The structure of growing social networks

Emily M. Jin[1,2], Michelle Girvan[2,3], and M. E. J. Newman[2]
[1] *Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138*
[2] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501*
[3] *Department of Physics, Cornell University, Ithaca, NY 14853–2501*

We propose some simple models of the growth of social networks, based on three general principles: (1) meetings take place between pairs of individuals at a rate which is high if a pair has one or more mutual friends and low otherwise; (2) acquaintances between pairs of individuals who rarely meet decay over time; (3) there is an upper limit on the number of friendships an individual can maintain. Using computer simulations, we find that models that incorporate all of these features reproduce many of the features of real social networks, including high levels of clustering or network transitivity and strong community structure in which individuals have more links to others within their community than to individuals from other communities.

I. INTRODUCTION

Many real-world systems take the form of networks—nodes or "vertices" joined together by links or "edges." Commonly cited examples include communication networks such as the Internet or the telephone network, information networks such as the World-Wide Web, transportation networks such as airline routes or roads, distribution networks such as the movements of delivery trucks or the blood vessels of the body, and other naturally occurring networks such as food webs or metabolic networks. In the last few years there has been a substantial amount of interest in network structure and function within the physics community; see Refs. [1–3] for reviews. In particular, it turns out that many of the techniques of statistical physics, such as scaling and renormalization group methods, Monte Carlo simulation, and mean-field theory, are well suited to the study of these systems.

One specific question which has received a large amount of attention in the physics literature concerns the structure of networks which are evolving over time. While many networks, such as metabolic networks or blood vessels, are fundamentally static and do not change their topology, many others change substantially over time. The classic example is the World-Wide Web. The vertices in this network are Web pages and the (directed) edges between them are hyperlinks from one page to another. This network is certainly changing; pages are added to the Web at a rate of over a million pages a day, according to some estimates, while other pages disappear. It is widely believed that the rapid growth of the Web leaves a highly distinctive signature in the resulting network, including such characteristic features as power-law degree distributions [4,5], correlations between degree and age of vertices [6], and correlations between degrees of connected vertices [7,8]. A number of models of the growing Web have been proposed, which convincingly reproduce some or all of these features [6–11].

The Web however was not the first type of network to catch the eye of the physics community. In a seminal paper in 1998, Watts and Strogatz [12] discussed a number of features of social networks—networks of acquaintance between people, for instance—and introduced a simple model of a (static) social network, which has since been analyzed in depth in the physics literature [1,13–17]. Social networks also evolve, with new acquaintances forming between individuals, and old ones decaying. However, it is clear that the evolution of a social network is governed by very different processes from those that govern the evolution of the World-Wide Web. In this paper, we propose some new models of the evolution of social networks. In the same spirit as the highly successful models of Web growth (and indeed of most of statistical physics), these models are based on simple stochastic processes and do not attempt to capture the microscopic details of social dynamics. As we will see however, a number of non-trivial but intuitively reasonable results emerge from these models, including the formation of closely-knit communities within the network, and the development of a high degree of network transitivity, as we now describe.

II. MECHANISMS OF SOCIAL NETWORK GROWTH

The key elements in previous network growth models, such as models of the growth of the World-Wide Web, are (1) continual addition of both vertices and edges to the network as time passes and (2) preferential attachment, meaning that edges are more likely to connect to vertices of high degree than to ones of low degree. (The degree of a vertex is the number of other vertices to which it is connected.) Other features, such as removal of vertices or edges, or movement of edges to new positions in the network, can also be incorporated [18], but the crucial features of power-law degree distributions and correlations between vertex degrees are reproduced with only the elements 1 and 2 above.

Growth models of this type are, as mentioned above, quite inappropriate as models of the growth of social networks, for a number of reasons, as follows.

1. New vertices are of course added to social networks all the time: people are born and people travel around joining new networks and leaving old ones. However, the timescale on which people make and break social connections, which can be as short as hours or days, is much shorter than the timescale on which vertices join or leave the network, which is typically some years. For this reason, we expect that the addition and removal of vertices will not be a major factor determining the instantaneous structure of social networks, and to a first approximation these networks can therefore be treated using a model with a constant number of vertices but a varying number and arrangement of edges. This is in sharp contrast to models of Web growth.

2. The degree distribution of many acquaintance networks does not appear to follow a power-law distribution, as the degree distribution of the Web does. Instead the distribution appears to be strongly peaked around a certain mean degree (whose value depends on what definition of acquaintance one adopts) and is not noticeably right-skewed [19,20]. The typical explanation for this result is that there is a recurring cost in terms of time and effort to maintaining a friendship, and given limited resources people can only maintain a certain number of them. Indeed, in cases of networks in which there is little cost, or only a one-time cost, to increasing one's degree, e.g., in networks of sexual contacts [21], highly skewed and possibly power-law degree distributions are seen. In our work, we have assumed, as is usually the case, that there are costs to friendship, and hence vertex degrees are narrowly distributed.

