

SCIENTIFIC REPORTS



OPEN

Graph distance for complex networks

Yutaka Shimada¹, Yoshito Hirata², Tohru Ikeguchi¹ & Kazuyuki Aihara²

Received: 01 June 2016
 Accepted: 20 September 2016
 Published: 11 October 2016

Networks are widely used as a tool for describing diverse real complex systems and have been successfully applied to many fields. The distance between networks is one of the most fundamental concepts for properly classifying real networks, detecting temporal changes in network structures, and effectively predicting their temporal evolution. However, this distance has rarely been discussed in the theory of complex networks. Here, we propose a graph distance between networks based on a Laplacian matrix that reflects the structural and dynamical properties of networked dynamical systems. Our results indicate that the Laplacian-based graph distance effectively quantifies the structural difference between complex networks. We further show that our approach successfully elucidates the temporal properties underlying temporal networks observed in the context of face-to-face human interactions.

Describing real complex systems as coupled dynamical systems is a basic and powerful approach for enhancing our understanding of their dynamical properties. If we can appropriately evaluate the distance between graphs or networks, this distance can be used as a fundamental and effective tool to characterise differences in the dynamical properties of two different coupled dynamical systems, because couplings or network structures affect collective behaviour of the coupled dynamical systems including convergence of consensus dynamics^{1–3}, controllability of coupled linear systems^{4–8}, and basin stability of coupled, especially nonlinear, dynamical systems^{9,10}.

Although several definitions of the distance between networks have been proposed¹¹, they all simply focus on the differences in the number of nodes and links. For example, the Hamming distance is the sum of the simple differences between the adjacency matrices of two graphs¹¹, and the graph edit distance is the minimum cost for transforming one graph into another by deletions and insertions of nodes and links¹². These distances simply focus on the differences in the number of nodes and links, and lack the perspective of both dynamical and structural properties of complex networks. Here, we propose a graph distance for complex networks that reflects both their structural and dynamical properties, focusing on the Laplacian matrix that represents the coupling topology. We define a graph distance between networks through comparing their Laplacian matrices. Because the Laplacian matrix contains significant information about the structural and dynamical properties of networks, the Laplacian-based graph distance directly reflects the differences in both the structural and dynamical properties characterised by the Laplacian matrix. We show that the Laplacian-based graph distance is effectively applied to mathematical models and several real networks, even in a situation where their sizes differ from one another. We further apply the proposed distance to a real temporal network whose structure changes with time^{13,14}. Our results indicate that the distance for complex networks is a key concept for understanding the dynamical properties of temporal networks.

Graph Distance for Complex Networks

Let A be an $N \times N$ adjacency matrix of a given undirected network in which $A_{ij} = 1$ if the node i connects to the node j and $A_{ij} = 0$ otherwise. The Laplacian matrix associated with the adjacency matrix A is defined by

$$L = A - K, \quad (1)$$

where K is the diagonal matrix, whose i th diagonal element is the degree of the i th node k_i . The Laplacian matrix appears in several contexts of network science, such as detecting community structures in the networks^{15,16}, consensus dynamics¹, and describing diffusive coupling of nonlinear dynamical systems^{9,10}. For example, a network consisting of connected N identical dynamical systems is governed by

¹Faculty of Engineering, Tokyo University of Science, 6-3-1 Nijjuku, Katsushika-ku, Tokyo 125-8585, Japan. ²Institute of Industrial Science, the University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan. Correspondence and requests for materials should be addressed to Y.S. (email: yshimada@rs.tus.ac.jp)

