly apart from a possible level constant.

It is therefore important in all further discussion to keep in mind that perfect informational transmission does not mean that the form of what is transmitted remains unchanged. The form may change radically without loss of information about whatever the original was about. For example, the input might be a meeting invitation in English and the output a good Chinese translation that conveys to a Chinese speaker all the information about where and when the meeting is to be held that the original would have conveyed to an English speaker.

D.2.3 Links and Channels

As repeatedly stressed above, all information is *about* something. When a node observes another node through a link, as B observes A in Figure D-11a, the observation necessarily is of a state of the observed node, A. But, as shown in Figure D-11a, A's state is likely to be influenced or even determined by some observation that A has made, whether it be of another node or of some state of the world outside the network (the "pragmatic embedding field" of the network, discussed in Annex C). Often, the observing node B gets information not only about the state of the observed node, A, but about the state of whatever A observed. The observed node then forms part of the channel between the entity of interest to the observer and the observer. It is a translator.

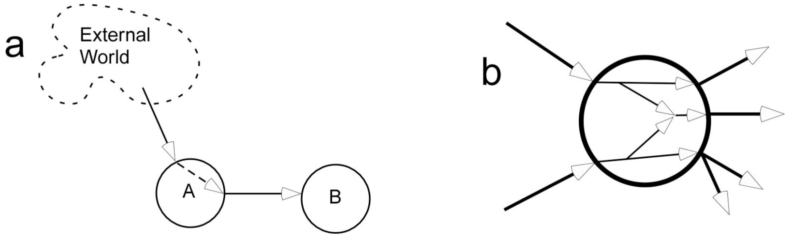


Figure D-11: Nodes as Components of Communication Channels: (a) Simple Onward Transmission, in which Node B Virtually Observes the External World Through the Observations Made by Node A; (b) Complex Multi-Channel Transmission, in which Two Non-Interacting Channels Pass Through the Node and Both Contribute to a Third Channel that Might be Observed by Another Node.

Figure D-11b illustrates that nodes can transmit information from more than one input to more than one output. Several non-interacting channels can pass through the same node, and may at the same time contribute to a novel possibility for another node to observe some combination of the two non-interacting inputs. The important point, though, is that whatever nodes observe the outputs of this complex node, they are obtaining information about what the complex node observes.

Within a network, all channels are composed of one or more links between nodes. The nodes themselves can be information generators or observers or both. The issues addressed in this chapter concern how information gets from one part of the network to another. If one node emits "traffic", when and how much could this event influence anything about another node? If the structure of the network changes in one region, when and to what degree might this influence the traffic or the node behaviour in another part of the network? This latter question also addresses issues of network robustness and vulnerability.

D.2.4 Complex Channels

In a network, a simple or basic channel consists of a single directed link between nodes. It has an input terminal (an “out-link” from the transmitting node) and an output terminal (an “in-link to the receiving node). However, in almost every real network, paths between at least some pairs of nodes can take multiple routes, and at least some of these routes will consist of a series of links. Such complexes of link structures that connect nodes are the constituents of complex channels. There is only one channel, but it has a complex structure.

Figure D-12 shows an observer B obtaining information about entity A by different routes (Figure D-12a) or using C as an intermediary (Figure D-12b). In either case, there is but one channel from A to the observer B, but it is a complex channel, labelled X in the figure. Now we consider its lag and capacity as determined by those of the component elementary channels or links.

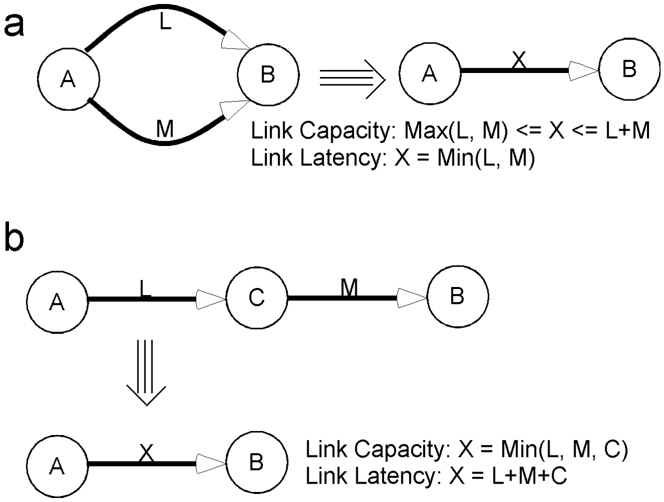


Figure D-12: The Two Constituent Components of Complex Channels.

The basic computational statements concern the relations of the latency and channel capacity of the complex channel to those of the constituent links. For the most part, these are self-evident, but they should be stated, nevertheless.

D.2.4.1 Parallel Links

If a channel consists of two parallel links, its capacity cannot be less than that of the higher-capacity link by itself unless one of the links is subversive, deliberately sending contradictory information. The channel capacity might be as high as that of the two links summed if they pass independent components of the information. For the time being, we will discount the subversive possibility. The actual capacity of the complex depends on the relationship between the two links, since if they always provide the same information, the lower-capacity link is redundant, and serves only as a backup in case the other is broken, whereas if they provide unrelated information (the mutual information between the information gained from each is zero), the channel bandwidth is the sum of the two individual bandwidths.

Quite often, the mutual information is non-zero (correlated but not identical), but less than the information gained from the lower-capacity link alone. In this case, the channel capacity is intermediate between that of the higher-capacity link and the sum of the capacities of the two links. In the analysis of more complex networks, maxima and minima can be described, but the actual channel capacities of complex paths cannot, unless the relationships between parallel paths are known.

In contrast, the computation of latency for a channel consisting of two parallel links is self-evident. It is the lesser of the latencies of the two links considered separately, since latency refers to the moment the first element of information arrives from the source.

D.2.4.2 Series Links

When we consider a path consisting of links in series, there is an added complication, namely the intervening node. In a dynamic network, nodes are more complex than just simple junctions where incoming traffic becomes outgoing traffic. They may have memories and perform computational processes on their inputs to provide their outputs. For example, if its inputs and outputs are values of a continuous variable, a node might output simply the integral of its input. Or, after having been provided with a complex set of data over several eons, it might output simply “42”.

Nodes have their own properties of throughput channel capacity and latency. For our purposes here, we will treat nodes as if they were complete sub-nets. Doing so allows us to treat defined sub-nets as though they were nodes, and simplifies the analysis of modular networks, where perhaps the modules represent physically distinct networks such as the electricity net and the rail network and their interlinks include vulnerabilities.

Having the node properties in mind, we can place limits on channel capacity and latency for a channel that consists of two links in series. The channel capacity is the least of the capacities of the three components – the two links and the transceiver node. The latency is the sum of the three latencies.

D.2.4.3 More Complex Path Structures

Since we can reduce any parallel pair to an equivalent single link, it follows that any number of parallel links can be successively reduced to a single virtual link. The same applies to series connections of more than two links. Accordingly, any complex pathway between two nodes can be reduced in a series of stages, first reducing all series paths that consist only of simple links, then reducing all parallel paths that consist of simple or virtual links, then again reducing series paths, and so on until what remains is one virtual link. This virtual link has a capacity and a latency, which we can denote by script letters C and L .

D.2.5 The Concept of Informational “Distance”

In the usual analysis of static networks, links have “strength” or “weight”, a concept that can be instantiated in a variety of ways, a few of which are discussed in the Framework Annex of this report. Seldom, except in a network with a geographic embedding field, does the concept of “distance” become associated with a link. Quite often, the length of a path between two nodes is considered to be the number of links constituting the path. In a traffic-carrying dynamic network, however, distance-like concepts are important.

In a geographically embedded network, distance is usually important only because it related to travel time for traffic. In a dynamic network, time is the important aspect of distance because, as discussed above, it limits the mutual information between nodes.

What measure of “distance” is suited to the dynamic network? Two obvious measures can plausibly take the place of “distance” in this calculation – latency and channel capacity. The longer the time delay, between an event at node A and its first effect at node B, the greater the distance. Alternatively, the lower the channel capacity of the virtual link between A and B, the greater the distance. The inverse of channel capacity is the time it takes to transfer a specified number of bits, so it makes sense to treat the inverse of the channel capacity as the capacity distance.

There is a third, and often more important, plausible measure that could take the place of “distance”. This measure is based on the uncertainty generation rate of the remote node, and L the latency distance of the channel connecting the observed to the observer node. From any node, another node is at a greater distance the greater the value of U. The less a node can know about another node, the further apart are the pair.

Put another way, channel latency puts lower bounds on the uncertainty one node can have about the current state of another node for a given uncertainty gain rate at the other node. Before making the observation, the observer has some estimate of the possibilities for the observation. The probability distribution across those possibilities defines the maximum uncertainty, Umax, the observer can have about the observed node. If the observation cannot reduce the observer’s uncertainty below Umax, the observation is valueless, as it provides no information about current state of the distant node. The expression Umax > G x L defines the region of the network over which a node can have some information about the current state of another node. If Umax < G x L, the distant node is beyond a “current state horizon”.

Events occurring at nodes beyond the current state horizon can eventually influence the observing node, and the observing node can gain information about past states of the observed node. No matter how long the latency, the effects of a specific event will over time reach the observer. Indeed, we could define yet more kinds of distance measure, based on the precision with which an observer can know about a distant event, precision that is influenced by both latency and channel capacity of the path between the event and the observer. We will not address such measures at this point.

We have defined three quite distinct possible measures of distance, each useful for different purposes. We can label them respectively “latency distance”, “capacity distance” and “uncertainty distance”.

These measures are not symmetric. The distance A®B is usually different from the distance B®A. This point leads to another important distance concept, the round-trip distance, and particularly the round-trip uncertainty distance. For now, we simply define the round-trip distance, leaving the discussion of its importance until we discuss the dynamics associated with cycles in a network.

To determine a round-trip distance between two nodes A and B (using any of the three measures), create the virtual links A®B and B®A (which may be real links). Create a single virtual link by treating these as though they were two links in series, and compute its distance measure. The round-trip distance measure is symmetric with respect to the two nodes, because whichever node is taken to be the starting point, the round-trip virtual link consists of A®B and B®A in series. Sometimes we call the round-trip distance the circumference of a loop that includes A and B.

We now have six distinct measures of distance, three directed (latency, channel capacity, and uncertainty) and three corresponding round-trip distances. Each of these may be useful for different purposes, though the round-trip measures are more closely tied to the network dynamics.

D.2.6 Network Structural Properties

D.2.6.1 Diameter

So far, we have considered only single real or virtual links and the nodes that they connect. One link does not a network make, but a few network properties do depend on the properties of single virtual links. One such is network diameter.

As with any structure that has more than one dimension, two significant diameters can be defined. There may be more, but there are always at least a minimum and a maximum diameter. In a network the minimum diameter can be specified by finding the distance from each individual node to each other. For each node there is a largest of these distances. The least of these largest distances is the minimum diameter of the network. The maximum diameter is the longest of these distances.

The same algorithm generates the maximum and minimum diameter of the network no matter which distance measure is used.

D.2.6.2 Feedback Diameter

"Feedback" refers to the influence of events at one location on its properties at a later time. Feedback can stabilize systems or can cause explosive dynamic changes, both of which can have dramatic influences on the behaviour of the surrounding network, so it is necessary to consider the conditions under which feedback can be important.

Between any two nodes we can compute round-trip distances using any of the three measures. The latency round-trip distance is a measure of the earliest moment that the effects of an event can influence the later properties of the node, the capacity round-trip distance is a measure of how abruptly the returning influence can affect the properties of the node, and most importantly, the uncertainty distance determines the extent to which the node's properties can be stabilized or amplified by feedback.

As noted above, the uncertainty distance $U = G \times L$. In the round-trip calculation, the information generation rates of two nodes are involved, the initiating node and the distant node. In certain cases, such as the junction between two wires, we can treat the distant node as though its information generation consisted entirely of what arrives on its in-links, all of which is reflected in its out-links, and as though it had infinite capacity and zero latency. In that case, the uncertainty distance is determined by the information generating rate of the initiating node and the round trip latency distance. If the distant node has other information-generation sources, or has its own latency and limited capacity, the round-trip distance will be increased in a way that can be computed as described above.

Every node has a maximum uncertainty value Umax determined by the possible states it could reach. If the round-trip uncertainty distance U to another node exceeds Umax, there is no possibility of feedback effects between these two nodes. Hence, for every node in the network, there is a sub-set of the network that includes only those other nodes for which the round-trip uncertainty distance allows feedback effects. In terms of the other five measures, this sub-net has a maximum and a minimum diameter.

Within the limiting feedback radius, there may be many distant nodes that could participate in feedback to the initiating node. All of these will influence the details of the feedback dynamics, but those with lower uncertainty distance will potentially have greater influence. For any value of U, some sub-net can be defined that contains only those nodes whose round-trip distances are less than U. Those nodes are likely to have the greatest effects on the dynamics of the network near the initiating node.

Feedback diameter occupies a place between being a property of the network and a property of a node. It is a variable that is associated with each individual node, but it depends on, and is associated with, the entire network.

D.2.6.3 Between-ness

In the analysis of static networks, many different measures have been given the name of "between-ness". The same will be true for information-based analysis of dynamic networks.

The two diagrams of Figure D-13 are topologically the same. The difference is that in the lower panel the paths L and M are shown as long and convoluted, whereas in the upper panel they are short and straight. This makes a difference to the eye, but does it make a difference to the analysis? In order to answer this question, we must ask what it means for a node to be between two others.

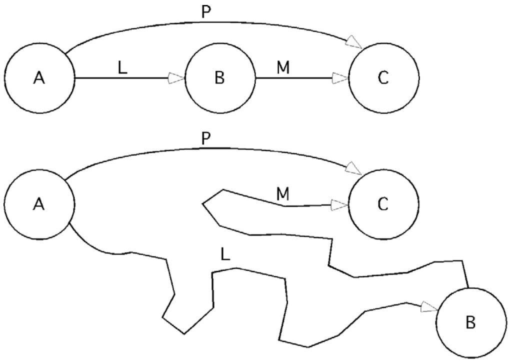


Figure D-13: To what extent are B between A and C?

There are several natural answers to what it means to be between. One of them is that a path exists between A and C that incorporates B. Another is that the most effective path between A and C pass through B. To give the notion of "between" a measure, one could determine all the paths between A and C that do not pass through any node twice, and find out what proportion of these go through B; or one could do the same, but weight the paths inversely according to their length. Probably many other measures could be developed, but these two offer a flavour of the possibilities.

The situation is in a way much simpler in an information-theoretic analysis. We can start by simplifying the network as described above, simplifying all paths from A to C that do not pass through B into the single virtual link "P" in Figure D-13, and the paths that do pass through B into the virtual links "L" and "M". Having done that, we take the appropriate distance measure, whether it is directed or round-trip, and whether it be latency, capacity, or uncertainty, and compute the distance from A to C through B and not going through B. The ratio of those two distances, appropriately scaled, is the degree to which B is between A and C. The scaling function will depend on the use to which the number is to be put.

The suggested procedure provides a measure of how much node B is between nodes A and C. But ordinarily, when the "between-ness" of a node is calculated, no specific pair of other nodes is of interest. The question is often about how likely it is that a particular node will be between two randomly chosen other nodes in the network. We can ask a similar question within the information-theoretic framework, by determining for every pair of nodes the distances for paths that go through the node in question and for paths that do not. Summing all the distances in each class and taking the ratio gives an information-based between-ness measure for that node. It is an index of the power of the node to influence what happens in other parts of the network.

Any measure that scales as N^2 becomes computationally impractical very quickly. However, the concept of the uncertainty event horizon can reduce the magnitude of the problem, since a node cannot usefully be between two nodes of which one is outside the event horizon of the other. Furthermore, the insertion of a node between two others is less important if they have a large uncertainty distance between them than if they are close in uncertainty distance. Between-ness should therefore be scaled according to the uncertainty distance of the path under consideration. For most practical purposes, the calculation of between-ness could be limited to a computationally feasible region of the network surrounding the node whose between-ness value is being calculated.

In everyday language “Knowledge is Power”, and nodes with high between-ness both gather knowledge and dispense it more freely than do those with low informational between-ness. If the measure is latency, a high between-ness node can obtain and dispense information in advance of other ways the receiving nodes can get it. If the measure is capacity, a high between-ness node can dispense more accurate information than the receiving nodes could otherwise get, and if the measure is uncertainty, it can enable nodes on either side to interact more successfully than they could do by ignoring it.

D.2.7 Intermediary Effects on Channel Capacity

Consider the channel capacity of the link by which node C observes node A in Figure D-14.

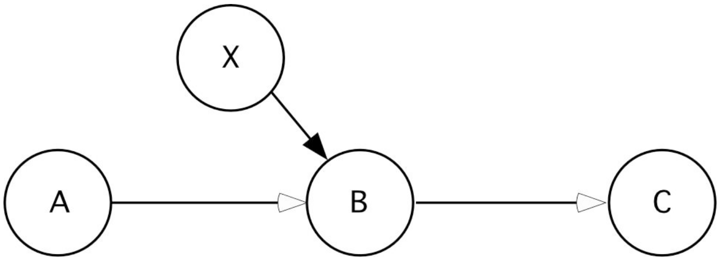


Figure D-14: Node B is Between Nodes A and C, but also Observes Node X.

Assuming that if X transmits nothing, B accurately passes to C everything it observes of A. In that case the channel capacity A®C is that of whichever link A®B or B®C has the lower capacity. But if X has a certain information loss rate R(X) and B observes X so as to maintain its level of uncertainty about X that information could interact with the information B gets about A. In the simplest case, B observing X is equivalent to adding noise at a rate R(X) into the channel A®C. The channel capacity of the link A®C is reduced by exactly that amount.

In practice, with complex nodes, information from X may not interact at all with the information B gets from A and passes on to C. The information from X might not be about the same thing as is the information C acquires from A through B. B might have enough processing capability to treat the sources A and X quite independently, being a simple sink for information from X and a simple transceiver for the A®B connection. However, if C is acquiring information about something from A, and X is providing information about the same thing, then X fully interacts with the A®C transmission. This interaction could have any effect from that of noise, reducing the channel capacity through B by R(X), to that of a parallel channel, adding R(X) to the capacity of the virtual link between C and the thing both A and X are observing.

D.2.8 Centrality

“Centrality” could mean a variety of different things. To be “most central” might mean:

- To be furthest from the perimeter of an object. Networks have no perimeter unless the connections to the world outside the network are taken to represent the perimeter. These connections are represented as source and sink nodes. One corresponding property for a node in a network might mean to determine the shortest path to the node from all source nodes; the longer this shortest path, the more central the node. This is a “receptive centrality”. A corresponding “effective centrality” exchanges. “To” and “from” in the algorithm. It would also be possible to define a “general centrality” as the product of the receptive and effective centralities.
- To be the most informative supplier of information to other nodes. This concept is almost the opposite of the “furthest from the perimeter” concept. To be most informative means to minimize the uncertainty of remote nodes about the supplier node – not necessarily about the current state of the supplier node, as a time capsule provides information about the state of a place decades or centuries earlier.
- To acquire the most information from other nodes, though not necessarily about their current states.
- To transmit the most information between nodes (this is effectively equivalent to one version of “between-ness”).
- To be in the best position to interact with other nodes.

Other similar concepts might be adduced, but these probably are sufficient to suggest the range of meanings that might be attached to the word “centrality”. For a dynamic network, they require definitions in information-theory-based terms. “To be furthest from the perimeter” implies that one of the six distance measures is used to assess the path length. Latency distance presents no problems, as it can increase without limit, in an additive manner. Capacity distance is interesting, since the channel capacity from one node to another is that of a virtual link between the nodes. If there is only one path between the nodes, the channel capacity is the least of the channel capacities of the links and nodes that form the path, and if that least capacity is on a nearby link, all nodes beyond that bottleneck are at the same capacity distance. Uncertainty distance increases with latency, for a given uncertainty generation rate of the observed node. All three distance types, whether unidirectional or round-trip, are applicable to “distance from the perimeter” centrality.

“To be the most informative supplier” has different implications depending on the distance measure used. None of the round-trip measures seem to apply, since the node is considered to supply information, not conduct interchanges about it. However, the three directed distance measures all have potentially useful implications. Minimum latency implies getting the information out fastest, maximum channel capacity (minimum capacity distance) allows the information to be precise most quickly once the latency period has passed, and minimum uncertainty distance gets out the most information about the current state of the sending node. In each case, a variety of algorithms are possible. Some examples might be median distance to all nodes in the network, number of nodes within a specified distance, maximum distance to the furthest node. The appropriate algorithm depends on the use to which the result will be put. The same comments apply to a measure based on acquiring information from other nodes, and need not be repeated.

“To transmit the most information between nodes” implies that both acquisition and distribution are good, suggesting that a centrality measure based on transmission must be a function of those two centrality measures. Since both acquisition and distribution centrality measures can be developed by a variety of algorithms, a transmission centrality measure also must have a variety of possible constructions. Again, it would depend on the use for the measure which construction algorithm would be most meaningful.

"To be in the best position to interact with other nodes" implies that the underlying distance measure is a round-trip measure, unlike the other centrality types mentioned above. Although there may be occasions on which the latency or capacity distances are appropriate, it will usually be the uncertainty distance that determines how effectively two nodes can interact. High interaction centrality must be a function of the number of nodes with which the test node can interact and the round-trip uncertainty distance for each of those interactions. A node is more central if it can interact with many nodes, even if those interactions are weak, having large uncertainty distances, and it is also more central if it has strong interactions with a few other nodes.

D.2.9 Common Source Information

Imagine two entities, A and B observed by an analyst as suggested in Figure D-2 and Figure D-3. The analyst could obtain a certain amount of information by observing A. This is the prior uncertainty of A, U(A). Similarly, by observing B the analyst could obtain a certain amount of information, U(B). The question is how much information, U(A,B), could the analyst obtain by observing both A and B.

To put this question in its simplest form, suppose A and B both report measurements of a person's height to X. A reports 180.2 cm, which from prior experience with A X treats as 180.2 ± .5, and B also reports 180.2 cm. What should X make of this? Is it by chance that these two reports are the same, when measures of a person's height rarely agree with such precision, or did B hear from A what A measured, and is simply passing on the A's report? If B's report is of an independently made measurement, then X should take the correct height as being 180.2 ± .3, a more precise estimate than could be obtained from either A or B alone. But if X knows that B is helpfully passing on A's report, he should ignore what B says.