3. The lack of a power-law degree distribution in acquaintance networks also suggests that the preferential attachment mechanism is not an important one. Since most people have about the same number of friends, it makes little difference whether people with more friends attract new ones at a higher rate.

4. Lastly, and perhaps most importantly, social networks show "clustering," also called "transitivity" in the sociological literature. Clustering is the propensity for two of one's friends to be friends also of each other, and is very common in social networks. Growth models of the Web show weak clustering—the probability C that two neighbors of a given vertex will be neighbors also of one another, also called the clustering coefficient, is greater by a factor of about five than in the corresponding baseline network, a random graph in which edges are assigned to vertices completely at random [3]. However, in social networks C can be thousands or millions of times greater than in the corresponding random graph [12,22]. The importance of this result has been emphasized extensively in the literature [23], and certainly any reasonable model of social network growth should incorporate it.

Taking each of these points into account, we propose the following as a minimal set of features that a model of social network evolution should have.

1. Fixed number of vertices: we consider a closed population of fixed size.

2. Limited degree: the probability of a person developing a new acquaintance should fall off sharply once their current number of friends reaches a certain level.

3. Clustering: the probability of two people becoming acquainted should be significantly higher if they have one or more mutual friends.

4. Decay of friendships: Given that the number of vertices is fixed, and the degree is limited, friendships must be broken as well as made if the evolution of the network is not to stagnate.

In the following sections we propose and study two models which have these properties. The first model is quite general in its formulation, allowing for arbitrary functional forms representing people's propensity to form friendships. This model makes a reasonable stab at realism in its representation of network evolution, but turns out to be cumbersome to simulate and analytically intractable. So we also propose a second model, a much simplified version of the first, which reproduces the characteristic features of the first model, albeit in stylized form, and which can be simulated with considerably greater efficiency. This second model is similar in its level of sophistication to the previously studied models of growth of the Web and other networks, and may be similarly amenable to analytic treatment, although we have not attempted an analytic treatment here.

III. MODEL I

We consider the following mechanism for the growth of social networks. Pairs of individuals meet with a probability per unit time which depends on how many mutual friends they have. If they have no mutual friends, then there is only a very small chance of their meeting, but if a pair have a friend in common, then their chance of meeting is increased substantially. In the particular case of networks of collaboration between scientists, the existence of this effect has been verified by direct empirical measurement [24]. The presumed mechanism which drives it is a social one: people often introduce pairs of their friends to one another, either deliberately, or simply by virtue of bringing them together in the same place.

We also place a limit on the number z of friends that people can have by arranging for the probability of their forming new friendships to fall off beyond some cutoff point z^*.

If only these two mechanisms were in place, we would get a network which would grow until all or most people had about z^* friends, and then stop growing. The structure of the community would not change after its initial formation. In fact, Watts [23] has described just such a model, his "α-model," in which a hard upper limit is placed on the number of acquaintances an individual can have, and the model does indeed stop evolving once everyone has this many. In the real world, however, social networks do not stop evolving. Although there really does appear to be an upper limit to the numbers of people's friends, the network continues to change because friendships are broken as well as made. To account for this, we propose an obvious mechanism: we propose that even after a pair of people become acquainted, they still need to meet regularly in order to maintain that acquaintance. If they cease meeting, their acquaintance ceases as well. (Many people say that they have friends they rarely see but with whom they nonetheless remain acquainted. We discount such friendships from our model, since there is essentially no cost to such a friendship, and hence it does not fall under the influence of our upper limit on friendship number.)

Thus our model has three components: (1) friendships form when people meet, which happens preferentially between pairs of people who have one or more mutual acquaintances; (2) the number of a person's friendships is limited; (3) friendships decay and disappear if the two people in question do not meet in a regular basis. In detail we implement these components as follows.

The probability per unit time p_{ij} of a given two people, i and j, meeting depends on two factors: (1) the number of friends z_i and z_j each person already has and (2) the number m_{ij} of mutual friends shared by both. We represent these factors by functions f and g thus:

$$p_{ij} = f(z_i)f(z_j)g(m_{ij}). \tag{1}$$

The function $f(z)$ is presumably large and roughly constant for small z, but falls off sharply around the transition value z^*. One possible functional form with these properties is the Fermi function:

$$f(z_i) = \frac{1}{e^{\beta(z_i - z^*)} + 1}, \tag{2}$$

and we have used this form for the simulations described here. The temperature-like parameter β controls the sharpness of the fall-off at z^*.

The function $g(m)$ represents the expected increase in the likelihood that two people will meet if they have one or more mutual friends. In recent studies of collaboration networks [24] this function was measured directly, and found to be well fit by the simple exponential form:

$$g(m) = 1 - (1 - p_0)\mathrm{e}^{-\alpha m}, \tag{3}$$

where p_0 represents the probability of a chance meeting between two people with no mutual acquaintances and the parameter α controls the rate at which $g(m_{ij})$ increases.

The forms for f and g chosen here are somewhat *ad hoc,* but we have experimented with other forms and found the qualitative predictions of the model to be the same. Amongst other things, this provides some justification for the simpler model presented in Section IV, which does not contain arbitrary functions of this sort.

