# The Nature of Networks:

# A Structural Census of Degree Centrality

# Across Multiple Network Sizes and Edge Densities

by

Benjamin Elbirt

2007

A thesis submitted to the

Faculty of the Graduate School of State

University of New York at Buffalo

In partial fulfillment of the requirements

For the degree of

MASTER OF ARTS

Department of Communication

Acknowledgements

I would like to thank my advisors Dr. Joe Woelfel and Dr. George Barnett. Special thank you to Dr. Frank Tutzauer for providing assistance with the mathematical formulae and methodologies here in. I would also like to thank CCR (Buffalo Super Computers) for providing super-computing power. Special thanks to friends and family who supported my efforts.

## Author's Note

This document contains diagrams and tables that do not print on any standard format paper. The special page sizes are used to allow for the proper presentation of the materials.

Please contact the author for any data sets or graphic files. Pages 38 through 82 are non-standard page sizes.

Abstract

This thesis examines the mathematical properties of networks, specifically degree centrality at the actor (node) and group (network) level. An algorithm is presented for the creation of all possible edge, node, chain and group degree structures for a given network size and edge density. The census of networks size five through fifteen are used to investigate degree distributions, degrees of freedom and effects of size and density on actor and group degree. Variability (entropy) of information based on actor and network degree centrality structure variations are provided as insight into the complexity of networks. Results indicate an underlying structural influence irrelevant of context suggesting residual data as the contextual behavior element. Power law, fat tail and low density distributions are empirically produced through non-contextual network census suggesting the current behavioral models as structural influence rather than human influence. Finally a general theory for autonomic structural influence is presented with implications for past, present and future research in the area.

Table of Contents

List of Tables

List of Figures

# Chapter 1
# Theory and Research

## Introduction

Network scientists use matrix algebra (Sabidussi, 1966, Wasserman & Faust, 1994) and measures of network structure as indicators of system behavior (Bavelas, 1948, 1950; Freeman, 1977, 1979, 1980; Bonacich 1972, 1987). These measures and methods have been used on networks of all sizes from small group (Leavitt, 1951) to hyperlink networks (Rosen, et. al. 2003; Barnett & Park, 2005, Barnet, Chon, et. al. 2001) and varying contexts including Canadian provinces (Barnett & Sung, 2003), organizations (Carley, 1996), information entropy (Tutzauer, 2007) and semantic analysis of textual information (Doerfel & Barnett, 1999).

Among network measures, nodal degree distributions are highly prevalent. Although usage of degree has been promiscuous over the last half century there has been little work investigating the properties and characteristics of the measure under perfect conditions. Given a matrix of any size with any set of connections there are an absolute number of possible combinations and resulting centrality measures; yet modern research considers random distributions of large data sets arguing for power-law distributions of nodal centrality only[1].

This paper focuses on the properties of degree centrality under conditions of varying network size, edge density, edge strength and directionality. Readers are expected to be familiar with matrix algebra and its use with social networks and graph theory. Seminal works on the topic include Erdos & Renyi (1960), Freeman (1979), and Wasserman & Faust (1994).

[1] See Social Networks volume 29, issue 2 (May, 2007) for a recent overview of p* methodologies.

The remainder of Chapter 1 will discuss degree centrality, research about the measure and the mathematics behind chain-link degree centrality structures. Chapter 2 will focus on the methods of analysis to create the results in Chapter 3. Chapter 4 will provide discussion of the results and insights into how this research may be taken further.

Point/Degree Centrality (DC)

Degree Centrality[2] (DC) indicates the direct connectivity of one actor of the network to all other actors in the network. It has been described as an index of socio-metric status (Friedkin, 1991) and ego density (Wasserman & Faust, 1994). The measure was introduced to social science by Bavelas (1948) and Sabidussi (1966). Freeman's seminal paper on centrality (1979) standardized the formulae which are referred to via. Wasserman & Faust:

$$C_D(n_i) = \sum_{j=1}^{N} x_{ij} \qquad [1.1]^{3}$$

The base formula assumes a binary, non-reflexive, symmetrical network due to the original context of human relationships. It was assumed that two persons who communicated did so bi-directionally and without talking to themselves. To standardize Formula [1.1] divide by the total possible connections.

$$C'_D(n_i) = \frac{\sum_{j=1}^{N} x_{ij}}{(N-1)} \qquad [2.1]$$

Under such conditions, the over-all network, or group degree centrality, can be measured as [3.1]. Freeman (1979) provided a proof for the denominator.

$$C_D = \frac{\sum_{i=1}^{N} [C'_D(n^*) - C'_D(n_i)]}{(N-2)(N-1)} \qquad [3.1]$$

[2] Degree centrality is called "valence" or simply "degree" in graph theory.
[3] Formula (5.2) page 178.

