Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 20 (2013) 510 – 515
Procedia Computer Science

Complex Adaptive Systems, Publication 3
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2013- Baltimore, MD

Resilience Notions for Scale-Free Networks

Gunes Ercal*, John Matta

Southern Illinois University Edwardsville, Edwardsville, IL, 62026

Abstract

Much of traditional graph theoretic analysis of networks had focused on regular or near-regular network models such as random Erdos-Renyi models, where the degree distribution is either the same for every node or highly concentrated about the mean. As such, much beautiful theoretical machinery exists to analyze various properties including analysis of network resilience via eigenvalues of the adjacency matrix representing the network. However, it has been recently observed that real world networks tend to be scale-free, which usually implies a high variance and power-law degree distribution. This poses a problem in applying existing theoretical machinery to some problems, particularly that of the resilience of networks to node attacks. In this work we examine networks in which the greatest discrepancy arises in attempting to apply previous resilience notions, and we tailor a new mathematical notion of resilience that works for scale-free networks in the presence of node attacks.

Selection and peer-review under responsibility of Missouri University of Science and Technology
Keywords: Networks; conductance; expansion; resilience; scale-free; theory

1. Introduction and Motivation

Whether we speak of cellular networks, the Internet, social networks, or road networks, it is clear that networks are a ubiquitous, inescapable aspect of daily life. In addition to such networks relevant to our daily lives, networks arise in several academic domains as models of biological, physical, and social phenomena. A number of important problems are relevant across the varied network domains. For example, the problem of network congestion clearly affects both traffic networks and the Internet. Another problem that is relevant to many different networks is that of tolerance to errors and attacks. We would like to be able to guarantee that most naturally arising networks are resilient against attacks as well as to engineer artificial networks to be so. Both congestion and error tolerance relate to network resilience.

While it is clear that diverse networks share fundamental problems in common, it may be somewhat surprising that they often share common structural properties as well. For example, the "six degrees of separation" phenomenon that any two people who are not directly acquainted with each other are nonetheless connected by a short chain of acquaintances applies not only to the actual social network, but also to online social networks (e.g.

* Corresponding author. Tel.: 618-650-3348.
E-mail address: gercal@siue.edu

Selection and peer-review under responsibility of Missouri University of Science and Technology
doi:10.1016/j.procs.2013.09.311

Facebook, Twitter), co-author citation networks (as exemplified by the "Erdos number"), the link network of the World Wide Web, and even completely non-social networks such as the physical Internet [18]. Such observations of common properties and problems have led to the emergence of a new area called network science, which presents a unified approach to networks across diverse domains [16]. The word "science" is appropriate in this context, particularly to distinguish the new area from its far older mathematical counterpart, namely graph theory. While graph theory presents a rich language to speak of networks rigorously, network science probes what actual properties and dynamics tends to arise, given a domain, or across domains. It is clear that a network scientist must be well-versed in the language of graph theory to make rigorous statements. However, network science also provides the graph theorist with a plethora of more relevant questions and problems to answer. Therefore, that there should be a healthy feedback between these two areas is clear. And, yet, as we shall soon express, not only have major discrepancies arisen already, but other gaps continue to exist. The aforementioned major discrepancy concerns the seeming ubiquity of scale-free degree distributions in many networks, a particular preferential attachment (PA) model known to generate such degree distributions, and the resilience or lack-thereof in such PA scale-free networks. Before proceeding, we must explain some terms.

When we speak of the structure of any type of network, we are speaking of the graph $G = (V, E)$ that represents the network. The degree of a node v in a network is the number of nodes that v is connected to, that is the number of edges involving the node v. The distribution of node degrees in a network is clearly an important structural property of the network. A network in which all nodes have similar degrees has a homogeneous degree distribution whereas a network with clearly variant node degrees has a heterogeneous degree distribution. An extreme example of a high variance, heterogeneous degree distribution is that of a power-law degree distribution, where the frequency of nodes with degree d is inversely proportional to d^c for a constant $c > 1$. A power-law degree distribution is a particularly important class of the more general heavy-tailed or scale-free degree distributions, all of which share the property of very high variance. It turns out that, excepting a few notable exceptions (e.g. wireless sensor networks), a scale-free degree distribution is another unified property that apparently arises from across these various network domains [1, 10, 11, 16].

