IEICE Proceeding Series

Detecting growth of groups in networks by using a transformation method from networks to time series

Yutaka Shimada, Yoshito Hirata, Tohru Ikeguchi, Kazuyuki Aihara

Vol. 2 pp. 231-234
Publication Date: 2014/03/18
Online ISSN: 2188-5079

Detecting growth of groups in networks by using a transformation method from networks to time series

Yutaka Shimada[1,2], Yoshito Hirata[3], Tohru Ikeguchi[4,5], and Kazuyuki Aihara[3]

1. FIRST, Aihara Innovative Mathematical Modelling Project, JST
2. Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan
3. Institute of Industrial Science, The University of Tokyo
4. Graduate School of Science and Engineering, Saitama University 255 Shimoohkubo, Sakura-ku, Saitama-shi, Saitama, 338-8570 Japan
5. Saitama University, Brain Science Institute

Abstract—A novel framework to analyze nonlinear time series and complex networks has recently been proposed. In the framework, nonlinear time series are transformed to networks and analyzed through the transformed networks. On the other hand, the networks are also transformed to time series and analyzed through the obtained time series. In this paper, we show that the transformation method from the networks to the time series can be applied to analyze the evolution of sizes of communities in the networks. We apply the transformation method to an evolving network model and show that the method can detect the temporal change of the evolution of the citation network.

1. Introduction

A network, or a graph, is one of effective tools to describe various real systems. To analyze real networks, several measures that quantify their structural properties have been proposed, and thereby one can obtain structural and dynamical properties of the real systems (see for example [1]). Recent researches on nonlinear time series analysis incorporate the measures of the networks and build a novel framework to analyze nonlinear time series [2]. In the framework, the time series are firstly transformed into networks by several methods [3–8]. These methods enable us to analyze the networks transformed from the time series by the measures of the networks. The properties of the transformed networks characterize their original time series.

On the other hand, we have proposed an opposite method for transforming the networks into the time series to analyze the networks through their time series [9]. In the method, the structures of the networks are embedded in their time series: lattices correspond to periodic time series, small world networks to noisy periodic time series, and random networks to random time series. We can quantify the structural properties of the networks through the obtained time series.

In this paper, we focus on the transformation method from the networks to time series [9]. We then find that the method closely relates to one of community detection methods, and that the time series obtained from the networks include the information about communities in the network. Applying the transformation method to an evolving network model, we show that the transformation method can detect communities. By combining the temporal information and the information about the detected communities, the temporal growth of the sizes of the communities is evaluated.

2. Method for transforming networks to time series

In Ref. [9], we proposed a method for transforming networks to time series. The method has three steps: (i) defining a distance between nodes, (ii) arranging the nodes in the Euclidean space by the classical multidimensional scaling (CMDS) such that the given distance relations hold, and (iii) tracking the obtained coordinate vectors of the nodes in the order of time when each node is added to the network. The distance d_{ij} is defined on the basis of the adjacency relation between nodes i and j. If two nodes connect with each other, $d_{ij} = 1$, otherwise $d_{ij} = w\ (> 1)$.

Let an $N \times N$ symmetric matrix $A = (a_{ij})$ be an adjacency matrix of the undirected network in which $a_{ij} = 1$ if the node i connects to j by an link, otherwise $a_{ij} = 0$. The CMDS arranges the nodes on the Euclidean space by using the following relationship between the innerproduct and the Euclidean distance [10]

$$\boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{x}_j = -\frac{1}{2} \left[d_{ij}^2 + \sum_{r=1}^{N} \sum_{s=1}^{N} \frac{d_{rs}^2}{N^2} - \sum_{r=1}^{N} \frac{d_{rj}^2}{N} - \sum_{s=1}^{N} \frac{d_{is}^2}{N} \right], \tag{1}$$

where d_{ij} is the Euclidean distance and $\boldsymbol{x}_m = (x_{m1}, x_{m2}, \ldots, x_{mh})^{\mathsf{T}}$ is an h-dimensional coordinate vector of the mth node. By this relation, one can obtain coordinate vectors only from the distances between the nodes by the following equation:

$$X^{\mathsf{T}}X = (\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{x}_j) = -\frac{1}{2}JD^{(2)}J, \quad (2)$$

where $D^{(2)} = \left(d_{ij}^2\right)$, $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)^{\mathsf{T}}$, $J = E - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathsf{T}}$, $\mathbf{1}$ is a column vector with N ones, and E is an $N \times N$ unit matrix. By using the eigenvalues and the eigenvectors of the matrix $G \equiv -\frac{1}{2}JD^{(2)}J$, we can estimate the matrix X as follows

$$G = V\Lambda V^{\mathsf{T}} = (\Lambda^{(1/2)}V^{\mathsf{T}})^{\mathsf{T}}(\Lambda^{(1/2)}V^{\mathsf{T}}) = X^{\mathsf{T}}X. \quad (3)$$

where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_h)$, λ_i ($\lambda_1 \geq \cdots \geq \lambda_h > 0$) is the ith positive eigenvalue of G, $\boldsymbol{v}_i$ is the ith eigenvector corresponding to λ_i, and $V = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_h)^{\mathsf{T}}$. Assuming that the node t is added to the network at time t, we can obtain the ith time series $y_i(t)$ defined by

$$y_i(t) = \sqrt{\lambda_i}v_{ti}. \quad (4)$$

3. Relationship between time series and communities

Time series obtained from a network by the transformation method give us a certain division of the network. In the following, we present the relationship between the transformation method and one of community detection methods that uses eigenvectors of a matrix which is called a modularity matrix.

3.1. Community detection method

A method based on the modularity is one of effective methods to find groups, or communities, in a network. The basic idea of the modularity-based community detection method is to find a particular division of a given network by comparing the number of links included in each group of the given network with that of random networks. Because the random networks usually have no groups, namely densely connected subgraphs, if the given network has M groups, the total sum of the differences Q between the number of links in the groups of the given network and that of the random networks (null model) takes large value. The sum Q is called the modularity and described by

$$Q = \sum_{r=1}^{M}\sum_{i=1}^{N}\sum_{j=1}^{N}(a_{ij} - P_{ij})u_{ir}u_{jr} = \mathrm{Tr}(U^{\mathsf{T}}BU), \quad (5)$$

where the null model P_{ij} be the expected number of links between nodes i and j in a random network, and

$$u_{ij} = \begin{cases} 1 & \text{if the group } j \text{ includes the node } i, \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

In Eq. (5), $\mathrm{Tr}(U^{\mathsf{T}}BU)$ shows the matrix form of Q, where $B = (b_{ij})$, $b_{ij} \equiv a_{ij} - P_{ij}$, and the $N \times M$ matrix $U = (u_{ij})$. The matrix B is called the modularity matrix.

Indeed, the null model P_{ij} is arbitrary, but P_{ij} needs to satisfy the conditions $\sum_{j=1}^{N} P_{ij} = k_i$ and $\sum_{i=1}^{N}\sum_{j=1}^{N} P_{ij} = 2L$, where L is the total number of links in the network. When the null model P_{ij} is given, we can obtain groups in the network by the maximization of Q. Because B is a real symmetric matrix, the eigenvalues and their eigenvectors can be calculated. By the eigenvalues and the eigenvectors, Eq. (5) is rewritten by

$$Q = \sum_{i=1}^{N}\sum_{j=1}^{M}\lambda_i(\boldsymbol{v}_i^{\mathsf{T}}\boldsymbol{u}_j)^2 = \sum_{i=1}^{N}\lambda_i(\beta_i^j)^2, \quad (7)$$

where $\boldsymbol{u}_j$ is written as a linear combination of the normalized eigenvectors $\boldsymbol{u}_j = \sum_i \beta_i^j \boldsymbol{v}_i$ in the second equality. Equation (7) implies that Q can be maximized by deciding the value of u_{ij} such that if $v_{ij} \geq 0$, $u_{ij} = 1$, but if $v_{ij} < 0$, $u_{ij} = 0$. In other words, $\boldsymbol{u}_i$ should be decided as parallel to the eigenvector with the ith largest eigenvalue as possible.

