

Assortative Mixing in Networks

M. E. J. Newman

*Department of Physics, University of Michigan, Ann Arbor, MI 48109-1120
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*

(Received 20 May 2002; published 28 October 2002)

A network is said to show assortative mixing if the nodes in the network that have many connections tend to be connected to other nodes with many connections. Here we measure mixing patterns in a variety of networks and find that social networks are mostly assortatively mixed, but that technological and biological networks tend to be disassortative. We propose a model of an assortatively mixed network, which we study both analytically and numerically. Within this model we find that networks percolate more easily if they are assortative and that they are also more robust to vertex removal.

DOI: 10.1103/PhysRevLett.89.208701

PACS numbers: 89.75.Hc, 64.60.Ak, 87.23.Ge, 89.20.Hh

Many systems take the form of networks—sets of vertices joined together by edges—including social networks, computer networks, and biological networks [1,2]. A variety of models of networks have been proposed and studied in the physics literature, many of which have been successful at reproducing features of networks in the real world [3–5]. However, there is an important element missing from these models: many networks show “assortative mixing” on their degrees, i.e., a preference for high-degree vertices to attach to other high-degree vertices, while others show disassortative mixing—high-degree vertices attach to low-degree ones. (The degree of a vertex is the number of other vertices to which it is attached.) In this paper we first demonstrate the presence of assortative mixing in a wide variety of networks by direct measurement, and then argue, using exactly solvable models and numerical simulations, that assortative mixing can have a substantial effect on the behavior of networked systems. Models that do not take it into account will necessarily fail to reproduce correctly many of the behaviors of real-world networked systems.

Consider then a network, represented in the simplest case by an undirected graph of N vertices and M edges, with degree distribution p_k . That is, p_k is the probability that a randomly chosen vertex on the graph will have degree k . Now consider a vertex reached by following a randomly chosen edge on the graph. The degree of this vertex is not distributed according to p_k . Instead, it is biased in favor of vertices of high degree, since more edges end at a high-degree vertex than at a low-degree one. This means that the degree distribution for the vertex at the end of a randomly chosen edge is proportional $k p_k$, rather than just p_k . In this paper, we will usually be interested not in the total degree of such a vertex, but in the *remaining degree*—the number of edges leaving the vertex other than the one we arrived along. This number is one less than the total degree and hence is distributed in proportion to $(k+1)p_{k+1}$. The correctly normalized distribution q_k of the remaining degree is then

$$q_k = \frac{(k+1)p_{k+1}}{\sum_j j p_j}. \quad (1)$$

Following Callaway *et al.* [6], we now define the quantity e_{jk} to be the joint probability distribution of the remaining degrees of the two vertices at either end of a randomly chosen edge. This quantity is symmetric in its indices on an undirected graph $e_{jk} = e_{kj}$, and obeys the sum rules

$$\sum_{jk} e_{jk} = 1, \quad \sum_j e_{jk} = q_k. \quad (2)$$

In a network with no assortative (or disassortative) mixing e_{jk} takes the value $q_j q_k$. If there is assortative mixing, e_{jk} will differ from this value and the amount of assortative mixing can be quantified by the connected degree-degree correlation function $\langle jk \rangle - \langle j \rangle \langle k \rangle = \sum_{jk} jk(e_{jk} - q_j q_k)$, where $\langle \cdots \rangle$ indicates an average over edges [6]. This correlation function is zero for no assortative mixing and positive or negative for assortative or disassortative mixing, respectively. For the purposes of comparing different networks, it is convenient to normalize it by dividing by its maximal value, which it achieves on a perfectly assortative network, i.e., one with $e_{jk} = q_k \delta_{jk}$. This value is equal to the variance $\sigma_q^2 = \sum_k k^2 q_k - [\sum_k k q_k]^2$ of the distribution q_k , and hence the normalized correlation function is

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k), \quad (3)$$

which is simply the Pearson correlation coefficient of the degrees at either ends of an edge and lies in the range $-1 \leq r \leq 1$. For the practical purpose of evaluating r on an observed network, we can rewrite (3) as

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}, \quad (4)$$

where j_i, k_i are the degrees of the vertices at the ends of the i th edge, with $i = 1, \dots, M$.

