Article
Chinese Science Bulletin

Statistical Physics and Mathematics for Complex Systems
December 2011 Vol.56 No.34: 3671–3676
doi: 10.1007/s11434-011-4795-2

Effect of spatial distance on search information correlation in complex systems

ZHAO TingTing* & CAI Xu

Complexity Science Center, Institute of Particle Physics, Central China Normal University, Wuhan 430079, China

Received May 12, 2011; accepted July 24, 2011

Spatial distance has a remarkable effect on the attended mode of a network embedded in a certain space. First, we investigate how spatial restriction leads to information-information correlation that is strong, linear and positive in real networks. We then construct a two-dimensional space, define the action radius R for nodes of networks, and propose a class of models that depend on spatial distance. Information correlation of the models is consistent with that of real networks. The spatial distance plays a leading role in generating assortative mixing by degree, while the generation of disassortative mixing relies on both the degree of preferential attachment and spatial restriction.

spatial distance, search information, action radius, assortative mixing

Citation: Zhao T T, Cai X. Effect of spatial distance on search information correlation in complex systems. Chinese Sci Bull, 2011, 56: 3671–3676, doi: 10.1007/s11434-011-4795-2

Watts et al. [1] pointed out that social networks are searchable, and they constructed social space using a coordinate vector. As the unit of society, human beings have cognitive competence that is lacking in other systems. However, the searching phenomenon is not particular to the social network. Thus, even if cognition can improve search efficiency [2], searching should not rely on cognition. We hold the opinion that not only the social network but also any network that can be described in a certain space has the ability to search. The space can be geographic space, in which transportation systems [3–7], communication networks [8], and infrastructure networks [9] have been well described. Alternatively, the space can be social space [1,10] in which there is, for example, a collaboration network, an epidemic or a network for spreading rumors [11,12]. However, whether in geographic or social space, all networks have one thing in common, which is that each node owns a location. Therefore, we can simply give coordinates to locate nodes. Once a network is embedded in a space, there is a spatial distance between different parts of the network. To maintain global functionality, a logical network should be searchable, and the spatial distance should have an effect.

In the first section, we study the correlation of search information for real networks. To verify the importance of spatial distance in producing a searchable network, we construct two-dimensional (2D) space as many real networks are embedded in 2D or 3D space [1,13]. We present distance-dependant models in subsequent sections.

It is found that spatial distance is a dominant factor that affects the ability to search. Moreover, a network that relies more on spatial distance will be assortative mixing by degree or disassortative.

1 Information-information correlation

We first review the concept of search information, which describes the searching ability of a network [14].

$$S(i \to b) = -\log_2 \left(\sum_{p(i,b)} \mathcal{P}\{p(i,b)\} \right), \qquad (1)$$

*Corresponding author (email: zhaottchina@gmail.com)

$$\mathcal{P}\{p(i,b)\} = \frac{1}{k_i} \prod_{j \in p(i,b)} \frac{1}{k_j - 1}, \quad (2)$$

where k_i is the degree of node i. $p(i,b)$ is one of the degenerate paths from i to b, and j indicates the node that appears on path $p(i,b)$. $S(i \to b)$ is the search information from i to b, and it reflects the difficulty of searching. The average search information of a whole network with N nodes is

$$S = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{b=1}^{N} S(i \to b). \quad (3)$$

Real networks always have larger average search information S than their random counterparts (see [15] and Figure 1), which means that searching in real networks is more difficult. Could it be that real networks are not searchable? Indeed, Rosvall et al. [15] showed that a relatively large value of S reflects a separation of neighborhoods. Because of the presence of neighborhoods, there is greater interaction in a local area. Additionally, larger information S represents greater diversity of the attended mode according to the principle of maximum entropy, which can be visualized practically with the concept that "All roads lead to Rome". Thus, a real network itself is searchable in this sense.

Since the average search information S varies with the size and evolution of the network, it cannot incarnate structural differentiation of various types of systems. Therefore, we attempt to obtain the local property of real networks from an information perspective; i.e. we investigate how search information depends on real network connections.

The sum of the search information from node i to other nodes in a network is access information $A_i = \sum_b S(i \to b)$, and the sum of the search information from other nodes to node b is hide information $H_b = \sum_i S(i \to b)$ [14]. To explore the relationship between a node and its neighbors, we define the average access information and hide information of neighbors of any node i as $A_{nn,i}$ and $H_{nn,i}$:

$$A_{nn,i} = \frac{\sum_j A_j a_{ij}}{k_i}, H_{nn,i} = \frac{\sum_j H_j a_{ij}}{k_i}, \quad (4)$$

where a_{ij} is the adjacent matrix element of the network.