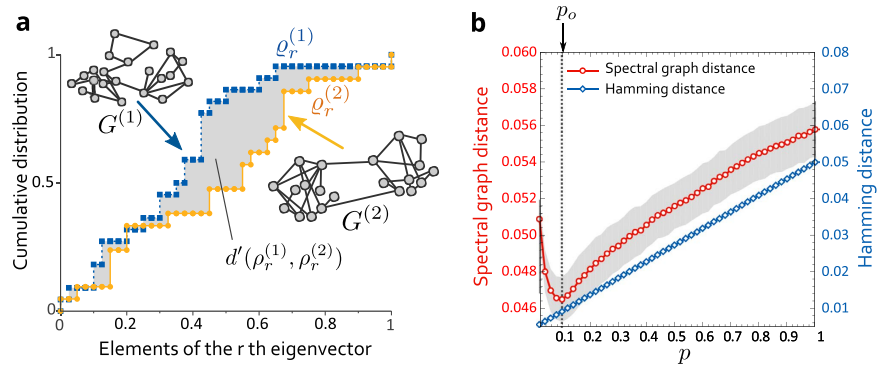


Figure 1. (a) Schematic diagram of the spectral graph distance described by Eqs (3) and (5). The distributions $\rho_r^{(1)}$ and $\rho_r^{(2)}$ are obtained from the r th eigenvectors of the Laplacian matrix $L^{(1)}$ corresponding to the network $G^{(1)}$ and $L^{(2)}$ corresponding to $G^{(2)}$. (b) The spectral graph distances and the Hamming distances between the networks $G(p_o)$ and $G(p)$ generated from the WS model, where the value of p is varied and p_o takes the fixed constant value ($p_o = 0.1$). The number of nodes $N = 500$, and each node has on average $k = 10$ links. The grey area indicates \pm standard deviation.

$$\dot{\mathbf{x}}_i = \mathbf{F}(\mathbf{x}_i) - \varepsilon \sum_{j=1}^N J_{ij} \mathbf{H}(\mathbf{x}_j), \tag{2}$$

where \mathbf{x}_i is the multidimensional state vector of the i th dynamical system, ε is the coupling strength, \mathbf{F} determines the intrinsic dynamics of each isolated uncoupled dynamical system, J_{ij} is a coupling coefficient which satisfies $\sum_{j=1}^N J_{ij} = 0$, and \mathbf{H} is the coupling function that defines the dynamics of the links. Equation (2) covers several coupling forms and indicates that the Laplacian matrix is crucial for estimating the stability of the various coupled dynamical systems⁹. In this paper, even though we mainly focus on the situation where the coupling matrix $J = \{J_{ij}\}$ is just the Laplacian matrix L , our approach can be naturally extended to other types of the coupling matrix J .

On the other hand, the detection problem of communities in networks is formulated as one of discrete optimisation problems. A well-known technique for detecting communities is spectral partitioning based on the direct relation between the structural properties of the network and the spectrum of the Laplacian matrix¹⁶. In this sense, the Laplacian matrix reflects both structural and dynamical properties of the networked dynamical systems.

In our approach utilising the spectrum of the Laplacian matrix, we propose a graph distance between two networks by calculating the distance between the spectra of the corresponding Laplacian matrices. Let $G^{(i)}$ be the i th network with $N^{(i)}$ nodes and $L^{(i)}$ be its Laplacian matrix. The Laplacian matrix $L^{(i)}$ is rewritten by $L^{(i)} = V^{(i)} \Lambda^{(i)} (V^{(i)})^T$, where $\Lambda^{(i)}$ is the diagonal matrix $\text{diag}(\lambda_1^{(i)}, \dots, \lambda_{N^{(i)}}^{(i)})$, $\lambda_r^{(i)}$ is the r th eigenvalue ($\lambda_1^{(i)} = 0 \leq \lambda_2^{(i)} \leq \dots \leq \lambda_{N^{(i)}}^{(i)}$), $V^{(i)} = (\mathbf{v}_1^{(i)}, \dots, \mathbf{v}_{N^{(i)}}^{(i)})$ and $\mathbf{v}_r^{(i)}$ is the eigenvector corresponding to the r th eigenvalue of $G^{(i)}$. The Laplacian matrix is a semi-definite matrix, whose smallest eigenvalue, λ_1 , is zero and the eigenvector corresponding to λ_1 is $\mathbf{1}$ which is a column vector with N ones. Let $\rho_r^{(i)}$ be the distribution of the elements in the r th eigenvector of the i th network. Given two networks $G^{(i)}$ and $G^{(j)}$, calculating the distance between pairs of the distributions $\rho_r^{(i)}$ and $\rho_r^{(j)}$, we define the distance between the two networks as

$$d(G^{(i)}, G^{(j)}) \equiv \frac{1}{M_{ij} - 1} \sum_{r=2}^{M_{ij}} d'(\rho_r^{(i)}, \rho_r^{(j)}), \tag{3}$$

