

CHANGING CORRELATIONS IN NETWORKS: ASSORTATIVITY AND DISSORTATIVITY*

R. XULVI-BRUNET AND I.M. SOKOLOV

Institut für Physik, Humboldt Universität zu Berlin
Newtonstraße 15, 12489 Berlin, Germany

(Received January 5, 2005)

Dedicated to Professor Andrzej Fuliński on the occasion of his 70th birthday

To analyze the role of correlations in networks, in particular, assortativity and dissortativity, we introduce two algorithms which respectively produce assortative and dissortative mixing to a desired degree. In both procedures this degree is governed by a single parameter p . Varying this parameter, one can change correlations in networks without modifying their degree distribution to produce new versions ranging from fully random ($p = 0$) to totally assortative or dissortative ($p = 1$), depending on the algorithm used. We discuss the properties of networks emerging when applying our algorithms to a Barabási–Albert scale-free construction. In spite of having exactly the same degree distribution, different correlated networks exhibit different geometrical and transport properties. Thus, the average path length and clustering coefficient, as well as the shell structure and percolation properties change significantly when modifying correlations.

PACS numbers: 05.50.+q, 89.75.Hc

1. Introduction

Our interest to complex networks was strongly motivated by our studies of models of infection spread, since it got clear that many effects can only be described when taking into account that an infecting agent spreads not in a homogeneous population of infectives, but over a complex network of contacts [1–3]. In many cases the exact knowledge of the network structure is necessary, in other situations one can rely on simple model assumptions, which, however, mirror the properties of a real networked system.

* Presented at the XVII Marian Smoluchowski Symposium on Statistical Physics, Zakopane, Poland, September 4–9, 2004.

Examples of systems described as complex networks are abundant in many disciplines of science and have received much attention in the last few years. Thus, technological systems such as the World Wide Web, Internet, and electrical power grids, as well as other natural and social systems like chemical reactions in the living cell, different social collaboration networks, *etc.*, have been successfully described through scale-free networks, networks with power-law degree distributions $P(k) \sim k^{-\gamma}$ [4, 5]. The degree distribution $P(k)$ is one of the principal measures used to capture the structure of a network and represents the probability that a node chosen at random is connected with exactly k other vertices of the network. The network of human sexual contacts also has a property of being scale-free [6].

It was recently pointed out that real networks exhibit a degree of correlations among their nodes [7–20]. Thus, in social networks nodes having many connections tend to be connected with other highly connected nodes [8, 10]. This characteristic is usually referred to as assortativity, or assortative mixing. On the other hand, technological and biological networks show the property that nodes having high degrees are preferably connected with nodes having low degrees, a property referred to as dissortativity [7, 11].

Correlations play an important role in the characterization of the topology of networks, and therefore, they are essential to understand spreading of information or infections, as well as their robustness against targeted or random removal of their elements [21–26]; in the case of infection such removal corresponds to immunization of a part of the population. In order to determine the exact influence of correlations several authors have proposed procedures to build correlated networks [7, 28–30]. The most general procedures are the ones proposed by Newman [7] and Boguñá and Pastor-Satorras [30], who suggested two different ways to construct general correlated networks with predetermined correlations. Following the same goal, we, however, adopt a different perspective. Instead of putting in correlations “by hand” we propose to use link-restructuring (“rewiring”) processes [31] satisfying conditions such as “nodes with similar degree connect preferably” (assortative mixing) or “nodes with low degree try to connect with highly connected nodes” (dissortativity), leaving all other properties random.

Such processes, which do not change the degree distribution of networks and do avoid the appearance of multiple and self-connections, can be viewed as ergodic Markovian chains defined on the set of the network’s configurations. Repeated application of a rewiring step leads to distributions of link configurations which converge to a stationary distribution with desired correlation properties, independently of the initial correlations of the network. In this work we thus introduce two procedures to change correlations based on the rewiring of links, which produce assortative and dissortative mixing, respectively. Both algorithms are governed by a single parameter, p , and

are capable to change the degree of assortativity (dissortativity) to a desired amount allowing us to generate networks ranging from fully uncorrelated to totally assortative (dissortative). In the present work we focus on undirected networks. Some of the results were earlier reported on in Refs. [29, 33].

2. The assortative model

2.1. Algorithm

Starting from a given network, two links of the network connecting four different nodes are randomly chosen at each step. We consider the four nodes associated with these two links, and order them with respect to their degrees. Then, with probability p , the links are rewired in such a way that one link connects the two nodes with the smaller degrees and the other connects the two nodes with the larger degrees; otherwise the links are randomly rewired (Maslov–Sneppen algorithm [16]). In the case that one or both of these new links already existed in the network, the step is discarded and a new pair of edges is selected. This restriction prevents the appearance of multiple edges connecting the same pair of nodes. A repeated application of the rewiring step leads to an assortative version of the original network. Note that the algorithm does not change the degree of the nodes involved and thus the overall degree distribution in the network. Changing the parameter p , it is possible to construct networks with different degrees of assortativity.

2.2. Correlations and assortativity

Let \mathcal{E}_{ij} be the probability that a randomly selected edge of the network connects two nodes, one with degree i and another with degree j . The probabilities \mathcal{E}_{ij} determine the correlations of the network. We say that a network is uncorrelated when

$$\mathcal{E}_{ij} = \mathcal{E}_{ij}^r = (2 - \delta_{ij}) \frac{iP(i)jP(j)}{\langle i \rangle \langle j \rangle}, \quad (1)$$

i.e., when the probability that a link is connected to a node with a certain degree is independent from the degree of the attached node. Here $\langle i \rangle = \langle j \rangle$ denotes the first moment of the degree distribution, assumed to be finite.

Assortativity means nodes with similar degrees tend to be connected with a larger probability than in the uncorrelated case, *i.e.*, $\mathcal{E}_{ii} > \mathcal{E}_{ii}^r \quad \forall i$. The degree of assortativity of a network can thus be characterized by the quantity [7]

$$\mathcal{A} = \frac{\sum_i \mathcal{E}_{ii} - \sum_i \mathcal{E}_{ii}^r}{1 - \sum_i \mathcal{E}_{ii}^r}, \quad (2)$$

which takes the value 0 when the network is uncorrelated and the value 1 when the network is totally assortative. (Note that in a finite network the constraint that no pair of vertices is connected by more than one edge bounds \mathcal{A} from above by the values lower than 1 [27].)

Now, starting from the algorithm generator, we can obtain a theoretical expression for \mathcal{E}_{ij} as a function of p . Let E_{ij} be the number of links in the network connecting two nodes, one with degree i and another with degree j , so that $\mathcal{E}_{ij} = E_{ij}/L$, where L is the total number of links of the network. (Since undirected networks satisfy $E_{ij} = E_{ji}$, the restriction $i \leq j$ can be imposed without loss of generality.) We now define the variable

$$F_{ln} = \sum_{r=l}^n \sum_{s=r}^n E_{rs}, \quad r \leq s, \quad l \leq n. \quad (3)$$

Every time the rewiring procedure is applied, the values of F_{ln} either increase or decrease by unity, or do not change. We can then calculate the probabilities of change, *i.e.*, that $F_{ln} \rightarrow F_{ln} + 1$ or $F_{ln} \rightarrow F_{ln} - 1$. The effect of multiple edges can be disregarded since they are rare in the thermodynamical limit. Taking all corresponding possibilities into account, we obtain the following expressions for the probabilities of change:

$$(X_{ln} - f_{ln})^2 + p(X_{ln} - f_{1n} + f_{1,l-1})^2, \quad \text{for } F_{ln} \rightarrow F_{ln} + 1,$$

$$f_{ln} [(1-p)(1-2X_{ln}) + p(X_{1,l-1} - f_{1,l-1} - f_{1n}) + f_{ln}], \quad \text{for } F_{ln} \rightarrow F_{ln} - 1.$$

