IEICE Proceeding Series

Classification of real networks by using classical multidimensional scaling

Yong Gao, Kaori Kuroda, Yutaka Shimada, Kantato Fujiwara, Tohru Ikeguchi

Vol. 2 pp. 126-129
Publication Date: 2014/03/18
Online ISSN: 2188-5079

Classification of real networks by using classical multidimensional scaling

Yong Gao†, Kaori Kuroda†, Yutaka Shimada⋆⋄, Kantato Fujiwara† and Tohru Ikeguchi†‡

† Graduate School of Science and Engineering, Saitama University
255 Shimo-Ohkubo, Sakura-ku, Saitama-city, Saitama, 338-8570 Japan
‡ Brain Science Institute, Saitama University
255 Shimo-Ohkubo, Sakura-ku, Saitama-city, Saitama, 338-8570 Japan
⋆ FIRST, Aihara Innovative Mathematical Modelling Project, JST
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan
⋄ Institute of Industrial Science, the University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan
Email: gaoyong@nls.ics.saitama-u.ac.jp, kuroda@nls.ics.saitama-u.ac.jp, yshimada@sat.t.u-tokyo.ac.jp,
kantaro@mail.saitama-u.ac.jp, tohru@mail.saitama-u.ac.jp

Abstract—Real systems can be described by networks: Internet, WWW, neural networks, and human relationships. A novel framework which combines the complex network theory and the nonlinear time series analysis becomes one of the useful tools to understand characteristics of these complex networks and reveal hidden structures underlying in these complex networks. In the framework, using the classical multidimensional scaling, complex networks can be transformed into time series. In this paper we investigated the distribution of values of the time series transformed from the networks by the transformation method of the framework. We compared the distributions obtained from real networks with those from networks generated from the Watts-Strogatz model and the Barabási-Albert model to discuss characteristic properties of the real networks.

1. Introduction

Networks are one of the useful tools to describe various real systems, such as WWW[1], humans relationship[2], neural network[3], traffic networks[4] and so on. Networks consist of vertices and edges. In these real networks, vertices could be websites, neurons, human, trains; edges could be links, friendship, axons or railways. Recent researches on these complex networks have revealed that the real networks have common properties such as small-world property[3], scale free property[5] and so on.

In the past decade, to analyze structural features of the real complex networks, several measures have been proposed, for example, characteristic path length, clustering coefficient, and degree correlation. Even though these measures can effectively characterize the structural features of the real networks, it is still important to evaluate real networks from various points of view because of a wide variety of the real networks.

A transformation method from complex networks to time series can be a new effective tool to evaluate the structures of the networks from the perspective of time series[6]. For example, networks show intrinsic distributions of the values of the time series[7]. Although the distributions of the timeseries have been numerically calculated from the time series of networks generated from a few mathematical models in Ref. [7] it is still unclear. However, the distributions time series transformed from real complex networks are evaluated.

In this paper, we analyzed six real networks using the transformation method[6]: the western states power grid network of the United States[3], a word adjacency network of common adjectives and nouns[8], a neural network of C. elegans[3], an American college football network which consists of American football games between Division IA colleges 2000[9], a jazz musicians network[10], and a US air lines network[11].

2. From networks to time series

To transform networks into time series, we used a method of Ref.[6]. In this method, the classical multidimensional scaling(CMDS)[12] is used as a main tool to transform the networks to time series.

Let $A = (a_{ij})$ be an $N \times N$ adjacency matrix of an undirected and unweighted network with N vertices. If vertices v_i and v_j are adjacent, $a_{ij} = 1$, otherwise $a_{ij} = 0$. In Ref. [6], the quasi-distance d_{ij} between vertices v_i and v_j is defined as follows: if $a_{ij} = 0$ $(i \neq j)$, $d_{ij} = w$ (> 1) otherwise $d_{ij} = a_{ij}$. By using the CMDS[12], we calculated coordinate values that satisfy the defined quasi-distance as accurately as possible. At first, A is transformed into a squared distance matrix $D = (d_{ij}^2)$. Next, D is transformed by $G = -\frac{1}{2} J_N D J_N^T$, where $J_N = E - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$, where E is an $N \times N$ unit matrix and $\mathbf{1}_N$ is a column vector with N ones.