Actor degree can be described as "in" and "out" degree centrality for directed networks. Out directionality represents the actor's direct connectivity to others in the network. In directionality represents the actor's popularity with others. The formulae are represented as follows:

$$C_{Din}(n_i) = \sum_{j=1}^{N} x_{ji} \quad [1.2]$$

$$C_{Dout}(n_i) = \sum_{j-1}^{N} x_{ij} \quad [1.3]$$

The measures are standardized as before:

$$C'_{Din}(n_i) = \frac{\sum_{j=1}^{N} x_{ji}}{N-1} \quad [2.2]$$

$$C'_{Dout}(n_i) = \frac{\sum_{j=1}^{N} x_{ij}}{N-1} \quad [2.3]$$

Finally, directed group degree centrality can be calculated and compared to other networks using:

$$C_D = \frac{\sum_{i=1}^{N} [C'_D(n^*) - C'_D(n_i)]}{(N-1)^2} \quad [3.2]^4$$

The group measure [3.2] can be reduced to [3.3] which illustrates the measure is a function of the maximum actor degree centrality of the network and the sum of all actor degree centralities (Donninger, 1986). Donninger also pointed out the alteration in the denominator for directed networks.

$$C_D = \frac{[(N)(C'_D(n^*))] - \sum_{i=1}^{N} C'_D n(i)}{(N-1)^2} \quad [3.3]^5$$

[4] Adapted from Wasserman and Faust Formula (5.5) page 180. This version uses standardized vs. non-standardized values.
[5] $C'_D(n^*)$ is the maximum $C'_D(n_i)$.

Reflexivity is defined as the ability for an actor to communicate with itself. Although this may seem ridiculous there are circumstances where an actor might do so such as computer networks and networks of networks (aggregation). The formulae are altered to accommodate for the non-zero diagonal as follows:

$$C'_{Din}(n_i) = \frac{\sum_{j=1}^{N} x_{ji}}{N} \quad [2.4]$$

$$C'_{Dout}(n_i) = \frac{\sum_{j=1}^{N} x_{ij}}{N} \quad [2.5]$$

$$C_D = \frac{[N(C'_D(n^*))] - \sum_{i=1}^{N} C'_D\, n(i)}{N^2} \quad [3.4]$$

In non-binary networks the edge strengths can and should be pre-standardized relative to the total strength of the network. The standardized edges are now proportional and can be compared across networks. Under conditions of non-binary networks the formulae are updated as follows:

$$C'_{Din}(n_i) = \frac{\sum_{j=1}^{N} x_{ji}}{\sum_{k=1}^{N}\sum_{m=1}^{N} x_{km}} \quad [2.6]$$

$$C'_{Dout}(n_i) = \frac{\sum_{j=1}^{N} x_{ij}}{\sum_{k=1}^{N}\sum_{m=1}^{N} x_{km}} \quad [2.7]$$

$$C_D = \frac{[(N)(C'_D(n^*))] - \sum_{i=1}^{N} C'_D\, n(i)}{(C'_D(n^*))^2} \quad [3.5]$$

Non-reflexive non-binary group degree centrality:

$$C_D = \frac{[(N)(C'_D(n^*))] - \sum_{i=1}^{N} C'_D\, n(i)}{[(C'_D(n^*))^2] - [(N)C'_D(n^*)]} \quad [3.6]$$

Investigations of Degree/Point Centrality

Network topologies – groups of similar network structures – have been an important area of research. Degree centrality has been compared across topologies include *ring lattice, small world, Erdos random, core-periphery, scale free and cellular* (Airoldi & Carley, 2005). Attention has been given to actor level degree distributions in large networks (Barabasi & Albert, 1999) using random sampling of network structures (Newman, Strogatz, Watts, 2001).

Erdos and Renyi (1960) are credited as seminal research using Poisson distribution random sampling of networks. This distribution has been challenged by research suggesting real-world networks are scale-free with actor distributions following a power-law curve (Newman, Watts, Strogatz, 2002).

Butts (2001) utilized existing theories of complexity to describe the properties of networks and nodes in networks. Conclusions state real networks are similar in complexity to random networks and the results suggest a hypothesis: *the aggregate distribution of empirically realized social networks is isomorphic with a uniform distribution over the space of all graphs, conditional on graph size and density* (Butts, pg 67). Further research into this phenomenon has not been published.

Analysis has been conducted considering the effect of error on degree centrality measures across multiple network sizes and densities (Carley & Butts, 1999; Borgatti & Carley, 2006). The "robustness" of a measure is a function of how well it handles reporting error. Investigation into the effect of adding and removing actors and edges as reporting error show accuracy declines with increasing error (Borgatti & Carley, 2006) and the topology of the network influences the amount of error (Frantz & Carley, 2005).

Understanding the nature of degree centrality does not require prediction or random sampling. The following section will provide a mathematical and algorithmic solution for calculation of degree centrality based on network size and edge density. By way of this algorithm a probability distribution for group and actor degree centrality values can be achieved at any aggregation[6] and used to understand the centrality measure under controlled conditions. Further, the probability distribution for any actor or group degree across the *population* is possible.

Chain-Link Structures

Basic properties must be identified mathematically to create the degree centrality census. The following formulae are derived from the matrix algebra used to represent networks. Current considerations are based on a non-reflexive binary network. Considerations for reflexive and strength based networks follow.

Let N = the number of nodes/actors in the network.