And what happens once two people meet? Friendships do not merely exist or not exist: we have friends whom we see every week, once a month, or whom we gradually lose touch with. We represent this in our model by giving each friendship a strength. When two people i and j meet, the strength s_{ij} of the connection between them is set to 1. Then as time passes and they do not meet again the strength decays exponentially $s_{ij} = \mathrm{e}^{-\kappa\Delta t}$, where Δt is the time since they last met and κ is an adjustable parameter of the model. If they do meet again, s_{ij} is set back to 1. Thus the time averaged strength of a connection is measure of how often people meet.

For the purposes of constructing pictures of our networks, we normally place a threshold on the connections, and consider only those whose strength is greater than that threshold to be active friendships. For the figures in the following section the threshold value used was 0.3. The same criterion is used for counting numbers of mutual friends and for calculating clustering coefficients.

A. Results

We have simulated the model described in the previous section for networks of up to 1000 vertices. In order to pick pairs of individuals with the correct probability per unit time, Eq. (1), we use a continuous-time Monte Carlo method (also called an "n-fold way" algorithm) [25]. Simulations can be initialized in a variety of ways: one can for example start with a random graph in which each vertex has average degree z^*. In our simulations, we started with an empty network, having no edges, and then allowed edges to appear with the decay parameter κ set to zero, or a very small value. After each individual has formed about z^* friendships, the evolution of the network then stagnates because no more edges can be either added or removed. At this point we set κ to a larger, more realistic value and watch the subsequent evolution of the network. Statistics such as the clustering coefficient C and the average path length are measured as a function of time.

Figure 1 shows a snapshot of the network from a simulation with $N = 250$ vertices with $\kappa = 0.01$, $\beta = 5$, and $z^* = 5$. There are a number of interesting features of this network. First, it has a high clustering coefficient of $C \simeq 0.45$. The clustering coefficient for a random graph of the same size and number of edges is roughly

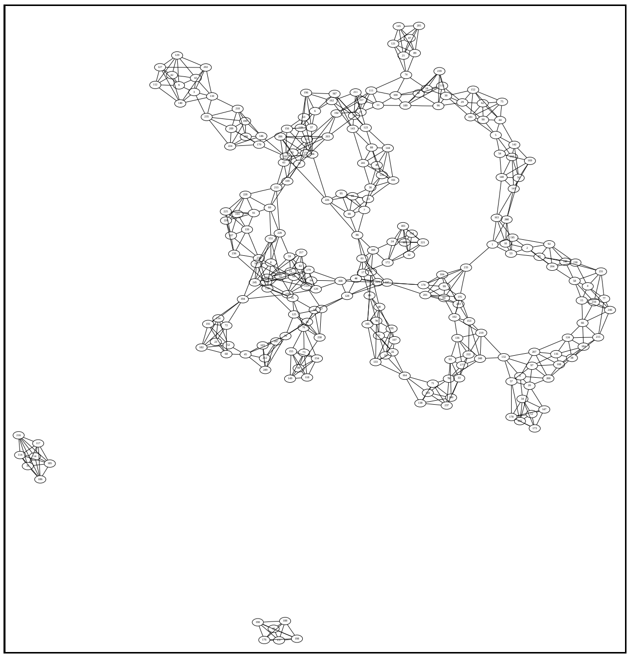

FIG. 1. Sample network generated by model I. In this simulation there where $N = 250$ vertices, and $\kappa = 0.01$, $\beta = 5$. During the course of the simulation the isolated components did not join the main component.

$\frac{z^*}{N} = 0.02$. Thus our model clearly reproduces the strong clustering of real social networks. This however is no great surprise; the primary mechanism of network evolution in the model—the meeting of pairs of people with high numbers of mutual friends—is clearly geared precisely towards creating such strong clustering. Nonetheless, our results provide a demonstration that such mechanisms can produce clustering in social networks.

A less trivial outcome which emerges from our model is the formation of clearly defined communities. As Fig. (1) shows, there are groupings of vertices in the network among which there is a high density of connections, and between which there are few connections. Most of these communities are joined together in one large connected component, but there are also a small number of communities which have no connection with the main body of the graph (although as we will see shortly, the existence of such islands depends on the precise choice of the parameters in the model).

One way to examine the community structure quantitatively is to assign a "connection strength" to every pair of vertices in the network and then examine the component structure of the graph as edges are added between vertex pairs in order of decreasing strength, starting from a graph with no edges. Here we use this method with a connection strength which is a weighted sum of the number of different paths through the network between vertex pairs, with shorter paths weighted more heavily than longer ones [26]. (The paths we consider need not be node- or edge-independent, although connection strengths based on node-independent paths have been considered elsewhere [27].) To visualize the community structure extracted by this calculation, we draw a hierarchical tree showing the order in which components form and are joined together. Such trees have been used widely in social network analysis, where they are sometimes called "dendrograms" [28,29], and occasionally in physics too [30].