where $d'(\rho_r^{(i)}, \rho_r^{(j)})$ is the distance between $\rho_r^{(i)}$ and $\rho_r^{(j)}$, and $1 < M_{ij} \leq \min(N^{(i)}, N^{(j)})$. Throughout this paper, $M_{ij} = \min(N^{(i)}, N^{(j)})$. To calculate $d'(\rho_r^{(i)}, \rho_r^{(j)})$, we introduce the Kruglov distance with the absolute value function from the viewpoint of computational convenience (see Methods and Supplementary Information). On the right-hand side of Eq. (3), $d'(\rho_1^{(i)}, \rho_1^{(j)})$ is not included in the summation because both eigenvectors corresponding to zero eigenvalues are $\mathbf{v}_1^{(i)} = \mathbf{v}_1^{(j)} = \mathbf{1}$, and do not give information concerning differences between two networks. Figure 1a shows the schematic diagram of Eq. (3). We call the distance defined by Eq. (3) the *spectral graph distance* (See Supplementary Information for detail of properties of the spectral graph distance).

Results and Discussion

Classifying network structures.

To evaluate how the spectral graph distance performs, we first apply our method to the Watts–Strogatz (WS) model, which can generate small-world and random networks by rewiring edges in the initial ring lattices with a probability p ¹⁷. If $p = 0$, the WS model generates ring lattices, but if $p = 1$, the WS model generates random networks. In the experiments, a network $G(p_o)$ is generated from the WS model with a fixed rewiring probability, p_o , in advance. We then calculate distances between $G(p_o)$ and networks $G(p)$, which are generated from the WS model with several values of the rewiring probability, p . If the proposed distance

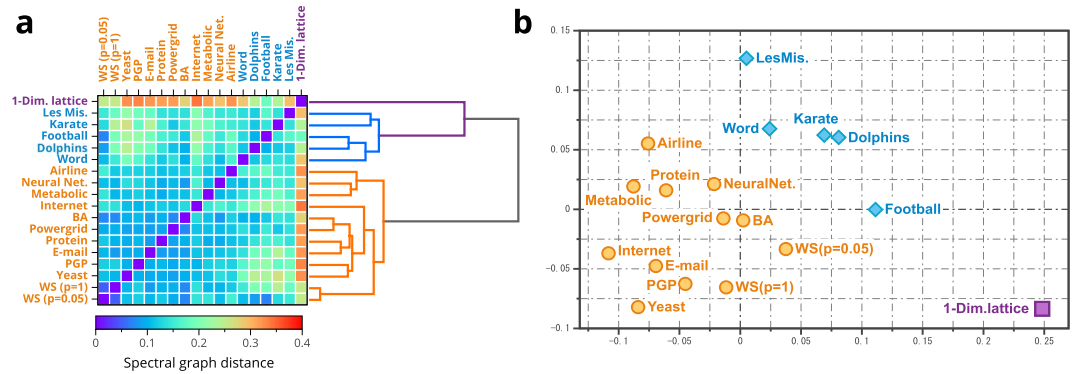


Figure 2. (a) Spectral graph distances between real networks and networks obtained from mathematical models. The figure shows the distance matrix, with colours representing the spectral graph distances. The tree shown in Fig. 2a is the hierarchical clustering tree calculated by the classical hierarchical clustering method on the basis of the distance matrix. (b) Two-dimensional visualization of networks obtained by multidimensional scaling (MDS)¹⁸. See Methods and also Supplementary Information for further details.

appropriately evaluates the distance between $G(p_o)$ and $G(p)$, $d(G(p_o), G(p))$ should take a minimal value when $p = p_o$.

Figure 1b shows the results of the distances between $G(p_o)$ and $G(p)$. Here, the spectral graph distance takes its minimum value when $p_o \approx p$. On the other hand, the Hamming distance does not take its minimum value even when the value of p is close to that of p_o . This result clearly shows that our distance effectively evaluates the difference between the two networks, whereas a simple distance focusing on the number of nodes and links cannot appropriately reflect the differences between the networks' structural properties.

Next, we apply our method to real networks. Figure 2a shows the distance matrix in which the real networks as well as mathematical models are listed on the vertical and horizontal axes, and colours show their inter-network distances. In Fig. 2b, each network is arranged in the two-dimensional Euclidean space such that their spectral graph distances are preserved as much as possible using multidimensional scaling (see for example ref. 18 and Supplementary Information). The hierarchical tree described on the right-hand side of Fig. 2a is obtained by the classical hierarchical clustering method (see also Methods).