Let E = the number of edges in the network.

$$E = \overline{GD} \qquad [4]^{7}$$

Let G = the maximum number of edges possible in the given network.

$$G = N(N-1) \qquad [5]$$

Let D = the edge density of the network.

$$D = \frac{E}{G} \qquad [6]$$

Let M = maximum edges place-able on a given actor.

[6] The limitation becomes one of computing power.
[7] The line above the right half of the equation signifies the ceiling function. Ceiling rounds up any decimal values to the nearest integer irrelevant of the decimal. Partial edges cannot exist in a network and are considered full edges.

$$M = N - 1 \quad [7]$$

Let $C_E$ = edge structural combinations.

$$C_E = \binom{G}{E} = \frac{G!}{E![(G-E)!]} \quad [8]$$

Formula [8] is the number of combinations for *E* edges on *G* possible edges using standard combinatorial mathematics. The result is the total possible edge placement structures of binary networks.

Given a network of size *N* and density *D* it is known that *E* edges must be placed where no more than *M* edges can exist on a given actor. Further, if a number of edges *e* is placed on an actor *n* there exist that many less edges available to place on other actors and one less actor in the network available to place them on. This process of placement can be described as an edge to actor link in a chain

$$e_1\text{-}n_1, e_2\text{-}n_2, \ldots e_k\text{-}n_k \quad [9]$$

where *k* is the number of edge-actor links in the chain, *e* is the number of edges being placed and *n* is the number of actors receiving *e* edges in the link. The chain can be read as "*e* edges placed on *n* actors" for a total of *(e)(n)* edges placed. The resulting chain of edge-actor links has the properties below in order to satisfy network size and density constraints.

$$M \geq e_k \quad [10]$$

$$e_k > e_{(k+1)} \quad [11]$$

$$N \geq \sum_{i=1}^{k} n_i \quad [12]$$

$$E = \sum_{i=1}^{k} (e_i)(n_i) \quad [13]$$

Formula [10] indicates no more than M edges can ever be placed on any actor in the chain. Formula [11] indicates the edge placement volume must decrease as the chain grows; creation of a chain *4-1, 4-1* (four edges on one actor, four edges on one actor) is the same as 4-2 (four edges on two actors) and the final order *4-2, 4-1* vs. *4-1, 4-2* is irrelevant to the mathematics[8].

Formula [12] indicates the total number of actors that have edges must never exceed the total number of actors in the network; however some actors can be unused. They are isolates. Formula [13] indicates that the sum of all links of the chain multiplied must equal the total edges placed in the network. All chains created will fit the specified network size and density of measure by using these properties.

The $C_E$ structural combinations can also be calculated using the chain methodology.

$$C_E = \prod_{i=1}^{k} \binom{\alpha}{n_i} \binom{M}{e_i}^{n_i} \qquad [14.1]$$

$$\alpha = N - \sum_{j=1}^{(k-1)} n_j \qquad [15]$$

where $\alpha$ is the number of available actors for edge placemen as indicated in Formula [15]. Formula [14.1] can be reduced to Formula [16] for actor isomorphic structures $C_N$ if edge placement is not considered. This value represents the perspective that degree distribution is important, but the specific actor relationship to the distribution is not.

$$C_N = \prod_{i=1}^{k} \binom{\alpha}{n_i} \qquad [16]$$

[8] The use of this property is especially important for a computational algorithm. By forcing the edge size to decrease over time chain redundancy is eliminated. In other words, 4-1, 4-2 is the same as 4-2, 4-1 so do not process it twice. This occurs across all chains – order is not important, simply the edge to actor placement. Recursion can be used when creating the algorithm because of this property.

Value $e_k$ is the out degree centrality for all $n_k$ in the chain. Group degree centrality is based solely on the largest actor value in the chain as per Formula [3.3]. This value is *always* $e_1$. Using Formula [3.6] we can determine the group degree centrality of any chain to be

$$C_D = \frac{[(N)(e_1)] - \Omega}{\Psi} \qquad [3.7]$$

where $\Omega$ is either the number of edges (binary) or the sum of strengths and $\Psi$ is the proper denominator based on reflexivity and strengths. Therefore, if multiple chains with the same $e_1$ exist, those chains also share the same group degree centrality value (another level of aggregation).

Non-Binary / Strength of Ties

To accommodate for strength values, formula [14.1] is replaced by:

$$C_E = E! \prod_{i=1}^{k} \binom{\alpha}{n_i} \binom{M}{e_i}^{n_i} \qquad [14.2]$$

Every edge has a non-zero positive strength that can be placed. When using the standardized strengths from Formulae [2.6] and [2.7] the maximum and minimum are defined as 0 <= Min <= Max <= 1 without consideration for used strengths; however this can be reduced based on the chain structures. Chain structures dictate the edge to actor distributions. The maximum strength of any edge to actor link in the chain is the sum of the maximum edge strength values for the number of edges to be placed on the actor.

The minimum possible strength of any edge to actor link in the chain is the sum of the minimum edge strength values for the number of edges to be placed on the least connected actor or zero if there are actors without edges. As the links in the chain

progress the number of available strengths for placement reduces until none are available at the same rate as the edge placement on actors.

The minimum / maximum group degree centrality can be used to determine a final range of possible values as shown in Formulae [17.1] and [17.2] respectively.