The most popularized model generating scale-free degree distributions is the preferential attachment model: Given a small constant integer parameter k, a small initial graph G_0, and an ordering of the vertices $v_0, v_1, v_2, \ldots, v_n$, consider vertex v_t to enter the graph at time t. Node v_t chooses k neighbors to connect to randomly with probabilities proportional to the existing in-degrees of the nodes $v_0, v_1, v_2, \ldots, v_{t-1}$ (normalized by the total indegrees). Thus, a node with more incoming neighbors (i.e. higher degree) after time t-1 will have a higher likelihood of being connected to at time t as well, thus exhibiting the "rich get richer" phenomenon of scale-free distributions. However, it should be noted that preferential attachment is not the only type of dynamic that yields scale-free degree distributions. Indeed, one may obtain a random scale-free graph by first generating the degree distribution $d_0, d_1, d_2, \ldots, d_n$ according to a power-law generator, and then allowing a node v_i to choose its d_i neighbors randomly.

Although both PA scale-free graphs and random scale-free graphs may share the same degree distributions, they differ according to other structural properties. And, it should be noted that there are many other mechanisms by which scale-free networks may arise. Nonetheless, the preferential attachment dynamics of generating a scale-free distribution was popularized and put forward by a series of papers as the mechanism under which complex networks emerge [1, 2, 7, 14, 16]. And, such assumption led to a very popular work [1] which claimed that the Internet is robust against random attacks but highly fragile to targeted attacks; in other words, that the Internet exhibits an "Achilles' heel". This claim, in turn, was followed by a highly oppositional response [7] which challenged a number of assumptions of [1], from the inherent bias of degree data obtained via traceroute methods, to the PA model that is assumed to have generated the network. That, and subsequent works noted that different scale-free graphs can have different resilience properties, and that there is a lack of evidence in claiming that the Internet is PA. Although the problem of trace-route bias was also noted to open the question of whether the internet is scale-free at all, other evidence and subsequent work even by original critics of the PA scale-free hypothesis indicate that the scale-free model remains highly relevant in explaining and modeling network structure [2]. This leads us to our main question, which is how to characterize the robustness of scale-free networks, under various forms of attack, and for various generative models of such networks. We stress that the node-based resilience notion that we formulate will not only be applicable to scale-free networks, but to any network susceptible to node attacks. However, our formulation is particularly relevant for heterogeneous degree networks such as scale-free networks.

2. Existing Mathematical Characterizations of Network Robustness

There are existent, well-studied formal measures associated with a graph's resilience, the most important one of which is a graph's conductance [6, 17]. For homogeneous degree graphs, the conductance is the minimum of the ratio |Cut(S, V-S)|/|S|, taken over all (non-majority) subsets S of the vertex set V, where |Cut(S, V-S)| is the size of the cut-edges separating S from the remainder of the graph V-S. In other words, it is a worst-case measure of the number of edges whose failure (or "cut") would disconnect a proportionally large region of the graph: If more edges are required to be cut in order to cause a significant disconnection to the graphs, then the graph is more resilient and has a correspondingly higher conductance. Formally, due to the beautiful relationship between the mixing time of random walks on graphs and the edge-based resilience of graphs, the notion of conductance Φ can be expressed more generally for any graph G =(V, E) as follows [6, 17]:

Conductance of a Markov Chain: $\Phi = \min \Phi(S)$ over all subsets S of V, where $\Phi(S) = Q(S, V\text{-}S)/\pi(S)$

where $\pi(S) = \Sigma_{v \text{ in } S} \pi(v)$ is the total probability density of S under the stationary distribution π of a random walk (more generally, for a Markov chain) on G, and $Q(S, V\text{-}S) = \Sigma_{e \text{ in Cut}(S, V\text{-}S)} Q(e)$ is the sum of the edge-crossing probabilities $Q(v,u) = \pi(v)P(v, u)$ for every edge along Cut(S, V-S). These random walk based notions correspond more directly to resilience based notions when considering the natural random walk for undirected graphs, and in that situation note that the following hold: $\pi(v) = d_v/(2|E|)$ and $P(v, u) = 1/d_v$ where is the degree of v. We call the matrix P in such case the *normalized adjacency matrix* of G. Therefore, we may obtain the following simplification:

