Massachusetts Institute of Technology
Engineering Systems Division

Working Paper Series

Short Takes on the Literature

ESD-WP-LIT-2003-02

LINKS BETWEEN STRUCTURE AND BEHAVIOR IN COMPLEX NETWORKS: WHAT THE LITERATURE SAYS

Daniel E Whitney

July 2003

LINKS BETWEEN STRUCTURE AND BEHAVIOR IN COMPLEX NETWORKS: WHAT THE LITERATURE SAYS[1]

Daniel E Whitney

"The study of networks pervades all of science, from neurobiology to statistical physics. The most basic issues are structural: how does one characterize the wiring diagram of a food web or the Internet, or the metabolic network of the bacterium *e. coli*? Are there any unifying principles underlying their topology? From the perspective of nonlinear dynamics, we would also like to understand how an enormous network of interacting dynamic systems – be they neurons, power stations, or lasers – will behave collectively, given their individual dynamics and coupling architecture. Researchers are only now beginning to unravel the structure and dynamics of complex networks." [Strogatz]

Most literature on complex networks looks at networks that have evolved naturally, even if humans are involved. Protein networks in cells and the WWW are examples, as are networks of actors, co-authors, and power grids. None can be said to have been designed according to conscious principles or to have a supervisor that directs how they are put together. Nevertheless they display internal structure. There is in the literature a considerable amount of speculation as to what properties or behaviors these structures impart to the networks, as well as considerable disagreement among researchers. If a network analysis were to be done on their papers and reference lists, it is likely that two disjoint clusters would emerge, one centered on Doyle and Carlson (called DC below) the other centered on Barabasi, Strogatz, and Watts (called BSW below). Even when these researchers discuss the same topic, such as percolation or power laws, they never cite each other and almost never cite the same third parties.

In ESD we are interested in both naturally evolving man-made systems like power grids and consciously directed systems like highways, cars, and aircraft. Typical network literature does not deal with the consciously directed ones. The closest related literature is that surrounding the Design Structure Matrix (DSM) [Steward] (which is equivalent to a network) when it is used to model parametric relationships in aircraft jet engines [Hague] or task relationships in complex product development processes [Guivarch]. Studies of product architecture [Ulrich and Eppinger] address creation consciously created network structures, but this literature is too large to consider here.

[1] This topic is very complex and it is very possible that what follows contains some mistakes or misstatements. My apologies in advance.

"Natural network" researchers (Amaral, Albert and Barabasi, Carlson, Doyle, Strogatz, Watts, and others) have studied a number of network metrics and sought to infer from them certain behaviors. The root of most of this research is graph theory, although Carlson and Doyle combine it with control theory while Barabasi combines it with network growth dynamics. A variety of metrics is used to characterize the networks:

- Diameter d, the maximum separation between two randomly selected nodes
- Clustering coefficient C, a measure of the existence of tightly linked subgraphs or cliques
- Average path length ●, the average distance between two randomly selected nodes
- Nodal degree k, the number of links each node has, and its average $\bar{k}$

From these metrics, certain behaviors can be derived, some of which read directly on the networks and others of which are proxies for more complex phenomena that we would like to understand about complex systems.

Networks have also been classified into at least three types:

1. Random graphs with N nodes: The probability of two nodes being connected is p. The nodal degree k has mean $\bar{k} = pN$ and is distributed according to the Poisson distribution $P(k) = e^{-\bar{k}} \frac{\bar{k}^k}{k!}$ for large N. Most nodes have nodal degree near the average.
2. Scale free networks: The distribution of nodal degree follows a power law $P(k) \sim k^{-\gamma}$ with γ having values in the range of 1 to 3. Many nodes have nodal degree far larger than the average.
3. A hybrid of these in which the graphs are built randomly but subject to obeying a power law.

In random graphs, many properties scale with the number of nodes N whereas in scale free graphs, many properties are independent of N.