Let $S^L = \{s_1, s_2, \ldots, s_E\}$ the ordered set of edge strengths ordered largest to smallest.

$$C_D = \frac{[(N)\Lambda] - \Omega}{\Psi} \quad [17.1]$$

$$C_D = \frac{[(N)(C'_D(n^*))] - \Omega}{\Psi} \quad [17.2]$$

where $\Psi$ is either [3.5] or [3.6] depending on reflexivity and:

$$\Lambda = C'_D(n^*) \text{ satisfying } \underline{\underline{\sum_{i=1}^{e_k} S_i^L}} \geq \sum_{j=1}^{e_{(k+1)}} S_j^L \text{ and } i \neq j \quad \text{for all } i, j, k. \quad [17.3]$$ [9]

$$C'_D(n^*) = \sum_{j=1}^{e_1} S_j^L \quad [17.4]$$

$$\Omega = \sum_{j=1}^{E} S_j^S \quad [17.5]$$

Formula [17.3] represents the smallest maximum actor degree centrality possible. The inequality conditional requires the largest actor degree centrality be associated with the first link in the chain and all subsequent sum actor centralities in the link/chain be smaller or equal to the first. The smallest possible total strength for the largest placed edges per actor satisfying the conditional will be the smallest possible actor degree centrality used for determining group degree centrality.

[9] The double lines under the left side of the inequality represents minimum; i.e. it should be the smallest possible value that satisfies the entire inequality.

Formula [17.4] signifies the true maximum as the sum of the largest strengths for the largest edge placement ($e_1$). This provides the maximum possible group degree centrality for the given set of strengths.

Reflexivity

Reflexivity requires increasing *G* to $N^2$ from *N(N-1)* [4], [5], [6], [8] and *M* changes from *(N-1)* to *N* [7], [10], [14].

Summary

This chapter began with an outline of degree centrality with a complete description of the formulae involved for strength of ties, binary ties, and reflexivity. An overview of the literature investigating properties of degree/point centrality followed. Finally, the chain-link methodology for generating a complete census of structural aggregations at the edge, node, chain and group degree were presented. The following chapter will discuss the method of census data creation and will be followed by results and discussion.

# Chapter 2

# Method

## Creation of Census

An algorithm was written to generate all possible chains for all networks of a given size and edge density. Binary, non-reflexive networks were assumed for this study. Chains of every edge count possible given network sizes five (5) through fifteen (15) were created. In other words, a census of possible edge counts were generated.

A unique file was created containing all chains for a given network size and edge count. Group degree [3.7], edge structures [14.1] and node structures [16] were calculated for each chain.

## Structure Distributions

An algorithm was used to create the structure volume distributions across each network size and edge count. The resulting values are the total edge structure, node structure, chain structure and distinct group degree centrality value for each network size and edge count.

MatLab 7.1 Student™ was used to create actor degree centrality distributions utilizing the unique chains to determine actor degree centrality appearance volumes. Edge structure volumes were calculated as the sum of the number of actors ($n_i$) with the given centrality value ($e_i$) multiplied by the number of edge structures [14.1] represented by the chain. A proportional value was created by dividing by the total possible values; that is the total structure volume [8] multiplied by the network size ($N$). MatLab 7.1 Student™ also calculated node structures with the same method using [16] in place of [8] as the total possible values.

In addition, MatLab 7.1 Student™ calculated actor volume and chain volume. Actor volume was calculated as the sum of $n_i$ for the given $e_i$ with total values as the number of chains multiplied by the network size. Chain volume was calculated as the number of chains in which the $e_i$ appears irrelevant of the $n_i$ with a total possible values of the chain volume.

The four resulting values represent the proportion of structures/volumes that contain the given actor degree centrality based on network size and edge count. These proportions also represent probability distributions based on aggregation methodology.

Group degree structure distributions were created for each network size and edge count. The resulting data file contains the edge, node and chain structure counts for each unique group degree centrality associated with each network size and edge count. Results indicate the volume of each structure type for the specific group degree centrality based on the network size and edge count.

Proportional distributions were calculated for the group degree values of each network across the edge, node and chain structure distributions. The results indicate the probability/proportion of each group degree for each network size and edge count combination across edge, node and chain structure distributions.

MatLab 7.1 Student™ was used to create edge, node and chain structure appearance probabilities (densities) for many of the structural distributions using

$$p(C_D)^{N,E} = \frac{S_{N,E}}{\sum_{E=1}^{N(N-1)} S_{NE}} \qquad [18]$$

where $p(C_D)^{N,E}$ is the probability or appearance density of the specific $C_D$ group degree value given a network of size $N$ with edge count $E$. $S_{N,E}$ is the specific structure count variable; i.e. edge, node or chain structure count.

Aggregates of the group centralities across all network sizes and edge counts, based on group degree centrality value, were created based on the edge, node and chain structure counts. Results indicate the over-all distributions of centrality values across networks size 5 through 15.