How can this be treated in information-theoretic terms? The key is in the definitions of uncertainty and information. Before A makes a report, X has a certain distribution of probabilities for the height of the person, and hence for what A is likely to report. That distribution is X's expectation, and using Shannon's metric it implies an uncertainty of $U = -\Sigma p(i) \log p(i)$ where each "i" represents one of the things A might report. Afterwards, X has a new distribution, usually more sharply peaked. If A might have reported a wide range of values, say from 150 to 190 cm, a report of 180.2 would have a prior probability on the order of 1/300 or 1/400 of fairly representing the height of the measured person, and if X assumed a range of ±0.5 cm on the measurement, the value of 0.2 would have a probability around 1/10. The report would have provided approximately 5 bits of information (5 bits would represent a factor of 32 increase in precision).

Now B makes a report of 180.2. If X believes B is passing on a measurement made by A, X has a prior probability near 1.0 that B will say 180.2, so X's probability distribution for the height of the person measured is not changed by B's report. If, however, X believes B to have made an independent measurement, 180.2 might have a prior probability (after A's report) on the order of 0.5, with values nearby having decreasing prior probabilities. The report of 180.2 would provide about 1/2 bit of information (reducing the standard deviation by a factor of about $\sqrt{2}$). Clearly, although the report of either A or B alone could give X 5 bits of information about the height of the person, the report of both does not provide 10 bits. In fact, if X knows B got the measure from A, hearing both reports provides the same 5 bits as does hearing either one alone. The question at issue is X's understanding of the mutual information between A and B. If X does not observe that A and B communicate and does not know they are measuring the same person, then X's uncertainty U(A,B) about their two reports will be U(A)+U(B). If X knows they are measuring the same person but does not know they communicate, then X's uncertainty about the joins report (A, B) is U(A)+U(B|X knows A), which from X's point of view is U(A)+U(B)-U(A:B), and if X does know about the communication between A and B, X, ignoring miscommunication between A and B, will assume that U(A:B) is U(A), reducing U(A,B) by the same amount. Since U(A,B) is the most information X could get by observing A and B, the mutual information between them is critical in X's refinement of the precision of his probability distribution about the thing he learns from them.

D.2.10 Hypernodes

As discussed in the Framework Annex of this report, hypernodes are a way of grouping sets of nodes that are similar in ways of interest to a user [2]. A parameter determines the requisite degree of similarity, which can be varied from so strict that each node is its own hypernode to so lax that the entire network constitutes one hypernode. In Bjørke, Nilson and Varga, the index of similarity was based on local connectivity of each node. Those nodes with most similarity in their linkage would be grouped into first-level hypernodes. Second level hypernodes would be created in the same way, treating the first-level hypernodes as nodes, and similarly for third and further levels of node structure.

In an information-theoretic view, one of the criteria for judging similarity is the distance between nodes, whether this distance be latency distance, capacity distance, or uncertainty distance. All of these might be useful for one purpose or another.

The index of similarity determines just what nodes are combined into what hypernode, but whatever the index of similarity, nodes that belong to the same hypernode at some level are more likely to obtain their information from the same sources than are nodes that belong to different hypernodes at that level. The implication is that if a node X observes some entity by way of getting information from nodes A and B, X is more justified in treating the information as being independently derived if A and B are in different hypernodes than if they are in the same hypernode, whatever the level of hypernode analysis. If A and B are in the same first-level hypernode, and offer the same information about the entity of interest, X would be justified in ignoring one or the other, rather than treating them as mutually confirmatory.

As mentioned frequently above, information is always *about* something. The fact that both node B and node C observe node A would tend to place B and C into the same hypernode. But if what B observes about A is different from what C observes, these links are informationally independent, and have no bearing on whether B and C belong in the same hypernode. On the other hand, we can treat all the cases in which information is about topic 1 in creating one hypernode structure, those in which it is about topic 2 as creating another, and so forth. Accordingly, the hypernode structure is informationally layered, as suggested in Figure D-15.

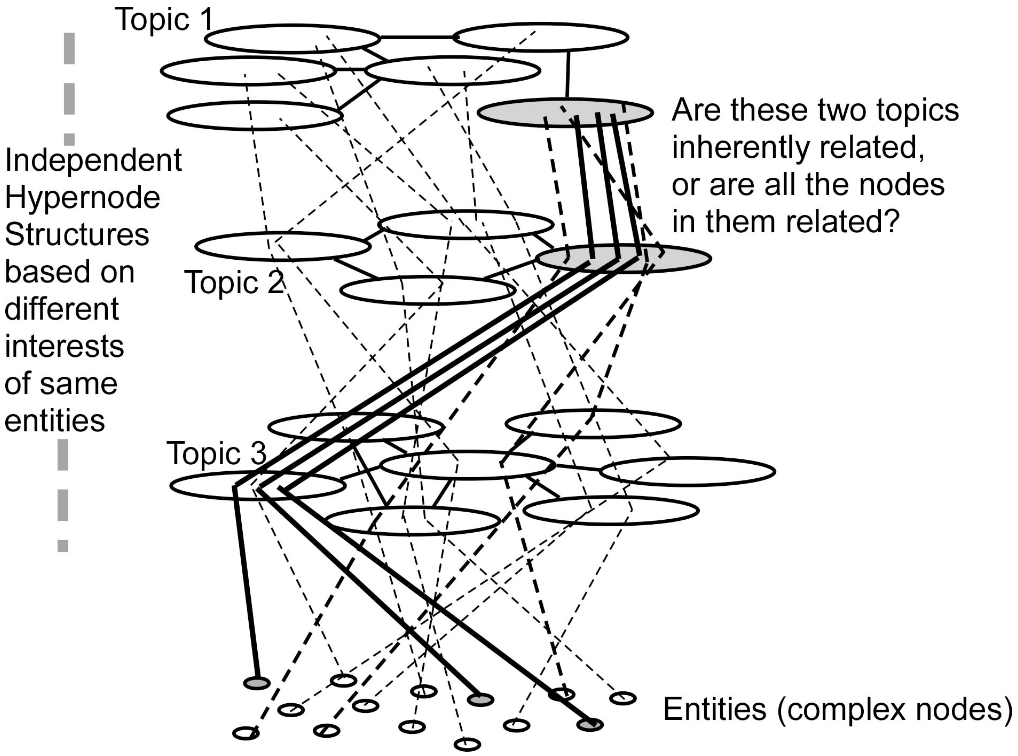


Figure D-15: Different hypernode structures may be developed from different similarities of interest in different topics, just as they can be developed from other different properties of nodes. If a group of different nodes coalesce together in several of the independent hypernode structures, it may indicate that some other factor relates those nodes.

In Figure D-15, there is a different hypernode structure for each of three possible topics of interest that define the observations of the different nodes. The structures for two topics each contain a hypernode that incorporates essentially the same set of entities. Three of those entities participate in the same hypernode for all the topics. It is reasonable to suppose that these three entities have some characteristic in common that is not represented, and that the mutual information between their structures is a good proportion of their total structural uncertainty. One casual way of saying this is that the three entities probably belong to the same club and that information on other topics is likely to be correlated across them, not independently obtained.

We pursue the hypernode concept further when we consider cycles, because it is quite likely that entities that belong to the same hypernode for different topics will also communicate with each other to form a well-connected sub-net in which cycles are common.

D.2.11 System Dynamics of Traffic Carrying Networks

Traffic-carrying networks without cycles have no traffic dynamics, though they may have structural dynamics, which we consider later. Traffic simply appears at source nodes and disappears at sink nodes. There may be issues of channel capacity, if too much traffic attempts to use some link, and such limits might cause backups, where information does not get through in timely fashion, but any dynamical issues are local, attributable to specific nodes or links. Considerations of system dynamics must therefore be a discussion of cycles and feedback processes.

D.2.12 Cyclicity

A cycle is a path that leads from a node back to the same node. How many of the outlinks from a node participate in a cycle that includes the node? How many do not? One possible measure of cyclicity for the node is the proportion of outlinks (or of inlinks or of both) that do. The cyclicity of the network might then be the average of these values. Another possible measure might be the number of distinct cycles in which a node participates, averaged over the network. Or again, for the network as a whole, the cyclicity might be represented by the proportion of nodes that participate in at least one cycle, or the number of links that do, or the number of cycles in which a link participates averaged over the network, or, or… The possibilities are numerous.

All the above suggestions apply to the structure of a static net. Here we are interested in dynamic nets, and therefore we are interested in information-theoretic measures.

D.2.13 Cycle Length and Feedback

A non-infinite round-trip distance from one node to another implies that the two nodes form part of at least one cycle, and every pair of nodes that participate in a cycle will have a non-zero round-trip distance between them. Indeed, if they participate ion no other cycles, every pair will have the same round-trip distance between them, that being the distance right around the cycle. Of course, if two nodes participate in more than one cycle in common, their round-trip distance will be that of the shorter of the two cycles in which they both participate, as suggested in Figure D-16a.

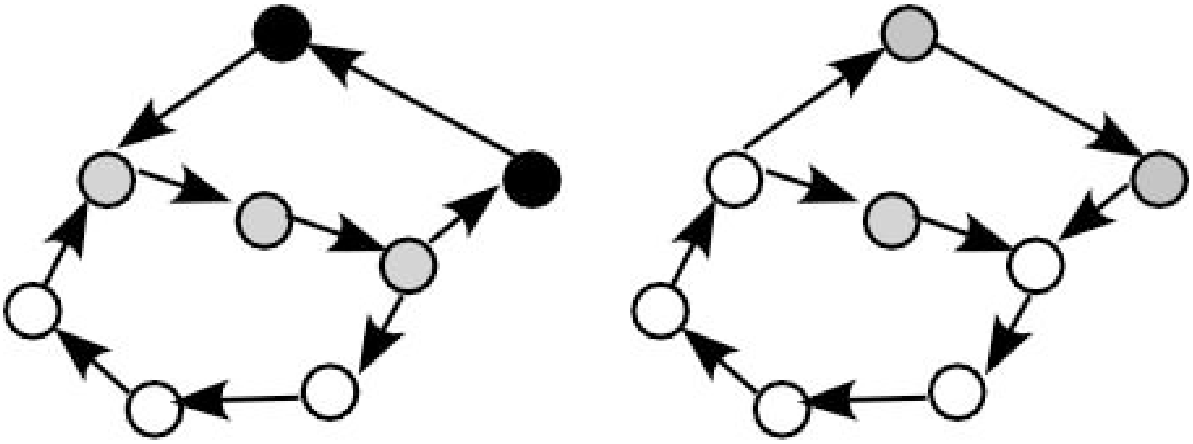


Figure D-16: (a, Left) The white and black nodes belong to one cycle each, the gray nodes to two; (b, Right) There is only one cycle, but the grey nodes may have uncertainty distances around the cycle different from the white nodes and from each other in the two paths.

Figure D-16b illustrates that if part of the cycle consists of parallel paths, the channel capacities of the parallel segments may differ, which may mean that the round-trip distances from the grey nodes differ from those of the white nodes. The cycle distance is the least of these possibly different distances.

A cycle has the property that information sent out by a node may be received later by the same node, in other words, it provides feedback. However, feedback will be ineffective if the uncertainty distance C_U around the cycle is too great. If the state of a node when the information returns is unrelated to its state when the information was originally transmitted, there can be no useful interaction between the past and present states.

The limit depends not only on the latency distance around the cycle C_L, but also on the information loss (uncertainty gain) rate of all the nodes in the cycle. Although one might say that there is feedback if the uncertainty cycle distance exceeds the limit, the feedback has no more dynamic significance than does insertion of data from an arbitrary source. In a dynamical sense, feedback exists only if C_U is finite.

Imagine a cycle of three nodes (a “3-cycle”) such as that in Figure D-17, in which B observes A, C observes B, and A observes C but the reverse is not true in any of the three cases. Additionally, each of the nodes is influenced by nodes in other parts of the network. How do we determine the uncertainty distance around the cycle?

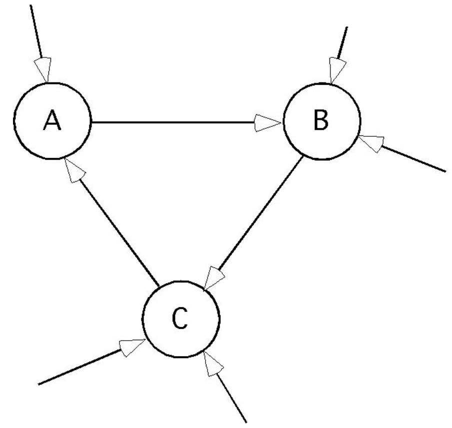


Figure D-17: A 3-Cycle with External Sources.

Remember that “information loss rate” refers to the mutual information between an observed entity (such as A) and its observer (B). The loss could be caused by autonomous activity in the observed entity, or it could be caused by influences from outside the entity affecting the state that had been observed. In a network, that means influences from the observed node’s in-links. Whatever the source of the information loss, the information that returns to the originator around a cycle may well not be the same as was originally transmitted, as a well-known party game illustrates (one person whispers a message such as the old English example “Send reinforcements; we are going to advance” to another, who repeats it to the next, and so around the circle, until it returns as “Send three and four pence; we are going to a dance”).

In the party game, the originator knows that the message that will be received derives from the message that was originally sent. In most real-world cases, this knowledge is not available, but if it is, the effect of the returning information is quite different in information-theoretic terms than if it is not. To explain this, we must return to a consideration of mutual information.

The uncertainty distance for B observing A in Figure D-17 is the information loss rate of A times the latency of the A®B link. Ignoring the possibility that A generates autonomous changes in its state, its information loss rate is the sum of the information received from the external part of the network plus that received from C. Since that received from C is in part due to observing B, which observes A, we have to consider the information loss rates for B and C as well. Uncertainty generation in B and C is the rate of information received from observing the sources outside the cycle, each of which subtracts from the channel capacity of the connection through that node. The cycle channel capacity is therefore the basic link capacity of the weakest link minus the sum of the externally derived information that influences the cycle information.

Now we can return to consideration of the uncertainty distance around a cycle. We initially defined uncertainty distance between two nodes as G x L, where G is the information loss rate at the observed node. How does this translate when around the cycle there are many nodes, each with in-links from outside the cycle, and each in-link may contribute to the loss rate for the cycle? We must be careful here, because as shown earlier, nodes may transmit information from a variety of non-interacting channels. In every analysis of a cycle, it is important that the information that returns to a node is *about* the same thing as was the information transmitted by that node. If it is about something different, the cycle is not complete. Even if the cycle is complete to the eye of the analyst, if the originating node is aware that the returning information is a possible degraded version what was originally sent, the incoming information will not influence the node's probability distribution about the topic. Only if the originating node treats the incoming information as independent will it (improperly) influence its probability distribution. If you tell Joe a piece of gossip, and later you hear the same thing from Betty, you may believe it more strongly than you originally did, but if you know Betty got it from Joe it should not affect your belief.

This is not to say that all around the cycle the information is overtly about the same thing. It need not be, as the example of a node that integrates its input to produce its output illustrates. However, it is often convenient to assume that it is. What matters is that information about something else at a node does or does not contribute directly to the information loss rate seen by an observer of that node further around the cycle. Put simply, I am unlikely to change my opinion of today's weather because I am informed that today is Beethoven's birthday. Anyone who asks me about the weather will get the same answer both before and after I received the information about Beethoven. However, if someone looks out of the window and tells me they see that it is raining, I may change my previous opinion that the weather is fine. The information loss rate for someone observing my opinion about the weather is affected by the input from the person who looked out of the window, but not by the input from the person who told me about Beethoven.

One way of looking at the question of the uncertainty distance around the cycle is to note that the information loss rate at each node detracts from the channel capacity through the node, and that the channel capacity around the loop is the minimum channel capacity of the nodes and links that form the loop. If there is even one node or link for which the channel capacity is less than the greatest uncertainty generation rate of any node in the cycle, C_U becomes infinite and the cycle is informationally broken. Even though the static connections indicate that a cycle exists, dynamically it does not.

In many cases, the uncertainty generation rate of a node is determined by the node's observation through links that are not part of the cycle.

D.2.14 Contradiction, Control, and Islands of Stability

Before making a particular observation or receiving a message, the observer may have a distribution of probabilities for the different possible states of an entity. After making the observation, the probability distribution has changed, and therefore so has the observer's uncertainty about the entity's true state.

Although there is no mathematical difference between observations that concentrate the probability distribution near some particular state ("It is about 180") and observations that reduce the probability near a particular state ("It is not near 180"), it will be convenient for the following discussion to distinguish between the two cases. Both single out a specific state from among the set of possibilities, but while first tends to concentrate the probability distribution around the identified state, the second has a more diffuse effect, driving the probability distribution away from that state to the other possibilities, as suggested in Figure D-18.

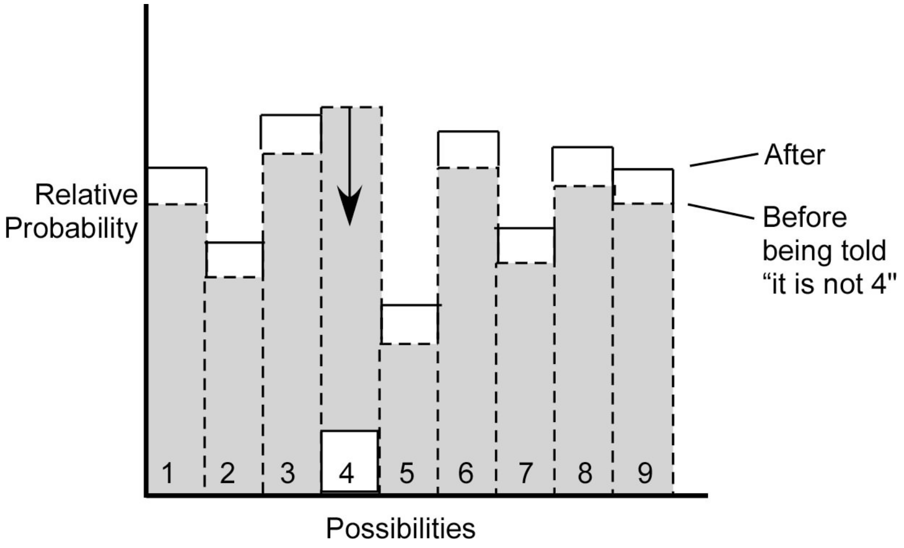


Figure D-18: Revision of probabilities after a denial that one of the possibilities is the correct one.

Figure D-18 shows a hypothetical distribution of relative subjective probability among nine possible states of something, say the next number to come up in a lottery. The observation is that someone tells the observer that the number is not “4”. If the observer has reason to believe that the speaker has inside information and always tells the truth, the observer’s subjective probability for “4” should go to zero. However, the observer may think it only likely that the speaker has inside information, and that is the situation shown in Figure D-18. The observer’s probability for “4” has been reduced only to the probability that the speaker does not have access to the facts.

Now suppose someone of equal credibility tells the observer that the number will be “4”. The observer has no reason to believe one of the informants over the other, and should revert the probability distribution to what it was before receiving either message, as shown in Figure D-19.

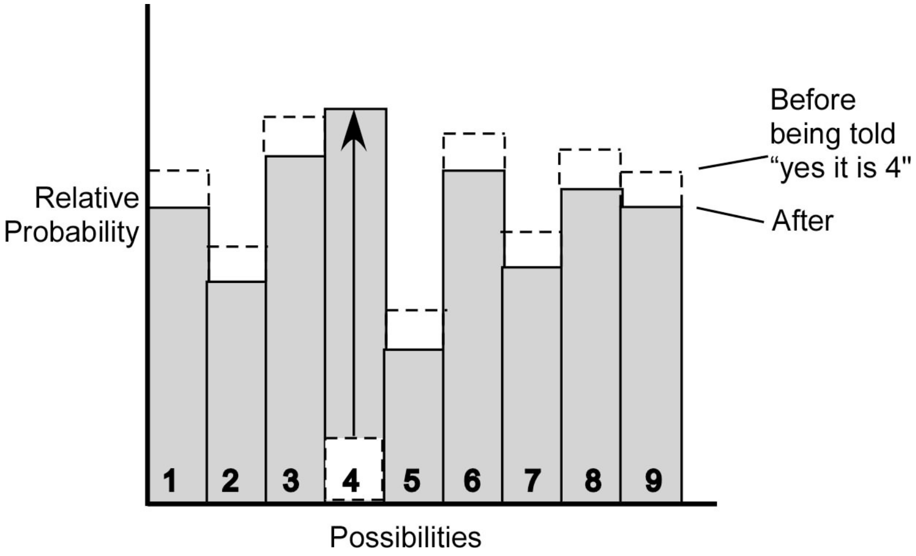


Figure D-19: Further Revision of Probabilities After a Contradiction.

At this point the first informant says: “I mis-spoke. I meant that the number will be 4.” Now the observer has a confirmation of what the second informant had said. What probability should the observer assign to “4”. Before getting either message, the probability was roughly 0.15. Each informant is assigned a probability of around 0.15 of not in fact knowing what the number will be. Assuming that their sources of information were independent, the probability that they are both wrong is 0.15 x 0.15, or .0225. The resulting probability distribution for the possible numbers now looks very different, as suggested in Figure D-20, which shows absolute probabilities rather than relative probabilities. The observer now believes rather strongly that the number is “4”.