Here $f_{ln} = F_{ln}/L$, and X_{ln} is given by

$$X_{ln} = \frac{1}{\langle k \rangle} \sum_{k=l}^n kP(k), \quad l \leq n.$$

Using these expressions, we can calculate the expected value of f_{ln} . The process of repeatedly applying our algorithm corresponds to an ergodic Markov chain, and the stationary solution is given by the condition

$$\begin{aligned} & (X_{ln} - f_{ln})^2 + p(X_{ln} - f_{1n} + f_{1,l-1})^2 \\ &= f_{ln} [(1-p)(1-2X_{ln}) + p(X_{1,l-1} - f_{1,l-1} - f_{1n}) + f_{ln}], \end{aligned} \quad (4)$$

for all $l > 1$. For $l = 1$ this condition reduces to

$$(1+p)(X_{1n} - f_{1n})^2 = (1-p)f_{1n} [1 - 2X_{1n} + f_{1n}]. \quad (5)$$

Using Eq. (4) and Eq. (5) we calculate f_{ln} . The solution reads

$$f_{ln} = \frac{X_{ln}^2 + (B_n - B_{n-1})^2}{(1 - p)/2 + pX_{ln} + B_n + B_{n-1}}, \quad l \leq n,$$

with

$$B_n = \sqrt{\left[pX_{1n} + \frac{1 - p}{4} \right]^2 - pX_{1n}^2 \left(\frac{1 + p}{2} \right)}.$$

Applying the definition, Eq. (3), we obtain the correlations

$$\mathcal{E}_{ij} = f_{ij} - f_{i,j-1} - f_{i+1,j} + f_{i+1,j-1}. \quad (6)$$

We note that Eq. (6) reduces to the corresponding uncorrelated case \mathcal{E}_{ij}^r when $p = 0$, and reduces to

$$\mathcal{E}_{ij} = \delta_{ij} \frac{iP(i)}{\langle i \rangle}, \quad (7)$$

for the case $p = 1$.

2.3. Properties of an assortative network

Let us start this section with drawing a small network to show how our algorithm works. The initial network is a Barabási–Albert scale-free construction [32] with only $N = 200$ nodes and $L = 400$ links, see Fig. 1(a). To obtain other networks with exactly the same degree distribution but different degree of assortativity we apply the algorithm discussed. Fig. 1 shows the changes in the network with varying parameter p . In the figure we have placed the nodes in such a way that nodes of degree 2 are shown in the left part of each panel, all nodes of degree 3 lie to the right of any node of degree 2, all nodes of degree 4 lie to the right of any node of degree 3, *etc.* The nodes of the same degree are randomly spread within the corresponding area of the figure to better show the links.

The network corresponding to the maximal assortativity is shown in Fig. 1(d). In this network almost all nodes with the same degree are linked only between themselves. Panel (d) shows that all nodes with degree $k = 2$ form separated clusters (a more careful analysis unveils that there are three “pearl necklace” clusters with $N = 23, 30$, and 48 nodes). All nodes with $k = 3$ are linked between themselves except for one, which is linked to a node of connectivity $k = 4$. Note that since there are $N_3 = 41$ nodes with $k = 3$ in our network their links cannot be redistributed within the set. If this would be possible, the overall number of links would be $41 \times 3/2 = 61.5$, since each node bears 3 links and each of these links is counted twice in the

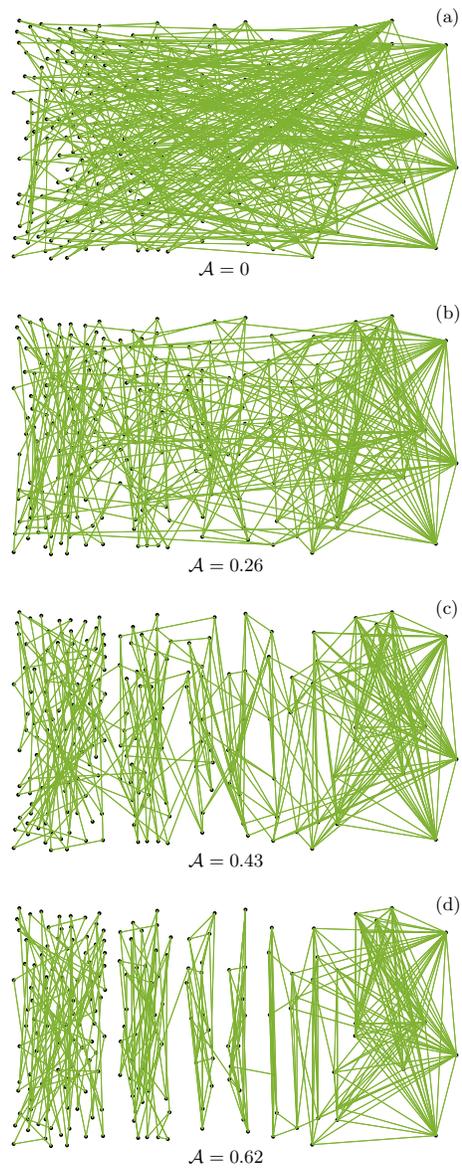


Fig. 1. Scale-free networks for different degrees of assortativity (see text for details). The nodes of the same degree are grouped together; the degree is nondecreasing from left to right. The panels show: (a) $\mathcal{A} = 0$ (uncorrelated network), (b) $\mathcal{A} = 0.26$, (c) $\mathcal{A} = 0.43$, (d) $\mathcal{A} = 0.62$ (maximal assortativity).

set. All nodes with degree $k = 4$ form a single cluster, with two outgoing links, one to the cluster of nodes with $k = 3$, and one to a cluster of nodes of connectivity $k = 5$. In fact, the network is not a set of isolated clusters of nodes with the same connectivity only due to the restrictions imposed by the given degree distribution. These restrictions are also responsible for the fact that $\mathcal{A} < 1$ (for our network the maximal assortativity is $\mathcal{A}_{\max} = 0.62$).

In the present work we apply the algorithm to the Barabási–Albert construction with the number of links being twice the number of nodes $L = 2N$, just like in our example on Fig. 1. We measure \mathcal{E}_{ij} as functions of p , and use them to calculate the corresponding values of \mathcal{A} . All simulation results are averaged over ten independent realizations of the algorithm as applied to the same original network.

Fig. 2 shows the relation between the parameter p and the coefficient of assortativity \mathcal{A} . The lower curves correspond to the measured assortativity for two networks of different size ($N = 10^4$ and $N = 10^5$); the upper curve corresponds to our theoretical prediction (pertinent to an infinite network). All curves coincide for small values of \mathcal{A} . However, whereas the theoretical curve reaches the value $\mathcal{A} = 1$ when $p \rightarrow 1$, the measured assortativity increases until a maximal value smaller than one. Thus, the central curve in the figure, corresponding to the Barabási–Albert network with 10^5 nodes, shows a maximal value $\mathcal{A} = 0.917$ when $p \rightarrow 1$, and for the network with $N = 10^4$ this value is even smaller, reaching only $\mathcal{A} = 0.864$. This was expected [27], and is due to the finite-sized effects mentioned above.

