

Impact of Information Loss and Information Error on Network-enabled Decision-making

**Michael K. Martin, Geoffrey P. Morgan,
Kenny Joseph, and Kathleen M. Carley**

September, 2010
CMU-ISR-10-127

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213



Center for the Computational Analysis of Social and Organizational Systems
CASOS technical report.

This work was supported by the Network Science Collaborative Technology Alliance: No. W911NF08R0013 from the Army Research Lab, and by the NSF IGERT 9972762 in CASOS. Additional support was provided by the center for Computational Analysis of Social and Organizational Systems and the Institute for Software Research at Carnegie Mellon University. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Lab, the National Science Foundation, or the U.S. government.

Keywords: Multi-agent simulation, virtual experiment, decision-making, network-enabled organization, information networks, communication networks, socio-cognitive networks, Construct, network science, dynamic network analysis, network metrics, robustness, uncertainty, information error, communications loss, network representation, representation of uncertainty

Abstract

This report describes a preliminary Virtual Experiment (VE) utilizing multi-agent simulation that explores how information loss and information error impact decision accuracy in organizations. Results indicate that information loss and error exert interactive influences on decision accuracy. Moreover, the pattern of interaction suggests that information loss due to break-downs in communication networks does not adversely affect decision-making accuracy in situations characterized by moderate degrees of error in information networks. The results of our VE and the trends observed in our exploratory analyses imply that a line of research concerned with controlling access to information in a manner that depends on its veracity can potentially improve organizational decision-making. Put differently, given estimates of information error, communication networks and socio-cognitive network topology can potentially be used as “throttles” to control access in a manner that improves organizational performance. Along these lines, we include an appendix that describes our initial approach to the representation and measurement of error in communication and information networks.

Table of Contents

1	Introduction	1
1.1	Motivating Scenarios.....	2
1.1.1	Scenario One – Military Patrols in Iraq	2
1.1.2	Scenario Two – Deepwater Horizon oil spill (pre-explosion)	4
1.1.3	Summary	6
1.2	Network Interactions	6
2	Method.....	7
2.1	Virtual Scenario.....	7
2.2	Experimental Design	8
2.3	Performance Variable.....	8
2.4	Operational Definitions	9
3	Results	11
4	Discussion.....	12
5	Appendix: Representing and Measuring Information Loss and Error	18
5.1	Theoretical Perspective: Dynamic Network Analysis	18
5.2	DyNetML Network Representation Language	19
5.3	Communication Loss.....	20
5.4	Information Error	20
5.5	Conclusion.....	21
6	References	21

1 Introduction

The general goal of the Army's Network Science Collaborative Technology Alliance (NS-CTA) is to enhance human performance in network-enabled organizations by identifying and exploiting synergies in research performed by scientists who study social/cognitive, information, and communication networks. Accordingly, researchers in the NS-CTA have defined socio-technical systems in the battlespace as being comprised of three interacting genres of networks, communication networks, information networks, and socio-cognitive networks.

Our research for the NS-CTA is concerned with how information loss in communication networks and information errors in information networks impact decision makers in socio-cognitive networks. Information loss in communication networks occurs when one or more nodes or links in physical communication networks disappear (e.g., offline servers or cell phone towers). The resulting functional impact on decision-makers is a decrease in information accessibility. Information error in information networks arises from occurrences of missing or extraneous nodes or links within the networks of concepts formed by an organization's databases, websites, reports, etc. Information error may also arise from missing, extraneous or incorrect attributes that are used to describe the nodes and links within such networks (e.g., weighting, labeling, or directionality of links). The functional impact of information error on decision-makers is a decrease in the veracity of information on which decisions are based. These errors tend to propagate across levels, as seen in our high level concept diagram, Figure 1, below.

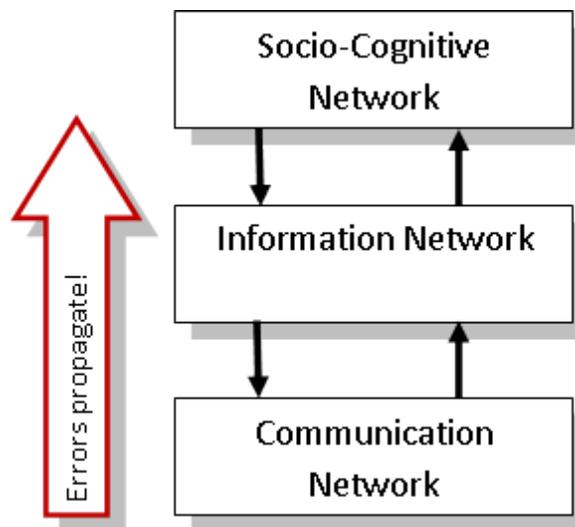


Figure 1. Three genres of interacting battle-space networks

Although our research topic implies that three mutually exclusive networks operate in the battlespace, it would be a mistake to characterize socio-cognitive networks as a particular “level” of this multi-level framework. Rather, it is more productive to see the socio-cognitive networks that exist in this system as a flexible web that constrains and enables communication networks (e.g., helping govern when, how and for what purpose specific communication devices and approaches are used) and information networks (e.g., helping establish the salience and useful

interpretation under differing circumstances by various groups and individuals). Socio-cognitive networks can also be usefully seen as a dynamic set of pipes that enables decisions to be made and activity to continue despite uncertainty caused by information loss at the communication layer and errors at the information level.

1.1 Motivating Scenarios

The following scenarios are designed to convey issues with information loss and error that may develop in the real world, however, the scenarios themselves are entirely fictional. While they may be loosely based on facts and/or real-world events, there is no intention of modeling a specific real-world situation in which these events have actually played out. Certain concepts, such as procedures for oil rig wellhead shutdown, may very well be incorrect.

1.1.1 Scenario One – Military Patrols in Iraq

Socio-cognitive Level: Platoon leader A, Platoon leader B, C2 Dispatcher 1, C2 Dispatcher 2, IED Recon Team

Information Level: Information on C2 Server

Communication Level: Radio, GPS, C2 Server

At 0100, Platoon leader C completes his debriefing of his patrol of Transport Route 5, a road commonly used to transport materials to and from base. He indicates that there was no movement in the area. While eating a meal in the mess hall, Platoon leader C sits down with Platoon leader A, who patrols the same route starting at 0300. Leader C comments on the fact that he hasn't seen anything interesting on Route 5 in weeks. Nonetheless, Platoon leader A notices some suspicious activity under an underpass during his patrol some time around 0600. He immediately radios in to Command and Control, giving his location and details of the activity. C2 Dispatcher 1 takes the call and logs the information. Dispatcher 1 confirms the location of Platoon leader A via GPS linked to his humvee. In addition, he enters Platoon leader A's visual description of his location. Platoon leader A is then given permission to investigate the activity further, but finds nothing of interest by the time he assembles his team and reaches the area below the underpass.

At 1200, Platoon leader B is taking his team through the same area along Transport Route 5, also on a routine patrol. Platoon leader B notices an irregular lump of dirt on the left side of the road. He radios in to C2, gives his location along Transport Route 5, adding that the suspicious mound is located underneath an overpass. C2 Dispatcher 2, who takes the call, confirms the location of Platoon leader B and his team via GPS, and discovers that logs from earlier in the morning that morning indicate suspicious activity underneath an overpass in the same area. In fact, Dispatcher 2 had already been made aware of a possible situation in this area, as addition, Dispatchers 1 and 2, because they work in the same room, had been discussing events that occurred on Dispatcher 1's shift that morning. Dispatcher 1 indicated that he thought "something fishy" might be going on near an underpass under Route 5. Dispatcher 2 therefore had direct knowledge of the previous issues in the area because of his face-to-face discussion with Dispatcher 1. Dispatcher 2 orders Platoon leader B to cordon off the area, and sends an IED Recon Team in to search the area for explosives. Two IEDs are recovered later that afternoon from the area of interest.