Figure 2 shows positive linear information correlation.

Making use of the Pearson product-moment correlation coefficient (hereinafter referred to as Pearson's r) r_A of A_n and A_{nn} and r_H of H_n and H_{nn}, expressed as

$$r_A = \frac{\sum_i \left(A_i - \frac{\sum_i A_i}{N} \right) \left(A_{nn,i} - \frac{\sum_i A_{nn,i}}{N} \right)}{\sqrt{\sum_i \left(A_i - \frac{\sum_i A_i}{N} \right)^2 \sum_i \left(A_{nn,i} - \frac{\sum_i A_{nn,i}}{N} \right)^2}}, \quad (5)$$

$$r_H = \frac{\sum_i \left(H_i - \frac{\sum_i H_i}{N} \right) \left(H_{nn,i} - \frac{\sum_i H_{nn,i}}{N} \right)}{\sqrt{\sum_i \left(H_i - \frac{\sum_i H_i}{N} \right)^2 \sum_i \left(H_{nn,i} - \frac{\sum_i H_{nn,i}}{N} \right)^2}}, \quad (6)$$

is a more intuitive way to determine correlation characteristics.

Table 1 lists Pearson's r for the networks. The upper part of the table gives results for several real networks, with all values of r_A and r_H ranging from 0 to 1 and being closer to 1. Therefore, a linear correlation for the real network not only exists but is strong. Although the networks we refer to are undirected, they turn out to be directed from an information perspective. Furthermore, Table 1 details the asymmetry of the out (access) and in (hide) information, and r_A is always larger than r_H. This situation is quite different from that for a regular lattice, which has almost linear correlation (see the bottom of Table 1).

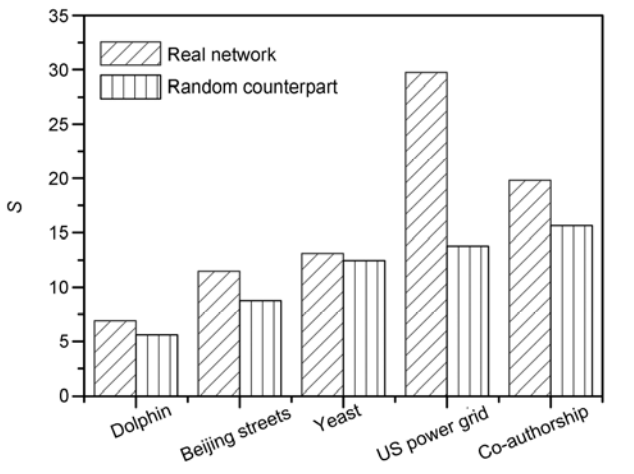

Figure 1 Average search information S of real networks and their random counterparts. The random counterpart has the same degree sequence as the real network. The network of Beijing streets is constructed by taking a dual approach [7]. Networks of dolphins, protein-protein interactions in yeast, the power grid in the United States and co-authorships are taken from http://www.cs.helsinki.fi/kurssit/syventavat/582488/MACN2006/data-code.html [16–19].

2 Analysis of network size

In the previous section, we mentioned that the average search information could not help us determine the network type. Therefore, we give Pearson's r of information-information correlation and find that real networks have remarkable linear relevance. Considering that some systems grow, we need to know the stability of Pearson's r.

Figure 3 shows the variations in average search information and Pearson's r with network scale N for the BA model. r_A and r_H remain robust for large system size. In contrast, S is divergent. We find that S is much larger when $\langle k \rangle$=2 as seen in Figure 3(a). This means that it is much harder to search in a sparse network. Figure 3(b) shows three pairs of Pearson's r with a different mean degree $\langle k \rangle$.