In Table I we show values of r for a variety of real-world networks. As the table shows, of the social networks studied (the top five entries in the table) all have significant assortative mixing. By contrast, the technological and biological networks (the middle six entries in Table I) are all disassortative.

We have also calculated r analytically for three models of networks: (1) the random graph of Erdős and Rényi [14], in which edges are placed at random between a fixed set of vertices; (2) the grown graph model of Callaway *et al.* [6], in which both edges and vertices are added at random at constant but possibly different rates, the ratio of the rates being denoted δ ; (3) the grown graph model of Barabási and Albert [4], in which both edges and vertices are added, and one end of each edge is added with linear preferential attachment, meaning that the probability of its attaching to a vertex is proportional to the current degree of that vertex.

For the random graph, since edges are placed at random without regard to vertex degree it follows trivially that $r = 0$ in the limit of large graph size. The model of Callaway *et al.*, however, although apparently similar in construction, shows markedly different behavior: using

TABLE I. Size n and assortativity coefficient r for various networks: collaboration networks of (a) scientists in physics and biology [7], (b) mathematicians [8], (c) film actors [3], and (d) businesspeople [9]; (e) connections between autonomous systems on the Internet [10]; (f) undirected hyperlinks between Web pages in a single domain [4]; (g) protein-protein interaction network in yeast [11]; (h) undirected (and unweighted) synaptic connections in the neural network of the nematode *C. Elegans* [3]; undirected trophic relations in the food webs of (i) Ythan Estuary, Scotland [12], and (j) Little Rock Lake, Wisconsin [13]. The last three lines give analytic results for model networks in the limit of large network size: (u) the random graph of Erdős and Rényi [14]; (v) the grown graph model of Callaway *et al.* [6]; (w) the model of Barabási and Albert [4].

Network	n	r
Physics coauthorship (a)	52 909	0.363
Biology coauthorship (a)	1 520 251	0.127
Mathematics coauthorship (b)	253 339	0.120
Film actor collaborations (c)	449 913	0.208
Company directors (d)	7 673	0.276
Internet (e)	10 697	-0.189
World-Wide Web (f)	269 504	-0.065
Protein interactions (g)	2 115	-0.156
Neural network (h)	307	-0.163
Marine food web (i)	134	-0.247
Freshwater food web (j)	92	-0.276
Random graph (u)		0
Callaway <i>et al.</i> (v)		$\delta/(1 + 2\delta)$
Barabási and Albert (w)		0

results from Ref. [6], we find that $r = \delta/(1 + 2\delta)$ for this model. Thus the model shows significant assortative mixing, with a maximum value of $r = \frac{1}{2}$ in the limit of large δ . This agrees with intuition: in the grown graph the older vertices have higher degree and also tend to have higher probability of being connected to one another, simply by virtue of being around longer. Thus one would expect positive assortative mixing.

The model of Barabási and Albert [4] provides an interesting counterexample to this intuition. Although this is a grown graph model, in which again older vertices have higher degree [15], it shows no assortative mixing at all: making use of results due to Krapivsky and Redner we can show that $r \rightarrow 0$ as $(\log^2 N)/N$ as N becomes large. The model of Barabási and Albert has been used as a model of the structure of the Internet and the World-Wide Web. Since these networks show significant disassortative mixing, however (Table I), it is clear that the model is incomplete. It is an open question what type of network evolution processes could explain the values of r observed in the real-world networks.