<i>Possible Sources of Loss/Error</i>	<i>Network Involved</i>		
	<i>Communication</i>	<i>Information</i>	<i>Socio-cognitive</i>
<u>Human Error</u>			
-Platoon leader A fails to report suspicious activity			*
-Platoon leaders A or B neglect to visually describe their whereabouts			*
-C2 Dispatcher 1 is the only dispatcher on duty at the time, and forgets to log information given by Platoon leader A			*
<u>Communication Network Error</u>			
-C2 Dispatcher 2 does not have the ability to communicate with any IED Recon Teams	*		
-Platoon leader B cannot give the correct location of his patrol group	*		
-Due to the inability of the server to handle multiple simultaneous requests, C2 Dispatcher 1 is unable to log information from Platoon A's routine patrol until 1300, when a planned mission requiring the entire throughput of the system is finished	*	*	
-Platoon leader A's radio is malfunctioning, and C2 Dispatcher 1 is not able to log the suspicious activity because he cannot get all of the details	*	*	
-C2 Dispatcher 1 cannot determine the exact location of Platoon leader A because GPS is malfunctioning	*		
-Platoon leader B's radio is malfunctioning, so C2 Dispatcher 2 cannot log his account	*		
-C2 Dispatcher 2 is unable to access the system immediately to look for prior reports of the given location	*		
-The C2 Server goes offline for system maintenance when either Dispatcher 1 or Dispatcher 2 is trying to access it	*		
<u>Information Network Error</u>			
-Platoon leaders A and/or B give the wrong location, perhaps indicating the issue is well beyond the overpass. The IED recon team continues past the IED in search of the specified area.		*	*
-Platoon leader B mischaracterizes the mound of dirt as a rock. C2 dispatcher is aware that IEDs are rarely placed under rocks- they are usually much more well disguised. In addition, areas under underpasses are often rocky.		*	*
-C2 Dispatchers 1 and/or 2 enter incorrect information into the system, perhaps indicating that all suspicious activity and objects were confirmed clear. They might also enter the activity as having occurring on the wrong date		*	*
-Dispatcher 1 had indicated that Transport Route 5 sounded "all clear" when talking with Dispatcher 2		*	*
-The C2 database gets corrupted, and the information entered by Dispatcher 1 is changed.		*	
-Platoon leader A listens to Platoon leader C and ignores his patrol duties			*
Correct Decision			
Send IED team to remove explosives from the scene			
Possible Incorrect Decisions			
-Platoon leader B is told to investigate			
-Platoon leader B is told not to investigate, but IED Recon Team is never called			

1.1.2 Scenario Two – Deepwater Horizon oil spill (pre-explosion)

Socio-cognitive Level: Mechanics, Mechanics’ Supervisors, Engineer, Engineer’s Supervisor, Rig Supervisor, All Workers on Rig, All BP Employees

Information Level: BP Internal web server, BP internal mail server, Onboard wireless connections

Communication Level: Radio, Employee blackberries, Wellhead pressure monitor

Throughout the first three weeks of April 2010, mechanics aboard the Deepwater Horizon rig sent several emails to their supervisors denoting strange readings stemming from a wellhead that was being constructed for use onboard. These supervisors, in turn, forwarded these emails to the chief supervisor of the rig, who in turn sent his own report to head BP officials and engineers. The internal BP web server was updated to reflect these issues so that engineers could continue to collaborate on what they might mean.

On the morning of April 20th, an engineer monitoring the pressure systems on one of the wellheads aboard the rig noticed a steep rise in pressure. The engineer quickly radioed his supervisor, explaining the steep jump and requested a protocol for action. The engineer’s supervisor quickly logged onto BP’s internal server to check the latest developments in the ongoing research about the bulkhead. The pressure value that the engineer gave lay above the “safety threshold” that had been established by a BP engineer working out of an office in New Orleans, indicating that a complete shutdown of the wellhead might be necessary. The engineering supervisor quickly emailed the rig supervisor, who had a reputation amongst rig workers as being constantly connected to his email. In fact, all rig workers were required to have company phones that were used to send out emergency alerts and other information, in addition to having push-to-talk capabilities necessary for rig operation. The engineering supervisor needed the approval of the rig supervisor to shut down the wellhead. Upon seeing the engineering supervisor’s email, the rig supervisor, who had years of experience on the rig, knew that disaster might be imminent. He quickly gave permission, via email, to the engineering supervisor to shut down the wellhead. The engineering supervisor then radioed to his engineer to shut down the wellhead. As a preventative measure, an email was also sent to all workers on the rig to stay clear of the area where the wellhead was located, and to prepare for emergency evacuation pending the sound of the rig-wide evacuation alarm.

<i>Activity</i>	<i>Network Involved</i>		
	<i>Communication (Info Loss Issues)</i>	<i>Information (Info Error Issues)</i>	<i>Socio-cognitive (Decision-making)</i>
<u>Communication Network Error</u>			
<i>Communication Technology Errors</i>			
-Any one of the many emails sent throughout the scenario was ignored	*		*
-Internal BP server did not allow the engineering supervisor to access the Deepwater Horizon page as BP was only leasing the rig, the workers were not all BP employees.	*		*
-Any one of the many emails sent throughout the scenario was not sent immediately, or at all, due to misconfiguration or overload of the internal company mail server	*	*	*
-The engineering supervisor could not access the internal company server detailing protocol for the malfunctioning bulkhead	*		
-The pressure monitor for the bulkhead was not functioning so there were no pressure readouts		*	
-The engineer monitoring the pressure system was unable to radio his supervisor due to issues with the radio	*		
-The wireless aboard the ship was broken, or too slow to access BP's server in time	*		
<i>Face-to-Face Access (including worker beliefs of the network structure)</i>			
-Any one of the many emails sent throughout the scenario were never sent, due to a lower-level worker's belief that, given the organization's social structure, no one would ever read it anyway			*
-The engineer that calculated the threshold for pressure for the bulkhead was not included in the study of the wellhead		*	*
-The onboard engineer monitoring the pressure system did not notify his supervisor, instead assuming that the dangers of the wellhead were already known, and that they were being monitored at higher levels of the organizational chain		*	*
<u>Information Network Error</u>			
<i>Report Errors</i>			
-The engineering supervisor misread the data on the web			*
-Engineers that calculated the "safety threshold" for the pressure of the wellhead were incorrect			*
-The engineer monitoring the pressure of the wellhead misread the value		*	*
<i>System Errors</i>			
-The system monitoring the pressure monitor for the wellhead was buggy, and therefore gave incorrect readings when the pressure breached a certain value.		*	
<i>"Rumor" Errors</i>			
- One of the emails sent in the scenario contained incorrect information about the wellhead, that propagated up the chain of command. For example, engineering supervisor, when contacting the rig supervisor, might misconstrue the situation as being "nothing to worry about".		*	*
-The rig supervisor wasn't actually always connected to his email. He would eventually have had to make a decision on whether or not to shut down the rig. Without seeing an email of the current readings, he may have had to make that decision with incorrect and/or outdated.	*		
Correct Decision			
-Shut down the wellhead			
Possible Incorrect Decisions			
-First-level mechanics and engineers choose not to pass on information about issues with the wellhead			
-Supervisors try to mediate the problem by withholding information from superiors in a "wait and see" type situation			

1.1.3 Summary

The goal of these scenarios is to show how much interaction may occur between socio-cognitive agents (i.e. decision makers) that require the use of the information and communication layers to make the correct decision. In addition, an attempt is made to show that an agent's perception of the network, and his face-to-face interactions, may also play a role in his ability to make decisions.

1.2 Network Interactions

In addition to governing the usage of the communication and information networks, socio-cognitive networks also play their own role in facilitating information access. Decision-makers have two general methods for accessing information in organizations. The first method, electronically mediated access, requires the infrastructure provided by communications networks. In this vein, communication networks are particularly important for decision-makers who require access to mass media sources (e.g., organizational websites or databases; email or radio broadcasts from superiors), and for socio-cognitive agents who use electronically mediated point-to-point communications (e.g., phone, email and texting). However, communication networks are not required for the second method of information access – face-to-face communication with co-workers. Because this process exists solely within the socio-cognitive networks of the multi-level framework, this can be used as further justification that the system presented above cannot be considered three mutually exclusive networks performing mutually exclusive tasks.

Furthermore, the point-to-point access of information known by co-workers requires more than communication networks or shared locations- it requires knowing who to contact. In practice, such knowledge will tend to be a combination of knowledge about formal organizational structure and roles, and informal knowledge of what one's social network of acquaintances and peers know (i.e., transactive memory), neither of which can be represented in the communication or information networks. In contrast, socio-cognitive networks represent these two types of knowledge. More specifically, socio-cognitive networks are multimodal, multiplex networks of decision-making agents and their knowledge.