3.2. How time series relate to the community detection

To clarify the relationship between our method and the community detection method, we redefine the distance d_{ij}: if the node i connects to j, $d_{ij}^2 = \alpha - 1$, otherwise $d_{ij}^2 = \alpha$ (> 1). This distance implies that the distances between connected nodes are shorter than those between unconnected nodes, and the distance takes only three values, zero, $\sqrt{\alpha}$, and $\sqrt{\alpha - 1}$. In this sense, the redefined distance is essentially the same as the distance in [9]. The distance d_{ij} has the following another form

$$d_{ij}^2 = (1 - \delta_{ij})(\alpha - a_{ij}), \quad (8)$$

where if $i = j$, $\delta_{ij} = 1$, otherwise zero. Substituting Eq. (8) into Eq. (1), we obtain the element of G, g_{ij}, as follows

$$\begin{aligned} g_{ij} &= -\frac{1}{2}\Bigg\{(1 - \delta_{ij})(\alpha - a_{ij}) \\ &\quad + \frac{1}{N^2}\sum_{r=1}^{N}\sum_{s=1}^{N}(1 - \delta_{rs})(\alpha - a_{rs}) \\ &\quad - \frac{1}{N}\sum_{r=1}^{N}(1 - \delta_{rj})(\alpha - a_{rj}) \\ &\quad - \frac{1}{N}\sum_{s=1}^{N}(1 - \delta_{is})(\alpha - a_{is})\Bigg\}, \\ &= \frac{1}{2}a_{ij} - \frac{1}{2N}\left\{k_i + k_j - N\alpha\delta_{ij} - \frac{2L}{N} + \alpha\right\} \\ &= \frac{1}{2}\left(a_{ij} - P'_{ij}\right), \end{aligned} \quad (9)$$

where

$$P'_{ij} \equiv \frac{1}{N}\left\{k_i + k_j - N\alpha\delta_{ij} - \frac{2L}{N} + \alpha\right\}. \quad (10)$$

In Eq. (10), P'_{ij} satisfies the conditions, $\sum_{j=1}^{N} P'_{ij} = k_i$ and $\sum_{i=1}^{N}\sum_{j=1}^{N} P'_{ij} = 2L$. This relation implies that P'_{ij} can be used as a null model in the modularity, and we define the modularity Q' as $\mathrm{Tr}(U^{\intercal}GU)$. As shown in Eq. (4), the obtained time series are proportionally equal to eigenvectors, and thus the time series corresponding to the largest eigenvalues can partition a given network such that the modularity Q' is maximized.

4. Results

To confirm the relation between time series generated from networks and community structures, we investigate networks generated from the model proposed by Klemm and Eguíluz (KE model) [11]. The KE model is realized by the following algorithm:

1. Starting from a complete graph with m_0 nodes at time $t = 0$. The states of these nodes are active.
2. Increasing time t by unity ($t \leftarrow t + 1$).
3. At time t, the $(m_0 + t)$th node with m_0 links is newly added to the network. The state of the new added node is active.
4. The new added node connects to m_0 active nodes with probability $1 - \mu$ and randomly connects to m_0 nodes with probability μ by the preferential attachment rule in which each link of the new node is connected to the pre-existing node i with probability $k_i / \sum_{j=1}^{m_0+t-1} k_j$.
5. One of active nodes is deactivated. The ith node is deactivated with the probability $k_i^{-1} / \sum_{j\in I} k_j^{-1}$, where I is a set which includes indices of all active nodes.
6. Repeating the steps 2–5 until time $T(= N - m_0)$.

In this model, the indices of nodes directly indicate the order of time when each node is added to the network. By using this information, we transform the networks into time series by Eqs. (2)–(4).

In the KE model, old nodes are not so easy to connect with young nodes because of the deactivation, and the nodes whose indices are similar are likely to connect with each other. This process of the deactivation involves groups of densely connected nodes, and hence the networks generated from the KE model include some groups. In the following numerical simulations, we show that our method can detect these groups.

Figure 1 shows a typical example of an adjacency matrix of the network generated from the KE model, which the parameters are set to $N = 4,000$, $m_0 = 10$, and $\mu = 0.2$. This network has typical three hub-nodes with over 500 degree, the 139th node (A), the 1159th node (B), and the 2301th node (C). We apply our method ($\alpha = 2$) to this network and then obtain time series.