Turning now to theoretical developments, we propose a simple model of an assortatively mixed network, which is exactly solvable for many of its properties in the limit of large graph size. Consider the ensemble of graphs in which the distribution e_{jk} takes a specified value. This defines a random graph model similar in concept to the random graphs with specified degree sequence [5,16], except for the added element of assortative mixing.

Consider a typical member of this ensemble in the limit of large graph size, and consider a randomly chosen edge in that graph, one end of which is attached to a vertex of degree j . We ask what the probability distribution is of the number of other vertices reachable by following that edge. Let this probability distribution be generated by a generating function $G_j(x)$, which depends in general on the degree j of the starting vertex. By arguments similar to those of Ref. [5], we can show that $G_j(x)$ must satisfy a self-consistency condition of the form

$$G_j(x) = x \frac{\sum_k e_{jk} [G_k(x)]^k}{\sum_k e_{jk}}, \quad (5)$$

while the number of vertices reachable from a randomly chosen vertex is generated by

$$H(x) = xp_0 + x \sum_{k=1}^{\infty} p_k [G_{k-1}(x)]^k. \quad (6)$$

The average size of the component to which such a vertex belongs is given by the derivative of H : $\langle s \rangle = H'(1) = 1 + \sum_k k p_k G'_{k-1}(1)$. Differentiating Eq. (5) we then get

$$\langle s \rangle = 1 - z \mathbf{q} \cdot \mathbf{A}^{-1} \cdot \mathbf{q}, \quad (7)$$

where z is the mean degree, \mathbf{q} is the vector whose elements are the q_k , and \mathbf{A} is the asymmetric matrix with elements $A_{jk} = k e_{jk} - q_k \delta_{jk}$.

Equation (7) diverges at the point at which the determinant of \mathbf{A} is zero. This point marks the phase transition at which a giant component forms in our graph. By considering the behavior of Eq. (7) close to the transition, where $\langle s \rangle$ must be large and positive in the absence of a giant component, we deduce that a giant component exists in the network when $\det \mathbf{A} > 0$. This is the appropriate generalization for a network with assortative mixing of the criterion of Molloy and Reed [16] for the existence of a giant component.

To calculate the size S of the giant component, we define u_k to be the probability that an edge connected to a vertex of remaining degree k leads to another vertex that does not belong to the giant component. Then

$$S = 1 - p_0 - \sum_{k=1}^{\infty} p_k u_{k-1}, \quad u_j = \frac{\sum_k e_{jk} u_k^k}{\sum_k e_{jk}}. \quad (8)$$

To test these results and to help form a more complete picture of the properties of assortatively mixed networks, we have also performed computer simulations, generating networks with given values of e_{jk} and measuring their properties directly. Generating such networks is not entirely trivial. One cannot simply draw a set of degree pairs (j_i, k_i) for edges i from the distribution e_{jk} , since such a set would almost certainly fail to satisfy the basic topological requirement that the number of edges ending at vertices of degree k must be a multiple of k . Instead, therefore we propose the following Monte Carlo algorithm for generating graphs.

First, we generate a random graph with the desired degree distribution according to the prescription given in Ref. [16]. Then we apply a Metropolis dynamics to the graph in which on each step we choose at random two edges, denoted by the vertex pairs, (v_1, w_1) and (v_2, w_2) , that they connect. We measure the remaining degrees (j_1, k_1) and (j_2, k_2) for these vertex pairs, and then replace the edges with two new ones (v_1, v_2) and (w_1, w_2) with probability $\min[1, (e_{j_1 j_2} e_{k_1 k_2}) / (e_{j_1 k_1} e_{j_2 k_2})]$. This dynamics conserves the degree sequence, is ergodic on the set of graphs having that degree sequence, and, with the choice of acceptance probability above, satisfies detailed balance for state probabilities $\prod_i e_{j_i k_i}$, and hence has the required edge distribution e_{jk} as its fixed point.