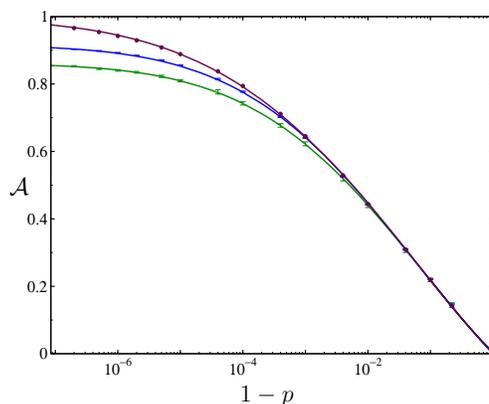


Fig. 2. Assortativity \mathcal{A} as function of the parameter p . The two lower curves correspond to the measured assortativity \mathcal{A} of our simulations for two different Barabási–Albert networks, one with $N = 10^4$ nodes (the lowest curve) and another with 10^5 nodes. The upper curve corresponds to the theory (thermodynamical limit). We note that all curves coincide for small \mathcal{A} , whereas for large values of \mathcal{A} finite-size corrections get important, leading to a value of $\mathcal{A} < 1$ for $p \rightarrow 1$.

To assess the goodness of the Eq. (6), we compare the simulations with the theoretical values of \mathcal{E}_{kk} , given by Eq. (6), in the Fig. 3. Here the results correspond to a Barabási–Albert network with 10^6 nodes. The points are the outcomes of the simulations and the curves are the corresponding theoretical results obtained based on the actual degree distribution of our particular discussed network. We note that the agreement is really excellent.

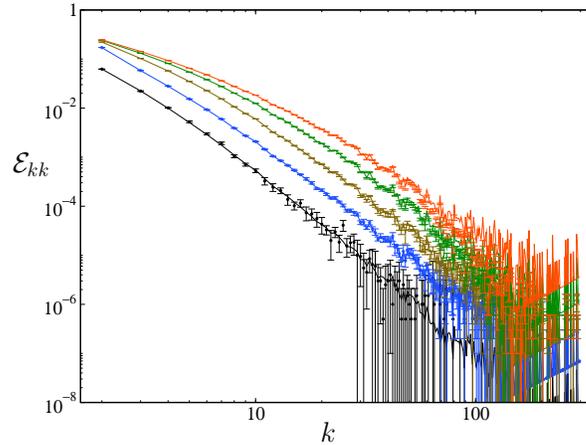


Fig. 3. \mathcal{E}_{kk} as a function of k for different values of \mathcal{A} . From bottom to top: $\mathcal{A} = 0$, 0.221, 0.443, 0.640, and 0.777. The points are the results of the simulations while the curves are calculated using the theory.

Average path length — The average path length of a network is the average distance between every pair of vertices of the network, being defined as the number of edges along the shortest path connecting them. Uncorrelated scale-free networks show a very small path length, typically growing logarithmically with the network’s size (a small-world behavior).

Our simulations show that the average path length l grows rapidly when the assortativity of the network increases so that it becomes some hundreds times larger than in an uncorrelated network when the coefficient of assortativity tends to its maximal value. In Fig. 4 we plot l as function of \mathcal{A} for networks with $N = 10^4$ and $N = 10^5$ nodes. We observe that l increases following the expression $l \propto (\mathcal{K} - \mathcal{A})^{-\gamma}$, with $\mathcal{K} = 0.864$ ($N = 10^4$) and $\mathcal{K} = 0.917$ ($N = 10^5$), corresponding to the maximal values of \mathcal{A} attainable in the networks, and with $\gamma = 1.2$. The inset of the figure show this behavior of l for the network with $N = 10^5$ nodes.

Although assortative networks present large mean path lengths, they are still small worlds, *i.e.*, they are exhibiting the logarithmic dependence of l on the network’s size N . Fig. 5 shows this behavior for three different values of \mathcal{A} . The error bars result from averaging over ten realizations of

the algorithm. This small-world behavior is preserved for all tested values of $\mathcal{A} \leq 0.6$ (this maximal value of \mathcal{A} is considerably larger than the values found in real assortative networks, where \mathcal{A} range between 0 and 0.4). Thus, assortative networks are the “large” small worlds.

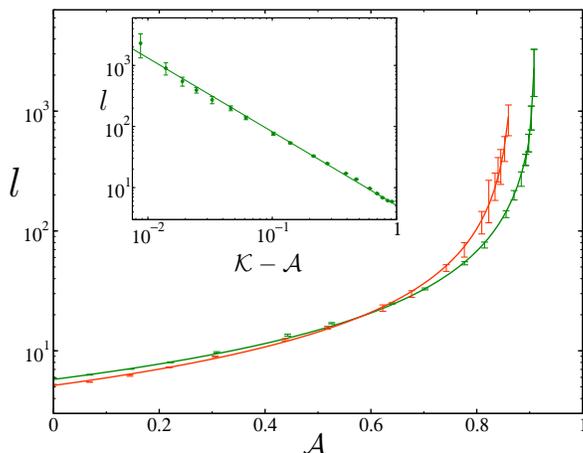


Fig. 4. Average path length l vs coefficient of assortativity for two Barabási–Albert networks, one with $N = 10^5$ nodes (the curve that reach the largest value of \mathcal{A}) and another with $N = 10^4$. We note that l grows rapidly when \mathcal{A} increases. The average path length is plotted on double logarithmic scales as function of $\mathcal{K} - \mathcal{A}$ for the larger network ($N = 10^5$) in the inset. Here is $\mathcal{K} = 0.917$. The slope of the straight line is -1.2 .

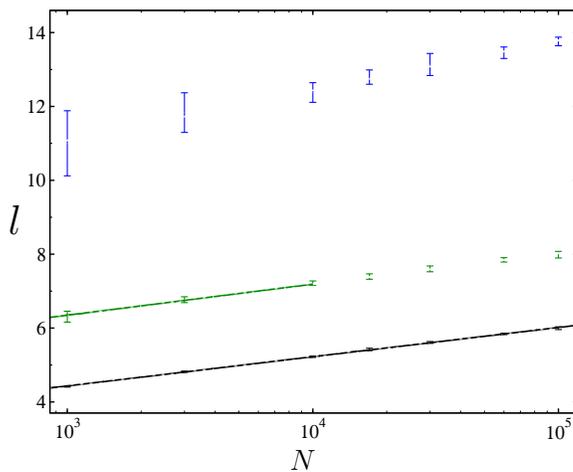


Fig. 5. The average path length l is plotted as function of N for three values of the coefficient of assortativity \mathcal{A} . From bottom to top: $\mathcal{A} = 0, 0.221$ and 0.443 . Note the logarithmic scale.

Natural networks, like different co-authorship networks (physics, biology, mathematics, *etc.*), the film actor collaboration network, *etc.* (all of them assortative networks) seem to show somewhat smaller average path lengths than the ones found here [4, 7]. We attribute this finding to the fact that the mean degree of such networks is 2 to 4 times larger than in our case ($\langle k \rangle = 4$). Therefore, one has to be cautious about comparing absolute numerical values.

Clustering coefficient — Clustering coefficients of a network are a measure of the number of loops (closed paths) of length three. The notion has its roots in sociology, where it was often used to analyze the groups of acquaintances in which every member knows every other one. To discuss the concept of clustering, let us focus first on a vertex, having k edges connected to k other nodes termed as nearest neighbors. If these nearest neighbors of the selected node were forming a fully connected cluster of vertices, there would be $k(k-1)/2$ edges between them. The ratio between the number of edges that really exist between these k vertices and the maximal number $k(k-1)/2$ gives the value of the clustering coefficient of the selected node. The clustering coefficient of the whole network C is then defined as the average of the clustering coefficients of all vertices. One can also speak about the clustering coefficient of nodes with a given degree k , referring to the average of the clustering coefficients of only this type of nodes. We shall denote this degree-dependent clustering coefficient by $\bar{C}(k)$, to distinguish it from C .