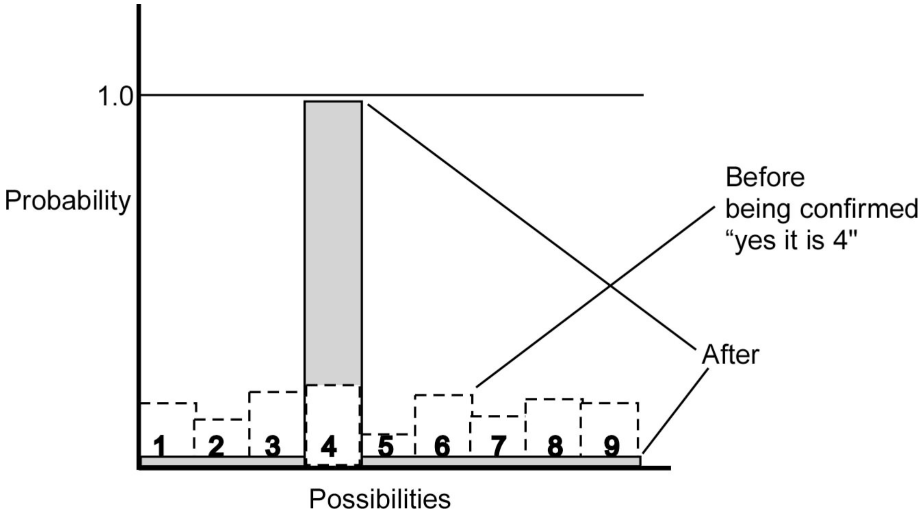


Figure D-20: Probability Distribution After Reversal of the Denial to Create a Confirmation.

Figure D-18, Figure D-19, and Figure D-20 illustrate one important generalization: denial has a less precise effect on the probability distribution, and hence uncertainty, of an uncommitted observer than does confirmation. For simplicity, let us suppose that there are 8 numbers rather than 9, all assigned equal probability. The associated uncertainty is 3 bits. If the observer gets a completely credible denial that the number is 4, seven numbers are left, each with a probability 1/7. The associated uncertainty is 2.8 bits. The observer received only 0.2 bits of information. However, if the observer gets a completely credible assertion that the number is 4, the uncertainty goes to zero, and the observer has received 3 bits of information.

On the other hand, denial can have a strong effect if the observer originally has a weak belief in the thing denied. Suppose the observer originally assigned a probability 0.875 that the number was 4, and then received a denial that it was 4 to which the observer had assigned a probability 7/8 of being based on true data. After receiving the denial, the observer’s uncertainty about the correct number is 3 bits, whereas beforehand it was 0.89 bits. The observer has received – 2.11 bits of information. (A negative value of information simply means that the observer is more uncertain after the observation than before it. The magnitude of the effect of the information is independent of sign).

Even though a previously uncommitted receiver of a denial message receives less information than the uncommitted receiver of an equivalent assertion, the transmitter’s selection process is no different. In either case, the transmitter must select “4” from the range of possibilities, and attach the equivalent of “Yes” or “No” to the message. This difference between transmitter and receiver provides yet another indication that the information in a message depends on the recipient, not on the transmitter or on the form of the message as it is transmitted.

D.2.15 Cliques and the Madness of Crowds

"Extraordinary Popular Delusions and the Madness of Crowds" is the title of a book by Charles Mackay [9]. In it, he describes several cases in which large numbers of people, even whole populations, have come to believe something that in retrospect was quite ridiculous, and in which bad things have happened as a consequence of the mass delusion. We argue that this kind of thing is a consequence of undetected cycles in networks. However, before dealing with the Madness of Crowds, we must consider the formation of cliques.

In everyday life, a clique is a group of people who tend to act and associate together more than they do with people outside the clique. In the dynamic network, a similar definition applies. Members of a clique are more likely to associate in cycles that are largely contained within the clique than to belong to cycles that largely involve nodes outside the clique.

For a cycle to exist, for every node in the cycle the information returned to it must be *about* the same thing as the information it transmitted. The form of the information does not have to be the same, but it must be about the same thing. As noted earlier, it is quite possible for two streams of information to be superficially quite different (as are a waveform and its integral) and yet convey exactly the same information. The question arises as to what the originating node does with the returned information.

All information is about something, which we call "the observed entity", whether the observation is direct or by way of intervening nodes. There are three main possibilities:

- The information is treated as an indirect observation of the originating node around a cycle: "I told your friend that, and I guess you got it from her".
- The information is treated as being about the same thing but as being obtained from an independent source, such as a direct observation of the pragmatic embedding field, or from another node uninfluenced by the originating node, and as such affects the originating node's probability distribution over the possible states of the observed entity: "I thought it was 25, but you think it is 22, which probably means it is somewhere in the low 20 s".
- The information is treated as being from an independent source about the same entity, but as being about a differently timed observation of the entity. As such, it is information about the time variation of the observed entity: "When I last heard, it was 25, but you tell me it is now 22, so I believe it is decreasing rather rapidly".

D.2.15.1 Case 1: The Originating Node Recognizes that the Information Completes a Cycle

In the first case, in which the originating node knows that the cycle exists, observing the return from the last node of the cycle should have little effect on the originating node's probability distributions for the state of the entity that is the topic of the information. There is, however, an exception to this, usually manifest in a 2-cycle.

A 2-cycle is a direct dialogue between two nodes, which we may call A and B. If B observes A and gets some information about something B already knows, B may thereby increase the precision of the probability distribution about that something, assuming that A's information is independently derived. Now B returns the same information to A, along with the information that B already knew. A now has reason to increase his precision, on the assumption that B got the information directly (Figure D-21a). Both now have increased precision, or strength of belief, in something that might well have been obtained from a common source (Figure D-21b). The source of their increased precision is the presumption that the two external sources are independent.

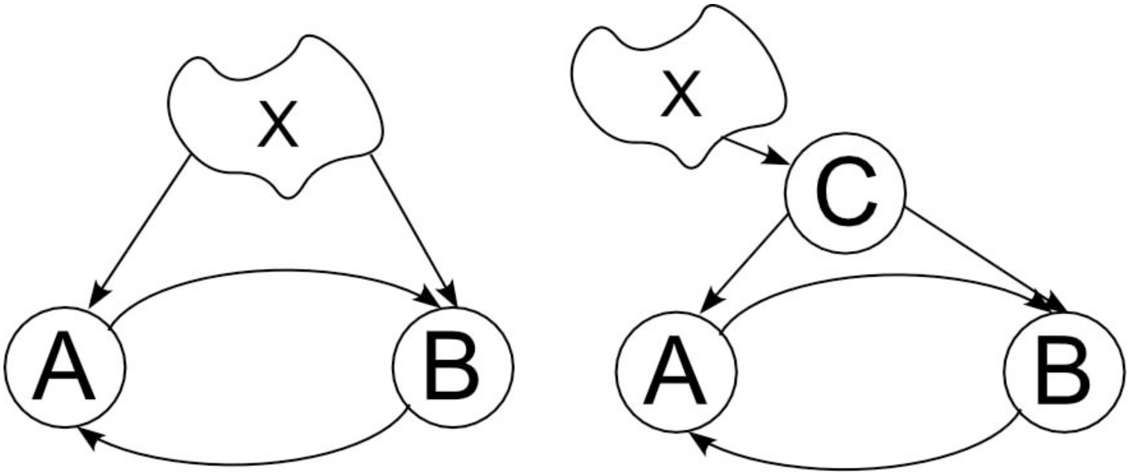


Figure D-21: (a, left) Two nodes forming a 2-cycle independently observe the same thing, X; (b, right) The two nodes believe each other has independent information about X, but both get it from the same source.

The same effect of increasing strength of belief can occur in cycles with more than two nodes, even if some of the nodes know that the cycle exists, since the return of the same information indicates that the other nodes in the cycle do not have inputs from elsewhere that significantly alter their probability distributions for the possible states of the entity. To put this less abstractly, if I tell my friend that the moon is made of green cheese, and later I hear the same from someone else who I know belongs to a group that also contains my friend, I may be reasonably sure that nobody in the links between my first and second friends has significant belief to the contrary, and that, in itself may allow me to increase my strength of belief in the correctness of what I said.

A 2-cycle has another common use, as an ask-answer protocol. A asks B about a state of B, which to B is an observation of A indicating that A does not know the value of that state. B responds with the value. Both observations are about the same thing (the state of B), the result tending toward a state in which A and B have similar probability distributions for the value of the state. A typical computer protocol has the form “Are you ready” “Yes, I am ready”, at which point both systems are prepared for some further action.

As with any cycle, the effects are stronger the shorter the uncertainty distance around the cycle, even if the originator knows that the cycle exists.

D.2.15.2 Case 2: Hidden Cycle, Information about the Same Thing Circulates

Now we consider cycles that are not known to the originating node. Figure D-22 shows a simple cycle in which the informational topic is some property or state of an entity we call “E”, which could be a node in the network or something in the pragmatic embedding field. Call the originating node that observes E “A”, the node that originally observes A “B”, and the last node in the cycle “Z”. Somewhere between B and Z are one or more intermediate nodes collectively labelled “X” in Figure D-22. A does not know whether there is a link between B and Z, and so, even if Z reports exactly what A reported to B, A may treat the observation of Z as independent information about E.

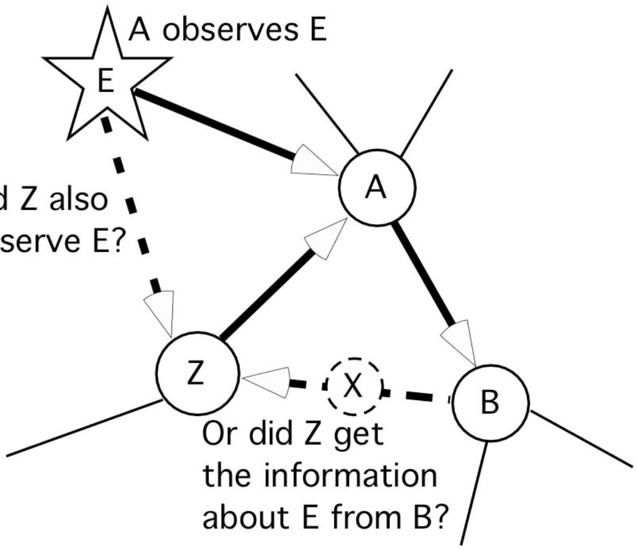


Figure D-22: A cycle exists that starts with A®B and ends with Z®A. A observes something (E) and also gets information about E by observing Z. Z might have made an independent observation of E or might have got it by directly or indirectly observing B.

Informationally, if Z made an independent observation of whatever A communicated to B and reports the same as what A reported to B, A should increase the strength or precision of belief about that thing. As discussed above, even if A knows the cycle exists, the fact that the same information is returned indicates that none of the nodes in the cycle have made independent observations that contradict A's own observation.

It is possible, however, that the information returned by Z directly contradicts the information originally observed by B. A observes E, another node in the network, obtains the information that the sky is clear, and reports to B "The sky is clear", but Z reports "the sky is heavily overcast". How should A treat this situation? A's uncertainty about the state of the sky is increased, and if A's history with E and Z is that they are equally reliable, A's uncertainty is as though A had observed neither E nor Z. We simply note this possibility here, and discuss it in a later section on contradiction.

D.2.15.3 Case 3: Hidden Cycle, Information about a Different Time of Observation

Assume that A continues to observe some state of E as it changes with a certain uncertainty generation rate R bits/sec. A's uncertainty about the present state of E depends on the link latency L, as discussed above; it is L x R. Now A receives information from Z about the same topic, information that differs from A's original observation. Unless the time of the original observation is available to A (in which case A would know whether the information from Z had been gathered independently), it is inherently impossible for A to know whether the information from Z should be used to refine A's estimate of the state of E or to reduce A's uncertainty about the rate of change of the observed property of E.

As A makes more observations of the interesting property of E, the existence of trends over time allows A to alter the estimate of the probability that the differences from one observation to the next are due to measurement uncertainty (noise) or to real changes in the measured value. It is impossible for A ever to be sure which is the case, but the probability of one or the other can change, and with that change comes a change in how A should apply the observations to the uncertainty of the value or of the rate of change of value.

D.2.16 Confirmation and Denial Around a Cycle

As noted above, if the observer starts from a position of high uncertainty, a precise confirmation has a much larger effect on uncertainty than does a precise denial. On the other hand, mutually contradictory information from two sources can counterbalance each other, and a precise denial of a previously held belief can increase uncertainty. These effects determine the influence of cycles in a network in two ways; observations by a node of sources outside the cycle may insert noise or contradiction, and a perverse node may simply deny what came from within the cycle. The term "perverse" here need not imply that the node or cycle is useless. Indeed, a perverse node is required if a cycle is to implement control.

D.2.16.1 Control and Stability

"Control" means keeping some property in some desired state regardless of influences that would otherwise disturb it. Informationally, if the controlled property is a state of a node, the controller must contradict the disturbing influences, providing counter-influences that sustain the desired state. Informationally, the details of the outputs of the positive link and the negative link to the controlled node should be identical. Any deviation from identity means that the disturbing influence is able to alter the state of the controlled node.

Figure D-23 shows the salient features of contradiction and control. In the left part of the figures, a source S is observed by node A, but node B supplies contradictory information. The state of A is undetermined. Also, a question arises as to how B knows what to contradict, without itself observing S. To observe S is impractical in almost all real-life cases, because S represents all the influences that could affect the interesting state of A.

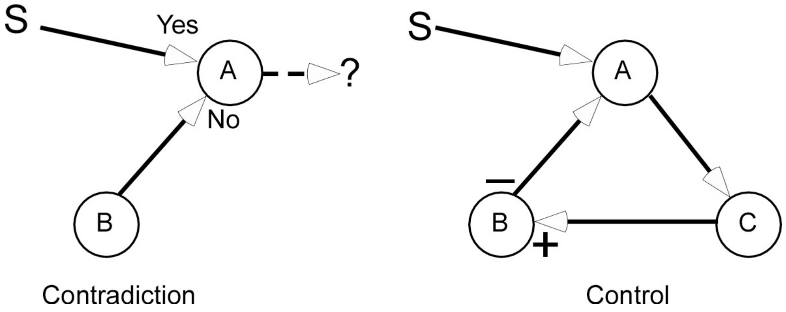


Figure D-23: Contradiction and Control. (Left) node A observes both S and node B, and both observations provide information about some topic X. The two observations oppose one another, leaving A with no information on the topic; (Right) Node C observes the state of A relevant to the topic and determines that it differs from what C wishes A to believe. C reports this difference to B, and B provides A with information to alter A's belief structure toward C's reference value. A's state is controlled to match the reference value of C. A thus appears as a stable point in the network.

The right part of Figure D-23 shows the way out of the problem. Node C cannot observe S, for the same reason that node B cannot. But node C can observe node A and see how the state of node A is changing from its desired value. Node C can report this deviation to node B, which can then influence node A to change state in the appropriate direction, countering the changes influenced by S. This is "control". Nodes C and B might, of course, be physically the same node, but what matters is the existence of a loop with an odd number of "perverse" nodes. If B and C actually is the same node, that node acts as a monitor trying to "keep A on the straight and narrow".

The critical measures for the ability of this cycle to control the state of one of its nodes are G(S), the uncertainty production rate of S, and the uncertainty distance around the cycle. If there is no uncertainty generation at nodes B and C and the links are noise-free, L x G(S) is the minimum residual uncertainty of the value of the state of A, where L is the latency from A to B through C.. This is a measure of the degree of control achievable by an ideal control system based on observation of the deviation of the state of A from its ideal condition.

To see the effect of control in the network dynamics, consider the situations represented in Figure D-24. In all four diagrams, node X is observing some state of node P. In Figure D-24a, the observation is direct. In Figure D-24b, it is by way of the mediating node A, which we will assume to simply observe the state of P that X is observing, and replicates it for X to observe. X's observation of P is unaffected by the mediation if A is immediately responsive to P and is noise-free. The situation is different if A is influenced by something else besides P. Figure D-24c shows the state of node A being influenced by node B as well as node P. The influence of B acts as a noise source in the channel through which X observes P. The effect is to reduce the bandwidth of the channel through A.

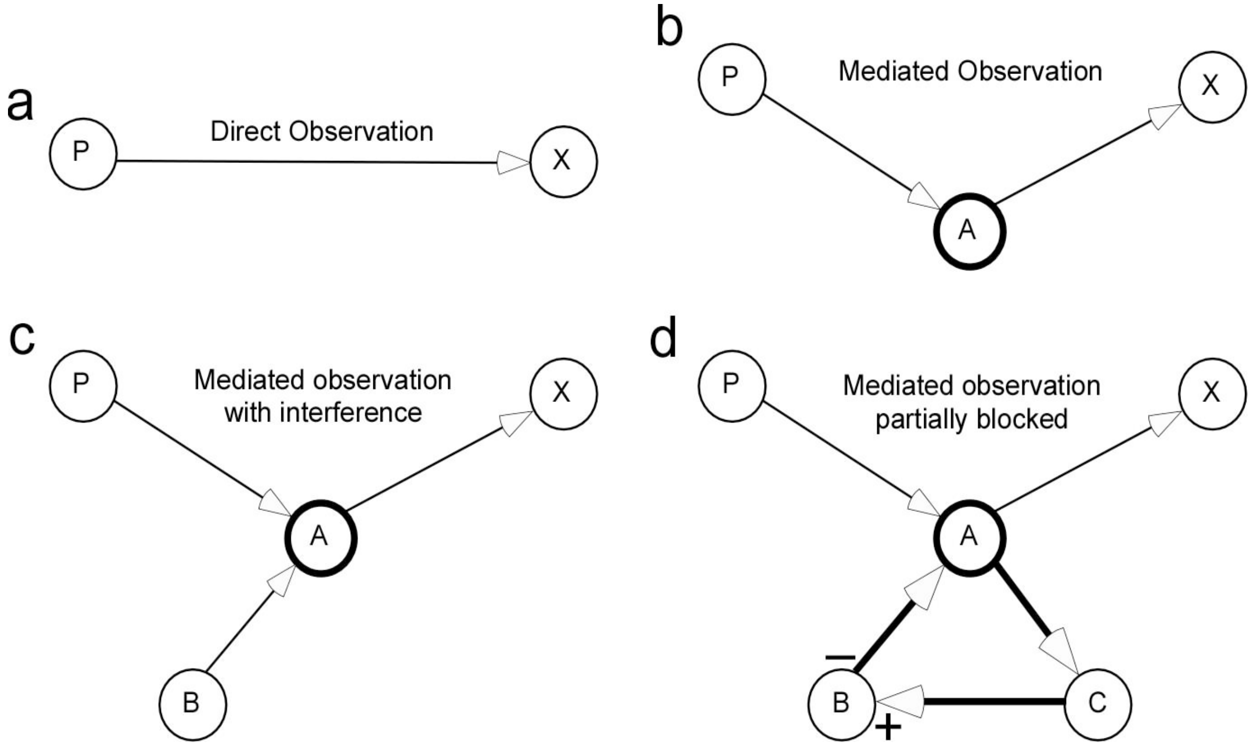


Figure D-24: The observation of P by X is partially blocked if the mediating node is controlled: (a) Direct observation; (b) Mediated observation; (c) Mediated observation with an interfering noise source; (d) Mediated observation partially blocked by control of the mediating node.

The situation is more complex if the interference of B is such as to directly contradict the influence of P on A, as shown in Figure D-24d. If the uncertainty distance around the ABC cycle (A observed B, B observes C, and C observes A) is near zero, the state of A will be effectively uninfluenced by the state of P, and X will no longer be able to observe P at all. In electronic circuitry, the effect is called a "clamp". If the voltage at some circuit element is clamped, any signals that would have been passed by that element are blocked.

Any real cycle has some non-zero uncertainty circumference. If such a cycle implements a control system controlling some state of a node, as in Figure D-24d, control cannot be perfect, and X is not blocked entirely from observing P. In particular, X can observe changes in A that are composed of changes in P mixed with changes in B that are related to earlier states of A.

D.3 NETWORK ANALYSIS PART 2: NETWORK STRUCTURE

When we talk about network structure we are considering links as well as nodes. In analyzing the traffic-carrying network, we were concerned mostly with what one node can learn by observing another through a link, not with observation of the links themselves, except inasmuch as the link character affected the ability of the node to make the observation. When we talk about network structure, the Observer is the Analyst, and even if we consider the variations in effects of traffic on a node caused by the node's observation of network structure, the node is taking on at least part of the role of an analyst. Hence, in what follows, when we talk about an observation, the observation is made by an Analyst who is not an ordinary part of the network, unless otherwise specified.

D.3.1 Traffic-Free Networks

Network structure is important for traffic-carrying networks in which nodes can get information by observation, but there is an important class of networks in which this concept does not apply, A family tree is an example of such a network, as are more abstract networks such as the organizational structure of a business, in which the nodes do not represent physical entities, but represent instead abstract concepts such as "Team Leader", and links represent relationships among the abstract concepts.

In a traffic-free network, all the information is that obtained by the Analyst, who we will call "Andy" from observations of the network. The information-theoretic analysis must therefore be concerned with Andy's possible uncertainties about aspects of the network structure. As discussed in the introductory section of this chapter, uncertainty is always *about* some aspect of the thing observed. Accordingly, we should consider what about network structure might reasonably be of interest to Andy. Many of these properties will be the same as or analogous to those appropriate to static networks, well studied under the rubric "Social Network Analysis". Others, however, apply only to dynamically varying networks, and those are the properties for which information-theoretic analyses are most likely to be useful.

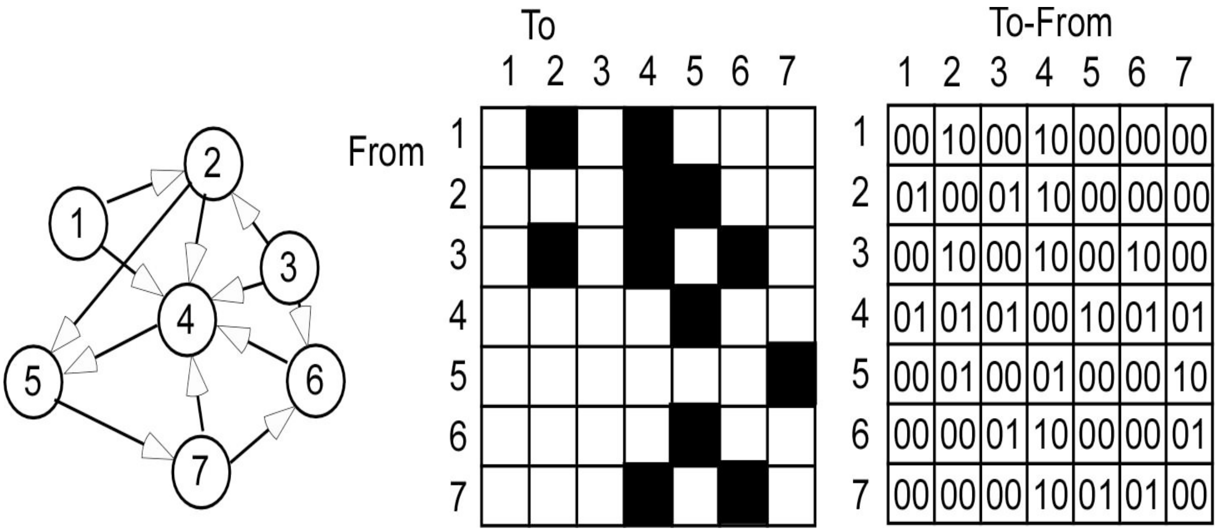
As noted above, the Shannon measure of uncertainty is $-\Sigma p(i)\log_2 p(i)$, where "i" indexes the possible macrostates of the entity in question. A macrostate is defined by the momentary interest of the observer. If a difference between two microstates matters to (and is observable by) the observer, they are in different macrostates. If it does not, they are in the same macrostate. This uncertainty measure can be applied to anything to which a probability can be assigned, but the same formula can also be applied when "p(.)" is any partition of data such that the partitions do not overlap and their sizes are normalized so that they sum to unity.

