Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 338 (2004) 119–124
PHYSICA A
www.elsevier.com/locate/physa

Social network growth with assortative mixing

Michele Catanzaro[a,*], Guido Caldarelli[b], Luciano Pietronero[b,c]

[a] Departament de Física i Enginyeria Nuclear, Univeristat Politécnica de Catalunya, Campus Nord, Módul B4, 08034 Barcelona, Spain
[b] INFM UdR ROMA1 Dipartimento di Fisica, Università di Roma La Sapienza, Piazzale A. Moro 2, 00185 Roma, Italy
[c] CNR Istituto di Acustica, O.M. Corbino, via Fosso del Cavaliere 100, 00133 Roma, Italy

Abstract

Networks representing social systems display specific features that put them apart from biological and technological ones. In particular, the number of links attached to a node is positively correlated to that of its nearest neighbours. We develop a model that reproduces this feature, starting from microscopical mechanisms of growth. The statistical properties arising from the simulations are in good agreement with those of the real-world social networks of scientists co-authoring papers in condensed matter physics. Moreover, the model highlights the determinant role of correlations in shaping the network's topology.

PACS: 89.75.−k; 89.75.Hc; 89.65.−s

Keywords: Social networks; Assortativity

1. Social networks

The research developed in the field of networks has revealed their pervasive presence in technology, biology and society [1]. Systems as different as the Internet, the neurons of the brain, food-chains and scientific literature can all be represented as *graphs* [2], i.e., geometrical objects composed by *nodes* connected by *edges*. While the statistical analysis of these networks has revealed some nearly ubiquitous features [1], social networks (where nodes are people and edges are interactions) display a specific behaviour that distinguish them from the others. Apart from the intrinsic interest in the principles governing social interactions, such systems provide information about phenomena like

* Corresponding author.
E-mail address: michele.catanzaro@upc.es (M. Catanzaro).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2004.02.033

the spreading of information and illnesses that take place on top of them [1]. Examples of social graphs that have been analized are the network of actors co-starring in movies, that of scientists co-authoring papers, that of businessmen sitting in common boards of directors, and that of individuals with sexual interactions [1]. Between the aspects that place them apart from technological and biological networks, we have focused on the emergence of special correlations in the properties of the nodes at the ends of an edge.

2. Assortativity in networks

A network is called *assortative* (*disassortative*) with regards to a certain property if one can observe a positive (negative) correlation in that property when considering adjacent nodes. In this work, we characterize the social networks, in a first approximation, with a scalar discrete property: the degree, i.e., the number of edges a node has. In other words, we ask ourselves if nodes with a certain degree tend to connect with others with similar or different degree. Indeed, it results that while the majority of technological and biological networks appear to be disassortative with respect to the degree, social networks are generally assortative [3]. Such a result gives a significant insight into the specificity of social networks. Indeed, far from being a secondary feature, degree-assortativity has important consequences, both on their topology and on the dynamics that can take place on top of them: assortative networks percolate more easily, are more resistant to attacks, and provide an ideal reservoir for epidemics [3].

From a quantitative point of view, indications of assortativity can be drawn from the *neighbour connectivity* [4], i.e., the average nearest neighbour degree of a node of degree k

$$knn(k) = \sum_{k'} k' P(k' \mid k) , \tag{1}$$

where $P(k' \mid k)$ is the conditional probability that an edge belonging to a node of degree k points to a node of degree k'. Such a function is increasing, decreasing or constant if the network is, respectively, assortative, disassortative or non-assortative. Indeed, in the first case as the degree of a node increases, the average degree of the nearest neighbours increases as well. The opposite happens in the second case, while in the third, no correlation is observed. Another measure is the *assortativity coefficient* r. While a rigorous definition can be found in Ref. [3], to the aims of this work it is enough to say that it is a quantity corresponding to the correlation coefficient of the degrees of the nodes at the extremes of an edge.