Figure 2 shows the hierarchical tree for a network generated by our model with parameters as in Fig. 1. The tree reveals strong community structure in the network: substantial connected components appear early in the clustering process (lower down in the tree) and persist until late (higher up). By contrast, a typical hierarchical tree for a random graph shows a few small components forming early in the process but these quickly combine into one giant component, with subsequent edges only serving to connect individual nodes to the giant component. The strong communities seen in Fig. 2 are absent in the random graph.

The formation of communities is of course seen in real-world social networks but was not a specific design feature of our model. We can however explain it in terms of the model's dynamics as follows. If, during the growth of our network, a region forms in which there is a higher than average density of connections between vertices, then there will also be more pairs of vertices in that region which have common acquaintances. Hence, new friendships will preferentially form between those pairs and so the density of connections in the region will become higher still. Thus small initial fluctuations in network density can form the seeds for the growth of tightly connected communities.

Furthermore, communities in our model are self-sustaining structures. Within communities many pairs of people necessarily have mutual friends, and the communities thus contain a high density of "triangles" of friendship. (The clustering coefficient can in fact be defined precisely as a measure of the density of such triangles [31].) A triangle is a self-sustaining structure in our model. Each pair of vertices in a triangle has a mutual neighbor in the third vertex, and as a result meetings between each pair take place at a much higher rate than between randomly chosen pairs of vertices in the graph. Thus the strength of the connection between each pair of vertices is repeatedly reinforced. This means that edges within a community have a greater lifetime on average than those between communities—the community structure is both created by mutual friendships and helps to sustain them.

B. Other behaviors of the model

The behavior described above is typical of a large region of the parameter space of this first model. However,

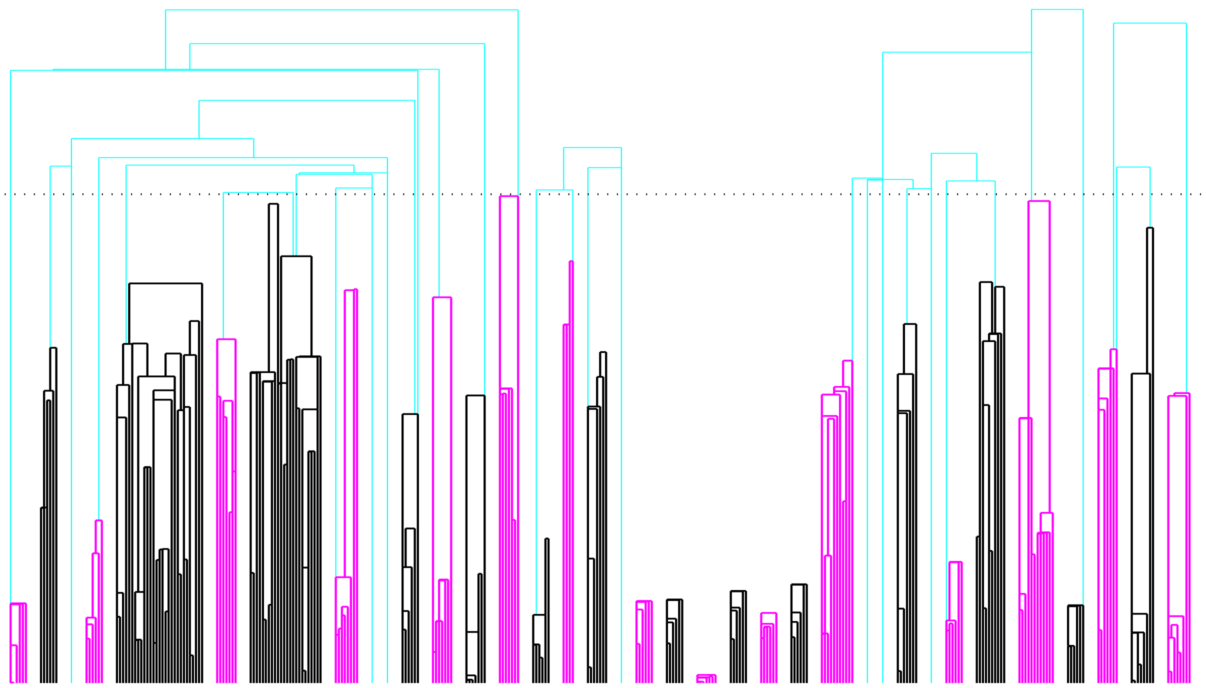

FIG. 2. The hierarchical tree or dendrogram showing community structure for a network with $N = 250$, $\kappa = 0.01$, $\beta = 5$, calculated as described in the text. In this case, we have designated separate communities by whether their lowest connecting path in the tree falls above or below a specified threshold, indicated by the horizontal dotted line, and the components have been spaced out and colored to illustrate this designation. The threshold value is chosen using a criterion based on the density of edges within components as described in Ref. [26].

for extreme values of the parameters other behaviors are seen, most of them rather unlike the behaviors of real social networks.

