10-30-2016

Confidence Investigation of Discovering Organizational Network Structures Using Transfer Entropy

Joshua V. Rodewald
Air Force Institute of Technology

John M. Colombi
Air Force Institute of Technology

Kyle F. Oyama
Air Force Institute of Technology

Alan W. Johnson
Air Force Institute of Technology

Recommended Citation

Rodewald, J., Colombi, J. M., Oyama, K. F., & Johnson, A. W. (2016). Confidence Investigation of Discovering Organizational Network Structures Using Transfer Entropy. Procedia Computer Science, 95 (Complex Adaptive Systems Los Angeles, CA November 2-4, 2016), 66–72. https://doi.org/10.1016/j.procs.2016.09.294

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 95 (2016) 66 – 72

Procedia Computer Science

Complex Adaptive Systems, Publication 6
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2016 - Los Angeles, CA

Confidence Investigation of Discovering Organizational Network Structures Using Transfer Entropy

Joshua Rodewald[a]*, John Colombi[a], Kyle Oyama[a], Alan Johnson[a]

[a]*Air Force Institute of Technology, 2950 Hobson Way, WPAFB, OH 45433, USA*

Abstract

Transfer entropy has long been used to discover network structures and relationships based on the behavior of nodes in the system, especially for complex adaptive systems. Using the fact that organizations often behave as complex adaptive systems, transfer entropy can be applied to discover the relationships and structure within an organizational network. The organizational structures are built using a model developed by Dodd, Watts, et al, and a simulation method for complex adaptive supply networks is used to create node behavior data. The false positive rate and true positive rates are established for various organizational structures and compared to a basic tree. This study provides a baseline understanding for the accuracy that can be expected when discovering organizational networks using these techniques. It also highlights conditions in which it may be more difficult to successfully discover a network structure using transfer entropy and bounds confidence levels for practitioners of such methods.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

Keywords: Complex adaptive systems; transfer entropy; organizational network; network structure; network discovery; confidence;

1. 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

* Corresponding author. Tel.: +1-937-255-2992; fax: +1-937-255-4981.
E-mail address: joshua.rodewald@afit.edu

1877-0509
Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2016.09.294

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.

2. Background

2.1. 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.[12]

Anderson was perhaps the first to apply CAS research to the strategic management of organizations.[1] In his seminal work, he proposed that organizations take on the characteristics of CAS. Furthermore, he went on to prescribe methods that 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.[1]

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.[4]

These complex organizational structures can be studied using a model introduced by Dodd 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).[2] 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 1. 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.

2.2. 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[15] such as mutual information (Eq. 1), and conditional mutual information (Eq. 2) between processes X, Y, and Z.