As Fig. 2a,b shows, the real networks are mainly classified on the basis of their spectral graph distances into two groups: one consists of biological and technological networks, and the other consists mainly of social networks, including the dolphin social network, the network of words and the network of football games. These results agree with the previous research¹⁹, which discussed why the social networks differ from other types of real networks. On the other hand, the one-dimensional lattice is relatively far from other networks because its structure is regular and different from those of many real biological, technological and social networks.

Analyzing temporal properties in face-to-face human interactions. We next apply our method to temporal networks. One example is a contact network in which face-to-face interactions between human individuals are recorded using radio frequency identifiers^{20,21}. The temporal network is described as a set of networks $\mathcal{G} = \{G^{(1)}, \dots, G^{(T)}\}$, where $G^{(t)}$ is a set of nodes and links that are observed within a certain period of time from $t \times \Delta t$ to $(t+1) \times \Delta t$. The index, t , of $G^{(t)}$ corresponds to the discrete temporal index. In real contact networks, individuals are likely to have many contacts in the daytime, but few during the night. If we simply focus on temporal changes in the number of links and nodes, these temporal networks seemingly have a one-day cycle (see also Supplementary Information). However, even if the temporal changes in the number of nodes and links are periodic, this does not directly indicate that network structures of the temporal networks also change periodically with time. This periodicity in the structures of the temporal networks is considerably important, especially when we evaluate the dynamical properties of temporal networks, because the structural properties characterised by the Laplacian matrix directly affect the dynamical properties^{2,14,22}.

If the temporal change in the structures of the contact networks is periodic, there exists a non-zero positive constant τ^* such that the network structure of $G^{(t)}$ is equivalent to that of $G^{(t+\tau^*)}$ for all values of t . We distinguish this periodicity in the structures of the temporal networks from the periodicity simply observed from the temporal changes in the number of nodes and links, and then we call the former *structural periodicity*. The final purpose here is to evaluate the structural periodicity of these real temporal networks.

The spectral graph distance enables us to determine whether the temporal networks have structural periodicity, by evaluating temporal differences between the network, $G^{(t)}$, and its m -nearest neighbours for all values of t as follows:

$$\tau(t, t') \equiv |t - t'|, \quad G^{(t)} \in \mathcal{N}(G^{(t)}), \quad (4)$$

where $t \neq t'$ and $\mathcal{N}(G^{(t)})$ is a set of m -nearest neighbours of $G^{(t)}$ given by the spectral graph distance. If the contact network exhibits structural periodic behaviour, the values of $\tau(t, t')$ should be $c\tau^*$ ($c = 1, 2, \dots$), where τ^* is its period, but if the contact network exhibits random behaviour, $\tau(t, t')$ takes several values. We simply describe $\tau(t, t')$ as τ hereafter.

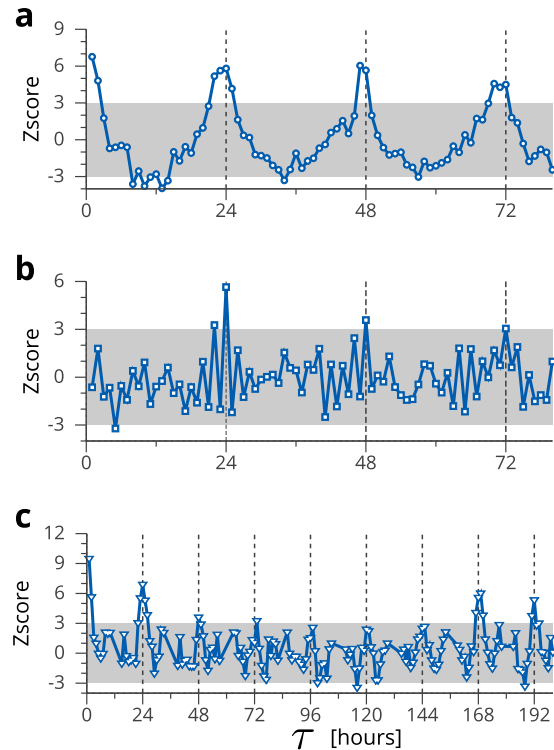


Figure 3. Z score results $[P(\tau) - \langle P(\tau_r) \rangle] / \sigma(\tau_r)$ for the hospital (a), high school (b) and gallery (c). The grey shading shows the area where $-3 < Z \text{ score} < 3$. The number of nearest neighbours, m , is set to 20% of the number of networks, T . The distances between the networks are calculated using the maximum connected component of $G^{(t)}$.

