A dynamic model of time-dependent complex networks

Scott A. Hill[1,*] and Dan Braha[2]

[1]Department of Physics, Southern Methodist University, Dallas TX 75275

[2]New England Complex Systems Institute, Cambridge MA 02138, and University of Massachusetts

(Dated: January 28, 2009)

The characterization of the "most connected" nodes in static or slowly evolving complex networks has helped in understanding and predicting the behavior of social, biological, and technological networked systems, including their robustness against failures, vulnerability to deliberate attacks, and diffusion properties. However, recent empirical research of large dynamic networks (characterized by connections that are irregular and evolve rapidly) has demonstrated that there is little continuity in degree centrality of nodes over time, even when their degree distributions follow a power law. This unexpected dynamic centrality suggests that the connections in these systems are not driven by preferential attachment or other known mechanisms. We present a novel approach to explain real-world dynamic networks and qualitatively reproduce these dynamic centrality phenomena. This approach is based on a dynamic preferential attachment mechanism, which exhibits a sharp transition from a base pure random walk scheme.

A substantial number of papers have extended the study of universal properties in physical systems to complex networks in social, biological, and technological systems [1, 2, 3, 4]. Through theory and experiment, they have characterized the structural properties of such networks, their mechanisms of formation, and the way these underlying structural properties provide direct information about the characteristics of network dynamics and function. Of particular interest are "scale-free" networks, where the degree (i.e., the number of nodes adjacent to a node) is distributed according to a power law or other long-tail distribution, implying the existence of 'hubs'— highly connected nodes of the network [5]. The long-tail aspect of these networks leads to distinct behavior: for example, it has been shown that a long-tail degree distribution makes real-world networks particularly robust against random failures [6, 7], but particularly vulnerable to deliberate attacks and epidemics [8].

Most studies of large social networks have accumulated data over the entire time of observation [1, 2, 3, 4] or have examined the evolution of networks at the macro– and mesoscopic levels [9]. Such studies treat the network as either static, or changing more slowly than the dynamical processes occurring on the network. In general, however, network connections can be fluid, evolving on the same time scale as the underlying dynamical processes; even when there exists an underlying fixed topological structure, connections between nodes can become active or inactive over time. Examples include travel on a transportation network, message exchanges over communication networks, interactions between molecules that can bind to each other, and the switching of genes "on" and "off": all are networks with an underlying structure that can change quickly, and whose changes are relevant to the behavior and functionality of the system.

The existence of a dynamic interaction structure has strong implications for existing work in network transport [6, 7]. Research in epidemics has suggested that an effective disease or computer-virus prevention strategy would be to identify and vaccinate the high-degree nodes of a network, which would inhibit the spread of infection [10, 11, 12, 13]. Similarly, "popular influencer" marketing techniques (closely related to word-of-mouth or viral marketing) are based on the premise that focusing marketing activities on the hubs of social networks increases the likelihood of a cascading adoption of products or services— a type of social epidemic. The underlying assumption of both models is that a node which is well-connected or popular now will continue to be a hub later in time; in other words, the topology of the network is assumed to be static. A dynamic network structure, however, invalidates this assumption: well-connected nodes today may be only weakly connected (or even disconnected) tomorrow. Such a scenario calls for a radical rethinking of the interplay between link/structural dynamics and the dynamical processes underlying the time-dependent complex networks, and may call for modified strategies in predicting and preventing (or encouraging) epidemics.