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

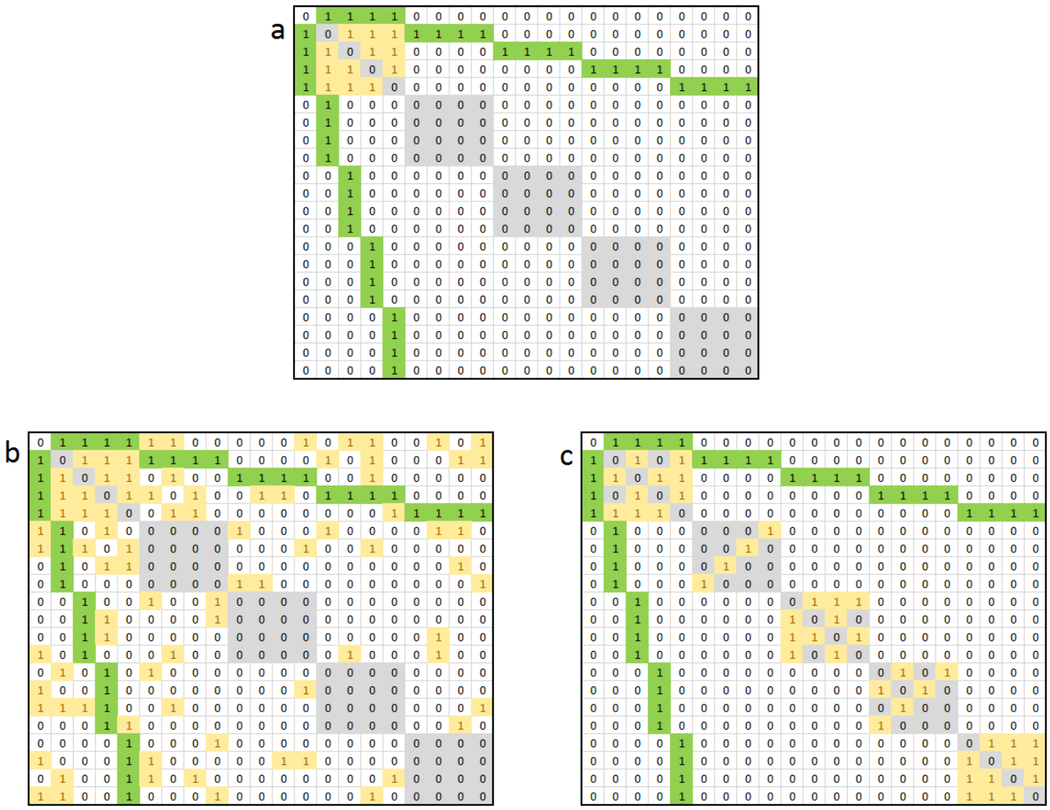

Fig. 1. Example adjacency matrices for a 2-level tree with a branching factor of 4 (shown in green) with links added (shown in yellow) according to Dodd et. al. demonstrating (a) core-periphery structure with links more likely within the top tier of the tree; (b) random inter-divisional structure with links more likely within different divisions (shaded in gray); (c) local teams structure with links more likely within their own division.

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

Transfer entropy[10] is perhaps the most applied of the information dynamics and has been used extensively in the study of many varied complex systems including neuroscience[3,18], social networks[13], finance[8], and others. Lizier[5] 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. 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[7] or an observed correlation between two processes rather than direct effect[6].

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

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 H_0 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 H_0 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.[5]

3. 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. 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.

4. Analysis

Table 1 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 2 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 2 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 1-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 1. 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 1 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.

