12-22-2016

An Information Theoretic Investigation Of Complex Adaptive Supply Networks With Organizational Topologies

Joshua V. Rodewald

AN INFORMATION THEORETIC INVESTIGATION OF COMPLEX ADAPTIVE SUPPLY NETWORKS WITH ORGANIZATIONAL TOPOLOGIES

DISSERTATION

Joshua V. Rodewald, MS, Civilian

AFIT-ENV-DS-16-D-029

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-DS-16-D-029

AN INFORMATION THEORETIC INVESTIGATION OF COMPLEX ADAPTIVE SUPPLY NETWORKS WITH ORGANIZATIONAL TOPOLOGIES

DISSERTATION

Presented to the Faculty

Department of Systems Engineering & Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Joshua V. Rodewald, MS, Civilian

December 2016

AFIT-ENV-DS-16-D-029

AN INFORMATION THEORETIC INVESTIGATION OF COMPLEX ADAPTIVE SUPPLY NETWORKS WITH ORGANIZATIONAL TOPOLOGIES

Joshua V. Rodewald, MS, Civilian

Committee Membership:

John M. Colombi, PhD
Chair

Lt Col Kyle F. Oyama, PhD
Member

Alan W. Johnson, PhD
Member

Brent T. Langhals, PhD
Member

ADEDJI B. BADIRU, PhD
Dean, Graduate School of Engineering and Management

AFIT-ENV-DS-16-D-029

Abstract

Supply networks exist throughout society in manufacturing and knowledge-intensive industries as well as many service industries. Organizations have been noted to behave as complex adaptive systems or information supply networks with both formal and informal structures. Thoroughly understanding supply network structure and behavior are critical to managing such organizations effectively, but their properties of complex adaptive systems make them more difficult to analyze and assess, forcing researchers to rely on unrealistic data or assumptions of behavior. This research proposes an information theoretic methodology to discover such complex network structures and dynamics while overcoming the difficulties historically associated with their study. Indeed, this was the first application of an information theoretic methodology as a tool to study complex adaptive supply networks. Moreover, managing these complex networks with formal and informal structures poses additional challenges because the effects of intervention can result in even more unpredictable effects. Noting that two primary functions of organizational networks are to transfer information between nodes and store information in the network, this research quantifies the effects of increased and decreased node performance on the ability of multiple organizational network topologies to accomplish these tasks. Multiple qualitative observations from previous researchers are quantitatively analyzed using information theoretic modeling and simulation. Results show an increased ability in local teams to store information within the network as well as a decreased ability by core-periphery networks to respond to increased information rates.

Acknowledgments

I would like to express my sincerest gratitude to…

my faculty advisor, Dr. John Colombi, for his guidance and patience while an aero engineer and electrical engineer dove into information theory and organization science;

my management who supported this endeavor;

my son who is too young to comprehend how motivating and uplifting he is;

most importantly, my wife for her support and flexibility during the past three years and putting up with, well, me. We can do fun things again!

<div align="center">Joshua V. Rodewald</div>

Table of Contents

List of Figures

ix

List of Tables

AN INFORMATION THEORETIC INVESTIGATION OF COMPLEX ADAPTIVE SUPPLY NETWORKS WITH ORGANIZATIONAL TOPOLOGIES

I. Introduction

Problem Statement

The National Air and Space Intelligence Center (NASIC) Plans and Programs group submitted the following problem statement around 2012:

"A full understanding of the types of data collected and used for intelligence assessments provides optics on the chain of dependencies of the intelligence analytical process. Knowing the sources of collected data and the value of specific data within the analytical process allows more informed decisions on resource prioritization of collection assets and methods. Analyze the data dependencies of a select number of functional areas (e.g. counter-space, ballistic missiles, etc) and determine the value (quantity, quality, timeliness, etc) of collected data."

General Issue

This problem is representative of a larger problem concerning a system's ability to transfer and store information within its network of actors. The actors in the organization use formal and informal networks to coordinate, share, exchange, and access information. [1] These are common activities in organizational networks and will be discussed later. Communication within an organization, especially due to informal communication networks, is very difficult to track, but the informal network contributes significantly to the organization's goals of information transfer and storage. [1] The

1

information dependencies within the organization (i.e., which actors rely on which other actors, both formally and informally), if better understood, could be used by managers of the network to identify opportunities in the network where intervention such as shifting resources or directed collaboration may lead to improved system performance (i.e., increased information transfer or storage).

Although much of this research was performed with the sub-context of an intelligence production network, the fundamental principles may apply to other networks as well. Many military (non-intelligence) operations involve communication networks, and optimizing their architecture to enhance information transfer or storage may yield operational benefits. Swarming networks (both military and civilian) rely on information transfer between entities and storage within individual entities. Public news/media organizations and real estate agencies are additional examples of networks with a goal of transferring information within a network to achieve a goal of quickly disseminating information relevant to the greater public. In any of these examples, better understanding the information transfer and storage behaviors would help direct managers or architects of these networks to improve network performance. [2]

Research Objectives and Questions

Information theory and information dynamics are actively being researched in many fields. However, they have yet to be applied to complex adaptive supply networks (CASNs). Therefore, the primary goal of this dissertation is to propose a methodology for conducting analysis of CASNs using information-theoretic principals and identify key information-theoretic indicators that are useful to understanding and managing these

CASNs. This will provide both researchers and practitioners with another tool with which to study CASNs. Additionally, information theoretic methodologies will be applied to networks with organizational topologies to study their information transfer and information storage properties. This methodology is further developed in three proposed research questions:

1. How can information theory be applied to CASNs to model and analyze the static and dynamic structures of those networks?

2. What are the relationships between network structure and information-theoretic indicators of information transfer and information storage?

3. How do organizational network structures differ in their inherent ability to transfer and store information within the network?

Following the stated goals of this sponsored research project, and proposed research questions, my research hypothesis is: an effective methodology for analyzing CASNs based on information theory and information dynamics can be used to understand and manage these systems.

Methodology

As discussed more thoroughly in the Background sections of the various chapters, CASNs have primarily been studied using either agent-based models or social networking analysis. However, both of these approaches have their inherent drawbacks. Agent-based modeling simulates the underlying processes believed to be responsible for the global pattern. [3] Therefore, this approach requires a thorough knowledge of the underlying processes which can be difficult to grasp in a truly complex network. Social

networking analysis requires a thorough understanding of the underlying network structure. Again, the underlying network structure can be difficult to determine in a complex and dynamic network.

It is from here that inspiration is drawn from other disciplines to tackle the problem of analyzing information flow through a network without prior knowledge of that network structure or behavior. Recognizing that the supply chain can be modeled as a complex adaptive system (CAS) [3], and that organizations also behave as CAS [4] it is possible to look to CAS techniques to analyze these CASNs.

One of the best examples of a CAS is the brain. Researchers in the neuroscience field have been using information theory principles to study the brain for decades. Dimitrov, Lazar, & Victor [5] provide an overview of information theory in neuroscience. Soon after Shannon's "A Mathematical Theory of Communication" in 1948 [6], neural information flow was proposed in 1952 based on information theory. The author notes that interest was high in the topic area until the 1980s during which time the applicability of information theory was hampered by difficulty in measuring and interpreting information-theoretic quantities. This changed in 1988 when researchers provided a straightforward way to estimate information-theoretic quantities and remove biases due to finite sample sizes. Following this, the idea of directed information transfer (transfer entropy) around 2000 [7] set the stage for the current state of information theory application to neuroscience. For example, researchers applied transfer entropy as a measure of information transfer to functional imaging data to determine the direction of information flow between brain regions. In the mid-2000s, researchers gained great interest in the concept that information-theoretic tools could probe the structure of

4

network firing patterns, and they were able to determine structure from information using these tools.

Neuroscience researchers determine structure from information by placing electrodes at various recording sites on the brain. [8] They expose the subject to stimuli and record the frequency output of each of the electrodes in what are referred to as "spike trains". Then the researchers calculate the transfer entropy between each pair of recording sites, determine a threshold of significant entropy values, and finally create a directed network of information flow between the recording sites.

When considering applications to information-intensive intelligence production enterprises, research methods used in neuroscience are being proposed to create parallel methods to study these enterprises. After all, the brain is a complex, information processing entity with sensory inputs; complex computation abilities including information transfer between regions, information storage within regions, and information modification; as well as expected output possibly in the form of movement, speech, or further information gathering. An information-intensive organization has parallel characteristics: information flows into the organization; must be processed by storing it, modifying it, or transferring it to other entities; and then some type of output is expected. Instead of synapses an organization would utilize communication channels. Instead of neurons, an organization would have production elements. Information that flows through synapses between neurons in the brain is comparable to information exchanged through communication channels. By observing the spike trains of organizational elements, just as in neuroscience, one can determine the information flow patterns through the organization. [9]

5

Transfer entropy is an important measure used in neuroscience to discover network structures. Transfer entropy [7] is the amount of information a source process' past provides about the next state of a destination (or target) process in the context of the destination's past. [10] Other useful definitions of transfer entropy describe it as a measure of deviation from independence [11] or as an observed correlation between two processes rather than a direct effect [12].

Transfer entropy techniques were demonstrated using social media as an application area. [13] In this study, the authors observed the spike train time histories of a group of Twitter users. The data spiked when the user tweeted a message. Using transfer entropy, they calculated the strengths of the influence between each pair of users, applied a threshold, and created a network map showing how information flows within this social network. They validated the results by reviewing the content of the messages. The authors were able to determine the network completely with the exception of a few links that can be traced back to sampling effects. A similar technique could be performed to study the information flows through an information-intensive organization. An information supply network is essentially a communication network. [14] Then by observing the output of each information production element, performing transfer entropy calculations, and applying a threshold (either using the transfer entropy value or p-value) to the results, it should also be possible to determine the information supply network within an information-intensive organization.

Assumptions/Limitations

This dissertation relies on certain assumptions encompassing the underlying information theory.

One of the benefits of using information theory to uncover a network structure is it requires no assumptions about the structure to proceed. However, a fundamental assumption within information theory and especially information dynamics is to assume the processes being evaluated are statistically independent (i.e., $P(A \cap B) = P(A)P(B)$). This is done to simplify the underlying calculations. This assumption can be counterintuitive given transfer entropy (for example) can be defined as the deviation of two processes from independence, so the assumption of independence would appear to breakdown at higher transfer entropy values where the processes are more dependent. However, previous work has justified this assumption by noting transfer entropy remains a measure of observed (conditional) *correlation* rather than direct effect. [12] Nevertheless, the assumption of independence should be kept in mind.

Implications

The methodology presented in this dissertation will show that by using transfer entropy it is possible to reverse engineer a complex adaptive supply network's structure and dynamics with only time-series production data about network nodes and use local transfer entropy to analyze the network structure's dynamics. This approach requires minimal information about the network. Only the existence of possible nodes with their node histories are needed; underlying behaviors and relationships between the nodes are not needed, modeled, or assumed. Being able to uncover network structures, either static

or dynamic, from real-world data overcomes two significant barriers to studying CASN: that static and dynamic structures or underlying node behaviors must be completely known and that node behavior must be modeled or assumed.

Furthermore, applying information theory in the context of such supply networks will open up new ways of studying these complex systems by enabling other information theoretic measures such as information storage or information modification. Other disciplines have done significant research relating information theoretic measures with network behaviors such as self-organization and emergence [15] [16], stability/instability [17] [18], and distributed computation [19], all of which remain open areas of research for CASNs. Focusing supply network research on these relationships could reveal new approaches to network management and business analytics in networks adapting to their environment.

This dissertation also presents analysis of information transfer and information storage in organizational networks using the information theoretic methodology. Given that organization performance is largely impacted by actors in an organization using formal and informal networks to coordinate effort, share goals, exchange information, and access resources [20], the measures of information transfer and information storage provide quantification of how organizational network topologies differ in their performance.

I put forth multiple observations which provide a quantitative basis for either empirical or qualitative assertions made by previous researchers. The underlying model allows the testing of these assertions and apply them across 5 organizational network topologies, which has not previously been done. In this way, I also study the risks of

8

decreased communication and opportunities of increased communication in these various

organizational networks. Finally, I provide a quantitative study of the effects of informal

network links and their ability to affect the transfer and storage of information within the

network.

Again, although much of this research was performed with an eye toward

intelligence production networks, the fundamental principles may apply to other

networks as well. Examples of other networks which may be able to apply these

methodologies and principles include military (non-intelligence) operations, swarming

networks, public news/media organizations, and real estate agencies. In any of these

systems, a better understanding of the information transfer and information storage

behaviors would help direct managers or architects of these networks to improve network

performance. [2]

Preview

This dissertation is arranged so that the chapters build upon each other.

Chapter II presents a motivation using a very basic application of transfer entropy

to discover simple networks. This is intended to show the ability of the production

dynamics model to reliably simulate communication in a network such that information

theoretic principles can rediscover the network. This chapter was peer-reviewed and

published in the proceedings for the Complex Adaptive Systems Conference in 2015. [9]

Chapter III is an extension of Chapter II, introducing a methodology and applying

it to more complex networks, network dynamics, and a real-world network example using

actual communication/production data. This chapter was peer-reviewed and published in Entropy in 2016. [2]

Chapter IV extends the methodology of Chapter III to networks with organizational topologies and studies the ability of the communication/production data simulation technique to discover those topologies. A statistical approach is used and best-practice values are presented to enable to practitioners to set limits which discover significant network links with desired probabilities of true/false positives. This chapter was peer-reviewed and published in the proceedings for the Complex Adaptive Systems Conference in 2016. [21]

Chapter V investigates organizational network topologies to analyze their capacity to both transfer and store information. Various node communication throughput values are studied as a means of analyzing the effects of increased or decreased resources in a node. These results are compared with empirical and qualitative organization science research to provide quantitative examples of their findings. This chapter has been submitted to the Journal of Complex Networks and is under peer review for publication consideration.

Finally, Chapter VI concludes the dissertation with a summary of research, significance, and recommendations for future research.

II. Motivation for Simulation of CASN Using Information Theory

Introduction

Supply networks exist throughout our society in manufacturing and knowledge-intensive industries as well as many service industries. Examples of these industries include product development, real-estate, healthcare, news/media, and investment services. Thoroughly understanding supply network behavior is critical to managing such systems effectively. Unfortunately, supply networks often take on the behavior of complex adaptive systems making them more difficult to analyze and assess. Being able to fully grasp how material and information flow through a complex adaptive supply network is key to being able to make more informed management decisions and prioritize resources and production throughout the network.

Insights from information theory and information dynamics (heavily leveraged in neuroscience) provide a novel way to overcome many of the difficulties associated with modeling complex adaptive supply networks. The notion of transfer entropy is especially useful in mapping the production network merely by observing the nodal output.

Background

Complex Adaptive Supply Networks

A complex adaptive system (CAS) is a network of dynamical elements where the states of both the nodes and the edges can change, and the topology of the network itself often evolves in time in a nonlinear and heterogeneous fashion. [3] A supply network is a network of firms that exist upstream to any one firm in the whole value system where the value stream could represent material or knowledge flow and behaves as a CAS. A

11

complex adaptive supply network (CASN) is a collection of firms that seek to maximize their individual profit and livelihood by exchanging information, products, and services with one another. [22] Supply-chain networks form CASs because they display the following: structures spanning several scales, strongly coupled degrees of freedom and correlations over long length and timescales, coexistence of competition and cooperation, nonlinear dynamics involving interrelated spatial and temporal effects, quasi-equilibrium and combination of regularity and randomness (i.e. interplay of chaos and non-chaos), emergent behavior and self-organization, and adaption and evolution. [3]

Previous researchers have posited that successful modeling efforts of large-scale CASN would require a solid empirical base and that purely abstract mathematical contemplation would be unlikely to lead to useful models. It is also unlikely that a single model can capture all the aspects of supply-chain processes, and therefore the modeling process should occur at multiple levels. Several options exist for modeling CASN: system dynamics, agent-based modeling, deterministic models, and network models. A drawback to system dynamics models is that the structure has to be determined before starting the simulation. Agent-based modeling, from complexity theory, is a bottom-up approach which simulates the underlying processes believed to be responsible for the global pattern. A network view has proven useful in modeling a supply-chain for patterns of interaction. Queuing theory can be used to analyze the steady state operation of a network, and mathematical programming can solve problems of resource allocation. [3]

Li et al. proposed a model for CASN evolution using the principles of CAS and fitness landscape theory. They modeled the evolution of the CASN by modeling the environment, the firm, and the supply network evolution. For simulating the CASN

12

evolution they used a multi-agent architecture, interaction of agents and two different design experiments: the first for structure dynamics of CASNs and the second for dynamic evolution of firm's fitness. Their work resulted in some interesting managerial implications. Their primary take-away for managers of CASNs was that evolution is a self-organizing process and that any planning and regulation of the market and firm may be undermined by the fact that the outcomes are both open and unknowable. [23]

This result was similar to the implications of applying CAS research to the strategic management of organizations, namely that a manager should not attempt to make sweeping enterprise-wide changes because the system's nonlinear response is too difficult to predict and control. Instead, the managers should set boundaries or constraints on the system and observe the outcome. Then they would be able to tune the system by modifying the constraints and/or changing the amount of energy allowed into the system. The primary role of a strategic organizational architect is to influence the extent of improvisation, the nature of collaboration, the characteristic rhythm of innovation, and the number and nature of experimental probes by changing structure and demography. One factor complicating the use of CAS models in strategic management is that no theory exists to help managers predict how their actions may cascade through the CAS and affect emergence. [4]

There are at least three major challenges critical to CASN research. First, the complexity of supply networks pushes the limits of researchers' ability to understand the internal interactions between constructs and mechanisms or larger-scope phenomena. Second, operations management and supply-chain management lack metrics for evolution and dynamism in supply networks. Third, developing robust theories in the presence of

13

adaptation is a formidable task. Supply-chain management theory can be built by identifying CAS phenomena and that future CASN theories show be built by viewing the properties associated with entities, topology, system, and environment as interrelated constructs. Other research issues include system scale and unit of analysis, environmental scope, and leveraging models, measurements, and methodologies for validation. [24]

Information Dynamics

To address the issue of how information moves between nodes of the supply network, we appeal to the concepts of information dynamics. Information dynamics arises out of concepts in information theory [6] such as conditional entropy, mutual information (Eq. 1), and conditional mutual information (Eq. 2) between processes *X*, *Y*, and *Z*.

$$I(X;Y) = \sum_{x \in \alpha_x} \sum_{y \in \alpha_y} p(x,y) log_2 \frac{p(x|y)}{p(x)} \tag{1}$$

$$I(X;Y|Z) = \sum_{x \in \alpha_x} \sum_{y \in \alpha_y} \sum_{z \in \alpha_z} p(x,y,z) log_2 \frac{p(x|y,z)}{p(x|z)} \tag{2}$$

Transfer entropy [7] is perhaps the most applied of the information dynamics and has been used extensively in neuroscience research. Lizier [25] describes transfer entropy (Eq. 3) as the amount of information that a source process provides about a destination (or target) process' next state in the context of the destination's past.

$$T_{Y \to X}(k,l) = I\left(Y_n^{(l)}; X_{n+1} \Big| X_n^{(k)}\right) \tag{3}$$

Excess entropy is another information dynamics measure that is useful in studying CASN. Excess entropy (Eq. 4) is the mutual information between the semi-infinite past and semi-infinite future of a single process, X. It captures the information in the process' past that is useful to predicting the process' future. [25] For this reason, excess entropy is also referred to as predictive information or stored information (information storage). [15]

$$E_X = \lim_{k \to \infty} I\left(X_n^{(k)}; X_{n+1}^{(k+)}\right) \tag{4}$$

Active information storage (Eq. 5) follows from excess entropy except that only the next state of X is considered, rather than the semi-infinite future of X. Active information storage measures how much of the information from the past of process X is actively in use computing the next state of X.

$$A_X = \lim_{k \to \infty} I\left(X_n^{(k)}; X_{n+1}\right) \tag{5}$$

Separable information is one attempt to measure the information modification within a system. This issue remains an open problem, but Lizier has implemented one attempt at capturing information modification (Eq. 6). This separable information is the limit of the linear combination of active information storage and all transfer entropies

into source X. It aims to quantify how sources within a process combine to predict the next state of X.

$$S_X = \lim_{k \to \infty} \left[I\left(X_n^{(k)}; X_{n+1} \right) + \sum_{Y \in V_X \setminus X} T_{Y \to X}(k, l_Y) \right] \tag{6}$$

When considering these information dynamics measures within the context of CASNs, they can provide useful insights into the system behavior, especially for knowledge-intensive CASNs. Transfer entropy is useful for determining dependencies within the network, and by altering the process history lengths, can be used to uncover dependencies which change over time. Excess entropy (or stored information) is information which although not being used for the current process, will be used to determine future process states. This provides insight to managers of CASN which may have previously only considered products "in-hand" rather than those in progress. The active information storage takes on a more immediate role to the managers who want to know what is currently in progress within the CASN. Finally, although information modification is the least understood of the information dynamics, it measures how sources combine to predict future states. This may be useful to managers who want to understand what-if scenarios should a source be delayed or removed.

