Importance of extremists for the structure of social networks

Philippe Blanchard and Santo Fortunato
Fakultät für Physik & BiBoS, Universität Bielefeld, D-33501 Bielefeld, Germany

Tyll Krüger
Fakultät für Mathematik & BiBoS, Universität Bielefeld, D-33501 Bielefeld, Germany

(Received 27 January 2005; published 20 May 2005)

The structure and the properties of complex networks essentially depend on the way nodes get connected to each other. We assume here that each node has a feature which attracts the others. We model the situation by assigning two numbers to each node, ω and α, where ω indicates some property of the node and α the affinity towards that property. A node A is more likely to establish a connection with a node B if B has a high value of ω and A has a high value of α. Simple computer simulations show that networks built according to this principle have a degree distribution with a power-law tail, whose exponent is determined only by the nodes with the largest value of the affinity α (the "extremists"). This means that the extremists lead the formation process of the network and manage to shape the final topology of the system. The latter phenomenon may have implications in the study of social networks and in epidemiology.

DOI: 10.1103/PhysRevE.71.056114 PACS number(s): 89.75.Hc, 05.10.−a

The study of complex networks is currently one of the hottest fields of modern physics [1–3]. A network (or graph) is a set of items, called *vertices* or *nodes*, with connections between them, called *edges*. Nodes linked by an edge are neighbors and the number of neighbors of a node is called *degree*. Complex weblike structures describe a wide variety of systems of high technological and intellectual importance. Examples are the Internet, the World Wide Web (WWW), social networks of acquaintance, or other connections between individuals, neural networks, food webs, citation networks, and many others.

One of the crucial questions concerns the formation of these structures. Complex networks are in general systems in evolution, with new nodes/edges that get formed and old ones that get removed or destroyed. The currently accepted mechanism finds its roots in an old idea of Price [4], based on the so-called preferential attachment, which means that a newly formed node A builds an edge with a preexisting node with a probability that is proportional to the degree of the latter node. Networks constructed in this way [5–7] have a degree distribution with a power-law tail, as observed in real networks. This simple rule, however, makes implicitly the strong assumption that each node is at any time informed about the degree of all other nodes, which is certainly not true, especially for gigantic systems which contain many millions of nodes, like the WWW. We rather believe that the key behind the building of a connection between a pair of nodes lies essentially in the mutual interaction of the two nodes (almost) independently of the rest of the system: two persons usually become friends because they like each other.

In this paper we have social networks in mind, but nevertheless we will speak generally about networks, as we believe that our model has a more general validity. The mechanism we propose is that any node has some *property* (beauty, richness, power, etc.) by which the others are *attracted*. We indicate the property with a positive number ω, the attractiveness by another positive number α. We assume that high values of ω correspond to a high degree of the property (the most beautiful people, for instance). Both ω and α are attributes of single nodes/individuals, so they take in general different values for different nodes. What we need is a knowledge of the distribution of ω and α in the network. For the property ω, distributions that vanish for high values of ω, like exponentials or power laws, are realistic. As far as the affinity α is concerned, it is less clear which distributions can be considered plausible, therefore we tested several possibilities. We remark that the idea that the nodes have individual appeal already exists in the literature on complex networks. Bianconi and Barabási [8] assigned a parameter η called "fitness" to each node of the network and the linking probability becomes proportional to the product of the degree with the fitness of the target node. In the same framework, Ergün and Rodgers [9] proposed a different ansatz for the linking probability, which in their case is proportional to the sum of the fitness and the degree of the target node. Both models, however, are based on preferential attachment. Caldarelli *et al.* [10] proposed instead a variation of the fitness theme which eliminates preferential attachment, so that the formation principle of networks lies in the attraction nodes exert on each other by virtue of their individual quality/importance, which is actually in the spirit of our work. So, in [10], the linking probability is simply a function of the fitness values of the pair of nodes, and several possible choices for this function are introduced and discussed.