Density Distributions

MatLab 7.1 Student™ was used to create density versions of the aggregate output files. The edge count at a given density is determined using [4]. A list of all edge counts for networks size five (5) through fifteen (15) were generated for edge densities one through one hundred (see Table 1 Appendix B). The resulting list was used to create aggregate files in which shared edge densities were used rather than all possible edge combinations.

Information Entropy Distributions

Information entropy (Shannon, 1948) is a mathematical means of calculating the amount of complexity (Butts, 2001), uncertainty or variability in a variable if the values are treated as random; i.e. there is no history to the selection. This formula is represented as:

$$H(X) = -\sum_{i=1}^{N} p_i \log p_i \qquad [19]$$

where $p_i$ is the appearance probability of a given value in variable $X$.

Using the entropy formula [19] and actor degree/group degree percentile/proportional distributions the variability, measured in log base $e$, of

proportional/probability distributions were calculated. The results indicate the entropic variability of the data; i.e. how ordered/chaotic it is. The larger the resulting value the greater the variability in the information. Zero means completely ordered (without variability).

Formula [20] was used to create the entropy values for actor and group degree centrality for a given Network across all edge densities of the given network size.

$$H(CN) = -\sum_{e=1}^{N(N-1)} p(\Phi)\log(p(\Phi)) \qquad [20]$$

where *H(CN)* is the entropy of actor/group degree value *C* given network size N. $\Phi$ is the probability/proportion of structures represented by the specific *CN* combination for edge count *e*. The result is the variability/entropy as measured in log base *e* of a given actor/group degree for a specific network size across all edge densities of that network.

Graphs & Charts

MatLab 7.1 Student™ and Microsoft Excel™ 2007 were used to create all graphical visualizations provided.

# Chapter 3

# Results

## Structure Distributions

Figures [1] A through D illustrate the edge, node, chain and group degree centrality structural combinations possible for networks size five through fifteen across edge densities 1 through 99. A logarithmic transformation was necessary for structure count due to the rapid growth when looking at edge, node and chain structures. The y-axis labels illustrate the rate of aggregation. The full data set for these figures are found in Appendix B tables 2 through 5.

## Actor degree centrality Distributions

Figures [2] through [16] represent the actor degree centrality distribution across edge structure, node structure, chain structure and actor volume for each network size, across densities one through ninety-nine, organized by actor degree centrality value. Each line of the graph represents a network size. The X-Axis of the graph represents the network edge density by percentage; i.e. what percentage of the network is non-zero. The Y-Axis represents the proportion of actors, based on structure volumes, that contain the given actor degree value. These results represent the impact of network size and edge count on actor degree centrality likelihood.

Figures [17] through [27] represent the actor degree centrality distributions as shown in Figures [2] through [16]; however these are grouped by network size with each line representing a specific actor degree centrality value. These results allow for the comparison of actor degree centrality curves as a function of the specific network size and edge density.

Group Degree Centrality Distributions

Figure [28] contains the aggregate distributions of group degree centrality values across networks size 5 through 15. Figure [29] contains the same information as percentages/proportions or probabilities. Both aggregate across all densities for the given network size and group degree centrality value.

Figures [30] and [31] represent the edge, node and chain structural probabilities for group degree centrality values given edge count and network size for networks size 5 through 15. These graphs represent the likelihood of a given group degree centrality for each network size and edge count.

Entropy Distributions

Figure [32] represents the actor degree centrality entropy distribution, for edge structures, node structures, chain structures and actor volumes, across networks size 5 through 15 and actor degree values 0 through 14. The entropy, as measured in log base *e*, represents the variability in structural distribution for the given network size and actor degree value. The larger the number the more variation exists within the network.

Figure [33] represents the group degree centrality entropy distribution, for edge, node and chain structures, across networks size 5 through 15 and the complete group degree centrality range of zero to one. The entropy, as measured in log base *e*, represents the variability in structural distribution for the given network size and group degree value. The larger the number the more variation exists within the network.

The curve that manifests toward the right side of the graphs is not a phenomenon of any kind. Rather, it is a result of the actor degree centrality range increasing as

network size increases. Networks of size *N* have *N* actor degrees possible including zero when measuring non-reflexive matrixes.

Summary

The results of the census creation indicate a consistent pattern of probabilities with regard to structure likelihood with consistent impacts of both edge density and network size. Power law, fat tail and low density distributions are manifested within the data irrelevant of context. This is important toward forming theories of network behavior because the common behavioral theories are contextually based, yet the phenomena of these theories manifest within the census that is context irrelevant. The results suggest a structural impact on behavior as an underlying influence that is further augmented by context. The following chapter will discuss the implications of these results and describe the results in further detail.

# Chapter 4

## Discussion

Degree centrality is a basic measure/indicator of network structure and organized behavior among any system measured by matrix algebra. This measure is heavily used within many matrix oriented disciplines (Social Networks, Graph Theory, etc.) and many types of distributions are described and explained in the literature presented in Chapter 1.

The use of the chain-link algorithm provides an insight into the variations in structure that occur at multiple aggregation levels based on mathematical phenomena. This study has completed the network census for sizes 5 through 15 to determine if any trends or patterns exist within the aggregations. The following are a description of the various trends found within the data as depicted by the graphs and charts in Appendix A and B respectively; described in Chapter 3 - Results. These are by no means exhaustive and there are many other analyses that can be performed on the current data set.