Following the above reasoning, recent studies [14, 15, 16] have considered the dynamics of the activity of network connections. The experiments detailed in [14] studied the time-dependent structure of a large-scale email network between 57,158 users based on data sampled over a period of 113 days from log files maintained by the email server at a large university. They found, among other things, that the daily networks (as well as the total aggregate network) were scale-free, but with local hubs that changed from day to day. That is, whether or not a particular node is a "hub" changes over time (popular today, anonymous tomorrow) and a node which is among the most connected on one day, may or may not be among the most connected in the aggregate network. This phenomenon, called *dynamic centrality*, was a surprise: normally hubs in networks develop when one particular node becomes a little more popular than the rest, thus attracting more nodes, making it more popular, and so forth: a process called *preferential attachment*. Such a process typically works over a long period of time, so how do the new hubs in this email network develop so quickly, only to disappear the next day? Similar results [15] were found in the interactions caused by the spatial proximity of personal Bluetooth wireless devices, recording the interactions between pairs of students over the period of 31 days.

In this Letter, we will investigate the type of network seen in Ref. [14, 15], which we will call *dynamic scale-free* (DSF) networks. We will show that the primary features of DSF networks (i.e., scale-free networks with a scale-free aggregate and dynamic centrality) can be naturally explained as arising from a series of vertex-weighted random walks on a finite scale-free network, accompanied by a rule of *dynamic preferential attachment*. Starting with a vertex-weighted random walk on the network, we add weight (according to some scheme) to the vertices every time they are visited, constantly altering the jump probabilities in such a way that the walk will tend to revisit vertices already visited [17]. This proposal is inspired by the way messages on actual communication networks are part of larger-scale conversations. For instance, suppose Alice sends out an email to Bob, who then responds to Alice and also forwards her message to Claire. Claire could in turn either forward it to someone new, reply to either Alice or Bob, or do nothing. Assuming these actions are roughly equal in likelihood, there is a greater probability that Claire sends an email to either Alice or Bob, then to some other specific individual: that is, people who are already part of the email chain are more likely to become part of the chain again.

In our approach we start with the aggregate network, which is called the *underlay*; it is reasonable to assume that the aggregate network represents the long-term relationships between individuals, which form in the usual ways [1], and is thus scale-free as most acquaintance networks tend to be. For ease of comparison and computation, we use a single scale-free underlay network throughout the paper, generated using the Barabási-Albert algorithm [5] with parameters $m_0 = 5$ (initial number of nodes; also the minimum degree of every node in the underlay) and $N = 50,000$ (the size of the network). Each "daily" network is a subnetwork of the underlay, built on top of the underlay via a node-reinforced random walk[18]. Starting with a randomly chosen node, we consider that node's neighbors i in the underlay, each neighbor being weighted according to

$$w_i = 1 + CV_i, \tag{1}$$

where V_i is the number of visits the node has received so far, and C is a parameter of the model. Neighbor i is then chosen as a target with probability $\Pi_i = w_i / \sum_j w_j$. Once a target is chosen, the link between the current node and the target is added to the sub-network, the algorithm

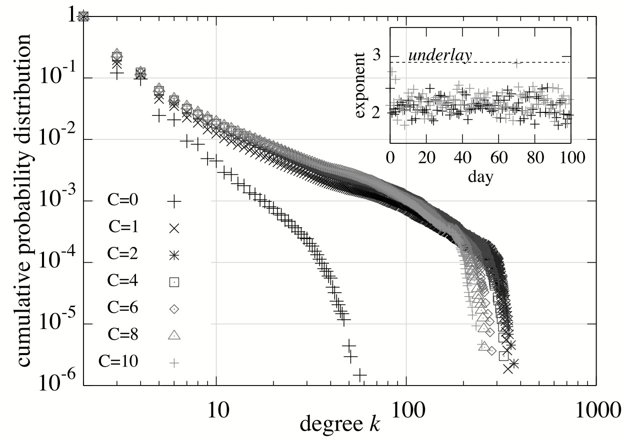

FIG. 1: The log-log plots of the cumulative degree distributions of daily networks, averaged over 100 days, built with single chains of $n_s = 8000$ steps. Each histogram shows the probability that a member node of the daily network has degree k or larger. For $C > 0$, we find power-law regimes with a fast-decaying tail in both cases. By comparison, the cumulative degree distribution for $C = 0$ suggests a single scale more characteristic of exponential decay. The inset shows the change of scaling exponents for the individual daily networks where $C = 1$ (in black) and $C = 10$ (in gray), demonstrating the robustness of our result; the dashed line marks the exponent 2.9 of the underlay.