Transfer entropy methods have been applied in several fields to study complex systems: neuroscience [26] [27] [28], social networks [13], medicine [29], finance [30], climate [31] [32], transportation [33], and others.

Methodology

  To begin applying information-theoretic principles to CASN, the authors devised a simple conceptual supply network made up of two primary supply chains. The network consists of 6 nodes (2 source nodes, 3 intermediary nodes, and 1 final production node), see Figure 1. In this network, the output of one node feeds into another node which produces its own output based on that input.
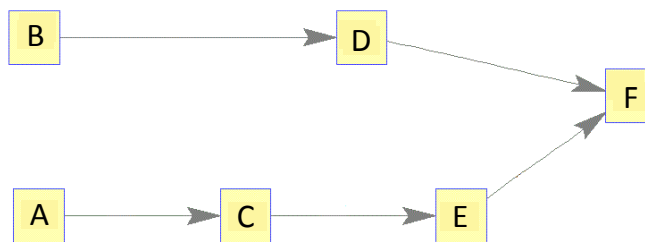
Figure 1. Conceptual supply network

  The nodes were also given varying production schema as shown in Table 1. The source nodes produced at random time intervals with either high or low frequency production, while the various intermediary nodes had production schema dependent on the preceding node in the network. Nodes were either given random or fixed offset times from the preceding node; fixed offsets represented deterministic production turnaround times while random offsets represented varying production times. The final production node essentially combined the production of the two source streams by employing a fixed production time from the maximum of the two preceding nodes.

Table 1. Simulated network production schema

Node	Node role	Production schema
A	Source	Random, high-frequency times (rand(10000,1)>0.3)
B	Source	Random, low-frequency times (rand(10000,1)>0.7)
C	Intermediary	A (every other) + 1
D	Intermediary	B + randbetween(1,4)
E	Intermediary	C + randbetween(1,2)
F	Final production	D|E + 1

Using these simulated production times, the transfer entropy was calculated for every pair of nodes in the network using the JIDT code published by Lizier [25]. The resulting output was then filtered to remove network links of lesser significance. The authors retained edges between nodes with a transfer entropy of 0.03 bits or greater. Again, this value was empirically determined, but the transfer entropy between the random source nodes was a good starting point (since they should be uncorrelated). The resulting network is shown in Figure 2.
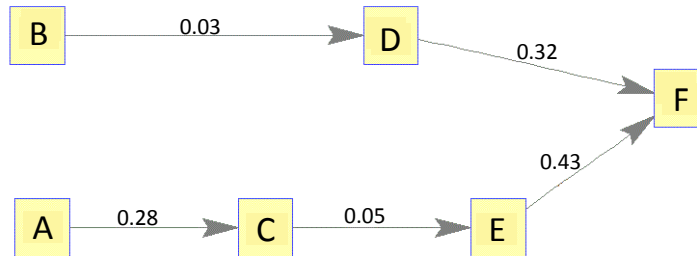
Figure 2. Resulting supply network links and transfer entropy values

Additionally, the authors investigated the filtering threshold used to generate the network graph and how the threshold changed as the amount of time history data changed. To do this, the authors calculated the transfer entropy on the network for every pair of nodes with the first 10 data points and incrementally added the next 10 data points, observing how the transfer entropy values changed with each addition. The results are shown in Figure 3. Note that the solid lines are transfer entropy on edges which were in the conceptual supply network while the dashed lines are transfer entropy on edges not in the conceptual network.
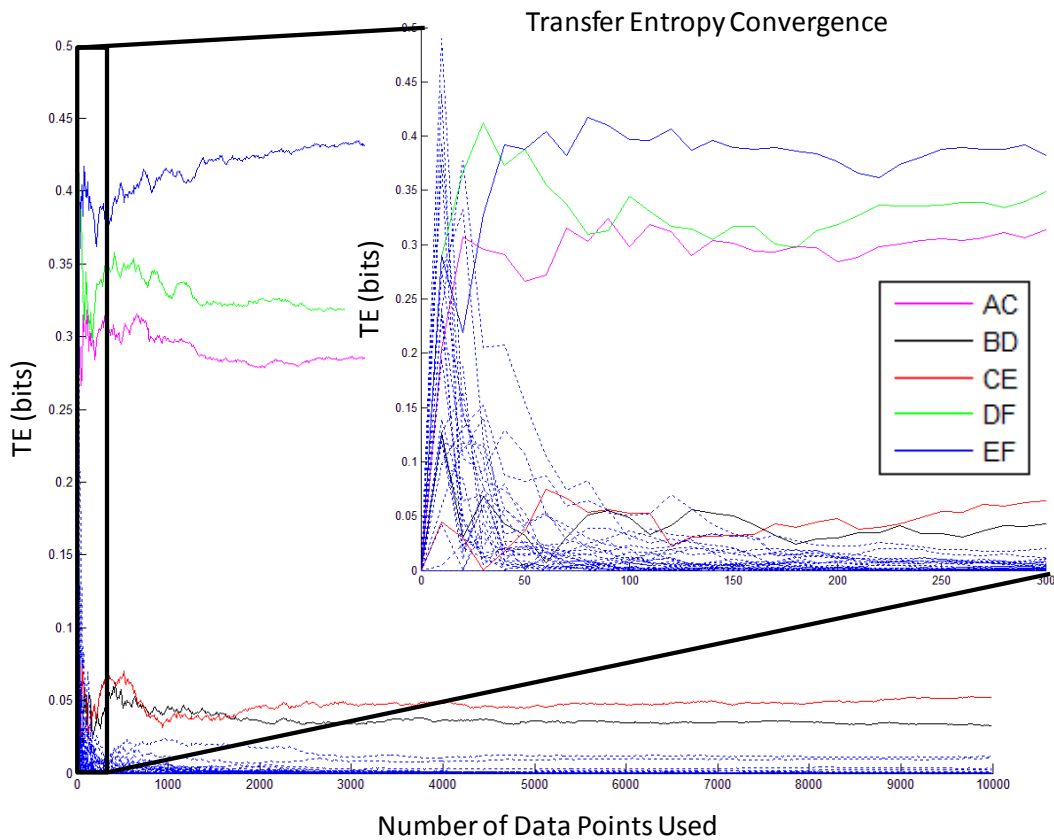
Figure 3. Edge transfer entropy changes with time history length

Analysis

For this simple supply network, transfer entropy was able to uncover the network structure using only the node production histories. The results were dependent on the filter chosen in the calculations. Once the data leveled off (around 1,000 data points), the filter threshold value did not vary significantly based on the length of time history used to calculate the transfer entropies. However, with fewer data points (especially *<200*), the TE results would not be able to accurately return the underlying network.

As seen in Figure 3, there was essentially no distinction between the edges in the conceptual supply network (solid lines) and those not in the conceptual supply network (dashed lines) when the number of data points used was less than 200 and only became clearer when more than 200 data points were used. As more data points were used, the less-significant edges (those not in the original concept network) remained lower than the actual edges in the network. This behavior is expected as the nodes should not be correlated. This could complicate efforts to filter the edges in the network when the network is not known and relatively short time histories of data are available.

An additional way to visualize the threshold is to discreetly integrate the data plotted in Figure 3. Although not shown here, this plot produces a much more apparent "natural" break in the transfer entropy values of the edges and allows visualization of the threshold to lower number of data points.

This analysis leads the authors to propose three heuristics for choosing filter thresholds on transfer entropy data:

- *Heuristic #1*: From two known randomly correlated nodes, calculate the transfer entropy. Filter any edges with transfer entropy less than or equal to this value.

- *Heuristic #2*: From the graph of all transfer entropies, remove the lowest valued edge until a graph of desired (or expected) connectivity is reached.

- *Heuristic #3*: From plots of transfer entropy vs. time history length (or its discreet integral), look for natural breaks in the data and filter any edges below this threshold.

Conclusions

CASNs have challenged researchers' ability to study them because of their dynamic nature and complex behaviors. Information theory and information dynamics provides a way to overcome many of the difficulties associated with modeling CASNs. The notion of transfer entropy is especially useful in mapping the production network merely by observing the nodal output.

A simple conceptual supply network was created using simulated production timing for 6 nodes. Transfer entropy calculations completely revealed the network structure based only on input of the production times. Filtering thresholds were studied for varying lengths of time history and 3 heuristics were proposed to set these thresholds.

Future work will study larger, more complex networks which change over time. Simulations will help determine methods and techniques for analyzing dynamic network

structures and revealing links which change with time. Additionally, organizational data

will demonstrate these techniques on actual production systems.

III. Methodology for the Analysis of Structure and Dynamics of CASN Using Information Theory

Introduction

Supply networks exist throughout society in manufacturing and knowledge-intensive industries as well as many service industries. Manufacturing and material production supply networks are arguably more straightforward to understand than knowledge- and information-intensive networks as relationships tend to be more clearly defined. Examples of these knowledge- and information-intensive networks include product development, real-estate, healthcare, news/media, and investment services. In these networks, information and knowledge is more difficult to track as it moves through the network. Thoroughly understanding supply network behavior and dependencies are critical to managing such systems effectively. Unfortunately, supply networks often take on the behavior of complex adaptive systems making them more difficult to analyze and assess. Being able to fully understand both the static and dynamic structures of a complex adaptive supply network is critical to being able to make more informed management decisions and prioritize resources and production throughout the network.

Complex Adaptive Supply Networks

As a starting point, Choi et al. [22] state that a complex adaptive supply network (CASN) is a collection of firms that seek to maximize their individual profit and livelihood by exchanging information, products, and services with one another. They further elaborate that a larger component of the order in a CASN is the emergent, dynamic, and unpredictable order that arises when the nature of interactions between

firms determines the behavior of the network as a whole. In these ways, these interconnected entities exhibit adaptive action in response to changes in both the environment and the system of entities itself.

According to Surana et al. [3] in a complex adaptive system (CAS), the network elements are dynamic with the states of both the nodes and edges changing with time. They further argue, supply-chain networks form a CAS because they display the following characteristics: structures spanning several scales, strongly coupled degrees of freedom and correlations over long length and timescales, coexistence of competition and cooperation, nonlinear dynamics involving interrelated spatial and temporal effects, quasi-equilibrium and combination of regularity and randomness (i.e. interplay of chaos and non-chaos), emergent behavior and self-organization, and adaption and evolution.

These are similar to the properties noted by Pathak et al. [24]: a CAS consists of entities that may evolve over time and interact with other entities and the environment by following a set of simple decision rules; a CAS is self-organizing as a consequence of interactions between entities; a CAS coevolves with its environment to the edge of chaos; a CAS is recursive by nature and recombines and evolves over time.

Surana et al. noted several options which exist for modeling CASN: system dynamics, agent-based modeling, deterministic models, and network models. A drawback to system dynamics models is that the structure has to be determined before starting the simulation.

Agent-based models and NK models offer additional approaches to studying supply chains as complex adaptive systems. Mari et al. [34] used agent-based simulation to test various models of resiliency metrics in complex supply chains. Li et al. [23]

proposed a model for CASN evolution using the principles of CAS and fitness landscape theory. Giannoccaro has used agent-based models to simulate learning in adaptive supply chains [35] and NK simulation to study supply chain integration and governance [36]. NK simulation has been also used to study supply chain interdependence/trust [37] [38].

Hearnshaw et al. [39] provided a conceptual approach to supply chain network theory drawing on an understanding of complex network theory and adaptive systems. They provide a justification for applying advances in complex network science to supply chain management, but given their conceptual approach, stopped short of empirically validating the many propositions they presented in the research. Providing a quantitative validation of their propositions through modeling or empirical analysis remains an open issue.

However, according to Pathak et al., there are at least three major challenges critical to CASN research. First, the complexity of supply networks pushes the limits of researchers' ability to understand the internal interactions between constructs and mechanisms or larger-scope phenomena. Second, operations management and supply-chain management lack metrics for evolution and dynamism in supply networks. Third, developing robust theories in the presence of adaptation is a formidable task.

Additionally, Bellamy et al. [40] identified other research gaps in the area of network science and supply chain management. Overall, they noted that future work should focus on supply chain network structure, supply chain network dynamics, and supply chain network strategy. Since then, recent studies of supply network structures have primarily utilized social network analysis methods. For example, Basole et al. [41] used a network analysis approach to study the relationships between network structure,

25

risk diffusion, and network health. Bellamy et al. [42] also used a social network approach to examine how supply chain network structure impacts innovation within the firm.

The obvious drawback of the social network analysis approach is it requires complete knowledge of the network structure in order to be effective. However, in a CASN where both the edges and nodes are dynamic, a complete understanding of the network structure may not be known. To address these difficult issues of complex network structures and dynamics, the authors appeal to the concepts of information dynamics.

Information Dynamics

Information dynamics arises out of concepts in information theory [6] such as mutual information, *I(X;Y)*, (Eq. 7) and conditional mutual information, *I(X;Y|Z)*, (Eq. 8). In these equations, processes *X*, *Y*, and *Z* have time-series realizations *x*, *y*, and *z* over all possible states α_x, α_y, and α_z. Lizier [25] describes mutual information as a measure of the information contained in *X* about *Y* (or vice versa) and the conditional mutual information as the mutual information between *X* and *Y* when *Z* is known.

$$I(X;Y) = \sum_{x \in \alpha_x} \sum_{y \in \alpha_y} p(x,y) log_2 p(x|y) \tag{7}$$

$$I(X;Y|Z) = \sum_{x \in \alpha_x} \sum_{y \in \alpha_y} \sum_{z \in \alpha_z} p(x,y,z) log_2 \frac{p(x|y,z)}{p(x|z)} \tag{8}$$

Additionally, Lizier [25] describes transfer entropy (Eq. 9) [7] as the amount of information a source process' past, $Y_n^{(l)}$, provides about the next state of a destination (or target) process, $X_{(n+1)}$, in the context of the destination's past, $X_n^{(k)}$. In this equation k and l are the history lengths of X and Y, respectively, while n is the current time index. Other useful definitions of transfer entropy describe it as a measure of deviation from independence [11] or as an observed correlation between two processes rather than a direct effect [12].

$$T_{Y \to X}(k, l) = I\left(Y_n^{(l)}; X_{n+1} \middle| X_n^{(k)}\right) \tag{9}$$

Additionally, local transfer entropy (Eq. 10) [25] defines transfer entropy at each time step n using the past realizations $x_n^{(k)}$ and the past realizations $y_n^{(k)}$ to quantify the information contained in the source, Y, about the next state of the destination, X, at time step $n+1$. The local transfer entropy measure gives insight into network structure dynamics and how the correlation and influence or two processes are changing over that time series. This is because local transfer entropy provides a time history of transfer entropy values which quantifies the observed correlation (or influence) between two processes.

$$t_{Y \to X}(n+1) = \lim_{k \to \infty}\left(\log_2 \frac{p\left(x_{n+1} \mid x_n^{(k)}, y_n\right)}{p\left(x_{n+1} \mid x_n^{(k)}\right)}\right) \tag{10}$$

27

Transfer entropy has been implemented with a method to determine the statistical significance of the process relationships being studied. A null hypothesis testing approach has been taken in the Java Information Dynamics Toolkit (JIDT) [25]. In order to determine if the transfer entropy calculated between two processes is statistically different from 0 (i.e., no relationship exists), the author forms a null hypothesis *H0* which states there is no relationship between the processes between studied. He then uses the null hypothesis to create distributions of surrogate measurements using the same statistical properties of *Y* with any potential correlation with *X* removed. Knowing what the distribution for the measurements would look like if *H0* were true, he then calculates a *p*-value, sampling the actual measurement from this distribution. If the test fails, the alternate hypothesis is accepted stating there is a statistically significant relationship between the processes.

Transfer Entropy Application

Transfer entropy has been applied extensively in neuroscience [26] [27] [28] [5] to uncover underlying network structures based solely on node behavior. Many other fields have used transfer entropy to infer network structure and/or relational dependencies between entities in complex networks: social networks [13] [43], medicine [29], finance [30], climate [31] [32], transportation [33], and biology [44] [45], just to name a few.

Here, the authors are proposing information dynamic methods including transfer entropy and local transfer entropy as an alternate perspective to the analysis of CASN. Where previous attempts to study CASN relied on agent-based models marked by numerous modeling assumptions or on social network analysis requiring complete knowledge of the network structure and dynamics, the information dynamic approach

28

allows a researcher to proceed with analysis of CASN under minimal assumptions and instead rely more heavily on real-world data to extract meaningful insights. This proposed analysis approach, which does not make assumptions about node behavior, their relationships, or their dynamics, overcomes one of the primary challenges inherent to the study of CASN: the complex structure and internal interactions can be found instead of assumed.

Additionally, the proposed information theoretic methodology opens up other opportunities when analyzing CASN. Information theoretic measures such as information storage or information modification could prove useful to CASN analysis as well. Although these ideas require more research relative to CASN specifically, other disciplines have found correlations between information theoretic measures and the concepts of self-organization and emergence [15] [16], network stability [17] [18], interdependencies [46], and distributed computation [19] in many types of networks. Being able to quantify these behaviors and characteristics for CASN would enable significant advances in a research area that has been primarily grounded in empirical data, models and simulations, and social network analysis. Moreover, as information theory provides insight into emergent behaviors or network stability, this could spawn new approaches to business analytics or network management as a network adapts to its environment.

Materials and Methods

The general methodology used to analyze CASNs using information dynamics is as follows:

1. Create conceptual network graph

2. Simulate production data on network

3. Apply transfer entropy to production data for static analysis

4. Apply local information transfer to production data for dynamic analysis

It should be noted that the first two steps are only necessary for simulating conceptual networks for academic study. Analysis of real-world networks using information dynamics begins at the third step. For the purpose of this article, the 'Materials and Methods' includes the description of the process used to create the conceptual networks and simulate production data on the networks. It also includes validation of the methodology applied to both static and dynamic structures of simulated networks. The 'Results' focuses on the application of the information dynamics methodology as would occur for a real-world application.

Conceptual Static Supply Network Simulation

In preparation to analyze CASNs using information-theoretic principles, the authors first devised a conceptual information supply network structure and then simulated knowledge production timing schema as behaviors of each of the nodes. The conceptual network structure consisted of 10 nodes (4 source nodes, 5 intermediary nodes, and 1 final production node), shown in Figure 4. A simple tree structure was chosen for this case, although other structures would have been equally suitable. The methodology does not necessitate any specific number of source nodes, intermediary nodes, or final destination nodes. In fact, when applied in practice, these labels are not applied, and the methodology detects the network structure as present.
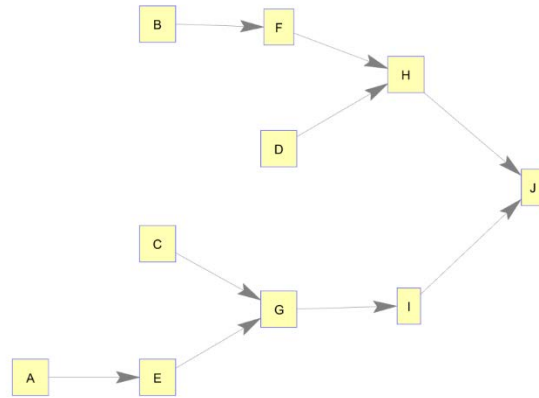
Figure 4. Conceptual static supply network structure

For the production schema on this network, the output of one node became an input to a second node which produced its own output based on input(s) received. All nodes' knowledge production was governed by a simple 1-unit time delay. Essentially, a node produced information and passed it to the next node which used that for its own production 1 time unit later. In the cases in which a node had dual input sources, that node produced knowledge based on both input nodes' information; if either source node produced information, that node also produced 1 time unit later. In the cases of the source nodes (those without any inputs), production timing was simulated as a random series with approximately 30% production frequency.

