

IEICE **TRANSACTIONS**

on Communications

VOL. E102-B NO. 2
FEBRUARY 2019

The usage of this PDF file must comply with the IEICE Provisions on Copyright.

The author(s) can distribute this PDF file for research and educational (nonprofit) purposes only.

Distribution by anyone other than the author(s) is prohibited.

A PUBLICATION OF THE COMMUNICATIONS SOCIETY



The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Information Propagation Analysis of Social Network Using the Universality of Random Matrix

Yusuke SAKUMOTO^{†a)}, *Member*, Tsukasa KAMEYAMA^{††b)}, *Nonmember*, Chisa TAKANO^{†††c)},
and Masaki AIDA^{†d)}, *Members*

SUMMARY Spectral graph theory gives an algebraic approach to the analysis of the dynamics of a network by using the matrix that represents the network structure. However, it is not easy for social networks to apply the spectral graph theory because the matrix elements cannot be given exactly to represent the structure of a social network. The matrix element should be set on the basis of the relationship between persons, but the relationship cannot be quantified accurately from obtainable data (e.g., call history and chat history). To get around this problem, we utilize the universality of random matrices with the feature of social networks. As such a random matrix, we use the normalized Laplacian matrix for a network where link weights are randomly given. In this paper, we first clarify that the universality (i.e., the Wigner semicircle law) of the normalized Laplacian matrix appears in the eigenvalue frequency distribution regardless of the link weight distribution. Then, we analyze the information propagation speed by using the spectral graph theory and the universality of the normalized Laplacian matrix. As a result, we show that the worst-case speed of the information propagation changes up to twice if the structure (i.e., relationship among people) of a social network changes.

key words: social network, information propagation, random matrix, spectral graph theory, Wigner semicircle law, Laplacian matrix

1. Introduction

The emergence of social networking services (SNSs) and the widespread of mobile devices promote people interaction beyond anticipation in the society. As the results, the people interaction has strong capability to propagate the information submitted by someone to the whole society. In the recent years, it is important for the success of a new product and a new spot to propagate its information all over the social network through not only face-to-face offline conversation but also online communication via SNS (e.g., Twitter and Instagram). Therefore, the understanding of the information propagation property on social networks is essential to design marketing strategy of new products and new spots.

Spectral graph theory [1] is widely used to analyze the

properties of network dynamics by using the eigenvalues and the eigenvectors of a matrix (e.g., Laplacian matrix) that represents the structure of the network. However, when applying the spectral graph theory to social network analysis, two problems arise. First, social networks are huge. To analyze them using spectral graph theory, the eigenvalues and eigenvectors of the huge matrices must be calculated but this calculation is impossible because of its computational cost. Second, the relationship between persons in a social network is complex. It is hard to quantify the relationship accurately from obtainable data (e.g., call history and chat history). To represent the social network structure by a matrix, the matrix elements must be given exactly on the basis of the relationship but we need to perform the difficult task of quantifying the relationship. Therefore, before applying the spectral graph theory to the social network analysis, we should discuss the way around the above-mentioned problem.

A random matrix is a matrix whose elements are random variables; it is often utilized to analyze large-scale and complex structures in quantum mechanics [2], [3]. In quantum mechanics, there is a method to derive the electron orbital around an atomic nucleus by using a matrix that represents the atom structure. However, for a large atom (e.g., uranium) having many electrons with complex orbitals, it is impossible to give the matrix elements exactly. Hence, quantum mechanics gives up representing such large and complex atom structure exactly, and analyzes electron orbital property using the universality when the matrix elements are given by random variables. The analysis using the universality of the random matrix has had great success in quantum mechanics. The circumstance of the large and complex atom analysis in quantum mechanics is like the social network analysis, so the random matrix is expected to solve the fundamental problem in the social network analysis.

In this paper, we first investigate the universality of random matrices with the feature of social networks. As such a random matrix, we use the normalized Laplacian matrix for a network where link weights are randomly given. We clarify that the universality (i.e., the Wigner semicircle law) of the normalized Laplacian matrix appears in the eigenvalue frequency distribution regardless of the link weight distribution in random networks generated with the popular models (i.e., ER (Erdős–Rényi) model [4] and BA (Barabási–Albert) [5] model), which are also used in several studies [4]–[9]. Then, we analyze the information propagation speed in social net-

Manuscript received April 1, 2018.

Manuscript revised July 9, 2018.

