PHYSICAL REVIEW E **73**, 065101(R) (2006)

Competitive cluster growth in complex networks

André A. Moreira, Demétrius R. Paula, Raimundo N. Costa Filho, and José S. Andrade, Jr.
Departamento de Física, Universidade Federal do Ceará, 60451-970 Fortaleza, Ceará, Brazil
(Received 10 March 2006; published 1 June 2006)

In this work we propose an idealized model for competitive cluster growth in complex networks. Each cluster can be thought of as a fraction of a community that shares some common opinion. Our results show that the cluster size distribution depends on the particular choice for the topology of the network of contacts among the agents. As an application, we show that the cluster size distributions obtained when the growth process is performed on hierarchical networks, e.g., the Apollonian network, have a scaling form similar to what has been observed for the distribution of a number of votes in an electoral process. We suggest that this similarity may be due to the fact that social networks involved in the electoral process may also possess an underlining hierarchical structure.

DOI: 10.1103/PhysRevE.73.065101 PACS number(s): 89.75.Hc, 05.50.+q, 89.75.Da, 89.75.Fb

The sedimentation of new trends and ideas in large social communities can have a profound impact in the life of individuals. An instance where the dynamics of opinion formation may be of major importance is the democratic elections of representatives. In addition, in the electoral process every agent is called to give her/his opinion in an anonymous way and the statistical results are easily accessible [1]. This makes elections ideal systems to researchers interested in studying the process of opinion formation. Since, typically, an individual is more likely to listen to someone they have a personal contact with, the process may also be driven by a mouth-to-mouth interaction besides from massive political campaigns based on media programs. In this way, the spreading of an opinion follows a pattern similar to the spreading of an epidemic process [2]. One should also expect that the particular structure of the network of contacts among the agents of a society may have an impact on the way the opinions propagate.

The intricate structure of the interactions of many natural and social systems has been the object of intense research in the new area of complex networks. Most of the effort in this area has been directed to find the topological properties of real world networks [3–7] and understanding the effects that these properties cast on dynamical processes taking place on these complex networks [8–11]. For instance, the small-world characteristic [3], where each node of the network is only a few connections apart from any other, permits a quick spreading of information through the network, being fundamental in processes of global coordination [12] and feedback regulation [13].

Another property commonly studied in complex networks is the degree distribution $P(k)$, that gives the probability with which an arbitrary node is connected to exactly k other nodes. One relevant characteristic often observed in complex networks is a scale-free degree distribution [4], namely, a distribution that follows a power law, $P(k) \sim k^{-\gamma}$, with an exponent typically in the range $2 < \gamma < 3$. Such broad degree distribution has a dramatic effect in many dynamical processes. In the spreading of infectious diseases, for example, it has been shown that when the infection is mediated by a scale-free network, any infection rate above zero results in a positive fraction of infected individuals [14]. It was recently suggested that another universal characteristic of real-world networks is a structure of communities, where smaller communities in the network are joined to larger communities by highly connected nodes that play the role of local hubs [15]. This structure may be related to the self-similar characteristic observed in some complex networks [16].

In this Rapid Communication we investigate a dynamic process of competitive cluster growth in complex networks. In this process many alternative and self-excluding states are accessible to the nodes of the network. We say that nodes in the same state belong to a cluster, and each of these clusters competes with the others to reach a larger part of the network. This idealized mechanism can be thought of as a model for a variety of different processes that take place in real networks. For example, one can think of each cluster as the part of a population that has been infected with a certain strain of a virus. Alternatively, the clusters may represent alternative opinions in a social group. We study the distribution of the fraction of the network occupied by an arbitrary cluster. Our results show that the network topology has great influence over the behavior of the cluster size distributions.

Our model for competitive cluster growth is described as follows: In a fixed underlying network of interactions, each agent starts undecided and eventually takes the opinion of one of its decided neighbors. For simplicity, we assume that, once the agent decides for one opinion, this opinion remains unchanged during the growth process. The process is mediated by a substrate network with N nodes; in the first moment, a number n_s of nodes is chosen at random to be the seeds of the spreading process, with the density of seeds being n_s/N. Each seed will be the first node of a cluster. Then, the clusters grow by incorporating nodes that are neighbors of these seeds and have not yet been assigned to any other cluster. Once a node is incorporated to a cluster it will stay in this cluster until the end of the process, and only the nodes not belonging to any of the existing clusters are accessible to the growth process. We will refer to these as accessible nodes. The growth process takes place in discrete steps. At each step, we randomly select a pair of connected nodes, one belonging to a cluster and the other that is accessible to growth. The accessible node is subsequently incorporated to the same cluster as its neighbor. Figure 1 presents a pictorial description of our growth model. We suggest that

