Physics Letters A 381 (2017) 3377–3383

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Common neighbour structure and similarity intensity in complex networks

Lei Hou [a], Kecheng Liu [a,b,*]

[a] Informatics Research Centre, Henley Business School, University of Reading, Reading, Berkshire RG6 6UD, United Kingdom
[b] Data Science and Cloud Service Research Centre, Shanghai University of Finance and Economics, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 11 May 2017
Received in revised form 21 August 2017
Accepted 21 August 2017
Available online 25 August 2017
Communicated by C.R. Doering

Keywords:
Complex networks
Vertex similarity
Common neighbours

ABSTRACT

Complex systems as networks always exhibit strong regularities, implying underlying mechanisms governing their evolution. In addition to the degree preference, the similarity has been argued to be another driver for networks. Assuming a network is randomly organised without similarity preference, the present paper studies the expected number of common neighbours between vertices. A symmetrical similarity index is accordingly developed by removing such expected number from the observed common neighbours. The developed index can not only describe the similarities between vertices, but also the dissimilarities. We further apply the proposed index to measure of the influence of similarity on the wring patterns of networks. Fifteen empirical networks as well as artificial networks are examined in terms of similarity intensity and degree heterogeneity. Results on real networks indicate that, social networks are strongly governed by the similarity as well as the degree preference, while the biological networks and infrastructure networks show no apparent similarity governance. Particularly, classical network models, such as the Barabási–Albert model, the Erdös–Rényi model and the Ring Lattice, cannot well describe the social networks in terms of the degree heterogeneity and similarity intensity. The findings may shed some light on the modelling and link prediction of different classes of networks.

1. Introduction

Networks can efficiently describe a wide range of complex systems, such as social systems, biological systems and infrastructure systems [1–4]. Since most real-world networks are either incomplete or evolving, to understand the dynamics and growing patterns of networks has attracted increasing attentions [5–11].

The degree preference has been considered as the key attractiveness driving the evolution of networks [12–15] since the finding of scaling phenomena [16]. However, real networks are also found to be highly clustered [17] and with dense community structure [18,19] which cannot be explained by the preferential attachment mechanism alone. Accordingly, vertex similarity is also argued to be a driver for networks [20] and has been applied to study the formation and evolution of different networks [21–24]. While the ground-truth similarities among vertices are mostly unknown, a number of similarity indices have been developed by evaluating either the adjacency matrix or the common neighbour structure of the network [25–29]. Normally, the vertices that share the same neighbours (adjacent vertices) are considered to be similar to each other. However, the similarities quantified by these indices mostly have systematic bias regarding the vertex degree [6,26,30] that hub vertices tend to have more common neighbours with others due to their rich connectivities. As a consequence, it is difficult to determine whether the common neighbours are due to the similarity between vertices or just random mechanism. Additionally, most indices give only positive values without an indication of neutral similarity. Even with a same similarity value, the meaning would be different in different scenarios such as the degrees of the measured vertices and the degree distribution of the given network. For example, two vertices α and β having 5 common neighbours could indicate that they are extremely similar to each other if their degrees are $k_\alpha = k_\beta = 5$, but could also be interpreted as extremely dissimilar if their degrees $k_\alpha \approx N, k_\beta \approx N$ where N is the network size, because they are expected to have a lot more common neighbours. Therefore, the key question needs to be answered is that how many common neighbours two particular vertices are expected to share in a given network. Finally, to what extent does the similarity shape the structure and evolution of a given network is still an open question due to the lack of an unbiased and symmetrical similarity index.

In this paper we study the expected number of common neighbours between two vertices which is shown to be determined by

* Corresponding author.
E-mail addresses: l.hou@pgr.reading.ac.uk (L. Hou), k.liu@henley.ac.uk (K. Liu).

http://dx.doi.org/10.1016/j.physleta.2017.08.050
0375-9601/

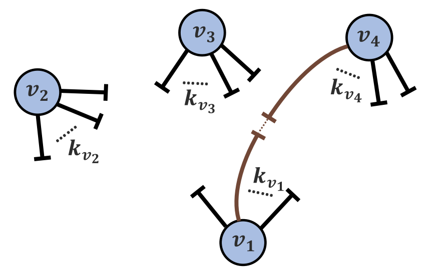

Fig. 1. (Colour online.) Illustration of the random rewiring. Each vertex v in the network has k_v half-edges to be paired with others' and each pair of half-edges has equal chance to be connected. Obviously, vertices with more half-edges are more likely to be connected to each other.