As described above, socio-cognitive networks play an important role in point-to-point information access. The pattern of links among nodes in socio-cognitive networks (i.e., topology), along with one's position in such networks, determines the ease or difficulty of directly and indirectly accessing information. Thus, information access is constrained and enabled by socio-cognitive networks, regardless of the presence or absence of communication networks.

Logically, one would assume that information loss in communication networks should tend to decrease the accuracy of time-sensitive decisions in evolving situations. Decisions that require information from authoritative sources (e.g., C2 nodes) or mass media sources would be particularly vulnerable to fluctuations in information accessibility. Situation updates, for example, may be unobtainable when communication networks are inoperable. This forces decision-makers to rely on information that is potentially out-dated, erroneous or incomplete. The degree and frequency of information loss in communication networks, therefore, should be

key accessibility characteristics that can degrade decision accuracy; the degradation should be especially prominent for those decisions that require authoritative, time-sensitive information.

Face-to-face communications may obviate the detrimental effects of information loss in communication networks on decision accuracy. However, the efficacy of such communications will depend on at least two other characteristics of the decision-making situation. The first concerns the topology of decision-makers' socio-cognitive network, because it determines the point-to-point accessibility of information. The second concerns the veracity of local information sources (e.g., local co-workers' knowledge). When communication networks are down, gaps and errors in the information used for decisions can be corrected only to the degree that locally accessible information accurately augments information already known by decision-makers.

Locally accessible information is a product of information diffusion processes that take place between nodes within socio-cognitive networks (i.e., point-to-point information exchange), and between nodes in socio-cognitive networks and information networks (e.g., use of authoritative information sources). Consequently, the obviating potential of locally accessible information depends to some extent on the prevalence of error in information networks. Incomplete, erroneous, or out-dated information obtained from information networks by decision-makers will tend to diffuse throughout an organization via point-to-point exchange processes (e.g., spread of rumors). Socio-cognitive network topology will influence the extent to which information error is propagated through these secondary, point-to-point channels; socio-cognitive network size and the density of links will influence the extent of propagation. The prevalence of information error in information networks, therefore, is a key characteristic of information networks that can degrade decision accuracy – by decreasing the veracity of diffused information.

The primary purpose of the Virtual Experiment (VE) reported in this paper is to explore how interactions among these key characteristics, described above, of information loss and information error impact decision-making accuracy.

2 Method

We conducted a Virtual Experiment (VE) using Agent-Based Simulation (ABS), cognitive agents (i.e., agents with knowledge representation and decision-making capabilities), and network-constrained information diffusion processes.

The ABS model of information diffusion employed is called Construct (Carley, 19xx). In Construct, cognitive agents probabilistically interact to exchange information. The probability of interaction depends on homophily in terms of the knowledge shared between pairs of agents, expertise seeking in terms of knowledge not shared between pairs of agents, and whether a tie exists between two agents in a social network.

2.1 Virtual Scenario

Our VE scenario required each agent in an organization to make 13 decisions. Six items of information were associated with each decision. Three items positively influenced decision accuracy (i.e., provided accurate information) and three negatively influenced decision accuracy (i.e., provided inaccurate information). According to the decision model used, correct decisions required at least one of the positive information items.

At the start of each simulation, agents' prior knowledge of decision-relevant information varied randomly. During the simulation, agents had 365 time periods to obtain additional decision-relevant information. Additional information could be obtained by interacting with other agents or three types of mass media sources: radio, organizational website, or email broadcasts.

For each simulation, agents were randomly placed in 1 of 12 organizations, which were formed by varying the size, topology, and density of social networks. The organizations were either small or large, employing 500 or 2000 agents, respectively. The topology of ties among agents within each organization – hence the availability of information from other agents – was structured as being random, scale-free, or cellular. The density of ties among agents within each organization was either 0.005 or 0.05.

The availability and veracity of information from mass media sources (radio, website, email) were varied exogenously due to the lack of communication and information network models. Thus, communication networks correspond to the physical access points of actors to special media sources, such as web sites and radio. Information networks correspond to the type and quality of the information available from these media sources, such as web sites and radio channels. Socio-cognitive networks correspond to the people to people connections within the environment, and the connections between each person-agent and its knowledge.

2.2 Experimental Design

Our experimental design was a three-factor complete factorial that manipulated Communication Network Intermittency, Communication Network Accessibility, and Degree of Information Error. The first two factors are associated the availability of communication networks, which connect agents to mass media sources of information. The third factor is associated with how much of the information provided by mass media sources is misleading, not helpful, or missing (cf. negatively influences decision accuracy).

We used four levels of each experimental factor (0, 33, 67, 100) to create a 4 x 4 x 4 factorial design. The four levels of Communication Network Intermittency correspond probabilistically to the percentage of time periods during which mass media sources were unavailable (cf. frequency of communication network outages). The four levels of Communication Network Accessibility correspond probabilistically to the percentage of agents within an organization that could not interact with mass media sources (cf. extent of information loss due to communication network outages). The four levels of Degree of Information Error correspond probabilistically to the percentage of erroneous/misleading information provided by mass media sources.

2.3 Performance Variable

At a basic level, we are interested in how information loss and error influence decision accuracy in a given organizational context. Thus, the dependent variable of interest is ostensibly the proportion of correct decisions out of the 13 decisions made by each agent. However, using agents as our unit of analysis does not take into account the dependencies created by the size, topology, and density of the organizations in which agents were placed. Therefore, we defined organizations as our basic unit of analysis, and used the mean proportion of correct decisions per

agent within organizations as our dependent variable (i.e., the mean proportion of correct decisions per simulation).

Table 1. Communication and Information Network Error Virtual Experiment

Factor	# of Combinations	Combinations
Socio-cognitive Network Characteristics		
Topology	3	ER Random, Scale-Free, Cellular
Size	2	500, 2000
Density	2	.005, .05
Communication Loss Characteristics		
Intermittency of Access to Mass Media Sources	4	0%,33%,67%,100%
Extent of Access to Mass Media Sources	4	0%,33%,67%,100%
Information Loss Characteristics		
Information Error	4	0%,33%,67%,100%
Controlled Variables		
Actor Perception	1	Perfect
Knowledge Network Distribution	1	Random
Knowledge Network Size	1	Constant
Belief Network Distribution	1	Random
Belief Network Size	1	Constant

Total Combinations: $(3 * 2 * 2 * 4 * 4 * 4) = 768$ (but topology, size, density are treated as random variables for this VE)

2.4 Operational Definitions

- Socio-cognitive network – Stylized networks drawn from the *ORA stylized network generator - these agent x agent networks are defined by the combination of topology, density, and size of the networks.
- Information network -- exogenous component of our simulation, modeled by parameters of special agents (see Mass Media Information Sources).
- Communication network – exogenous component of our simulation, modeled by parameters of special agents (see Mass Media Information Sources).
- Decision-maker – a node in the socio-cognitive (i.e., Agent x Agent) network
- Mass Media Information Sources – special agents including radio, mail (same as e-mail), and website. Special agents are used to exogenously model communication and information networks in our simulation model
- Information Loss in Communication Networks

- Intermittency – proportion of time periods in which special agents are inactive.
 - Defined negatively, how often are the special agents available to interact with?
 - So, Total Intermittency means that the special agents are not available after the first time-step at all.
 - High Intermittency means that the special agents are not available more often than not. (suggest we go with ~66%, meaning that special agents are active only every third time-period)
 - Low intermittency means that the special agents are more available than not. Suggest we go with ~33%, meaning special agents are active 2 out of every 3 time periods
 - When information is not intermittent, it means it's always available. (i.e., 0%)
- Degree – proportion of agents that special agents can interact with during a single time period. (I'm unsure how to scale this for time course of the simulation.)
- Information Error in Information Networks
 - Amount – proportion of incorrect/misleading knowledge facts transmitted by special agents.
 - Of the information the special agents possess (from Information Completion above), how much of it is accurate/correct in relation to the 13 decision tasks? Total accuracy means that all information provided is correct. The inaccuracy condition implies that of the information provided, some fixed percentage of the information is wrong.
 - Suggest proportions of 0, .33, .66, 1
- Socio-cognitive Network Size – the number of decision-makers in the socio-cognitive network.
- Socio-cognitive Network Density – the level of interconnectedness between decision-makers – for the Cellular topology, this will be a close approximation through tuning the “intra-cell” and “extra-cell” densities.
- Interaction Sphere. The Information Network is operationalized to refer to special agents that provide information to the ‘regular’ agents. We’re measuring Socio-Cognitive network accuracy through accuracy on the classification-choice decision tasks.
- Actor Perception – The agents will have zero-error Transactive Memory – they know what they know, and have accurate perceptions of what their interaction partners know.
- Knowledge Network Distribution – We will use a uniform random distribution to seed the people to knowledge network. This distribution will randomize every run.
- Knowledge Network Size – 500.
- Belief Network Distribution – We will use a uniform random distribution to seed the people to belief network. This distribution will randomize every run.
- Belief Network Size – This is linked to the existing IRS definitions.