Consider, for instance, the extreme cases where the decay rate κ is either very slow or very fast. Figure 3 shows the time evolution of the clustering coefficient C for simulations with $\kappa = 0.001$ and $\kappa = 20$. With an extremely slow decay (top panel in the figure), established connections decay very little before being reinforced by a repeat "meeting" of the two corresponding vertices. Thus connections rarely disappear once established, and the evolution of the network stagnates. We still get a high clustering coefficient, as the figure shows, but it has almost no fluctuation with time, because the topology of the network is not changing. This roughly reproduces the behavior of Watts' α-model [23]. At the other extreme, very rapid decay of connections prevents the formation of any lasting friendships, producing a network which is essentially a random graph with no clustering or community structure (lower panel in Fig. 3).

Between these two extremes variation of the parameters produces slight variations on the basic behavior discussed in Section III A. In Fig. 1, for example, we saw the formation of well connected communities, some of which could be isolated from the rest of the graph. The length of time for which this isolation persists depends on the decay parameter κ as well as the parameter p_0 which governs the probability of a chance meeting between two people with no common acquaintances. If κ is increased, then friendships decay more quickly, leaving some vertices with room for an extra edge. And if p_0 is sufficiently high, then edges will occasionally be formed between two isolated components of the graph. Once one such edge forms, there exist other pairs of vertices in the two components which have a common neighbor, and hence more edges will quickly form between the components. In other words, once a single friendship forms between different communities, others usually follow. Note however that, as we saw above, higher decay rate κ leads to a lower clustering coefficient, and in fact the decrease in the clustering coefficient can be seen as clear "steps" when different communities in the graph merge (see Fig. 4). Thus it appears that communities which are less tightly connected internally (lower C) allow for new connections to appear more easily between separate communities.

We can also vary the value of the temperature parameter β. Decreasing β allows a vertex more flexibility about its degree—it can add an extra edge more easily. If β is decreased while keeping the other parameters fixed, we find that the basic community structure of the graph remains roughly constant, as does the clustering coefficient, but the pattern of connections within communities is continually changing. New edges are added to vertices occasionally, and edges are removed to bring the mean degree

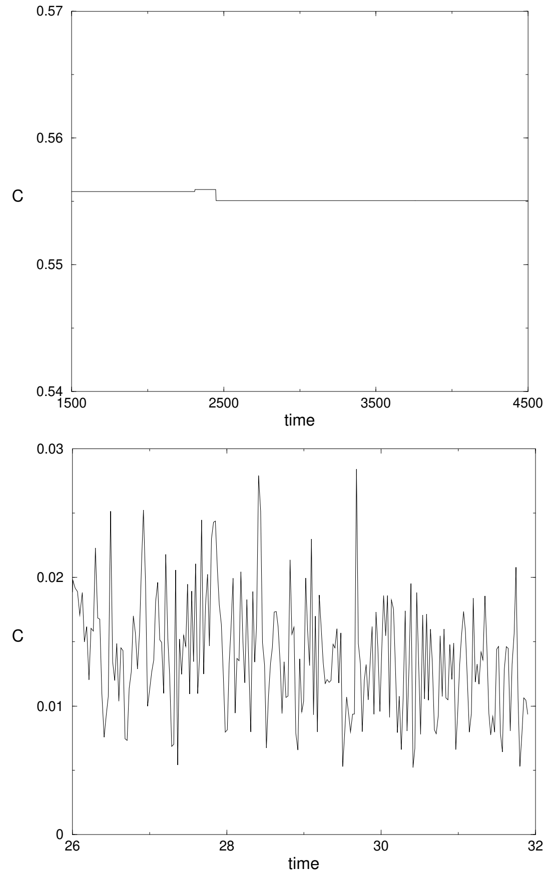

FIG. 3. Clustering coefficient as a function of time for $\kappa = 0.001$ (top) and $\kappa = 20$ (bottom). In the former case the clustering coefficient is high, but hardly fluctuates, since the network topology is almost constant. In the latter case, there is much fluctuation, but the value of C rarely rise above that for a random graph of the same size and edge density. (C would take the value 0.019 in the random graph.

back to about z^*. But the edges which are removed are not necessarily the same ones that were added. In Fig. 5 we show how the pattern of edges evolved in one such community during one of our simulations. This simulation seems to mimic a situation in which the exclusivity of communities is maintained, but the friendships within those communities are brief and casual, which may be a reasonable representation of certain types of social organization.

IV. MODEL II

The model described in the first part of this paper has many adjustable parameters, as well as the functions f and g, whose forms are infinitely variable. While a large number of free parameters allows us a lot of flexibility to study the behavior of the model and may, in addition, make the model a more accurate representation of the real world, we find in fact that the selection of behaviors which we get from the model is limited to a few general classes, as described above. This suggests that it may be possible to formulate a less baroque model, one whose definition and dynamics are simpler, and still retain most of the interesting behavior. In this section we do just this.