Fig. 6 shows the variation of both clustering coefficients with the assortativity of the network. Fig. 6(a) corresponds to a Barabási–Albert network with $N = 10^5$ nodes while Fig. 6(b) corresponds to a one with $N = 10^6$ nodes. We see that the clustering coefficient C increases with the assortativity (insets of the figure). However, typical values of clustering coefficients found in our simulations are still much smaller than the ones observed in real networks ($C \geq 0.1$) [4]. The latter ones might, however, have a much more intricate structure, partly governed by the metrics of the underlying space, as in the models discussed in [34].

The variation of $\bar{C}(k)$ shows more interesting features. Our simulations indicate that for small k the values of $\bar{C}(k)$ grow with the degree of nodes k producing a peak whose height increases with the assortativity of the network. The peak (probably a finite size effect) moves to larger k when the size of the network increases. Thus, our simulations show a peak around $k = 90$ for the network with 10^5 nodes and one around $k = 185$ for the network with 10^6 nodes. Assortative networks show a strong tendency of clustering (for relatively large values of k) compared to uncorrelated networks, where $\bar{C}(k)$ does not depend on k [18]. We also observe that $\bar{C}(k=2) = 0$ when $\mathcal{A} \simeq 1$ ($k = 2$ correspond to the minimal degree of the vertices). This is

not surprising since in a strongly assortative network almost all nodes with $k = 2$ are connected between themselves, forming one or several large loops of length larger than three. This means that all nodes having this minimal degree (in our simulations the half of the total number of vertices) do not tend to contribute to the clustering coefficient C at all.

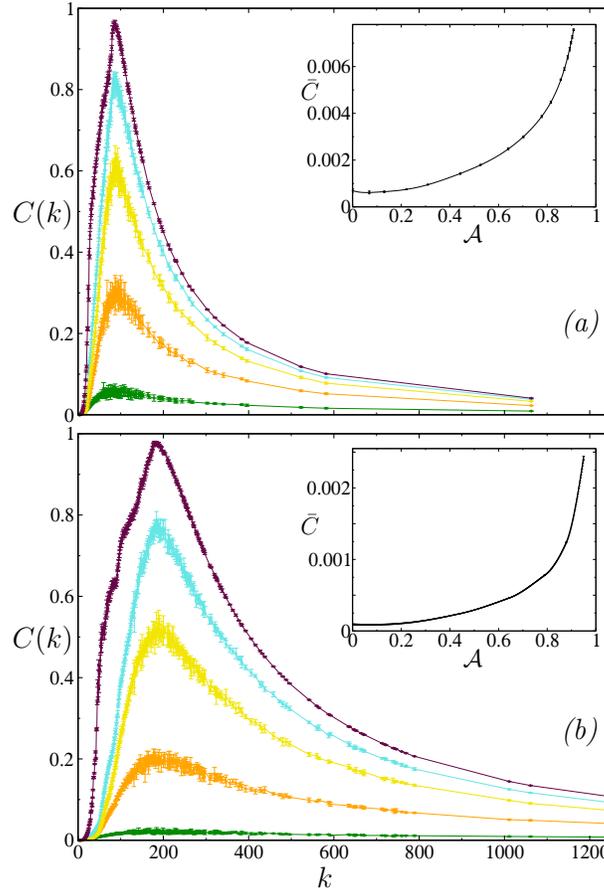


Fig. 6. $\bar{C}(k)$ as a function of the degree of nodes k . (a) $N = 10^5$, (b) $N = 10^6$. In both panels different curves correspond to different values of \mathcal{A} . From bottom to top $\mathcal{A} = 0, 0.221, 0.443$; $\mathcal{A} \simeq 0.640, 0.78$ and $\mathcal{A} = \max$ ($\mathcal{A} = 0.917$ for (a), $\mathcal{A} = 0.864$ for (b)). Insets: clustering coefficients C versus the degree of assortativity \mathcal{A} .

Tomography — Tomography is a useful tool to examine the local structure of networks. How can a computer virus spread in the Internet or a certain information in social networks from the original node to the others? This depends clearly on the distribution of vertices around the node from which the spreading starts, *i.e.*, on the structure of shells around the orig-

inal node. Thus, Cohen *et al.*, examined the shells around the node with the highest degree for uncorrelated networks. In our study we use a different perspective, and apply the procedure to each node of the network. The initial node (the root) is assigned to shell number 0. Then all links starting at this node are followed and all vertices reached are assigned to shell number 1. Then all links leaving nodes in shell 1 are followed, and all nodes reached that do not belong to previous shells are labeled as nodes of shell 2, *etc.*, until the whole network is exhausted. We then get $N_{l,r}(k)$ as the number of nodes with degree k in shell l for root r . The repetition of the whole procedure starting at all N vertices of the network gives us $P_l(k)$, the degree distribution in shell l . We define $P_l(k)$ as

$$P_l(k) = \frac{\sum_r N_{l,r}(k)}{\sum_{k,r} N_{l,r}(k)}. \quad (8)$$

We are interested in the average degree $\langle k \rangle_l = \sum_k k P_l(k)$ of nodes of the shell l . In the epidemiological context, this quantity can be interpreted as a disease multiplication factor after l steps of propagation. It describes how many neighbors a node can infect on the average. Note that such a definition of $P_l(k)$ gives us the following degree distribution in the first shell

$$P_1(k) = \frac{\sum_r N_{1,r}(k)}{\sum_{k,r} N_{1,r}(k)} = \frac{k N_k}{\sum_k k N_k} = \frac{k P(k)}{\langle k \rangle}, \quad (9)$$

where $P(k)$ and N_k are the degree distribution and the number of nodes with degree k in the network respectively. We bear in mind that every link in the network is followed exactly once in each direction. Hence, we find that every node with degree k is counted exactly k times. From Eq. (9) follows that $\langle k \rangle_1 = \langle k^2 \rangle / \langle k \rangle$. This quantity plays an important role in the percolation theory of networks [35] and depends only on the first and second moment of the degree distribution, but not on the correlations. Of course $P_0(k) = P(k)$.

A similar study for more distant shells gets complicated because of correlations and closed loops in the network. Let us discuss for example the degree distribution in the second shell. In this case we find that every link leaving a node of degree n is counted $n - 1$ times. Let $P(k|n)$ be a probability that a link leaving a node of degree n enters a node with degree k . Neglecting the possibility of short loops (which is always appropriate in the thermodynamical limit $N \rightarrow \infty$) we have

$$P_2(k) = \frac{\sum_n n P(n) (n - 1) P(k|n)}{\sum_n n P(n) (n - 1)}, \quad (10)$$

an expression that explicitly depends on the correlations. For uncorrelated networks, where the probability that a link connects to a node with a certain degree is independent from whatever is attached to the other end of the link, $P(k|n) = kP(k)/\langle k \rangle$. On the other hand, in the assortative case, *i.e.* when nodes attach to nodes with similar degree more likely than in uncorrelated models, $P(k|n) > kP(k)/\langle k \rangle$ for $k \approx n$. Inserting this in Eq.(10), and calculating the mean, one finds for weakly assortative networks $\langle k \rangle_2 > \langle k \rangle_2^c$ as a first approximation. Note that for strongly assortative ones the short loops could play an important role.

Fig. 7(a) shows $\langle k \rangle$ as a function of the shell number l . Here, the simulations are based on a Barabási–Albert network with $N = 30000$ nodes. We compare the shell structure for different assortative versions of the original network, ranging from an uncorrelated version ($\mathcal{A} = 0$) to strongly assortative correlated one ($\mathcal{A} = 0.777$). Note that tomographic properties can be only investigated on fully connected networks, or on connected clusters of nodes. For larger values of assortativity our initially connected network breaks into clusters, with a non-negligible part of nodes not belonging to the largest connected one. Moreover, for these extremely assortative networks the largest connected cluster exhibits a different degree distribution than the original network. Fig. 7 suggests that, independently on the degree of the initial root, any spreading phenomenon on weakly assortative networks (a realistic case) will rapidly reach highly connected vertices, and then propagate to nodes with smaller and smaller degree. On the other hand, when the assortativity increases the propagating agent does not reach the highly connected nodes so fast, and the spreading on distant shells, where the less connected vertices are found, is slower. Thus, the spreading agent infects the whole network more rapidly if the network is uncorrelated.