Manuscript publicized August 17, 2018.

[†]The authors are with Tokyo Metropolitan University, Hino-shi, 191-0065 Japan.

^{††}The author is with Nochu Information System Co., Ltd, Tokyo, 135-0061 Japan.

^{†††}The author is with Hiroshima City University, Hiroshima-shi, 731-3166 Japan.

a) E-mail: sakumoto@tmu.ac.jp

b) E-mail: kameyama445@gmail.com

c) E-mail: takano@hiroshima-cu.ac.jp

d) E-mail: aida@tmu.ac.jp

DOI: 10.1587/transcom.2018EBP3098

works by using the clarified universality. In this analysis, we model the information propagation by a random walk in light of the resemblance between their characteristics. Since random walks are too slower than the information propagation in a social network, our analysis focuses on the worst-case situation. As a metric for the information propagation speed, we use the expected value of first arrival times of the random walker for each node. As the result of our analysis using spectral graph theory and the clarified universality, we show that the worst-case speed of the information propagation changes up to twice if the structure (i.e., relationship among people) of a social network changes.

This paper is organized as follows. In Sect. 2, we describe the normalized Laplacian matrix and the Wigner semicircle law as the preliminary of our discussion. In Sect. 3, we generate the random matrix with the social network feature, and investigate its universality. Section 4 analyzes the information propagation property using the universality clarified in Sect. 3. Finally, in Sect. 5, we conclude this paper and discuss future works.

2. Preliminary

2.1 Normalized Laplacian Matrix

We denote an undirected network with n nodes by $G(V, E)$ where V and E are the sets of nodes and links, respectively. Let $A = (A_{ij})_{1 \leq i, j \leq n}$ be the adjacency matrix, which represents the link structure of network G . A_{ij} is defined by

$$A_{ij} := \begin{cases} w_{ij} & ((i, j) \in E) \\ 0 & ((i, j) \notin E) \end{cases}, \quad (1)$$

where $w_{ij} > 0$ is the weight of link (i, j) . Since network G is undirected, adjacency matrix A is symmetric ($A_{ij} = A_{ji}$). Let $D = \text{diag}(d_i)_{1 \leq i \leq n}$ be the degree matrix where $d_i = \sum_{j=1}^n w_{ij}$ is the weighted degree of node i . Using adjacency matrix A and degree matrix D , Laplacian matrix L for network G is defined by

$$L := D - A. \quad (2)$$

Laplacian matrix L represents the node and link structure of network G .

Normalized Laplacian matrix N is also used to represent the network structure. N is defined by

$$N := D^{-1/2} L D^{-1/2}. \quad (3)$$

Since N is symmetric, N can be always diagonalized. Hence, N is also given by

$$N = P \Lambda P^{-1}, \quad (4)$$

where $\Lambda = \text{diag}(\lambda_l)_{1 \leq l \leq n}$ and $P = (\mathbf{q}_l)_{1 \leq l \leq n}$. λ_l and \mathbf{q}_l are l -th eigenvalue of N and the eigenvector for λ_l , respectively. In this paper, we arrange eigenvalue λ_l in ascending order (i.e., $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_n < 2$). Note that we assume $\lambda_2 \neq 0$. This means that G is connected. Namely,

there is a path between every pair of nodes in G . Since N is symmetric, eigenvector \mathbf{q}_l is the orthonormal basis. Namely, $\sum_{i=1}^n q_k(i) q_l(i) = \delta_{kl}$ where δ_{kl} is the Kronecker delta. According to Eq. (4), the combination of P and Λ also represent the network structure equivalent to N .

In Sect. 4, we discuss the relation between the random walk and the information propagation in a social network, and analyze the information propagation based on N .

2.2 Wigner Semicircle Law

The Wigner semicircle law [3] is the universality that appears in the eigenvalue density distribution of random matrices. Let $X = (X_{ij})_{i \leq j \leq n}$ be a $n \times n$ real symmetric matrix where X_{ij} is a random variable. X_{ij} for $j \geq i$ follows an independent identical distribution where all odd-order moments are zero and all even-order moments are finite amounts. We denote k -th eigenvalue for a sample of X by λ_k ($k = 1, \dots, n$). We look at the eigenvalue density $\rho_n(\lambda)$, which is given by

$$\rho_n(\lambda) = \frac{1}{n} \sum_{k=1}^n \delta\left(\lambda - \frac{\lambda_k}{\sqrt{n}}\right), \quad (5)$$