Many alternative information production timing schema could have been simulated. For example, the authors could have chosen to vary the time delays among the nodes, create randomly variable (but bounded) time delays, or to ignore some number of inputs to determine production timing. [9] However, for this initial application, the deterministic approach of a 1-unit time delay was chosen to more clearly demonstrate the method without adding unnecessary noise.

It should be noted, that although this data is simulated, real-world data may already exist for many business applications and be readily available from workflow management software such as those discussed in Van der Aalst et al. These workflow process mining techniques address a similar problem of discovering a business process from event logs, but the method is an application of Petri nets and has limitations as to the network structures that can be discovered. [47] It would be an interesting area of future research to compare the workflow mining techniques from those authors to the information theoretic methodologies presented here.

To complete the analysis, the authors used these simulated production times to calculate the transfer entropy for every pair of nodes in the network using the JIDT code published by Lizier. [25] P-values were also calculated using the JIDT code which allows the user to compute the significance of the transfer entropy. The resulting output was then filtered to remove network links of lesser significance as determined by the p-value. As an example, a network link could be accepted as significant if $p < 0.05$, as is typical of many p-value tests. Because this was a deterministic case, only links with $p = 0$ were retained. In the case where uncertainty is added into the production timing data, the p-value would need to increase accordingly.

The result of this analysis gives the user a network of nodes with significant relationships identified and weighted by transfer entropy values. As discussed previously, these transfer entropy values can be interpreted in multiple ways. Looking at relationships between individual nodes, the transfer entropy is the amount of information the source node provides about the future state of the target node (given the target's past). This means higher transfer entropy indicates that the source node better correlates with

future activity of the target. Put another way, when a source and target node pair have high transfer entropy between them, if the source node produces information, the user can reasonably expect the target node to produce information during the next time step. Additionally, transfer entropy can be viewed as the deviation from independence for two processes. Therefore, in this network example, a source and target node pair that relate with a high transfer entropy indicate that the target node is highly dependent upon information from the source node. These relationships are especially important to understand in a CASN where dependencies between nodes are continually evolving and emerging as the network adapts to its environment.

Static Network, Structure Validation

In the case of the 10-node, static supply network, transfer entropy calculations successfully uncovered the underlying conceptual network structure using only the simulated node production histories. This validated the methodologies used to simulate production dynamics within the network as well as the methodology used to analyze the static network structure. Resulting transfer entropy values and corresponding p-values are shown in Table 2 and Table 3, respectively, with significant ($p = 0.00$) values bolded. Because transfer entropy is a directional measure, it produces an adjacency matrix which is not symmetric enabling us to construct a directed network graph. This adjacency matrix of transfer entropy values is shown in node/edge format in Figure 5.

Table 2. Transfer entropy values as edge weights in adjacency matrix for static supply network (significant values in bold).

Node	A	B	C	D	E	F	G	H	I	J
A	0	0.000	0.000	0.000	**0.887**	0.000	0.000	0.000	0.000	0.000
B	0.000	0	0.000	0.000	0.000	**0.882**	0.000	0.000	0.000	0.000
C	0.000	0.000	0	0.000	0.000	0.000	**0.371**	0.000	0.000	0.000
D	0.000	0.000	0.000	0	0.000	0.001	0.000	**0.390**	0.000	0.000
E	0.000	0.000	0.000	0.000	0	0.000	**0.386**	0.001	0.000	0.000
F	0.000	0.000	0.000	0.000	0.000	0	0.000	**0.377**	0.000	0.000
G	0.000	0.000	0.001	0.000	0.000	0.000	0	0.000	**0.999**	0.000
H	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0.000	**0.304**
I	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	**0.302**
J	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0

Table 3. *P*-values for corresponding edges in adjacency matrix for static supply network (significant values in bold).

Node	A	B	C	D	E	F	G	H	I	J
A	1	0.375	0.612	0.989	**0.000**	0.455	0.129	0.287	0.252	0.780
B	0.556	1	0.763	0.978	0.736	**0.000**	0.868	0.317	0.280	0.046
C	0.315	0.937	1	0.074	0.016	0.204	**0.000**	0.643	0.686	0.005
D	0.688	0.860	0.814	1	0.367	0.566	0.807	**0.000**	0.202	0.143
E	0.197	0.598	0.122	0.612	1	0.369	**0.000**	0.456	0.135	0.584
F	0.407	0.441	0.002	0.917	0.537	1	0.145	**0.000**	0.822	0.944
G	0.645	0.635	0.386	0.612	0.815	0.203	1	0.001	**0.000**	0.034
H	0.646	0.164	0.387	0.055	0.437	0.921	0.775	1	0.058	**0.000**
I	0.399	0.252	0.059	0.361	0.627	0.696	0.900	0.410	1	**0.000**
J	0.609	0.199	0.870	0.113	0.670	0.046	0.298	0.273	0.816	1
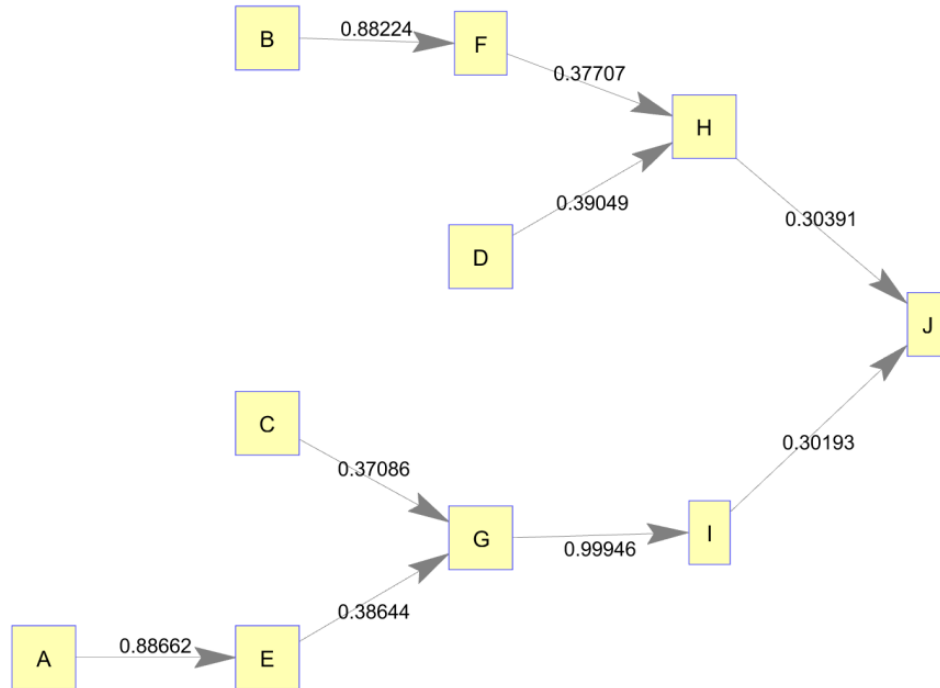
Figure 5. Significant node links for static supply network weighted by transfer entropy values.

Conceptual Dynamic Supply Network Simulation

The network shown in Figure 5 was considered to be static in its nodes' dynamics; that is, the behavior of each node followed the same pattern over the entire duration of the simulation. The authors next considered a dynamic case in which a node varied which inputs it used to produce its information output. Although this was chosen as an academic case, it could represent a real-world situation in which a knowledge producer continually choses between multiple information suppliers for its production based on price, availability, et cetera. For this case, the authors created a network structure with 5 nodes (3 source nodes, 1 intermediary node, and 1 final production node), shown in Figure 6. Here, source nodes A-C produced randomly, as before. Node E

35

deterministically produced based on the production of node D with a 1-unit time delay. Node D, however, dynamically chose its input source from among nodes A, B, & C. In this case, node D selected a source for 20 time steps, followed by another source (or the same one) for 20 time steps, and so on. Simulated production times are shown in Figure 7.
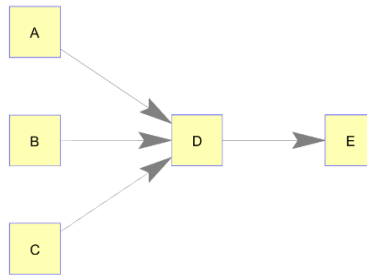
Figure 6. Conceptual dynamic supply network structure.
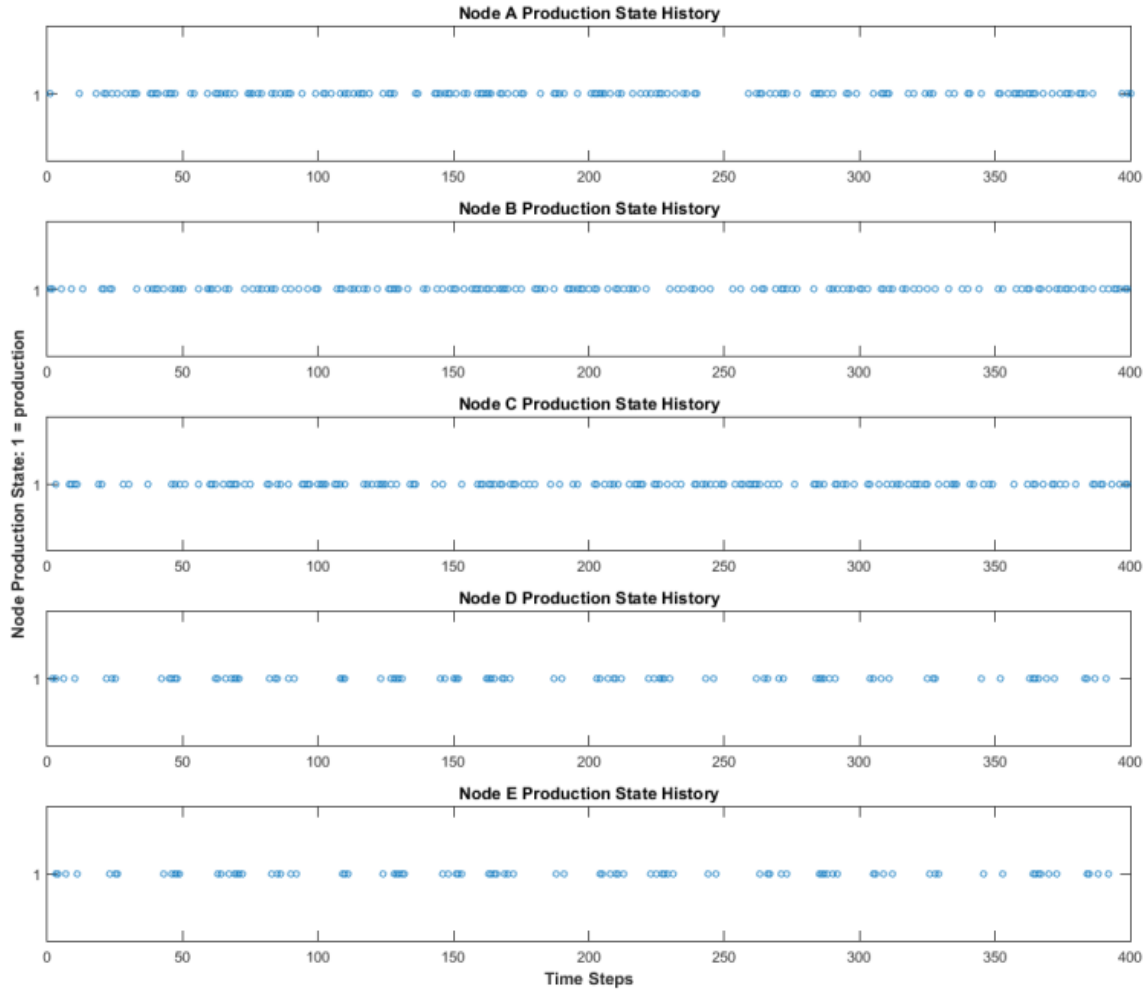
Figure 7. Conceptual dynamic supply network nodes' simulated production histories.

Using these simulated production times, the transfer entropy was calculated for every pair of nodes in the network as before. The resulting output was then filtered to remove network links of lesser significance. Only links with *p < 0.001* were retained. Additionally, using these simulated production times, the local transfer entropy was calculated using JIDT for every pair of nodes deemed to be significant. This was done to study the time-series dynamics of the nodes, especially node D which varied its relationships over time.

37

The result of this analysis gives the user the local transfer entropy of the relationships in the network. Again, local transfer entropy is a time history of the transfer entropy values, providing dynamic insight into the specific interactions between nodes and when they occur. The same interpretation of transfer entropy values applies to local transfer entropy analysis, except that the user now has the network relationship interactions at each time step in the time series, instead of the interaction averaged across all time steps.

Dynamic Network, Static Structure Validation

In the case of the dynamic supply network, transfer entropy calculations again successfully uncovered the underlying conceptual network structure using only the simulated node production histories. This validated the methodologies used to simulate production dynamics within the network as well as the methodology used to analyze the static network structure. Resulting transfer entropy values and corresponding p-values are shown in Table 4 and Table 5, respectively, with significant ($p = 0$) values bolded. This adjacency matrix of transfer entropy values is shown in node/edge format in Figure 8.

Table 4. Transfer entropy values as edge weights in adjacency matrix for dynamic supply network (significant values in bold).

Node	A	B	C	D	E
A	0	0.002	0.005	**0.072**	0.005
B	0.003	0	0.003	**0.051**	0.001
C	0.005	0.004	0	**0.020**	0.002
D	0.002	0.009	0.001	0	**0.743**
E	0.002	0.008	0.006	0.008	0

Table 5. *P*-values for corresponding edges in adjacency matrix for dynamic supply network (significant values in bold).

Node	A	B	C	D	E
A	1	0.431	0.155	**0.000**	0.159
B	0.254	1	0.358	**0.000**	0.793
C	0.160	0.245	1	**0.000**	0.400
D	0.467	0.026	0.639	1	**0.000**
E	0.538	0.034	0.076	0.045	1
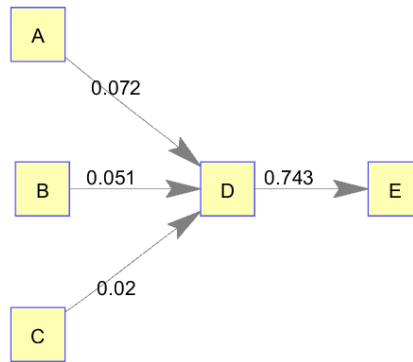
Figure 8. Significant node links for dynamic supply network weighted by transfer entropy values.

Dynamic Network, Dynamic Structure Validation

In the case of the dynamic supply network, local transfer entropy calculations again successfully uncovered the underlying conceptual network dynamics using only the simulated node production histories. This validated the methodologies used to simulate production dynamics within the network as well as the methodology used to analyze the dynamic network structure.

Because local transfer entropy can fluctuate greatly, a moving average was applied to make trends more apparent. Moving averages of 15 time units are shown in

Figure 9. This simulation was fashioned to determine whether or not changing influences in a network can be detected using information theory. Recall that node E's production deterministically followed the production of node D. Therefore, it is intuitive that the local transfer entropy from node D→E would remain the highest of all links. Indeed, it can also be observed that this link's average local transfer entropy is equal to the transfer entropy observed in the static network analysis between the two nodes. Also recall that node D selected inputs from nodes A, B, & C every 20 time steps. This is also apparent in the moving averages of local transfer entropy which display periodic behavior. By choosing max {A→D, B→D, C→D} for each time step, it is possible to determine the source node used by node D as input in the simulation.
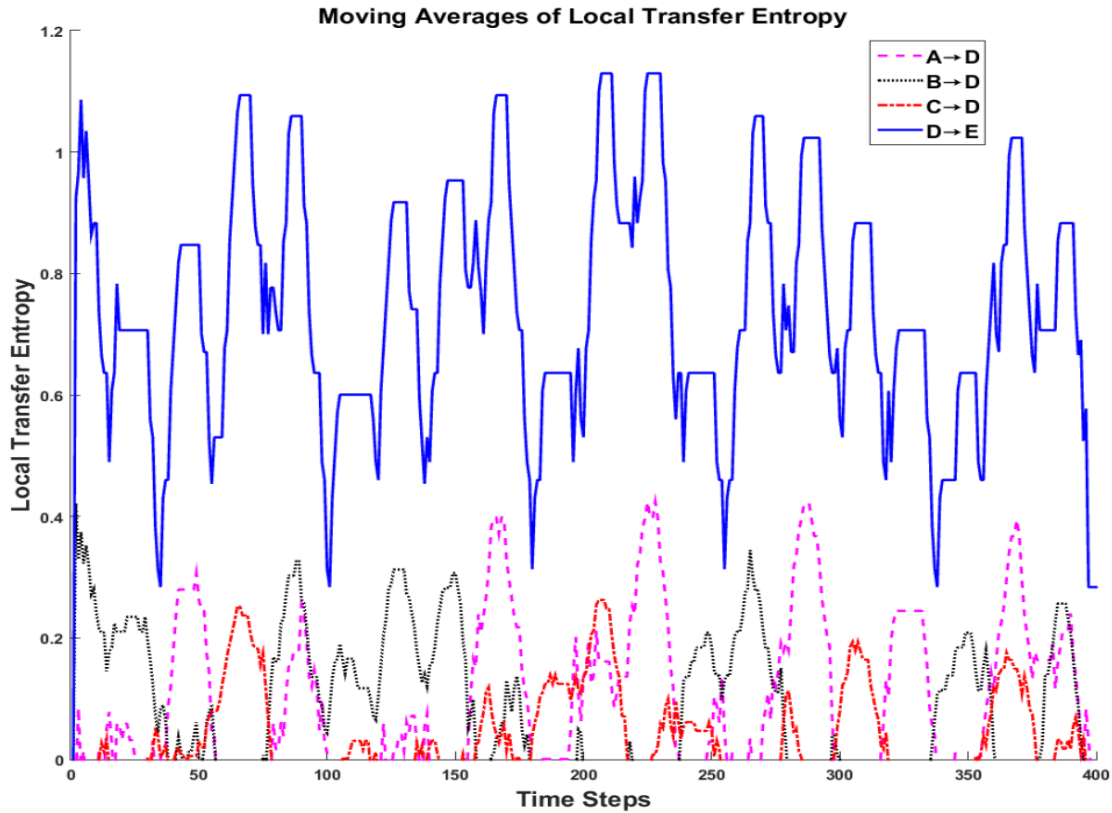
Figure 9. Local transfer entropy moving averages (15 time units) for significant links in dynamic supply network.

Results

Real-World, Dynamic Supply Network

Finally, these concepts were applied to a real-world network with an unknown structure. Data was obtained for an information supply network in which nodes shared information with each other in the form of knowledge production. In this real-world network, all nodes are geographically local to an organization, but nodes A-D are generally considered source nodes while node E is a hub which generally aggregates information and then passes it outside of the organization. The data for this network consisted of monthly production timing for a single topic area, shown in Figure 10. It

should be noted that the methodology does not presume any existing relationships or labels for nodes (source, sink, etc.), but instead determines all nodes' inputs and outputs based on their behavior. This means a node that is generally considered a source node may actually exhibit behaviors of an intermediary node in the network. This arises frequently in the context of information networks as knowledge tends to be highly collaborative. It is these differences between assumptions and real-world analysis that is of great interest to a manager of the network.