$$r \propto \langle ij \rangle - \langle i \rangle \langle j \rangle . \tag{2}$$

3. The model

The model we define in this work was thought to reproduce the assortative character of social network as a result of a few elementary mechanisms of formation of the network at a microscopic level. This model can be thought as a generalization of the Barabási–Albert preferential attachment one [5], and as the assortative version of the one presented in Ref. [6]. While in the Barabási–Albert model the only allowed microscopical mechanism was the addition of new nodes, we include as well *mixing*, i.e., the addition of new links between nodes already existing in the network. This mechanism is crucial in social networks, since the average life of a node (a human or professional life) is usually much longer than that of an edge (a social relation). We will compare the statistical properties of the network emerging from this rules with those of a real-world social graph, i.e., the co-authorship network of condensed matter physicists (*cond-mat*) [7].

The mechanism of growth is the following. At every time step:

(1) with probability p a new node is wired to an existing one; the choice of the destination node is left to Barabási–Albert preferential attachment rule ('rich gets richer'). Thus, the probability of adding a new node and connecting it to an old node i is

$$p\,\frac{k_i}{\sum_{j=1,N} k_j}\ . \tag{3}$$

(2) with probability $(1-p)$ a new edge is added (if absent) between two existing nodes. These are chosen on the basis of their degree. In other words, the probability of adding an edge between nodes 1 and 2 is a $\tilde{P}(k_1,k_2)$. This can be written as $P_1(k_1)P_2(k_2|k_1)$, being the second factor a conditioned probability. $P_1(k_1)$ is the rule for choosing the first of the two nodes, and again it is determined by the preferential attachment. The functional form of $P_2(k_2|k_1)$ can be chosen so as to favour links between similar or different degree. In this way, the probability of adding a new edge and connecting two old non-linked nodes is

$$(1-p)\frac{k_i}{\sum\limits_{j=1,N} k_j}\,P_2(k_2\,|\,k_1)\ . \tag{4}$$

In the limit of $p=1$ the model reduces to a traditional BA tree. In order to reproduce the assortative behaviour we have explored two different functional forms: an inverse dependence

$$P_2(k_2\,|\,k_1) \propto \frac{1}{|k_1-k_2|+1}$$

and an exponential dependence, which clearly has a stronger effect

$$P_2(k_2\,|\,k_1) \propto \mathrm{e}^{-|k_1-k_2|}.$$

4. Simulation of the model

We performed extensive simulations of the model for the two considered functional forms, and for a wide range of values of their parameters. The first issue to address

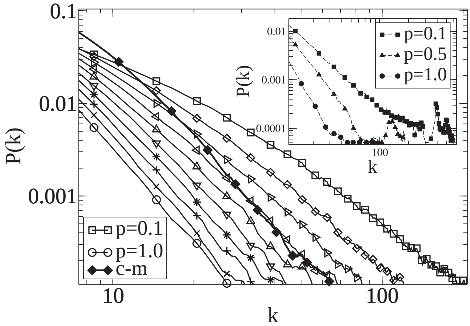

Fig. 1. Degree distribution in the inverse case. The distribution for cond-mat is reported for comparison. In the inset, degree distribution in the exponential case.

Table 1
Assortativity coefficients emerging from the simulation

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
r_{inv}	0.052	0.056	0.062	0.063	0.076	0.067	0.048	0.028	0.022	−0.01
r_{exp}	0.694	0.690	0.638	0.590	0.540	0.440	0.338	0.230	0.121	−0.01

is the emergence of assortativity. In Fig. 1 we reported the average nearest neighbours degree versus the degree for a few realizations of the model, while in Table 1 the assortativity coefficients are reported. Both results indicate that the model is able to reproduce assortativity (at a stroger level in the exponential case). Mixing plays a central role: as p is decreased (so that the rate of mixing is increased), both the slope of the curve and the value of the coefficient grow. This implies, as well, that, as far as this model is concerned, the exponent and the coefficient carry the same information. The *cond-mat* co-authorship network seems to be some way in the middle between the two functional forms, since the value of its coefficient is comparable to that of the exponential case in the limit of low mixing, while its slope is similar to that of the inverse case, in the limit of high mixing.