The coordinate values of vertices are calculated by decomposing G into eigenvalues and eigenvectors:

$$G = P\Lambda P^T = (P\Lambda^{\frac{1}{2}})(P\Lambda^{\frac{1}{2}})^T = XX^T,$$

where

$$\begin{aligned}\Lambda_N^{\frac{1}{2}} &= \mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_N}),\\ P &= (\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_N)^T,\\ \boldsymbol{p}_m &= (p_{m1},p_{m2},\cdots,p_{mN})^T(\lambda_1 \geq \lambda_2 \geq \cdots, \geq \lambda_m).\end{aligned}$$

Let h be the number of nonzero eigenvalues of G. Then, the coordinate matrix X is described by $X = (\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N)^T$, and $\boldsymbol{x}_m = (x_{m1},x_{m2},\cdots,x_{mh})^T$ is the h-dimensional coordinate value of the vertex v_m. The number of nonzero eigenvalues h corresponds to the dimension of the Euclidean space where the vertices are arranged. Finally, we defined the patterns of coordinate values of N vertices, or the eigenvectors, as a time series $x_{mt} = \sqrt{\lambda_m}p_{mt}$ $(1 \leq m \leq h, 1 \leq t \leq N)$.

This means that the indices of vertices are treated as quasi-discrete time. In this paper, we focused on the distribution of values of the time series, x_{ij}. Thus, we did not use the temporal information. Namely, we investigated distributions of coordinate values of time series transformed from the real networks, then compared the results of the real networks with those of network models. In the following results, we used 100 eigenvectors whose eigenvalues correspond to $\lambda_1,\lambda_2,\cdots,\lambda_{100}$. We used small values of w which is close to unity according to Ref.[6].

2.1. Network model

We first used two network models: the Watts-Strogatz (WS) model[3] and the Barabási-Albert (BA) model[5]. The WS model starts from a ring-lattice, and can generate small-world networks and random networks. In the Watts-Strogatz model, an initial state is a ring-lattice network with N vertices whose degree is k. Then, the edges in the ring-lattice are rewired at random with a probability p. When $p = 0$, the network is the same as the initial ring-lattice network that has a large characteristic path length L and a small clustering coefficient C. When the probability p gets larger, the generated networks change to a small-world network which has a small L and a large C. When $p = 1$, all edges are rewired, and the generated network is a random network which has a small L and a small C. In numerical simulations, we set N to 1,000 and k to 10.

Networks generated by the BA model have a power-law degree distribution which is also found in many real networks, for example, the World Wide Web[5], power grids[5] and so on. In the BA model, we start from a complete graph with m_0 vertices. Next, at every time step t, a new vertex with m edges is added to the network repeatedly. The new vertex then connects to m different pre-existing vertices with the following probability:

$$P_i(t) = k_i(t)/\sum_{j=1}^{m_0+t} k_j(t).$$

The new vertices are added until the number of vertices becomes N. In this paper, we set N to 5,000, m_0 to 11 and m to 10.

2.2. Real networks

First, we used six real networks. The first one is the western states power grid of United States[5]. The vertices can be generators or power substations and the edges can be electric cables. The second one is a word-adjacency network which adjectives and nouns in the novel of David Copperfield by Charles Dickens[8]. The vertices are nouns and adjectives. For example, in a phrase of “the big green bus,” the adjective “big” is connected to the noun “bus.” The third one is a neural network of C. elegans[5] in which the vertices are neurons and the edges are axons. The fourth one is an American college football network in which the vertices correspond to teams and the teams are connected if they have games in a regular season[9]. The fifth one is a jazz-musicians network which the vertices are musicians who are connected by an edge if two musicians have played in the same band[10]. The sixth one is a US airline network[11] in which the vertices are airports connected by direct flights. We show summarized the clustering coefficient C, the characteristic path length L, the network size N and the average degree $<k>$ of these networks in Table 1.