This network is similar to those information networks of news/media organizations or real-estate agencies where local sources report information to local hubs, some of which gets passed to regional hubs and then national hubs. As an example, the topic area could be likened to a real-estate topic of 'single-family homes'. Each time a node (or office) created a report about a single-family home, that information was counted as production and added to the production data time-series for that node. In this situation only information about the most marketable single-family homes generated by a local office would be passed to a local hub, and then the information about the most prestigious of those homes would be passed to regional hubs.

Another parallel can be drawn here with the research area of workflow process mining. The data records available in that area of research are often tagged with information about specific tasks, specific cases, and a timestamp. [48] The topic area of this real-world example could relate to the specific case in workflow process mining, the nodes in the network relate to the tasks in workflow logs, and the timestamp in a workflow log is captured by the time-series for that node. Again, although the authors did not choose to perform a workflow mining analysis here, the area could provide an

interesting comparison for future research. It is also possible that workflow process

mining tools such as EMiT (Enhanced Mining Tool) could provide alternate analyses and

additional perspectives on the dataset.
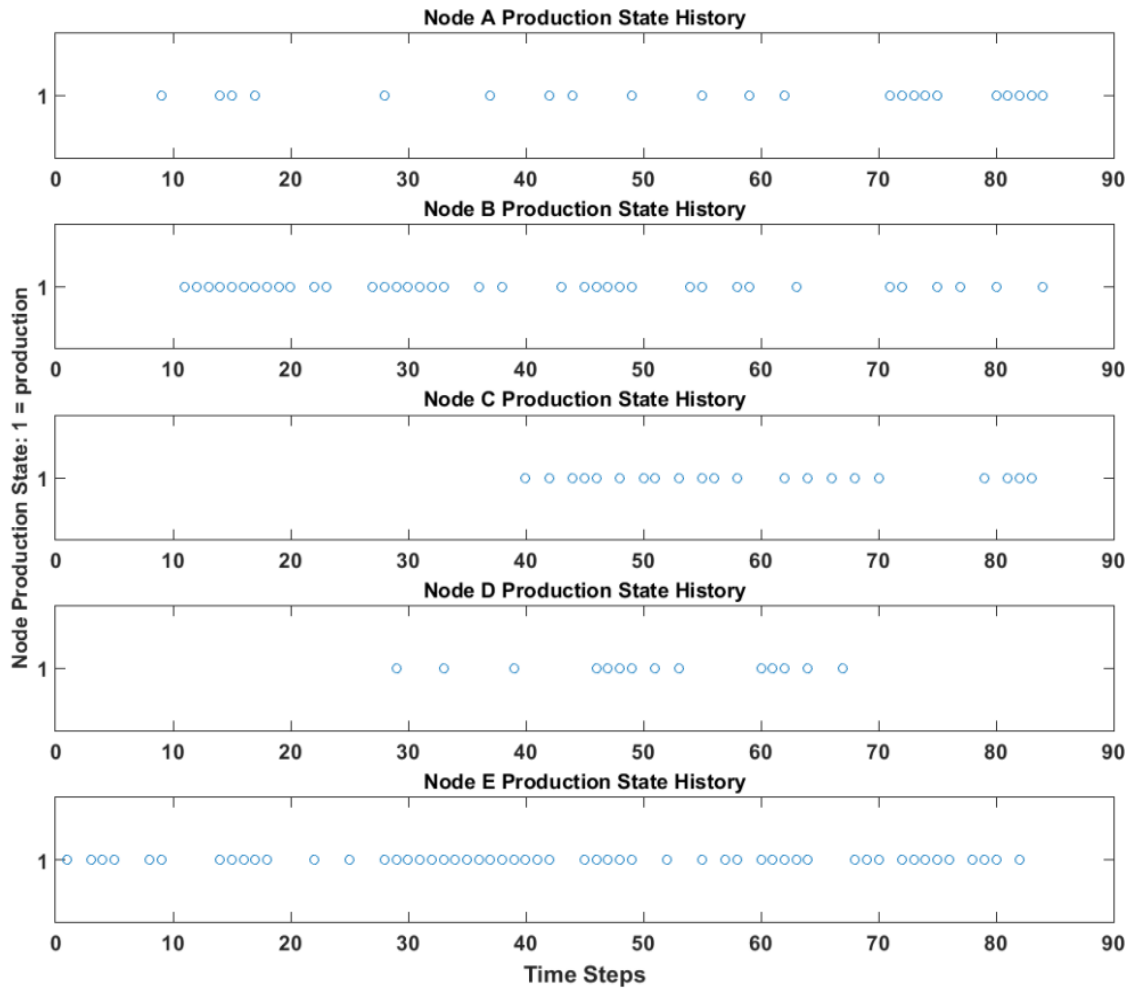
Figure 10. Real-world supply network nodes' production history data.

Based on this production timing, the transfer entropy was calculated for every pair

of nodes in the network as before to determine the underlying network structure. The

resulting output was then filtered to remove network links of lesser significance. Because

this real-world case contained a considerable amount of variability, links with $p < 0.15$ were retained. Additionally, using these simulated production times, the local transfer entropy was calculated for every pair of nodes deemed to be significant. This was done to study the time-series dynamics of the nodes to determine how their relationships varied over time.

Further, this data was then analyzed from the perspective of a manager of node E, the non-source node (or hub), which generally relies on production of the source nodes A-D. Based on all available sources, the manager was interested to learn how his organization's dependence on those data sources had changed over the time period studied. Both transfer entropy and local transfer entropy were used to analyze node E's dependencies, first from an average perspective and then as a dynamic time-series.

Real-World Network, Static Structure

In the case of the real-world supply network, transfer entropy calculations revealed which links between network nodes were significant using only the actual node production histories. Resulting transfer entropy values and corresponding p-values are shown in Table 6 and Table 7, respectively, with significant ($p < 0.15$) values bolded. This p-value was chosen in order to produce a connected graph in which every node was traversable. The adjacency matrix of transfer entropy values is shown in node/edge format in Figure 11. Qualitative analysis of the resulting network graph shows the nodes A-D sharing significant amounts of information as this is where some of the most significant network links reside. This would appear to indicate that these nodes have a strong tendency to collaborate amongst themselves and possibly incorporate each other's information in their production. However, these strong connections may potentially be

explained by unobserved, common triggers. For example, nodes D and C may gather

information from an unobserved common origin which may alternatively explain their

seemingly strong link.

Table 6. Transfer entropy values as edge weights in adjacency matrix for real-world

supply network (significant values in bold).

Node	A	B	C	D	E
A	0	**0.052**	**0.033**	0.018	**0.024**
B	**0.026**	0	0.008	**0.025**	0.006
C	0.017	**0.034**	0	0.022	**0.024**
D	0.022	0.003	**0.086**	0	0.011
E	0.011	0.002	**0.026**	**0.028**	0

Table 7. *P*-values for corresponding edges in adjacency matrix for real-world supply

network (significant values in bold).

Node	A	B	C	D	E
A	1	**0.014**	**0.067**	0.220	**0.141**
B	**0.118**	1	0.518	**0.122**	0.587
C	0.234	**0.061**	1	0.157	**0.136**
D	0.157	0.790	**0.001**	1	0.409
E	0.412	0.867	**0.114**	**0.101**	1
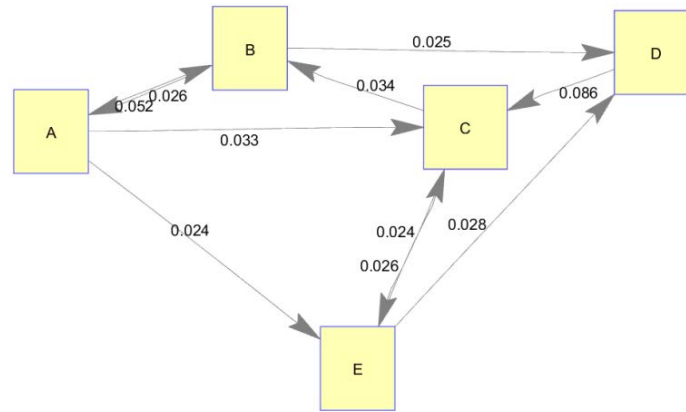
Figure 11. Significant node links for real-world supply network weighted by transfer entropy values.

Because the actual information sharing network (i.e., truth data) was not known in this case, a subject matter expert on the topic area and a network manager familiar with all production offices were consulted to determine whether these new insights into the network were likely to be valid. In general, the resulting network (derived from transfer entropy) matched expectations very well for this topic area. Experts expected to see strong collaboration among nodes A-D and also expected to observe node E to be strongly dependent on nodes A and C. However, it was expected that node B would have more influence on node E's production. It was concluded that this also made sense as information produced by node B did not typically meet a significance threshold to trigger node E's production unless it was first combined with information from other sources.

Further, the managers noted the lack of connection between nodes A and D, which they assumed should be connected. The question arose as to whether this was an issue with processes, personnel, or some other factor. Some amount of investigation into

the issue revealed poor working relationships between the individuals in nodes A and D which undoubtedly hampered collaboration. With this knowledge, the managers were able to assess paths forward to foster a connection between these offices.

These management observations highlight one of the primary characteristics of a complex adaptive system in that the agents in the network work autonomously to contribute to the overall functioning of the network. This autonomy meant the management team did not have total transparency into the inner workings of these actors without additional tools. In fact, it is these inner workings and hidden relationships that the information theoretic tools are able to discover.

Dynamic aspects of this network are considered in the next section by means of local transfer entropy.

Real-World Network, Dynamic Structure

For the real-world information supply network, the next step of the methodology was to calculate the local transfer entropy for each of the links in the network. However, this time, instead of choosing all significant links as determined by the transfer entropy calculations in the static network analysis, the authors chose to evaluate the dynamics of all potential suppliers to node E. This analysis would be useful to a manager of node E to determine how the information dependencies of his organization are changing with time.

Figure 12 shows the results of the local transfer entropy calculations with a moving average of 4 time units applied. Again by choosing max {A→E, B→E, C→E, D→E } for each time step, it becomes apparent how the information dependencies of node E changed with time. Early on node E relied on significant contributions from nodes A and C, later relying heavily on node C for over 10 time units. In the second half of the

time period studied, nodes A and C continued to be influential at times as well as several key spikes from node D. It is interesting to note the local transfer entropy on B→E remained low but consistent throughout this time period.
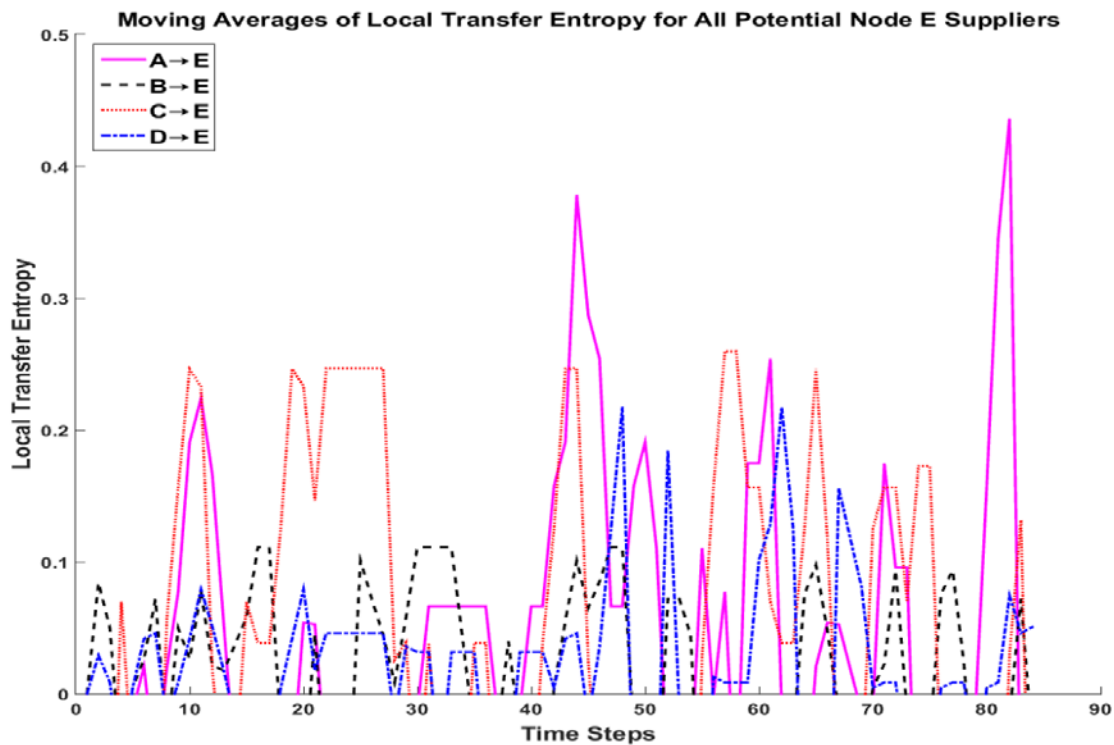
Figure 12. Local transfer entropy moving averages (4 time units) for all potential node E suppliers in real-world supply network.

Again, no truth data was available for this network so the same subject matter expert and manager were consulted to critique these results. As before, they expected nodes A and C would be the most influential early in the time period with node C's influence remaining high. They also expected node D would not have significance contributions until the second half of the time period at which point nodes A and C would

continue with spikes of significance. As was the case with the static network analysis of this graph, it surprised the experts that node B did not have higher spikes but rather remained low throughout. This was attributed to the fact that node B does not generally produce information with enough significance to directly influence node E's production. Instead, node B's production was often included in the production of nodes A, C, & D upon which node E is very dependent (as evidenced by the static network structure). Overall the subject matter experts were pleased with the analysis and were excited to apply the framework to other topic areas within their organization.

Discussion

CASNs have challenged researchers' ability to study them because of their dynamic nature and complex behaviors. Information theory and information dynamics provides a novel framework to overcome some of the difficulties associated with modeling CASNs. Specifically, transfer entropy can be used to determine the network structure while local transfer entropy provided insights into dynamics of those structures when applied to time-series production data.

First, the authors created a simple (10-node) conceptual supply network and simulated information production timing as a means of verifying the methodology. Transfer entropy calculations and their corresponding *p*-values accurately revealed the network structure based only on time-series input of the production times. Next, the authors created a dynamic conceptual supply network and again simulated production timing, also to verify the methodology. The focus of this study was to observe the local transfer entropy time series for a node which dynamically varied the inputs for its

production. A static network analysis using transfer entropy and their corresponding *p*-values was able to uncover this network structure of significant relationships while local transfer entropy calculations determined the dynamic switching behavior of the node being studied.

Finally, a real-world information supply network with actual production time-series data was studied using this information dynamic framework. A static network analysis using transfer entropy returned a network structure which was validated by subject matter experts who have experience with the topic area and management of the real-world network. Likewise, a dynamic network analysis using local transfer entropy revealed a time-series of the changing relationships in the information sharing network. Although experts validated the analysis, the analysis also provided these managers with novel and unexpected insights into their network, especially concerning strength of relationships and relative influence between nodes in the network.

In general, managers of the network had pre-conceived notions of how their network behaved, but could not completely know how the agents in the network behaved due to the agents' autonomy. The managers discovered they had made assumptions about how nodes communicated and passed information that may have been unfounded. This approach provided managers with new insights to an information network that they had assumed they understood and challenged the mechanisms that they assumed grounded the network. They assumed specific nodes to act as sources and specific nodes to act as sinks. However, this information theoretic methodology revealed that no nodes acted as pure source nodes and instead collaborated significantly among the other nodes. The dynamic network resulting from the methodology revealed time-history dependencies between

nodes that were very close to what was expected by the managers. However, in the static structure, many relationships between the nodes surprised the managers.

Based on the static network graph produced by the methodology, the managers identified specific offices in which personnel and/or process could be fostered to increase collaboration with others. They also had assumed various relational dependencies between the nodes (which nodes' information production could accurately predict another nodes' information production) which did not appear significant based on these methods. Following this analysis, they had more information available in order to make resource decisions based on actual information dependencies within the network than they did based on their prior assumptions of network behavior.

These simulations and the real-world analysis showed that by using transfer entropy it is possible to reverse engineer the supply network's structure with only time-series production data about network nodes and use local transfer entropy to analyze the network structure's dynamics. This approach required minimal information about the network. Only node histories were needed; underlying behaviors and relationships between the nodes were not needed, modeled, or assumed. Being able to uncover network structures, either static or dynamic, from real-world data overcomes two significant barriers to studying CASN: that static and dynamic structures or underlying node behaviors must be completely known, or that node behavior must be modeled or assumed.

Furthermore, applying information theory in the context of such supply networks opens up new ways of studying these complex systems by enabling other information theoretic measures such as information storage or information modification. Other

disciplines have done significant research relating information theoretic measures with network behaviors such as self-organization, emergence, stability/instability, and distributed computation, all of which remain open areas of research for CASNs. Focusing supply network research on these relationships could reveal new approaches to network management and business analytics in networks adapting to their environment.

IV. Application of Methodology to Complex Organizational Networks

Introduction

Organizational networks have long been studied as complex adaptive systems (CAS) due to their inherent formal (i.e., chain of command) and informal (i.e., communication) structural makeup. Understanding organizations through this lens is critical to managing such systems effectively. Unfortunately, the behavior of complex adaptive systems makes them more difficult to analyze and assess. Leveraging techniques from information theory and information dynamics provides additional methods for analyzing and assessing these complex systems. The notion of transfer entropy is especially useful in mapping the organizational network merely by observing the behavior of individual nodes. These techniques, however, may have limitations within certain network structures, complicating the problem of discovering the network structure accurately. The fundamental issue being addressed is how well various organizational network structures can be expected to be discovered using transfer entropy principles.