From Fig.2(a), the first time series $s_1(t)$ divides the network into two groups. One group consists of the node

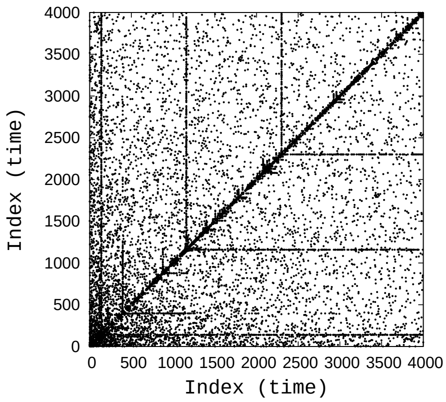

Figure 1: Adjacency matrix of the network generated from the KE model as a typical example.

i which satisfies $s_1(i) > 0$, and this group mainly consists of two types of nodes: nodes connecting A and nodes not connecting A, B, and C [purple rectangles in Fig.2(a)]. By using $s_2(t)$, we can further distinguish these types of nodes. In Fig.2(a), 98% of nodes which connect to A [blue circles in Fig.2(a)] are included in the region that satisfies $s_1(t) < 0$ and $s_2(t) < 0$, while 80% of nodes which connect to A, B, and C [purple rectangles in Fig.2(a)] are included in the region that satisfies $s_1(t) < 0$ and $s_2(t) > 0$.

On the other hand, most nodes which satisfy $s_1(t) < 0$ connect the hub-nodes B and C. In this case, although we cannot distinguish two types of nodes, namely nodes connecting B and nodes connecting C, only from $s_1(t)$ and $s_2(t)$, $s_3(t)$ or $s_4(t)$ can divide them [Fig.2(b)]. In Fig.2(b), most nodes which satisfy $s_3(t) > 0$ (or $s_4(t) > 0$) connect to C, but most nodes which satisfy $s_3(t) < 0$ (or $s_4(t) < 0$) do not connect to C. From these results, we can obtain the four sets of indices of nodes: the group-1 $G_1 = \{i|s_1(i) > 0 \text{ and } s_2(i) > 0\}$ [purple rectangles in Fig.2(a)], the group-2 $G_2 = \{i|s_1(i) > 0 \text{ and } s_2(i) < 0\}$ [blue circles in Fig.2(a)], the group-3 $G_3 = \{i|s_1(i) < 0 \text{ and } s_3(i) < 0\}$ [red circles in Fig.2(b)], and the group-4 $G_4 = \{i|s_1(i) < 0 \text{ and } s_3(i) > 0\}$ [yellow rectangles in Fig.2(b)].

Combining the temporal information of the time series and the information of communities, we can track the temporal evolution of the size of each group. From Eq. (4), the time series obtained from the network have the temporal information about the order of time when nodes are added to the network. From this information, the number of nodes included in the group j at time t is calculated by the following equation

$$N_j(t) = \sum_{i\in G_j} H(t - i), \quad (11)$$

where the function $H(x)$ is the Heaviside step function in which if $x < 0$, $H(x) = 0$, but if $x > 0$, $H(x) = 1$.

Figure 3 shows temporal changes of the size of the groups. After that the hub-node A appears, the size of the

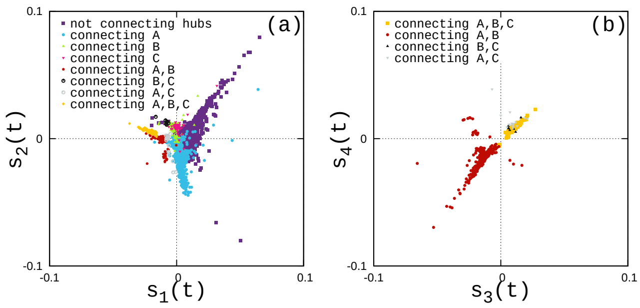

Figure 2: Divisions by (a) the fist time series $s_1(t)$ and the second time series $s_2(t)$ and (b) the third time series $s_3(t)$ and the fourth time series $s_4(t)$.