Table 1: Clustering coefficients L, characteristic path lengths L, network sizes N and average degrees $<k>$ of the real networks. The clustering coefficients and the characteristic path lengths are normalized by corresponding randomized networks C_R and L_R.

network	C/C_R	L/L_R	N	$<k>$
ring lattice	9.2	3.0	112	8.0
	53.6	2.3	1,000	10.0
small-world	1.2	1.4	112	8.0
	51.1	1.0	1,000	10.0
random	1.0	1.0	112	8.0
	1.0	1.0	1,000	10.0
scale-free	5.0	0.9	5,000	19.9
power grid	16.0	1.5	4,941	2.7
adjective and nouns	3.0	1.0	112	7.6
C. elegans	5.8	1.0	297	14.5
American football	5.1	1.1	115	10.7
Jazz musicians	4.5	1.2	198	27.7
US airlines	19.7	1.3	332	12.8

3. Results

We first show the distribution of coordinate values of time series transformed from the WS network model. We show the distribution of the coordinate values when the rewiring probability p is changed in Fig. 1. In Fig. 1, we found that the distribution of the coordinate values change from a bimodal distribution ($p = 0$) to a trapezoid-like distribution ($p = 0.01$) as the rewiring probability p increases. Finally, it becomes an unimodal distribution ($p = 1$). (see also Fig. 2)

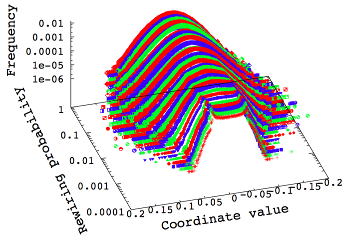
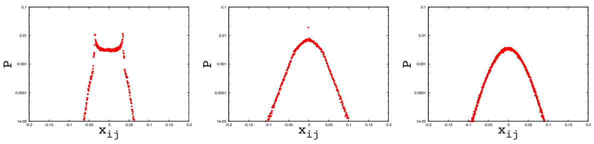


Figure 1: Distributions of the coordinate values of the time series transformed from the WS model when the rewiring probability p is changed.



(a) (b) (c)

Figure 2: Distributions of the coordinate values of the WS model: (a) a ring-lattice network, (b) a small-world network, and (c) a random network. The results are averaged over 50 tria

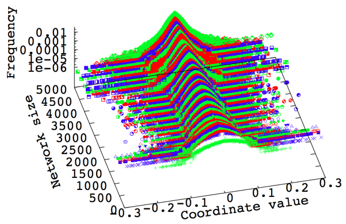
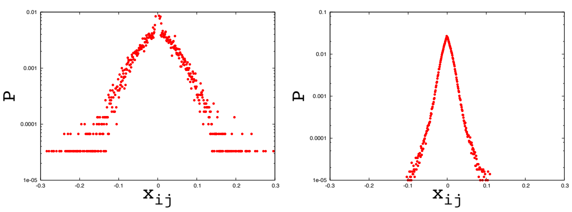


Figure 3: Distributions of the coordinate values of the BA model when the network size N is changed from 300 to 5,000.

Second, in the case of the BA model, we investigated the distribution of the coordinate values when the network size increases (Fig. 3). The distributions have unimodal shape, which is the same tendency as the random networks generated from the WS model. However, the peak of the distribution of the BA model becomes sharp as the network size increases (see also Fig. 4).



(a) (b)

Figure 4: Distributions of the coordinate values of the BA model: scale-free networks which have (a) 300 vertices and (b) 5,000 vertices. The results are averaged over 30 trials.