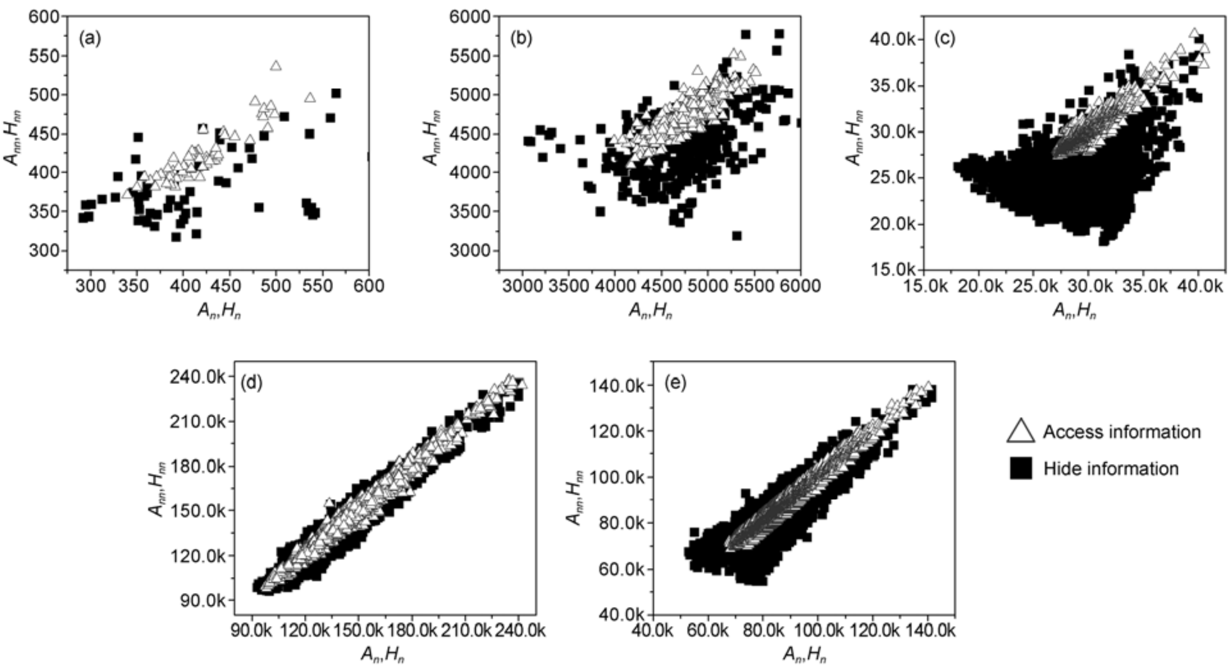

Figure 2 Scatterplots for information-information correlation. The abscissa represents access information A_n and hide information H_n of a node. The ordinate represents access information A_{nn} and hide information H_{nn} of the neighbors of the node. Access information is marked with open triangles, and hide information is marked with filled squares. (a)–(e) Real networks of dolphins, Beijing streets, protein-protein interactions in yeast, the United States power grid and co-authorship, respectively.

Table 1 Pearson's r of access and hide information r_A and r_H and the number of nodes and links (N,L)[a)]

Network (N, L)	r_A	r_H
Dolphin [16,17] (62, 159)	0.905604	0.500918
Beijing streets (415, 1136)	0.852729	0.443243
Yeast [18] (2361, 6646)	0.928872	0.258598
US power grid [19] (4941, 6594)	0.989780	0.962079
Co-authorship (7955, 10055)	0.986055	0.865384
Random counterpart of dolphin	0.478886	0.103039
Random counterpart of Beijing streets	0.442335	0.054288
DDPC counterpart of dolphin	0.909720	0.571272
DDPC counterpart of Beijing streets	0.892961	0.317105
Regular lattice I (20×20)	0.985184	0.997980
Regular lattice II [20] (20×20)	0.985191	0.994754
BA model (2010, $\langle k\rangle \approx 4$)	0.510250	−0.122566
DDPG model (2010, $\langle k\rangle \approx 4$)	0.920435	0.735093
Generalized BA model (2010, $\langle k\rangle \approx 4$)	0.895596	0.511169

a) Only the largest connected component in the network is considered. Lattice I is a simple tetragonal lattice and lattice II is a highly clustered regular lattice.

r_A and r_H both decrease with $\langle k\rangle$ increasing. Furthermore, we find that a sparse graph also has large Pearson's r, such as in the case of a BA model with small average degree (shown in Figure 3(b)) and even in the case of a sparse random network. However, why do relatively dense real networks have large Pearson's r for search information also? The following section provides an answer.