group-2 rapidly increases. However, the growth of the size of the group-2 becomes slow after that the hub-node B appears, while the size of the group-4 quickly increases. In the same way, the growth of the size of the group-4 becomes slow after that the hub-node C appears. On the other hand, the size of the group-3 quickly increases. From these results, by using the time series from the network, we can effectively detect the evolution of the sizes of the citation network. In the above numerical simulations, we show the results for a typical network generated from the KE model. Even if we apply our method to other networks generated from the KE model with several parameters, our method can work well. Here, if μ is close to unity, the KE model is essentially the same as the BA model, and the networks generated from the KE model are almost random. In this case, groups do not usually emerge and thus our method does not naturally detect the groups.

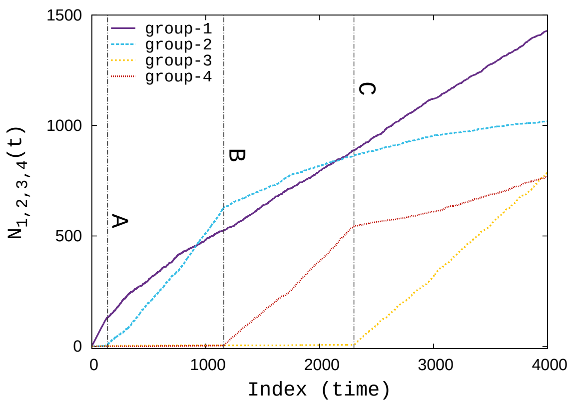

Figure 3: The cumulative number of nodes in each group at time t. The first vertical dashed line shows the appearance time of the paper A, the second vertical dashed line shows the appearance time of the paper B, and the third vertical dashed line shows the appearance time of the paper C,

5. Conclusion

In this paper, we focused on the transformation method from networks to times series and showed that the transformation method is interpreted as one of community detection methods. We can then divide the networks into some groups by using the time series as well as a community detection method that uses the eigenvectors of the modularity matrix. Utilizing two information in the transformed time series, namely the community information and the temporal information, we proposed a method to detect the evolution of sizes of groups in the networks. To confirm the validity of the method, we analyzed the network generated from the KE model. We then showed that our method can effectively detect the large groups in the network and can track the temporal evolution of the sizes of the groups. As a future work, we will analyze the differences between traditional community detection methods and our method in detail.

This research is supported by the Aihara Innovative Mathematical Modelling Project, the Japan Society for the Promotion of Science (JSPS) through the "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)," initiated by the Council for Science and Technology Policy (CSTP). The research of T.I. is supported by Grant-in-Aid for Exploratory Research (No. 24650116) from JSPS.

References

[1] Strogatz S.H. (2001) *Nature* 410:268–276; Albert R., Barabási A.-L. (2002) *Rev. Mod. Phys.* 74:47–97; Newman M.E.J. (2003) *SIAM Rev. Soc. Ind. Appl. Math* 45:167–256.

[2] Donner R. et al. (2010) *New J. Phys.* 12:033025.

[3] Zhang J. and Small M. (2006) *Phys. Rev. Lett.* 96:238701.

[4] Lacasa L. et al., (2008) *Proc. Nat. Acad. Sci. USA* 105:4972–4975.

[5] Shimada Y., Kimura T., and Ikeguchi T. (2008) *Lect. Notes Comp. Sci.* 5163:61-70.

[6] Shimada Y., and Ikeguchi T. (2010) *Phys. Lett. A* 374:3170–3176.

[7] Xu X., Zhang J., and Small M. (2008) *Proc. Nat. Acad. Sci. USA* 105:19601–19605.

[8] Marwan N. et al. (2009) *Phys. Lett. A* 373:4246–4254.

[9] Shimada Y., Ikeguchi T., and Shigehara T. (2012) *Phys. Rev. Lett.* 109:158701.

[10] Cox T.F. and Cox. M.A.A. (2000) *Multidimensional Scaling*, Chapman & Hall/CRC, Boca Raton, Florida.

[11] Klemm K. and Eguíluz V. (2002) *Phys. Rev. E* 65:057102.

[12] Newman M.E.J. (2006) *Phys. Rev. E* 74:036104.