After presenting trends the discussion section of this paper will conclude with a theory of autonomic structural influence and suggested future research for the topics discussed.

### Impact of Structural Aggregation

Edge, node and chain structure volumes center on 50% edge density and decline as density approaches the tail ends [Figures 1A, 1B, 1C]. Group degree centrality structure volumes are an exception to the distribution curve with smaller edge densities containing larger structural volumes [Figure 1D]. Further, as the network size increases the structural volume shifts toward increasingly smaller edge densities.

Aggregation of network structures reveals extreme variation reduction as aggregation occurs. Figure [1] shows edge structure distributions are $10^{50}$ larger in volume from node structures which are $10^{13}$ larger than chain structures which are $10^{6}$ larger than group structures. This is an extreme reduction in terms of degrees of freedom. There are $9.05x10^{61}$ edge structures to 14 distinct group degree centrality structures for networks size 15 (see tables 2 and 5).

Indicators for Smaller Density; Larger Structure Volume

Network edge density is a group level measure that indicates the amount of interconnection between members of the network. The larger likelihood of smaller densities for network structure, when looking at group degree centrality structures, suggests a general structural phenomenon for larger volumes of smaller densities in networks.

This trend is further illustrated by Figure [29], which depicts structural probabilities for group degree centrality values given network size. Node structure distributions are centered on the 0.50 group degree centrality value. The impact of increasing network size is to flatten the probability curve.

Chain structure distributions start slightly skewed toward smaller group degree centrality values; however the trend tends to center toward 0.50 and flatten as network size increases. The conflict, however, is in edge structure probabilities where group degree centrality values tend to decrease as network size increases and are already skewed heavily toward to smaller group degree value.

The importance of skewing toward smaller group degree values is not readily apparent, but can become so when reviewing Formula [3.2]. As Donninger (1986) found

the formula illustrates the group degree measure is a function of the maximum actor degree centrality of the network. For there to be a tendency toward smaller group degree values there *must* be a tendency toward smaller actor degree centrality values resulting in a need for smaller edge densities. Figures [2] through [16] illustrate smaller actor degree centrality values are more prevalent in smaller edge density networks.

Indicators for Power Law Distributions

Power law distributions (Barab´asi, et. al., 1999) are prevalent throughout edge structure Figures [2] through [16]; actor degree distributions based on network size and edge density. The larger actor degree values exist in larger densities and smaller actor degree values exist in smaller edge densities. Further, the power law occurs in both directions; i.e. at larger densities the power law favors the larger actor degree for the given network size; the smaller group degree values are favored by the smaller densities.

Mathematically this phenomenon should be expected. The presence of smaller actor degree values in larger densities decreases as more edges are non-zero. Further, as density decreases edge placement decreases thus decreasing actor degree values. This phenomenon is partly responsible for the “fat tail” phenomenon discussed later in this paper.

An interesting exception is the chain structure distribution for actor degree zero in Figure [2]. As previously stated, chain volume is calculated as the number of chains in which the $e_i$ appears irrelevant of the $n_i$ with a total possible values of the chain volume. Given this methodology, the zero actor degree is and should be the only degree distributed as shown. This is because it is the only degree centrality value that appears as

a result of missing edges; i.e. all other degrees require placing an edge and the zero degree centrality requires a lack of edges.

Indicators for "Fat Tail" Distributions

The "fat tail" phenomenon is most obvious when reexamining the edge structure actor degree distributions in Figures [17] through [27]. These figures show the actor degree centrality values simultaneously for the given network size. Increasing network size causes a greater distinction in the tails and the tails occur across all actor degrees. The fatter side of the tails is consistent with power law indicators such that the smaller actor degree values are more prevalent in smaller edge densities while larger actor degrees are more prevalent in larger edge densities. Further, the larger group degree values are found in the smaller tail consistent with other phenomenon.

Figures [30] and [31] allow the phenomenon to be seen in group degree centrality results. Figure [31] contains the best representation as data points are fine enough to see phenomena at network size 15. Notice the two tail ends of the edge structure variation are more probable than those between and the phenomenon becomes more distinct as network size increases.

Information Entropy

As network size increases so does the entropy of all the measures (Figure [32]). This increase tends to slow as network size increases suggesting a maximal point of entropy. All actor degrees report near equal entropies suggesting an even distribution of actor degree centralities across all edge densities of a given network.

Figures [33] through [35] represent the edge, node and chain structure group degree entropy distributions. Structural aggregation results in a smoothening of the

curve. However, there is greater entropy in node structure variation versus edge and chain structure variation. This is a result of structural aggregations reducing the complexity of the structures and restricting the variations to a more specific and variant set. Edge structures have the greatest amount of repetition and reduce down to a very specific, but highly variant set of node structures. The chain structure reduction from node structures results in an additional smoothing. This smoothing has the effect of reducing complexity without much of an effect in reducing repetition.

In contrast to actor degree entropy distributions, group degree entropy distributions follow a very linear decline as group degree value approaches one. This is a related phenomenon of the low density high structure volume for group degree values. Given the low structural volume there should be smaller variations possible and thus much smaller entropy.