3 Results

Data from 875 simulations were used to calculate our measure of decision accuracy, operationalized as the mean proportion of correct decisions per agent for each simulation run. Decision accuracy served as the dependent variable in a 4 (Communication Network Intermittency) x 4 (Communication Network Accessibility) x 4 (Degree of Information Error) ANOVA. The ANOVA yielded a main effect of Degree of Information Error, $F(3, 63) = 95.18, p < .05$. Degree of Information Error also interacted with Communication Network Intermittency, $F(9, 63) = 10.01, p < .05$. No other effects were reliable.

As can be seen in Figure 3, agents tended to make accurate decisions approximately 2/3 of the time. Moreover, decision accuracy decreased as Degree of Information Error increased – unless the communication network was down for the entire simulation (i.e., Communication Network Intermittency = 1).

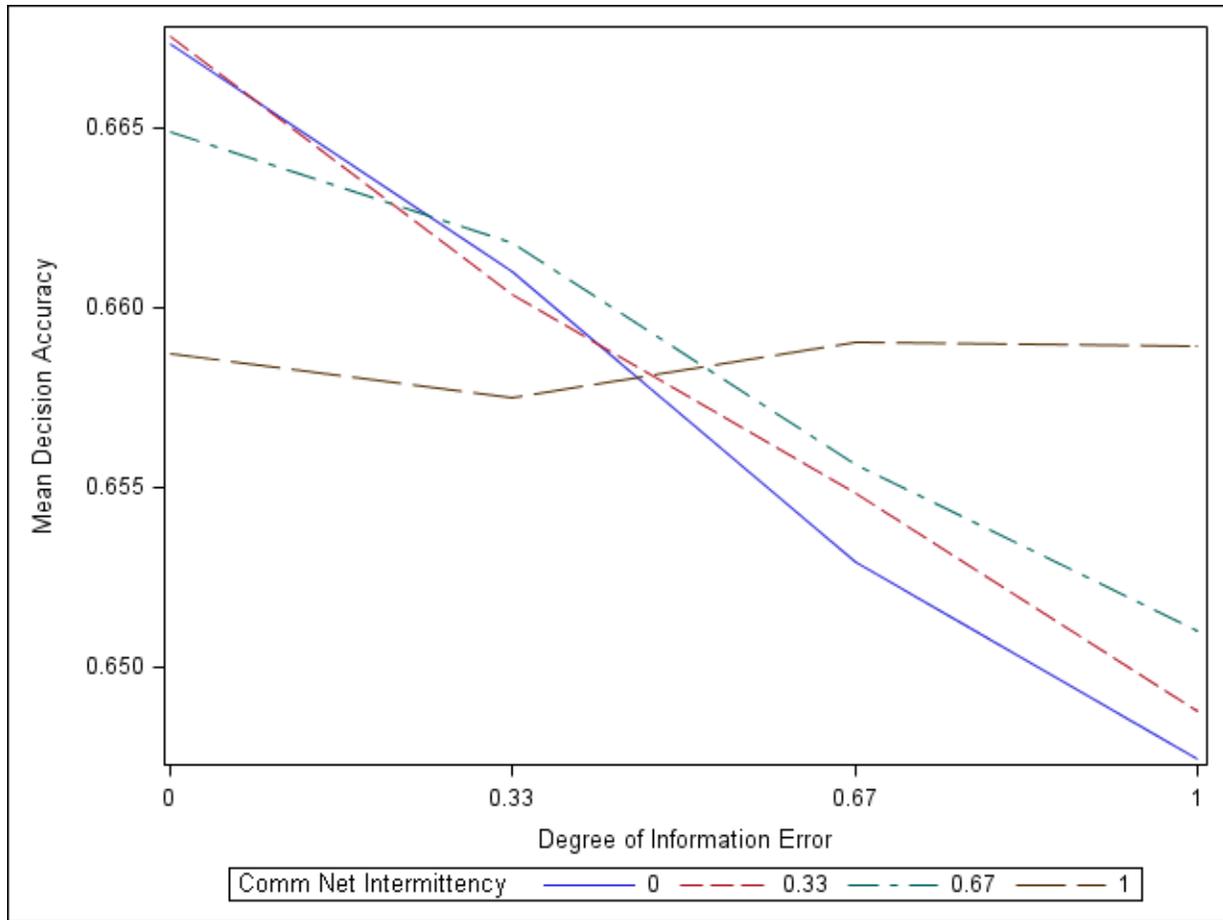


Figure 2. Decision Accuracy varies as a Function of Communication Network Intermittency and Degree of Information Error.

Closer inspection of the Degree of Information Error x Communication Network Intermittency interaction indicates that Communication Network Intermittency affected decision accuracy only when the Degree of Information Error was extreme. When all information provided by mass media sources was accurate (i.e., Degree of Information Error = 0), decision accuracy increased

in presence of a communication network relative to the absence of a communication network (i.e., Communication Network Intermittency = 1). Conversely, when all information provided by mass media sources was inaccurate (i.e., Degree of Information Error = 1), decision accuracy decreased in presence of a communication network relative to the absence of a communication network (i.e., Communication Network Intermittency = 1). The effects of Communication Network Intermittency for moderate Degrees of Information Error were unreliable.

4 Discussion

On the one hand, our initial VE demonstrated a relatively straightforward dependence of decision accuracy on information loss in communication networks and information error in information networks (cf. the Communication Network Intermittency x Degree of Information Error interaction). Communication networks foster the diffusion of information regardless of whether the information conveyed over them is correct. But we observed that the impact of communication networks on decision accuracy is substantive only in those situations characterized by extreme accuracy or inaccuracy in the information conveyed.

On the other hand, our initial results also imply that in some situations organizations may be able to tolerate – or even benefit from – information loss due to problems with communication networks. In situations involving moderate Degrees of Information Error, decision accuracy was not reliably affected by Communication Network Intermittency. In situations where information error was rampant (i.e., Degree of Information Error = 1), the absence of communication networks improved decision accuracy by interfering with the diffusion of incorrect information. Thus an interesting line of future research would more carefully explore how tolerance for information error varies as a function of situational and socio-technical system characteristics.

Follow-on research should continue to explore how interactions among the three network genres impact decision accuracy. A logical next VE would partially replicate and extend the VE discussed in this report by manipulating organizational parameters (i.e., topology, density, and size of the socio-cognitive network), along with the communication and information network parameters manipulated in our preliminary study. If the finding of an apparent “tolerance” for information loss given moderate degrees of information error is found again, subsequent studies can be designed to more carefully delineate the boundaries of information loss tolerance, and whether it is influenced by the type of information error (e.g., misinformation versus missing information). It may also be of interest to examine whether information loss and error differentially impact decisions of varying complexity. Finally, future research should begin to address how different information seeking policies influence decision accuracy in different organizations experiencing different degrees of information loss/error.

To generate additional hypotheses for follow-on research we examined patterns between organizational/socio-cognitive network parameters (i.e., topology, size, and density) and information loss and error. In Figure 3 below, we plotted decision accuracy as a function of Communication Network Intermittency and socio-cognitive network topology. This plot averages over Degree of Information Error and thus represents trends in the effects of information loss and socio-cognitive network topology given that the Degree of Information Error is moderate on average.