Background

Complex Organizational Networks

A complex adaptive system (CAS) is a network of dynamical elements where the states of both the nodes and the edges can change, and the topology of the network itself often evolves in time in a nonlinear and heterogeneous fashion. [3]

Anderson was perhaps the first to apply CAS research to the strategic management of organizations. [4] In his seminal work, he proposed that organizations take on the characteristics of CAS. Furthermore, he went on to prescribe methods that

53

these CAS organizations should be managed. Through his research he posited that a manager should not attempt to make sweeping enterprise-wide changes because the system's nonlinear response is too difficult to predict and control. Instead, the managers should set boundaries or constraints on the system and observe the outcome. Then they would be able to tune the system by modifying the constraints and/or changing the amount of energy allowed into the system. The primary role of a strategic organizational architect is to influence the extent of improvisation, the nature of collaboration, the characteristic rhythm of innovation, and the number and nature of experimental probes by changing structure and demography. One factor complicating the use of CAS models in strategic management is that no theory exists to help managers predict how their actions may cascade through the CAS and affect emergence. [4]

Li et al. reached similar conclusions in the context of complex adaptive supply network (CASN). They proposed a model for CASN evolution using the principles of CAS and fitness landscape theory. They modeled the evolution of the CASN by modeling the environment, the firm, and the supply network evolution. For simulating the CASN evolution they used a multi-agent architecture, interaction of agents and two different experiment designs: the first for structure dynamics of CASNs and the second for dynamic evolution of firm's fitness. Their primary take-away for managers of CASNs was that evolution is a self-organizing process and that any planning and regulation of the market and firm may be undermined by the fact that the outcomes are both open and unknowable. [23]

These complex organizational structures can be studied using a model introduced by Dodds et al. which is able to produce a spectrum of organizational networks for the

purpose of studying communication within the organization. This model begins with a tree network structure and adds links between nodes as prescribed by probabilistic functions. It is these links which form the informal organizational structure and add additional complexity to the system. The authors observed five primary network structures which emerge when varying their network building parameters: random, random inter-divisional (RID), multi-scale (MS), local teams (LT), and core periphery (CP). [49] An illustration of these network types is provided in the original article, but viewing them as adjacency matrices allows further clarification of the differences between the network structures, as in Figure 13. The random structure is assumed to be familiar to the reader and is therefore not shown; the MS structure displays attributes of each of the other structures and is, again, not shown.
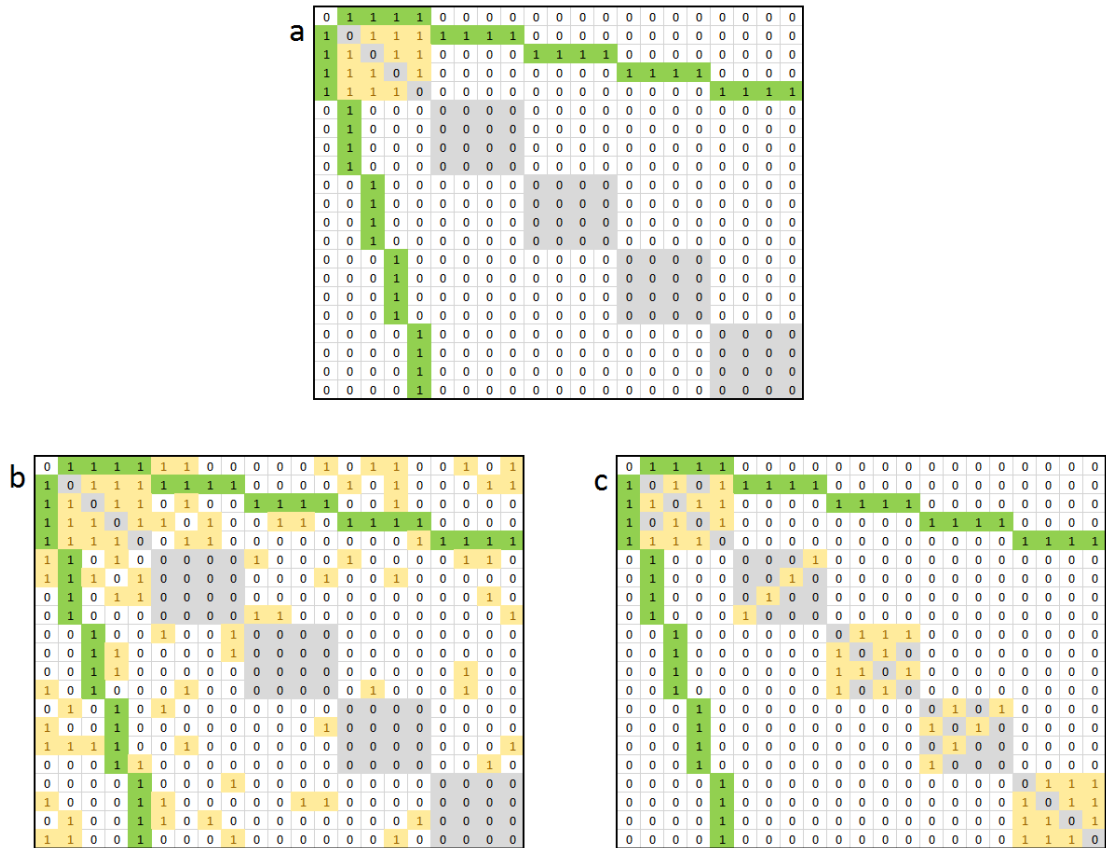
Figure 13. Example adjacency matrices for a 2-level tree with a branching factor of 4 (shown in green) with links added (shown in yellow) according to Dodd et al. demonstrating (a) core-periphery structure with links more likely within the top tier of the tree; (b) random inter-divisional structure with links more likely within different divisions (shaded in gray); (c) local teams structure with links more likely within their own division.

Information Dynamics

To address the issue of how information moves between nodes of the organizational network, we appeal to the concepts of information dynamics. Information

dynamics arises out of concepts in information theory [25] such as mutual information

(Eq. 11), and conditional mutual information (Eq. 12) between processes X, Y, and Z.

$$I(X;Y) = \sum_{x \in \alpha_x} \sum_{y \in \alpha_y} p(x,y) \log_2 \frac{p(x|y)}{P(x)} \tag{11}$$

$$I(X;Y|Z) = \sum_{x \in \alpha_x} \sum_{y \in \alpha_y} \sum_{z \in \alpha_z} p(x,y,z) \log_2 \frac{p(x|y,z)}{P(x|z)} \tag{12}$$

Transfer entropy [7] is perhaps the most applied of the information dynamics and

has been used extensively in the study of many varied complex systems including

neuroscience [26] [28], social networks [13], finance [30], and others. Lizier [25]

describes transfer entropy (Eq. 13) as the amount of information that a source process

provides about a destination (or target) process' next state in the context of the

destination's past. In this equation k and l are the history lengths of X and Y, respectively,

while n is the current time index. Other useful definitions of transfer entropy describe it

either as a measure of deviation from independence [11] or an observed correlation

between two processes rather than direct effect [12].

$$T_{Y \to X}(k,l) = I(Y_n^{(l)}; X_{n+1} | X_n^{(k)}) \tag{13}$$

Lizier created the Java Information Dynamics Toolkit (JIDT) which implements

many aspects of information dynamics including the transfer entropy calculations used

here. This toolkit also implements a method to determine the statistical significance

between the process relationships being studied. This null hypothesis testing approach determines if the transfer entropy calculated between two processes is statistically different from 0 (i.e., no relationship exists). The author first forms the null hypothesis *H0* stating there is no relationship between the processes between studied. He then uses the statistical properties of *Y*, removes any correlation with *X*, and creates surrogate measurement distributions. Knowing what the distribution for the measurements would look like if *H0* were true (i.e., the processes are each random), the *p*-value can be calculated and an actual measurement sampled from this distribution. If the test fails, the alternate hypothesis is accepted stating there is a statistically significant relationship between the processes. This test allows one to compute the transfer entropy between all possible node combinations in a network and keep only those network relationships or links that are likely to be statistically appropriate. [25]

Methods

The fundamental issue being addressed in this research concerns how well transfer entropy calculations are able to discover the various organizational network structures put forth by Dodd et al. [49] The authors extended the approach from Rodewald et al. which first created a known network structure, simulated production data (or communications) on the network, and then used only the simulated data to discover the original structure. [9]

Here, a significant number of network structures (*n=100*) were created for each of the topologies from Dodd et al. Then production/communication data was simulated on each network using the approach from Rodewald et al. Transfer entropy was calculated

for each pair of nodes in the network (in both directions) using the JIDT toolkit. ROC curves were then generated for each network structure by varying the acceptable *p*-value, and the area under the curve (AUC) was calculated for each ROC curve. The minimum, maximum, mean, and standard deviation for AUC were noted and compared for each of the network structures.

This procedure was done on networks of 156 nodes (*N=156*) with a branching factor of 5 (*b=5*), first with 50% added edges (*m=78*), and then with 100% added edges (*m=156*). For the core-periphery networks, only 20 nodes were added (*m=20*) due to the additional computational time required to add links outside of the top tier of the network. A basic tree (*m=0*) was also studied as a baseline using the same techniques.

Analysis

Table 8 shows the results of the AUC analysis of transfer entropy calculations on the various network structures with both 50% added edges (*N=156, m=78*) and 100% added edges (*N=156, m=156*). Figure 14 presents minimum and maximum ROC curves for each of the network structure types only for the case of 50% added edges. In the baseline case of a tree with *m=0* added edges, the resulting ROC curves had an AUC of *0.93-0.99*. To bound the range of *p*-values that would be used to threshold transfer entropy values in practice, two cases were considered. The first case attempts to eliminate false positives; the second case attempts to maximize true positives while still keeping the false positive rate within approximately 20%. Table 9 shows recommended *p*-values of between *p<0.001* and *p<0.07* for all network structures. For the baseline case of the tree, this results in true positive rates between *0.80-1.00* and false positive rate between

0.00-0.17. The *p*-value can be adjusted between these values to refine the trades between

the rates for true positives and false positives. For most applications these results are

probably sufficient.

Table 8. AUC results of network structures with *m=78* and *m=156*: minimum, maximum,

mean, and standard deviation.

(N=156, m=78)	Area Under Curve (AUC)				(N=156, m=156)	Area Under Curve (AUC)			
Structure	Min	Max	Mean	Std Dev	Structure	Min	Max	Mean	Std Dev
Random	0.854	0.959	0.909	0.022	Random	0.734	0.864	0.803	0.029
LT	0.760	0.920	0.850	0.035	LT	0.607	0.796	0.713	0.032
MS	0.842	0.959	0.899	0.024	MS	0.653	0.856	0.781	0.034
CP (m=20)	0.848	0.942	0.887	0.020	CP (m=20)	0.848	0.942	0.887	0.020
RID	0.854	0.967	0.906	0.024	RID	0.758	0.918	0.821	0.028
Tree (m=0)	0.933	0.992	0.971	0.011	Tree (m=0)	0.933	0.992	0.971	0.011

Additionally, Table 8 shows that the LT network structure is the most difficult to

discover using transfer entropy principles. CP and RID networks appear to be yield more

accurate results with random networks close behind. However, in the case of CP

networks, the observed effects could be a result of only adding 20 edges, as adding edges

appears to lower the accuracy in general as observed by comparing the *m=78* and *m=156*

cases.
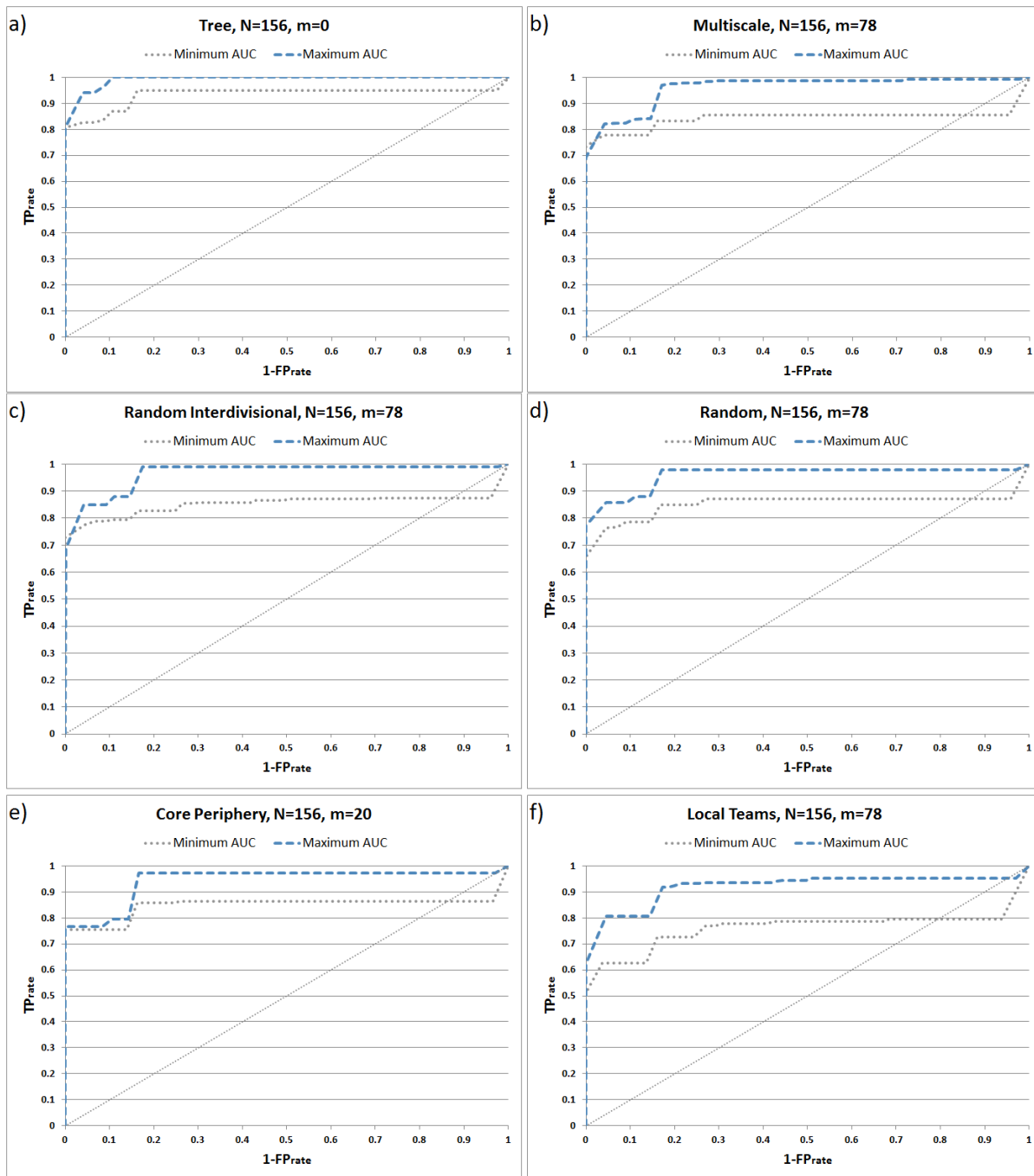
Figure 14. ROC curves of min and max AUC for various organizational network

structures: (a) tree; (b) MS; (c) RID; (d) random; (e) CP; (f) LT.

Conclusions

Organizational networks have long been posited to take on the behaviors of complex adaptive systems. Because of these complex behaviors, it is often desirable to study the informal organization of these networks rather than a strictly formal org-chart representation of them. The informal organization, which could be developed using communication and/or production data, can take on various structure topologies from a more formal tree network to a completely random network. Transfer entropy from information theory provides a way of discovering these informal organizational structures and gain insight into the complex behaviors of these networks.

Using a model of building organizational structures from Dodds et al., various networks were created to simulate the primary organization topologies: random, LT, MS, CP, and RID. Production/communication data was simulated on these networks, and then transfer entropy was used to discover the original network. ROC curves were generated for each of the network types by varying the accepted p-values of transfer entropy and compared using an AUC metric.

More edges added to the baseline tree structure appeared to make it more difficult to accurately discover the underlying network solely from production/communication data. Additionally, if these edges tended to clump closely together (as in local teams), this tended to increase the difficulty as well. Those networks with edges connecting more distant nodes (as in random inter-divisional) tended to be easier to discover.

Table 9 gives recommendations for p-values when a manager may be trying to discover an informal organization structure. The manager should be aware of the expected rates of true positives and false positives and adjust their p-values according to

their specific problems. In the case of *p<0.001*, the false positive rate is practically zero, but the true positive rate would be expected to range from *0.50-0.80* depending on the network type present. In the case of *p<0.07*, true positive rates drastically increase from *0.75-1.00* but come with an approximately *0.16* rate of false positives. These expected values should inform the manager and bound their confidence in organizational networks discovered using these transfer entropy techniques.

Table 9. Recommended *p*-values to 1) minimize false positives and 2) maximize true positive rate while maintaining relatively low (*<20%*) false positive rates.

	P-value Recommendation #1						P-value Recommendation #2			
Structure	P-value	Min Exp TP rate	Max Exp TP rate	Min Exp 1-FP rate	Max Exp 1-FP rate	Structure	P-value	Min Exp TP rate	Max Exp TP rate	Min Exp 1-FP rate	Max Exp 1-FP rate
Random	<0.001	0.66	0.77	0.00	0.00	Random	<0.07	0.85	0.98	0.16	0.17
LT	<0.001	0.51	0.62	0.00	0.00	LT	<0.07	0.73	0.92	0.16	0.17
MS	<0.001	0.69	0.73	0.00	0.00	MS	<0.07	0.83	0.97	0.16	0.17
CP	<0.001	0.75	0.77	0.00	0.00	CP	<0.07	0.86	0.97	0.16	0.17
RID	<0.001	0.69	0.73	0.00	0.00	RID	<0.07	0.83	0.99	0.16	0.17
Tree	<0.001	0.80	0.81	0.00	0.00	Tree	<0.07	0.95	1.00	0.16	0.17

V. Analysis of Information Transfer and Storage in Complex Organizational Networks

Introduction

It has long been noted among researchers that organizations form complex adaptive systems consisting of both formal and informal structures in a network. The exact topologies of the structures have organizational consequences that are difficult to predict. Managing these complex networks creates additional challenges because the effects of intervention can result in even more unpredictable outcomes.

Noting that two primary functions of organizational networks are to transfer information between nodes and store information in the network, this article studies the effects of node communication behavior on the ability of the network to accomplish these tasks. Information theory measures are used to quantify the ability of organizational network topologies to both transfer and store information. Similarly, quantitative aspects of communication between formal and informal networks are explored by maintaining a comparison to a baseline tree network.

Background

Complex Organizational Networks

A complex adaptive system (CAS) is a network of dynamic elements where the states of both the nodes and the edges can change, and the topology of the network itself often evolves over time in a nonlinear and heterogeneous fashion. [3] This interconnected network of multiple entities (or agents) exhibits adaptive changes in response to both the environment and the system of entities itself. [22] Collective system performance or

behavior emerges as a nonlinear and dynamic function of the large number of activities made in parallel by interacting entities. [24]

Anderson [4] was one of the first to apply CAS research to the strategic management of organizations. In his seminal work, he proposed that organizations take on the characteristics of CAS. Furthermore, he went on to prescribe methods by which these CAS organizations should be managed. Through his research he posited that a manager should not attempt to make sweeping enterprise-wide changes because the system's nonlinear response is too difficult to predict and control. Instead, the managers should set boundaries or constraints on the system and observe the outcome. Then they would be able to tune the system by modifying the constraints and/or changing the amount of energy allowed into the system. The primary role of a strategic organizational architect is to influence the extent of improvisation, the nature of collaboration, the characteristic rhythm of innovation, and the number and nature of experimental probes by changing structure and demography. One factor complicating the use of CAS models in strategic management is that no theory exists to help managers predict how their actions may cascade through the CAS and affect emergence.

Li et al. [23] reached similar conclusions in the context of complex adaptive supply networks (CASN). A supply network is fundamentally a CAS because they exhibit adaptivity and can exist in a complex environment with myriad relationships and interactions. [24] Li et al. proposed a model for evolution using the principles of CAS and fitness landscape theory. They studied the evolution of the organization by modeling the environment, the firm, and the supply network evolution. For simulating the evolution they used a multi-agent architecture, interaction of agents and two different design

65

experiments: the first for structure dynamics of the networks and the second for dynamic evolution of a firm's fitness. Their primary recommendation for managers was that evolution is a self-organizing process and that any planning and regulation of the market and firm may be undermined by the fact that the outcomes are both open and unknowable. [23]

Continuing research into complex organization networks has been focused along two primary paths: agent-based models and social network analysis. Examples of agent-based approaches are found in [50] [51] [52] [53] [54]; examples of social network analysis approaches are found in [55] [56] [57] [58] [59] [60]. Other modeling options include systems dynamics and deterministic models. Surana et al. [3] posited that successful modeling efforts of large-scale CASN would require a solid empirical foundation.

Dodds et al. [49] introduced a model to study complex organizational structures across a spectrum of topologies for the purpose of studying communication within an organization. This model begins with a tree network structure and adds links between nodes as prescribed by probabilistic functions. It is these links which represent the informal organizational structure and add additional complexity to the system. They observed five primary network structures which emerge when varying their network building parameters: random, random inter-divisional (RID), multi-scale (MS), local teams (LT), and core periphery (CP).

In relation to their model, Dodds et al. discussed the distinction between a formal and informal network. In their model, the initial tree structure formed the hierarchical formal organization, while the added edges used to create the various topologies form the

informal organization structure. [49] Both the formal and the informal organizational elements generate patterns of interactions through which organizational actors coordinate effort, share goals, exchange information and access resources that affect performance. [20]

This concept was also studied by Soda et al. [1] and tested in context of a financial services organization. They found that overlap between the formal and informal network created a positive performance effect. However, inconsistency between the formal and informal networks exerted two direct and opposing effects on performance: it negatively impacted performance by reducing coordination, but also improved performance by enhancing access to potentially valuable and diverse information, ideas and knowledge dispersed within the organization.

One major contribution of this article is the use of information theory measures to quantify the ability of Dodd et al.'s organizational structure topologies to both transfer and store information within the network types. This allows the risks to be quantified with respect to increased or decreased node performance in a network and observe the change in ability of the network to transfer or store information. Furthermore this article provides a model to quantitatively test Anderson's early assertion that changing the amount of energy allowed into the system is a major control mechanism for the organization. Similarly, this article explores quantitative aspects of communication between formal and informal networks by maintaining a comparison to a baseline tree network. These effects are quantified in terms of information transfer and information storage as two primary behavior goals of an organizational network.