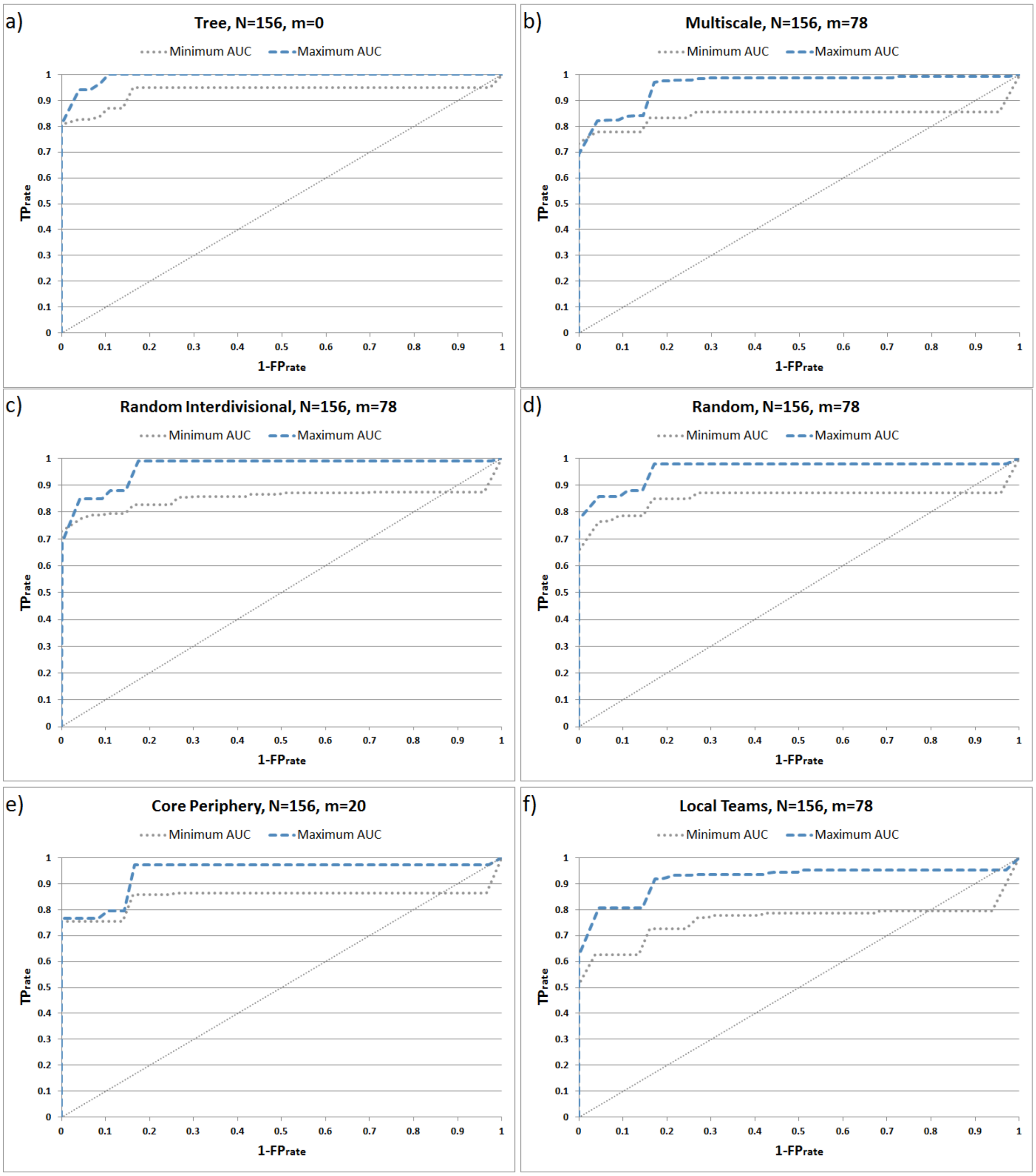

Fig. 2. ROC curves of min and max AUC for various organizational network structures: (a) tree; (b) MS; (c) RID; (d) random; (e) CP; (f) LT.

5. 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 Dodd 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 2 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 2. 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					
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
LT	<0.001	0.51	0.62	0.00	0.00
MS	<0.001	0.69	0.73	0.00	0.00
CP	<0.001	0.75	0.77	0.00	0.00
RID	<0.001	0.69	0.73	0.00	0.00
Tree	<0.001	0.80	0.81	0.00	0.00

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
Random	<0.07	0.85	0.98	0.16	0.17
LT	<0.07	0.73	0.92	0.16	0.17
MS	<0.07	0.83	0.97	0.16	0.17
CP	<0.07	0.86	0.97	0.16	0.17
RID	<0.07	0.83	0.99	0.16	0.17
Tree	<0.07	0.95	1.00	0.16	0.17

References

1. Anderson, Philip. "Perspective: Complexity theory and organization science."*Organization science* 10.3 (1999): 216-232.
2. Dodds, Peter Sheridan, Duncan J. Watts, and Charles F. Sabel. "Information exchange and the robustness of organizational networks." *Proceedings of the National Academy of Sciences* 100.21 (2003): 12516-12521.
3. Gómez, Carlos, et al. "Reduced predictable information in brain signals in autism spectrum disorder." *Frontiers in neuroinformatics* 8 (2014).
4. Li, Gang, et al. "The evolutionary complexity of complex adaptive supply networks: a simulation and case study." *International Journal of Production Economics* 124.2 (2010): 310-330.
5. Lizier, Joseph T. "JIDT: an information-theoretic toolkit for studying the dynamics of complex systems." *arXiv preprint arXiv:1408.3270* (2014).
6. Lizier, J. T., & Prokopenko, M. (2010). Differentiating information transfer and causal effect. *The European Physical Journal B 73, no. 4*, 605-615.
7. Lizier, J. T., Prokopenko, M., & Zomaya, A. Y. (2014). A framework for the local information dynamics of distributed computation in complex systems. *Guided self-organization: inception*, 115-158.

8. Marschinski, Robert, and H. Kantz. "Analysing the information flow between financial time series." *The European Physical Journal B-Condensed Matter and Complex Systems* 30.2 (2002): 275-281.
9. Rodewald, Joshua, et al. "Using Information-theoretic Principles to Analyze and Evaluate Complex Adaptive Supply Network Architectures." *Procedia Computer Science 61 (2015)*: 147-152.
10. Schreiber, Thomas. "Measuring information transfer." *Physical review letters* 85.2 (2000): 461.
11. Shannon, C. E. and W. Weaver (1949), *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press.
12. Surana, Amit, et al. "Supply-chain networks: a complex adaptive systems perspective." *International Journal of Production Research* 43.20 (2005): 4235-4265.
13. Ver Steeg, Greg, and Aram Galstyan. "Information transfer in social media." *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012.
14. Vicente, Raul, et al. "Transfer entropy—a model-free measure of effective connectivity for the neurosciences." *Journal of computational neuroscience* 30.1 (2011): 45-67.