To create an example at random, it would be possible to define a measure of "distributional variation", where a distribution number for a node is the ratio of outgoing links to incoming links. All such ratios are rational fractions, and there are a finite number of them. The "i" in the formula would be the individual fractions, such as 5/2; if n(5/2) was the number of nodes that had five outlinks for every two inlinks, and N was the number of nodes in total, then the "p(5/2)" value would be n(5/2)/N. The formula $-\Sigma p(i) \log p(i)$ would indicate the analyst's uncertainty about the distribution ratio for a node before observing that node, and that uncertainty could be an index of "distributional variation". Whether such an index would be useful is another question. For different purposes, many distributional measures based on the mathematics of information might be defined.

D.3.2 Hypernode Structure and Uncertain Links

The method of analysing a static network structure into a nested set of levels of "hypernodes" has been described in Annex C and elsewhere [2], and the IST-059 final report], but it may be helpful to repeat a brief summary here, since the dynamic context is a little different.

The simplest kind of network description merely indicates for each pair of nodes whether there exists a directed link in each direction. The description might take the form of a diagram or of a matrix, as shown in Figure D-25.

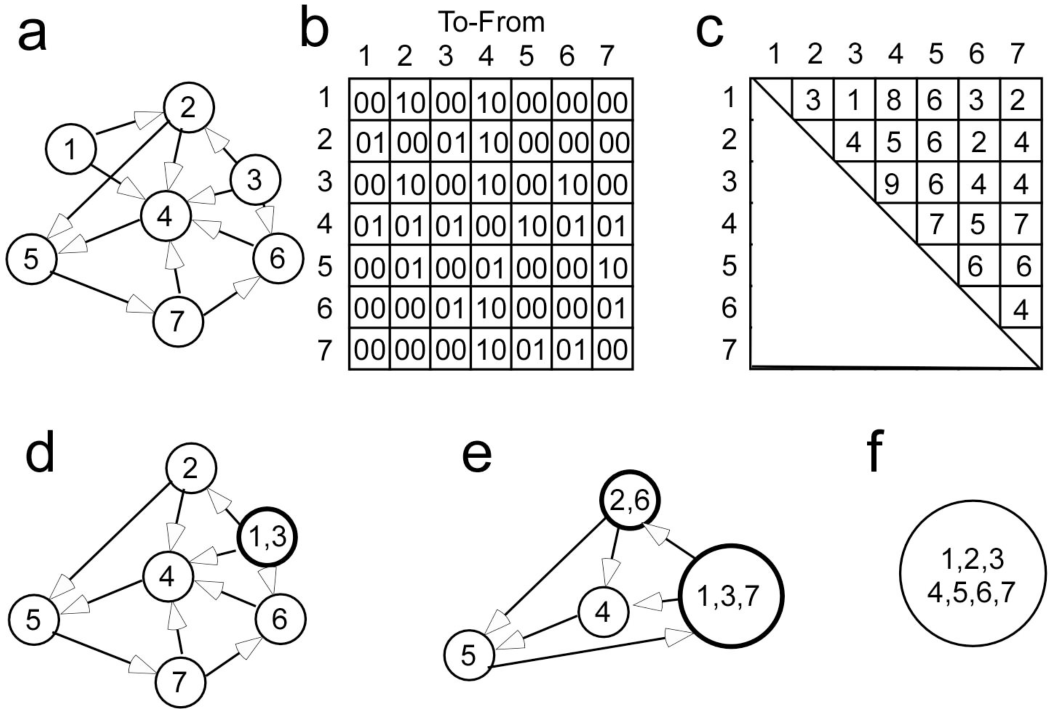


	To						
From	1	2	3	4	5	6	7
1		■		■			
2				■	■		
3		■		■		■	
4					■		
5							■
6					■		
7				■		■	

	To-From						
	1	2	3	4	5	6	7
1	00	10	00	10	00	00	00
2	01	00	01	10	00	00	00
3	00	10	00	10	00	10	00
4	01	01	01	00	10	01	01
5	00	01	00	01	00	00	10
6	00	00	01	10	00	00	01
7	00	00	00	10	01	01	00

Figure D-25: Three Ways of Representing a Simple Network. (Left) Diagrammatically; (Middle) As a matrix with blacked out cells indicating the presence of links from the cell listed along the left to the cell listed across the top; (Right) As a matrix in which each cell indicates both whether there is a link to and a link from the node listed on the left to or from the node listed across the top.

To simplify the structure using hypernodes, one combines nodes that are most similar. Similarity here implies maximum mutual information between the linkage patterns of the nodes. In this simplest case, it is possible to substitute for mutual information the number of cases in which the linkage patterns of the two nodes differ. For example, in the network of Figure D-25 and Figure D-26, nodes 1 and 3 differ in only one link, both having out-links going to nodes 2 and 4 and having no in-links. No other pair of nodes is so similar, and thus in the first simplification, nodes 1 and 3 are combined, as shown in Figure D-26d. Subsequent stages of simplification use less rigorous criteria for similarity (an intermediate stage is shown in Figure D-26e), until finally all the nodes are collected into a single hypernode, as shown in Figure D-26f.



b — To-From

	1	2	3	4	5	6	7
1	00	10	00	10	00	00	00
2	01	00	01	10	00	00	00
3	00	10	00	10	00	10	00
4	01	01	01	00	10	01	01
5	00	01	00	01	00	00	10
6	00	00	01	10	00	00	01
7	00	00	00	10	01	01	00

c

	1	2	3	4	5	6	7
1		3	1	8	6	3	2
2			4	5	6	2	4
3				9	6	4	4
4					7	5	7
5						6	6
6							4
7							

Figure D-26: Hypernode Simplification of the Network: (a and b) From Figure D-24; (c) The cells show in how many places the strings in (b) differ for each pair of nodes; (d, e, f) stages in the hypernode simplification of the network.

The structure of a network consists of nodes that may be connected pair wise by directed links. Nodes and links may have types, For example, if there are three nodes A, B, C, two of which are of type “person” and one of type “dog”, a link type “is father of” might exist between A and B but not A and C (C being the “dog”). If “is father of” links A to B, that type of link cannot connect B to A, but a link of type “is child of” must link B to A. If a link “owns” connects A to C, no link of type “owns” can go to C from any other node.

Hypernode structures can be developed on the basis of any observable property or set of properties. In particular, they can be based on commonality of temporal variation. If two nodes are fed from the same source, they are likely to change state synchronously, but if one influences the other there will be a distinct time lag between them. If the analyst can track state changes across time, it may be possible to derive hypernode structures that indicate common information sources, even if those sources are not visible in the part of the network known to the analyst. Elements that vary together, with or without time-delays, are reasonably treated as being coordinated “objects”, and when the hypernode structure is based on common variation, these “objects” are seen as hypernodes at the kinds of different structural level illustrated in Figure D-26.

If the nodes are complex, which they are if they represent persons, many independent hypernode structures can be built using the same nodes, based on different kinds of links among the nodes. Such independent structures can be used in conjunction with each other to make inferences about hidden properties of the entities (persons, for example), as illustrated in Annex C.

Hypernodes can often be treated as representing Boltzmannian macrostates. As such, the total uncertainty of the structure is partitioned into the uncertainty between the hypernodes and the uncertainty within them. A particular analyst at a particular moment might be interested in the structure of the network connecting the hypernodes built on a certain property, and at another moment on the resemblances and differences among the nodes contained within a particular hypernode.

As discussed in Annex C, Entities may have many different properties, each of which could be used as the basis for a different Hypernode restructuring of the network. Groups of entities that belong to the same hypernodes in structures based on different properties share a considerable amount of mutual information. This statement is somewhat tautological, since mutual information between entities on the basis of linkage structure is the measure that allows the construction of the syntactic hypernodes in the first place. Nevertheless, it is useful to emphasize the point, since the mutual information can suggest areas of interest for the analyst to explore. Places in the network where links should exist but do not may lead to the discovery of deliberately hidden nodes, especially if the entity that "should" be linked in the network shows up in a structured way at other hypernode levels.

"Missing links" may be links that have negative fuzzy membership values. They do not exist, but may exist in the future. The non-existence of a "missing link" suggests investigation of the pragmatic embedding field, to judge whether interesting structural changes in the network may be expected.

D.3.2.1 Network Stiffness

In Annex C we discussed the idea of a "phase space" to represent the dynamics of network structure. We also considered fuzzy membership and the effort it takes to move the fuzzy membership of a node or link. This led to the concept of negative fuzzy membership, the degree of negativity being indicated by the amount of political, structural, conceptual, financial and other effort needed to bring the link or node to a point where any more effort would bring its fuzzy membership above zero. Nothing about this was quantifiable. Information theory suggests an approach that might allow the concept to be quantified.

The ideas in what follows are derived from Bagno [1], a paper in economic theory that should have become seminal, but that has been largely forgotten. Bagno correctly noted that the process of manufacture was the creation of structure, otherwise known as the reduction of entropy or the production of information for a user of the manufactured product. Structure represents information, the reduction of uncertainty about the relationships among the components of the structure. The user either produces the structure by making decisions and acting on them, or pays for this information with money, at a rate that is variable over time and across commodities and services. However, averaged over the wide range of transactions that occur every day, the exchange rate of money for structure changes very slowly. In 1955 Bagno estimated it to be about 250 bits to the cent. Today structure might be ten or twenty times more expensive in numerical dollars, say 25 bits to the cent.

In the average transaction, the purchaser does not require all the structure created by the producer, which means that information loss is associated with the transaction. Furthermore, structure decays over time if it is not maintained, requiring the introduction of information to retain the structure, or the loss of information if the structure is not maintained. Bagno's argument is more thoroughly developed than this short abstract might suggest, but the upshot is that money must continuously be created by government debt that leads to a necessary inflation, or the economy will gradually decline. Here we use Bagno's line of argument in connection with the structure on a network.

In Annex C, the concept of a phase space for the network structure was introduced. Phase spaces work well if the changes of location of the phase point are continuous, but not if they involve the kind of instantaneous jump that occurs in a road network when the mayor cuts the ribbon to open a new bridge. The bridge may have taken many years to build. Someone had to create the concept, to sell the idea politically, to organize the financial backing, to engage the designers and engineers, to hire the contractors, and to do the physical work. An outside observer might see the bridge coming ever closer to being a link in the network.

Annex C introduced the concept of a shadow network that manifests this continuous movement toward the point in phase space where the network would arrive after the ribbon-cutting ceremony. Bagno's analysis allows us to quantify the work in progress, and to assess in comparable terms how difficult it is to change a link or a node from one state to another. It quantifies distances in the phase space in a way that is not possible when considering only the abrupt changes that occur in the actual network.

Moreover, the same approach allows for quantification in bits of the negative fuzzy membership that is implied by the difficulty of building a barely useable link. The values may be infinitesimal if the network exists only in someone's mind, but very high if the link under construction (or maintenance) is a six-lane bridge over a wide river. The constraints lie in the embedding fields of the network, particularly in the pragmatic embedding field.

When links are hard to change, the shadow network's location in the phase space moves slowly. We can say that the network is "stiff". Social networks do not tend to be as stiff as those involving physical infrastructure, but it does take effort to create and maintain a social relationship. Changes in a social network cannot always be discerned in a shadow network as easily as they can in a physical network, because the embedding fields tend to be hidden in the minds of the people concerned. Nevertheless, in principle if not in practice, the same considerations apply, and it may be possible to discern different degrees of stiffness based on the participation of the people in more or less formal social institutions.

If one is exploring (see Annex C for an explanation) a network, the stiffness of the network is important. The result of Exploration is valuable only over a time span commensurate with the rate of information loss of the thing studied. A stiff network retains its location in phase space longer than a fluid one, so the more fluid the network, the shorter the half-life of information acquired by Exploring, and the more actual Monitoring will be required to retain situation awareness of its important aspects.

Many networks have internal constraints, some of which were described above when considering feedback loops of short Uncertainty circumferences. Such constraints are likely to increase the stiffness of the parts of a network that participate in the feedback loops. The immediate consequence is that one would expect naturally evolved large networks to consist of a range of relatively stiff sub-nets linked by flexible connections. In other words, modularity should be the usual case. In social networks, we may call such modules "cliques", "churches", "clubs", "committees", "governments", and so forth.

If a network can be identified as having specific stiff modules, it may be possible to resolve relevant problems by considering the modules as nodes in a reduced network, in much the same way as the hypernode construction process allows groups of nodes to be identified as being of the same kind so that the analysis can proceed in terms of the kinds rather than treating each individually. It should be useful to explore stiff modules, even if the network as a whole is insufficiently stiff to allow the results of Exploring to survive over the time span of a problem of interest.

D.4 CONCLUSION

Throughout this annex, the question of actually calculating uncertainty and information values has been deliberately avoided. Conditions for limits on uncertainty and information have been described, but no algorithms for computing these limits. There are two reasons for this deliberate omission. Firstly, as noted in many places, information is always about something, and depends on the observer's prior understanding both of the something and of the interpretation of the data obtained from the observation. Quantification therefore depends intimately on the observer's current state. The actual Uncertainty about the "something" observed also necessarily depends on the observer, but for Uncertainty there is a limit that could be known to the external analyst privy to the possible variations in what might be observed. This limit, signified as Umax, is not observer-dependent. But in most cases, neither is it useful, because the observer's interest in the different possible observable properties changes over time; and neither is it knowable, because the number of properties of an observable entity that might vary is often undefined. Even if the computation is restricted to a defined set of variable properties of interest to an observer, there are usually unknown correlations among elements of what is observed, and these may affect Umax.

Given all these restrictions that put difficulties in the way of actual computation of uncertainty and information values, it may be surprising that information-theoretic measures have any use at all in the study of networks. However, Shannon [10] showed that although the entropy (Umax) of a continuous variable was defined by the choice of coordinates for observing the variable, nevertheless the change of entropy consequent on receipt of a message was well defined in continuum space. Likewise, many of the useful conclusions developed in this annex depend on the direction and relative magnitude of change of uncertainty, rather than on its actual quantity.

In this annex, we have suggested six different distance measures for networks. These distance measures refine the specifications of path length sometimes used in Social Network Analysis. Informationally, distances may relate to latency, channel capacity, or to the minimum uncertainty about a remote node possible at the current node. When these distances are used in respect of a round-trip loop that returns to the current node, they define the limits of the network over which feedback effects influence the dynamic behaviour of the network. In particular, the latency round-trip circumference of the loop influences possible oscillatory behaviour, while both channel capacity and uncertainty distance affect in different ways the possibilities for stabilizing or destabilizing the network.

Informational criteria also suggest differences in display techniques for the different kinds of perceptual task. Monitoring/Controlling usually refers to a pattern in the data space that has a very low uncertainty as compared to the uncertainty of the data space; it is often based on the mutual information among many elements that individually may have quite high uncertainties. Hence, good displays for Monitoring/Controlling probably should emphasise the correlations that correspond to the pattern, and de-emphasise the variations to which the pattern is indifferent. Computerized algorithms that result in low-dimensional displays such as meter dials or bar charts can be useful for this purpose.

Similar considerations apply to Search, which is always a sub-task for Monitoring/Controlling. Exploring, however, is another matter. When one is Exploring, one is usually ignorant of what one might find, which suggests that the display should be agnostic about what it emphasises. Exploring is the main mode in which visualisation is important, so the display should be dense, as opposed to the uncluttered displays best suited for Monitoring/Controlling. Nevertheless, the possibilities for Exploring can be enhanced algorithmically if the user is able to interact with the computer through the "Engines" of the VisTG Reference Model in order to specify particular aspects of the data space currently of interest. The result of Exploring is the discovery of patterns of mutual information that are likely still to be valid when they are needed for Monitoring/Controlling.

Although the original concept of a hypernode was based on the local connectivity of the network nodes, it can be extended to any property for which there is sufficient possible variation to allow meaningful differences in mutual information between nodes. Patterns of common membership in hypernodes based on different properties may usefully identify different "kinds" of nodes or sub-nets within a large net. Such identification might help the analyst to identify places in an uncertain network where important links or nodes might exist. Especially if those elements were deliberately hidden, such hints might prove important.

Information about the future state of the network may be obtained from changes in the pragmatic embedding field. The nature of this information was characterized as taking the form of a "shadow network". The continued reduction of uncertainty about the real network implied by the solidifying of the shadow network might be quantified, at least in a relative sense, by using the ideas introduced to Economic Theory by Bagno [1]. These same ideas might prove relevant to the quantification of positive and negative fuzzy membership values, but to investigate this possibility is a matter for further work.

D.5 REFERENCES

[1] Bagno, S. (1955). The Communication Theory Model and Economics, *IRE Convention Record* 1955 Part 4, 162-181. (Reproduced in facsimile by permission of IEEE at http://www.mmtaylor.net/Economics/Bagno/index.html, accessed October 26, 2012).

[2] Bjørke, J.T. (†), Nilsen, S. and Varga, M. (2010). Visualization of network structure by the application of hypernodes, *Int. J. Approximate Reasoning*, 51, 275-293.

[3] Boltzmann, L. (1866). Über die Mechanische Bedeutung des Zweiten Hauptsatzes der Wärmetheorie. *Wiener Berichte* 53: 195-220. (Reference taken from Wikipedia, accessed 2012.11.10).

[4] Chaitin, G.J. (1987). *Algorithmic Information Theory*, New York: Springer-Verlag.

[5] Denning, P.J. and Bell, T. (2012). The Information Paradox, *American Scientist*, 100, 470-477.

[6] Garner, W.R. (1962). *Uncertainty and Structure as Psychological Concepts*, New York, Wiley.

[7] Garner, W.R. and McGill, W.J. (1956). The relation between Information and Variance Analyses. *Psychometrika*, 21, 219-228.

[8] Gibbs, J.W. (1875-1878). "On the Equilibrium of Heterogeneous Substances," *Transactions of the Connecticut Academy of Arts and Sciences*, 3, 108-248, 343-524. (Reference taken from Wikipedia, accessed 2012.11.10).

[9] Mackay, C. (1841, republished 1932). *Extraordinary Popular Delusions and the Madness of Crowds*, New York: Farrar, Straus and Giroux.

[10] Shannon, C.E. and Weaver, W. (1949, reprinted 1998). *The Mathematical Theory of Communication*, U. of Chicago Press.

Annex E – INTERACTIVE HYPOTHESIS VISUALIZATION

E.1 INTRODUCTION

Hypothesis visualization is concerned with supporting the construction of an argument – it shows:

- Relevant information and can highlight items of importance;
- The temporal development of facts and arguments; and
- How pieces of information support or refute alternative deductions.

It is important that hypothesis visualization shows any uncertainty inherent in the information, as well as duplicate, redundant and implausible pieces of information. Effective hypothesis visualization will draw attention to flaws in an argument, contradictory information, identify missing information and make clear the balance of the argument. The main benefit of a graphical or structural display of the network of information relations is that it enables the user to understand (mentally visualize) the issues more effectively than any other means.

This annex will consider the roles of information, and the relationships between items of information, in supporting/refuting a hypothesis or hypotheses, and the role of the uncertainty inherently associated with information. The PAGAn hypothesis visualization tool is used to illustrate aspects of this methodology.

E.2 HYPOTHESIS VISUALIZATION OVERVIEW

There has been a growing recognition of the benefits of using computer software tools for hypothesis management and support that allow the user to construct and visualize arguments in the form of graphs of nodes and links [6]. These tools are designed to make argument construction and visualization easier for the users as well as to guide the users in the evidential reasoning process. However, in general in the existing tools and approaches, no representation/differentiation has been made to acknowledge and represent the different types of evidence or facts, and that they vary in the degree to which they supporting or refuting a hypothesis. Furthermore, the majority of existing systems have not been concerned with handling dynamically developing situations nor the issues of representing and handling the uncertainties which are inherent in any data or information.

E.3 UNCERTAINTY MANAGEMENT

In any analysis a high degree of analytical judgement is required to manage the inevitable and inherent uncertainties in the data or information [4]. It is important to manage these uncertainties in a consistent manner, and that the collected evidence is used effectively and comprehensively to support the final decision made by the user, taking into account the uncertainties.

This can be achieved, for example, if the users or decision makers are able to estimate the relevance and the value of evidence for or against each of a number of competing hypotheses and to use each item of evidence in the analytical process accordingly.

In the case of textual information, such as a report, it is important to know the source, the when and who generated the information, etc. Furthermore, sometimes it is useful to see if there are other related reports associated with the report in question, past or present. Indeed, lots of supporting evidence is beneficial but lots of supporting evidence from independent sources is even better and more influential in supporting/refuting the hypothesis in question.

There is therefore a need for tools for exploiting textual data/information that enable:

1) The extraction of evidence;
2) Handling evidence relevance, validity and uncertainty; and
3) The generation, manipulation, visualization and the evaluation of evidence and hypotheses.

E.4 WIGMORE CHART

A Wigmore chart is a graphical method for the analysis of legal evidence in trials [1],[2],[5] and [8]. It is a landmark evidential method; it is the first system that charts arguments diagrammatically.