As can be seen, decision accuracy appears to be lowest for socio-cognitive networks with the most realistic topology. That is, in situations characterized on average by moderate degrees of information error, decision accuracy is lowest in cellular networks regardless of the information access constraints that stem from communication networks (i.e., Communication Network Intermittency). Furthermore, the inverted u-shaped pattern of decision accuracy in cellular networks implies the possibility of optimizing organizational decision-making performance by balancing the increased rate of information diffusion made possible by communication networks and the decreased rate of information diffusion between cells due to the “gate-keepers” who connect the various cells in an organization.

Put differently, cellular networks tend to produce a group-think phenomenon in which cell members’ knowledge and decisions (right or wrong) grow more similar over time. Communication networks can break this group-think phenomenon by providing access to the extra-cellular information typically obtained only through gate-keepers. But the increased access to extra-cellular information is a double-edged sword because communication networks facilitate access to information regardless of its accuracy. Thus, unfettered access to moderately erroneous information can be just as detrimental to average decision-making performance as the group-think that occurs in situations where gate-keepers provide the only access to extra-cellular information. Communication networks, therefore, have the potential of being used as a “throttle” that improves organizational decision-making by controlling the accessibility of extra-cellular information in a manner that is sensitive to estimates of the prevalence of information error.

The pattern of decision-making accuracy in scale-free socio-cognitive networks resembles that seen for cellular networks, but with slightly better accuracy. Barabosi and Albert (1999) postulated that scale-free networks arise in the real-world (e.g., the world-wide web) because preferential attachment mechanisms create small sets of highly connected “hubs” in otherwise sparsely connected networks. Agents positioned as hubs in scale-free socio-cognitive networks have privileged, highly influential opinions because of their many connections to the sparsely connected majority. More specifically, the power-law distribution of degree centrality in scale-free networks creates fault-tolerant hierarchies in which major hubs are followed closely by smaller hubs, which are followed by other nodes of smaller degree. In such networks, random losses of access to mass media information sources (cf., Communication Network Intermittency) are more likely to impact the many low-degree nodes than the few high-degree hubs. Thus, the majority of agents in scale-free networks can continue to access decision-relevant information (by way of hubs) despite communication network problems. This higher rate of information diffusion from hubs regardless of communication network functionality appears to increase the mean decision-accuracy for agents in scale-free networks relative to agents in cellular networks. Furthermore, the advantage of scale-free networks over cellular networks appears to increase when information access is unfettered by communication network intermittency.

Although scale-free socio-cognitive networks appear to offer a performance advantage over cellular networks, the notion of scale-free networks of human decision-makers is problematic. Socio-technical systems that rely on scale-free organizations of human decision-makers are unlikely to exhibit expected performance levels in the real world for two simple reasons. Attentional limitations and time constraints will impose upper limits on the amount of information that can be transmitted by a human hub. Attentional limitations in human decision-

makers will impose upper limits on the number of relations that can be managed effectively. Time constraints will limit the number of interactions available for the exchange of information. Consequently, socio-cognitive networks nominally designed to have a power-law distribution of degree centrality may functionally exhibit the performance characteristics of some other network topology. In future explorations of the effects of topology on decision-making, it may be fruitful to work toward examining how blends of well-known network topologies (e.g., cellular and scale-free) impact performance.

Figure 3 also shows a counterintuitive increase in decision accuracy as information loss increases for the least realistic socio-cognitive network topology – the Erdos-Renyi topology in which socio-cognitive nodes are randomly connected. The improvement in performance as information loss increases again indicates the potential of using communication networks to improve decision-making by controlling access to moderately erroneous information. In the absence of communication networks (i.e., Communication Network Intermittency = 1) random networks perform better than cellular and scale-free networks because the time required for any one agent to reach any other agent is on average shortest in random networks, and no agent's opinion is privileged. Consequently agents in random networks churn information through the group more quickly than do agents in cellular or scale-free networks, and on average more agents make better decisions.

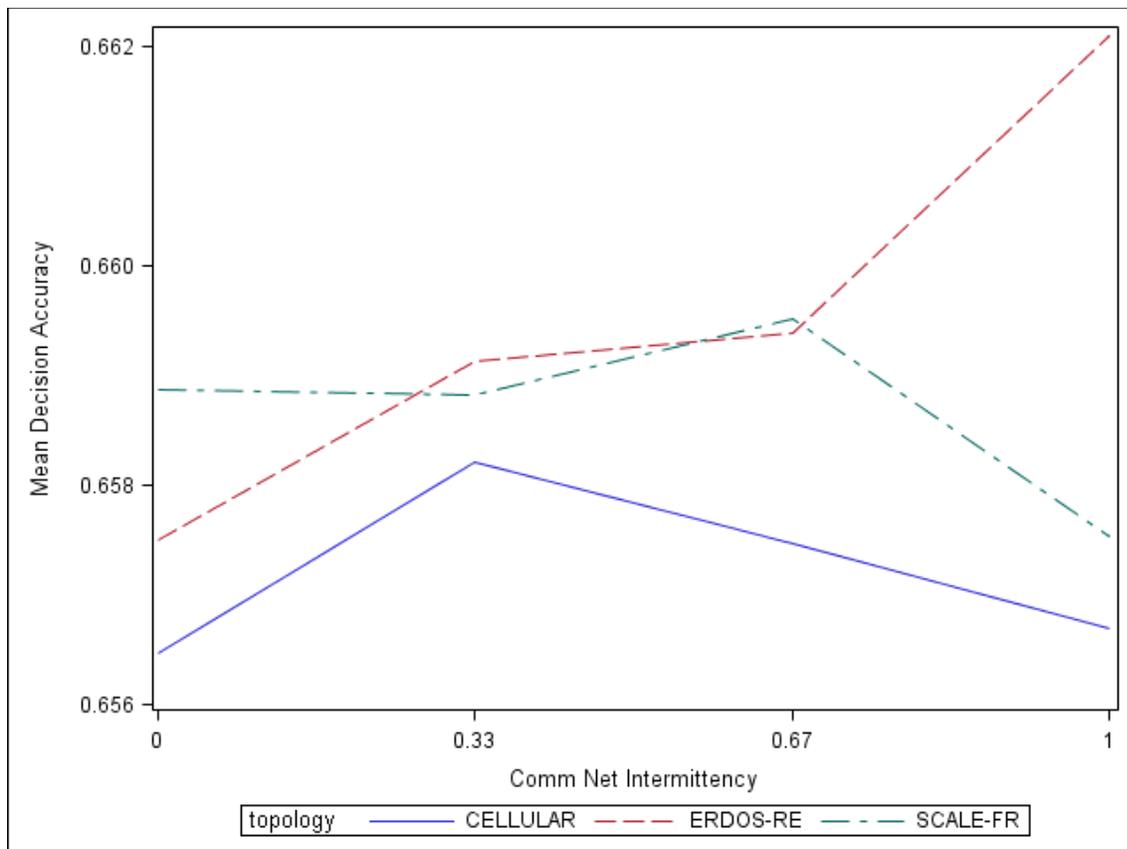


Figure 3. Communication Network Intermittency may Interact with Socio-cognitive Network Topology to influence Decision Accuracy.

In Figure 4 below is a plot of decision accuracy as a function of Degree of Information Error and socio-cognitive network topology. This plot averages over Communication Network Intermittency and thus represents decision accuracy given a moderate degree of information loss. The observed pattern implies that the effect of information error on decision accuracy is independent of socio-cognitive network topology.