In the experiments, we apply the proposed method to three types of temporal networks of contacts between individuals observed in a hospital, a high school and a science gallery^{20,21,23}. Each contact network is divided into T smaller networks every 60 min, and the number of networks, T , in the dataset of the hospital data is 97, that of the high school is 202, and that of the science gallery is 1,929. Evaluating the value of Eq. (4) for the networks $G^{(t)}$ ($t = 1, \dots, T$), we first obtain the probability distribution of the value of τ , $P(\tau)$. Next, we calculate the probability distribution, $P(\tau_r)$, of the temporal difference, τ_r , between $G^{(t)}$ and $G^{(t_r)}$, which is randomly selected from \mathcal{G} . In the real contact networks, due to few contacts in the night time, the contact networks observed in the daytime are likely to be selected, even in the case of random selections, and thereby, the occurrence frequency of τ is biased. To avoid such a bias, we calculate the Z score $(P(\tau) - \langle P(\tau_r) \rangle) / \sigma(\tau_r)$, where $\langle P(\tau_r) \rangle$ is the expected value of $P(\tau_r)$ and $\sigma(\tau_r)$ is its standard deviation.

From Fig. 3, we find peaks at $\tau = 24, 48$ and 72 in the results of the contact networks observed in the hospital and the high school. This result indicates that these contact networks exhibit recurring 24-h patterns. These results agree with intuition; most nodes in the contact network observed in the hospital are staff members including doctors and nurses who work there and patients who are admitted. Many of them participate in the temporal network every day. In the ward, the staff members work according to their fixed schedule and the temporal networks observed in the hospital are likely to exhibit periodicity. In the case of the high school, the students also behave according to a certain fixed schedule. In such cases, the network structure of $G^{(t)}$ is likely to be similar to that of $G^{(t+c\tau^*)}$, where $\tau^* (> 0)$ is the period of the contact network. The spectral graph distance effectively elucidates the structural periodicity in the contact networks. On the other hand, we also find the peaks at $\tau = 24, 48, 72, 168$, and 192 in the results of the science gallery to be against intuition; most nodes in the contact networks observed in the science gallery are visitors and their network structures seem to be randomly varied every day. However, our results reveal that even if the visitors change day by day, the temporal networks exhibit recurring patterns.

In summary, we have proposed the spectral graph distance that can evaluate the distance between complex networks on the basis of their spectra of the Laplacian matrices. The spectral graph distance was successfully applied to the classification of several networks and to the analysis of the contact networks, and has elucidated the temporal property underlying real contact networks, namely that which we call structural periodicity. In particular, revealing temporal properties underlying temporal networks would involve several applications, such as time series analysis of networks^{24,25}, prediction and prevention of the spread of infectious diseases on temporal networks^{26–28}, and elucidation of the relationship between functions and network structures in many real systems^{29,30}. The spectral graph distance could be quite an important tool for answering these open questions, and carrying out more in-depth discussion on the proper distance between complex networks would give us a novel perspective on

the dynamical properties of temporal networks. In addition to these applications, the notion of the spectral distance between networks is important and useful when we assess the properties of networks, for example, quantum networks^{31,32}. The application of our distance to the quantum networks is one of the most important future works.

Methods

Distance between distributions. In Eq. (3), we need to calculate the distances between distributions $\rho_r^{(i)}$ and $\rho_r^{(j)}$, which are the distributions of elements in the r th eigenvectors of the Laplacian matrices corresponding to the i th and j th networks, $G^{(i)}$ and $G^{(j)}$. Let $\varrho_r^{(i)}$ be a cumulative distribution function of the elements in the r th eigenvector, $\mathbf{v}_r^{(i)}$, of the Laplacian matrix of the i th network. The cumulative distribution function is calculated by $\varrho_r^{(i)}(y) = \frac{1}{N^{(i)}} \sum_{l=1}^{N^{(i)}} H(y - v_r^{(i)}(l))$, where $H(y)$ is the step function in which $H(y)$ is unity if $y \geq 0$, and zero otherwise. Although there are several choices for the distance between distributions, we use the following:

$$d'(\rho_r^{(i)}, \rho_r^{(j)}) = \int_{-\infty}^{\infty} |\varrho_r^{(i)}(y) - \varrho_r^{(j)}(y)| dy. \quad (5)$$

To numerically estimate the distribution function $\rho_r^{(i)}$, one needs to determine a suitable bin size, but this selection is complicated because of a wide variety of distributions of elements in the eigenvectors obtained from complex networks. However, the cumulative distribution function, $\varrho_r^{(i)}$, is directly calculated from the elements of the eigenvector $\mathbf{v}_r^{(i)}$ without any parameter. In our method, the elements in the eigenvector are normalised such that their minimum and maximum values are zero and unity. Transforming each eigenvector into the distribution function of its elements enables us to compare two eigenvectors of different sizes.