where $\delta(x)$ is the Dirac delta function of x . As the limit of $\rho_n(\lambda)$ with $n \rightarrow \infty$, $\rho_X(\lambda)$ follows by

$$\rho_X(\lambda) = \lim_{n \rightarrow \infty} \rho_n(\lambda) = \begin{cases} \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - \lambda^2} & (|\lambda| < 2\sqrt{\sigma^2}) \\ 0 & (\text{otherwise}) \end{cases}, \quad (6)$$

where σ is the standard deviation of the distribution of X_{ij} . Because of $\int_{-\infty}^{\infty} \rho_X(\lambda) d\lambda = 1$, $\rho_X(\lambda)$ is the density distribution of the eigenvalues. For matrix X , its eigenvalue density distribution satisfies the Wigner semicircle law if and only if it is given by Eq. (6).

3. Random Matrix with Social Network Feature

The existing study [10] clarified the universality of the well-known networks (ER network [4] and BA network [5]) with unweighted links. The clarified universality says that the eigenvalue frequency distribution of normalized Laplacian matrix N for an unweighted ER and BA network satisfies the Wigner semicircle law if the network fulfills the degree condition $k_{\min}^2 \gg k_{\text{ave}}$ where k_{\min} and k_{ave} are the minimum and average of degrees (numbers of links from a node), respectively.

To represent the structure of a social network, links should be weighted on the basis of the relationship between persons since people have diverse relationships in the social network. However, the relationships are complex, so we cannot give the link weights exactly. Hence, the universality for weighted networks should be needed for social network analysis.

In this section, we clarify the universality for weighted random networks to analyze the social network property. We

first generate normalized Laplacian matrix N for randomly-weighted random networks (ER network and BA network) as a random matrix, and then investigate the universality of N .

3.1 Generation Method of Random Matrix

As a random matrix with the feature of social networks, we use normalized Laplacian matrix N for randomly-weighted random networks (ER network and BA network) generated by the following steps. We denote the existing probability of links in the ER network by p . Then, let n_0 and n_a be the number of initial connected nodes and the number of adding links from a new node in the BA network, respectively.

1. Input expected average number k_{ave} of links of each node.
2. Generate an unweighted random network ($w_{ij} = 1$) based on the ER model or the BA model.
 - a. When we generate a ER network, we set existing probability p by $k_{ave}/(n - 1)$ so that each node has k_{ave} links in average.
 - b. When we generate a BA network, we set n_0 and n_a so that the BA network has about $n k_{ave}/(1 - q)$ links. After generating the BA network, we cut the links randomly with the probability q . Each node in the cut network has k_{ave} in average. We call the modified BA network as *cut BA network*.
3. Generate random values that follow a probability distribution (e.g., constant distribution, uniform distribution and exponential distribution), and set link weight w_{ij} for $i > j$ and $(i, j) \in E$ by the random value.

Figure 1 shows the degree (i.e., number of links) distribution of ER and cut BA networks obtained from the above steps with $n = 1,000$, $k_{ave} = 20$, $q = 0.5$, and $w_{ij} = 1$ for all $(i, j) \in E$. As reference, we drew the distribution of the original BA network [5] in Fig. 1. In the ER network and the cut BA network, each node has 20 links in average. On the contrary, the average links of the original BA network is about 40. According to Fig. 1, the degree distribution for large degree nodes in the cut BA network has the same scaling exponent (-3) of the original BA network. Hence, the cutting of links in step 2b keeps the scale-free property of original BA networks.

The reason why we cut the links of BA networks randomly in step 2b is described below. According to the BA model [5], minimum degree k_{min} and average degree k_{ave} of original BA networks are always the same value when keeping the configuration of n_0 and n_a . On the contrary, k_{min} and k_{ave} of ER networks are randomly changed when keeping the configuration of p . By cutting the links of BA networks randomly, the BA networks have different k_{min} and k_{ave} , and we can compare the results for BA networks and ER networks under comparable condition.