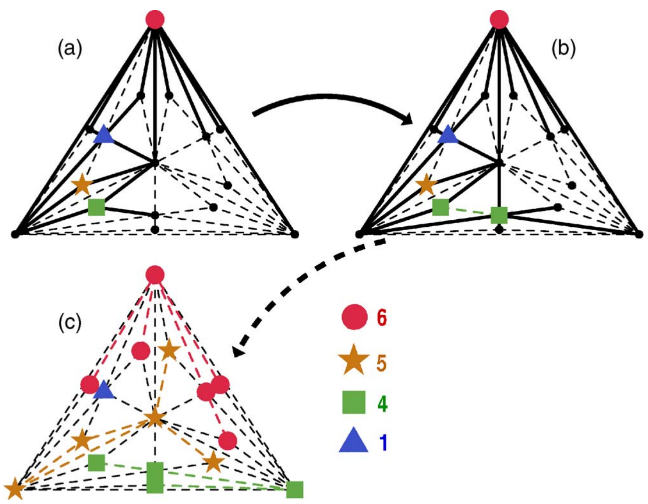

FIG. 1. (Color online) Pictorial description of our model for a competitive cluster growth process. To make a clear picture of our model, we used as a substrate network for the growth a third generation Apollonian network [19]. The matching graph form of this network makes it particularly suitable for a plane representation. In the first step of the growth model (a), a few nodes (the large circle, square, triangle, and star) are chosen to be the seeds of the growing clusters. Each of the seeds is the first node of a different cluster. All the remaining nodes (small dots) do not get assigned to any of the clusters in the beginning; these nodes will be the ones accessible to the growth. The dashed lines link either a pair of nodes that already belong to one of the clusters or a pair of nodes that are not in any cluster yet. At this time step these lines do not participate in the growth process. The thick lines link one of the seeds to an accessible node, and any of the thick lines can be chosen with the same probability to channel the growth of a cluster. In this way, the growth rate of a particular cluster is proportional to its perimeter, that is, the number of connections from one of the nodes already incorporated into the cluster to an accessible node. In the second step (b) a new node is incorporated into the cluster of squares and new thick lines are added to the perimeter of this cluster. The process continues until every one of the nodes has been incorporated into one of the clusters (c). At this point the size of all the clusters is computed.

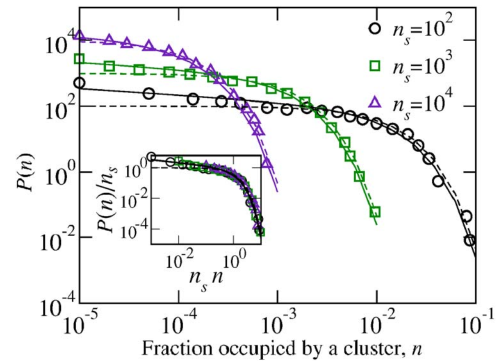

FIG. 2. (Color online) Normalized cluster size distributions obtained for the competitive growth model under a random ER network topology. These distributions were obtained for networks of 10^5 nodes. We performed a larger number of runs for the cases with less competing clusters in order to have 10^5 cluster samples for each case. The distributions obtained under this topology follow approximately an exponential distribution with a characteristic cluster size depending on the number n_s of competing clusters. The dashed lines are the functions $n_s e^{-n_s n}$. The continuous lines are a fit for the function $P(n)=C n_s^{1-\gamma} n^{-\gamma} \exp(-n_s n)$, with the parameters $C=0.85\pm0.03$ and $\gamma=0.2\pm0.05$ obtained for the collapsed data shown in the inset. This figure shows that for the random networks the form of the distribution is completely determined by the density of nodes chosen to be seeds for the growth process.

this mechanism can be representative of some sort of greedy process where large clusters, with many nodes, will have more connections to accessible nodes and therefore tend to grow faster and increase their perimeters even more. In this way, our model resembles the preferential attachment model for network growth [4]. However, the competitive growth has significant differences from that mode, in the sense that all seeds are present in the beginning of the process and also the fact that some clusters will start to grow before others, resulting in a variety of cluster sizes.

The choice of a particular topology for the network of contacts should affect the growth process. The simplest model for a network is probably the random network model proposed by Erdos and Reny (ER) [17]. In this model, any pair of nodes can be connected with probability ρ, and the degree distribution follows approximately a Poisson form with the average degree given by $\bar{k}=\rho(N-1)$. In Fig. 2 we show the distribution of the fractions of the network corresponding to each cluster $P(n)$ when the spreading process takes place on ER networks. In this case the cluster size distributions are power laws with a small exponent bounded by an exponential cutoff with a characteristic scale depending only on the density of competing clusters in the network.