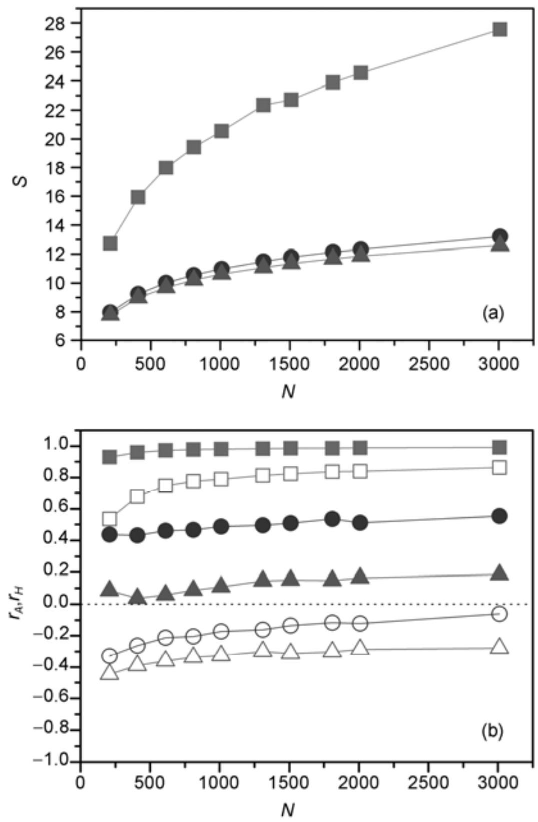

Figure 3 Squares, circles and triangles correspond to the BA model with average degree $\langle k\rangle$=2, $\langle k\rangle$=4 and $\langle k\rangle$=6, respectively. (a) Average search information S against network scale N; (b) Pearson's r for the BA model. The filled symbols represent r_A, and the open symbols represent r_H.

3 Effect of spatial distance on information-information correlation

Obviously, there must be mechanisms that lead real networks to act in an assortative manner according to search information. To investigate the mechanism, we compare a real network and its random counterpart with the same degree sequence. A real network not only has a larger value of search information than a random network (Figure 4) but also has much stronger linear correlation of search information. However, what shapes the intensive scattering (the random counterpart in Figure 4) into a strip-shaped distribution? This is investigated in the following.

A node has equal probability of connecting with any other in a random network. However, a real node does not act in this way. It chooses a more accessible connection. We know that two long streets are unlikely to intersect in an urban street network, and two persons are unlikely to be in the same group if they do not share common interests. Taking lattice networks I and II [20] (see Table 1) having linear correlation strictly as the reference, we infer that strong information correlation may partly come from local restrictions to connections. The restriction may be geographic isolation, individual differences, or limited acquaintance. On the basis of the concept of space, we place each node at a coordinate position, and simplify all restrictions as spatial restrictions.

In comparison with a random network, we present a distance-dependent preferential connecting (DDPC) network. The algorithm is as follows.

(1) Any node i in the network is embedded in 2D space at (x_i, y_i), $0<x_i, y_i<1$.

(2) A fixed degree sequence $(k_1, k_2, \ldots k_N)$ is given.

(3) Whether node i and node j are connected depends on the normalized spatial distance d_{ij}:

$$d_{ij} = \sqrt{\frac{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}{2}}. \tag{7}$$

We define the action radius R, $0<R<1$. The randomly connected nodes i and j must meet the condition that $d_{ij}<R$. Thus, every node has a fixed "home range", and the node itself is the center. In this way, we simplify the local restriction as a limited home range. Too small a value of R results in the network being disconnected, and the network approaches a random network as R approaches 1. To find a suitable value R_s, we compare simulation and empirical data.

In Figure 5, although only relatively small networks of dolphins and Beijing streets are considered, we can still see how Pearson's r varies with R. On the whole, Pearson's r decreases with R increasing. In the vicinity of R=0.2, the