Random graphs give each node the same probability of being connected to another node, whereas scale free graphs favor a very few nodes with very many connections. In scale free networks, ● is quite small, giving them what Watts and Strogatz call the "small world" property. Yet it is possible to construct random graphs in such a way that they exhibit the small world property, too. [Watts and Strogatz] build such random graphs by starting with a regular grid and randomly rewiring individual nodes. Even when the probability that a node will be rewired randomly is only 0.001, enough short cuts are created that the distance between nodes falls even though the clustering coefficient C remains nearly unchanged. C does not begin to fall until the rewiring probability rises to 0.1, "indicating that the transition to a small world is almost undetectable at the local level." [Watts and Strogatz]

Several properties of networks have been studied based on the above metrics. These include the communication consequences of the above-mentioned diameter and average path length as well as the robustness consequences of different rules for nodal degree distribution.

Communication consequences include

- How many/few nodes of an otherwise regular network with large d or λ must be randomly rewired to effectively bring all nodes into close proximity – equivalently a small value of d or ● (answer: a very small fraction)

- What is the critical interconnection probability needed to create random networks with superclusters in which every, or nearly every, node is connected to the others (answer: this is proportional to the inverse of the dimension of the network, that is, to the number of nodes that a node is allowed to link to)

Robustness consequences include

- What fraction of nodes must be removed before the network breaks into small isolated clusters (answer: high if nodes are removed randomly but low if the most highly connected nodes are removed first; scale free networks are less vulnerable to random removal and more vulnerable to focused removal – called "attacks")

For example, the notion of network diameter or average path length may be relevant to the number of "system-aware" people needed in an otherwise hierarchically organized product development process in order to prevent serious design errors of the "system accident" type from occurring.[2] The

---

[2] System accidents are variously defined as those involving many failures, many cascading or correlated failures, or rapidly cascading correlated hard to understand failures. The literature is huge, driven by social scientists, and nearly

existence of superclusters is relevant to the notion of cascading failures [Strogatz] and the spread of infectious diseases. The likelihood that a network will become disconnected if nodes are removed is relevant to robustness and survivability in the event of random failures or deliberate attacks.

The BSW school and the DC school agree (while ignoring each other) concerning the relationship between "robustness" and "fragility." They note that it is the heterogeneous nature of scale free (and many naturally occurring and, by assumption, well-adapted and survivable) networks that gives them their robustness. Each school has its own models for how this heterogeneity arises. BSW look at mechanisms by which the networks grow over time, either by rewiring or by accretion. DC look at deliberate construction to resist a known threat distribution subject to a resource constraint. Both schools are able to show that their mechanisms lead to scale free networks, although they (may – I can't tell yet) disagree as to what exponents to expect in the equation $P(k) \sim k^{-\gamma}$. Both agree, too, that while heterogeneity conveys certain strengths, it also comes with increasing vulnerabilities of certain kinds.

Less discussed is a matter of interest to us, namely modularity. Some authors seem to feel that clusters are modules but most agree that modules are defined by shared function rather than mere clustering. Thus they try hard to relate the network structures they observe to known functional clustering. [Ravasz et al] show that networks with scale free behavior can be constructed recursively from clusters ("scale free topology with embedded modularity"). After studying artificial networks constructed this way, they examine *e. coli's* metabolic reaction network and show that it has some resemblance to scale free behavior in its clustering coefficient (not the usual way of deciding if a network has scale free behavior), and that *e. coli's* metabolic network is characterized by "a natural breakdown of metabolism into several large modules, which are further partitioned into smaller but more integrated submodules."

The interest of Ravesz et al in modularity lies in "the notion that evolution may act at many organizational levels simultaneously. The accumulation of many local changes, which affect the small, highly integrated modules, could slowly impact the properties of larger, less integrated modules." They appeal then to "the emergence of hierarchical topology through copying and reusing existing modules and motifs" as a different kind of mechanism for development of scale free networks beyond the mechanisms suggested by BSW. In this sense, they align with the [Henderson and Clark] concept of modular innovation while BSW and DC explore mechanisms similar to architectural innovation.

---

devoid of network representations or calculations. [Perrow, Weick, et al, Rasmussen]

In the ESD world, [Baldwin and Clark] have studied the benefits of modularity from an economic and evolutionary point of view. They propose that modules create opportunities for innovation (evolution) in which "experiments" are performed at a cost per unit plus testing and integration cost. The benefit is that better systems emerge. More modules present more opportunities as well as more cost. The benefit is shown, by means of real options theory, to grow in proportion to the square root of the number of modules. The cost is due to the need to test, and rises in proportion to the number of modules if the behavioral influence of each on the whole system is easily seen, thus requiring no additional tests beyond the module level. But the cost rises exponentially if the influence of a module is hard to see, requiring every combination to be tested.

