Characterization of Efficiency-Robustness Trade-offs in Complex Networks

Ann Author[*] and Second Author[†]
Authors' institution and/or address

Charlie Author[‡]
Second institution and/or address
(Dated: June 13, 2009)

I.   INTRODUCTION

There is a lot of research interest in understanding the structural properties of complex networks, especially following the seminal work by Strogatz, Albert and Barabàsi [1–3]. Classes of complex networks have been studied in search of specific optimal properties such as diameter [4], degree distribution [1], clustering coefficient [5] and modularity [7]. Various real world complex networks have been studied in literature, such as metbolic networks [8, 9], airline networks [10, 11], food webs [**?**], protein interaction networks[citation needed], social networks [7] and trade networks [**?**].

The underlying structure or topology impacts not only the performance of a complex network, but also its survival in the phase of environmental uncertainties. Hence, a detailed understanding of the structures of complex networks is warranted. In this work, we analyze different classes of complex networks and study their structural properties with respect to two parameters critical to network performance: *efficiency* and *robustness*. [Need to say something about results here.]

A large body of literature exists that attempts to characterize complex networks inspired from ideas in statistical physics such as *emergence* and *self-organization*. The most popular model of self-organization is *self-organized criticality* (SOC) proposed by Bak [**?**]. The essential idea behind SOC is that systems with a large number of interacting components will "self-organize" to a critical point at which events follow scale-invariance leading to power laws. Since this behaviour "emerges" without the need for tuning parameters, system design has only a minor role.

Network generation models developed in efforts to explain seemingly universal motifs in complex networks such as scale-free degree distribution, community structures and small diameters use ideas similar to SOC. Barabàsi and Albert [2] propose *preferential attachment* as the underlying mechanism for the emrgence of power laws in nature and man-made networks. Watts and Stro-gatz [5] propose a model to generate networks that exhibit "small-world" properties with small average path lengths and high clustering coefficients leading to the formation of "communities". The Watts-Strogatz model uses a parameter β that determines the amount of randomness in the network. Kleinberg [6] provides a theoretical framework for small-world networks by developing exact conditions under which small-worlds occur. Newman [7] reports the existence of community structures in disparate complex networks.

In contrast to the statistical mechanical explanations of phenomema like power laws and scale-invariance as emerging due to simple underlying forces such as randomness and preferential attachment, there are recent efforts which propose that the occurrence of complex phenomena are a result of meticulous *design* for functionality and performance. A system's functional requirements are defined in terms optimality objectives. The occurrence of interesting motifs is thus the result of trade-offs between objectives. Alderson and Willinger [**?**] propose *highly optimized tolerance* (HOT) as a mechanism to generate power laws as a result of trade-offs between performance and robustness in a network. They also challenge the validity of models such as scale-free networks since such models ignore domain dependent trade-offs [**?**]. Further, the scale-free model attempts to characterize networks based solely on degree distribution, which might be inadequate [**?**].

Venkatasubramanian et al. [**?**] study the emergence of network topologies under different efficiency and robustness constraints. They propose a formalism based on the Darwinian evolutionary theory. Efficiency indicates the average case or short term survival of a complex network; whereas robustness indicates the worst case or the long term survival. A selection pressure variable decides the trade-off between efficiency and robustness. Using these measures, they let topologies evolve under different trade-offs through a genetic algorithm process. Efficiency is defined in terms of the average path length of the resulting graph topology and robustness in terms of the number (and size) of connected components a node deletion causes in the graph. The star topology emrges as most efficient and least robust, when the selection pressure is completely on efficiency. The circle topology emerges as the most robust and least efficient, when the selection pressure is completely on robustness. Hub and

———————
[*]Also at Physics Department, XYZ University.
[†]Electronic address: Second.Author@institution.edu
[‡]URL: http://www.Second.institution.edu/~Charlie.Author

spoke structures of diffrent types emerge for intermediate selection pressures.

We extend this formalism to analyze the efficieny and robustness of several classes of complex networks such as airline networks, food-webs and supply-chains. We use multiple measures of efficiency and robustness that characterize the different aspects of optimality in networks. We observe optimal motifs within classes of networks as well as across classes. We conjecture that these optimal motifs are potential underpinnings for the design of complex networks. [Need to be more specific.] [Should mention important results here.]