Our expression for the probability p_{AB} of a node A to build an edge with a node B is also a function of the individual attributes of the nodes, ω and α. We adopt the ansatz

$$p_{AB} = \frac{c_B}{[\phi(\omega_B)]^{\alpha_A}}, \tag{1}$$

where c_B is a normalization constant and $\phi(\omega)$ the distribution function of the property ω, whereas we will indicate with $\psi(\alpha)$ the distribution function of the affinity α. What (1) says is that the pairing probability is inversely proportional to the relative frequency of the property ω in the network. Thinking again of a social system, the idea is that there is a general tendency to be more attracted by those subjects who are characterized by high values of ω. In a network of sexual relationships, for instance, the best looking people usually have the greatest chances to be chosen as sexual partners. We believe that the choices of the people are not influenced by the absolute importance of ω, which is a vague and abstract concept, but rather by the perception of the importance of the property ω within the society, which is related to its distribution. This is why we associated the pairing probability to the relative frequency $\phi(\omega)$ and not directly to ω, at variance with [10]. Accordingly, the larger ω, the lower the occurrence $\phi(\omega)$ of that degree of the property in the network, and the edge-building probability gets higher. On the other hand, for a given node B, characterized by its property ω_B, the other nodes will feel an attraction towards B, which varies from one subject to another. This modulation of the individual attraction is expressed by the exponent α_A in (1). The probability p_{AB} increases with α_A, justifying the denomination of "affinity" we have given to the parameter α. The parameter α must be taken in the range $[0, 1]$ for normalization purposes [11]. The normalization constant c_B must be chosen such that the sum of the bond probability p_{AB} over all (target) nodes B be one, so c_B only depends on the number of nodes N and the affinity α_A. We also remark that the expression (1) is not symmetric with respect to the inversion $A \leftrightarrow B$. Most mechanisms of network growth, including that of Barabási and Albert, distinguish between the active node and the target node. This does not necessarily mean that the links are directed from the active to the target node; however, in this paper we will distinguish between in degrees and out degrees and when we speak of degree distribution, we mean the distribution of the in degrees (the out-degree distribution is a constant function).

Our simple model is a generalization of the so-called "Cameo principle," which has recently been introduced by two of the authors [12]. There, the affinity α was the same for all nodes and there was consequently no correlation between pairs of nodes. In this case it was rigorously proven that the network has indeed a degree distribution with a power-law tail and that the exponent γ is a simple function of α; more precisely,

(i) if $\phi(\omega)$ decreases as a power law with exponent β when $\omega \to \infty$, $\gamma = 1 + 1/\alpha - 1/\alpha\beta$;

(ii) if $\phi(\omega)$ vanishes faster than any power law when $\omega \to \infty$, $\gamma = 1 + 1/\alpha$.

We remark that the only feature of the distribution $\phi(\omega)$ that plays a role for the exponent of the resulting degree distribution is the way $\phi(\omega)$ vanishes at infinity; the behavior for low and intermediate values of ω can be arbitrarily chosen. This gives the result a wide degree of generality, which might explain why power-law degree distributions occur so frequently in real-world networks. Besides, the exponent γ does not depend on the value of the out degree m.

We also remark that the Cameo principle extends the concept of the random graph introduced by Erdös and Rényi,

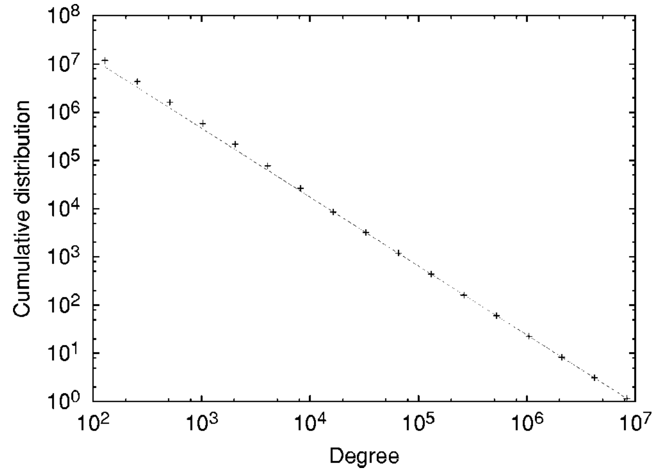

FIG. 1. Cumulative degree distribution for a network where $\phi(w) = e^{-\omega}$ and $\psi(\alpha)$ is constant in $[0, 0.7]$. In the double-logarithmic scale of the plot a power-law tail would appear as a straight line, as in the figure.

which is obtained in the limit when $\alpha \to 0$ for all nodes of the network. In a sense, our approach adds to the random graph's realistic features, which could make it suitable for applications [13]. The presence of the random variable ω in the standard Cameo principle completely changes the degree distribution of the graph, from Poissonian (Erdös and Rényi) to scale-free (Cameo). In the same way, the presence of the other random variable α, whose stochastic character is necessary for any realistic application of the model, could also deeply alter fundamental features of the network.