Hierarchical clustering. We used a classical hierarchical clustering to classify several complex networks. We first calculated the spectral graph distances between the networks. In the initial step, the individual networks are considered as clusters, that is, each cluster includes only one network. Next, using the obtained spectral graph distances, we merged the two closest clusters until only one cluster remained on the basis of the classical Ward's method³³. This hierarchical clustering is visualised by using a tree-like diagram.

We calculated the spectral graph distance between real networks and then applied the classical hierarchical clustering. Here, the real networks are the adjacency network of adjectives and nouns in the *David Copperfield* (Word)¹⁶, a social network of 62 dolphins (Dolphin)³⁴, human relations between 34 members of the Zachary karate club (Karate)³⁵, a co-appearance network between 77 characters in the famous novel *Les Misérables* (LesMis)³⁶, the neural network of *C. elegans* (NeuralNet)¹⁷, the power grid (Powergrid)¹⁷, the *C. elegans* metabolic network (Metabolic)³⁷, the Internet (Internet)³⁸, the e-mail network (E-mail)³⁹, the network of users of the Pretty-Good-Privacy algorithm (PGP)⁴⁰, the protein interaction network of yeast (Yeast)⁴¹, and the airline network (Airline)^{42,43}. We also used the networks generated from the WS model¹⁷ and the Barabási–Albert (BA) model⁴⁴. In the WS model, the networks are generated with the rewiring probabilities $p = 0$ (1-Dim. Lattice), $p = 0.05$ (WS($p = 0.05$)) and $p = 1$ (WS($p = 1$)), and the initial one-dimensional lattice has $N = 5,000$ nodes, with the degree of each node being 10. In the BA model, a single new node with ℓ links is repeatedly added to the initial complete graph consisting of ℓ nodes. The network grows until the number of nodes reaches 5,000, and ℓ is set to three (BA). The texts in parentheses in the above list of real networks correspond to the labels of real networks in Fig. 2.

References

- Olfati-Saber, R., Fax, J. A. & Murray, R. M. Consensus and cooperation in networked multi-agent systems. *P. IEEE* **95**, 215–233 (2007).
- Baronchelli, A. & Daz-Guilera, A. Consensus in networks of mobile communicating agents. *Phys. Rev. E* **85**, 016113 (2012).
- Estrada, E. & Vargas-Estrada, E. How peer pressure shapes consensus, leadership, and innovations in social groups. *Sci. Rep.* **3**, 2905 (2013).
- Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
- Galbati, M., Delpini, D. & Battiston, S. The power to control. *Nat. Phys.* **9**, 126–128 (2013).
- Sun, J. & Motter, A. E. Controllability transition and nonlocality in network control. *Phys. Rev. Lett.* **110**, 208701 (2013).
- Jia, T. *et al.* Emergence of bimodality in controlling complex networks. *Nat. Commun.* **4**, 2002 (2013).
- Menichetti, G., Dall'Asta, L. & Bianconi, G. Network controllability is determined by the density of low in-degree and out-degree nodes. *Phys. Rev. Lett.* **113**, 078701 (2014).
- Pecora, L. M. & Carroll, T. L. Master stability functions for synchronized coupled systems. *Phys. Rev. Lett.* **80**, 2109–2112 (1998).
- Menck, P. J., Heitzig, J., Marwan, N. & Kurths, J. How basin stability complements the linear-stability paradigm. *Nat. Phys.* **9**, 89–92 (2013).
- Deza, M. M. & Deza, E. *Encyclopedia of distances* 2nd ed. (Springer-Verlag, Berlin Heidelberg, 2013).
- Gao, X., Xiao, B., Tao, D. & Li, X. A survey of graph edit distance. *Pattern Anal. Appl.* **13**, 113–129 (2010).
- Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125 (2012).
- Masuda, N., Klemm, K. & Eguluz, V. M. Temporal networks: slowing down diffusion by long lasting interactions. *Phys. Rev. Lett.* **111**, 188701 (2013).
- Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
- Newman, M. E. J. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**, 036104 (2006).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- Cox, T. & Cox, M. *Multidimensional scaling* (Chapman & Hall/CRC, Boca Raton, 2001).
- Newman, M. & Park, J. Why social networks are different from other types of networks. *Phys. Rev. E* **68**, 036122 (2003).
- Isella, L. *et al.* What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theor. Biol.* **271**, 166–180 (2011).
- Vanhems, P. *et al.* Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* **8**, e73970 (2013).
- Fujiwara, N., Kurths, J. & Daz-Guilera, A. Synchronization in networks of mobile oscillators. *Phys. Rev. E* **83**, 025101 (2011).
- Fournet, J. & Barrat, A. Contact patterns among high school students. *PLoS ONE* **9**, e107878 (2014).
- Shimada, Y., Ikeguchi, T. & Shigehara, T. From networks to time series. *Phys. Rev. Lett.* **109**, 158701 (2012).