Next, we study the cluster growth process on a scale-free network. To build the scale-free network we use the so-called preferential attachment method [4]. Different from the random graph, the degree distribution of the networks built with this model has a power-law form, $p_k(k)\sim k^{-\gamma}$, with an exponent $\gamma=3$, followed by an exponential cutoff at a maximum degree, $k_{max}\sim N^{1/(\gamma-1)}$ [18]. Note that a cluster that incorporates one of the most connected nodes in the beginning of the growth process will increase its growth rate by a large factor. Thus, the long tail of this distribution may have a dramatic effect on the cluster growth process. Indeed, the cluster size distribution obtained when the spreading process is done in the preferential attachment network also shows a power-law tail, $p(n)\sim n^{-\alpha}$, with an exponent $\alpha=1.2$ for the case with $n_s=10^2$, as seen in Fig. 3. Similar results were obtained for the distributions of cluster sizes in the transient state of the Snadj model [20]. Interestingly, the competing growth dynamics magnifies the effect of the scale-free distribution producing distributions of cluster sizes that have smaller exponents, and as a consequence, a slower decay than that observed for the degree distribution. The scattering of the data close to the limit $n\rightarrow1$ is due to statistical fluctuations at this low frequency limit as well as to finite size effects. The implication of these heavy-tailed distributions is that now the average cluster size is not a characteristic scale for the process since one finds, with relatively high frequency, clusters that are orders of magnitude larger than the average.

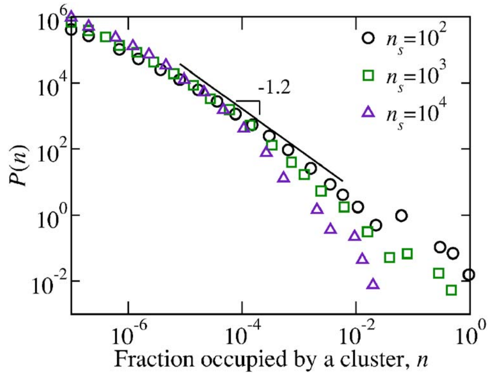

FIG. 3. (Color online) Normalized cluster size distributions obtained when the growth process is performed in networks built with the preferential attachment schema. We used networks with 10^7 nodes and obtained 10^5 cluster samples. Different from the case of random networks, here we cannot produce a data collapse for the distributions. We note, however, that for any value of n_s we have a broad distribution with clusters of all sizes ranging from just a few nodes to most of the network. Due to network size constraints, the distributions display a truncation in the limit of large cluster sizes. As observed in the ER networks, the exponential truncation moves to smaller cluster sizes as the number of seeds increases. The continuous line is a power-law fit, $P(n) \sim n^{-\alpha}$, with the exponent $\alpha=1.2$ obtained for the scaling region of the distribution for $n_s=10^2$. The slope of the distribution becomes more steep as the number of competing clusters in the network grows.

Although the preferential attachment model produces a degree distribution similar to what is found in several real networks, it does not display the self-similarity and hierarchical structure also observed in many of those networks. In the work of Ravasz *et al.* [15], it was suggested that the structure of a network could be probed in a quantitative way by studying the cluster coefficient of its nodes. The cluster coefficient of a node is defined as the probability that two of its neighbors taken at random are connected. A signature of the hierarchical structure would be a cluster coefficient proportional to the inverse of the nodes degree $c(k) \sim k^{-1}$ [15]. Both the preferential attachment and the random networks do not present this property. A model that shows small-world behavior, scale-free degree distribution, as well as a hierarchical structure is the recently proposed Apollonian network [19]. This network is obtained by simply connecting the center of the touching spheres that constitute an Apollonian tiling.

In order to test the effect of a hierarchical structure in the cluster growth we implemented our model in the topology of the Apollonian network. In Fig. 4 we show that the cluster size distributions obtained for the clusters grown in this topology follow power-law behavior with an exponent $\alpha=-1$. Such behavior can be understood with a simple scaling argument. Splitting the Apollonian network in the most connected hubs one finds three smaller networks corresponding to Apollonian networks of a lower generation. Each of these pieces could be split again, producing nine smaller networks with this hierarchical disassembly continuing down to the level of single nodes. After the growth process is performed in a large Apollonian network, one could measure the cluster sizes distribution for each of these generations. It should be expected that, as one goes to higher generations, the distributions approach a limiting form. Note also that only the clusters that reach the three hubs in the corner of the networks are merged when one moves one generation up, and only the size of these few clusters that reach these hubs change from one generation to another. Thus, this process can change the form of the distribution only in the limit of very large clusters, where the frequency is of the order of the inverse of the network size. This means that the distribution of cluster fractions should obey approximately the following similarity relation:

$$P(n) = 3P(3n). \tag{1}$$

A trivial solution for this similarity relation is the Dirac delta $\delta(n)$. We expect the distribution to assume this form if, with probability one, the growth process produces one giant cluster that incorporates a large fraction, $n \approx 1$, as N grows. Besides the delta function, any nontrivial function that satisfies the similarity relation (1) should be of the form C/n, where C is a constant. Although we have not proved that a giant cluster does not appear, our numerical results indicate that, at least for the densities of seeds we have investigated, this is unlikely.