As an example, consider the symmetric binomial form

$$e_{jk} = \mathcal{N} e^{-(j+k)/\kappa} \left[\binom{j+k}{j} p^j q^k + \binom{j+k}{k} p^k q^j \right], \quad (9)$$

where $p + q = 1$, $\kappa > 0$, and $\mathcal{N} = \frac{1}{2}(1 - e^{-1/\kappa})$ is a normalizing constant. (The binomial probabilities p and q should not be confused with the quantities p_k and q_k introduced earlier.) This distribution is chosen for analytic tractability, although its behavior is also quite natural: the distribution of the sum $j + k$ of the degrees at the ends of an edge falls off as a simple exponential, while that sum is distributed between the two ends binomially,

the parameter p controlling the assortative mixing. From Eq. (3), the value of r is

$$r = \frac{8pq - 1}{2e^{1/\kappa} - 1 + 2(p - q)^2}, \quad (10)$$

which can take both positive and negative values, passing through zero when $p = p_0 = \frac{1}{2} - \frac{1}{4}\sqrt{2} = 0.1464\dots$

In Fig. 1 we show the size of the giant component for graphs of this type as a function of the degree scale parameter κ , from both our numerical simulations and the exact solution above. As the figure shows, the two are in good agreement. The three curves in the figure are for $p = 0.05$, where the graph is disassortative, $p = p_0$, where it is neutral (neither assortative nor disassortative), and $p = 0.5$, where it is assortative.

As κ becomes large we see the expected phase transition at which a giant component forms. There are two important points to notice about the figure. First, the position of the phase transition moves lower as the graph becomes more assortative. That is, the graph percolates more easily, creating a giant component, if the high-degree vertices preferentially associate with other high-degree ones. Second, notice that, by contrast, the size of the giant component for large κ is smaller in the assortatively mixed network.

These findings are intuitively reasonable. If the network mixes assortatively, then the high-degree vertices will tend to stick together in a subnetwork or core group of higher mean degree than the network as a whole. It is reasonable to suppose that percolation would occur earlier within such a subnetwork. Conversely, since percolation will be restricted to this subnetwork, it is not surprising

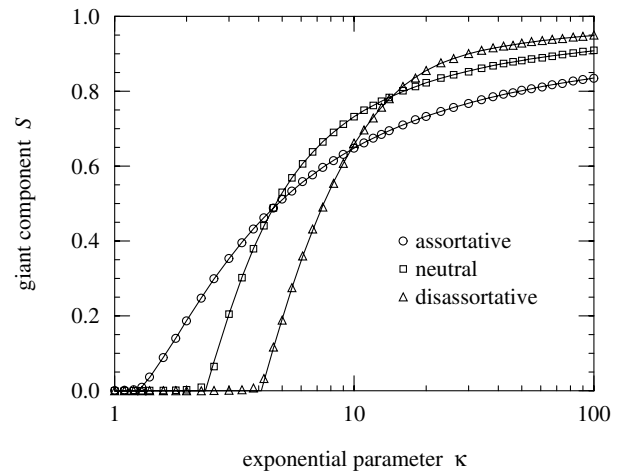


FIG. 1. Size of the giant component as a fraction of graph size for graphs with the edge distribution given in Eq. (9). The points are simulation results for graphs of $N = 100\,000$ vertices, while the solid lines are the numerical solution of Eq. (8). Each point is an average over ten graphs; the resulting statistical errors are smaller than the symbols. The values of p are 0.5 (circles), $p_0 = 0.146\dots$ (squares), and 0.05 (triangles).

that the giant component has a smaller size in this case than when the network is disassortative. These results could have implications, for example, for the spread of disease on social networks [17]—social networks being assortatively mixed in most cases, as Table I shows. The core group of an assortatively mixed network could form a “reservoir” for disease, sustaining an epidemic even in cases in which the network is not sufficiently dense on average for the disease to persist. On the other hand, one would expect the disease to be restricted to a smaller segment of the population in such cases than for diseases spreading on neutral or disassortative networks.