Third, we examined six real networks. All of the real networks are regarded as undirected and unweighted by symmetrization ($a_{ij} = a_{ji} = 1$,if $a_{ij} + a_{ji} \geq 1$). The results are summarized in Fig. 5 in the networks of C. elegans, the power grid and the jazz-musician, the distribution shows sharp unimodal distributions (Fig. 5(a),(c),(i),(k)). These distributions are similar to that of the BA model(Fig. 4).

From Fig. 5(e), the distribution of the coordinate values of the adjacency network of common adjectives and nouns has an unimodal distribution which is not sharp and similar to that of the random network of the WS model. The network of American football games (Fig. 5(g)) also has the unimodal distribution which is also similar to that of the random network.

We have focused on the word adjacency network and the American football network. These two networks have small L and large C compared with the random networks Table 1. However, we can distinguish these two networks by coordinate values.

In Fig. 6, we calculate the distributions by changing the number of eigenvectors. When we only use the eigenvectors corresponding to large eigenvalues ($0 \leq m \leq 200$), the two networks have intrinsic distributions and we cannot detect their differences. However, when we use the eigenvectors corresponding to not only small but also large eigenvalues, we can find their difference.

The number of the values of zero clearly increases in the word adjacency network (Fig. 6(b)), which is similar to the scale-free network (Fig. 6(d)). On the other hand, the values of zero do not increase in the American football network (Fig. 6(a)), which is similar to the small-world network (Fig. 6(c)). In addition, Fig. 6 indicates that each eigenvector might reflect important properties of the networks. In particular, the eigenvectors corresponding to large eigenvalues might show structural property of scale-free networks.

4. Conclusions

In this paper, we used the transformation method from networks to time series[6] to investigate real networks. We compared the distribution of the time series transformed from the real networks by the transformation method to discuss the results obtained from the real networks. Although they are different in size, average degree, clustering coefficient and characteristic path length, the real networks could have distributions of the coordinate values like the network models.

Acknowledgment

The authors would like to thank Mr. S. Ogawa and AGS Corp. for their encouragement on this research pro ject. The research of K. K. was supported by was supported by the Japan Society for the Promotion of Science (JSPS) research fellow (No. 24-9982). The research of Y. S. was supported by the Aihara Project, the FIRST program from