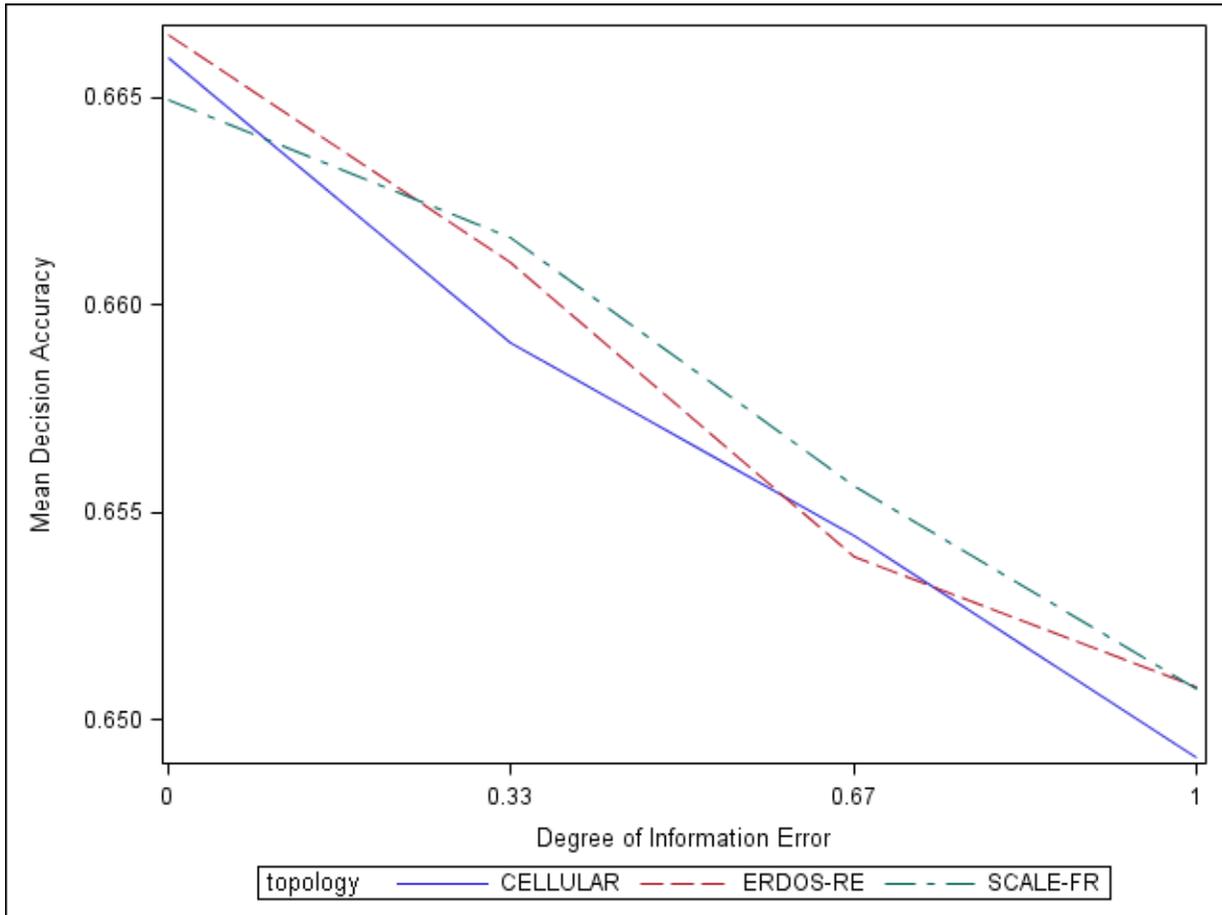


Figure 4. Effect of Information Error Degree on Decision Accuracy may vary Independently of Effects of Socio-cognitive Network Topology.

Figure 5 below plots decision accuracy as a function of information error and socio-cognitive network size, again averaging over information loss. The observed pattern implies that larger organizations may be more tolerant of the detrimental effect of information error on decision accuracy than smaller organizations.

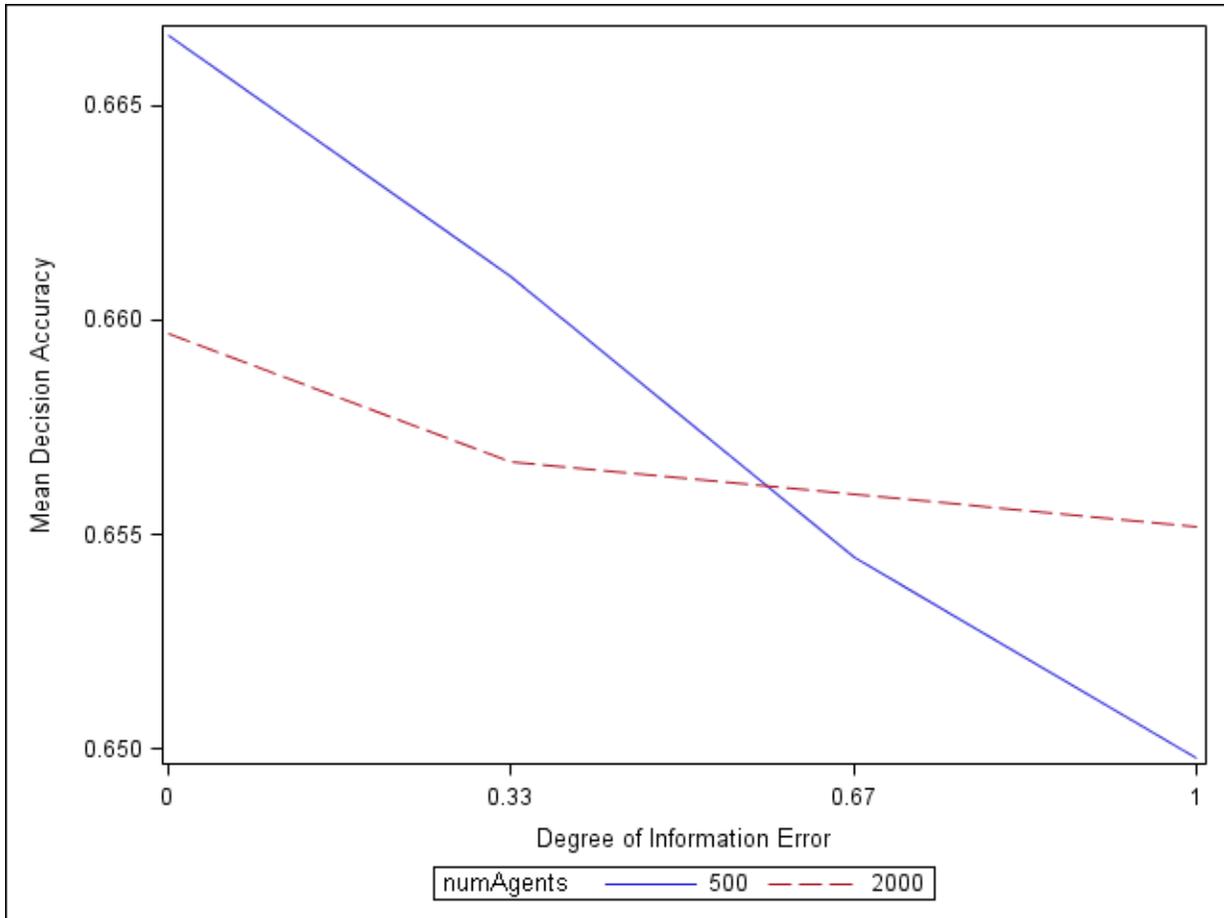


Figure 5. Degree of Information Error may Interact with Socio-cognitive Network Size to Influence Decision Accuracy.

In Figure 6 below is a plot of decision accuracy as a function of information error and socio-cognitive network density; it averages over information loss.