We studied our model numerically, by means of Monte Carlo simulations. In order to build the network we pick up a node A and build m edges with the other nodes of the network, with probability given by (1). The procedure is then repeated for all other nodes of the network. We remark that our construction process is static, i.e., all nodes of the network are there from the beginning of the process and neither nodes are added nor destroyed. However, the principle can as well be implemented in a dynamical way, with new nodes that are progressively added to the network [12].

We fixed the out degree m to the same value for all nodes, as it is done in the model of Barabási and Albert [5] (we set $m = 100$). The number N of nodes was always 10^6.

We have always used a simple exponential for $\phi(\omega)$. Figure 1 shows the cumulative degree distribution of the network constructed with a uniform affinity distribution $\psi(\alpha) = \text{const.}$, for α in the range $[0, 0.7]$. The cumulative distribution is the integral of the normal distribution. So, for a value k of the degree we counted how many nodes have a degree larger than k. The summation reduces fluctuations considerably and the analysis gets easier. If the degree distribution is a power law with exponent γ, its integral will be again a power law but with exponent $\gamma - 1$.

In Fig. 1 we see that indeed the cumulative distribution ends as a straight line in a double-logarithmic plot, so it has a power-law tail. We performed many trials, by varying the range of the uniform distribution $\psi(\alpha)$, and by using other kinds of distribution functions for α, like Gaussians, expo-

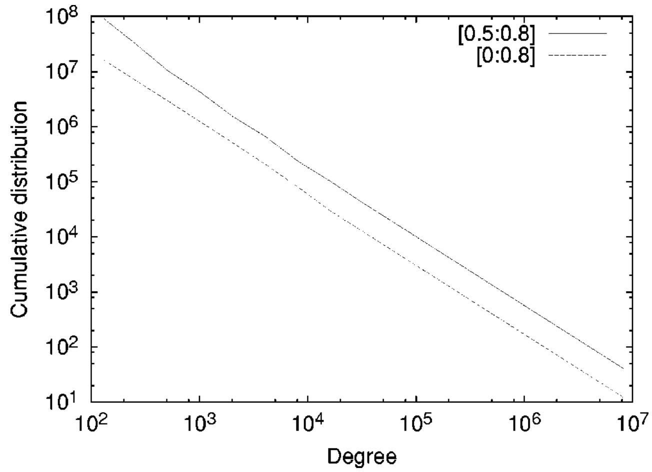

FIG. 2. Cumulative degree distribution for two networks characterized by uniform distributions $\psi(\alpha)$. The upper limit of the α range is the same in both cases. The power-law tails have the same slope.

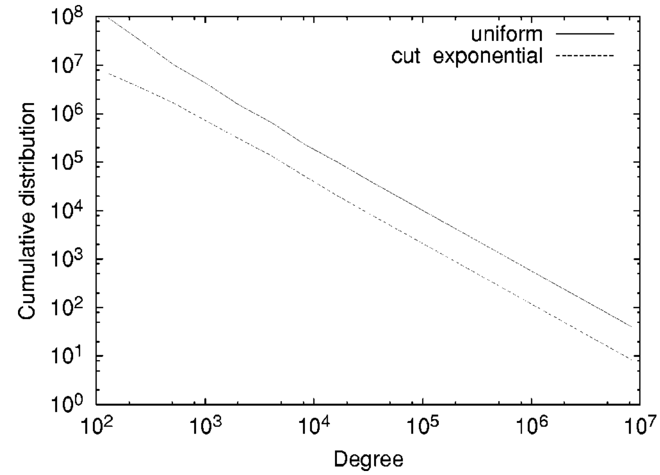

FIG. 3. Cumulative degree distribution for two networks characterized, respectively, by a uniform distribution of α in $[0.5:0.8]$ (upper curve of Fig. 2) and by an exponential distribution $\phi(\alpha) = C\exp(-\alpha)$, defined in $[0:0.8]$ (C is a normalization constant). The power-law tails have the same slope.

nentials, and power laws. We found that the result holds in all cases we considered. Beware that the use of Gaussians and exponentials does not mean that α can go to infinity, as it must be confined within the interval $[0:1]$; we just mean to consider distributions that fall like Gaussians and exponentials within this interval (they actually are "truncated" Gaussians and exponentials).