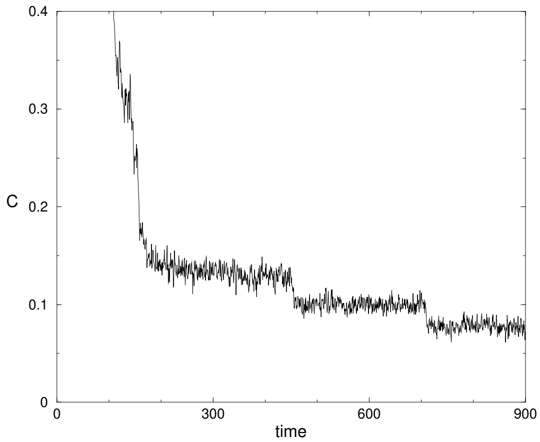

FIG. 4. Clustering coefficient as a function of time for $\kappa = 0.5$, $p_0 = 0.0001$, $\beta = 6.67$. The network settles into distinct groups that seem to be stable until individuals from separate communities become acquainted, causing two groups to merge and thus lowering C.

Our simplified model incorporates all four of the crucial features outlined in Section II, but in a simplified fashion as follows. First, all connections between vertices are only either present or absent—there is no longer any concept of connection strength. The exponential decay of connection strength from Section III is replaced by a constant probability γ per unit time that an existing connection will disappear. Thus out of any initial group of connections, $e^{-\gamma t}$ of them will remain after time t, in the absence of any other processes.

Second, "meetings" occur between pairs of individuals represented by vertices at a rate r which is simply linear in their number m of mutual friends: $r = r_0 + r_1 m$. If a pair meet and there is not already a connection between them, a new connection is established unless one of them already has z^* connections, in which case nothing happens. In other words, z^* forms a hard upper limit on the degree z of any vertex, beyond which no more edges can be added.

Apart from being conceptually much simpler than our first model, this model is also much easier to simulate. Instead of having to use a complicated and inefficient continuous time simulation method, the model can be simulated directly using the following algorithm.

Let $n_p = \frac{1}{2}N(N-1)$ be the number of pairs of vertices in the network. Let $n_e = \frac{1}{2}\sum_i z_i$ be the number of existing edges, where z_i is the degree of the i^{th} vertex. And let $n_m = \frac{1}{2}\sum_i z_i(z_i - 1)$ be the total number of mutual neighbors of pairs of vertices in the network.

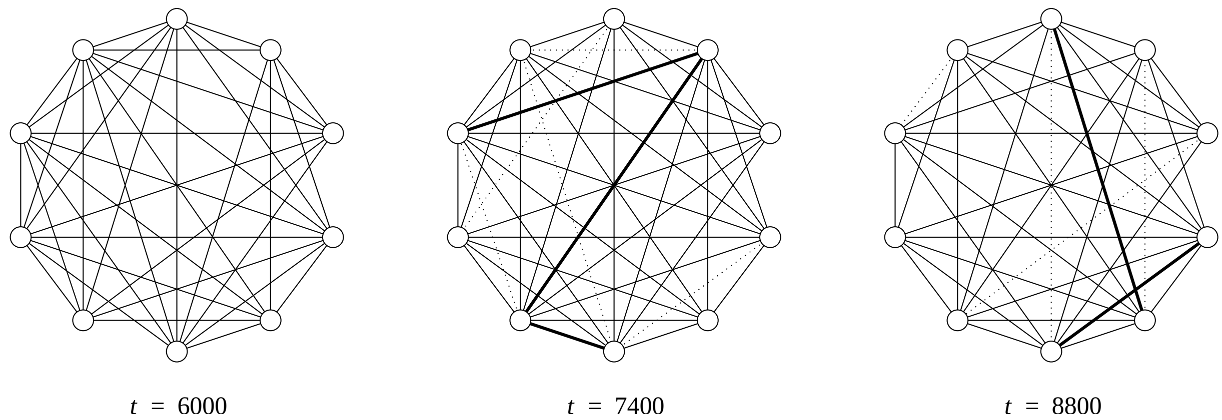

FIG. 5. Time evolution of the edges within one component for $\beta = 1.25$ and $\kappa = 0.01$. Dotted lines indicate connections that existed in the previous frame and have since decayed. Bold lines indicate new connections. All new connections are made with vertices already included in this group.

1. At each time-step, we choose $n_p r_0$ pairs of vertices uniformly at random from the network to meet. If a pair meet who do not have a pre-existing connection, and if neither of them already has the maximum z^* connections then a new connection is established between them.

2. At each time-step, we choose $n_m r_1$ vertices at random, with probabilities proportional to $z_i(z_i - 1)$. For each vertex chosen we randomly choose one pair of its neighbors to meet, and establish a new connection between them if they do not have a pre-existing connection and if neither of them already has the maximum number z^* of connections.

3. At each time-step, we choose $n_e \gamma$ vertices with probability proportional to z_i. For each vertex chosen we choose one of its neighbors uniformly at random and delete the connection to that neighbor.

It is straightforward to convince oneself that repetition of these steps simulates the dynamics of the model proposed above.