From Fig. 7(a) we could also conclude that the propagation changes initially quite abruptly when the assortativity increases starting from $\mathcal{A} = 0$, *i.e.*, when passing from uncorrelated networks to weakly assortative ones. Thus, weakly assortative networks present a jump in the value of $\langle k \rangle_2$ with respect to the uncorrelated ones (where $\langle k \rangle_2 = \langle k \rangle_1 = \langle k^2 \rangle / \langle k \rangle$, in the thermodynamical limit); the value of $\langle k \rangle_2$ then decreases slowly as the assortativity grows.

Real networks present assortativity only among highly connected nodes. Thus, the tomographical structure of real networks might be slightly different. However, our results indicate a general property of assortative networks that very probably is also pertinent to the realistic ones: under moderate assortativity, disease reaches highly connected nodes more rapidly than in an uncorrelated network.

Node percolation — Node percolation corresponds to removal of a certain fraction of vertices from the network, and is relevant when discussing

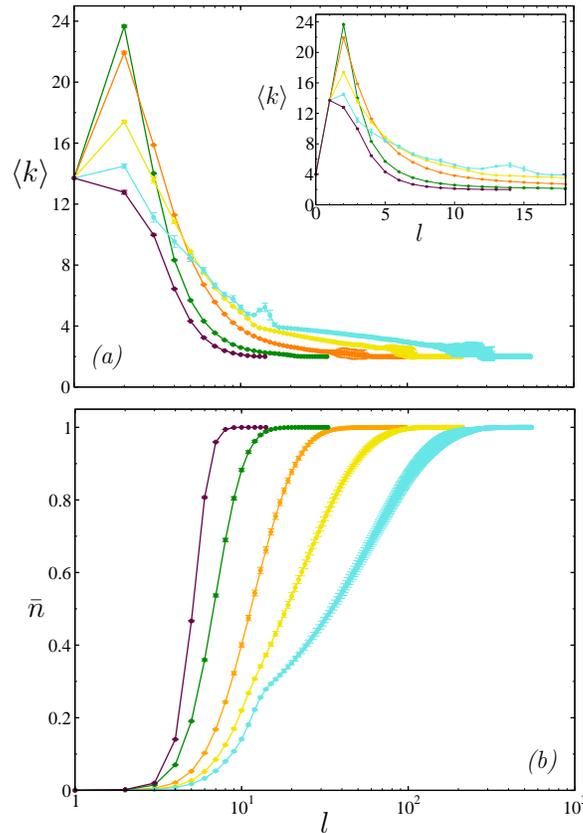


Fig. 7. (a) Average degree $\langle k \rangle_l$ as function of the number of the shell l . The four upper curves correspond, respectively, to the following values of the assortativity: $\mathcal{A} = 0.221, 0.443, 0.634$ and 0.777 . The lower curve corresponds to the uncorrelated version of the network $\mathcal{A} = 0$. Note the logarithmic scale in x . The inset shows the same curves on a linear scale, allowing to grasp the tomographical behavior close to $l = 0$. From both pictures we can see that $\langle k \rangle_1$ does not depend on correlations of the network. One also infers that the value of $\langle k \rangle_2$ decreases when the assortativity grows (see text for details). (b) $\bar{n}(l) = (\sum_{i=0}^l \sum_{r,k} N_{i,r}(k)) / (\sum_{r,k,i} N_{i,r}(k))$ as function of the shell number l . The picture shows clearly that spreading over the network gets slower the more assortative the network is. Logarithmic scale in l .

their vulnerability to a random attack (or immunization). Let q be the fraction of nodes removed. At a critical fraction q_c , the giant component (largest connected cluster) breaks into isolated clusters. Fig. 8 shows the fraction of nodes \mathcal{M} in the giant component as a function of q for different degrees of assortativity of the network. The four upper curves correspond to the values of assortativity found in natural networks. We note that the behavior of

$\mathcal{M}(q)$ changes gradually with \mathcal{A} from the uncorrelated case (upper curve) to a quite different behavior when $\mathcal{A} \rightarrow 1$ (lower curve), which indicates a very different topology in the network when it is strongly assortative. However, although the particular form of the \mathcal{M} -dependence is different for different degrees of assortativity, the absence of the transition at finite concentrations ($q_c = 1$) and the overall type of the critical behavior for correlated networks with the same $P(k)$ seems to be the same as for uncorrelated networks, namely the one discussed in Refs. [35, 36]. We thus see that this generic behavior in node percolation is only quantitatively affected by reshuffling, lowering \mathcal{M} at a given fraction of removed nodes. This quantitative behavior, however, might depend on the network's precise nature which fact has to be beared in mind when comparing our results with the ones for natural networks. We also point out that in case $\mathcal{A} \simeq 1$, a finite network is no longer fully connected: a part of the nodes does not belong to the giant component even for $q = 0$. The results suggest that, in the thermodynamical limit, the giant cluster at $q \rightarrow 0$ contains around a half of all nodes, and that its density then decays smoothly with q .

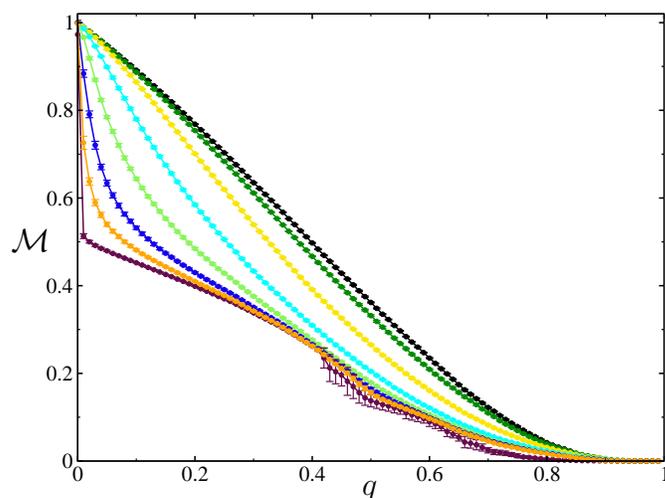


Fig. 8. Fraction of nodes \mathcal{M} in the giant component depending on the fraction of nodes removed from the network. The graph compares the results for different degrees of assortativity. From top to bottom: $\mathcal{A} = 0, 0.069, 0.221, 0.443, 0.640, 0.777, 0.856$ and 0.913 (maximal assortativity).

3. The dissortative model

3.1. Algorithm

A minor change in our algorithm can produce dissortative mixing. As before, we start from a given network and at each step we chose randomly two links of the network. We order the four corresponding nodes with respect to their degrees. Now, however, we rewire with probability p the edges so that one link connects the highest connected node with the node with the lowest degree and the other link connects the two remaining vertices; with probability $1 - p$ we rewire the links randomly [16]. In case that any of the new links already existed in the network the step is discarded and a new pair of edges selected. Varying the parameter p , it is possible to construct networks with different degrees of dissortativity. As before, the procedure does not change the degree distribution of the network and does not lead to the appearance of multiple and self-connections.