Another striking feature of our findings is shown in Fig. 2. Here we plot the cumulative degree distributions for two networks, where $\psi(\alpha)$ is uniform, and we chose the affinity ranges such that they share the same upper limit α_{max} ([0, 0.8] and [0.5, 0.8], respectively, so $\alpha_{max}=0.8$). We see that the tails of the two curves have the same slope, which suggests that γ only depends on α_{max}. We repeated this experiment several times for different ranges and taking as well truncated exponential and truncated Gaussian distributions for α. In Fig. 3 we compare the upper curve of Fig. 2, which corresponds to a uniform affinity distribution in $[0.5:0.8]$, with a simple exponential distribution in the range $[0:0.8]$. The upper limits of the two ranges coincide, and the plot shows clearly that the tails of the two degree distributions have the same slope. It would be interesting to check what happens if the affinity distribution vanishes continuously at the upper limit of the interval where it is defined. Unfortunately, the results of some tests we have performed show that to get reliable results one must go to much larger N, as in this case only a few nodes carry values of α close to α_{max}. The distribution we had used was

$$\psi(\alpha) = 2, \quad 0 \leqslant \alpha \leqslant 1/3;$$

$$\psi(\alpha) = 4 - 6\alpha, \quad 1/3 \leqslant \alpha \leqslant 2/3;$$

$$\psi(\alpha) = 0, \quad 2/3 \leqslant \alpha \leqslant 1.$$

We found that the slope of the resulting degree distribution is not steady if N is increased from 100 000 to 200 000 and finally to 1 000 000, which suggests that much larger sizes must be taken to safely extrapolate the result in the limit $N \to \infty$. So, we were not able to determine what happens when the mass of the affinity distribution is zero at the upper extreme α_{max}.

Figure 4 shows how the exponent $\gamma-1$ of the cumulative degree distribution varies with α_{max}. The pattern of the data points follows a hyperbola a/α_{max}, with a coefficient $a = 1.29$; this is very close to what one gets for the original Cameo principle [12], where $\gamma-1=1/\alpha$. It is likely that in the limit of infinite nodes the coefficient would indeed converge to one. Since α_{max} can be chosen arbitrarily close to zero, from the ansatz a/α_{max} we deduce that, within our model, we are able to build networks with any value of γ greater than (about) 2. This is fine, as for the great majority of complex networks $\gamma \geqslant 2$ as well.

So, we have discovered that the nodes with the highest affinity α, that we call "extremists" for obvious reasons, are responsible for the exponent γ of the power-law tail of the

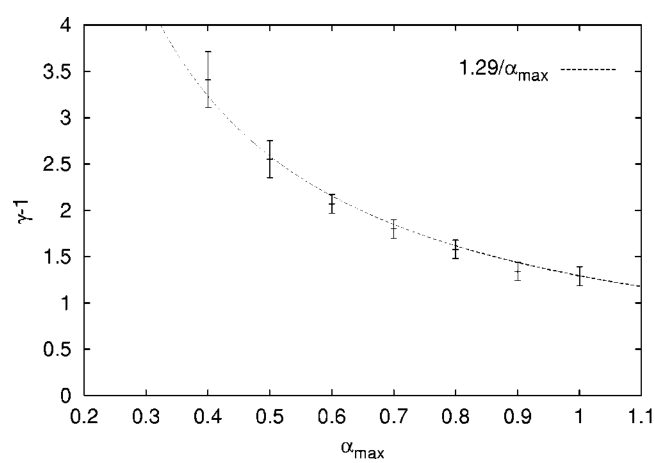

FIG. 4. Dependence of the exponent $\gamma-1$ of the cumulative degree distribution on the upper limit α_{max} of the affinity range. The data points can be fitted by the simple ansatz a/α_{max}.

degree distribution of the network. We found that this statement is true for several choices of the distribution $\psi(\alpha)$. In all our trials, however, the probability density at the upper bound of the affinity range was always different from zero. We do not know yet if and how the picture changes when the probability density vanishes at the extremum.

To translate our finding into a practical context, we can consider as an example a terrorism network. The leaders of the group (those with highest charisma/ω) are the hubs of the network, i.e., the most connected individuals, but their relative importance is determined by the most fanatic followers (those with largest α). We have then shown that there is a sort of time-dependent hierarchy among the nodes: the extremists lead the formation process, the hubs dominate the structure once the network is built.