As before, the network is initialized by starting with no edges, and running the first two steps (addition of connections) without the third (breaking any connections) until all or most vertices have degree z^*. Then all three steps are used for the remainder of the simulation.

Figure 6 shows a sample network from a simulation of this model with $N = 250$, $r_0 = 0.0005$, $r_1 = 2$ (about 4850 pairs per timestep), $\gamma = 0.005$, and $z^* = 5$. As with our first model, there are clearly identifiable communities in the network, mostly connected together in a single giant component, although there are also communities that are well-connected internally but disconnected from the rest of the graph. The values of γ and r_1 were chosen so that connections based on mutual friendship have some stability over time: even when they get broken, they are likely to be remade quickly. This mechanism replaces the "reinforcement" mechanism of the first model. However, there is always some possibility that broken links will not be remade and other links will appear instead, allowing for evolution of the network structure over time at a rate dependent on the parameter values. The network shown in Fig. 6 is also highly clustered, having a clustering coefficient of $C = 0.53$, where the corresponding random graph would have $C = 0.02$.

Most of the types of behavior seen in our first model can be reproduced by appropriate choices of parameter values in this second model. For example, extremely high or low values of the decay parameter γ produce either highly fluctuating structures with clustering not noticeably different from that of a random graph, or highly clustered graphs which are stagnant and barely evolve. Other parameter changes can affect the stability of the island communities in the graph over long periods, or vary the rate at which connection patterns within communities vary.

V. DISCUSSION

What can we learn from results of the type presented here? The primary lesson is that complex and intuitively reasonable patterns of social network structure and evolution can emerge from very simple rules. Furthermore, the general form of those patterns is not strongly influenced by the microscopic details of the rules, so that most of the interesting behaviors can be reproduced in a much simplified model which is clearly not a realistic representation of real-world social behaviors.

The crucial features which we find necessary to produce plausible networks are three in number: (1) meetings between pairs of individuals, giving rise to friendships, at a rate which is high if a pair has one or more mutual friends and low otherwise; (2) decay of friendships between pairs of individuals who no longer meet or

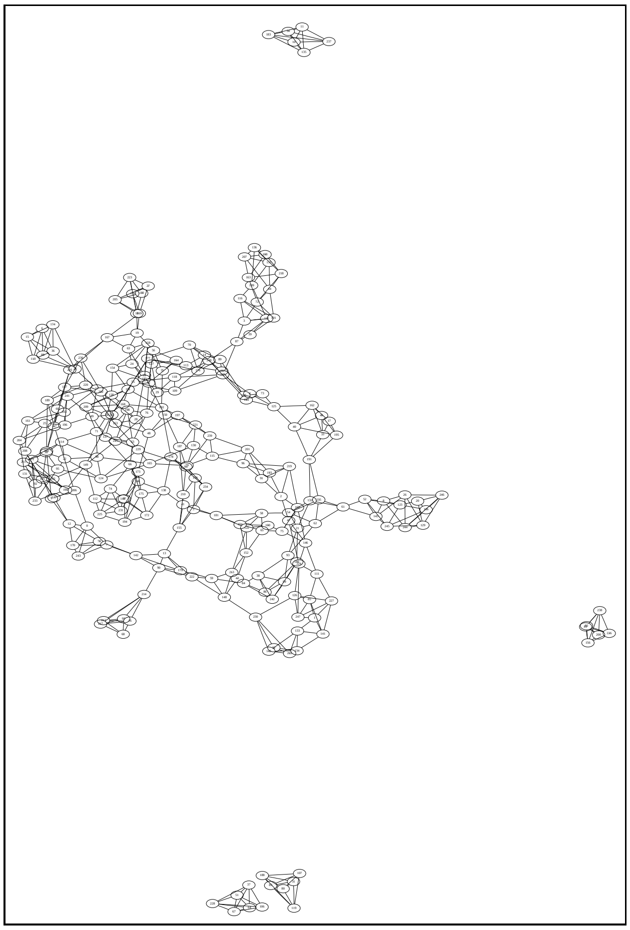

FIG. 6. Network structure generated in a run of our second model with $N = 250$, $r_0 = 0.0005$, $r_1 = 2$ (about 4850 pairs per timestep), $\gamma = 0.005$, and $z^* = 5$.

rarely do so; (3) an upper limit (either soft or hard) on the number of friendships an individual can maintain.

These rules are quite different from the rules which have been used to model the evolution of graphs in other arenas, such as the evolution of the World-Wide Web. While the evolution of the Web appears to be dominated by preferential attachment—vertices with many edges accrue new ones at a higher rate than those with few—we conjecture that social network growth is dominated by the introduction of future acquaintances to one another by mutual friends. As a result, almost everything about the resulting graphs is different between the two cases. Where preferential attachment yields a graph with a power-law degree distribution, the limit we place on vertex degree in our social networks creates a sharply peaked distribution. Where graphs grown with preferential attachment show clustering coefficients only slightly higher than the corresponding random graph, our social network models show very high clustering coefficients, similar to those seen in real-world social networks. And where the structure of the Web and similar networks is dominated by their rapid growth, the structure of our social networks is dominated by constant rewiring of connections between existing vertices, with the addition of new vertices not playing a major role.