We summarize the relation between the randomly-weighted random networks (the ER networks and the cut

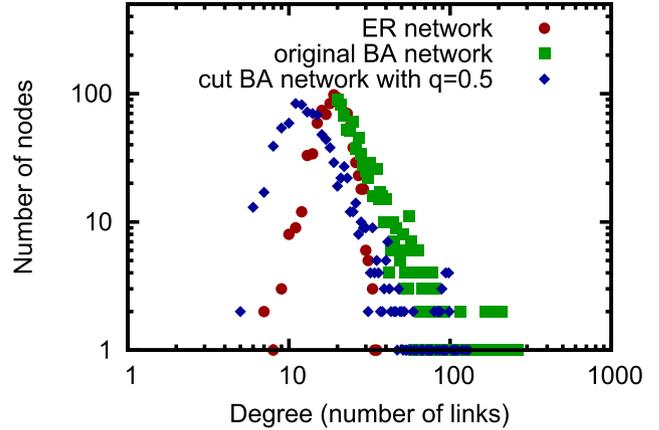


Fig. 1 Degree distribution of unweighted ER network and cut BA network.

Table 1 Default parameter configuration.

parameter	symbol	configuration
number of nodes	n	1,000
distribution of link weight w_{ij}		uniform distribution
average of link weights	\bar{w}	1
number of bins in $f_N(\lambda)$	n_h	50

BA networks) and social networks. Several actual social networks have been confirmed that they have scale-free property commonly [11]. The original BA model [5] represents the scale-free property, and so this model is widely used in several studies related to social networks [12], [13]. According to Fig. 1, the cut BA network also has the scale-free property like the original BA network. Hence, the cut BA network is associated with social networks in term of the scale-free property. On the other hand, the ER network is the simplest random network, and does not have the scale-free property. However, the ER network is useful to clarify the effect of the scale-free property in social networks by the comparison between the results with the ER network and the cut BA network.

3.2 The Universality and Its Applicable Condition

We experimentally investigate the eigenvalue frequency distribution $f_N(\lambda)$ of normalized Laplacian matrix N for the randomly-weighted random networks, and clarify the universality (the Wigner semicircle law) of $f_N(\lambda)$ and the applicable condition.

In the investigation, we use the constant distribution with $w_{ij} = \bar{w}$, the uniform distribution with the range $[\bar{w}/2, 3\bar{w}/2]$ and exponential distribution with the average \bar{w} to set link weight w_{ij} randomly. Note that these distributions have the same average of link weights, but a different variance of link weights. By the comparison of the results for the different distributions, we clarify the universality regardless of link weights in social networks. We repeat the generation of N 100 times, and calculate the average of these results. We use the parameter configuration shown in Table 1 as default.

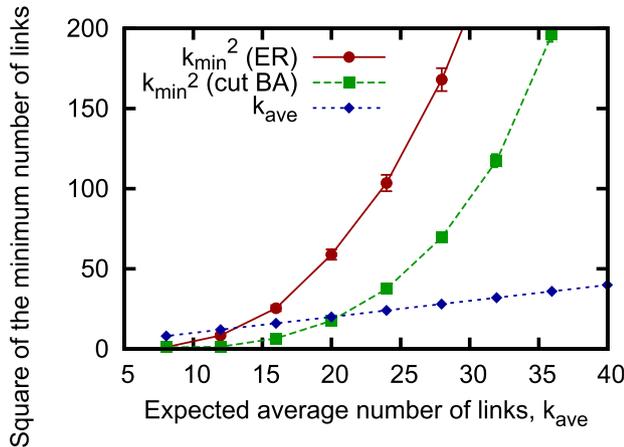


Fig. 2 The expected average number k_{ave} of links vs. the square of the minimum number k_{min} of links in randomly-weighted ER network and cut BA network.

We discuss the applicable condition of the universality on the basis of k_{min}^2 and k_{ave} like the existing study [10]. In Fig. 2, we first show k_{min}^2 and k_{ave} in randomly-weighted ER and cut BA networks with different expected average number of links, k_{ave} , and the uniform distribution. As k_{ave} increases, k_{min}^2 and k_{ave} increases simultaneously, but the increasing speed of k_{min}^2 is larger than that of k_{ave} . According to Fig. 2, in order to fulfill the degree condition $k_{min}^2 \gg k_{ave}$ in [10], $k_{ave} \geq 16$ for ER network and $k_{ave} \geq 24$ for cut BA network are needed at least, respectively. In this paper, we set the average of link weight w_{ij} to an amount equal to or greater than 1. Hence, if $k_{min}^2 \gg k_{ave}$ is fulfilled, $d_{min}^2 \gg d_{ave}$ is also fulfilled in average. Hence, we use $k_{min}^2 \gg k_{ave}$ instead of $d_{min}^2 \gg d_{ave}$ in order to clarify the applicable condition regardless of link weights.