Trade-offs have been studied in the context of several real world complex networks. Alderson *et al.* study the trade-offs at the router level as well as the autonomous system level in the internet [**?**]. Amaral *et al.* analyze the world wide air transportation network [11] and report an anamolous betweenness distribution, as a result of geo-political trade-offs. Martinez *et al.* study foodwebs extensively and demonstrate their *connectance* and degree distribution. William and Martinez develop a simple model of foodwebs [**?**]. [Need to add much more literature.]

[Need to do better positioning here.]

II.   A GRAPH THEORETIC FRAMEWORK

Networks are modelled as graphs. In the rest of the paper, the terms of graph and network are used interchangeably to mean the same.

A graph, $G(V, E)$ is a set of of vertices (or nodes) $V = \{1, 2, ..., n\}$ and a set of edges (or links) $E = \{e_{ij} : i \in V \ and \ j \in V\}$. A node can represent a machine, a human or a cell in a network. An edge can represent a *relation* between two nodes. For example, in a social network, an edge represents an "acquaintance" relation. In case of a food web an edge is a "predator-prey" relation. An edge can also represent *flow* of material and/or information as in supply chains and computer networks.

Edges can be *undirected* (or bidirected) or *directed*, indicating the direction of a relation or flow. Edge *weights* are used to convey additional (extra-topological) information. For example: in an airline network, weight on an edge can be distance, or estimated flight time.

Degree is the number of edges incident on a node. If the edges are directed, then we measure *indegree* (the number of incoming edges) and *outdegree* (the number of outgoing edges) of a node. The degree sequence of a graph is the sequence of node degrees. The degree distribution, $P(k)$, of a graph is the fraction of the nodes in the graph, n, that have a degree k.

A *path* in a graph is a sequence of edges from a source node to a destination node. The number of edges in a path is called its *pathlength*. If the edges have weights on them, then the pathlength of a path is the sum of the weights. A *shortest path* between a source and a destination is a path between the two nodes that has the smallest pathlength.

The *average path length* (APL) of a graph is computed by computing the average of the shortest pathlengths between all pairs of nodes in the graph. The longest of the all pairs shortest paths is called a diameter path, and its pathlength is called the *diameter* of the graph.

While the diameter is a property of a network, similar definitions at the level of individual nodes are also useful. For a node, the longest pathlength among its shortest paths is called the node's *eccentricity*. It is the greatest separation of a node from any other node in the network. Diameter is the highest eccentricity across all nodes in the network. Similarly, the smallest eccentricity across all nodes in the network is called the *radius* of the network.

Various measures of *centrality* of nodes in a network are used to determine the relative "importance" of the nodes within a network. Below we define some of the important ones.

Degree Centrality is usually the same as the degree sequence. However, it is often represented in a normalized measure. We normalize the degree centrality of a node against the total degree in the graph, $\sum_i k_i$, where k_i is the degree of i. The normalized degree centrality of a node i, $C_D(i)$, is defined as -

$$C_D(i) = \frac{k_i}{\sum_i k_i}$$

Betweenness Centrality of a node is measured in terms of the number of times the node occurs in shortest paths as an intermediate node. For a node i, and for every node pair (j, k) in a graph, the betweenness centrality of i, $C_B(i)$, is defined as -

$$C_B(i) = \sum_{j \neq i \neq k, j \neq k} \frac{\sigma_{jk}^i}{\sigma_{jk}}$$

Here, σ_{jk}, is the number of shortest paths between j and k. and σ_{jk}^i, is the number of shortest paths between j and k via i.

Betweenness Centrality of an edge is measured in terms of the number of times the edge occurs in shortest paths. For an edge e_{ij}, and for every node pair (k, l) in a graph, the betweenness centrality of e_{ij}, $C_B(e_{ij})$, is defined as -

$$C_B(e_{ij}) = \sum_{k \neq l} \frac{\sigma_{kl}^{e_{ij}}}{\sigma_{kl}}$$

Closeness Centrality of a node has the connotation of the average path length of that node. In other words, it is the average distance from the node to any other node. For a node i, the closeness centrality, $C_{cl}(i)$, is defined as -

$$C_{cl}(i) = \frac{\sum_{j, j \neq i} pl(i, j)}{n - 1}$$

Here, $pl(i, j)$ is the pathlength between i and j.