One of the strengths of the Wigmore approach is its handling and representation of the balance of view, c.f. bias.

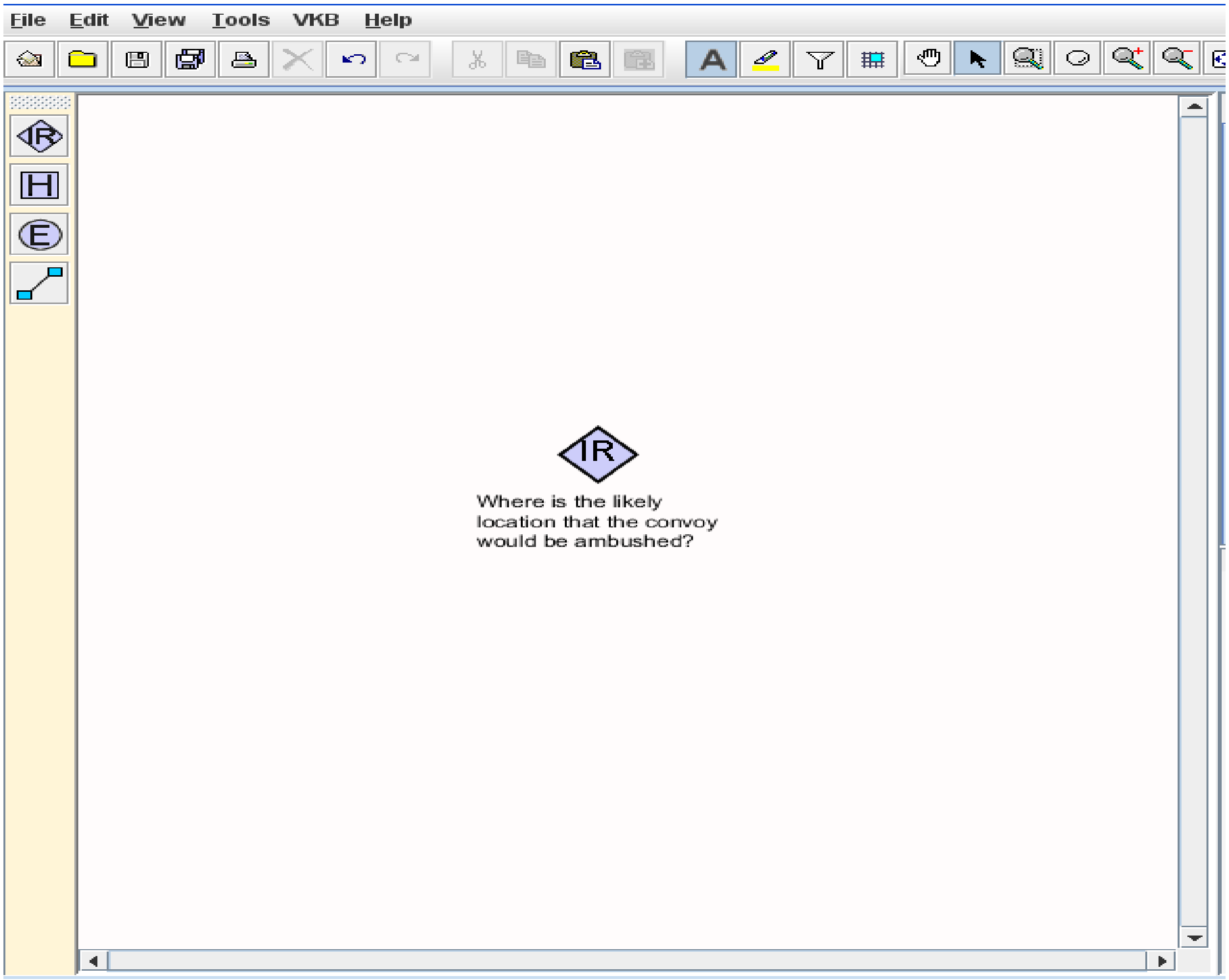
In a Wigmore chart, the various types and items of evidence gathered supporting and refuting a hypothesis are used so that the strength of the case can readily be observed. In particular, it can be seen if more evidence is required, e.g. to minimise the danger of self-confirming or to strengthen an aspect of the case/hypothesis.

E.5 A HYPOTHESIS VISUALIZATION EXAMPLE

This section shows an example of how hypotheses can be formulated and visualized so as to take into account the dynamic temporal and uncertain nature of evidence using a prototype hypothesis visualization system PAGAn [7]. It is a simplified network system based on the Wigmore concept; it is used here to illustrate some of the ideas of hypothesis visualisation.

An Intelligence Request (IR) concerning a convoy and where it is likely to be ambushed is shown in Figure E-1. An Analyst responding to such a request is to generate a number of hypotheses and gather evidence to support/refute the proposed hypotheses. In this example, the Analyst hypothesises the likely locations that the convoy could be ambushed. The Analyst gathers evidence to answer question such as:

- Where have recent ambushes been?
- Where are insurgents currently known to be operating?
- What types of ambushes are the insurgents capable of? Land or sea?
- Where and when is the convoy going, and what route will it take? What is the certainty that this route is not going to change (e.g. commander deciding to change the route, unexpected roadblock)?
- What is the weather forecast? Hurricane? Snowing?

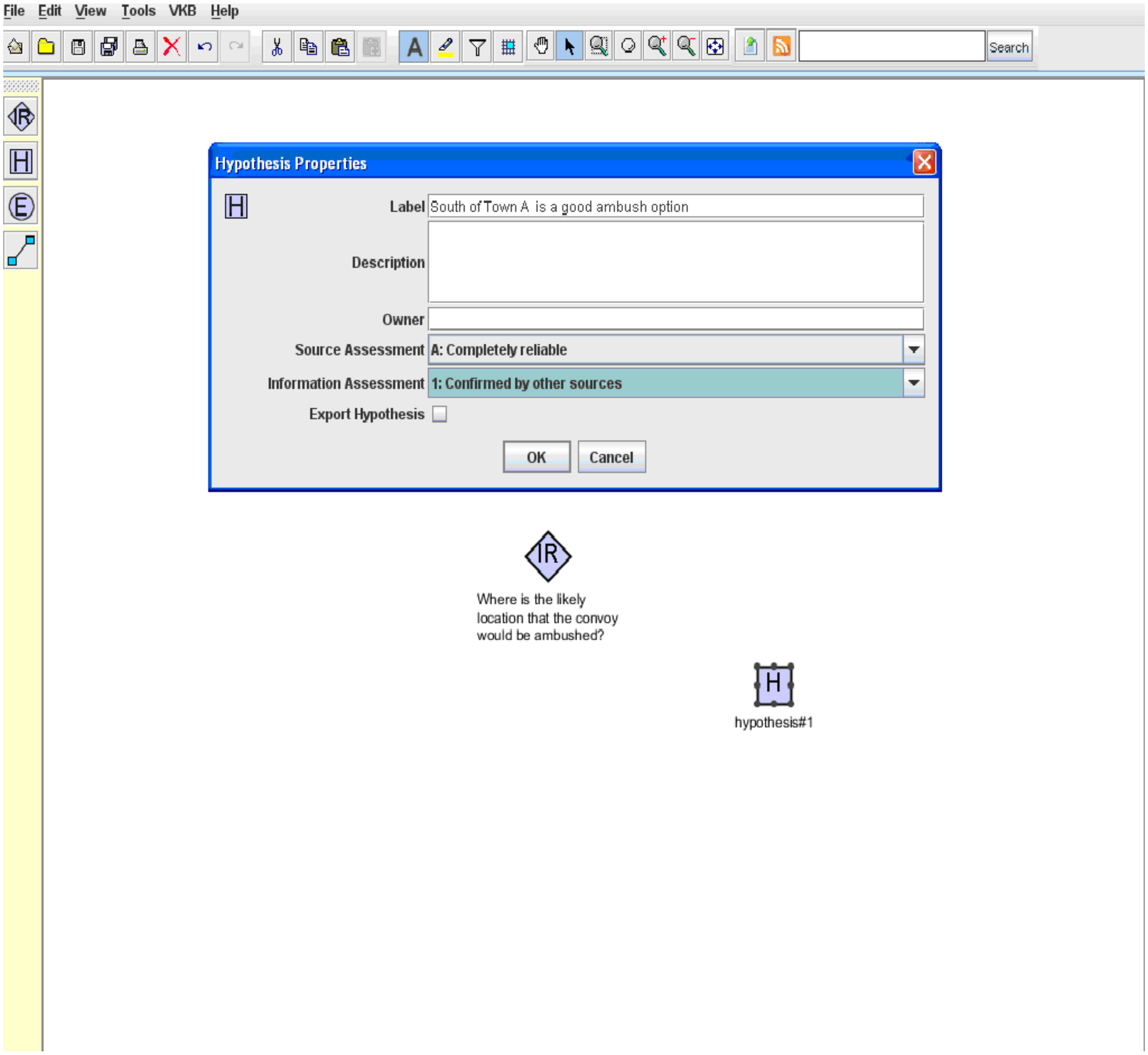


Where is the likely location that the convoy would be ambushed?

Figure E-1: An Intelligence Request (IR).

Figure E-2, a hypothesis was introduced that suggests that south of Village A is a good ambush option. The Analyst entered properties of the hypothesis by assessing its source (see Figure E-2), which can be rated as being:

1) Completely reliable;
2) Usually reliable;
3) Fairly reliable;
4) Not usually reliable;
5) Unreliable; and
6) Reliability cannot be judged.



Label: South of Town A is a good ambush option

Description:

Owner:

Source Assessment: A: Completely reliable

Information Assessment: 1: Confirmed by other sources

Export Hypothesis ☐

Where is the likely location that the convoy would be ambushed?

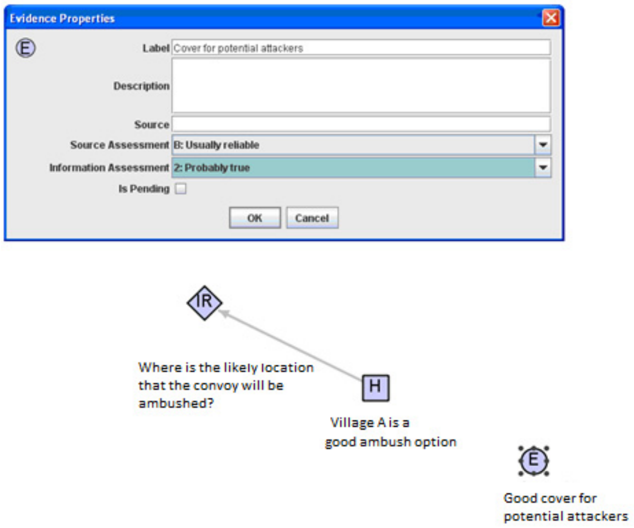
Figure E-2: Hypothesis of South of Village A is a Good Ambush Option and its Properties.

The information can also be assessed by different scores, such as:

1) Confirmed by other sources;
2) Probably true;
3) Possibly true;
4) Doubtful;
5) Improbable; and
6) Truth cannot be judged.

See again the Dialogue Box in Figure E-2. In this way each hypothesis that is being considered can be assessed so that effort is only made to gather evidence for hypotheses that are worth considering. In this example the hypothesis is assessed to be completely reliable and is based upon what can be confirmed by reliable sources. The Analyst can also export the hypothesis to share with other Analysts and make notes about the hypothesis as well as declare any ownership.

Next the Analyst is to gather evidence to support or refute this particular hypothesis. For example, there is supporting evidence to show that there is cover for the attackers at Village A, which makes Village A vulnerable, see Figure E-3. The properties of the evidence are entered using the same rating as the hypothesis – see the Dialogue Box in Figure E-3. The evidence in this example is believed to be usually reliable and probably true. More evidence will be gathered for this hypothesis. Similar processes will be used for other hypotheses. Furthermore, different types of evidence can be used; for example, significant trends might emerge from circumstantial evidence which can be correlated with other evidence. The 'Is Pending', checkbox allows the Analyst to represent evidence that someone has been tasked to collect. It can be added to the graph (and is shown in white) to be assessed later when the evidence is available.



Evidence Properties

Label: Cover for potential attackers

Description:

Source:

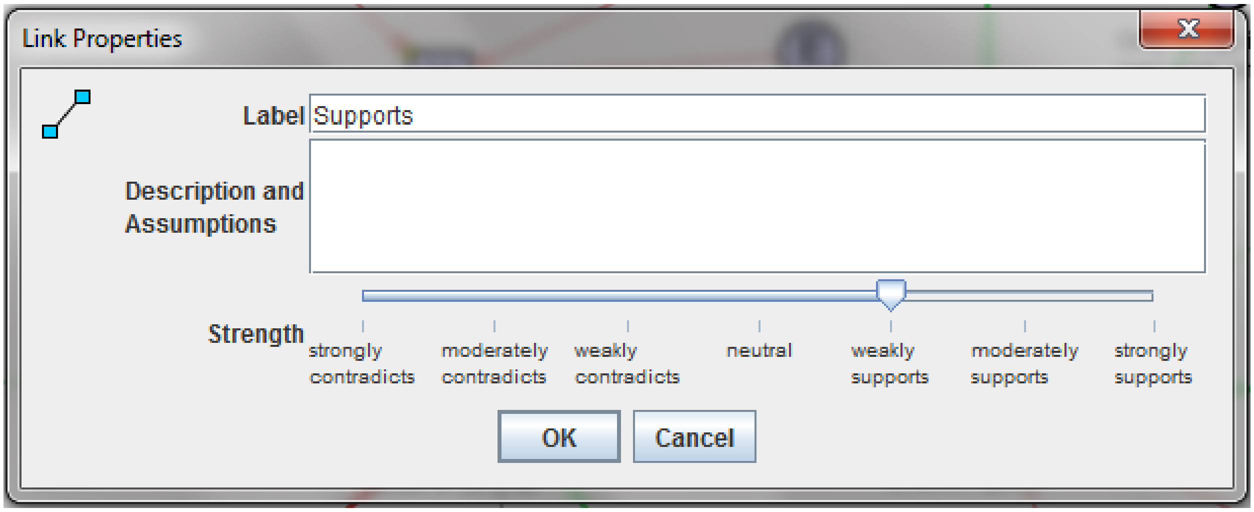
Source Assessment: B: Usually reliable

Information Assessment: 2: Probably true

Is Pending ☐

Figure E-3: Properties of the Evidence that there is a Good Cover for the Attackers.

The Analyst also enters the strength of the evidence, i.e. link in supporting and refuting the hypothesis, see Figure E-4. This varies from strongly/moderately/weakly contradicts, and not assessed, to weakly/moderately/strongly supports.

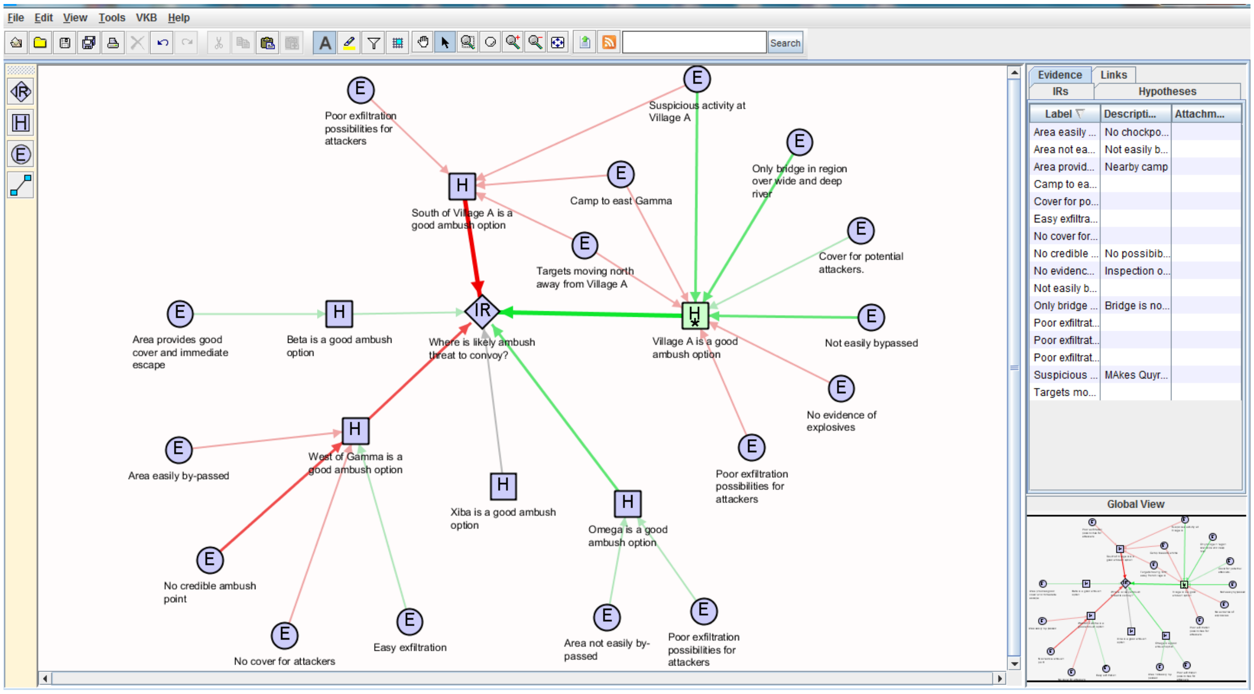


Label: Supports

Strength: weakly supports

Figure E-4: Link Properties.

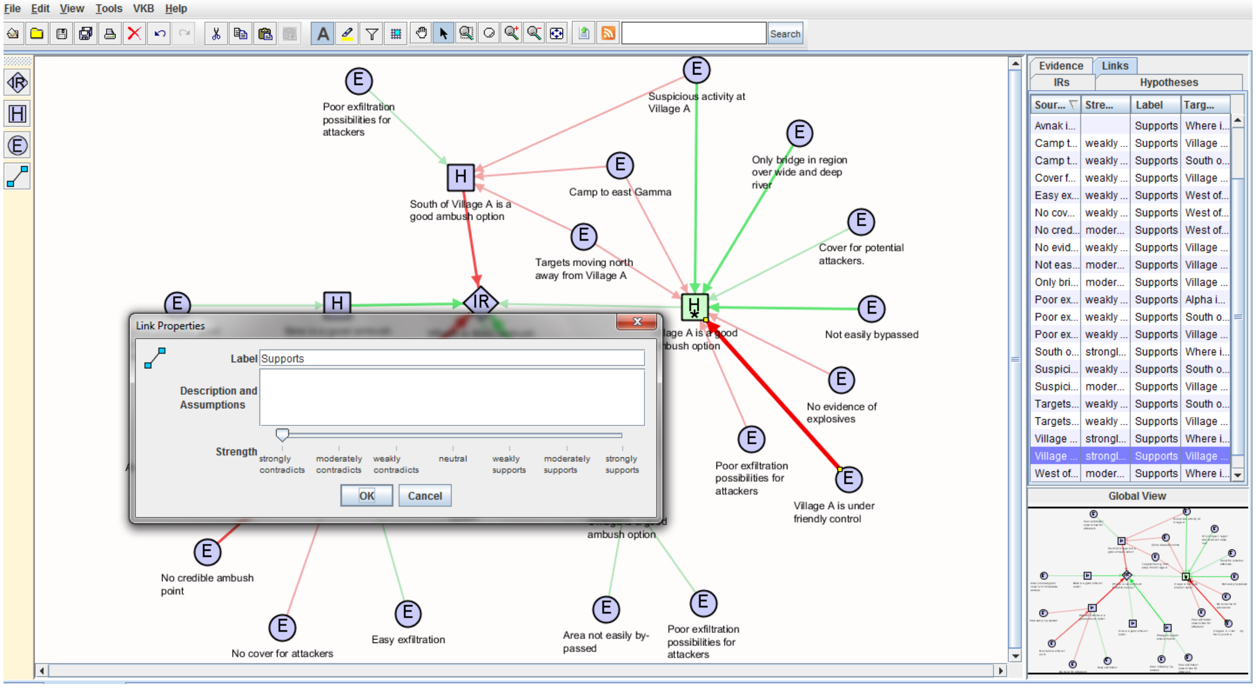
Figure E-5 shows a network of hypotheses and evidence built up in answering the Intelligence Request. Among the six hypotheses, no evidence has been collected for the hypothesis that Xiba is a possible ambush location. Therefore, there is either the need to gather evidence to address the proposed hypothesis or remove it from the view. It is always possible to display the hypothesis again when evidence becomes available and it is still a viable hypothesis. Thus, Figure E-5 could be interpreted as showing a balance of view as to what efforts have been invested in collecting evidence to support/refute any hypothesis, e.g. as the Xiba hypothesis has no supporting evidence. The upper right hand table allows all the evidence and their associated information to be examined. While the bottom right hand panel provides an overview facility for a network and can be used to navigate to different parts of the network. However, graphical displays only work up to certain size and complexity. Beyond this, the user is faced with 'spaghetti' that is hard to interpret, hence the need of filtering and/or simplification, see for example the Hypernode technique [3].



Label	Descripti...	Attachm...
Area easily ...	No chockpo...	
Area not ea...	Not easily b...	
Area provid...	Nearby camp	
Camp to ea...		
Cover for po...		
Easy exfiltra...		
No cover for...		
No credible ...	No possibib...	
No evidenc...	Inspection o...	
Not easily b...		
Only bridge ...	Bridge is no...	
Poor exfiltrat...		
Poor exfiltrat...		
Poor exfiltrat...		
Suspicious ...	MAkes Quyr...	
Targets mo...		

Figure E-5: Network of Six Hypotheses.

The Analyst can use the same evidence to support multiple hypotheses or support one hypothesis and refute another, for example, information such as “suspicious activity at Village A” may affect multiple hypotheses. In this way, it is not necessary to enter the same evidence for different hypotheses. This is an advantage so as not to create the wrong impression of the number of evidence items available; c.f. the same evidence appearing multiple times for different hypotheses. In Figure E-5, the influence of individual and combined evidence is analyzed automatically for the six hypotheses and the effect of new information on the hypothesis/hypotheses can be visualized instantly. The system calculates the strength of all the evidence of all the hypotheses and shows that Village A is a good ambush location, so the convoy should either alter its route or be prepared for the likely ambush. Hypothesis analysis is a dynamic process, so new hypotheses can be generated or removed when the situation changes or new evidence is gathered or there are changes in existing evidence. It is vital that new situations can be realised readily and easily, i.e. the hypotheses can be updated readily. Figure E-6 shows an example, here there is new evidence to show that Village A is under friendly control which results in immediate change in the situation and makes Village A less likely to be a good ambush site. In view of this new evidence the convoy is now more likely to be ambushed in Omega, see Figure E-7.