Finally, it should be noted how small the total entropy values are. Edge structures have a maximal entropy of slightly under 3.5, node structures have a maximal entropy of roughly 6.0 and chain structures have a maximal entropy of slightly more than 2.5 (Figures 33 through 35; Network Size = 15). The English language is reported to have upper bound entropy of roughly 1.75 bits (base 2) or roughly 1.21 base e (Brown et. al, 1992). This supports the reduction of complexity from edge to node structures and the smoothing result of node to chain structures. Further, the results support the massive redundancy in matrix structures under conditions of aggregation.

A Theory of Autonomic Structural Influence

Thus far the discussion has indicated the majority of degree centrality based structural phenomena commonly described in networks literature (Airoldi & Carley,

2005; Barabasi & Albert, 1999; Newman, Strogatz, Watts, 2001; Carley & Butts, 1999; Borgatti & Carley, 2006) are the result of matrix algebra. Further these phenomena are controlled for by limitations that exist in the configuration used regardless of the data contextual source. Fat tails, power laws and larger probabilities for smaller edge densities all manifest simultaneously across network data when context is removed suggesting a mathematical regularity.

This influence can be considered autonomic for any contextual structure as the influence is persistent across all contexts. The samples and results in current literature prove the influence is persistent within contextual networks albeit exceptions occur (Barnett & Elbirt, 2007). Although the exceptions are present, their probabilities are low enough to warrant investigation into long-term structure patterns and the likelihood of survival/maintenance of these structures over time.

Thus a theory of autonomic structural influence can be formulated for future contextual studies that use matrix algebra to represent inter/intra relations of systems; i.e. networks. This theory states *there is an underlying mathematical influence to system phenomena that can be removed as noise to determine non-structural influences on systems represented by matrix algebra.* These non-structural influences would be such things as the unique characteristics of the system or the behavior of individual actors.

This theory is a very important guide toward understanding the internal and external influences on social systems. Not only do the results provide a guideline for determining likelihood estimations for network structures, they also allow for the removal of structural patterns of the models of influence. Although the predictability of what is being removed is very high the information most needed for scientific understanding is

the overlying influences that are at odds or agreement with the underlying autonomic phenomena.

This is in contrast with current methodology; *P** (Robins, G., T. A. B. Snijders, et al. 2007), and *scale free* distributions (Newman, Watts, Strogatz, 2002). Often the residual information found in violation of the structural patterns (fat-tails, power-law distributions, low densities) is removed as noise when in fact it should be treated as real data and everything else noise. Further studies into the relationships that manifest through autonomic pattern removal are necessary for more information and determination of theory validity.

Future Research

Structural aggregation and autonomic influence are not exclusive to social phenomena. Chemistry, particle physics, biological proteins and other areas of science exhibit structural relationships among components that can and are represented by matrix algebra. Current investigations into the human genome rely on structural aggregations and pattern recognition algorithms (Gerstein and Jansen, 2000). Nanotechnology advocates have long term goals including the relationships of multiple nanos into nano-systems that will also inter-relate with each other and their environment as nano-ecologies (Roco, 2007). One might eventually create nano-societies.

Additional questions exist if the autonomic structural influence theory is applied and the structural patterns removed. What will the results of removing structural noise reveal? What are the significant influences provided by contextual behavior and are these influences predicted with greater accuracy? How much influence does structure

have on behavior? Does context influence long term structural patterns and behaviors or are these structures slaves to the algebra?

The degree centrality measure is only one of a handful of measures used in networks and matrix algebra. Social networks use measures of closeness, betweenness, eigenvector, entropy and structural equivalence to determine structural behavior. What aggregations exist for these measures? Does the autonomic structural influence exist for these measures or is it exclusive to degree centrality? Given the relations among these measures (Kincaid, 1993), one might expect a similar autonomic pattern.

Measurement is a key problem as evident by the large scale research efforts into atomic particle measurement devices and nano-visualization tools. Temporal factors, large data sets and costly computing needs continue to plague investigative efforts. Yet, temporal based studies using excessive amounts of data are necessary for validity and reliability. How does a structure that violates likelihood estimations survive or otherwise manifest in the first place? What level of influence do existing structural patterns have on future patterns given likelihood estimations? Can structures be manipulated?

## Conclusion

This paper offers an alternative perspective on network behavior analysis in which the autonomic behavior is isolated from the contextual. The results indicate a highly organized and regulated pattern to structures that underlies any contextual influences that may exist. Powerful phenomena described in network literature on contextual subjects have been recreated through non-contextual network structural analysis. This supports the notion of underlying structural influence on behavior and

suggests a re-evaluation of current research methodologies and results. Further research into structural aggregation and the impact of structural phenomena on behavior remains.

References

Airoldi, E. M. and K. M. Carley (2005). "Sampling Algorithms for Pure Network Topologies." ACM SIGKDD Explorations Newsletter 7(2): 13-22.

Barab´asi, A.-L. a. o. and R. Albert (1999). "Emergence of Scaling in random Networks." Science 286(5439): 509-512.