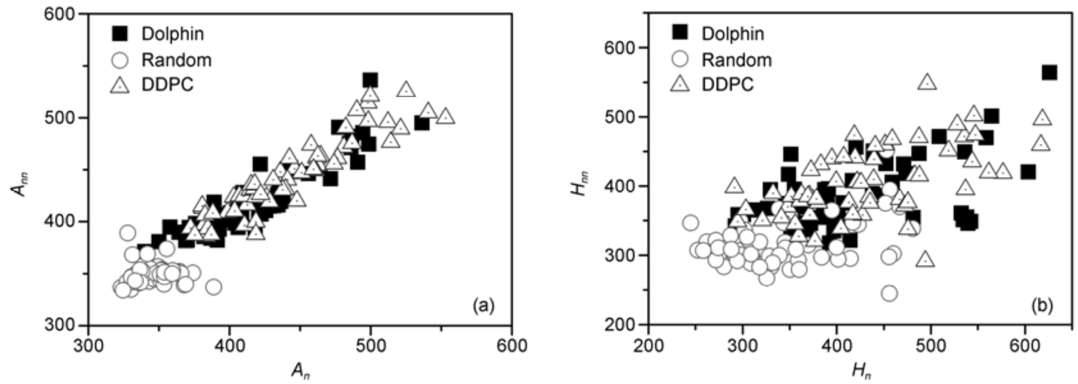

Figure 4 Comparison of a dolphin network with its randomized network and DDPC counterpart when R=0.2.

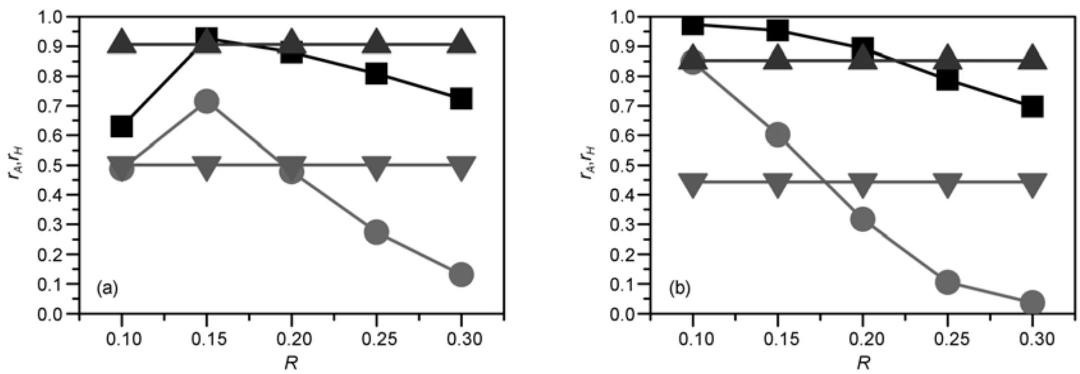

Figure 5 Pearson's r for the DDPC model. The upward and downward triangles represent r_A and r_H of a real network. The squares and circles show r_A and r_H of the DDPC counterpart. (a) Dolphins; (b) Beijing streets.

simulation results agree well with the empirical results. When R=0.1, the reproduced dolphin network collapses, and Pearson's r drops precipitously.

In Figure 6, the reproduced dolphin network has a structure most similar to the real one when R=0.2. Thus, R_s of the dolphin network approaches 0.2. Owing to spatial restrictions, some communities appear in the structure. However, when R=0.1, the network breaks into pieces. Figure 4 and Table 1 present the desired result that the DDPC model has more similar correlation distribution and Pearson's r to the real network.

Therefore, spatial restrictions result in a much larger Pearson's r for a relatively dense real network. Indeed, a suitable value for R_s varies from one kind of system to another. A small value of R_s indicates a strong local restriction.

The spatial distance between nodes can be easily understood for the network of streets and the power grid in geographic space, as well as for the networks of dolphins and co-authorship in social space. However, the network of protein-protein interaction in yeast is an exception, as we know that it is not a typical network depending on spatial distance during construction. We find that the average search information of the yeast network is closer to that of its random counterpart (see Figure 1), and the relevance of search information is less (see Table 1). We believe that this supports the view that the spatial distance affects searching in complex networks.

4 Evolving model

We present evolving models that depend on spatial distance. Taking only the spatial factor into account, we generate a distance-dependant preferential growing (DDPG) network first.

(1) The model is initialized with m_0 nodes, which are regarded as a small number of nodes.

(2) Any node i in the network is placed at a given location (x_i, y_i), $0<x_i, y_i<1$.

(3) A node j with m edges is added to the network during each time step t, $m<m_0$.

(4) Whether node i and node j are connected depends on the normalized spatial distance d_{ij}.

Considering the size effect, we assume that the number of nodes in the home range remains the same in each step. The action radius R therefore ought to vary with step t. Thus, $(m_0+t)R(t)^2=C$, where $R(t)=\sqrt{C/(m_0+t)}$ and C is an alterable parameter. The randomly connected nodes i and j also meet the condition that $d_{ij}<R(t)$ at step t. The smaller C is, the more obvious the aggregate nature of the node in the network will be. In this way, the space subdivision becomes increasingly fine as the model evolves. Moreover, to ensure the network is connected, we let the new node with no connection within its home range connect with its nearest neighbor. As shown in Table 1, the DDPG model has much stronger information correlation. However, only assortative mixing by degree is found in this model. Thus, the problem remains. There are in fact a large number of technological and biological networks that tend to be disassortative [21].