Link Properties

Label: Supports

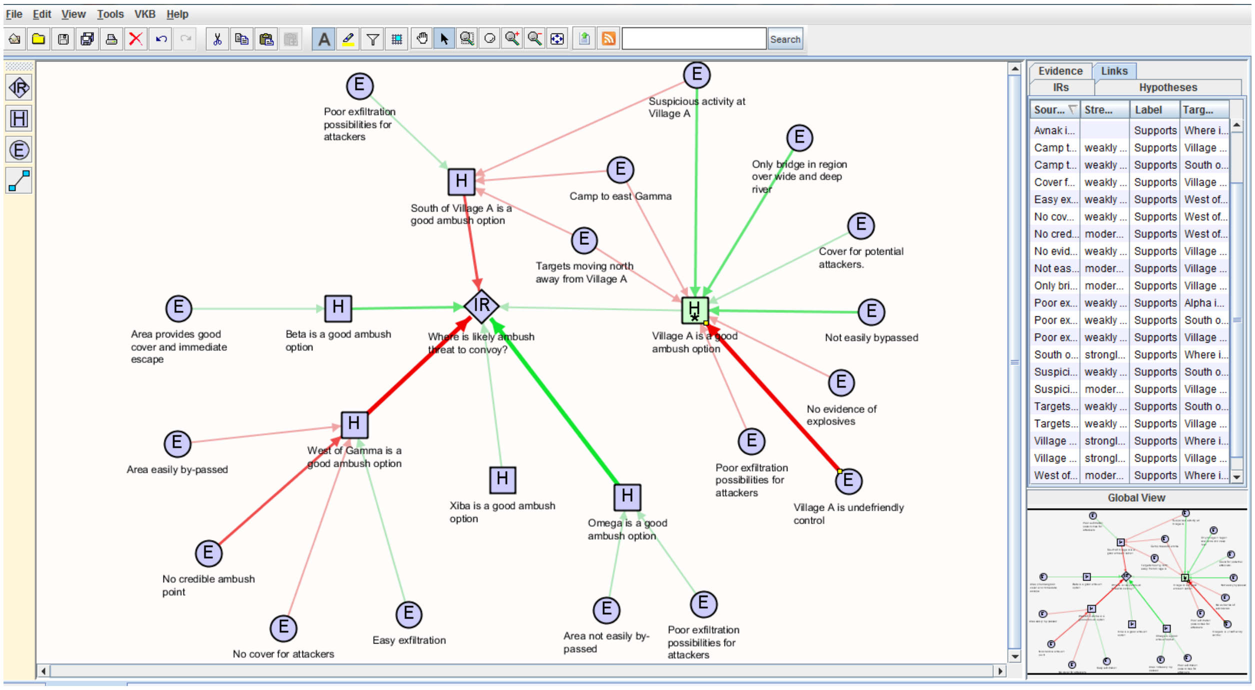
Description and Assumptions

Strength: strongly contradicts | moderately contradicts | weakly contradicts | neutral | weakly supports | moderately supports | strongly supports

OK
Cancel

Sour...	Stre...	Label	Targ...
Avnak i...		Supports	Where i...
Camp t...	weakly ...	Supports	Village ...
Camp t...	weakly ...	Supports	South o...
Cover f...	weakly ...	Supports	Village ...
Easy ex...	weakly ...	Supports	West of...
No cov...	weakly ...	Supports	West of...
No cred...	moder...	Supports	West of...
No evid...	weakly ...	Supports	Village ...
Not eas...	moder...	Supports	Village ...
Only bri...	moder...	Supports	Village ...
Poor ex...	weakly ...	Supports	Alpha i...
Poor ex...	weakly ...	Supports	South o...
Poor ex...	weakly ...	Supports	Village ...
South o...	strongl...	Supports	Where i...
Suspici...	weakly ...	Supports	South o...
Suspici...	moder...	Supports	Village ...
Targets...	weakly ...	Supports	South o...
Targets...	weakly ...	Supports	Village ...
Village ...	strongl...	Supports	Where i...
Village ...	strongl...	Supports	Village ...
West of...	moder...	Supports	Where i...

Figure E-6: Introduce new evidence and its role in supporting or refuting the hypothesis.



Sour...	Stre...	Label	Targ...
Avnak i...		Supports	Where i...
Camp t...	weakly ...	Supports	Village ...
Camp t...	weakly ...	Supports	South o...
Cover f...	weakly ...	Supports	Village ...
Easy ex...	weakly ...	Supports	West of...
No cov...	weakly ...	Supports	West of...
No cred...	moder...	Supports	West of...
No evid...	weakly ...	Supports	Village ...
Not eas...	moder...	Supports	Village ...
Only bri...	moder...	Supports	Village ...
Poor ex...	weakly ...	Supports	Alpha i...
Poor ex...	weakly ...	Supports	South o...
Poor ex...	weakly ...	Supports	Village ...
South o...	strongl...	Supports	Where i...
Suspici...	weakly ...	Supports	South o...
Suspici...	moder...	Supports	Village ...
Targets...	weakly ...	Supports	South o...
Targets...	weakly ...	Supports	Village ...
Village ...	strongl...	Supports	Where i...
Village ...	strongl...	Supports	Village ...
West of...	moder...	Supports	Where i...

Figure E-7: Omega a Good Ambush Location.

The hypotheses can be saved and re-used or modified if a similar request is made in the future.

E.6 CONCLUSION

This paper has discussed the interactive formation and visualization of dynamic hypothesis visualization using the PAGAn system. It shows the benefits in being able to interactively analyse and visualize the changes in evidence and hypotheses. Furthermore, it take into account the degree of uncertainty of each piece of evidence and its role in supporting and refuting different hypotheses which can be expressed by the user explicitly based on assessment of its source, assessment of the information, as well as consideration of the links between the evidence and hypotheses. Systems such as this provide invaluable support for informed and dynamic decision making.

E.7 REFERENCES

[1] Anderson, T. and Twining, W. (1998). Analysis of Evidence: How to Do Things with Facts Based on Wigmore's Science of Judicial Proof. Evanston, IL: Northwestern University Press.

[2] Anderson, T., Schum, D. and Twining, W. (2005). Analysis of Evidence, Second Edition, Cambridge University Press.

[3] Bjørke, J.T. (†), Nilsen, S. and Varga, M.J. (2010). Visualization of network structure by the application of hypernodes, International Journal of Approximate Reasoning 51, 275-293, 201.

[4] Bowman, N., Iswaran, N., Freemantle, Hines R., Swarbrick, A., Varga, M.J. and Young. R. (2008). Visualising the Propagation of Uncertainties, 9th NATO Network of Experts Workshop on Visualizing Network Dynamics, Malvern, UK, 4-6 November.

[5] Dawid, A.P. and Evett, I.W. (1997). Using a graphical method to assist the evaluation of complicated patterns of evidence. Journal of Forensic Sciences 42, 226-231.

[6] Lankenau, R.A., Eick, M.A., Decherd, A., Khailo, M., Paris, P. and Fugitt, J. (2006). DECIDE, IEEE, VisWeek, VAST Challenge.

[7] Varga, M.J. and Adams, K. (2010). Interactive hypothesis visualization, NATO Workshop on Visualising Networks: Coping with Change and Uncertainty, IST-093/RWS-015, Rome, New York, USA, 19-21 October.

[8] http://en.wikipedia.org/wiki/John_Henry_Wigmore.

Annex F – HYPERNODE

F.1 INTRODUCTION

Graphs are often used in the modelling of real-world network-related phenomena, such as computer and telecommunications networks, social networks, road and transport networks as well as the representation of abstract concepts for instance of information or in the analysis of systems, documents, images, etc. In many cases networks of data tend to be large and complex and it is all too easy, in attempting to analyze the information contained within a network to miss or overlook critical characteristics or clues, it is also hard to avoid human bias in the analysis of such networks.

The reduction of the complexity is an important enabler in the understanding, analysis, management and visualisation of networks with large and complex numbers of nodes and links [2],[3],[5],[7]-[13].

Bertin [1] introduced a graphical method to discover categories or groups within geographical data, which he termed Seriation. In his method the data is initially mapped to an image, which Bertin termed the reorderable matrix. The rows, or columns, of this image are then interchanged to generate different views of the data, and meaningful patterns in the data can be detected by visual interpretation of the reordered image. Bertin also constructed a mechanical permutation technique.

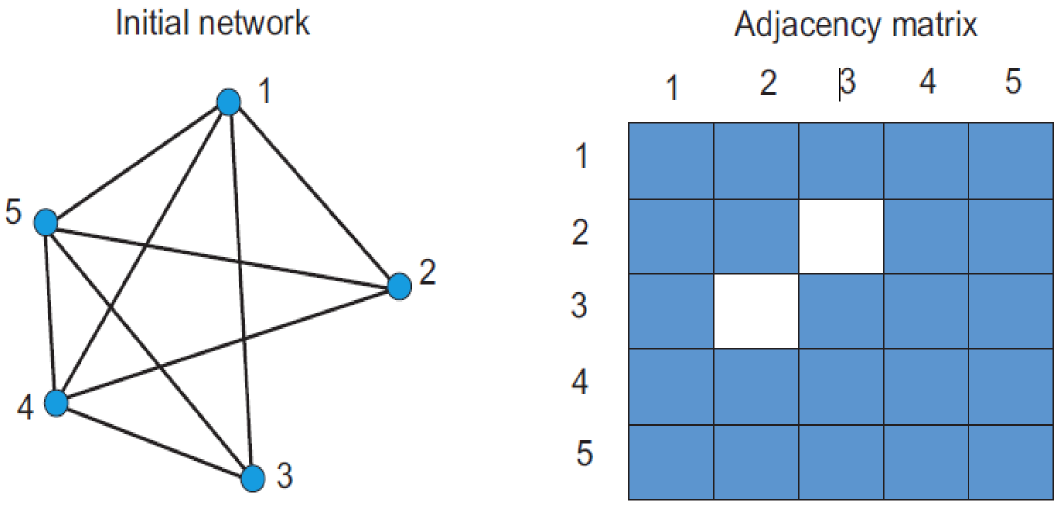
Based, also, on the idea of the re-orderable matrix Siirtola and Mäkinen [12] developed a tool for interactive cluster analysis. Bjørke and Smith [4] developed an algorithm to automate the reorganization (also termed seriation) of the reorderable matrix in which the seriation criterion is defined on the basis of the minimum entropy of a binary image. In the case of a network the re-orderable matrix is the adjacency matrix of the network – the adjacency matrix is a representation of which network nodes are adjacent to other nodes.

F.2 THE HYPERNODE TECHNIQUE

Hypernode is network abstraction or simplification technique; it is uses information theory to construct hierarchies of networks based on reorganization of the adjacency matrix of a network. It was developed to abstract complex networks; it aggregates nodes and links into hyper-nodes and hyper-links by reordering the adjacency matrix to generate hierarchies of hyper-networks. It thus enables networks to be represented at different levels of abstraction, and can operate as an interactive map that allows the user to zoom in and out so as to gain an understanding of the nature of a network, adapting the level of detail to the degree desired by the user. In other words it can transform a flat network into a hierarchical structure that can reveal and highlight the underlying structure/pattern of the network in an effective and intuitive manner. It therefore provides a highly effective means for exploring large and complex networks [5].

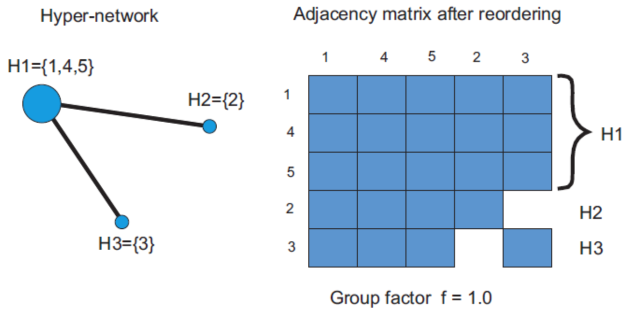
F.3 PRINCIPLES OF HYPERNODE

In this section the principles of the Hypernode technique are introduced. Figure F-1 shows a simple 5 nodes network and its associated adjacency matrix [6]. If there is a connection between two nodes in the network, the corresponding cell in the image of the Adjacency Matrix is coloured in blue, else it is white. Numerically, the binary property of the matrix can be represented by the numbers 1 or 0, i.e. a characteristic function. Figure F-2 demonstrates how reordering can be used to get a view of the adjacency matrix where groups of nodes can be derived. The network in Figure F-2 is constructed by clustering: grouping of the nodes requires a similarity measure, and an index on the scale [0, 1] is used to measure the minimum similarity for the rows or columns in a group. If the group factor, *f*, for example is 1, i.e. the rows must have a similarity measure of 1 in order to be mapped to the same group. The user sets or varies the group factor according to the tasks or to explore the characteristics of the networks. Furthermore, the size of the node is proportional to its connectivity. The thickness of the link represents the number of sub-links of a hyper-link. From the reordered matrix, three groups of nodes can be derived, i.e. a hyper-network with three hypernodes H1, H2 and H3, see Figure F-2.



	1	2	3	4	5
1	■	■	■	■	■
2	■	■	□	■	■
3	■	□	■	■	■
4	■	■	■	■	■
5	■	■	■	■	■

Figure F-1: A Network and its Adjacency Matrix. If there is a connection between two nodes in the network, the corresponding cell in the adjacency matrix is coloured blue (i.e. a value of 1), else it is white (i.e. a value of 0).



	1	4	5	2	3
1	■	■	■	■	■
4	■	■	■	■	■
5	■	■	■	■	■
2	■	■	■	■	□
3	■	■	■	□	■

Figure F-2: The Adjacency Matrix in Figure F-1 After Reordering. The group factor, *f,* is set to 1, i.e. rows must have similarity 1 (be identical) in order to be aggregated to a hypernode.

Hypernode H1 represents an aggregation of the three sub-nodes 1, 4 and 5. From the network of hypernodes, it can be observed that the initial network in Figure F-1 can be regarded as a tree structure, one mother node and two leaf nodes (Figure F-2). The mother node H1 is composed of the three sub-nodes (1, 4 and 5) which have strong connection, i.e. each node is connected to the other two. There is no direct connection between nodes 2 and 3, so Hypernodes H2 and H3 have no direct link.

In order to visualise the number of sub-nodes that a hypernode composes, the visual variable size is used. For example, in Figure F-2 it can be observed that Hypernode 1 with three sub-nodes is represented by a much larger circle than the other nodes. Similarly, the thickness of the line is proportional to the number of sub-links of a hyper-link.

F.3.1 An Example

This section demonstrates the application of hypernode on a more complex network using a grouping factor *f* = 0.5. Figure F-3 shows a network with 233 nodes [5]. It is assumed that an end-node has strong relations to its mother node so they can be mapped to their mother node, as shown in Figure F-4. This mapping will be performed before applying the seriation algorithm. Note that in Figure F-4, an isolated node at the bottom right was revealed that was previously hidden in the initial cluttered network.

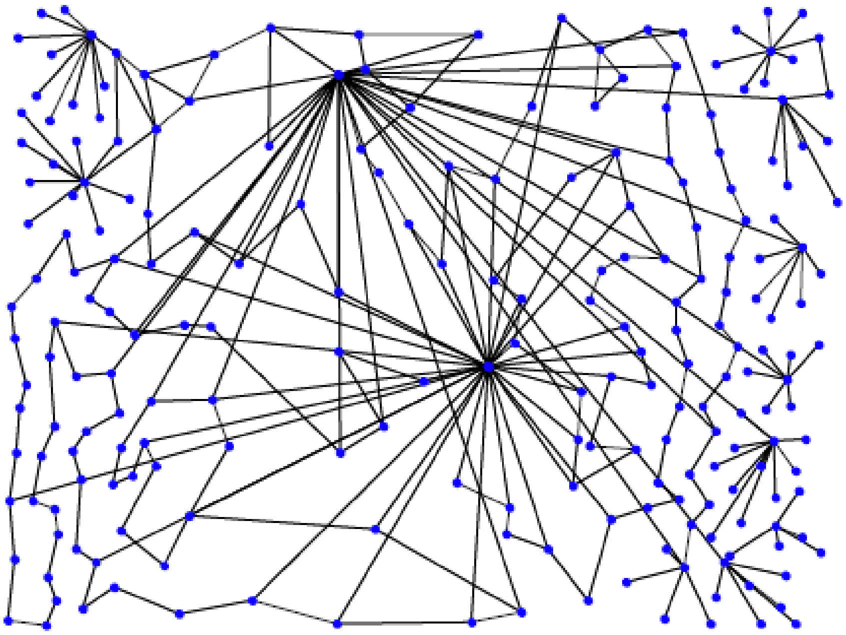


Figure F-3: An Initial Network with 233 Nodes.

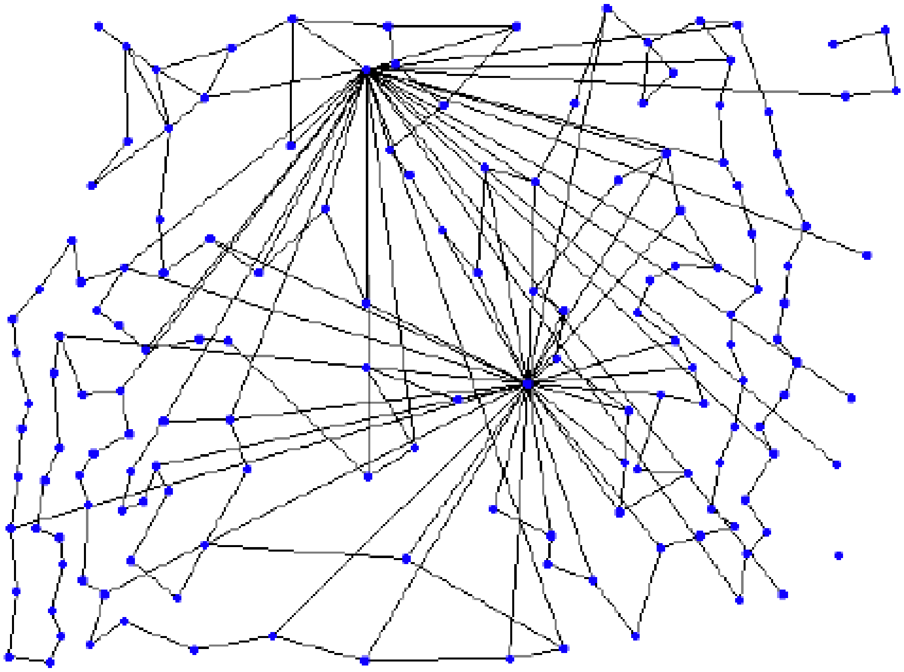


Figure F-4: Network After Mapping End Nodes to Their Mother Nodes.

Figure F-5 shows the sequence of hyper-networks that was generated by the hyper-network algorithm. Figure F-5.1 shows the initial network with 233 nodes, while Figure F-5.2 shows the network where the end nodes are mapped to their mother node. The isolated node labelled C is now clearly visible which was hidden in Figure F-5.1. The seriation algorithm was applied to network in Figure F-5.2. The recursive nature of the algorithm is demonstrated in Figure F-5.2 – Figure F-5.8. The hypernode in Figure F-5.4 is the result of applying the seriation to the network in Figure F-5.3 and so forth. The size of the circle and the thickness of the links are proportional to the number of sub-nodes and sub-links respectively. The process continues until it reaches the status in Figure F-5.8, which results in one large hypernode and a small isolated node C.

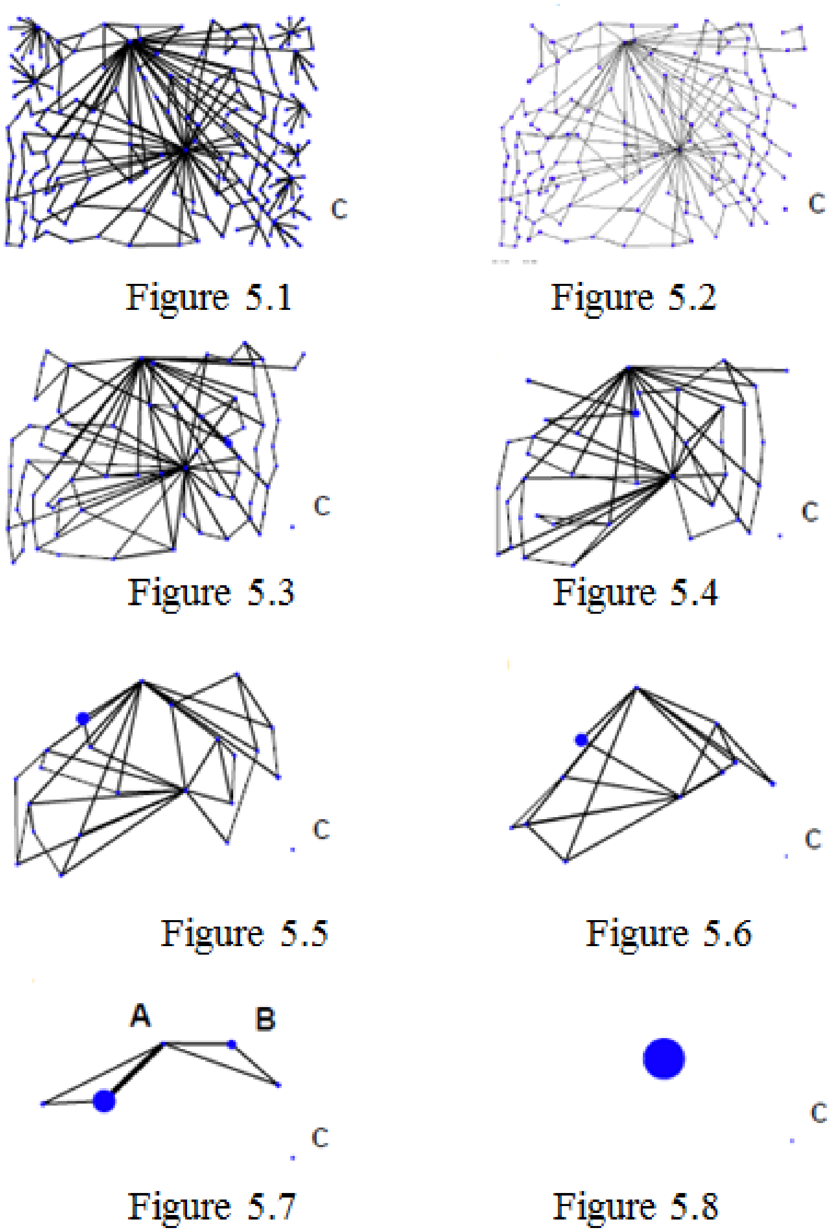


Figure 5.1 Figure 5.2

Figure 5.3 Figure 5.4

Figure 5.5 Figure 5.6

Figure 5.7 Figure 5.8

Figure F-5: A Hierarchy of Hyper-Networks with Group Factor 0.5.