Information Dynamics

Information theory [6] is a common tool used in the study of complex systems. It has been applied extensively in many fields to infer network structure and relational dependencies between entities as well as to identify emergent behavior of complex systems in: neuroscience [26] [27] [28] [5], social networks [13] [43], medicine [29], finance [30], climate [31] [32], transportation [33], and biology [44] [45], just to name a few.

Lizier et al. [11] proposed a framework for describing distributed computation of a complex system, such as an organization, using information storage, transfer, and modification at each point in the network. This research focuses on two information theoretic measures: information transfer and information storage. These measures will allow quantification of a network's or organization's ability to store information in nodes and transfer information between nodes. Further, these measures will allow for insight into self-organization [61] [62], which is a key feature of complex adaptive systems.

Transfer entropy is a means of measuring information transfer. First proposed by Schreiber [7], the transfer entropy between a source entity and a destination entity is the information provided by the source about the destination's next state that was not contained in the past of the destination. Other useful interpretations of transfer entropy describe it as a measure of deviation from independence [11] or an observed correlation (or influence) between two processes. Transfer entropy measures the directed, dynamic flow of information. [12]

The transfer entropy between a source process Y and a target process X at a time n, denoted by $T_{(Y \rightarrow X)}(k,l)$, is the expected amount of mutual information between the past

of process Y (with a history length of l) and the past of process X (with a history length of k). This measures how well these processes are able to predict the future state of X, denoted by X_{n+1}.

$$T_{Y \to X}(k, l) = I\left(Y_n^{(l)}; X_{n+1} \middle| X_n^{(k)}\right) \tag{1}$$

The active information storage (AIS) of an entity in a network is the amount of information in its past that is relevant to predicting its future. AIS is the stored information that is actively being used to compute the next state of the entity. [10] The active information storage of a process X, denoted by $A_X(k)$, is the expected mutual information between the past of process X (with a history length of k) and the future state of X, denoted by X_{n+1}.

$$A_X(k) = I\left(X_n^{(k)}; X_{n+1}\right) \tag{2}$$

Lizier et al. [40] analytically determined those structures in complex networks which naturally gave rise to the active information storage mechanism. Although Lizier et al. only considered structures up to a node size of three, they were able to show that directed cycles and feedforward loop structures dominated the mechanisms that created active information storage in a node. Directed cycles allowed a node to send information to its neighbors and receive it back at a later time. Feedforward loops captured dual paths

of different lengths in which the information is received by the node at different time steps. They also noted a strong correlation with the clustering coefficient of the network.

Lizier [25] has implemented a method within the Java Information Dynamics Toolkit (JIDT) to determine the statistical significance of the process relationships being studied. To determine if the transfer entropy calculated between two processes is statistically different from 0 (i.e., no relationship exists), the JIDT forms a null hypothesis H_0 which states there is no relationship between the processes between studied. It then uses the null hypothesis to create distributions of surrogate measurements using the same statistical properties of Y with any potential correlation with X removed. Knowing what the distribution for the measurements would look like if H_0 were true, it calculates a p-value, sampling the actual measurement from this distribution. If the test fails, the alternate hypothesis is accepted stating there is a statistically significant relationship between the processes.

Methods

Network Building Approach

To create various organizational network structures to study, the algorithm proposed by Dodds et al. [49] is employed. This algorithm began with a simple tree and added edges between nodes based on probabilistic distributions set by two dimensionless parameters: ζ and λ. These parameters represented characteristic lengths of two nodes' organizational distance and their lowest common ancestor, respectively. They were used to probabilistically add links in the network, but the parameters do not have any other interpretation within the resulting network. By varying these parameters Dodds et al.

70

created the five network structures, three of which are shown in Figure 15. In addition to these three structures, the algorithm also created random (R) and multi-scale (MS) topologies. Random networks should be familiar to the reader; multi-scale networks portray characteristics of each of the other network topologies.
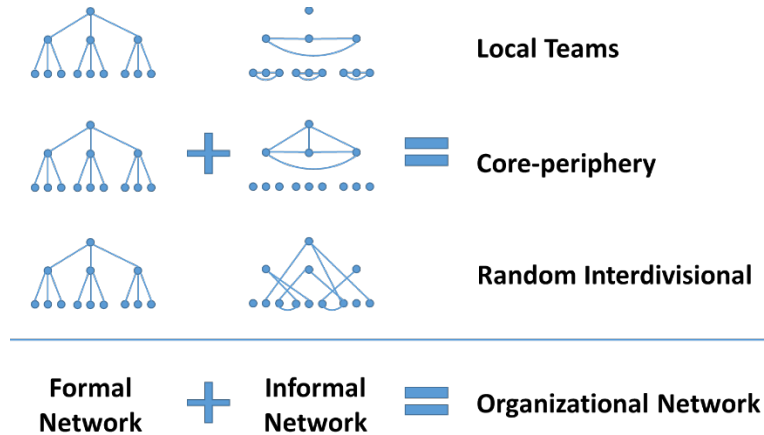
Figure 15. Dodds et al. algorithm's use of formal and informal networks to create organizational networks with random interdivisional (RID), core-periphery (CP), and local teams (LT) topologies.

For this analysis, the values shown in Table 10 were used to create the various organizational structures. Additionally, all network types began with a 156-node tree ($N=156$) and branching factor of 5 giving 4 levels which formed the formal network. An additional 78 edges ($m=78$) were added to form the informal network probabilistically based on Dodds et al.'s algorithm to arrive at the final network structure. Further, this process was repeated 150 times ($n=150$) to create a statistically significant number of trials for each network type.

71

Table 10. Values used for λ and ζ in Dodds et al.'s algorithm to create organizational

network structures.

Structure	λ	ζ
Random Interdivisional	0.1	0.9
Random	0.9	0.9
Multi-scale	0.5	0.5
Core-Periphery	0.1	0.1
Local Teams	0.9	0.1

Communication Simulation

After creating the network structures, simulated communication on these

networks was simulated. In contrast to Dodd et al's approach to studying communication

and distributed problem solving, this research used a communication simulation method

representative of those used to study communication and distributed computation with

complex adaptive systems using the information theory concepts discussed above. [9] [2]

The leaf nodes of the original tree were identified as source nodes. It is implied

that these source nodes are connected to the network's environment and create

information based on observations from the environment. The nodes randomly originated

communications and passed them to each node to which they shared a connection. The

amount of communication generated was determined by the communication rate. For

example, a 20% communication rate meant that at any time, t, a source node would have

a 20% probability of generating a communication. In turn, these nodes received that

communication and passed it to each node to which they shared a connection. In this way, communication flows from the original sources (leaf nodes) to the top of the tree.

Additionally, each node was given a communication throughput or effectiveness value, which is also generalized as the *node performance* for the purpose of this research. Communication throughput essentially acted as a filter through the node. For example, a 50% communication throughput indicated that each message a node received had a 50% probability of getting passed along by that node; otherwise the information was dropped. This communication throughput could represent multiple facets of communication in the network. First, it could represent a natural thresholding of information as it moved through the network. For example, a node may receive communication and deem that it is not significant enough to process and pass to its connections. Second, the communication throughput could represent the processing capacity of a node. A node with greater capacity (more resources) to process information would likely have a higher communication throughput. Therefore, by simulating a decrease in communication throughput, the node is modeled to have a decrease in capacity (through resources allocation or otherwise).

A new communication pattern was simulated for each of the 150 network structures *(n=150)* for a particular organization type. Therefore, observed variations in the average, maximum, and minimum properties of these cases are due to both the variation in network structure and the randomness inherent to the communication patterns. The exception to this is the baseline tree structure. There is only one way to create a 156-node tree, so the structure is common to all *n=150* cases. The only variation is the change in communication patterns.

73

Case Simulations

Any change in communication behavior in an organizational network could result in both local and global effects. Therefore, case studies were chosen to observe both local and global effects resulting from increases and decreases in node performance or throughput.

Local Effects

For the purposes of this analysis, local effects were considered to be the effects of transfer entropy on a node's direct input or output. The local effects would tell a manager of the network what risk or opportunities to information transfer or storage would be expected to result if node performance was increased or decreased, either by natural means or forced intervention, by modifying its communication effectiveness.

To study the sensitivity of transfer entropy coming in to a node, the number of inputs and communication rate of those inputs were varied. Baseline source communication rates of 20%, 50%, and 80% were chosen to model the baseline information load on the network. Between 1 and 4 inputs were studied with uniform communication throughputs ranging from 0-100%.

The p-values of the transfer entropy values input to the node were also measured. This case addresses the question of how much can the throughput of a node be decreased before its transfer entropy contribution is no longer significant in the network.

It should be pointed out that this modeling of inputs to a node can also be applied to spontaneous information generation within the node. For example, if a node can act as a source and generate its own information without receiving that communication from an input, this can be modeled using another input node that does not appear in the formal

network. So a generating node with 2 formal network inputs can be modeled as a node with 3 inputs with one of those inputs acting as a ghost source for information generation within that node.

To study the effects of transfer entropy leaving a node, the same cases were used except the output transfer entropy was calculated assuming a perfectly communicating receiving node. Baseline source communication rates of 20%, 50%, and 80% were chosen to model the baseline information load on the network. Between 1 and 4 inputs were studied with uniform communication throughputs ranging from 0-100%. The receiving node had a communication throughput of 100%. This case addresses the question of how much input is too much for a node after which point its transfer entropy shows a diminishing marginal increase.

For each of the local effects cases, this analysis used simple network graphs. Because these structures were simple, no active information storage was expected and indeed its value was zero for all cases. As discussed previously, information storage is only observed with certain network structures including directed cycles and feed-forward loops, none of which were present in these cases. [63]

Global Effects

The *global effects* were studied by analyzing the sum of transfer entropy across all edges in the network. The total sum of active information storage was also analyzed across all nodes in the network. This allowed us to quantify the overall network's ability to store information in nodes and transfer information between nodes and compare these values across organizational network types. In the literature, these are considered the two

primary functions of a complex network that distributes computation, and these are the two primary measures of these functions.

The nodes in the network were labeled based on their original tree structure. Source nodes were those which appeared as leaves in the initial tree used to build the network. The sink node was the top-most node of the initial tree. The intermediate nodes (or mid nodes) were the remaining nodes. The communication throughput for each type of node was increased or decreased from a baseline network-wide 50%, meaning half of any messages received were passed on as output of a node.

Each network type was analyzed with a baseline case of 50% throughput on that same network type. This was done so that a manager of a certain type of organizational network could understand what risks or opportunities to information transfer or storage may result from changing resource loads within his network. A similar comparison could have been performed between network types, but the matrix was too large for this analysis.

Table 11 shows the communication throughputs for each of the cases considered. Case 1 quantifies a risk to information transfer and storage of the network if the source nodes in a network decreased their throughput. Case 2 is its counterpart asking what opportunities to information transfer and storage of the network if the source nodes increased their communication throughput. Case 3 and 4 do the same for intermediate nodes. What risks or opportunities to information transfer and storage appear if the intermediate nodes in a network decreased or increased their throughput? Finally, cases 5 and 6 address management reactions to the risks observed in cases 1-4. Is it possible for a network manager to reduce the realized risk by "plussing-up" or adding resources to

either source or intermediate nodes if the other's communication throughput has been decreased?

Table 11. Communication throughput values (%) used for a baseline case and 6 cases to study global effects on transfer entropy and information storage within each organizational network structure.

	Communication Throughput		
Case	**Source Nodes**	**Mid Nodes**	**Sink Node**
Baseline	**50%**	**50%**	**50%**
1	**20%**	50%	50%
2	**80%**	50%	50%
3	50%	**20%**	50%
4	50%	**80%**	50%
5	**20%**	**80%**	50%
6	**80%**	**20%**	50%

It should be noted that the sink node's throughput was not modified in these cases. This was done for two reasons. First, in a 156-node network the overall effect on total transfer entropy or active information storage for a single node is minimal. Second, the effects of modifying the sink node's behavior are predictable based on the results of the local effects analysis above.

Analysis

    Local Effects

Local effects on transfer entropy and p-values due to modifying node behavior are shown in Figure 16. Each plot varies the communication throughput of the node under observation from 0-100% and graphs the associated transfer entropy and p-values of the input connections to the node. The graphs show the effects of 1-4 input connections and 20%, 50%, and 80% fundamental communication rate from the sources.
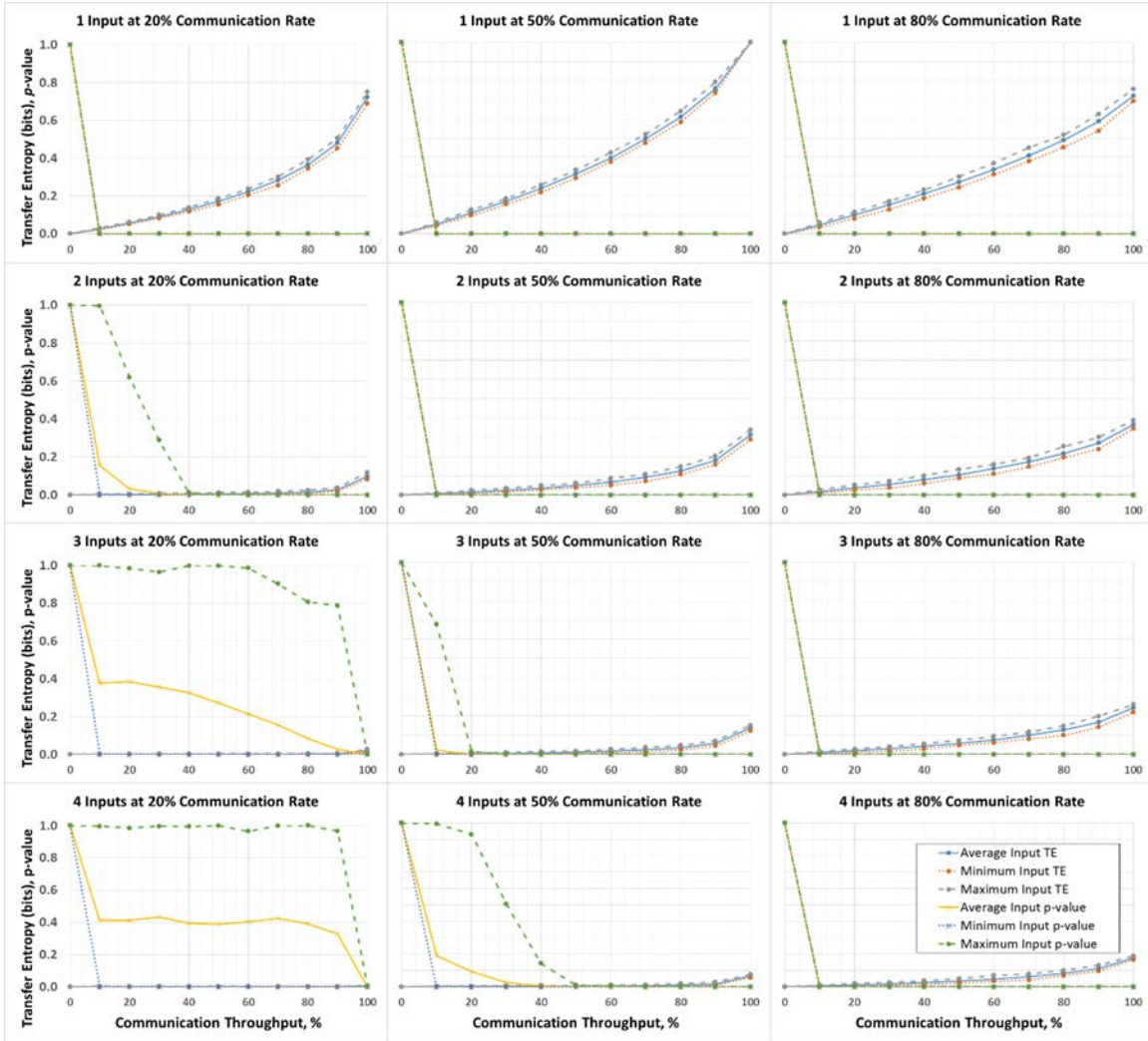
Figure 16. Transfer entropy values and *p*-values for input connections to node.
Plots vary fundamental network communication rate and number of inputs to node.

In general, several trends are noted. First, increasing the communication

throughput results in increased transfer entropy from the source nodes. Second,

increasing the number of input connections decreases the transfer entropy on each input

link. Third, lower communication throughput results in higher *p*-values, especially in the

cases of lower communication throughput or a large number of input connections. However, when *p*-values are high, the false transfer entropy values on those edges are very small (approximately *TE < 0.001*).

This last trend is what allows the transfer entropy to be summed over the entire adjacency matrix. Any connections that are not significant will tend to have insignificant values of transfer entropy so thus, their contribution to the overall sum will be negligible.

In the case of the outgoing connection from the node under observation, the simulation results are shown in Figure 17. Again, each plot varies the communication throughput of the node under observation from 0-100% and graphs the associated transfer entropy of the input connections to the node. The graphs show the effects of 1-4 input connections and 20%, 50%, and 80% fundamental communication rate from the sources.
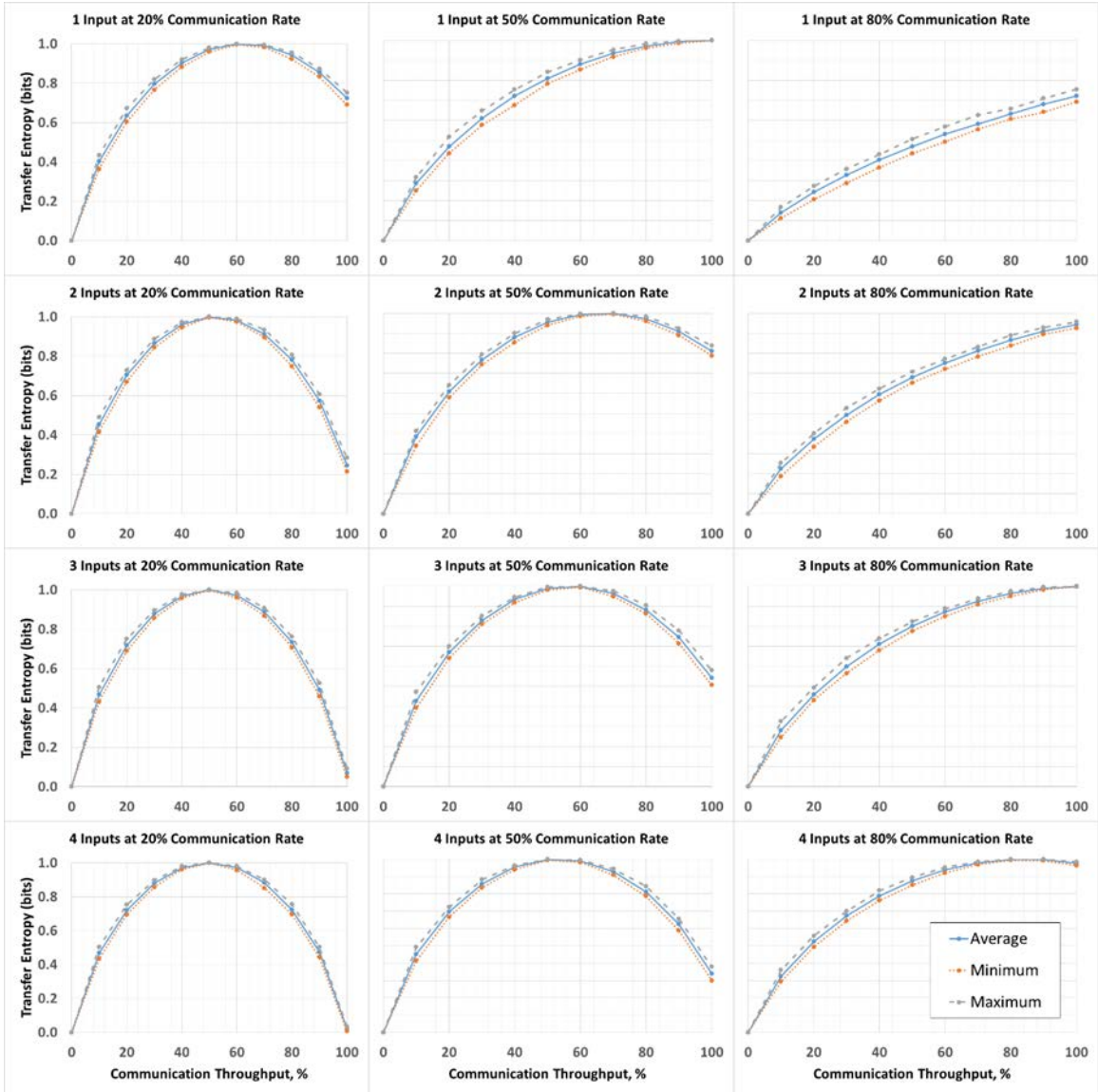
Figure 17. Node output transfer entropy sensitivity with various inputs and communication rates.