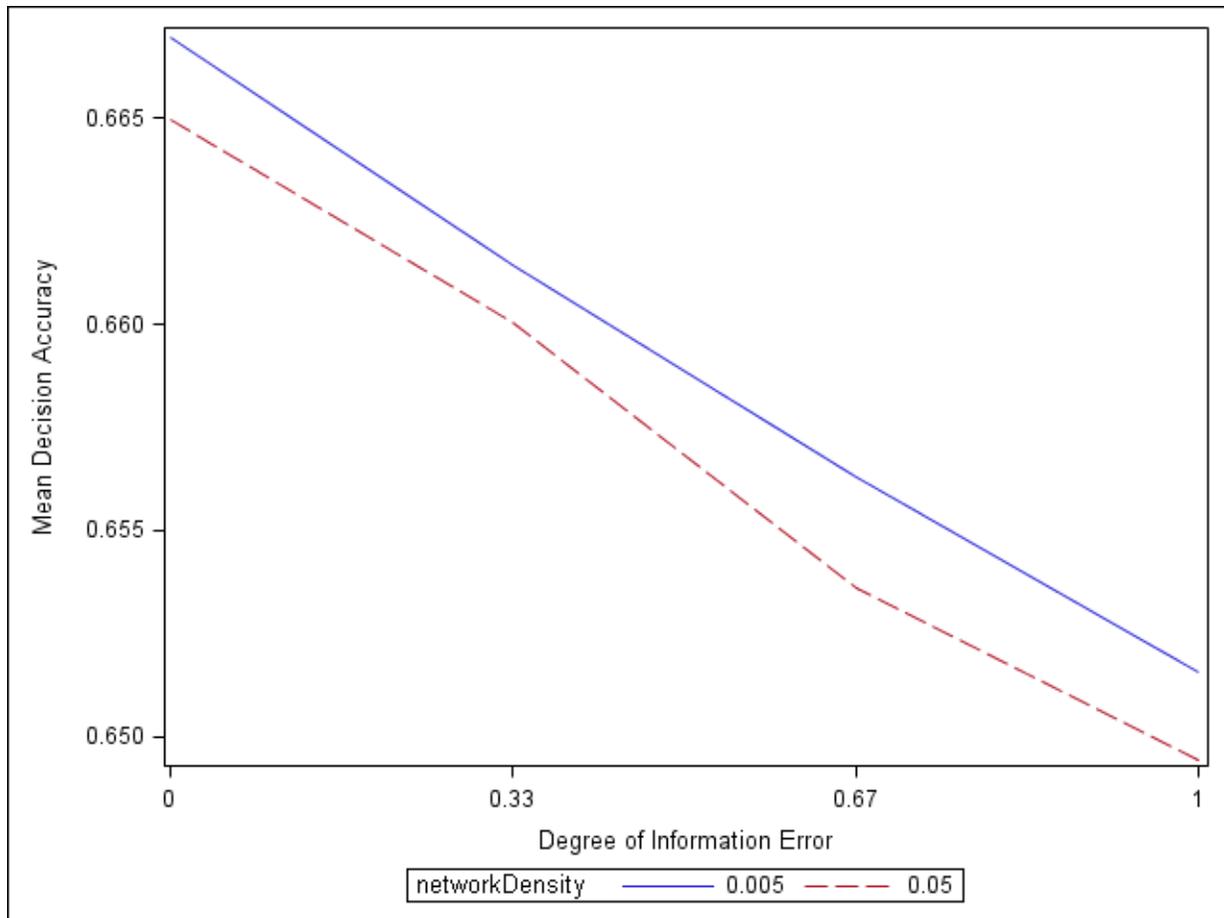


Figure 6. Decreased Socio-cognitive Network Density may Ameliorate Influence of Information Error Degree on Decision Accuracy.

Taken together, the results of our initial VE and the trends observed in our exploratory analyses imply that a line of research concerned with controlling access to information in a manner that depends on its veracity can potentially improve organizational decision-making. Communication networks provide one method for controlling access; socio-cognitive networks provide another. More specifically, a fruitful line of research will explore organizational decision-making by varying temporal and structural characteristics of both genres of networks – and the interactions among such characteristics –in situations that vary temporal and structural characteristics of information error.

Pursuing research along these lines will eventually require procedures for measuring and representing information loss and error in real-world networks, along with understanding their impact on the network metrics that warfighters will use to assess their situational understanding. In this vein, Appendix A describes our preliminary approach to the representation of errors in communication and information networks, and initial procedures for measuring error.

5 Appendix: Representing and Measuring Information Loss and Error

In this appendix, we begin to address the fundamental research question of how missing or incorrect network data affect the metrics used to describe and reason about multi-modal, multiplex networks. In particular, we provide an initial description of communication loss and information error; the representation of errors in communication and information networks, and procedures for measuring error.

5.1 Theoretical Perspective: Dynamic Network Analysis

We approach the representation and measurement of information loss and error from a Dynamic Network Analysis (DNA) framework (Carley, 2002). Complex social systems are often represented as dynamic networks that relate entities in a system (e.g., people, knowledge, actions) to each other over time. The emergent field of DNA centers on the collection, analysis, understanding and prediction of dynamic relations in such networks, and the enabling and constraining impacts of network structure on individual and group behavior. DNA facilitates reasoning about real groups as complex dynamic systems that evolve over time as boundedly rational agents learn and adapt. Within this field computational techniques such as machine learning and artificial intelligence, are combined with traditional graph and social network theory, and empirical research on and theories about human behavior, groups, organizations, and societies.

At the heart of DNA is the meta-network. A meta-network is a multi-mode, multi-link, multi-level network, often referred to as a network of networks (Krackhardt & Carley, 1998; Carley, 2002a). Meta-networks place the social network of who knows/works with/has some relationship with whom in an ecology of networks covering the linkages among who, what, where, how, why and when. More generally, the nodes classes and the links can come from anywhere in the socio-technical domain.

The types of nodes and links in a meta-network may vary with application; for example, we might use these types of nodes - people, organizations, resources such as databases and servers, actions, locations - to capture information on an international firm, and the links might be based on data extracted from newspapers (Carley, 2006). In another application, we might add cell-phones and supplement open-source information sensor information for defining the links.

A dynamic meta-network thus represents real-world socio-technical systems in which the state of the nodes and links, and even which nodes (and so links) exist changes with time. Faults in socio-technical systems appear as missing or extraneous nodes and links (or attributes thereof) which may or may not vary over time. Just how these faults ultimately impact warfighters' understanding of the battlespace is currently unknown. Using agent-based dynamic network simulations the impact of interventions on these dynamic meta-networks (Carley, 2002b, Carley 2006) and the impact of rumors and errors on trust (Prietula and Carley, 2001) and performance can be assessed (Lin and Carley, 1997).

What we do know is that network metrics are not robust to error (Borgatti, Krackhardt & Carley, 2006). For example, the presence of missing nodes has been shown to severely impact

betweenness centrality accuracy in scale-free networks, whereas superfluous nodes have a high impact on local clustering metrics in small world networks (Frantz, Cataldo, & Carley, 2009). Furthermore, we can logically presume that node-related error will be more serious than link-related error because node-related error necessarily includes link-related error.

False information can lead to or mitigate conflict (Jong, 2009; Carley & Prietula, 2009). Previous work on this topic includes efforts to make socio-cognitive networks robust to error by propagating information about how long a piece of information has been in the network (Rosvall and Sneppen, 2006). One interesting result from this work is that “hubs” in scale-free networks tend to accumulate and propagate larger amounts of false information than nodes in random (Erdős–Rényi) networks. However, this model of reducing the cost of information error does not take into account the needs of real-time decision making or a way to efficiently reduce the spread of information errors throughout the network. A large amount of work has also been done in how to detect information intended to deceive others in the network (Sabater & Sierra, 2002; Pradchayakool, Daengdej & Koolmanojwong, 2006; Koolanojwong et al.). This measure requires information about how “trustworthy” an agent is. Therefore, it is indicative of the maliciousness or lack of perception of an agent, and not necessarily indicative of the types of error that our research is focused on. Borgatti, Carley and Krackhardt (2006) suggests that, even with erroneous data in the network, centrality measures can still be somewhat indicative of the true measures of the actual network up to a given threshold. In addition, the work suggests that the accuracy of predicting these measures decreases systematically with known increases in error. This may indicate that, by predicting the extent to which you can “believe” the current system-view of the meta-network, you may be able to determine the rigor with which you search for erroneous information. The concept of allowing for some error has been studied by Olson and Carley (2010) in terms of latent space models – this work may help to assess the likelihood of error in a network.

Representing and measuring errors in multi-level meta-networks is described below.

5.2 DyNetML Network Representation Language

DyNetML is an expressive, extensible, open-source, XML-derived language for representing relational data consisting of multiple node and link types (e.g., the DNA ontology known as the meta-network model; see Tsvetovat, Reminga, & Carley, 2004). DyNetML provides a means to represent socio-technical systems as meta-networks using data obtained from one or more sources with characteristics such as reliability and pedigree, variable delays in positive identification of unknown entities (nodes), and nodes and edges that may appear, disappear, or change in strength over time as new information arrives. The resulting meta-networks include different types of nodes (i.e., multi-modal), multiple types of links among nodes (i.e., multi-link) and multiple types of networks (i.e., multi-level). Multiple link-types may exist between any pair of nodes, and multiple types of networks may occur because of variations in the granularity in collection of data such as that associated with individuals or groups.