F.4 CONCLUSIONS

This has shown that Hypernode provides an effective means to abstract complex networks, it enables users to zoom in and out of a network so as to explore and understand the network's characteristics. A technique of reordering the adjacency matrix of the network allows a hierarchy of hyper-networks to be derived. The re-ordering of the matrix is based upon the minimum entropy criteria. Furthermore, by varying the grouping factor used in the reordering process different aspects of the networks can be brought out and explored.

F.5 REFERENCES

[1] Bertin, J. (1981). *Graphics and graphic information processing.* Walter deGreuter: Berlin, 273 pp.

[2] Berge, C. (1973). *Graphs and Hypergraphs.* North-Holland, Amsterdam-London.

[3] Berge, C. (1989). *Hypergraphs: Combinatorics in Finite Sets.* Elsevier, New York.

[4] Bjørke, Jan T. (†) and Smith, B. (1996). Seriation: An implementation and case study. Computers, Environments and Urban Systems, 20(6): 427-438.

[5] Bjørke, Jan T. (†) (2006). Reduction of complexity: an aspect of network visualization. *"RTO-MP-IST-063 Visualising network information, NATO conference, Copenhagen, Denmark",* 9-1 to 9-10. NATO. ISBNs 92-837-1156-4 or 978-92-837-1156-8.

[6] Bjørke, J.T. (†), Nilsen, S. and Varga, M.J. (2010). Visualization of network structure by the application of hypernodes, International Journal of Approximate Reasoning 51, 275-293.

[7] Börner, K., Sanyal, S. and Vespignani, A. (2007). Network Science. In Blaise Cronin, editors, Annual Review of Information Science and Technology, Chapter 12, 537-607. Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, USA.

[8] Borkowski, A., Grabska, E. and Szuba, J. (1999). Visualisation of Graphs in ArchiCAD, Lecture Notes in Computer Science, Vol. 1774, 241-246.

[9] Flake, G.W., Tarjan, R.E. and Tsioutsiouliklis, K. (2004). Graph Clustering and Minimum Cut Trees. *"Internet Mathematics",* 1(4):385-408.

[10] Huang, X. and Lai., W. (2006). Clustering graphs for visualization via node similarities. Journal of Visual Language & Computing, 17:225-253.

[11] Lee-Kwang, H. and Cho, C.H. (1996). Hierarchical Reduction and Partition of Hypergraph, IEEE Transaction on Systems, Man, and Cybernetics – Part B: Cybernetics, Vol. 26, Number, April.

[12] MacEachren, A.M. (1994). Visualization in modern cartography: Setting the agenda. In MacEachren, Alan M. and Taylor, D.R. Fraser, Editors, Visualization in modern cartography, 1-12. Pergamon, Oxford.

[13] Siirtola, H. and Mäkinen, E. (2005). Constructing and reconstructing the reorderable matrix. *"Information Visualization,"* 4(1): 32-48.

Annex G – UNCERTAINTY

G.1 INTRODUCTION

Graphs are often used in the modelling of real-world network such as transportation, communications and social networks [1]-[3],[11] and [14], etc. One of the key issues for the visualization of real networks is that uncertainty is an aspect that must be considered. Indeed uncertainty is inherent in even the seemingly most certain networks such as road and rail networks (can you be sure that there is not a traffic jam or road block due to an accident, etc.). Uncertainty, thus poses a challenge that needs to be addressed. As the number of nodes and links increases, compounded with uncertainty, the representation of the network needs simplification in order to keep the clarity of the visualisation while taking into account, for example, the propagation of, and the degree of, uncertainties and their effects on the effective topological structure of the network.

G.2 HYPERNODE FOR WEIGHTED NETWORK

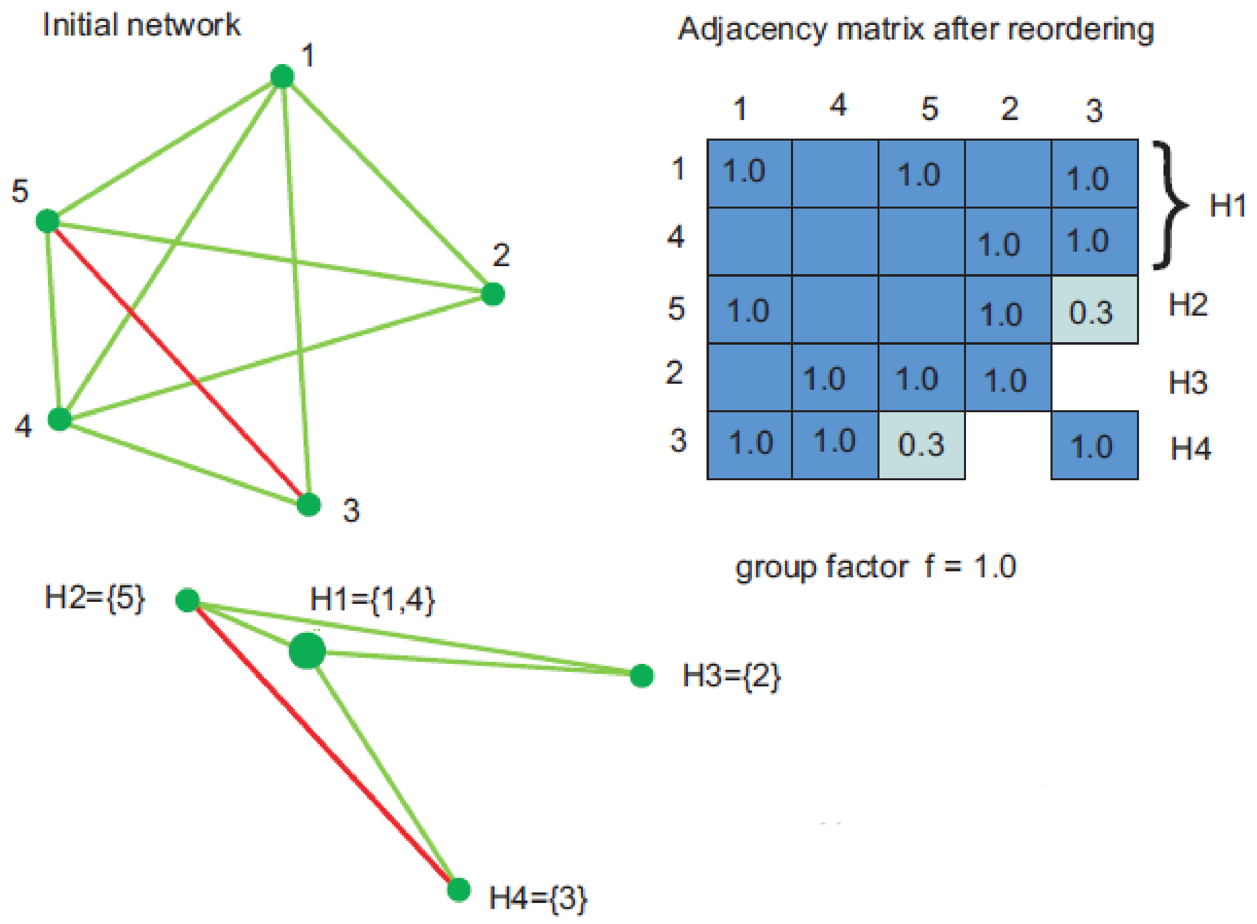
Hypernode is a network abstraction or simplification technique; it is uses information theory to construct hierarchies of networks based on reorganization of the adjacency matrix of a network, see Annex F. It was developed to abstract complex networks; it aggregates nodes and links into hyper-nodes and hyper-links by reordering the adjacency matrix to generate hierarchies of hyper-networks [4]-[6]. Annex F discussed the application of Hypernode to certainty-based networks.

In this annex the Hypernode algorithm is extended to handle uncertain relationships [7],[8] and [10]. There are two main categories of uncertainties in a network, namely uncertainties about the edges or links and uncertainties about the nodes; both will be discussed in this annex.

If uncertainty about an edge is mapped to a membership function in a class such as “perfect edge”, the concept of fuzzy relations can be applied [13] and [15]. Crisp relations can be described by their characteristic function, i.e. an edge in a crisp network is associated with the number 1 or 0, which depends on whether the edge exists or does not exist. In a fuzzy relation (binary) the edges are allowed to have varying degrees of membership within the relation [13]. Although uncertainty is quite distinct from fuzziness, such a mapping is considered appropriate in this context.

This annex introduces a simple approach to advance the crisp Hypernode algorithm to a weighted network such as a fuzzy network [7] and [9]. This is achieved by replacing the characteristic function with a membership function or a weight function.

In an example developed from that given in Annex F, the Hypernode algorithm is applied to the reordering and grouping of the adjacency matrix in Figure G-1. The group factor is still set to 1.0. Figure G-1 demonstrates how uncertainty in a network can be visualized by the application of colour coding. A traffic light colour scheme is used. For simplicity the effect of strong (certain) and very weak (highly uncertain) links is shown. The green nodes and links have membership value $\mu(r_{ij})$ =1.0, certain, and the red link$\mu(r_{ij})$ = 0.3, uncertain. Compared with the crisp network in Annex F, the hypernode H1 in the weighted case represents one sub-node less than in the crisp case, i.e. only Nodes 1 and 4 are included and not Node 5. This is due to the uncertain link between Node 5 and 3. Therefore, Node 5 is not aggregated to hypernode H1 in the weighted (fuzzy) case when the group factor is set to 1.0. The role of H1 changes here due to the ‘uncertain’ link between Nodes 3 and 5, i.e. it is no longer the most connected node. The topological structure and relationship can be visualised very clearly in this hypernode representation.



	1	4	5	2	3	
1	1.0		1.0		1.0	H1
4				1.0	1.0	
5	1.0			1.0	0.3	H2
2		1.0	1.0	1.0		H3
3	1.0	1.0	0.3		1.0	H4

Figure G-1: A Weighted Network, its Adjacency Matrix and its Network of Hypernodes. The group factor *f* is set to 1.0. The green nodes and links have membership value ($\mu_{i,j}$) = 1 and the red link ($\mu_{i,j}$) = 0.3

The group factor plays an important role in determining the structure of the hypernode and therefore must be set carefully to explore the network relationship. Figure G-2 shows the result of reducing the group factor from 1.0 to 0.8, i.e. relaxing the degree of required similarity. The result is dramatic, as there are now only two hypernodes, H1 and H2, c.f. 4 hypernodes in Figure G-1. This is because a group factor of 1 requires complete similarity, whereas group factor of 0.8 allows a lesser degree of similarity to form a group.

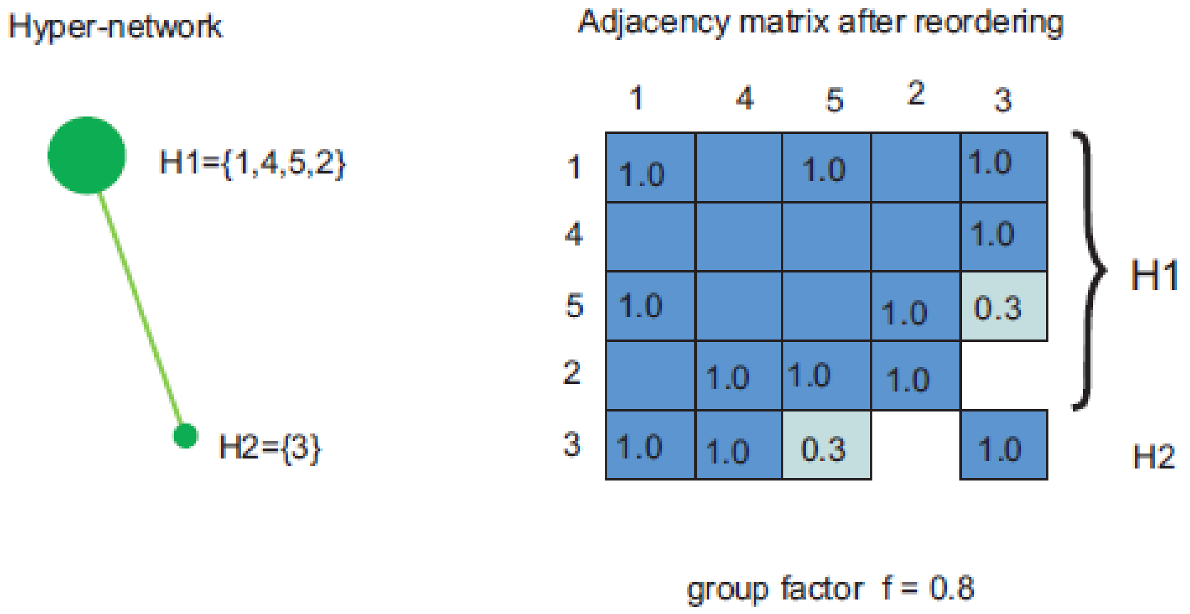


Figure G-2: The Hypernode Results from the Weighted Network in Figure G-1 with Group Factor *f = 0.8.*

The above examples show how the ‘weight’ (uncertainty) and group factor govern the resultant hypernode structure.

G.3 VISUALIZATION AND REPRESENTATION OF PROHIBITED OR UNLIKELY LINKS

In the above example a value 0 is assigned to a cell, which is then coloured in white, when there is no link between two nodes. There is no prior knowledge regarding the reason that there is no link and there is nothing to preclude there being a link between them. There are, however, cases where some links are highly unlikely or forbidden. Therefore, there is a need to differentiate and manage the information about links that just do not happen to exist and links that are prohibited.

There are many different ways that this can be addressed; one way is through the use of prior beliefs (e.g. that a link cannot exist or is highly unlikely) alongside measurements, for example, as in Bayes’ theorem [9]. In the example network shown in Figure G-3 there are no links between Nodes 2 and 3 and Nodes 3 and 4. Let it be assumed that there is *a priori* knowledge that it is impossible or prohibited (or at least highly unlikely) that Nodes 3 and 4 should be connected.

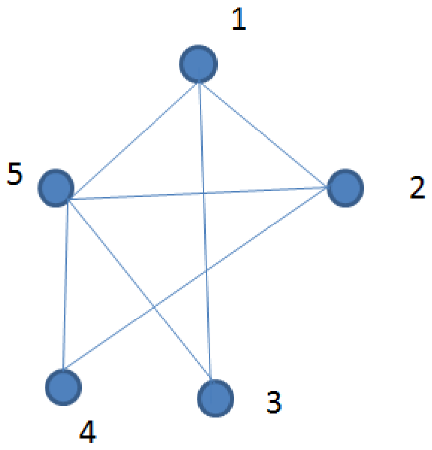


Figure G-3: Network with No Link Between Nodes 2 and 3 and Nodes 2 and 5.

Let it be defined that:

x = 1 means there is a link

x = 0 means the link is prohibited or highly unlikely and

y = 1 means a link is observed

y = 0 means a link is not observed

If we assign some reasonable prior probability measures as follows:

$p(x=1) = 0.9$

$p(x=0) = 0.1$

$p(y=0|x=1) = 0.1$

Then applying Bayes' Theorem $P(x|y) = P(y|x)P(x)/\Sigma\, xp(y|x)p(x)$ to create *posterior* probability values we get:

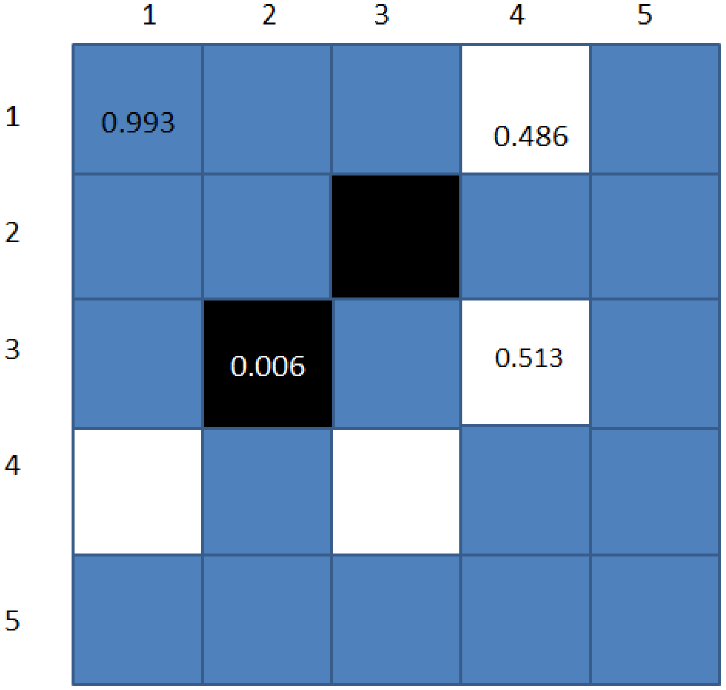
$P(x=1|y=1) = 0.993$

$p(x=1|y=0) = 0.486$

$p(x=0|y=0) = 0.513$

$p(x=0|y=1) = 0.006$

In addition, when a link is highly unlikely the corresponding square in the adjacency matrix, Figure G-4, can be coloured black (Nodes 2 and 3) and white when a link may exist but is not observed (Nodes 3 and 4). (Note: Node 1 is just a normal node).



	1	2	3	4	5
1	🟦 0.993	🟦	🟦	⬜ 0.486	🟦
2	🟦	🟦	⬛	🟦	🟦
3	🟦	⬛ 0.006	🟦	⬜ 0.513	🟦
4	⬜	🟦	⬜	🟦	🟦
5	🟦	🟦	🟦	🟦	🟦

Figure G-4: Adjacency Matrix with Prohibited Links.

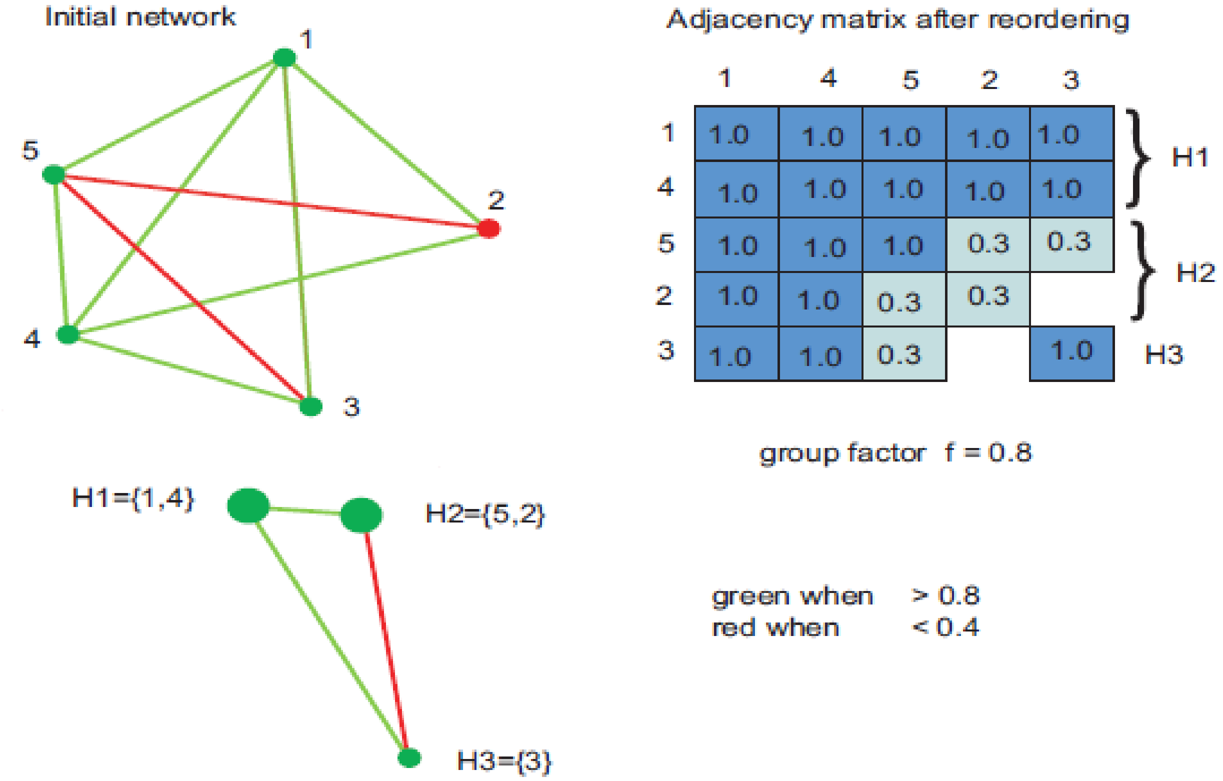
In this way we can extend the Hypernode algorithm using different link measures.

G.3.1 Propagation of Edge Uncertainties

This section will discuss the propagation of uncertainties of the edges in a network. Many possible approaches can be used to address this problem, but among them a possible strategy is to follow the concept of how the usual union operator in fuzzy set theory is constructed. Here, the maximum value is selected [13]. This means that when computing the strength of edge $r_{H1,H2}$, in Figure G-2, for example, one should select the strongest link that connects a sub-node in hypernode $H1$ to a sub-node of hypernode $H2$. Node 4 is a member of $H1$, Node 3 is a member of $H2$ and the link between Nodes 3 and 4 has membership valueμ $(r_{3,4})$ = 1.0, as shown in Figure G-2. Therefore, the edge $r_{H1,H2}$ has membership value 1.0.