The simulation results of the output are more difficult to interpret. In general, the transfer entropy responses are harder to predict. Increasing communication throughput can either increase or decrease the transfer entropy value of the outgoing connection

81

based on where it is starting. Likewise, increasing the number of input connections can either increase or decrease the output transfer entropy based on where it is starting.

One trend that can be noted is that increasing the number of input connections tends to shift the point of maximum transfer entropy to the left on communication throughput. Another trend is that decreasing the fundamental communication rate of the input while keeping the number of input connections the same also shift the point of maximum transfer entropy to the left.

The difficulty in predicting local transfer entropy effects in a network due to changes in the node behavior may make it more useful to consider the global effects of changes in a network instead.

Global Effects

Transfer Entropy

Global effects on transfer entropy due to modifying node behavior are shown in Figure 18. The plots cover the 6 cases detailed in Table 11.

Case 1 studied the risk to network information transfer of reducing the communication throughput of the source nodes to 20% compared to the baseline case of 50%. The effect on each of the organizational network types is shown in Figure 18. Likewise, case 2 studied the opportunities to information transfer by increasing the communication throughput of the source nodes to 80% compared to the baseline case of 50%. In case 1, total transfer entropy in the network increased marginally for all cases except the basic tree showing no risk to information transfer. In case 2, however, the total transfer entropy in the network increased significantly: around 80% on average for most

networks. This shows an opportunity to increase information transfer in the system. However, the core-periphery network only saw a modest 25% increase in case 2.

Case 3 studied the risk to information transfer of reducing the communication throughput of the intermediate nodes to 20% compared to the baseline case of 50%. Likewise, case 4 studied the opportunities to information transfer by increasing communication throughput of the intermediate nodes to 80%. The reduced throughput of the intermediate nodes resulted in a risk of approximately 5% decreased total network transfer entropy for most networks. The primary exception was the core-periphery network which experienced nearly 25% decrease in total network transfer entropy. In contrast, increasing the throughput of the intermediate nodes resulted in an opportunity of approximately 20% increase in total network transfer entropy for most structures. Again, the exception was the core-periphery network which saw a higher risk with more than a 20% decrease.

Cases 5 and 6 were designed as compensatory studies. If the source nodes decreased communication throughput to 20%, could the intermediate nodes compensate for that loss of information transfer by increasing their throughput? Likewise, if the intermediate nodes decreased communication throughput to 20%, could the source nodes compensate for that loss of information transfer by increasing their throughput?
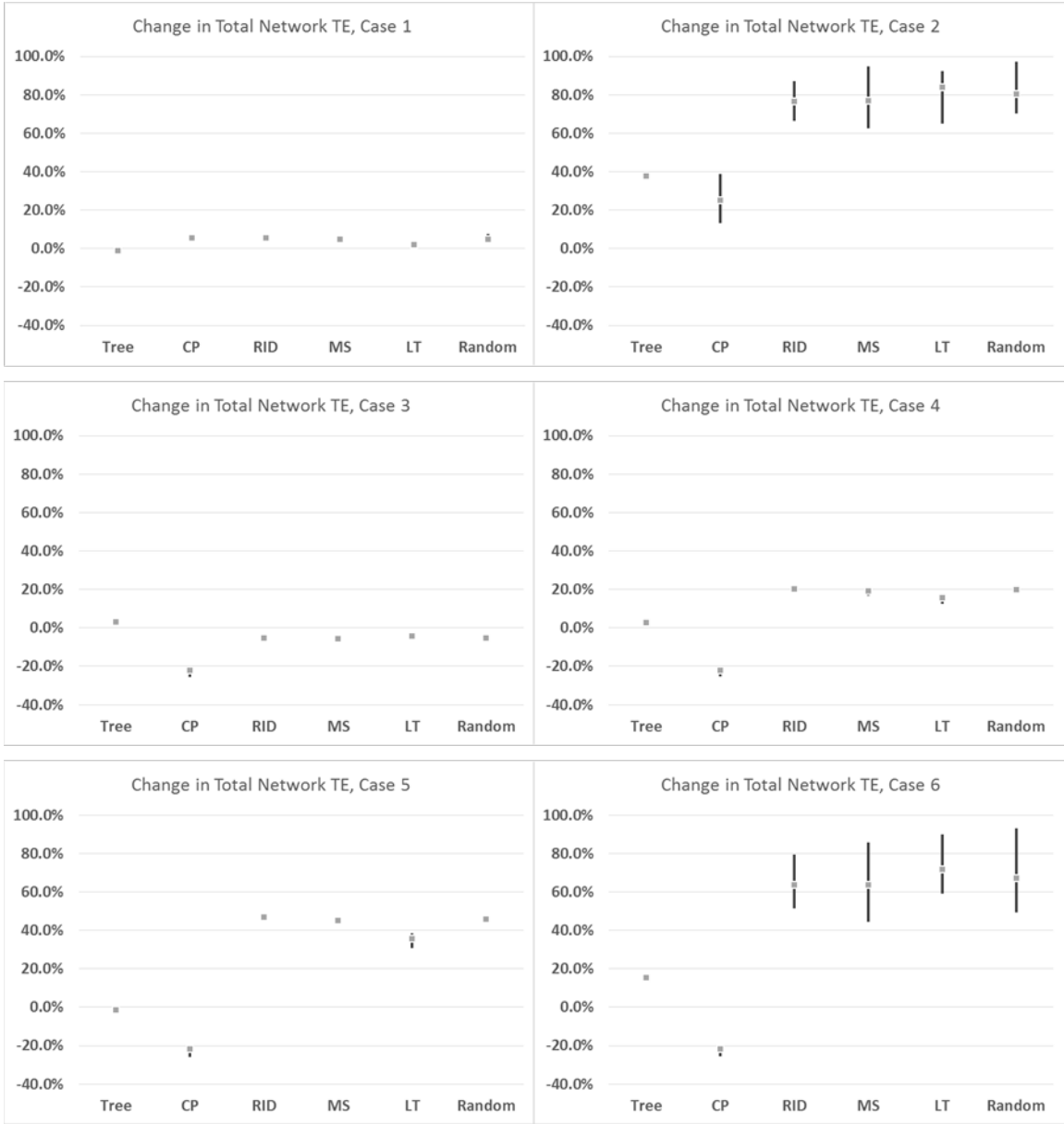
Figure 18. Percent change in transfer entropy from baseline case for various organizational network structures in each simulated case. See Table 11 for description of cases.

Case 5 should be compared to case 1. In case 1 the source nodes had a communication throughput of 20%. In case 5 the intermediate nodes have compensated for this decreased throughput in the sources by increasing their own throughput. The results show that in most network structures it was possible for the total transfer entropy in the network to recover from the risk realized by the source nodes' decreased throughput. This was done by the intermediate nodes increasing their throughput. Again, the exception was the core-periphery network which showed significantly decreased total network transfer entropy. It can also be seen that case 5 resulted in a higher overall network transfer entropy than even case 4 which increased the communication throughput of intermediate nodes without decreasing the throughput of the source nodes.

Case 6 should be compared to case 3. In case 3 the intermediate nodes had a communication throughput of 20%. In case 6 the source nodes have compensated for this realized risk of decreased information transfer by increasing their own throughput. The results show that in most network structures it was possible for the realized risk of information transfer as measured by decreased total transfer entropy in the network to recover when the intermediate nodes decreased throughput. This was done by the source nodes increasing their throughput. Again, the exception was the core-periphery network which showed significantly decreased total network transfer entropy. However, it can also be seen that case 6 was not able to recover overall network transfer entropy to the levels observed in case 2 which increased the communication throughput of source nodes without decreasing the throughput of the intermediate nodes.

Active Information Storage

Global effects on active information storage due to modifying node behavior are shown in Figure 19. The plots cover the 6 cases detailed in Table 11.

Case 1 studied the risk to information associated with reducing the communication throughput of the source nodes to 20% compared to the baseline case of 50%. The effect on each of the organizational network types is shown in the plot. Likewise, case 2 studied the risk to information storage of increasing the communication throughput of the source nodes to 80% compared to the baseline case of 50%. In case 1, total active information storage in the network was fairly neutral for most cases with average increases near 0%. Some networks did show higher and lower maximum and minimum values of +/-10%. In case 2, the total active information storage in the network again remained near 0% for most networks. However, the local teams network saw a significant opportunity to information storage with an increase of 75%.

Case 3 studied the risk to information storage due to reducing the communication throughput of the intermediate nodes to 20% compared to the baseline case of 50%. Likewise, case 4 studied the risk to information storage due to increasing communication throughput of the intermediate nodes to 80%. Both cases showed marginal impact to total active information storage in the networks. Again, the exception was the local teams network which saw nearly a 15% increase in active information storage.

Cases 5 and 6 were the compensatory cases. Either the intermediate or source nodes decreased their throughput to 20% and then the other attempted to compensate by increasing their throughput to 80%. Overall, these cases showed little variation with the average change in active information storage staying near 0%. However, in case 6 the

local teams network structure saw an increase of nearly 40%. This is similar to the effect

observed in case 2 in which increasing the communication throughput of the source

nodes resulted in significantly higher total active information storage in the network.
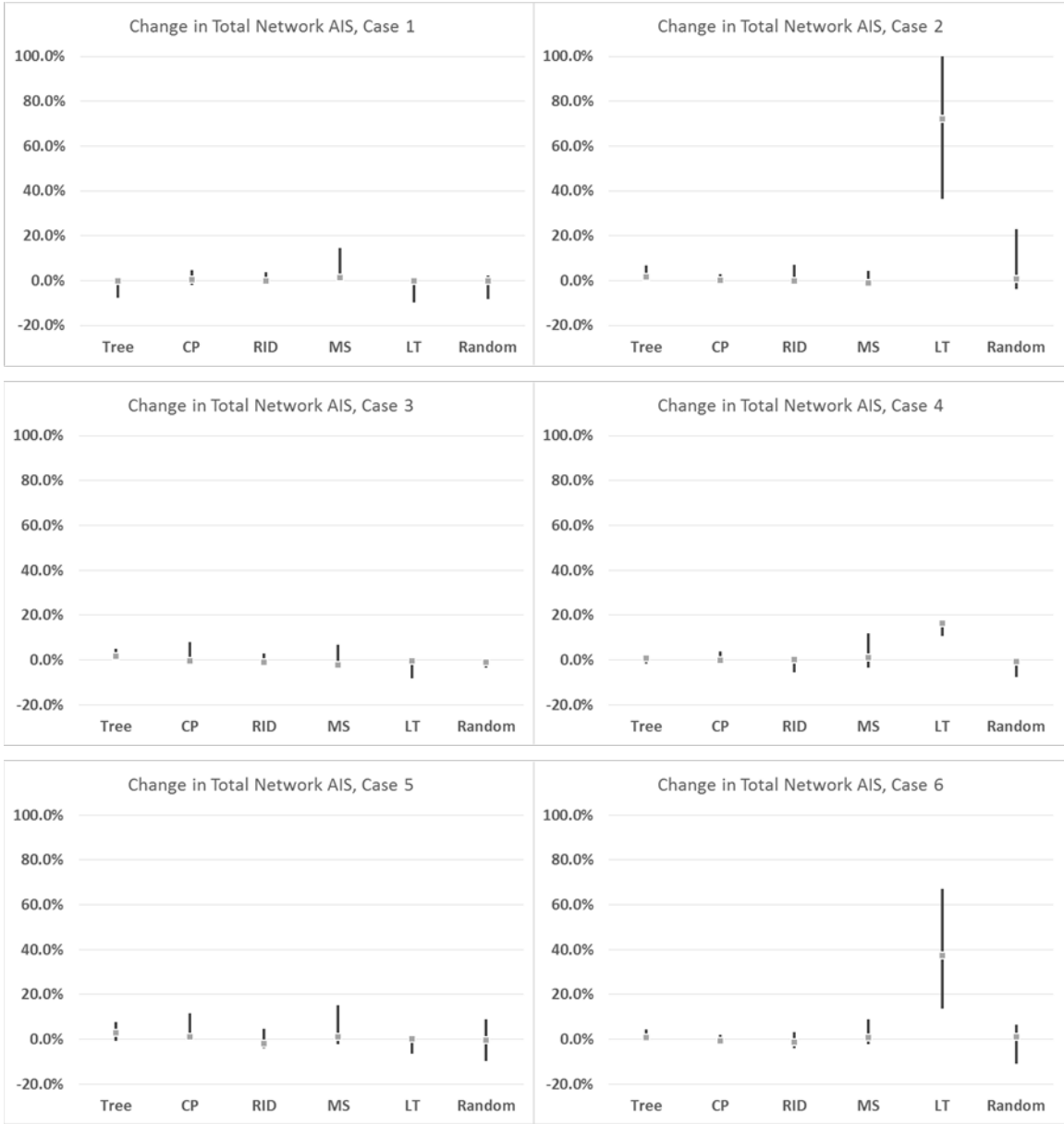
Figure 19. Percent change in active information storage from baseline case for various organizational network structures in each simulated case. See Table 11 for description of cases.

Discussion

Information transfer and information storage were used to quantify a network's ability to store information in nodes and transfer information between nodes. Given that organization performance is largely impacted by actors in an organization using formal and informal networks to coordinate effort, share goals, exchange information, and access resources [20], the measures of information transfer and information storage provided quantification of how organizational network topologies differed in performance.

When considering a network's ability to transfer information between nodes, in most cases studied, the informal organization contributed positively to that goal. This was seen by comparing the various organizational structures to a baseline tree network. One exception to this was the risk realized when intermediate nodes decreased their communication throughput. Another exception is core-periphery networks across nearly all cases which generally saw a decrease in information transfer capabilities. In general, the core-periphery network did not demonstrate significant ability to transfer information within the network. This seems to indicate that an organization with a tightly connected upper echelon has little ability to influence information transfer within the network. This could also provide a quantitative example to Soda et al.'s [1] empirical observation that informal connections which do not provide access to information are essentially a distraction in the organization. This upper echelon of actors stores relatively little information compared to the rest of the network and therefore informal connections between these actors are unlikely to yield a positive result on the organization's ability to process information.

Another effect that can be seen supports Anderson's [4] qualitative observation that a major lever to manage complex organizations is to control the amount of energy which enters the system. This is clearly seen in the simulated case in which the source nodes increase their communication throughput to 80% from their baseline of 50%. It is these nodes which interface with the environment, so a higher communication throughput results in more energy entering the system. Indeed, the network's ability to transfer information between nodes increased by more than 75% when the source nodes' communication throughput increased. However, in the case of decreased communication throughput by the source nodes (less energy entering the system) it was still possible to compensate and realize a nearly 50% increase in information transfer capability by increasing the communication throughput of the intermediate nodes.

This is further evidenced in the case of local teams networks and their ability to store information. Increasing the amount of energy entering a system by means of increasing the source nodes' communication throughput significantly increased the network's ability to store information when informal local team connections dominated the network structure. These appear to be direct cases which quantify the empirical observation made by Soda et al. [1] that the overlap of the authority network and the informal advice and information networks created a positive performance effect when hierarchically-ordered coordination (i.e. local teams) was involved. These local teams structures may have also improved performance by enhancing access to potentially valuable and diverse information, ideas and knowledge dispersed within the organization.

The above observations are evidence of Anderson's [4] assertion that managing a complex organization is best done at the macro or global level. To provide evidence for

90

his counter claim that micromanaging these networks yields unpredictable results, the local effects of information transfer should provide fair warning. In these results, increasing the communication throughput of a single node could either increase or decrease the ability of that node to transfer information in or out depending on its initial conditions. Likewise, increasing the number of connections to a node (forcing informal links) also had unpredictable results, again depending on the node's initial conditions. Further, adding too many input connections decreased information transfer significantly.

The information theory basis of the model used in this research provides quantification of network effects and support for qualitative network observations proposed by other researchers. The quantification of information transfer and storage in organizational network topologies allows for the study of the risks of decreased communication and opportunities of increased communication in these networks. By maintaining a baseline comparison with a simple tree, these information theory measures quantify the performance differences between the formal and informal organizational network.

VI. Conclusions and Recommendations

Chapter Overview

Chapter II presented a very basic application of transfer entropy to discover simple networks. This showed the ability of the production dynamics model to reliably simulate communication in a network such that information theoretic principles could rediscover the network.

Chapter III built from there, introducing a methodology and applying it to more complex networks, network dynamics, and a real-world network example using actual communication/production data.

Chapter IV introduced organizational network topologies and studied the ability of the communication/production data simulation technique combined with the information theoretic methodology to discover those topologies. A statistical approach was used and best-practice values were presented to enable the practitioners to set limits with which to discover significant network links haivng desired probabilities of true/false positives.

Finally, Chapter V presented a deeper analysis of the organizational network topologies as they were able to transfer and store information. Various node communication throughput values were studied as a means of analyzing the effects of increased or decreased resources in a node. These results were compared with empirical and qualitative organization science research to provide quantitative examples of their findings.

Conclusions of Research

The objectives and questions addressed in this research are:

1. How can information theory be applied to CASNs to model and analyze the static and dynamic structures of those networks?

2. What are the relationships between network structure and information-theoretic indicators of information transfer and information storage?

3. How do organizational network structures differ in their inherent ability to transfer and store information within the network?

The research presented in this dissertation shows that an information theoretic methodology allows a practitioner to reverse engineer a supply network's structure and dynamics with only time-series production data about network nodes. Additionally, local transfer entropy can be used to analyze the network structure's dynamics. This approach requires minimal information about the network. Only node histories are needed; underlying behaviors and relationships between the nodes are not needed, modeled, or assumed.

This methodology was demonstrated to be valid using simple static and dynamic networks with known structure and dynamics. Being able to uncover network structures, either static or dynamic, from real-world data overcomes two significant barriers to studying CASN: that static and dynamic structures or underlying node behaviors must be completely known, or that node behavior must be modeled or assumed. This addressed Question #1 by introducing a methodology, demonstrating the validity of the methodology, and then presenting a real-world example to apply the methodology to the

modeling and analysis of the static and dynamic structures of a real-world network. [9]
[2]

Furthermore, applying information theory in the context of such supply networks introduces new ways of studying these complex systems by enabling other information theoretic measures such as information storage or information modification. Indeed, this dissertation also presented analysis of information transfer and information storage in organizational networks using the information theoretic methodology. It first showed that the information theoretic methodology could be used to discover organizational network structures and presented p-value thresholds for practitioners.

Next, information transfer and storage for five organizational network structures with informal connections were compared to a baseline tree network with only formal connections. Given that organization performance is largely impacted by actors in an organization using formal and informal networks to coordinate effort, share goals, exchange information, and access resources, the measures of information transfer and information storage provided quantification of how organizational network topologies that include formal and informal structures differed in performance. This addressed aspects of both Questions #2 & #3 by comparing and contrasting the information transfer and storage behaviors of five organizational network structures with those of baseline tree. [21]

Finally, I put forth multiple observations which provide a quantitative basis for either empirical or qualitative assertions made by previous researchers regarding the management of complex organizations or behavior of organization networks. The underlying model allowed these assertions to be tested and applied across 5

organizational network topologies, which has not previously been done. In this way, I also studied the risks of decreased communication and opportunities of increased communication in these various organizational networks.