We report in Fig. 2 the degree distribution. In the inverse case, its shape can be fitted with a power law, indicating a scale-invariant behaviour. The slope is comparable with that of *cond-mat* for p about 0.5. Interestingly enough, in the exponential case a structure at high degrees is superimposed to the power law trend. This structure emerges as p gets smaller then 0.5, i.e., when events of mixing get more frequent than those of addition of nodes. Indeed the combination of strong assortativity, together with the “preferential attachment option” in mixing yealds a cluster of high degree nodes. Thus, a strong assortativity can brake the scale-invariant feature of the network, inducing a transition from a scale-free to a non scale-free graph.

Let us now consider the emerging networks from the point of view of the flow of information upon them. A measure for the centrality of a node for such flow is

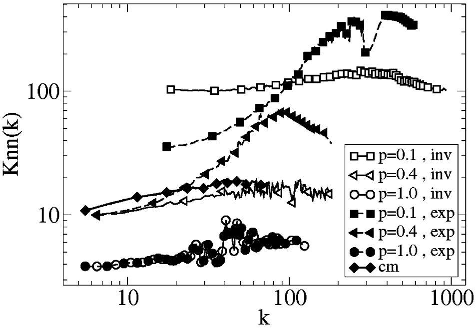

Fig. 2. Average nearest neighbour degree versus k in the inverse and exponential case, and for cond-mat.

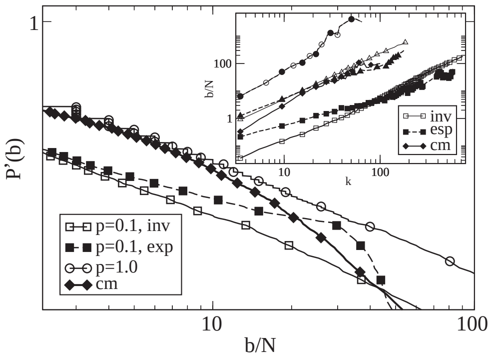

Fig. 3. Integrated site betweenness distribution in the inverse and exponential case, and for cond-mat. In the inset, b versus k in the inverse and exponential case, and for cond-mat.

the *site betweenness* (see Ref. [8] for a quantitative definition). In the inverse case (Fig. 3), the trend is a power law, in agreement with *cond-mat* data. Its slope is not a function of p, suggesting that assortativity does not affect the scaling of centrality. This scale-invariance is again broken in the exponential case, due to the positive correlations between degree and betweenness (inset of Fig. 3).

5. Conclusions

In conclusion, we succeeded in obtaining the macroscopical property of assortativity, a feature specific to social network, from the microscopical mechanism of growth called "mixing". The qualitative agreement of all trends emerging from simulation with the real-world comparison network (*cond-mat*) suggests that the model is able to catch some of the most important features of real social graphs. Moreover, when the parameters of the model are driven in the limit of strong assortativity, it comes out clearly how this feature can radically alter the properties of a network, by breaking its scale invariance.

References

[1] R. Albert, A.L. Barabási, Rev. Mod. Phys. 74 (2002) 47–97.
[2] B. Bollobás, Random Graphs, Academic Press, London, 1985.
[3] M.E.J. Newman, Phys. Rev. E 67 (2003) 026126.
[4] R. Pastor-Satorras, A. Vazquez, A. Vespignani, Phys. Rev. Lett. 87 (2001) 258701.
[5] A.-L. Barabási, R. Albert, H. Jeong, Physica A 272 (1999) 173.
[6] G. Caldarelli, P. De Los Rios, L. Pietronero, ArXiv:cond-mat/0307610.
[7] M.E.J. Newman, PNAS 98 (2001) 404.
[8] L.C. Freeman, Sociometry 40 (1977) 35.