In Figs. 3 and 4, we show eigenvalue frequency distributions $f_N(\lambda)$ of normalized Laplacian matrix N for randomly-weighted ER and cut BA networks, respectively. When we obtained $f_N(\lambda)$, we first counted the number of eigenvalues of N within $[\lambda - h_b/2, \lambda + h_b/2]$ where $h_b = (\lambda_n - \lambda_2)/n_h$. Then, we normalized the counted number so that $\sum_{i=1}^{n_h} f_N(\theta_i) = 1$ where $\theta_i = \lambda_2 + (i - 1/2)h_b$, and n_h is the number of bins in $f_N(\lambda)$. Note that we took out the minimum eigenvalue λ_1 of N when obtaining $f_N(\lambda)$ because λ_1 is always 0. In Figs. 3 and 4, we also draw the semicircle distribution, which is given by

$$\rho_N(\lambda) = \begin{cases} \frac{2}{\pi r^2} \sqrt{r^2 - (\lambda - 1)^2} & (\lambda_2 < \lambda < \lambda_n) \\ 0 & (\text{otherwise}) \end{cases}, \quad (7)$$

where r is the radius of the semicircle distribution, and is given by $1 - \lambda_2$ or $\lambda_n - 1$. According to the range of λ_2 and λ_n , r must be within $0 < r < 1$. Equation (7) is essentially equivalent to Eq. (6) in the Wigner semicircle law because there is just the difference of the semicircle center. Hence, we define that the Wigner semicircle law is satisfied if eigenvalue frequency distribution $f_N(\lambda)$ coincides with the semicircle distribution given by Eq. (7). According to

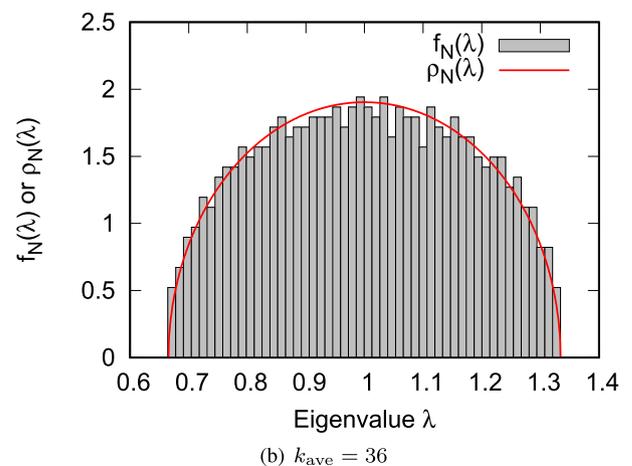
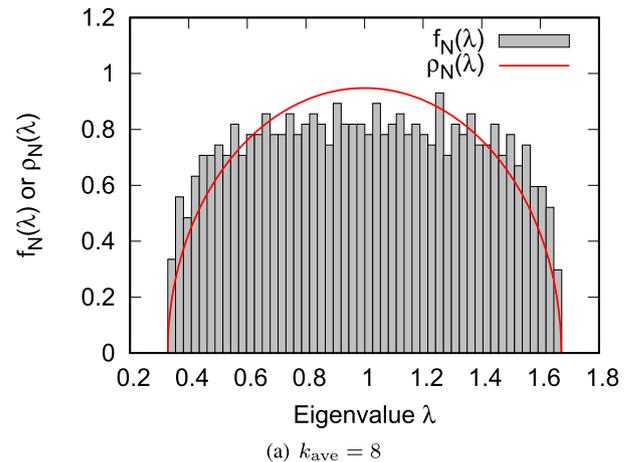


Fig. 3 The eigenvalue frequency distributions of N 's for randomly-weighted ER networks.

Figs. 3 and 4, eigenvalue frequency distribution $f_N(\lambda)$ of N for the randomly-weighted ER and cut BA networks with $k_{ave} = 8$ differs from the semicircle distribution (7), but the distributions for the randomly-weighted ER and cut BA networks with $k_{ave} = 36$ almost coincide with it. Hence, we expect that large k_{ave} is needed to satisfy the Wigner semicircle law.