DyNetML is attribute-enhanced network representation language. Thus, it provides a means to associate attributes with nodes and edges. Node attributes often represent details about the state of a node such as level of education, age, and so on. Link attributes often include details about the strength, direction, and reliability of a link. DyNetML also allows one to associate meta-data

attributes with networks. These meta-data attributes often include information about who generated the data, why, when, how they generated the data, coding constraints, information pedigree, credibility, level of classification, etc.

Given a network representation language such as DyNetML, we can define two general types of error. First, there may be missing or superfluous nodes or links. Split nodes are a special case of superfluous node-error in which a node is inadvertently included in a meta-network two or more times. This situation often arises when data are collected for situations in which a person uses his or her real name sometimes and an alias at other times. A second general type of error occurs in cases where there are missing or superfluous attributes on nodes or links. Weights on directional links could be missing or reduced, for example.

5.3 Communication Loss

Links or nodes in communication networks disappear for a variety of reasons. The partial loss of communication networks decreases bandwidth. Error in communication networks is always a loss. Thus the best way to deal with errors in communication networks is to represent information loss as a percentage reduction from one time period to the next. Shannon's information entropy metric provides an example of a metric that may prove useful.

Of the three types of networks under consideration in the NS-CTA, communication networks will be the easiest to describe. In terms of representation, it is questionable as to whether we must represent the uncertainty associated with particular nodes or links as values in a network representation language.

5.4 Information Error

Links or nodes may disappear in information networks just as they do in communication networks. This amounts to missing information just as in communication networks. But information networks can have several types of error not present in communication networks. First off, extraneous links or nodes may be present. This amounts to superfluous or erroneous information. In addition, the attributes of links or nodes may be missing or extraneous, link weights may be incorrect or link directionality may be incorrect.

In the case of numeric values, we may see an increase or decrease in the represented value on a node or link relative to the real value. For binary values, we may have missing or extra nodes or links in our data relative to the real world.

Information error, in addition to existing inside a meta-network as misinformation about actual people and events, may also persist through the creation of fictitious people and events. These erroneous nodes in the meta-network may develop either accidentally as the result of, for instance, a miscommunication between a soldier in the field and a dispatcher at headquarters, or intentionally as a result of, for example, false information given by a hostile who has infiltrated the network. False information inputted in various forms into different parts of the network may cause poor decisions to be made (e.g., preemptive engagement in a conflict). There is a need to study the spread of false information within a system and a means to efficiently filter it out.

For each of these types of error, we can represent our “confidence” in the accuracy of the represented value as a certainty value. Thus for each network characteristic (e.g., node, link, node attribute, link attribute), we could create triples of the form: network characteristic, value, certainty; where certainty ranges from 0 to 1.

In information networks, there occurs a subtle, different kind of error – the split node. Because two nodes might represent the same entity, we need a procedure for estimating when a split node is present. Ostensibly, a measure of structural equivalence could be used to identify split-nodes. Measures of structural equivalence, however, will work only for simple, obvious cases (e.g, multiple cell phones used for the same task).

Representationally, if one suspects two nodes code the same entity, we could use a meta-node that represents both of the suspected split node members, along with a certainty value that duplicate nodes exist, and certainty values on the membership links between the duplicate nodes and metanode. Notice that this procedure and representational scheme for split nodes could be extended to group membership.

The procedure for calculating information error gets trickier when trying to look at networks as trails through time. What do we do to calculate information error when performing analyses concerned with who was where when? Given the omnipresence of measurement error, we have a probability that a person was in a location and a probability that the person-location relation occurred at a particular time. The question is how to calculate certainty based on the relation of underlying probabilities. The relevant probabilities are not independent, but how do we calculate their dependencies?

5.5 Conclusion

Information uncertainty includes that uncertainty created by loss of information due to breakdowns in the communications layer, information layer errors, or cognitive uncertainty or forgetting. Information uncertainty can be created in many ways such as when a UAV’s camera malfunctions, mis-estimation of information value due to over- or under-trust in the source and the information, and the lack of full information due to incomplete, erroneous, biased, intentionally deceptive, inaccessible (e.g., due to security), or otherwise limited data availability in the information layer. Information uncertainty is likely to impact the network metrics on which future warfighters will base their understanding of the battlespace.

6 References

- Barabasi A. & Albert, R., 1999, “Emergence of scaling in random networks,” In *Science* Volume 286, Pp. 509-512. DOI: 10.1126/science.286.5439.509.
- Borgatti, S., K. Carley, and D. Krackhardt, 2006, “On the Robustness of Centrality Measures under Conditions of Imperfect Data.” *Social Networks* 28.2 (2006): 124-36.
- Carley, Kathleen M. 2002a, “Smart Agents and Organizations of the Future,” *The Handbook of New Media*. Edited by Leah Lievrouw and Sonia Livingstone, Ch. 12, pp. 206-220, Thousand Oaks, CA, Sage.
- Carley, Kathleen M., 2002b, “Computational Organization Science: A New Frontier,” In Arthur M. Sackler Colloquium Series on Adaptive Agents, Intelligence and Emergent Human

- Organization: Capturing Complexity through Agent-Based Modeling October 4-6, 2001; Irvine, CA, National Academy of Sciences Press 99(3): 7257-7262.
- Carley, Kathleen M., 2006, "A Dynamic Network Approach to the Assessment of Terrorist Groups and The Impact of Alternative Courses of Action," In *Visualising Network Information*. Meeting Proceedings RTO-MP-IST-063. Neuilly-sur-Seine, France: RTO, 12 - 14. Available from: http://www.vistg.net/documents/IST063_PreProceedings.pdf.
- Carley, Kathleen M. and Michael Prietula, 2009, Gossip Matters: Destabilization of an organization by Injecting Suspicion. In Alex Kott (Ed.) *Information Warfare and Organizational Decision Making*, 7: 159-190, Artech, Boston.
- Jong, S. 2009, "False information and the emergence of conflict." In the proceedings of BNAIC 2009, Eindhoven, the Netherlands.
- Koolmanojwong, S., Charoenvikrom, S., Moemeng, P., Daengdej, J. "An Approach for Dealing with Lying Agents. In the proceedings of the fourth International Conference on Intelligent Technologies (InTech'03), Chiangmai, Thailand, 2003
- Krackhardt, David and Carley, Kathleen M. (1998). "A PCANS Model of Structure in Organization," in Proceedings of the 1998 International Symposium on Command and Control Research and Technology. Conference held in June. Monterey, CA. Evidence Based Research, Vienna, VA, Pp. 113-119.
- Olson, J. and Carley, K. "Confident Network Indices with Latent Space Models." (December, 2009). Paper Accepted into *Analyzing Networks and Learning in Graphs, NIPS 2009*. Conference December 11, 2009 in Whistler, BC, Canada.
- Pradchayakool, P.; Daengdej, J.; Koolmanojwong, S. "An approach for detecting deception in agents," *Aerospace Conference, 2006 IEEE*, vol., no., pp.10
DOI: 10.1109/AERO.2006.1656066
- Prietula, Michael and Kathleen M. Carley, 1999, "Exploring the Effects of Agent Trust and Benevolence in a Simulated Organizational Task," *Applied Artificial Intelligence*, 13(3): 321-338.
- Rosvall, M., and K. Sneppen. "Self-assembly of Information in Networks." *Europhysics Letters (EPL)* 74.6 (2006): 1109-115.
- Sabater, J. and Sierra, C. 2002. Reputation and social network analysis in multi-agent systems. In *Proceedings of the First international Joint Conference on Autonomous Agents and Multiagent Systems: Part 1* (Bologna, Italy, July 15 - 19, 2002). AAMAS '02. ACM, New York, NY, 475-482. DOI= <http://doi.acm.org/10.1145/544741.54485>
- Tsvetovat M, Reminga J, & Carley K (2004) DyNetML: Interchange Format for Rich Social Network Data. CASOS Technical Report. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, CMU-ISRI-04-105.
- Lin, Zhiang & Kathleen M. Carley, 1997, "Organizational Decision Making and Error in a Dynamic Task Environment," *Journal of Mathematical Sociology*, 22(2): 125-150.
- Prietula, Michael J. and Kathleen M. Carley, 2001, "Boundedly Rational and Emotional Agents: Cooperation, Trust and Rumor," In *Trust and Deception in Virtual Societies*. Edited by Cristiano Castelfranchi and Yao-Hua Tan. Kluwer Academic Publishers. Ch. 9, pp. 169-194.