Barnett, G. A., B. S. Chon, et al. (2001). "The Structure of the Internet Flows in Cyberspace." Networks and Communication Studies **15**(1-2): 61-80.

Barnett, G. A. and B. S. Elbirt (2007). Soc.Culture. NNTP Newsgroup Network based on Cross-Group Authorship; A Preliminary Investigation; Unpublished.

Barnett, G. A. and H. W. Park (2005). "The structure of international Internet hyperlinks and bilateral bandwidth." Annales des Télécommunications **60**(9-10): 1115-1132.

Barnett, G. A., E. J. Sung, et al. (2003). An examination of Canada's Interprovincial Networks. Sunbelt Social Networks Conference. Cancun, Mexico.

Bavelas, A. (1948). "A Mathematical Model for Group Structures." Applied Anthropology 7(3): 16-30.

Bavelas, A. (1950). "Communication Patterns in Task-Oriented Groups." The Journal of the Acoustical Society of America 22(6): 725-730.

Bonacich, P. (1972). "Factoring and Weighting Approaches to Status Scores and Clique Identification." Journal of Mathematical Sociology 2(1): 113-120.

Bonacich, P. (1987). "Power and Centrality: A Family of Measures." The American Journal of Sociology 92(5): 1170-1182.

Borgatti, S. P., K. M. Carley, et al. (2006). "On the Robustness of Centrality Measures under Conditions of Imperfect Data." Social Networks 28(2): 124-136.

Borgatti, S., M. Everett, et al. (2002). UCINET 6 for Windows - Software for Social Network Analysis, Analytic Technologies, Inc.

Brown, P. F., V. J. D. Pietra, et al. (1992). "An estimate of an upper bound for the entropy of English." Computational Linguistics archive **18**(1): 31-40.

Butts, C. (2001). "The complexity of social networks: theoretical and empirical findings." Social Networks **23**(1): 31-71.

Carley, K. M. (1996). "A comparison of artificial and human organizations." Journal of Economic Behavior and Organization 31: 175-191.

Carley, K. M., C. Butts, et al. (1999). "The Interaction of Size and Density with Graph-level Indices." Social Networks 21: 239-267.

Doerfel, M. L. and G. A. Barnett (1999). "A Semantic Network Analysis of the International Communication Association." Human Communication Research 25(4): 589-603.

Donninger, C. (1986). "The Distribution of Centrality in Social Networks." Social Networks 8: 191-203.

Erdos, P. and A. Renyi (1960). "On the Evolution of Random Graphs." Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5: 17-61.

Frantz, T. L. and K. M. Carley (2005). Relating Network Topology to the Robustness of Centrality Measures, Carnegie Mellon University**:** 24.

Freeman, L. C. (1977). "A Set of Measures of Centrality Based on Betweenness." Sociometry 40(1): 35-41.

Freeman, L. C. (1979). "Centrality in Social Networks - Conceptual Clarification." Social Networks 1: 215-239.

Freeman, L. C., D. Roeder, et al. (1980). "Centrality in Social Networks: II Experimental Results." Social Networks 2: 119-141.

Friedkin, N. E. (1991). "Theoretical Foundations for Centrality Measures." The American Journal of Sociology 96(6): 1478-1504.

Gerstein, M. and R. Jansen (2000). "The current excitement in bioinformatics - analysis of whole-genome expression data: how does it relate to protein structure and function?" Bioinformatics **10**: 574-584.

Kincaid, D.L. (1993). Communication Network Dynamics: Cohesion, Centrality and Cultural Evolution. In Richards, W.D. & G.A. Barnett, Progress in Communication Science Vol. XII , pp.111-134. Ablex: Norwood, NJ.

Leavitt, H. J. (1951). "Some Effects of Certain Communication Patterns on Group Performance." Journal of Abnormal Social Psychology **46**(1): 38-50.

Newman, M. E. J., D. J. Watts, S. H. Strogatz (2002). "Random graph models of social networks." Proceedings of the National Acadamy of Science of the United States of America 99(1): 2566-2572.

Newman, M. E. J., S. H. Strogatz, D. J. Watts (2001). "Random graphs with arbitrary degree distributions and their applications." PHYSICAL REVIEW E 64: (17).

Robins, G., T. A. B. Snijders, et al. (2007). "An introduction to exponential random graph (p*) models for social networks." Social Networks **29**(2): 173-191.
Roco, M. C. (2007). New Frontiers for Nanotechnology. SUNY Buffalo. Buffalo, NY**:** Public Lecture.

Rosen, D., J. Woelfel, et al. (2003). "Procedures for Analyses of Online Communities." Journal of Computer-Mediated Communication 8(4): 30.

Sabidussi, G. (1966). "The Centrality Index of a Graph." Psychometrika 31(4): 581-603.

Shannon, C. E. and W. Weaver (1949,1964). The Mathematical Theory of Communication. Urbana, IL, The University of Illinois Press.

Tutzauer, F. (2007). Entropy as a Measure of Centrality in Networks: Characterized by Path-Transfer Flow. ICA 2007, San Francisco, CA.

Wasserman, S. and K. Faust (1994). Social Network Analysis: Methods and Applications. Cambridge, Cambridge University Press.