Conductance of a graph: $\Phi_G = \min \Phi_G(S)$ over all subsets S of V, where $\Phi_G(S) = |Cut(S, V\text{-}S)|/(\Sigma_{v \text{ in } S} d_v)$

Note that the further simplification that results in the case of d-regular graphs, wherein $d = d_v$ for all v, yields the earlier combinatorial characterization of conductance |Cut(S, V-S)|/|S| with an additional normalization factor d. If we wish to use conductance to compare two different regular graphs of the same degree, then we may omit the normalization by d. Whichever way that the measure is used, it is important to note that all such numeric measures should be taken comparatively and in context. In the subject of numeric measures, we must stress the combinatorial intractability of directly measuring conductance: An exact method that does not necessitate trying all subsets S of V is yet unknown. However, fortunately, conductance is closely approximable via eigenvalue computation, a fact which has been well-known and well-used in the graph theory community [6, 17]:

Theorem on conductance and spectral gap [6, 17]:
For any undirected graph, G the following inequality holds where λ is the second largest eigenvalue of P(G):

$$\Phi^2 \leq 1 - \lambda \leq 2\Phi$$

As the first (largest) eigenvalue of P is always 1 when P represents a natural random walk on an undirected graph, the quantity $1 - \lambda$ is often referred to as the *spectral gap*. Spectral gap may also refer to the second smallest eigenvalue λ_L of the graph *Laplacian* [6], which is a transformation on the adjacency matrix of a graph yielding zero for the first eigenvalue, and preserving the inequality of the above theorem $\Phi^2 \leq \lambda_L \leq 2\Phi$. In general, however, λ_L need not be equal to $1 - \lambda$. In any case, the approximation of conductance via spectral gap may be very useful for large graphs and infinite graph families. For small graphs, it is useful to calculate conductance directly.

To illustrate conductance of graphs, we first compare identical degree regular graphs with the same number of nodes: Figure 1 gives an example of a graph with a good conductance on the left, and an example of a graph with a bad conductance on the right, controlling for size and degree as both graphs have 10 nodes and homogeneous degree of 3. No single edge acts as a "bottleneck" in the first graph, not even two edges are sufficient to disconnect the graph at all. However, in the second graph there is a single edge whose cut disconnects the graph severely, yielding a conductance of 1/5. Conductance formalizes the notion of "lack of bottlenecks", as a bottleneck forms a critical edge set. In homogeneous degree graphs, the critical edge set has a direct mapping to the critical node set as well, namely the nodes adjacent to the critical edges whose removal disconnects the graph severely. Therefore, in homogeneous degree graphs, the property of having high conductance does map directly to being resilient against attacks.

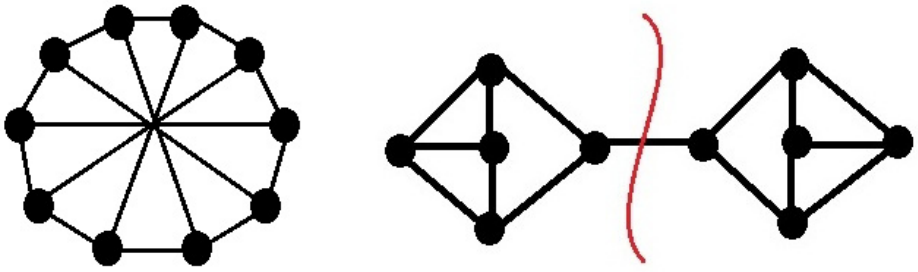

Fig. 1. The graph on the left has high conductance while the graph on the right has low conductance due to the "bottleneck" edge in the middle whose removal disconnects the graph. Both are 3-regular graphs on 10 nodes.