But perhaps the most intriguing feature of our models is that they show community structure in the networks they generate: there are groups of vertices with many connections between their members and few connections to vertices outside the group. For some parameter values, these communities even separate entirely and there are no connections between them at all. Community formation is certainly a feature of real social networks also, and it is interesting to see that communities can arise from simple local growth rules only. We are not aware of any study that has shown the existence of such communities in preferential attachment models. Interestingly, however, the real World-Wide Web *does* show community structure [32]. Perhaps then a realistic model of the growth of the Web should include some additional elements similar to those in our social network models in order to capture community formation fully.

This paper represents only a first attempt at modeling the evolution of the structure of social networks. There are many possible directions for further study. One can ask whether there are important mechanisms of network growth which we have missed out of the present models, or whether even our simplest model is still more complicated than it need be. Perhaps the three basic rules given here are not all necessary? It would also be useful to acquire a detailed understanding of how the parameters of the models relate to one another—what is the structure of the phase diagram for these models? And is an analytic approach to these or similar models possible? It would be helpful if we could understand the qualitative behaviors seen in our simulations in terms of analytic calculations, either approximate or exact. We hope that the first steps taken here will encourage others to look at these questions in more depth.

The authors would like to thank Cris Moore and Duncan Watts for useful and enjoyable conversations. This work was supported in part by the National Science Foundation and by Intel Corporation.

[1] M. E. J. Newman, "Models of the small world," *J. Stat. Phys.* **101**, 819–841 (2000).
[2] S. H. Strogatz, "Exploring complex networks," *Nature* **410**, 268–276 (2001).
[3] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.,* in press. Also `cond-mat/0106096`.
[4] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the world-wide web," *Nature* **401**, 130–131 (1999).
[5] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Ra-

jagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," *Computer Networks* **33**, 309–320 (2000).

[6] L. A. Adamic and B. A. Huberman, "Power-law distribution of the World Wide Web," *Science* **287**, 2115 (2000).

[7] P. L. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of growing random networks," *Phys. Rev. Lett.* **85**, 4629–4632 (2000).

[8] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, "Are randomly grown graphs really random?" `cond-mat/0104546`.

[9] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science* **286**, 509–512 (1999).

[10] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "The Web as a graph: Measurements, models and methods," Proceedings of the International Conference on Combinatorics and Computing (1999).

[11] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Structure of growing networks with preferential linking," *Phys. Rev. Lett.* **85**, 4633–4636 (2000).

[12] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature* **393**, 440–442 (1998).

[13] M. Barthélémy and L. A. N. Amaral, "Small-world networks: Evidence for a crossover picture," *Phys. Rev. Lett.* **82**, 3180–3183 (1999).

[14] M. E. J. Newman and D. J. Watts, "Renormalization group analysis of the small-world network model," *Phys. Lett. A* **263**, 341–346 (1999); "Scaling and percolation in the small-world network model," *Phys. Rev. E* **60**, 7332–7342 (1999).

[15] R. Monasson, "Diffusion, localization and dispersion relations on small-world lattices," *Eur. Phys. J. B* **12**, 555–567 (1999).

[16] A. Barrat and M. Weigt, "On the properties of small-world network models," *Eur. Phys. J. B* **13**, 547–560 (2000).

[17] M. E. J. Newman, C. Moore, and D. J. Watts, "Mean-field solution of the small-world network model," *Phys. Rev. Lett.* **84**, 3201–3204 (2000).

[18] S. N. Dorogovtsev and J. F. F. Mendes, "Scaling behaviour of developing and decaying networks," *Europhys. Lett.* **52**, 33–39 (2000).

[19] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proc. Natl. Acad. Sci.* **97**, 11149–11152 (2000).

[20] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," in preparation.

[21] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, "The web of human sexual contacts," *Nature,* in press.

[22] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2001).

[23] D. J. Watts, *Small Worlds,* Princeton University Press, Princeton (1999).

[24] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Phys. Rev. E,* in press. Also `cond-mat/0104209`.

[25] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics,* Oxford University Press, Oxford (1999).

[26] M. Girvan and M. E. J. Newman, in preparation.

[27] D. R. White and M. E. J. Newman, "Fast approximation algorithms for finding node-independent paths," in preparation.

[28] S. Wasserman and K. Faust, *Social Network Analysis,* Cambridge University Press, Cambridge (1994).

[29] J. Scott, *Social Network Analysis: A Handbook,* 2nd Edition, Sage Publications, London (2000).

[30] R. N. Mantegna, "Hierarchical structure in financial markets," *Eur. Phys. J. B* **11**, 193–197 (1999).

[31] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Phys. Rev. E,* in press. Also `cond-mat/0007235`.

[32] D. Gibson, J. Kleinberg, and P. Raghavan, "Inferring Web communities from link topology," in *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia* (1998).