Assortative mixing also has implications for questions of network resilience, the subject of much discussion in the recent literature [18–20]. It has been found that the connectivity of many networks (i.e., the existence of paths between pairs of vertices) can often be destroyed by the removal of just a few of the highest degree vertices, a result that may have applications in, for example, vaccination strategies [21]. In assortatively mixed networks, however, we find numerically that removing high-degree vertices is a relatively inefficient strategy for destroying network connectivity, presumably because these vertices tend to be clustered together in the core group, so that removing them is somewhat redundant. In a disassortative network with a similarly sized giant component attacks on the highest degree vertices are much more effective, these vertices being broadly distributed over the network and presumably therefore forming links on many paths between other vertices. For networks of the type described by Eq. (9) we find that the number of high-degree vertices that need to be removed to destroy similarly sized giant components is greater by a factor of about 5 to 10 in an assortative network ($p = 0.5$) than in a disassortative one ($p = 0.05$) for the typical parameter values studied here.

These considerations paint rather a grim picture: the networks that we might want to break up, such as the social networks that spread disease, appear to be assortative, and therefore are resilient, at least against simple targeted attacks such as attacks on the highest degree vertices. And yet at the same time the networks that we would wish to protect, including technological networks such as the Internet, appear to be disassortative, and are hence particularly vulnerable.

To conclude, in this paper we have studied assortative mixing by degree in networks—the tendency for high-degree vertices to associate preferentially with other high-degree vertices. We have defined a scalar measure of assortative mixing and used it to show that many social networks have significant assortative mixing, while technological and biological networks seem to be disassortative. We have also proposed a model of an assortatively mixed network, which we have solved exactly using generating function techniques, and also simulated using a Monte Carlo graph sampling method. Within this model

we find that assortative networks percolate more easily and that they are also more robust to removal of their highest degree vertices, while disassortative networks percolate less easily and are more vulnerable. This suggests that social networks may be robust to intervention and attack while technological networks are not.

The author thanks Duncan Callaway, Michelle Girvan, and Cris Moore for helpful comments, and László Barabási, Jerry Davis, Jennifer Dunne, Jerry Grossman, Hawoong Jeong, Neo Martinez, and Duncan Watts for providing network data used in the calculations for Table I. This work was funded in part by the National Science Foundation under Grant No. DMS-0109086.

-
- [1] S. H. Strogatz, *Nature (London)* **410**, 268 (2001).
 - [2] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [3] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998).
 - [4] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
 - [5] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
 - [6] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, *Phys. Rev. E* **64**, 041902 (2001).
 - [7] M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 404 (2001).
 - [8] J. W. Grossman and P. D. F. Ion, *Congressus Numerantium* **108**, 129 (1995).
 - [9] G. F. Davis, M. Yoo, and W. E. Baker, University of Michigan Business School report, 2001.
 - [10] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies* (IEEE Computer Society, New York, 2002).
 - [11] H. Jeong, S. Mason, A.-L. Barabási, and Z. N. Oltvai, *Nature (London)* **411**, 41 (2001).
 - [12] M. Huxham, S. Beaney, and D. Raffaelli, *Oikos* **76**, 284 (1996).
 - [13] N. D. Martinez, *Ecological Monographs* **61**, 367 (1991).
 - [14] B. Bollobás, *Random Graphs* (Academic Press, New York, 2001), 2nd ed.
 - [15] L. A. Adamic and B. A. Huberman, *Science* **287**, 2115 (2000).
 - [16] M. Molloy and B. Reed, *Random Struct. Algorithms* **6**, 161 (1995).
 - [17] M. Morris, *Nature (London)* **365**, 437 (1993).
 - [18] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **406**, 378 (2000).
 - [19] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).
 - [20] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, *Phys. Rev. E* **65**, 056109 (2002).
 - [21] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **65**, 036104 (2002).