3.2. Correlations and dissortativity

Dissortativity means that nodes with high degree tend to connect to ones with low degree with larger probability than in an uncorrelated network. Of course, for undirected networks, this means that lowly connected nodes connect preferably with highly connected vertices too, and thus that the nodes with moderate degrees tend to connect among themselves. In strongly dissortative networks this tendency is very strong. Let us assume that our network is scale-free. In an intuitive way, we could say that all nodes with the highest degree should be connected with nodes with the smallest degree. Once all nodes with the highest degree are exhausted, the nodes with the second highest degree should be also connected with nodes with the minimum degree (which is possible, since in a scale-free network the weakly connected nodes build an overwhelming majority). Also nodes with the third, fourth, *etc.*, highest degree might be connected with nodes having the smallest degree, until all nodes with the minimum degree are connected. After this the nodes with the second minimum degree should be connected with those vertices with the high degree which are not yet connected, and so on. Depending of the degree distribution $P(k)$, this intuitive procedure to construct a perfectly dissortative network produces actually a strong as-sortative mixing for few medium values of k . This peculiarity of dissortative networks is also evident in our simulations for $p = 1$.

The property discussed above makes theoretical analysis of correlations \mathcal{E}_{ij} in dissortative networks somewhat involved. In the thermodynamical limit the solution similar to one given in Sec. 2 shows that the expected correlations $\mathcal{E}_{ij} \rightarrow 0$ for $p \rightarrow 1$ for all finite values of i and j . This only indicates that all outgoing edges of nodes with a certain degree k tend to link up to

vertices with infinite connectivity, which makes this thermodynamical limit unrealistic. In finite networks a direct measurement of the \mathcal{E}_{ij} function is a rather complex task due to large statistical fluctuations. This discussion shows that it might be difficult to introduce a reasonable quantity to measure the degree of dissortativity in a network. In order to study dissortative correlations in networks, some authors study the quantity

$$\langle k_{nn} \rangle = \frac{\sum_j j(1 + \delta_{kj})\mathcal{E}_{kj}}{\sum_j (1 + \delta_{kj})\mathcal{E}_{kj}}, \quad (11)$$

the nearest neighbors' average degree of nodes with degree k . Eq. (11) corresponds to a constant function of value $\langle k_{nn} \rangle = \langle k^2 \rangle / \langle k \rangle$ for uncorrelated networks, whereas the function is decreasing when the network presents dissortative correlations. For assortative networks the function is an increasing one. The value $\langle k^2 \rangle / \langle k \rangle$ diverges in the thermodynamical limit for scale-free networks with diverging second moment of the degree distribution, but is finite for any finite network.

3.3. Topological properties

Let us show how does the function described by Eq. (11) vary when the correlations are changed (Fig. 9). All simulations in this section are based on a Barabási–Albert network with $N = 30000$ nodes and $L = 60000$ links.

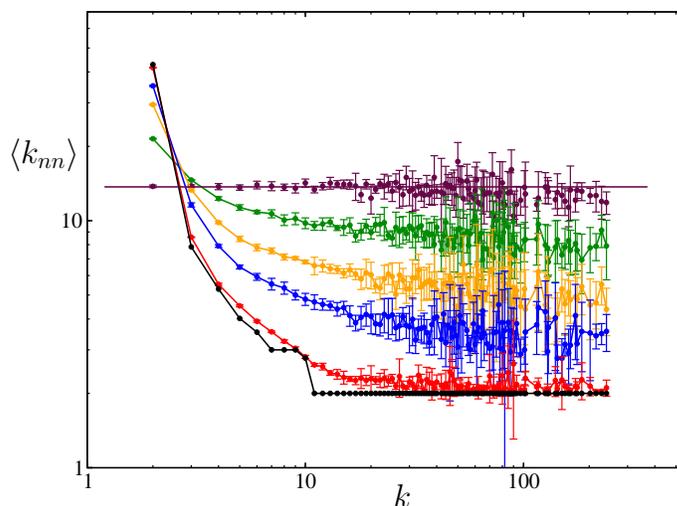


Fig. 9. Nearest neighbors' average connectivity $\langle k_{nn} \rangle$ as function of k for different dissortative correlated versions of a network. Using the parameter p to characterize the dissortativity, the picture shows from top to bottom, $p = 0$ (uncorrelated), 0.5, 0.78, 0.9, 0.999 and 1. The horizontal straight line corresponds to $\langle k_{nn} \rangle = \langle k^2 \rangle / \langle k \rangle$.

We see in Fig. 9 that $\langle k_{nn} \rangle$ stays constant and is equal to $\langle k_{nn} \rangle = \langle k^2 \rangle / \langle k \rangle$ for uncorrelated network (Eq. (1)). However, when the correlations are modified by the increase of the parameter p , this flat curve is transformed into a decreasing one, indicating the appearance of dissortative correlations. In fact, the larger the value of p , the smaller is $\langle k_{nn} \rangle$ for $k \gg 1$. This behavior corresponds exactly to what is expected. Moreover, since almost all highly connected nodes must be linked with nodes of minimum degree, the curve should exhibit a plateau for $k \gg 1$, as well as a high peak for the minimal value of $k = 2$. Both properties are revealed in our simulations (lower curve of Fig. 9).

Different dissortative real networks, as for example the Internet, show $\langle k_{nn} \rangle \propto k^{-\nu}$ [11]. Our results, plotted on a double logarithmic scale, do not reproduce this behavior. This means that dissortative networks which are random with all other respects do not reproduce all properties of dissortative real networks. We will, however, remark that real networks are also governed by the metrics of the underlying space, normally our Euclidean physical space.

A detailed analysis of strongly dissortative networks shows that they exhibit properties similar to ones we have described in our ideal dissortative case, such as the assortative correlations among nodes with medium degree. Fig. 10 shows \mathcal{E}_{4k} as function of k for different dissortative correlated networks. We observe, for example, that \mathcal{E}_{45} and \mathcal{E}_{46} increase as the dissortativity grows, in a such a way that $\mathcal{E}_{45} > \mathcal{E}_{45}^r$ as well as $\mathcal{E}_{46} > \mathcal{E}_{46}^r$. This indicates that nodes with degree $k = 4$ connect preferably with nodes with degrees $k = 5$ and $k = 6$, a clearly indication of assortativity.

Average path length — We now discuss the behavior of the average shortest path length l when the dissortativity of the network increases. Fig. 11 shows l as function of the parameter p . Although the parameter p is an internal parameter of the algorithm and does not immediately represent any property of the network, it is clear that any reasonably defined degree of dissortativity has to be an increasing function of p . Fig. 11 confirms that the average path length l always grows when the dissortativity of the network increases. At variance with the assortative case, the increase in l is moderate, and its maximal value is not much higher than in an uncorrelated network.

It is quite interesting to note the average minimal path length in all correlated networks studied in this work is larger than one for the uncorrelated ones; in might be that for a fixed degree distribution, the network with the minimum average path length would be exactly an uncorrelated one.

Clustering coefficient — Fig. 12 shows the behavior of the mean clustering coefficient C as function of p . Thus, in weakly dissortative networks the clustering coefficient does not change considerably compared to uncor-

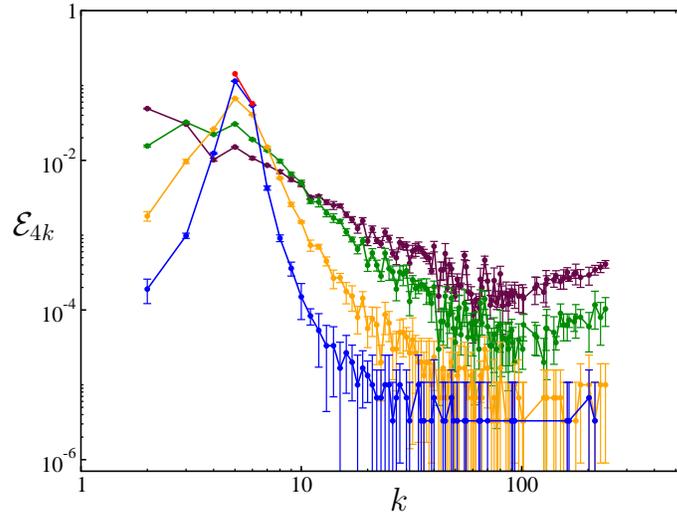


Fig. 10. \mathcal{E}_{4k} as function of k for different assortative correlated networks. From top to bottom: $p = 0$ (uncorrelated), 0.78, 0.9 and 0.999. The case $p = 1$ is also plotted; it corresponds to the curve with only two points (\mathcal{E}_{45} and \mathcal{E}_{46}) in the top part of the figure. The graph indicates the assortative behavior of nodes with moderate degree (see text).