steps to the target, and the process continues until the chain reaches the prescribed number of steps n_s. The parameter C captures the tendency of the random walk to revisit vertices— that is, the strength of dynamic preferential attachment. When $C = 0$, one has a pure random walk, a case we will refer to as a control, to demonstrate the importance of preferential attachment to our results. One undesirable aspect of this reinforcement scheme is that the process may not be recurrent and, in fact, could get "stuck" on a set of points visited infinitely often [17]. To minimize this effect, we introduce the rule that the same link cannot be activated twice in a row: nodes and links can be revisited multiple times during the simulation, but not immediately.

We will now show that the above reinforcement scheme is able to reproduce the main qualitative features of DSF networks. Figure 1 shows the cumulative probability distribution that a node has more than k links of a daily subnetwork created by a single chain of $n_s = 8,000$ steps, averaged over 100 days. With dynamic preferential attachment activated ($C > 0$), the distribution has a long tail, described well by a power law with an exponential cutoff. This suggests the existence of local hubs, an important element of the networks in [14, 15]. If we fit with a power-law the 100 individual non-cumulative daily degree distributions, we get an exponent (as shown in the inset) which fluctuates around a value of 2.1 ± 0.1, notably smaller than the exponent of the underlay's degree distri-

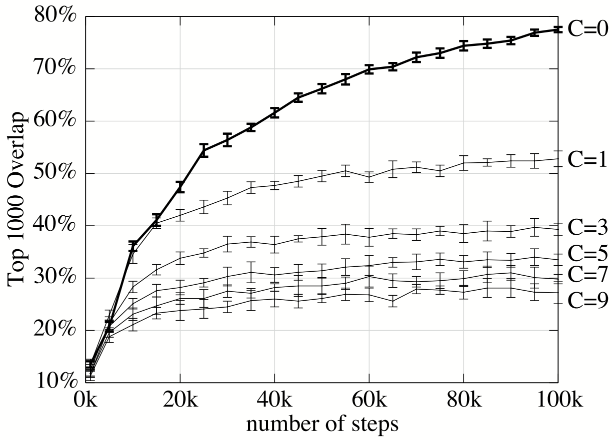

FIG. 2: The Top-1000 overlap, as defined in the text and Eq. 2. Data for various values of C are shown, and the error bars show the standard deviation of each point over 10 independent daily networks; the lines connect points of equal C, and are included for ease of reference.

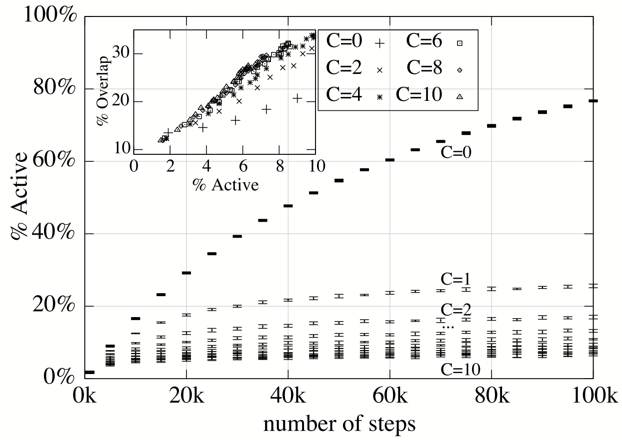

FIG. 3: The fraction of nodes in the underlay that are active (i.e., visited) in a daily network generated with different numbers of steps n_s and with $C = 0, 1, 2, \ldots, 10$. Higher percentages of active nodes correspond monotonically to lower values of C. The error bars show the standard deviation over 10 daily networks. The inset shows one portion of the same data, with the Top-1000 overlap plotted against the percentage of active nodes, demonstrating a strong positive correlation.

bution (which is 2.9). Without preferential attachment ($C = 0$), the degree distribution decays more quickly; the qualitative difference between the two cases is marked, and demonstrates that preferential attachment is an essential element of our model. We have verified (but not shown here) that these results hold true for larger and smaller values of n_s.