However, the story changes drastically for non-homogeneous degree graphs. Consider now Figure 2, a 10 vertex graph with highly variant degree distribution in which the center node has degree 9 whereas every other node has degree 1. As this graph also happens to be a tree, let us refer to the center node as the root and all the other degree 1 nodes as leaves. Cutting any edge results in only disconnecting its adjacent leaf from the remaining graph. Therefore, to disconnect x nodes from the remaining graph, x edges must be cut, yielding a quite high measure of resilience, numerically 1, if we were to apply the conductance measure stated above. However, an attack against a single node, namely the root, disconnects the entire graph. Thus, it is clear that the discrepancy arises in the consideration of edge failures versus node failures. Conductance captures resilience under a model of edge failures, and this also happens to coincide with a measure of resilience under node failures when the graph has a homogeneous degree distribution. But, as our example shows, conductance no longer captures resilience under a model of node failures when the graph is highly heterogeneous, and in particular scale free. In fact, theoretical work on the conductance of scale-free networks has shown that PA scale free networks exhibit maximal conductance [14] whereas random scale-free networks exhibit excellent though non-maximal conductance [13]. As the vulnerability of PA scale-free networks to targeted attacks is known [1], these are further indications that conductance does not fully capture resilience for scale-free networks under models of both edge and node attacks. In fact, perhaps, the scale-free graph family is extremal in its gap between edge-based resilience as measured via conductance and node-based resilience. As edge-based resilience is a well-studied theoretical problem, we wish to bridge this gap by establishing a theoretically sound framework for node-based resilience.

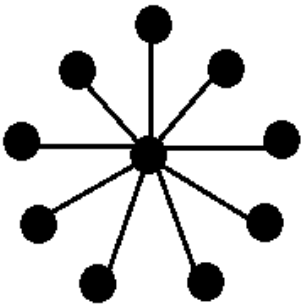

Fig. 2. A highly non-homogeneous degree graph: Any type of edge attack does proportionally little damage to the network, but a node attack targeting the root disconnects the network completely..

3. Our Proposed Mathematical characterization

What we really wish to measure is the following function:

$$s(G) = \min_{\{\text{over all subsets S of V}\}} |S|/|V\text{-}S\text{-}C_{max}|$$

where C_{max} is the largest connected component that remains in the graph G(V – S) (which is the remaining graph after the removal of S). Our measure s(G) is new to our knowledge, and yet also consistent with the combinatorial notion of conductance applied to node cuts rather than edge cuts. As an illustration, similarly to conductance for homogeneous degree graphs, s(G) would be 1/4 in the bad-conductance example at the right of Figure 1 and greater than 1/3 in the good-conductance example at the left of the same figure. However, unlike conductance, s(g) would be only 1/8 for the scale-free graph of Figure 2, better capturing the criticality of the central node. We are the first to propose the function s(G) as the appropriate measure of capturing node-based resilience of general networks.

Resilience and s(G): Suppose that an attack against a subset S of nodes occurred. What is the level of "disconnectedness" in the remaining network? To take the positive approach, when studying the emergence of connectivity from an initially disconnected network, one considers the growth of the size of the largest connected component. Therefore, naturally, larger C_{max} should correspond to better node-conductance given the same size of attacked set |S|. But if a small number of nodes |S| is sufficient to result in a small C_{max}, then that is a network that is not resilient against node attacks. Both extremal situations are captured by s(G). Moreover, we wanted a measure that is consistent with Φ in the case of regular or near-regular graphs. Note that the numerator Cut(S, V-S) in the expression for Φ can naturally represent the "attacked links" whose removal has disconnected some subset S from the remaining graph, and the level of disconnectedness (as measured via the denominator and controlling for same number of edge-attacks) is exactly the size of the disconnected set. Again, Φ like s measures the worst-case disconnectedness. Note that a big difference between edge-attacks and node-attacks is that attacking a single edge can create (detach) at most one new component whereas attacking one node can create several new components.