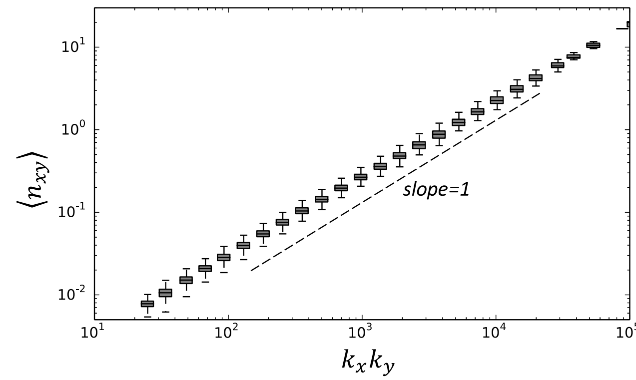

Fig. 2. Number of common neighbours between two vertices x and y, n_{xy} versus the product of the corresponding vertices' degrees $k_x k_y$ in BA networks. The dashed straight line has a slope of 1 in the log–log plot. The simulated network starts from a complete network of $m_0 = 6$ vertices. At each of the following step, one vertex is added to the network to connect to $m = 5$ existing vertices. The probability of each vertex being connected is proportional to its current degree, i.e. $p(v) \propto k_v$. Vertices are added continuously until the network size reach $N = 10^4$. Considering most vertex pairs would have no common neighbour at all in a single realisation of network, we average n_{xy} over 10^4 realisations of the generated BA network. We rewire the generated BA network as follows: a) select two from $N\langle k \rangle/2$ edges uniformly at random; b) chose one vertex from each edge and switch if this will not result in multi edges or self-loops; c) repeat a) and b) for $N\langle k \rangle$ times. In such way, the degree of each vertex will not be changed and we can average the number of common neighbours between two specific vertices accordingly.

the degree heterogeneity of the network. A vertex similarity index is thereby proposed by comparing the number of common neighbours with the expected number so that the random factors are removed. We further define the similarity intensity to quantify the governance of similarity in complex networks as the average similarity over all the connected vertex pairs in the network. The similarity intensities and degree heterogeneities of fifteen real networks are investigated and the social networks are found to be a special class which has both high degree heterogeneity and similarity intensity.

2. A balanced vertex similarity index

The vertices that share common neighbours are usually considered to be similar to each other. However, two vertices x and y that are not similar to each other at all, especially these with large degrees, could still have common neighbours by chance. For example, in a network of 10 vertices, x and y with degrees $k_x = k_y = 6$ should have at least 3 common neighbours, but having 3 common neighbours does not mean that they are similar. In other words, every pair of vertices x and y with no similarity are expected to have a certain amount of common neighbours n_{xy}^{exp} due to pure random mechanism. In a given network, if the observed number of common neighbours $n_{xy} = n_{xy}^{exp}$, we can consider these two vertices x and y to be neutral to each other. Accordingly, the difference between the observed and expected number of common neighbours $n_{xy} - n_{xy}^{exp}$ can be used to describe the tendency of x and y to connect the same vertices, which we argue is a more meaningful way to represent their similarity. Therefore, we calculate the expected number of common neighbours between two vertices with given degrees in a given network, so that we can remove the random-caused common neighbours from the observed number to estimate their similarity.

Consider a network consisting of a set of N vertices $V = \{v_1, v_2, \cdots, v_N\}$, and a set of M edges $E = \{e_1, e_2, \cdots, e_M\}$. The expected number of common neighbours between two vertices can be calculated by considering a random rewiring process of a network. Assume all the edges are broken into two half-edges (stubs) and thus each vertex v has k_v half-edges to be paired again with others as shown in Fig. 1. This process is normally referred as the configuration model [2,31] which generates random networks with a given degree sequence. In the rewiring process of the present paper, for each of v's half-edges, the paired half-edge is chosen randomly but from another vertex that has not been connected by v to avoid multi edges or self-loops. Therefore, the probability of the paired half-edge coming form vertex j is $k_j / \sum_v k_v$. Considering all the k_i edges that vertex i possessing, we have the probability of two random vertices i and j connecting with each other [32,33],

$$p(i \leftrightarrow j) = \frac{k_i k_j}{\sum_v k_v}. \tag{1}$$

Accordingly, the probability of a vertex i being a common neighbour for vertices x and y, i.e. connecting to both x and y, can be written as,

$$p(i \leftrightarrow x, y) = \frac{k_i(k_i - 1)}{(\sum_v k_v)^2} \cdot k_x k_y. \tag{2}$$

Considering all the possible common neighbours, we then have the expected number of common neighbours for x and y which reads,

$$n_{xy}^{exp} = \sum_i p(i \leftrightarrow x, y) = \frac{\sum_v k_v(k_v - 1)}{(\sum_v k_v)^2} \cdot k_x k_y. \tag{3}$$

Therefore, as suggested by Eq. (3), the neighbourhood size for two vertices x and y is expected to have a linear relation with the product of their degrees, i.e. $n_{xy}^{exp} \propto k_x k_y$. We test such relation using the Barabási–Albert (BA) network model [16]. The BA model is a random network model in which the edges are attached randomly according to the degree preference without predefined similarity. Accordingly, the vertices in a BA network are expected to be with no similarity and thus we should have $n_{xy}^{exp} = n_{xy}$. As shown in Fig. 2, the averaged number of common neighbours for two vertices x and y has the linear correlation with the product $k_x k_y$ as predicted by the Eq. (3).

Actually, one can find that, in Eq. (3), $\sum_v k_v$ can be given by the product of the network size and the average degree, $N\langle k \rangle$. Accordingly, we have also $\sum_v k_v(k_v - 1) = N\left(\langle k^2 \rangle - \langle k \rangle\right)$. Therefore, we can rewrite the expression for the expected number of common neighbours as

$$n_{xy}^{exp} = \frac{\langle k^2 \rangle - \langle k \rangle}{N \langle k \rangle^2} \cdot k_x k_y. \tag{4}$$

The parameter for the product of the degrees basically describes the degree distribution feature of the network. The component $\langle k^2 \rangle / \langle k \rangle^2$ is usually used to described a network's degree heterogeneity H [29,34]. With a unified degree for each vertex, a network has $\langle k^2 \rangle = \langle k \rangle^2$ and thus heterogeneity $H = 1$. The more heterogeneous the network's degree distribution is, the higher the value H would generally be. The BA network with the applied settings in