Next, we take the degree into account and give a generalized BA model by adopting the mechanisms of preferential attachment and growth. Node i has attachment probability Π_{ij} of connecting to another node j within its home range:

$$\Pi_{ij}=\frac{k_i}{\sum_i k_i}, d_{ij}<R(t). \tag{8}$$

Figure 7 shows search information correlation characteristics for the distance-dependant evolving models. The characteristics are quite different from those of the mentioned BA model as shown in Figure 3(b). Comparing with

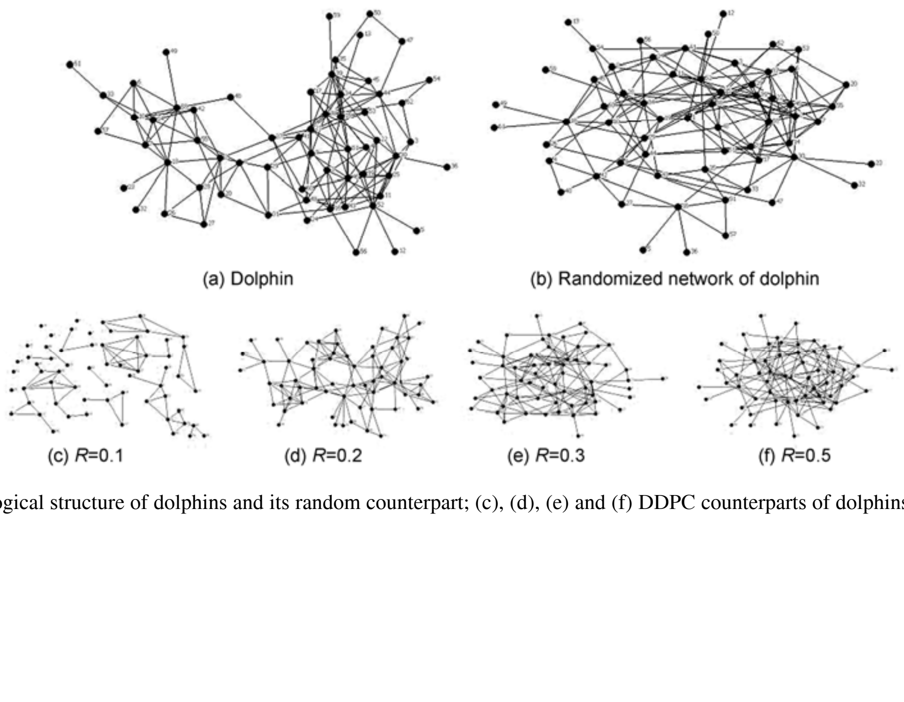

Figure 6 (a), (b) Topological structure of dolphins and its random counterpart; (c), (d), (e) and (f) DDPC counterparts of dolphins for R=0.1, R=0.2, R=0.3 and R=0.5, respectively.

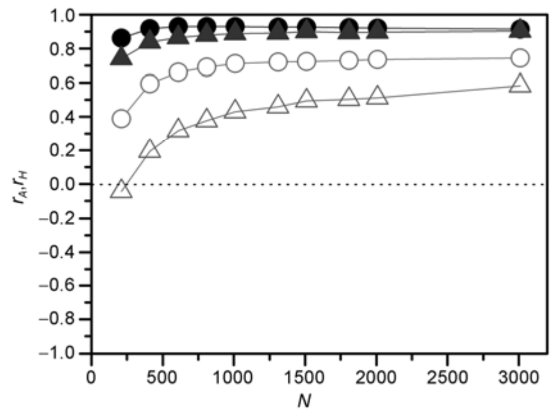

Figure 7 Pearson's *r* against network scale *N*. Circles and triangles represent the DDPG model and generalized BA model respectively; *m*=3, *C*=10. The filled symbols represent r_A and the open symbols represent r_H.

the DDPG model, the difference between r_A and r_H of the generalized BA model shows an increasing trend; moreover, we find that, under the same condition, the degree preferential attachment can lead to disassortative mixing by degree; there is only assortative mixing under the sole restriction of space. For example, when *t*=2000, the assortativity coefficient [21] is 0.24 and −0.03 for the DDPG model and generalized BA model in Figure 7. This discovery appears to support assortativeness, which quantifies the tendency for preferential association in various empirical networks. Social networks should be subjected to stronger spatial restrictions for there to be assortative mixing. In the generalized BA model, large *m* and small *C* usually enhance local aggregation and weaken the influence of the degree. Thus, for technological or biological networks, degree preferential attachment is more important, and there must be spatial restriction also for there to be strong information correlation.