We have shown that our model creates scale-fre subnetworks of a scale-free aggregate; now we wish to show that they exhibit dynamic centrality as well. One measure of dynamic centrality, used in [14, 15], is the percentage of the nodes with the n largest degrees in the underlay (aggregate) network, which are also among the top n daily nodes. We call this the "Top n overlap". Here, we use a slight modification of this measure to account for the possibility of ties in the nth slot. If we let D be the nth highest daily degree, then

$$O = \frac{1}{|U|}\left[|U \cap T| + |U \cap S|\,\frac{|U| - |T|}{|S|}\right] \qquad (2)$$

where U is the set of nodes with the top n underlay degrees, S is the set of nodes with daily degree D, and T is the set of nodes with daily degree greater than D. When $|S| = 1$, this reduces to the measure presented in [14, 15]; when S contains more than one element, then there are $|S|$ nodes which could conceivably be in the top n, and there are $|U| - |T|$ available slots remaining in the top n; $(|U| - |T|)/\,|S|$ is thus the fraction of nodes that "make the cut"[19].

In Fig. 2, we see that the overlap of the top 1000 nodes (i.e., the top 2%) is dramatically smaller with the introduction of dynamic preferential attachment, once the number of steps goes beyond $n_s = 10,000$, and the overlap continues to decrease as the preference weight C increases. For large number of steps, the $C > 0$ curves appear to increase at a very slow rate. This suggests that for a wide range of n_s values, highly connected nodes in the aggregate network only play a moderate role in the daily networks. Ref. [14] finds an overlap value of 30%, which fits the values found here.

To better understand the means by which preferential attachment suppresses overlap, we look at the number of nodes visited by the walk as one increases the value of C. Figure 3 shows that, with dynamic preferential attachment, most of the underlay network remains unvisited even after $n_s = 100,000$ steps, at which point the $C = 0$ case has explored three-quarters of the underlay. The inset to Fig. 3 shows the close correlation between the number of visited nodes and the Top 1000 overlap. There is a trade-off here between exploration and exploitation: for the pure random walk mechanism ($C = 0$), as the number of steps increases, the underlay network is rapidly explored (Fig. 3), the daily networks quickly become similar to the underlay network, and the daily local hubs become identical to the hubs of the aggregate network as well (Fig. 2). This result is in contrast to observations reported in [14, 15]. However, with dynamic preferential attachment ($C \neq 0$), the chain is more likely to revisit previously visited nodes (the essence of exploitation), and takes much longer to explore the entire underlay. In this case, if the majority of the underlay remains unexplored (Fig. 3), then it stands to reason that the overlap between the highly connected nodes (hubs) of the aggregate and daily networks will be dramatically suppressed as clearly seen in Fig. 2. This conclusion

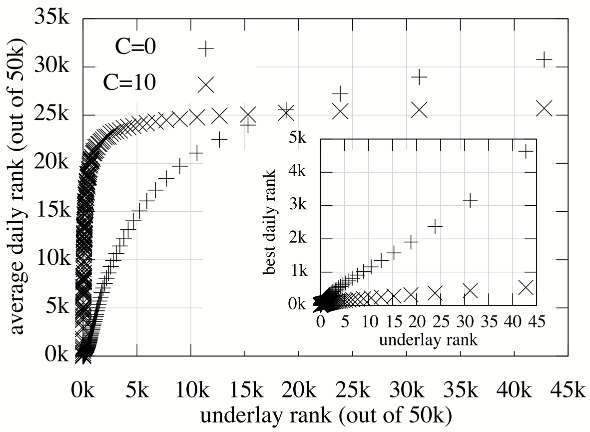

FIG. 4: Comparison of the underlay network with daily networks. We show the average rank of each node in roughly 1800 daily subnetworks (with $n_s = 64,000$ steps) versus their rank in the underlay network; nodes with higher degree have lower rank, and the top-ranked node has rank 0. In the case of ties, all tied nodes are assigned the rank that is the average value of the range they cover. The inset shows, not the average, but the best daily rank that node has over the 1800 days. Note that there are often many nodes with the same underlay degree; each point represents the average over all such nodes.