I provided a quantitative study of the effects of informal network links and their ability to affect the transfer and storage of information within the network. This addresses aspects of both Questions #2 & #3 by studying the inherent ability of five organizational network structures to transfer and store information while nodes in the network experienced increased or decreased communication performance capacity.

Significance of Research

This research was the first application of information theory to CASNs to study their static and dynamic structures. This was done by demonstrating the ability to discover both static and dynamic supply network structures. This approach does not require knowledge of the network structure (as in social network analysis approaches) or any assumed node behavior (as in agent-based modeling approaches). This introduced a new approach and methodology with which to study these complex networks.

This research was the first which looked at the statistical ability of an information theoretic methodology to discover organizational network topologies. It also verified the feasibility/applicability when applied to organizational networks and then identified thresholds for practitioners corresponding with desired probabilities of true/false positives.

This research was the first to identify the information transfer and storage properties of organizational network topologies. Additionally, by comparing these

networks to basic trees, I was able to observe the relative effect of informal communication networks to the overall goals of information transfer and storage and demonstrate improvement in networks' ability to transfer and store information with the addition of informal connections.

Finally, this research varied node performance in the organizational networks and observed the effects on information transfer and storage. These observations provided quantitative support for empirical or qualitative assertions made by previous researchers concerning the best management of complex organizations. Additionally, I identified compensatory measures which managers could enact to overcome the effects of decreased node performance in the network to recover information transfer or storage capabilities.

Recommendations for Action

Because organizational communication data can be very difficult to obtain, the ability to record and tag such information should be planned for early in a system's architecture development should that be a concern. This is especially true in systems where a manager may be interested in the informal communication network or the system's complex and dynamic structure.

Recommendations for Future Research

Other disciplines have done significant research relating information theoretic measures with network behaviors such as self-organization and emergence [15] [16], stability/instability [17] [18], and distributed computation [19], all of which remain open areas of research for CASNs. Focusing supply network research on these relationships

96

could reveal new approaches to network management and business analytics in networks adapting to their environment.

Further development of the communication simulation algorithms may allow for additional insights into more complex behaviors. The first priority would be to extend the simulation technique to allow for directed loops and two-way communication flows. The node behavior schemes could also be altered (as was began in Chapter II) to allow for variation in production schemes. Examples of this would be various probabilistic distributions of timing delays or alternate filtering algorithms.

Future research could also quantify maximum transfer entropy values within a network. Transfer entropy between two nodes is bounded by the number of states, but this relationship becomes more complex and non-linear when multiple nodes are involved. The maximum transfer entropy in a perfectly connected network of multiple nodes may be lower than that obtained by simply multiplying the edges by the number of states. The maximum transfer entropy would provide a bound by which researchers or practitioners could normalize and compare networks.

Additional research could be undertaken to relate transfer entropy and information storage to actual network performance or effectiveness measures. This would likely require some amount of real-world data beyond that which was available at the time of this research. This would also enable useful case studies to be performed.

Finally, instead of using organizational network topologies, more general network topologies could be studied. These would include small world, scale-free, and random graphs. Comparing the information transfer and storage capabilities of these network topologies would have much broader applicability and interest.

Summary

Supply networks exist throughout society in manufacturing and knowledge-intensive industries as well as many service industries. Organizations have been noted to behave as complex adaptive systems or information supply networks with both formal and informal structures. Thoroughly understanding supply network structure and behavior are critical to managing such organizations effectively, but their properties of complex adaptive systems make them more difficult to analyze and assess, forcing researchers to rely on unrealistic data or assumptions of behavior. This research proposes an information theoretic methodology to discover such complex network structures and dynamics while overcoming the difficulties historically associated with their study. Moreover, managing these complex networks with formal and informal structures poses additional challenges because the effects of intervention can result in even more unpredictable effects. Noting that two primary functions of organizational networks are to transfer information between nodes and store information in the network, this research quantifies the effects of increased and decreased node performance on the ability of multiple organizational network topologies to accomplish these tasks.

Bibliography

[1]   G. Soda and A. Zaheer, "A network perspective on organizational architecture: performance effects of the interplay of formal and informal organization," *Strategic Management Journal,* vol. 33(6), pp. 751-771, 2012.

[2]   J. Rodewald, J. Colombi, K. Oyama and A. Johnson, "Methodology for Simulation and Analysis of Complex Adaptive Supply Network Structure and Dynamics Using Information Theory," *Entropy,* vol. 18, no. 10, p. 367, 2016.

[3]   A. Surana, S. Kumara, M. Greaves and U. N. Raghavan, "Supply-chain networks: a complex adaptive systems perspective," *International Journal of Production Research 43.20,* pp. 4235-4265, 2005.

[4]   P. Anderson, "Perspective: Complexity theory and organization science," *Organization science 10.3,* pp. 216-232, 1999.

[5]   A. G. Dimitrov, A. A. Lazar and J. D. Victor, "Information theory in neuroscience," *Journal of computational neuroscience 30.1,* pp. 1-5, 2011.

[6]   C. E. Shannon and W. Weaver, "The mathematical theory of communication," *University of Illinois Press,* 1949.

[7]   T. Schreiber, "Measuring information transfer," *Physical review letters 85.2,* p. 461, 2000.

[8]   B. Gourevitch and J. J. Eggermont, "Evaluating information transfer between auditory cortical neurons," *Journal of Neurophysiology 97.3,* pp. 2533-2543, 2007.

[9]   J. Rodewald, J. Colombi, K. Oyama and A. Johnson, "Using information-theoretic principles to analyze and evaluate complex adaptive supply network architectures," *Procedia Computer Science,* vol. 61, pp. 147-152, 2015.

[10] J. T. Lizier, M. Prokopenko and A. Y. Zomaya, "Local measures of information storage in complex distributed computation," *Information Sciences,* vol. 208, pp. 39-54, 2012.

[11] J. T. Lizier, M. Prokopenko and A. Y. Zomaya, "A framework for the local information dynamics of distributed computation in complex systems," *Guided self-organization: inception,* pp. 115-158, 2014.

[12] J. T. Lizier and M. Prokopenko, "Differentiating information transfer and causal effect," *The European Physical Journal B 73, no. 4,* pp. 605-615, 2010.

[13] G. Ver Steeg and A. Galstyan, "Information transfer in social media," in *Proceedings of the 21st international conference on World Wide Web*, 2012.

[14] N. Jung and M. Matsumaru, "A Study on the Information Sharing in Supply Chain Network Based on Information Entropy," *日本経営工学会論文誌64, no. 2E,* pp. 317-324, 2013.

[15] M. Prokopenko, F. Boschetti and A. J. Ryan, "An information-theoretic primer on complexity, self-organization, and emergence," *Complexity 15.1,* pp. 11-28, 2009.

[16] R. Feistel and W. Ebeling, "Entropy and the self-organization of information and value," *Entropy,* vol. 18(5), p. 193, 2016.

[17] J. M. Miller, X. R. Wang, J. T. Lizier, M. Prokopenko and L. F. Rossi, "Measuring information dynamics in swarms," *Guided Self-Organization: Inception,* pp. 343-364, 2014.

[18] G. Nicolis and C. Nicolis, "Stochastic resonance, self-organization and information dynamics in multistable systems," *Entropy,* vol. 18(5), p. 172, 2016.

[19] J. T. Lizier, "The local information dynamics of distributed computation in complex systems," Doctoral Dissertation, School of Information Technologies, The University of Sydney, 2010.

[20] J. A. Nickerson and T. R. Zenger, "Being efficiently fickle: A dynamic theory of organizational choice," *Organization Science,* vol. 13(5), pp. 547-566, 2002.

[21] J. Rodewald, J. Colombi, K. Oyama and A. Johnson, "Confidence Investigation of Discovering Organizational Network Structures Using Transfer Entropy," *Procedia Computer Science 95,* pp. 66-72, 2016.

[22] T. Y. Choi, K. J. Dooley and M. Rungtusanatham, "Supply networks and complex adaptive systems: control versus emergence," *Journal of operations management 19.3,* pp. 351-366, 2001.

[23] G. Li, H. Yang, L. Sun, P. Ji and L. Feng, "The evolutionary complexity of complex adaptive supply networks: a simulation and case study," *International Journal of Production Economics 124.2,* pp. 310-330, 2010.

[24] S. D. Pathak, J. M. Day, A. Nair, W. J. Sawaya and M. M. Kristal, "Complexity and adaptivity in supply networks: Building supply network theory using a complex adaptive systems perspective," *Decision Sciences 38.4,* pp. 547-280, 2007.

[25] J. T. Lizier, "JIDT: an information-theoretic toolkit for studying the dynamics of complex systems," *arXiv preprint arXiv:1408.3270,* 2014.

[26] C. Gomez, J. T. Lizier, M. Schaum, P. Wollstadt, C. Grutzner, P. Uhlhaas and C. M. Freitag, "Reduced predictable information in brain signals in autism spectrum disorder," *Frontiers in neuroinformatics 8,* 2014.

[27] M. Lungarella and O. Sporns, "Mapping information flow in sensorimotor networks," *PLoS computational biology 2.10,* p. e144, 2006.

[28] R. Vicente, M. Wibral, M. Lindner and G. Pipa, "Transfer entropy—a model-free measure of effective connectivity for the neurosciences," *Journal of computational neuroscience 30.1,* pp. 45-67, 2011.

[29] M. G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak and R. Mrowka, "Identification of coupling direction: application to cardiorespiratory interaction," *Physical Review E 65.4,* p. 041909, 2002.

[30] R. Marschinski and H. Kantz, "Analysing the information flow between financial time series," *The European Physical Journal B-Condensed Matter and Complex Systems 30.2,* pp. 275-281, 2002.

[31] K. H. Knuth, D. Gencaga and W. B. Rossow, "Information-Theoretic Methods for Identifying Relationships among Climate Variables," *arXiv preprint arXiv:1412.6219,* 2014.

[32] J. Runge, V. Petoukhov and J. Kurths, "Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models," *Journal of Climate 27.2,* pp. 720-739, 2014.

[33] F. Wang, W. Chen, F. Wu, Y. Zhao, H. Hong, T. Gu, L. Wang, R. Liang and H. Bao, "A visual reasoning approach for data-driven transport assessment on urban roads," in *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, 2014.

[34] S. I. Mari, Y. H. Lee, M. S. Memon, Y. S. Park and M. Kim, "Adaptivity of complex network topologies for designing resilient supply chain networks," *International Journal of Industrial Engineering,* vol. 22, no. 1, pp. 102-116, 2015.

[35] I. Giannoccaro, "Adaptive supply chains in industrial districts: A complexity science approach focused on learning," *International Journal of Production Economics,* vol. 170, pp. 576-589, 2015.

[36] I. Giannoccaro, "Assessing the influence of the organization in supply chain management using NK simulation," *International Journal of Production Economics,* vol. 131, pp. 263-272, 2011.

[37] A. Capaldo and I. Giannoccaro, "Interdependence and network-level trust in supply chain networks: a computational study," *Industrial Marketing Management,* vol. 44, no. 1, pp. 180-195, 2015.

[38] A. Capaldo and I. Giannoccaro, "how does trust affect performance in the supply chain? The moderating role of interdependence," *International Journal of Production Economics,* vol. 166, pp. 36-49, 2015.

[39] E. J. Hearnshaw and M. M. Wilson, "A complex network approach to supply chain network theory," *International Journal of Operations & Productions Management 33, no. 4,* pp. 442-469, 2013.

[40] M. A. Bellamy and R. C. Basole, "Network analysis of supply chain systems: A systematic review and future research," *Systems Engineering,* vol. 16.2, pp. 235-246, 2013.

[41] R. C. Basole and M. A. Bellamy, "Supply network structure, visibility, and risk diffusion: a computational approach," *Decision Sciences 45.4,* pp. 753-789,

2014.

[42] M. A. Bellamy, S. Ghosh and M. Hora, "The influence of supply network structure on firm innovation," *Journal of Operations Management,* vol. 32.6, pp. 357-373, 2014.

[43] G. Ver Steeg and A. Galstyan, "Information-theoretic measures of influence based on content dynamics," in *Proceedings of the sixth ACM international conference on Web search and data mining*, 2013.

[44] L. R. Doyle, "Quantification of information in a one-way plant-to-animal communication system," *Entropy,* vol. 11(3), pp. 431-442, 2009.

[45] L. R. Doyle, B. McCowan, S. F. Hanser, C. Chyba, T. Bucci and J. E. Blue, "Applicability of information theory to the quantification of responses to antropogenic noise by southeast Alaskan humpback whales," *Entropy,* vol. 10(2), pp. 33-46, 2008.

[46] F. Rosas, V. Ntranos, C. J. Ellison, S. Pollin and M. Verhelst, "Understanding interdependencing through complex information sharing," *Entropy,* vol. 18(2), p. 38, 2016.

[47] W. M. Van der Aalst and B. F. van Dongen, "Discovering workflow performance models from timed logs," *Engineering and Deployment of Cooperative Information Systems,* pp. 45-63, 2002.

[48] W. Van der Aalst, T. Weijters and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering,* vol. 16(9), pp. 1128-1142, 2004.

[49] P. S. Dodds, D. J. Watts and C. F. Sabel, "Information exchange and the robustness of organizational networks," *Proceedings of the National Academy of Sciences,* vol. 100.21, pp. 12516-12521, 2003.

[50] V. Albino, L. Fraccascia and I. Giannoccaro, "Exploring the role of contracts to support the emergence of self-organized industrial symbiosis networks: an agent-based simulation study," *Journal of Cleaner Production,* vol. 112, pp. 4353-4366, 2016.

[51] A. Chaturvedi, B. Armstrong and R. Chaturvedi, "Securing the food supply chain: understanding complex interdependence through agent-based simulation," *Health and Technology,* vol. 4.2, pp. 159-169, 2014.

[52] G. Fioretti, "Agent-based simulation models in organization science," *Organizational Research Methods,* vol. 16(2), pp. 227-242, 2013.

[53] K. Mustapha, H. Mcheick and S. Mellouli, "Modeling and simulation agent-based of natural disaster complex systems," *Procedia Computer Science,* vol. 21, pp. 148-155, 2013.

[54] C. Watkins, D. Massey, J. Brooks, K. Ross and M. L. Zellner, "Understanding the mechanisms of collective decision making in ecological restoration: An agent-based model of actors and organizations," *Ecology and Society,* vol. 18(2), p. 21, 2013.

[55] A. Barberan, S. T. Bates, E. O. Casamayor and N. Fierer, "Using network analysis to explore co-occurrence patterns in soil microbial communities," *The ISME Journal,* vol. 6(2), pp. 343-351, 2012.

[56] R. C. Basole, "Topological analysis and visualization of interfirm collaboration networks in the electronics industry," *Decision Support Systems,* vol. 83, pp. 22-31, 2016.

[57] M. D. Conover, C. Davis, E. Ferrara, K. McKelvey, F. Menczer and A. Flammini, "The geospatial characteristics of a social movement communication network," *PloS One,* vol. 8(3), p. e55957, 2013.

[58] C. Ellinas, N. Allan and A. Johansson, "Exploring Structural Patterns Across Evolved and Designed Systems: A Network Perspective," *Systems Engineering,* vol. 19(3), pp. 179-192, 2016.

[59] H. L. Gainforth, A. E. Latimer-Cheung, S. Moore, P. Athanasopoulos and K. A. M. Ginis, "Using network analysis to understand knowledge mobilization in a community-based organization," *International Journal of Behavioral Medicine,* vol. 22(3), pp. 292-300, 2015.

[60] Y. Y. Liu, J. J. Slotine and A. Barabasi, "Control centrality and hierarchical structure in complex networks," *PloS One,* vol. 7(9), p. e44459, 2012.

[61] X. R. Wang, J. M. Miller, J. T. Lizier, M. Prokopenko and L. F. Rossi, "Measuring information storage and transfer in swarms," in *Proceedings of the eleventh European conference on the synthesis and simulation of living systems*, 2011.

[62] X. R. Wang, J. M. Miller, J. T. Lizier, M. Prokopenko and L. F. Rossi, "Quantifying and tracing information cascades in swarms," *PloS One,* vol. 7(7), p. e40084, 2012.

[63] J. T. Lizier, F. M. Atay and J. Jost, "Information storage, loop motifs, and clustered structure in complex networks," *Physical Review E,* vol. 86(2), p. 026110, 2012.

[64] J.-Y. Jung, C.-H. Chin and J. Cardoso, "An entropy-based uncertainty measure of process models," *Information Processing Letters 111.3,* pp. 135-141, 2011.

[65] X. Qu, Z. Hao and W. Sun, "The total entropy model of fractal supply chain network system," *Journal of Computers 6.11,* pp. 2454-2459, 2011.

[66] F. Yang, L. S. Sirish and D. Xiao, "Signed directed graph modeling of industrial processes and their validation by data-based methods," in *Control and Fault-Tolerant Systems (SysTol), Conference on*, 2010.

[67] J. T. Lizier, S. Pritam and M. Prokopenko, "Information dynamics in small-world Boolean networks," *Artificial Life 17.4,* pp. 293-314, 2011.

[68] P. Duan, T. Chen, S. L. Shah and F. Yang, "Methods for root cause diagnosis of plant-wide oscillations," *AlChE Journal 60, no. 6,* pp. 2019-2034, 2014.

[69] N. Ay and D. Polani, "Information flows in causal networks," *Advances in complex systems 11.01,* pp. 17-41, 2008.

[70] F. Yang and D. Xiao, "Progress in root cause and fault propagation analysis of large-scale industrial processes," *Journal of control science and engineering,* 2012.

[71] P. De Bruyn, P. Huysmans, G. Oorts, H. Mannaert and J. Verelst, "Using Entropy's Justificatory Knowledge for a Business Process Design Theory," in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, 2014.

REPORT DOCUMENTATION PAGE					*Form Approved* *OMB No. 074-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* 22-12-2016	2. REPORT TYPE Dissertation	3. DATES COVERED *(From – To)* Jan 2013 – Dec 2016

TITLE AND SUBTITLE An Information Theoretic Investigation Of Complex Adaptive Supply Networks With Organizational Topologies	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Rodewald, Joshua V., Civilian	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
    Air Force Institute of Technology
    Graduate School of Engineering and Management (AFIT/EN)
    2950 Hobson Way, Building 640
    WPAFB OH 45433-7765
	8. PERFORMING ORGANIZATION REPORT NUMBER
	
	AFIT-ENV-DS-16-D-029

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Intentionally left blank	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
    DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
Supply networks exist throughout society in manufacturing and knowledge-intensive industries as well as many service industries. Organizations have been noted to behave as complex adaptive systems or information supply networks with both formal and informal structures. Thoroughly understanding supply network structure and behavior are critical to managing such organizations effectively, but their properties of complex adaptive systems make them more difficult to analyze and assess, forcing researchers to rely on unrealistic data or assumptions of behavior. This research proposes an information theoretic methodology to discover such complex network structures and dynamics while overcoming the difficulties historically associated with their study. Indeed, this was the first application of an information theoretic methodology as a tool to study complex adaptive supply networks. Moreover, managing these complex networks with formal and informal structures poses additional challenges because the effects of intervention can result in even more unpredictable effects. Noting that two primary functions of organizational networks are to transfer information between nodes and store information in the network, this research quantifies the effects of increased and decreased node performance on the ability of multiple organizational network topologies to accomplish these tasks. Multiple qualitative observations from previous researchers are quantitatively analyzed using information theoretic modeling and simulation. Results show an increased ability in local teams to store information within the network as well as a decreased ability by core-periphery networks to respond to increased information rates.

15. SUBJECT TERMS
complex adaptive supply networks; supply chain management; network dynamics; information theory; transfer entropy; local transfer entropy; network structure; organizational networks; information transfer; information storage; network communication

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON John Colombi, AFIT/ENV
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU	118	19b. TELEPHONE NUMBER *(Include area code)* (937) 255-3636 x3347 john.colombi@afit.edu
U	U	U			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18