5 Conclusion

We studied information-information correlation in networks. It was found that all investigated real networks are positively information-information correlated; i.e. a node with a high value of access (hide) information tends to connect to nodes with a high value of information. This is quite different from the case for some typical models. Simulation showed that a spatial restriction is a main factor leading to this phenomenon. DDPC and DDPG models restrict the connection of each node to within a fixed home range, which can be considered to represent geographic coverage, a community, acquaintance, or similar in practice. As long as the node is active in a given area, the search information of the network is not disordered but positively related to neighbors in a linear way.

A network that is constructed depending on spatial distance has assortative mixing by degree only. Thus, one obtains the result that the spatial factor is dominant in generating an assortative network by degree. To obtain a disassortative one, we introduced the generalized BA model. Because of the degree preferential attachment and spatial restriction, the network tends to be disassortative by degree and has large Pearson's *r* for search information.

T. Zhao thanks Prof. Li and other members in their group for valuable suggestions. This work was supported by the National Natural Science Foundation of China (*10647125, 10635020, 10975057 and 10975062*) *and the Program of Introducing Talents of Discipline to Universities* (*B08033*).

1 Watts D J, Dodds P S, Newman M E J. Identity and search in social networks. Science, 2002, 296: 1302–1305
2 Wang W, Zeng G S. Bayesian cognitive trust model based self-clustering algorithm for MANETs. Sci China Inf Sci, 2010, 53: 494–505
3 Boccaletti S, Latora V, Moreno Y, et al. Complex networks: Structure and dynamics. Phys Rep, 2006, 424: 175–308
4 Li W, Cai X. Statistical analysis of airport network of China. Phys Rev E, 2004, 69: 046106
5 Li W, Cai X. Empirical analysis of a scale-free railway network in China. Physica A, 2007, 382: 693–703
6 Sergio P, Crucitti P, Latora V. The network analysis of urban streets: A primal approach. Environ Plan B, 2006, 33: 705–725
7 Sergio P, Crucitti P, Latora V. The network analysis of urban streets: A dual approach. Physica A, 2006, 369: 853–866
8 Barabāsi A L, Albert R. Emergence of scaling in random networks. Science, 1999, 286: 509–512
9 Albert R, Albert I, Nakarado G L. Structural vulnerability of the North American power grid. Phys Rev E, 2004, 69: 025103(R)
10 Bogunā M, Pastor-Satorras R, Díaz-Guilera A, et al. Models of social networks based on social distance attachment. Phys Rev E, 2004, 70: 056122
11 Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks. Phys Rev E, 2001, 63: 066117
12 Pastor-Satorras R, Vespignani A. Dynamics of rumor spreading in complex networks. Phys Rev E, 2004, 69: 066130
13 Kosmidis K, Havlin S, Bunde A. Structural properties of spatially embedded networks. Europhys Lett, 2008, 82: 48005
14 Sneppen K, Trusina A, Rosvall M. Hide and seek on complex networks. Europhys Lett, 2005, 69: 853–859
15 Rosvall M, Trusina A, Minnhagen P, et al. Networks and cities: An information perspective. Phys Rev Lett, 2005, 94: 028701
16 Lusseau D, Schneider K. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Behav Ecol Sociobiol, 2004, 54: 396–405
17 Lusseau D. The emergent properties of a dolphin social network. Proc R Soc Lond B, 2003, 270: S186–S188
18 Sun S W, Ling L J, Zhang N, et al. Topological structure analysis of the protein-protein interaction network in budding yeast. Nucleic Acids Res, 2003, 31: 2443–2450
19 Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. Nature, 1998, 393: 440–442
20 Jiang B, Claramunt C. Topological analysis of urban street networks. Environ Plan B, 2004, 31: 151–162
21 Newman M E J. Assortative mixing in networks. Phys Rev Lett, 2002, 89: 208701