25. Weng, T., Zhao, Y., Small, M. & Huang, D. D. Time-series analysis of networks: Exploring the structure with random walks. *Phys. Rev. E* **90**, 022804 (2014).
26. Holme, P. Epidemiologically optimal static networks from temporal network data. *PLoS Comput. Biol.* **9**, e1003142 (2013).
27. Holme, P. Information content of contact-pattern representations and predictability of epidemic outbreaks. *Sci. Rep.* **5**, 14462 (2015).
28. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
29. Iwayama, K. *et al.* Characterizing global evolutions of complex systems via intermediate network representations. *Sci. Rep.* **2**, 423 (2012).
30. Sporns, O. *Networks of the brain* (MIT Press, Cambridge, MA, 2010).
31. Paparo, G. D. & Martin-Delgado, M. A. Google in a Quantum Network. *Sci. Rep.* **2**, 444 (2012).
32. Paparo, G. D., Müller, M., Comellas, F. & Martin-Delgado, M. A. Quantum Google in a Complex Network. *Sci. Rep.* **3**, 2773 (2013).
33. Tan, P.-N., Steinbach, M. & Kumar, V. *Introduction to data mining* (Addison-Wesley, 2006).
34. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *P. Natl. Acad. Sci. USA* **99**, 7821–7826 (2002).
35. Zachary, W. W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **33**, 452–473 (1977).
36. Knuth, D. E. *The stanford graphBase: A platform for combinatorial computing* (Addison-Wesley, 1993).
37. Duch, J. & Arenas, A. Community detection in complex networks using extremal optimization. *Phys. Rev. E* **72**, 027104 (2005).
38. Newman, M. E. J. *Network data*. <http://www-personal.umich.edu/mejn/netdata/> (Date of access: 15th October 2013).
39. Guimerà, R., Danon, L., Daz-Guilera, A., Giralt, F. & Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **68**, 065103(R) (2003).
40. Boguná, M., Pastor-Satorras, R., Daz-Guilera, A. & Arenas, A. Models of social networks based on social distance attachment. *Phys. Rev. E* **70**, 056122 (2004).
41. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
42. Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M. & Morris, M. *Statnet: software tools for the statistical modeling of network data*. Seattle, WA. <http://statnetproject.org>. (Date of access: 11th March 2013) (2003).
43. Batagelj, V. & Mrvar, A. *Pajek datasets* <http://vlado.fmf.uni-lj.si/pub/networks/data/>. (Date of access: 15th October 2013) (2006).
44. Barabási, A. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

Acknowledgements

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 Y.S. is supported by Grant-in-Aid for Research Activity Start-up (No. 26880020) and Grant-in-Aid for Young Scientists B (No. 16K16126) from JSPS. The research of T.I. is partially supported by a Grant-in-Aid for Challenging Exploratory Research (No. 24650116) from JSPS. The research of T.I. and Y.S. is also supported by a Grant-in-Aid for Scientific Research (C) (Generative Research Fields) (No.15KT0112) from JSPS. The research of K.A. is also supported by CREST, JST and JSPS KAKENHI 15H05707.

Author Contributions

Y.S. conceived the study, performed the numerical simulations and prepared the manuscript. All authors discussed the results, drew conclusions and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Shimada, Y. *et al.* Graph distance for complex networks. *Sci. Rep.* **6**, 34944; doi: 10.1038/srep34944 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016