finds support in typical observations found in dynamic scale-free networks, and provides further evidence that the pure random-walk scheme cannot serve as a plausible explanation for experimental data, in contrast to dynamic preferential attachment schemes.

Our networks appear even more egalitarian when one looks at the rankings of all the nodes, not just the top 1000. To show this, we find the average rank of each node in roughly 1800 daily sub-networks, for $C = 0$ and $C = 10$, and compare that average rank with the rank of the node in the underlay. As illustrated in Fig. 4, we do see that the top-ranked underlay nodes tend to be the top-ranked daily nodes as well. However, remarkably, the results for $C = 10$ are highly egalitarian: all of the nodes outside of the top 5% have the same average rank, unlike in the $C = 0$ case where daily rank consistently increases with underlay rank. The inset to Fig. 4 shows each node's best daily rank (over 1800 days); for $C = 10$ it is possible for every node to be among the top 1000 daily nodes, while for $C = 0$ the bottom-ranked nodes never become very important. Again, these statistics are in agreement with the findings reported in [14, 15].

In conclusion, we have embarked on a research program designed to develop universal models that can recreate empirically observed phenomena in dynamic complex networks. We have shown that, using a suitable reinforced random walk on a "long-term" underlay network, one is able to produce instantaneous networks which reproduce qualitatively characteristic features of real world dynamic networks. This includes, in particular, the construction of scale-free sub-networks of the underlay network, whose local hubs substantially differ from sub-network to sub-network, and from those of the underlay network. We have presented evidence that dynamic preferential attachment (as opposed to a pure random walk) is a fundamental feature of dynamic scale-free networks: the preferential attachment mechanism limits the walk's exploration of the network, which gives early-visited nodes higher ranks than they would have with a more random traversal. We hope our model will stimulate further empirical and theoretical work, and provide a framework for analyzing the influence of link/structural dynamics on dynamical processes on complex networks.

* Electronic address: sahill@mailaps.org

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] L. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Nat. Acad. Sci. USA **97**, 11149 (2000).
[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chaves, and D.-U. Hwang, Physics Reports **424**, 175 (2006).
[4] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, Adv. Phys. **56**, 167 (2007).
[5] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[6] R. Albert, H. Jeong, and A.-L. Barabási, Nature **406** (2000).
[7] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[8] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 035108(R) (2002).
[9] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[10] Z. Dezső and A.-L. Barabási, Phys. Rev. E **65**, 055103(R) (2002).
[11] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002).
[12] R. Cohen, S. Havlin, and D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
[13] J. Gómez-Gardeñes, P. Echenique, and Y. Moreno, The European Physical Journal B **49**, 259 (2006).
[14] D. Braha and Y. Bar-Yam, Complexity **12**, 59 (2006).
[15] D. Braha and Y. Bar-Yam, in *Adaptive Networks: Theory, Models and Applications*, edited by T. Gross and H. Sayama (Springer, 2009).
[16] A.-L. Barabási, Nature **435**, 207 (2005).
[17] R. Pemantle, Probability Surveys **4**, 1 (2007).
[18] We have also considered mechanisms in which sub-networks are built out of multiple chains, or with different weighting schemes in Eq. 1, both of which provide essentially the same results.
[19] A more careful analysis would allow for the possibility of ties in the underlay degree as well, but since the underlay is (a) much larger, and (b) constant between iterations, we ignore this possibility.