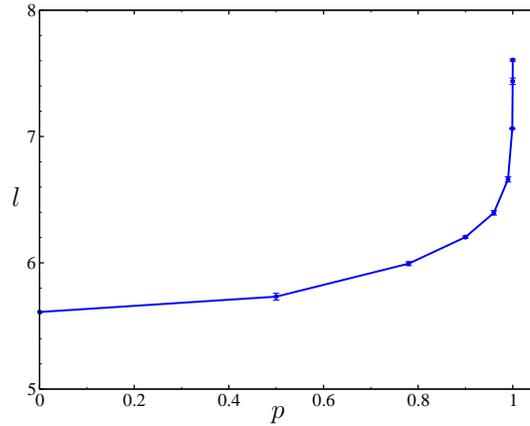


Fig. 11. Shortest average path length *versus* the parameter p , characterizing the degree of assortativity of the network.

related networks, while in strongly assortative networks C decreases and eventually vanishes. For networks exhibiting their maximum degree of assortativity (corresponding to $p = 1$), the simulations give $C = 0$: no loops of length three are present in the network.

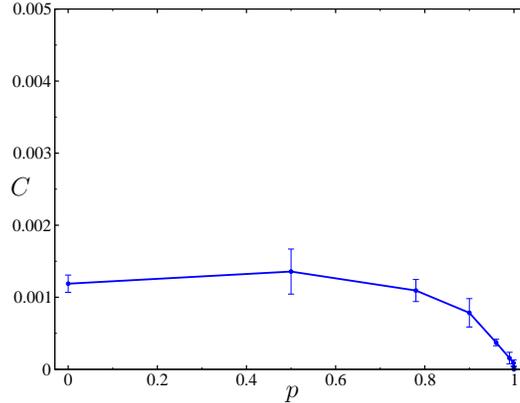


Fig. 12. Mean clustering coefficient C as function of the parameter p .

An analysis of the degree-dependent clustering coefficient $\bar{C}(k)$ shows that $\bar{C}(k)$ is smaller than in the uncorrelated case but remains practically constant. For perfectly assortative infinite correlated networks ($p = 1$) one would have $\bar{C}(k) = 0$ for all k .

Tomography — The study of tomography on the assortative versions of our original Barabási–Albert network, with $N = 30000$ nodes, yields interesting outcomes. In the Fig. 13(a) we plot $\langle k \rangle_l$ as a function of the shell number l . In uncorrelated networks the curve $\langle k \rangle_l$ reaches a maximum value of $\langle k^2 \rangle / \langle k \rangle$ in the first shell and then it smoothly decreases to $\langle k \rangle = 2$ corresponding to the minimum node degree. In a assortative network the behavior changes: the value of $\langle k^2 \rangle / \langle k \rangle$ oscillates as a function of l : now, the second shell exhibits a local minimum of $\langle k \rangle_2$, followed by a maximum for $l = 3$, decreases again for shell $l = 4$, *etc.* These jumps of the tomographical curve are typical for all assortative networks. The explanation is not complicated. Using Eq. (10), and supposing as a first approximation that there exists a large probability that a link leaving a node of degree n enters a node with “opposite” degree (*i.e.*, if n is large then k must be small, and vice versa), one can show that the calculation of the mean, $\langle k \rangle_2$, yields $\langle k \rangle_2 < \langle k \rangle_2^r$ (note that for assortative networks the number of short loops is small). Now, if the second shell possesses a large number of lowly connected nodes, the third shell must then be full of highly connected nodes, because of the assortative tendency of nodes to connect. For the same reason the shell $l = 4$ must contain mostly nodes with small degree, *etc.*

Thus, assortative correlations produce networks where the propagating agent more readily reaches nodes with small degree than in uncorrelated ones. In this networks the lowly connected nodes do not form the “periphery” of the network, and play a more important role in the spreading phenomena, since they represent bridges between the highly connected ones. The periph-

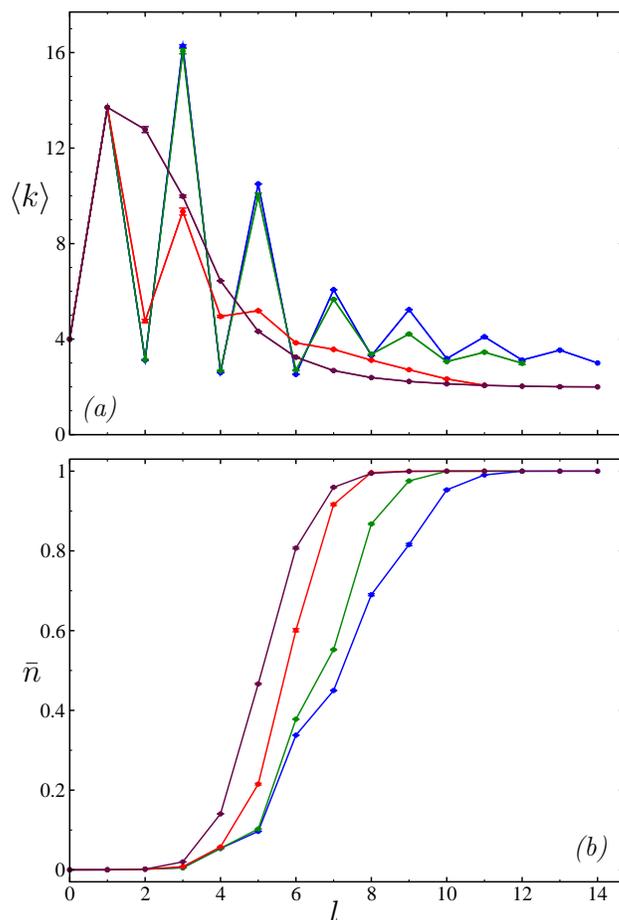


Fig. 13. (a) Average degree $\langle k \rangle_l$ as function of the number of the shell l . From top to bottom (or from more peaked to more smooth): $p = 1, 0.999, 0.9$ and 0 (uncorrelated). (b) $\bar{n}(l) = (\sum_{i=0}^l \sum_{r,k} N_{i,r}(k)) / (\sum_{r,k,i} N_{i,r}(k))$ as function of the shell number l .

ery of the network consists of a large fraction of the nodes of medium degree, which are the last ones to be affected by the spreading agent. Fig. 13(b) shows the cumulative distribution of the average number of nodes per shell

$$\bar{n}(l) = \frac{\sum_{i=0}^l \sum_{r,k} N_{i,r}(k)}{\sum_{r,k,i} N_{i,r}(k)} \tag{12}$$

as function of the shell number l . We note that the average path length increases when the degree of dissortativity grows and we can offer a qualitative

explanation for this result. We note that the difference between the curves appears starting from the shell $l = 3$. Under dissortative mixing a lot of nodes with small degree, located in the second shell, are connected to nodes with high degree in the third shell. The inverse is, however, not true since the number of highly connected nodes is small. Starting from a smaller number of nodes in the third shell (in comparison to an uncorrelated networks) hinders reaching a large number of further nodes and leads to a weaker population of the fourth shell, *etc.* A careful observation of the curves supports this argument: note also that the slope of the lines is smaller when passing from an even shell to a odd shell than when changing from an odd to an even one.