Often, the performance of a network depends on how well the network is connected. A network is said to be *strongly connected* if there exists a path from any node to any other node in the network. A network is said to be *disconncted* otherwise. Since node and edge failures can happen in a network, it is important to measure the resilience of the network in the face of failures. The *vertex cut* of a network is a set of vertices, whose removal renders the network disconnected. The *minimal vertex cut* is the smallest of such vertex cuts: it is the minimum number of vertices whose removal disconnects a network. Similarly, *edge cut* and *minimal edge* cut are defined.

The vertex or node connectivity of a network is the size of the minimal vertex cut. The edge connectivity of a network is the size of the minimum edge cut. There is a well known theorem by Menger which states that the maximum number of *independent paths* in a network is equal to the size of the minimal cut. A set of paths is said to be *vertex independent*, if there is no shared vertex. Similarly, a set of paths is *edge independent*, if there is no shared edge.

A. Critical Parameters

1. Efficiency

Efficiency is indicative of the short-term survival of a network. Survival of a network entails *communication* of a node with other nodes. A certain cost is associated with communication. Efficiency measures the cost of *communication* in a network. There are a number of ways in which efficiency can be defined. The *diameter* of a network is the upper bound on the communication cost in the network. The *average path length* (APL) gives the communication cost on average. Maximizing the symmetry in the distribution of distances between pairs of nodes can also serve as a useful measure of efficiency. Therefore, per node *eccentricity* (longest of all shortest paths from a node) and *closeness centrality* values can also be used to define efficiency.

The worst diameter for a connected graph of n nodes is $n-1$, which is the diameter of a straight line graph, in case of undirected graphs, and a circle in case of directed graphs. The best diameter is 1, which is the diameter of a clique (complete graph). In other words, a topology is most efficient if the diameter is 1, and least efficient if it is $n-1$. We map a diameter d that falls in the interval $[1, n-1]$, to a value of efficiency in the interval $[0, 1]$, as:

$$\eta = 1 - \frac{d-1}{n-2}$$

By this definition, a straight line topology has an η of 0, a circular topology has an η nearing 0.5, a clique has an η of 1 and so on.

The worst APL for a connected undirected graph of n nodes, which occurs again for a straight line, is $\frac{n+1}{3}$.

Similarly, for directed graphs, the worst case APL, the APL of the circle, is $\frac{n}{2}$. The best case APL in either case, is 1, for the clique. Thus, when measured in terms of APL, efficiency:

$$\eta = 1 - 3\frac{apl - 1}{n - 2}, \; for \; undirected \; graphs$$

$$\eta = 1 - 2\frac{apl - 1}{n - 2}, \; for \; directed \; graphs$$

2. Robustness

Robustness measures the resilience of a network in the face of node and edge failures. Robustness is often defined in terms of the *skew* in the *importance* of nodes and edges. As such, centrality measures viz; degree, node betweenness and edge betweenness distributions can be used. When there is a skew in the centrality measures, a small number of nodes and/or edges are more important than the others. Thus, their failure affects the network's performance much more than failures in the rest of the network. On the other hand, a symmetric centrality distribution ensures robustness to random as well as targetted node/edge failures.

One of the definitions of robustness we use is skew in degree centrality. We define this as the difference in the maximum degree in the graph ($\hat{p}$) and the mean degree of the nodes ($\bar{p}$). For a connected graph of n nodes, the worst skew occurs for the star topology. The central node has a degree of $n-1$ and all the nodes surrounding it have a degree of 1. Therefore, the worst skew is $\frac{(n-1)(n-2)}{n}$. The best skew is 0, when all the nodes have the same degree. This occurs when the topologies are *regular* graph topologies as in a circular topology or a clique. This holds for both directed and undirected graphs. Thus,

$$\rho = 1 - \frac{n(\hat{p} - \bar{p})}{(n-1)(n-2)}$$

Another way to measure robustness is in terms of *betweenness*.