Intriguingly, the scaling found for the distributions obtained with the Apollonian network is similar to those found in the distribution of the number of votes per candidate in the elections in Brazil [1]. This leads us to suggest that the universal behavior observed in the electoral processes may be driven by an underlying hierarchical structure of the social networks describing the interactions among voters.

In summary, we have introduced a model for competitive cluster growth in complex networks. By means of numerical

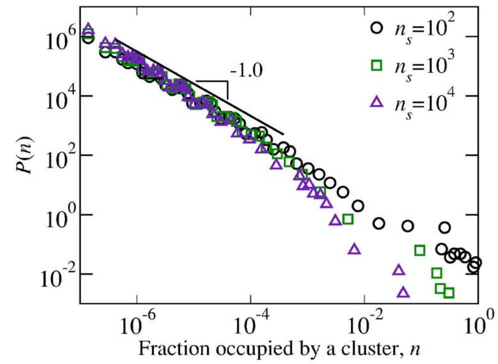

FIG. 4. (Color online) Normalized cluster size distribution when our model is performed on Apollonian networks. The distributions were obtained with the 15th generation of the Apollonian network that corresponds to $N=7174456$. Each curve was sampled from an ensemble of 10^5 clusters. We observe a power-law decay with an exponent $\alpha \approx 1$. The continuous line is a fit for the scaling region of data with $n_s=100$ initial seeds. As in the case of the preferential attachment networks, here we observe a truncation in the large cluster size region of the distributions, with an onset controlled by the number of seeds n_s. The periodical steps observed in the shape of the distributions are reminiscent of the self-similar structure resulting from the hierarchical construction of the Apollonian network.

simulations, we have shown that the fraction of the network accessed by each cluster follows a characteristic distribution that depends on the particular topology of the network. For the case of a random graph we found an exponential decay, while for scale-free networks we found power-law decaying distributions. These results show evidence that there is a straight connection between heavy-tailed degree distributions and power-law cluster size distributions. Although no general scaling law can be found for the exponential crossover in the large cluster region of the distribution, our results indicate that the onset of the crossover is controlled by the density of seeds n_s/N used in the growing process. In the particular case of a hierarchical scale-free network, we observed a decay with a governing exponent $\alpha=1$. Based on this fact, we suggested that the distributions we obtained with our model resemble those of the fraction of votes per candidates observed in the proportional elections in Brazil [1]. In a future work, we intend to investigate the presence of hierarchical structures in the social network governing the process by which voters reach their decisions.

ACKNOWLEDGMENTS

We thank A. D. Araújo, D. M. Auto, and H. J. Herrmann for all the discussions. The authors acknowledge the financial support of the Brazilian agencies CNPq, CAPES, and FUNCAP.

[1] R. N. Costa Filho *et al.*, Phys. Rev. E **60**, 1067 (1999); Physica A **322**, 698 (2003).
[2] M. Gladwell, *The Tipping Point* (Little, Brown, & Co., Boston, 2000).
[3] D. J. Watts and S. Strogatz, Nature (London) **393**, 440 (1998).
[4] L.-A. Barabási and R. Albert, Science **286**, 509 (1999).
[5] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999); H. Jeong *et al.*, Nature (London) **407**, 651 (2000).
[6] L. A. N. Amaral *et al.*, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[7] J. Camacho, R. Guimera, and L. A. N. Amaral, Phys. Rev. Lett. **88**, 228102 (2002).
[8] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[9] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).
[10] C. Castellano, D. Vilone, and A. Vespignane, Europhys. Lett. **63**, 153 (2003).
[11] M. C. Gonzalez, A. O. Souza, and H. J. Herrmann, Int. J. Mod. Phys. C **15**, 45 (2004).
[12] A. A. Moreira *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 12085 (2004).
[13] L. A. N. Amaral *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 15551 (2004).
[14] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[15] E. Ravasz *et al.*, Science **297**, 1551 (2002); E. Ravasz and A. L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[16] E. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).
[17] P. Erdös and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).
[18] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000); S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **63**, 062101 (2001).
[19] J. S. Andrade *et al.*, Phys. Rev. Lett. **94**, 018702 (2005).
[20] A. T. Bernardes, D. Stauffer, and J. Kertesz, Eur. Phys. J. B **25**, 123 (2002); F. Slanina and H. Lavicka, *ibid.* **35**, 279 (2003).