Another important property of dissortative networks compared to their assortative counterparts is that they always remain fully connected, independently of their degree of dissortativity. Our simulations show that even extremely dissortative networks ($p = 1$) are composed of a single connected cluster.

Node percolation — Fig. 14 shows the fraction of nodes \mathcal{M} in the giant component as a function of q (fraction of nodes removed) for different degrees of dissortative mixing of our Barabási–Albert network. Percolation properties of dissortative networks change when we vary the degree of dissortativity of the network.

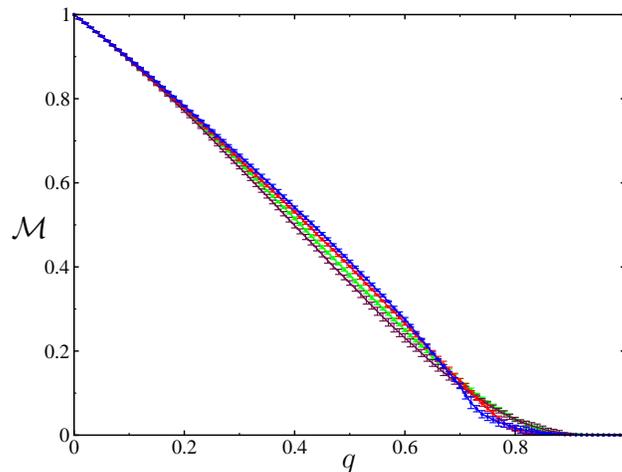


Fig. 14. Fraction of nodes \mathcal{M} in the giant component depending on the fraction of nodes removed from the network. The graph compares the results for different degrees of dissortativity. From bottom to top: $p = 0, 0.5, 0.9$ and 1 .

Although the differences between them and uncorrelated networks are not as large as in the assortative case, they could be important. Simulations show that as the dissortativity grows the robustness of the networks for small and moderate values of q increases. However, when q exceeds some characteristic value, the network tends to break more rapidly. While weakly dissortative networks exhibit a percolation behavior quite similar to one of uncorrelated ones, the behavior of strongly correlated networks seems to be different. However, our simulations do not allow us to make a definitive on the existence of a real percolation transition in such networks.

4. Conclusions

We present two algorithms based on the idea of rewiring of the preexisting network which are capable to change correlations in a network and produce assortative or dissortative mixing leaving all other properties of the network random. Both algorithms do not change the degree distribution of the network and avoid creating of multiple and self-connections. The algorithms are governed by a single parameter, p , whose variation changes the degree of assortativity (dissortativity) of the network to a desired amount. Using the algorithms we show that correlations have a drastic influence on the topological properties of networks. Assortative networks tend to form highly connected groups of nodes with similar degree which results in an increase of the average path length and clustering coefficient when the assortativity grows. The study of tomography and percolation on assortative networks shows that their transport properties differ from ones for uncorrelated networks. In dissortative networks, the tendency of hubs to connect nodes with low degree also produce changes in the topology; in comparison to uncorrelated networks they exhibit larger average path lengths and smaller clustering coefficients. All our results are pertinent to networks which do not exhibit any other correlations than the ones put by their mixing property. We have to note that the real ones might have also other types of correlations, and therefore, other statistical properties; geographical restrictions or simply peculiar properties of certain nodes may play an important role too.

IMS uses the possibility to thank the Fonds der Chemischen Industrie for the partial financial support.

REFERENCES

- [1] L.M. Sander, C.P. Warren, I.M. Sokolov, C. Simon, J. Koopman, *Math. Biosci.* **180**, 293 (2002).
- [2] L.M. Sander, C.P. Warren, I.M. Sokolov, *Physica A* **325**, 1 (2003).
- [3] L. Hufnagel, D. Brockmann, T. Geisel, *Proc. Natl. Acad. Sci. USA* **101**, 15124 (2004).
- [4] R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [5] S.N. Dorogovtsev, J.F.F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
- [6] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, Y. Aberg, *Nature* **411**, 907 (2001).
- [7] M.E.J. Newman, *Phys. Rev.* **E67**, 026126 (2003).
- [8] M.E.J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [9] A. Vázquez, M. Boguñá, Y. Moreno, R. Pastor-Satorras, A. Vespignani, *Phys. Rev.* **E67**, 046111 (2003).
- [10] A. Capocci, G. Caldarelli, P. De Los Rios, *Phys. Rev.* **E68**, 047101 (2003).
- [11] R. Pastor-Satorras, A. Vázquez, A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
- [12] A. Trusina, S. Maslov, P. Minnhagen, K. Sneppen, *Phys. Rev. Lett.* **92**, 178702 (2004).
- [13] M.E.J. Newman, J. Park, *Phys. Rev.* **E68**, 036112 (2003).
- [14] J. Berg, M. Lässig, *Phys. Rev. Lett.* **89**, 228701 (2002).
- [15] K.-I. Goh, E. Oh, B.Kahng, D. Kim, *Phys. Rev.* **E67**, 017101 (2003).
- [16] S. Maslov, K. Sneppen, *Science* **296**, 910 (2002).
- [17] P.L. Krapivsky, S. Redner, *Phys. Rev.* **E63**, 066123 (2001).
- [18] S.N. Dorogovtsev, *Phys. Rev.* **E69**, 027104 (2004).
- [19] D.S. Callaway, J.E. Hopcroft, J. M. Kleinberg, M.E.J. Newman, S.H. Strogatz, *Phys. Rev.* **E64**, 041902 (2001).
- [20] J. Park, M.E.J. Newman, *Phys. Rev.* **E68**, 026112 (2003).
- [21] M. Boguñá, R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* **90**, 028701 (2003).
- [22] V.M. Eguíluz, K. Klemm, *Phys. Rev. Lett.* **89**, 108701 (2002).
- [23] M. Boguñá, R. Pastor-Satorras, *Phys. Rev.* **66**, 047104 (2002).
- [24] N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabási, S. Havlin, *Phys. Rev.* **E66**, 015104 (2002).
- [25] A. Vázquez, Y. Moreno, *Phys. Rev.* **E67**, 015101 (2003).
- [26] Y. Moreno, J.B. Gómez, A.F. Pacheco, *Phys. Rev.* **E68**, 035103 (2003).
- [27] S. Maslov, K. Sneppen, A. Zaliznyak, *Physica A* **333**, 529 (2004).
- [28] A. Ramezanpour, V. Karimipour, A. Mashaghi, *Phys. Rev.* **E67**, 046107 (2003).

- [29] R. Xulvi-Brunet, W. Pietsch, I. M. Sokolov, *Phys. Rev.* **E68**, 036119 (2003).
- [30] M. Boguñá, R. Pastor-Satorras, *Phys. Rev.* **E68**, 036112 (2003).
- [31] I. Farkas, I. Derényi, G. Palla, T. Vicsek, *Lect. Notes Phys.* **650**, 163 (2004).
- [32] A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999).
- [33] R. Xulvi-Brunet, I.M. Sokolov, *Phys. Rev.* **E70**, 066102 (2004).
- [34] L.M. Sander, C.P. Warren, I.M. Sokolov, *Phys. Rev.* **E66**, 056105 (2002).
- [35] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).
- [36] R. Cohen, D. ben-Avraham, S. Havlin, *Phys. Rev.* **E66**, 036113 (2002).
- [37] M.E.J. Newman, I. Jensen, R.M. Ziff, *Phys. Rev.* **E65**, 021904 (2002).