The latter may be closer to the truth for several reasons. First, Baldwin and Clark assume that the system is decomposed hierarchically, so that each level needs to know only about the modules directly below it in the decomposition. This means that non-hierarchical connections, allowed by the scale-free network theorists, are not counted in computing the number of tests. Second, all systems have sneak paths or emergent interconnections, implying a further under-count. All this means that complex systems may not be favored as much by being modular as Baldwin and Clark state. What does favor them is not at all clear. Some possibilities are explored by [Whitney 2001].

[Sharman] explores the applicability of Baldwin and Clark's theory to systems with integral components. He concludes that the basic theory is likely to be applicable as long as the components' effect on the system is measurable, but the calculations are difficult, as is the task of obtaining the necessary data.

DSMs often display an average nodal degree of around 6 ±0.5 for matrices ranging in size from 10 to 10000 or more rows (rows in the matrix correspond to nodes while marks in the matrix correspond to edges). [Guivarch] analyzed the design process for a complex automotive component having over 400 parts and 595 distinct work packages. Most work packages required input from several previously addressed work packages, and one work package served or contributed to more than one subsequent work package. The DSM contained 595 rows and 2001 package delivery marks, yielding a ratio of about 3.36. But information exchanges of a less formal kind, such as requests for comment, amounted to another 1000, raising the ratio to about 5. In a similar vein, [Dong] found that DSMs created by reading documents had average nodal degree less than 4 but DSMs for the same products formed by interviewing people averaged nodal degree of around 6, indicating that documents and official information transfers may account for less than half of the interconnection knowledge required to design complex products in typical companies.

Some researchers have challenged the view that natural networks obey scale free rules, noting that at very large values of nodal degree they have fewer highly connected nodes than they should. Various natural limits have been proposed, including the ageing of nodes so that they stop accumulating links, or rising cost of maintaining many links. [Amaral et al] In a separate study [Whitney 2002] looked at mechanical assemblies to see if they displayed scale free behavior. He calculated the average nodal degree of liaison diagrams (network models of assemblies in which parts are nodes and deliberate connections between parts are links) of about two dozen consumer products such as paper shredders, cordless screwdrivers, Tonka toys, and so on. No scale free behavior was found, although most assemblies have too few parts to permit a really reliable calculation. The main finding is that the average number of links divided by number of parts (half the average nodal degree $\bar{k}$) rarely exceeded 2 regardless of how many parts the assembly had, including assemblies with over 100 parts. Similar data for another two dozen products was obtained from [Greer]. For both Whitney's data and Greer's, the average over the assemblies studied was about 1.6.

The reason, apparently, is that if a part is connected to too many other parts, the assembly will become over-constrained. The Kutzbach criterion for constraint from classical kinematics was used to show that properly constrained assemblies have, under mild assumptions, a value of $\bar{k} \approx 1.5$. Many authors in the field of mechanical design [e.g., Blanding, Whitehead] point out the disadvantages of over-constraint in assemblies, including extreme sensitivity to small variations in the parts, unpredictable assembly times, great skill required for assembly, locked-in stress and its consequences like fatigue and corrosion, and poor operation in general.[3]

In addition to the limit imposed by mechanical constraint (a limit not found in non-physically constrained networks), the deliberate hierarchical nature of complex assemblies may preclude the shortcuts and small average path lengths that characterize scale free networks. A car engine may be highly clustered inside, with many parts attached to the cylinder block. Similarly, car seats, sheet metal bodies, and instrument panels each contain such clusters. But there are no short cuts from the connecting rod to the speedometer, or between the shift lever and the seat cushion. Engineers have imposed such limits in pursuit of their own kind of evolutionary robustness objectives, such as the ability to design, make and test each of these items separately, or the need to use separate different

[3] In fact, the one assembly with nodal degree far in excess of 2 was a Chinese puzzle, which is highly over-constrained.

design and manufacturing skills, materials, and processes.[4] Each of these items has become more complex along its own technological trajectory.