Another way to measure robustness is in terms of *connectivity* (λ). Connectivity is the minimum number of nodes or edges whose removal renders the network disconnected. In case of an undirected graph, the tree topologies have the worst connectivity of 1, and the circle has the worst connectivity of 1 in directed graphs. For both cases, the clique has the best connectivity, $n-1$. Thus, robustness, when defined in terms of connectivity is:

$$\rho = \frac{\lambda - 1}{n - 2}$$

3. Cost

The minimum number of edges (e_{min}) required to have a connected undirected graph is $n - 1$ and n is case of a directed graph. We do not associate any cost to a minimally connected graph. Any "extra" edge has an associated cost. All extra edges cost the same. An undirected clique has the highest cost, with $\hat{e} = \frac{n(n-1)}{2}$ (and $\hat{e} = n(n-1)$, for directed) number of edges. Thus, the cost of a topology is defined as a ratio of the number of extra edges in a topology to the number of extra edges in the clique with the same number of nodes.

$$k = \frac{e - e_{min}}{\hat{e} - e_{min}}$$

In case of weighted graphs, cost is measured in terms of the total weight over edges.

4. Selection Pressure Estimate

Venkatasubramanian et al. [?] propose the selection pressure variable α that decides the trade-off between efficiency and robustness during the design process of a complex network. In this work, we propose a complementary measure β which is an *estimate* of α that might

have been used to design a particular network topology.

III. CLASSES OF COMPLEX NETWORKS

A. Foodwebs

B. Supply Chain Networks

C. Airline Networks

IV. PERFORMANCE ANALYSES

V. RELATED LITERATURE

VI. FUTURE WORK

VII. CONCLUSION

The conclusion goes here.

Acknowledgment

[1] R. Albert and A.-L. Barabási. *Statistical mechanics of complex networks*. Review of Modern Physics, 74, pp. 47-97, 2002.

[2] A.-L. Barabási. *Scale-Free Networks*. Scientific American, 288, pp. 60-69, May 2003.

[3] S.H. Strogatz. *Exploring complex networks*. Nature, Vol. 410, Issue 6825, pp. 268-276, 2001.

[4] R. Albert, H. Jeong and A.-L. Barabási. *Diameter of the world-wide web*. Nature, Vol. 401, Issue 6749, pp. 130-131, 1999.

[5] D.J. Watts and S.H. Strogatz. *Collective dynamics of 'small-world' networks*. Nature, Vol. 393, Issue 6684, pp. 440-442,v 1998.

[6] J.M. Kleinberg. *Navigation in a small world*. Nature, Vol. 406, Issue 6798, pp. 845, 2000.

[7] M.E.J Newman. *Modularity and community structure in networks*. In proc. National Academy of Sciences, Vol. 103, No. 23, pp. 8577-8582, 2006.

[8] R. Guimera, M. Sales-Pardo and L.A.N. Amaral. A network-based method for target selection in metabolic networks. Bioinformatics 23, 1616-1622 (2007).

[9] Smart, AG, Amaral, LAN, Ottino, JM. Cascading failure and robustness in metabolic networks. Proc. Natl. Acad. Sci. U. S. A. 105, 13223-13228 (2008).

[10] Modeling the worldwide airport network.

[11] Airline network.

TABLE I: Summary of properties of various domestic airline networks.

Ion	D^1_{4h}		D^5_{4h}	
	1st alternative	2nd alternative	1st alternative	2nd alternative
K	$(2e) + (2f)$	$(4i)$	$(2c) + (2d)$	$(4f)$
Mn	$(2g)^a$	$(a) + (b) + (c) + (d)$	$(4e)$	$(2a) + (2b)$
Cl	$(a) + (b) + (c) + (d)$	$(2g)^a$	$(4e)^{\mathrm{a}}$	
He	$(8r)^{\mathrm{a}}$	$(4j)^{\mathrm{a}}$	$(4g)^{\mathrm{a}}$	
Ag		$(4k)^{\mathrm{a}}$		$(4h)^{\mathrm{a}}$

[a]The z parameter of these positions is $z \sim \frac{1}{4}$.