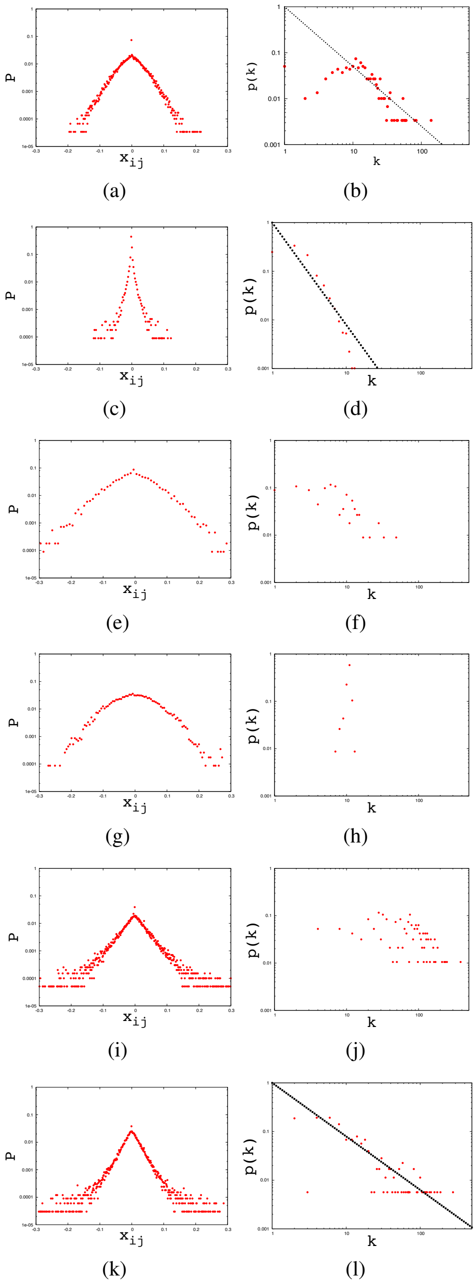


Figure 5: On the left side is the distribution of the coordinate values of (a) the neural network of C. elegans, (c) the western states power grid of the United States, (e) the adjacency network of common adjectives and nouns in the novel of David Copperfield by Charles Dickens, (g) the network of American football games, (i) the network of jazz-musician and (k) the network of US air lines. The right column shows the degree distributions corresponding to each network in the double logarithmic scale. (b) The dotted lines in (b), (d) and (l) have slopes of $-1.3, -2.1$ and -1.1 respectively.

JSPS, initiated by CSTP. The research of K. F. was supported by Grant-in-Aid for Young Scientists (B) 24700221 from MEXT of Japan. The research of T. I. was supported by a Grant-in-Aid for Exploratory Research (Grants No.

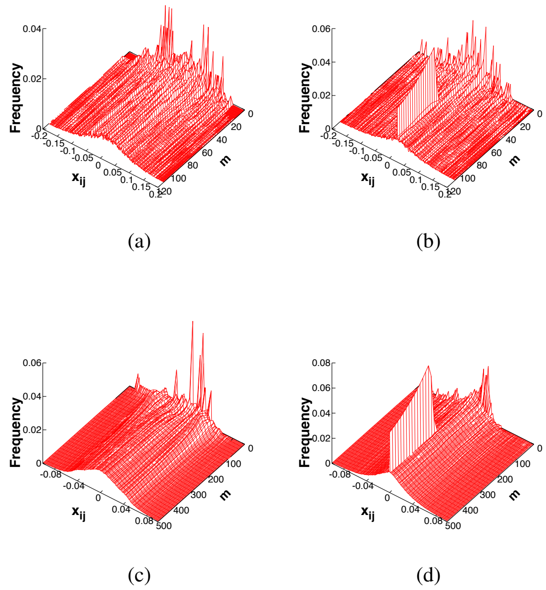


Figure 6: Distributions of the coordinate values in different amount of eigenvectors. (a) the network of American football games,(b) the adjacency network of common adjectives and nouns in the novel of David Copperfield by Charles Dickens, (c) a small-world network of WS model, (d) a BA model.

24650116) from JSPS.

References

[1] Barabási, Albert and Jeong, *Physica A*, **281**, 69-99, 2000.

[2] Travers and Milgram, *An experimental study of small world problem.*, Sociometry, 1969.

[3] Watts and Strogatz, *Nature*, **393**, 440-442, 1998.

[4] Sen, Dasgupta, Chatterjee, Sreeram, Mukherjee and Manna *Phys. Rev. E*, **67**, 3, 2003.

[5] Barabási and Albert, *Science*, **286**, 509-512, 1999.

[6] Shimada, Ikeguchi and Shigehara, *Phys. Rev. Lett*, **109**, 158701, 2012.

[7] Shimada, and Ikeguchi, *Proc. IEICE Soc. Conf.*, A-2-6, 2011.

[8] Newman, *Phys. Rev. E*, **74**, 036104, 2006.

[9] Girvan and Newman, *Proc. Natl. Acad. Sci. USA*, **99**, 7821-7826, 2002.

[10] Gleiser and Danon, *Adv. Complex Syst.*, **6** 565, 2003.

[11] Vladimir and Andrej, *Pajek datasets*, URL: http://vlado.fmf.uni-lj.si/pub/networks/data/.

[12] T. F. Cox, and M. A. A. Cox, *Multidimensional Scaling*, Chapman&Hall/CRC, 2000.