G.3.2 Propagation of Node Uncertainties

In a network, uncertainty applies as much to the nodes as to the edges. Figure G-5 shows a network with an uncertain node, i.e. Node 2 which is represented in red. The nodes are represented by the diagonal of the adjacency matrix. The uncertainty of a node can be related to the knowledge about the existence or reliability of a node. In the case considered in Figure G-5, there are two red links in the network, i.e. uncertain links and one uncertain node, i.e. Node 2. The resulting reordered adjacency matrix and the corresponding network of hypernodes is shown at the lower part of the figure. Here, the edge $r_{H2,H3}$, for example, is red since there exists no strong links between the sub-nodes of $H2$ and $H3$. A property of the algorithm is that the uncertain node is mapped to hypernode $H2$. Since $H2$ is composed of a certain and an uncertain node, the maximum membership principle leads to the result that hypernode H2 is a certain node, despite containing the uncertain Node 2.



	1	4	5	2	3	
1	1.0	1.0	1.0	1.0	1.0	H1
4	1.0	1.0	1.0	1.0	1.0	
5	1.0	1.0	1.0	0.3	0.3	H2
2	1.0	1.0	0.3	0.3		
3	1.0	1.0	0.3		1.0	H3

Figure G-5: The Weighted Network in Figure G-2 is Modified by Introducing Edge $r_{2,5}$ as an Uncertain Edge and Node 2 as an Uncertain Node. The membership values are shown in the adjacency matrix. The group fatore f is 0.8.

G.4 AN EXAMPLE

In Figure G-6 a network with 88 nodes is shown. A traffic light scheme is used to represent the network. The large number of edges makes it impossible to understand and analyze the network structure, its characteristics or pattern. After four iterations the initial clustered 88 nodes network becomes a network with 10 hypernodes, see Figure G-7, where the visual separation of the components of the network is clear and well presented (note that the iterative process stops when there are no further nodes that can be grouped together for the give group factor criterion). Figure G-7 shows a network structure which is very significantly clearer than the original highly cluttered network. It is easy to observe that the network is made up of a single connected component, i.e. there are no separate networks, etc.

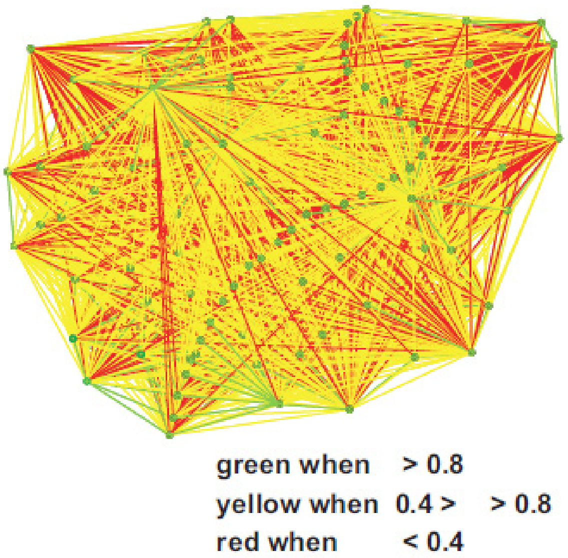


Figure G-6: Initial Network with 88 Nodes, Green Represents Strong or Certain Link, Yellow Represents Medium Strength Link While Red Means a Weak or Uncertain Link.

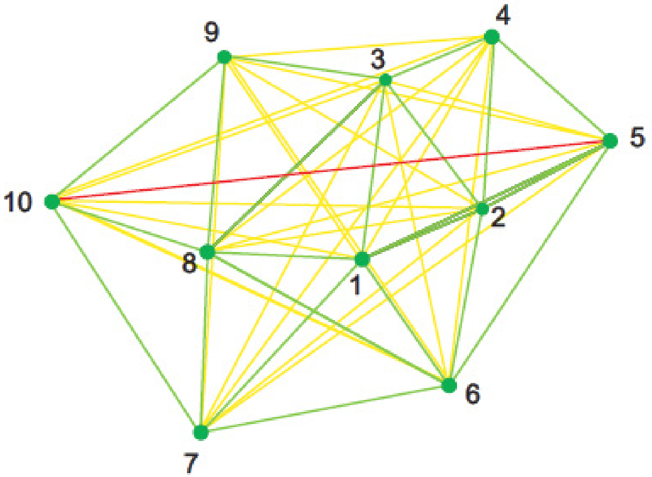


Figure G-7: Hypernode at Level 4.

The group factor, *f,* thus controls the clustering of the nodes. In order to demonstrate the effect of the group factor, hierarchies of hypernodes can be computed for different values of *f*. The result of this computation is shown in Figure G-8 which shows how the number of levels (iterations) and the number of hypernodes at the top level of the hierarchy depend on *f*. When *f* is greater than 0.9 the number of point clusters at the top level increases rapidly with increasing value of *f*. When *f* is less than 0.81, the hypernode at the top level contains all the nodes of the original network. The number of levels of the hierarchy reaches a maximum when $f = 0.87$. It can be seen that the group factor is a case and application dependent parameter which plays a significant role in the determination of the resulting hypernode structure.

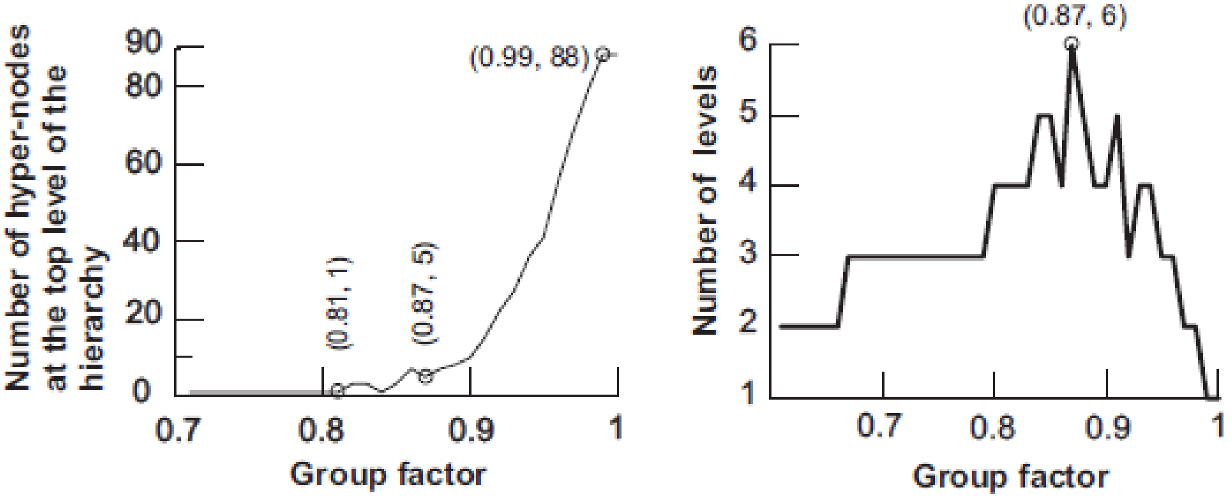


Figure G-8: Hypernodes Computed for Different Group-Factors of the Network in Figure G-6.

It is also possible to analyze specific properties of a network, for example, by 'lifting' the certain and uncertain parts from the hypernode network as shown in Figure G-9. This provides an effective means of studying and understands the uncertainties in the network. In Figure G-9A it can be seen that there are many strong and certain links in this 10 hypernode network, and by comparing Figure G-9A with Figure G-9B it can be seen that the medium strength links are more connected than the strong links. However, as shown in Figure G-9C, there is only one uncertain link between Hypernodes 5 and 10. It is also possible to zoom in and out (to different iteration levels) to get the details and higher levels of information about the network. Furthermore, the traffic light scheme provides an intuitive means of extracting information from the network. Alternatively, it is possible to visualise the separation in a layered or hierarchical network, see Figure G-10 [9].

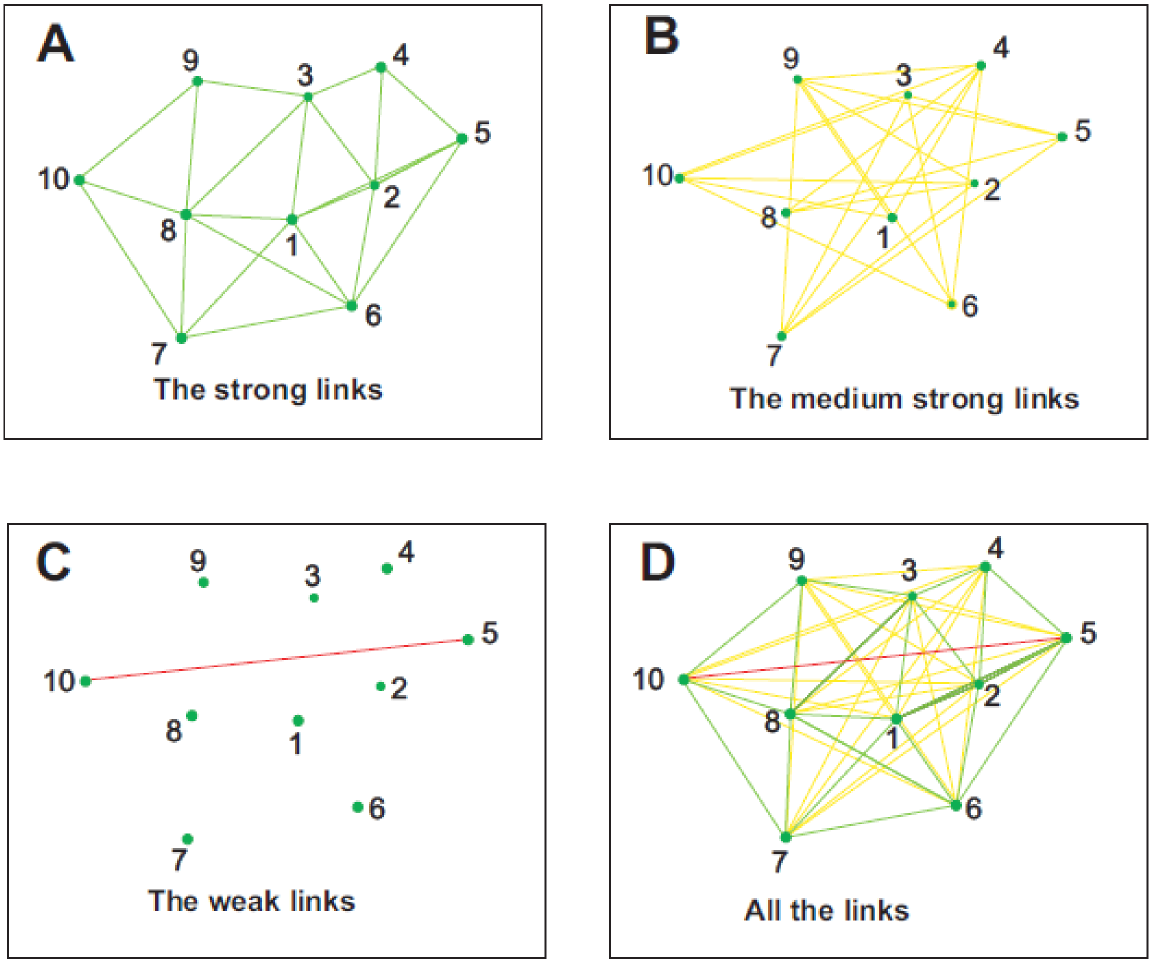


Figure G-9: The Network of Hypernodes in Figure G-6 is Divided into Three Separate Displays, i.e. One for Each Class of Links.

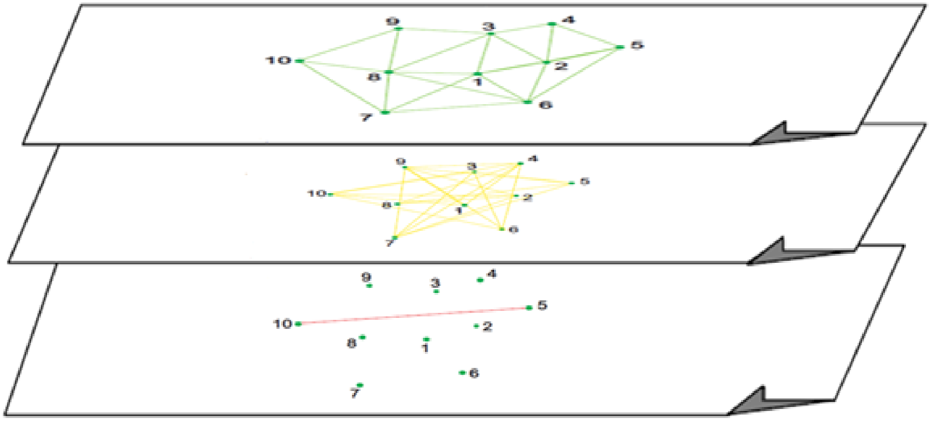


Figure G-10: An Example Layered Network.

G.5 CONCLUSIONS

The application of Hypernode algorithm to weighted networks has been described and demonstrated. The algorithm discussed considers the strength of the relationship; it can be used to construct networks of hypernodes of weighted (or uncertain) as well as crisp relations. The crisp case comes out as the special case where the membership values are either 0 or 1. The introduction of membership values allows uncertainty in network relationships to be studied.

Hypernodes can be constructed at different levels of abstraction, i.e. the degree of generalization or abstraction increases with the number of iterations the algorithm. The Hypernode iteration process terminates either when there is one single hypernode left at a particular level or when the selected group factor does not allow further grouping of the nodes. The mapping of the nodes to hypernodes is a many to one mapping, i.e. one sub-node can be mapped to only one hypernode at a certain level.

The algorithm allows networks to be studied at different levels of abstraction and provides a formalism in which the effects of uncertainty can be managed, analyzed and understood. Indeed the Hypernode algorithm transforms a flat network into a hierarchical structure that reveals the underlying structure and pattern of the network in an effective and intuitive manner. The use of hypernodes supports users in obtaining a very high level understanding of a network's characteristics.

G.6 REFERENCES

[1] Berge, C. (1973). *Graphs and Hypergraphs.* North-Holland, Amsterdam-London.

[2] Berge, C. (1989). *Hypergraphs: Combinatorics in Finite Sets.* Elsevier, New York.

[3] Bertin, J. (1981). *Graphics and Graphic Information Processing.* Walter de Gruyter, Berlin, New York,. 273 pages.

[4] Bjørke, J.T. (†) (1996). Framework for entropy-based map evaluation. *Cartography and Geographic Information Systems,* 23(2):78-95.

[5] Bjørke, J.T. (†) and Smith, B. (1996). Seriation: An implementation and case study. *Computers, Environments and Urban Systems,* 20(6): 427-438.

[6] Bjørke, J.T. (†) (2006). Reduction of complexity: an aspect of network visualization. *"RTO-MP-IST-063 Visualising network information, NATO conference, Copenhagen, Denmark",* 9-1 to 9-10,. NATO. ISBNs 92-837-1156-4 or 978-92-837-1156-8.

[7] Bjørke, J.T. (†) and Varga, M.J. (2007). Algorithm to construct fuzzy hyper-networks. NATO IST-059/ RTG-025 Network of Experts (N/X) Workshop: Network Analysis and Visualisation for Simulation and Prediction, El Segundo, CA, USA.

[8] Bjørke, J.T. (†) and Varga, M.J. (2008). Weight (uncertainty) propagation in networks of hypernodes, NATO Network-of-Experts Workshop on Visualisation of Network Dynamics, UK, November.

[9] Bjørke, J.T. (†) and Varga, M.J. (2010). Chapter 4 – Hypernodes: Their application to visualization of networks, A Framework for Network Visualisation: NATO Final Report of IST-059, RTO-TR-IST-059 AC/323(IST-059) TP/262, February, ISBN 978-92-837-0084-5.

[10] Bjørke, J.T. (†), Nilsen, S. and Varga, M.J. (2010). Visualization of network structure by the application of hypernodes, International Journal of Approximate Reasoning 51, 275-293.

[11] Borkowski, A., Grabska, E. and Szuba, J. (1999). Visualisation of Graphs in ArchiCAD, Lecture Notes in Computer Science, Vol. 1774, 241-246.

[12] Fabrikant, S.I., Montello, D.R., Ruocco, M. and Middleton, R.S. (2004). The distance similarity metaphor in network-display spatializations. *"Cartography and Geographic Information Science,"* 31(4):237-252.

[13] Klir, G.J. and Yuan, B. (1995). Fuzzy sets and fuzzy logic: Theory and applications. Prentice Hall P T R, New Jersey, 574 pages.

[14] MacEachren, A.M. (1994). Visualization in modern cartography: Setting the agenda. In MacEachren, A.M. and Taylor, D.R. Fraser, editors, Visualization in modern cartography, 1-12. Pergamon, Oxford.

[15] Pedrycz, W. and Gomide, F. (1998). An introduction to fuzzy sets: Analysis and design. The MIT Press, Cambridge, Massachusetts, 465 pages.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	STO-TR-IST-085 AC/323(IST-085)TP/561	ISBN 978-92-837-0211-5	UNCLASSIFIED/ UNLIMITED

5. Originator

Science and Technology Organization
North Atlantic Treaty Organization
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title

Interactive Visualization of Network Dynamics

7. Presented at/Sponsored by

Final Report of Task Group IST-085.

8. Author(s)/Editor(s)	9. Date
Multiple	June 2014

10. Author's/Editor's Address	11. Pages
Multiple	218

12. Distribution Statement

There are no restrictions on the distribution of this document. Information about the availability of this and other STO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Cognitive analysis
Framework
Hypernodes
Information analysis
Network analysis
Network dynamics
Networks
Visualization

14. Abstract

For decision-making in the areas of: network discovery; simulation and prediction supporting adaptive operations, political effects, public health and safety, and security issues; and uncertain environments and abstract concepts, one needs to understand how to visualise the changes taking place within a network (dynamics) and the trends within that change. Group objectives included to compare the utility of various interactive visualisation styles for providing the user knowledge of the dynamics of a network and subsequent trends and to produce a report highlighting interactive visualisation methods that facilitate and make more effective the analysis of network dynamics in applications such as netcentric warfare, counterterrorism including bioterrorism, peacekeeping, public security, and peace support operations. Anticipated security benefits include a better understanding of how interactive visualisation should be used to discover, simulate and predict network dynamics, and how such interactive visualisation may aid military command decision-making, public health and security operations as well as intelligence network analysis tasks.

NORTH ATLANTIC TREATY ORGANIZATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cso.nato.int

SCIENCE AND TECHNOLOGY ORGANIZATION

DIFFUSION DES PUBLICATIONS
STO NON CLASSIFIEES

Les publications de l'AGARD, de la RTO et de la STO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la STO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la STO, de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre est la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la STO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (http://www.sto.nato.int/) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
for Defence, National STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

CANADA
DSIGRD2 – Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9e étage
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESPAGNE
SDGTECIN (DGAM)
C/ Arturo Soria 289
Madrid 28033

ESTONIE
Estonian Ministry of Defence
Estonian National Coordinator for NATO STO
Sakala 1
Tallinn 15094

ETATS-UNIS
Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc - BP 72
92322 Châtillon Cedex

GRECE (Correspondant)
Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE
Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25, H-1885 Budapest

ITALIE
Centro Gestione Conoscenza
Secretariat General of Defence
National Armaments Directorate
Via XX Settembre 123/A, 00187 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research
Establishment, Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

PAYS-BAS
Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE
Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

REPUBLIQUE TCHEQUE
Vojenský technický ústav s.p.
CZ Distribution Information
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

ROUMANIE
Romanian National Distribution
Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353 Bucharest

ROYAUME-UNI
Dstl Knowledge and Information
Services
Building 247
Porton Down, Salisbury SP4 0JQ

SLOVAQUIE
Akadémia ozbrojených síl gen.
M.R. Štefánika, Distribučné a
informačné stredisko STO
Demänová 393
031 06 Liptovský Mikuláš 6

SLOVENIE
Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
Başkanlığı
06650 Bakanliklar – Ankara

AGENCES DE VENTE

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2
CANADA

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications STO, RTO et AGARD figurent dans le « NTIS Publications Database » (http://www.ntis.gov).

NORTH ATLANTIC TREATY ORGANIZATION

NATO
OTAN

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cso.nato.int

SCIENCE AND TECHNOLOGY ORGANIZATION

S&T organization

DISTRIBUTION OF UNCLASSIFIED STO PUBLICATIONS

AGARD, RTO & STO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all STO reports, or just those relating to one or more specific STO Panels, they may be willing to include you (or your Organisation) in their distribution.

STO, RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of STO reports as they are published, please visit our website (http://www.sto.nato.int/) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research for Defence, National STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30
1000 Brussels

CANADA
DRDKIM2 – Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street, 9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

DENMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESTONIA
Estonian Ministry of Defence
Estonian National Coordinator for NATO STO
Sakala 1, Tallinn 15094

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc - BP 72
92322 Châtillon Cedex

GERMANY
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GREECE (Point of Contact)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY
Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25, H-1885 Budapest

ITALY
Centro Gestione Conoscenza
Secretariat General of Defence
National Armaments Directorate
Via XX Settembre 123/A, 00187 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY
Norwegian Defence Research Establishment, Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

POLAND
Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

ROMANIA
Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6, 061353 Bucharest

SLOVAKIA
Akadémia ozbrojených síl gen M.R. Štefánika, Distribučné a informačné stredisko STO
Demänová 393
031 06 Liptovský Mikuláš 6

SLOVENIA
Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

SPAIN
SDGTECIN (DGAM)
C/ Arturo Soria 289
Madrid 28033

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Knowledge and Information Services
Building 247
Porton Down, Salisbury SP4 0JQ

UNITED STATES
Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

SALES AGENCIES

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2
CANADA

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of STO, RTO and AGARD publications are given in "NTIS Publications Database" (http://www.ntis.gov).

ISBN 978-92-837-0211-5