Attempts and future work: Unfortunately, it is unclear how to exactly compute s(G) without considering all possible subsets of vertices, yielding an intractable (NP-Hard) problem. Simulations on up to 10 nodes already require hours just to compute s(G). However, the meager results thus far confirm similarity of s and Φ for small regular graphs. Recall that while conductance suffers the same intractability problem, conductance is well-approximable with eigenvalues [6, 17]. Attempts at constructing a transformation of the normalized adjacency matrix in such a way that the linear algebraic machinery similar to that used towards conductance has proven unsuccessful in getting away from edge-based notions, and certainly is not yield information on the value C_{max}. We have also considered the notion of vertex expansion and vertex separators [12] which are more related to our work than conductance but also more inapproximable than conductance. Moreover, vertex expansion does not tightly bound s(G) despite the relationship. In any case, because meaningful scale-free network require large number of nodes, approximations of s(G) must be formulated. We consider the most important directions in this regard to be consideration of betweenness centrality properties [15], as it is clear that for the case of a single node attack an attacker would benefit most by attacking a high-betweenness node. A problem in generalizing this case to multi-node attacks arises because the sum total betweenness of a set of nodes no longer corresponds to their group betweenness, and we do not wish to calculate group betweenness for all possible groups as that returns us to an intractable approach. Therefore, what is most crucial is measuring and using the statistical properties of scale-free generative models, and using such statistical properties both as parameters to guide an approximation heuristic for s(G) and as ways of stochastically bounding s(G) for the model in question so that we may compare different generative models of scale-free networks using the appropriate node-based resilience measures. In our ongoing and future work, we attempt to discover methods of approximating s(G) and, particularly, providing statistical bounds for the measure for scale-free models.

References

1. Albert, R., H. Jeong, and A-L. Barabasi. “The Internet’s Achilles’ Heel: Error and attack tolerance of complex networks.” Nature, 2000.
2. Alderson, D., L. Li, W. Willinger, and J. Doyle. “Understanding Internet Topology: Principles, Models, and Validation.” IEEE/ACM Trans. on Networking. 13(6):1205-1218, 2005.
3. Avin, Chen and Gunes Ercal. “On the cover time of random geometric graphs.” Proceedings of the International Colloquium on Automata, Languages, and Programming, 2005.
4. Avin, Chen and Gunes Ercal. “Bounds on the mixing time and partial cover of ad-hoc and sensor networks.” Proceedings of the European Workshop on Wireless Sensor Networks, 2005.
5. Avin, Chen and Gunes Ercal. “On the cover time and mixing time of random geometric graphs.” Theoretical Computer Science, 380(1-2):2–22, 2007.
6. Chung, Fan R. K.. Spectral Graph Theory. American Mathematical Society, February 1997.
7. Doyle, John C., David L. Alderson, Lun Li, Steven Low, Matthew Roughan, Stanislav Shalunov, Reiko Tanaka, and Walter Willinger. “The ‘robust yet fragile’ nature of the internet.” PNAS, October 2005.
8. Ercal, Gunes. “Small Words and Rapid Mixing with a Little More Randomness on Random Geometric Graphs.” Proceedings of Networking, 2011.
9. Ercal, Gunes. "More Benefits of Adding Sparse Random Links to Wireless Networks: Yet Another Case for Hybrid Networks," International Journal of Distributed Sensor Networks, 2012.
10. Fabrikant, A., Elias Koutsoupias, and Christos Papadimitriou. “Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet.” Proceedings of the International Colloquium on Automata, Languages, and Programming, 2005.
11. Faloutsos, C., M. Faloutsos, and P. Faloutsos. “On power-law relationships of the internet topology.” In ACM SIGCOMM, August 1999.
12. Feige, Uriel, MohammadTaghi Hajiaghayi, and James R. Lee. “Improved approximation algorithms for minimum-weight vertex separators.” Proceedings of ACM Symposium on Theory of Computing, 2005.
13. Gkantsidis, Christos, Milena Mihail, and Amin Saberi. “Conductance and congestion in power law graphs.” SIGMETRICS Perform. Eval. Rev., 31(1):148–159, 2003.
14. Mihail, Milena, Christos Papadimitriou, and Amin Saberi. “On certain connectivity properties of the internet topology.” Proceedings of the IEEE Symposium on Foundations of Computer Science, 2003.
15. Newman, M. E. J.. “A measure of betweenness centrality based on random walks.” arXiv, 2003.
16. Newman, M. E. J., A-L Barabasi, and D. J. Watts. The Structure and Dynamics of Networks. Princeton University Press. 2006.
17. Sinclair, Alistair and Mark Jerrum. “Approximate counting, uniform generation and rapidly mixing markov chains.” Inf. Comput., 82(1):93-133, 1989.
18. Watts, D. and S. Strogatz. “Collective Dynamics of Small-World Networks .” Nature 393, 440-442 , 1998.