Based on the analytical derivation of [12], we can give an argument that may justify the result of this work. Let us consider the simple case of a discrete distribution $\Psi(\alpha)$ of the form

$$\Psi(\alpha) = \sum_{i=1}^{m} \lambda_i \delta(\alpha - \alpha_i), \tag{2}$$

with $\lambda_i > 0$ and $\Sigma_{i=1}^{m}\lambda_i = 1$. So, the fraction of nodes x with $\alpha(x) = \alpha_i$ is $\lambda_i > 0$. The validity of the result lies in the fact that the global degree distribution is given by a superposition of the degree distributions associated with nodes with the same α_i. Since each of those distribution has a power-law tail [12], the overlap is dominated by the term having the fattest tail, i.e., the smallest exponent γ, which corresponds to the maximum $\alpha_{\max}$ of the α's, due to the relation $\gamma = 1 + 1/\alpha$. In this way, any function can be considered as the limit of a sum like (2), when the number of terms goes to infinity. Indeed, the same result emerges if we randomize the model of Barabási-Albert, in that a new node can choose to make m links with preexisting nodes, with m being a discrete random variable instead of a constant. In this case, like in ours, it is easy to show that the exponent of the degree distribution depends on the maximum value $m_{\max}$ that the random variable m can take.

We know that the exponent γ is a crucial feature of complex networks in many respects. For epidemic spreading, for example, there is no nonzero epidemic threshold [14] so long as $\gamma \leqslant 3$, which would have catastrophic consequences. If the network is in evolution, to control the extremists would mean to be able to exert an influence on the future topology of the network, which can be crucial in many circumstances.

From a practical point of view, it is not obvious how to model things like attractiveness (or fitness), which usually are out of the domain of quantitative scientific investigations. However, our result on the leading role of the extremists is quite robust, as it does not depend on the specific function $\psi(\alpha)$ that one decides to adopt.

In conclusion, we have introduced a simple criterion for the nodes of a complex network to choose each other as terminals of mutual connections: each node has a property ω which attracts the other nodes to an extent that depends on another individual parameter α. Networks built in this way are always characterized by a degree distribution with a power-law tail. We remark that this generalization of the result presented in [12] is much more fundamental than the original "Cameo principle"; it is the only possible implementation of the principle to real systems, as in real populations people have individual attitudes, and there is no *a priori* reason why it should work. The regular appearance of power-law degree distributions in spite of our freedom to choose the two functions $\Psi(\omega)$ and $\phi(\alpha)$ suggests that our simple procedure may have to do with the general principle responsible of the evolution of real networks. Moreover, we hit another nontrivial result which may have far-reaching implications, i.e., the fact that the exponent of the power law seems to be determined uniquely by those nodes that are most sensible to the property ω, except perhaps when the relative fraction of these nodes vanishes. Acting on such nodes could be an effective way to control the structure of evolving networks.

As we said in our introduction, the aim of this work was to devise a simple mechanism of network growth which could take into account general reasonable features of social systems. We did not have any concrete network in mind, as we hoped to define a principle of a possibly wide applicability [15]. The results we gathered in this first phase of our research line seem to us very promising, and in the immediate future we plan to devise concrete models, inspired by our mechanism, to look for a quantitative matching with real data.

S. F. gratefully acknowledges the financial support of the Volkswagen Foundation; T. K. acknowledges the financial support of the DFG-Forschergruppe FOR 399/2.

[1] A. -L. Barabási, *Linked: The New Science of Networks* (Penguin Books, London, 2003).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] D. J. de S. Price, J. Am. Soc. Inf. Sci. **27**, 292 (1976).
[5] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[6] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[7] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[8] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).
[9] G. Ergün and G. J. Rodgers, Physica A **303**, 261 (2002).
[10] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[11] In fact, the total probability of a node A to get joined to any of the others must be 1, and the integral of (1) over all the network is convergent only if $\alpha_A < 1$.

[12] Ph. Blanchard and T. Krüger, J. Stat. Phys. **114**, 1399 (2004).
[13] We recall that real-world networks cannot, in most cases, be described as random graphs.
[14] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[15] We emphasize that we are not proposing any concrete model for attractiveness. Our viewpoint is rather: "If one would have a model, then …".