Where linkages between these items occur, they are of two types. One is distributive energy, such as vibrations and electro-magnetic interference, while the other is distributive information, such as sensory, control, and actuation signals. Here the potential for interactions is hardly limited. Higher energy creates higher excitation frequencies, while weight limits bring in new materials with higher natural response frequencies. Higher information flows and content create new dynamic modes, possibly coupling with the vibration phenomena or other dynamic events. The deliberately constructed modules remain, but the emergent connections grow by themselves.

Naturally, we want to know how to manage this. There is the temptation to use biological models because they seem to be "good," due to their evident survivability and robustness at the system level in spite of apparent weakness at the module level. Time will tell if this is fruitful or not.

References

[Albert and Barabasi] Albert, Reka, and Baarabasi, Albert-Laszlo, "Statistical Mechanics of Complex Networks, " Rev Mod Phys, v 24, Jan 2002, pp 47-97.

[Amaral et al] Amaral, L. A. N., Scala, a., Barthelemy, M., and Stanley, H. E., "Classes of Small-World Networks," Proceedings National Academy of Science, vol 97 no 21, Oct 20, 2000, pp 11149-11152

[Baldwin and Clark] Baldwin, C.Y. & Clark, K.B., *Design Rules*, MIT Press, 1999

[Blanding] Blanding, D., *Exact Constraint Design*, ASME Press, 1999.

[Carlson and Doyle] Carlson, J. M., and Doyle, J., "Complexity and Robustness," Proceedings National Academy of Science vol 99 Feb 19, 2002, pp 2538-2545

[Dong] Dong, Q., "Predicting and Managing System Interactions at Early Phase of the Product Development Process," PhD Thesis, MIT Dept of Mech Eng, Feb, 2002.

[Greer] Greer, Monty, private communication

[4] [Baldwin and Clark] call these "modularity in design" and "modularity in manufacture," respectively.

[Guivarch] Givarch, Antoine, " Concurrent Process Mapping, Organizations, Project and Knowledge Management in Large-Scale Design Projects Using the DSM Method," MIT TPP Thesis, June, 2003.

[Hague] Hague, D. C., "Description of a Turbofan Engine Product Development Process," MIT SDM Thesis, 2001

[Henderson and Clark] Henderson, R. M., and Clark, K. B., "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," Admin. Sci. Q., vol 35 (1990), pp 9-30.

[Perrow] Perrow, Charles, *Normal Accidents*, (revised) Princeton University Press, 1999.

[Rasmussen] Rasmussen, J., "Human Error and the Problem of Causality in Analysis of Accidents," Phil Trans R Soc of London B Biol Sci 1990; 327: 449-60

[Ravasz et al] Ravasz, E., Somera, A. l., Mongru, D. A., Oltvai, Z. N., and Barabasi, A.-L., "Hierarchical Organization of Modularity in Metabolic Networks," Science 297, 30 Aug 2002 pp 1551-1555.

[Sharman] Sharman, D. M., "Valuing Architecture for Strategic Purposes," MIT SDM Thesis, 2002

[Steward] Steward, D. V., "The Design Structure System: A Method for Managing the Design of Complex Systems," IEEE Trans Eng Mgt, vol EM-28 no 3, Aug 1981, pp 71-74

[Strogatz] Strogatz, S., "Exploring Complex Networks," Nature, vol 410, 8 March 2001, pp 268-276

[Ulrich and Eppinger] Ulrich, K. T. and Eppinger, S. D., *Product Design and Development*, 2nd Ed, Mc Graw-Hill Irwin, 2000.

[Watts and Strogatz] Watts, D. J. and Strogatz, S., "Collective Dynamics of 'Small-World' Networks," Nature, vol 393, 4 June 1998, pp 440-442

[Weick, et al] Weick, Weick, K. E., Sutcliffe, K. M., and Obstfeld, D., "Organizing for High Reliability: Processes of Collective Mindfulness," Res Org Beh, v 21, pp 81-123, 1999

[Whitehead] Whitehead, T. N., *The Design and Use of Instruments and Accurate Mechanism*, New York: Dover Press, 1954.

[Whitney 2001] Whitney, D. E., "Limits to Physical Modularity," ESD Internal Symposium, 2001

[Whitney 2002] Whitney, D. E., "Network Properties of Mechanical Assemblies," submitted for publication.