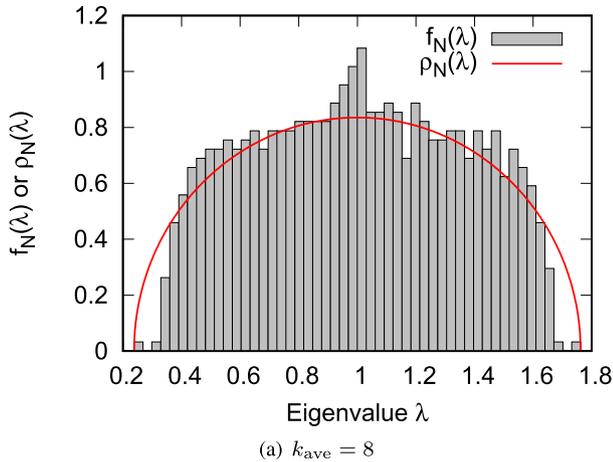
In order to clarify whether the eigenvalue frequency distribution $f_N(\lambda)$ satisfies the Wigner semicircle law, we investigate the difference between $f_N(\lambda)$ and $\rho_N(\lambda)$. As the definition of the difference, we use relative error ϵ . Relative error ϵ is given by

$$\epsilon = \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{|f_N(\theta_i) - P(\theta_i)|}{P(\theta_i)}, \quad (8)$$

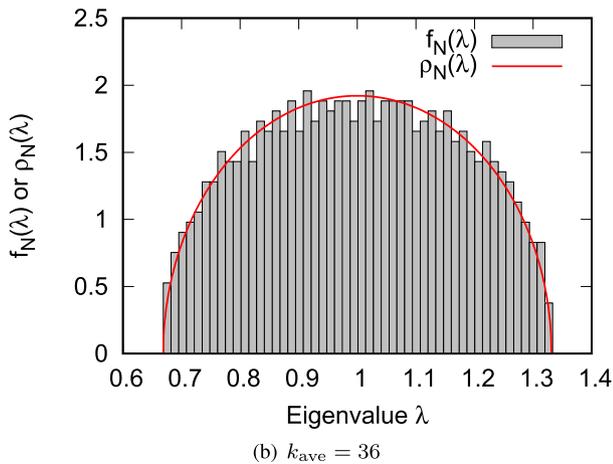
where $\theta_i = \lambda_2 + (i - 1/2)h_b$, and $P(\theta_i)$ is the probability that the eigenvalues of N is within $[\theta_i - h_b/2, \theta_i + h_b/2]$. Namely, $P(\theta_i)$ is defined by

$$P(\theta_i) = \int_{\theta_i - h_b/2}^{\theta_i + h_b/2} \rho_N(\lambda) d\lambda. \quad (9)$$

Note that $P(\theta_i)$ should become the true value of $f_N(\theta_i)$ for



(a) $k_{ave} = 8$



(b) $k_{ave} = 36$

Fig. 4 The eigenvalue frequency distributions of N 's for randomly-weighted cut BA networks.

$n \rightarrow \infty$ and $n_h \rightarrow \infty$.

In Fig. 5, we show the results of ϵ for randomly-weighted ER and cut BA networks with the constant distribution, the uniform distribution, and the exponential distribution of link weight w_{ij} . As k_{ave} increases, ϵ decreases regardless of the link weight distribution and network topology (ER or cut BA). The reason why ϵ does not approach to 0 is that there are discretization error due to the finite setting of n and n_h . Moreover, ϵ for the uniform and exponential distributions approaches to that for the constant distribution. According to [10], $f_N(\lambda)$ with the constant distribution ($w_{ij} = 1$) satisfies the Wigner semicircle law when the degree condition $k_{min}^2 \gg k_{ave}$ is fulfilled. Hence, $f_N(\lambda)$ with the uniform and exponential distributions also satisfies the Wigner semicircle law. However, it is ambiguity how $k_{min}^2 \gg k_{ave}$ is fulfilled. According to Figs. 2 and 5, the range of $k_{min}^2 \gg k_{ave}$ is different from the range of $k_{min}^2 > k_{ave}$. ϵ shown in Fig. 5 helps to understand how $k_{min}^2 \gg k_{ave}$ is fulfilled. According to Fig. 5, ϵ is convergent at $k_{ave} > 36$. Hence, $k_{ave} > 36$ is needed to fulfill $k_{min}^2 \gg k_{ave}$ for the example in this section.

From the above results, we can find that if the degree

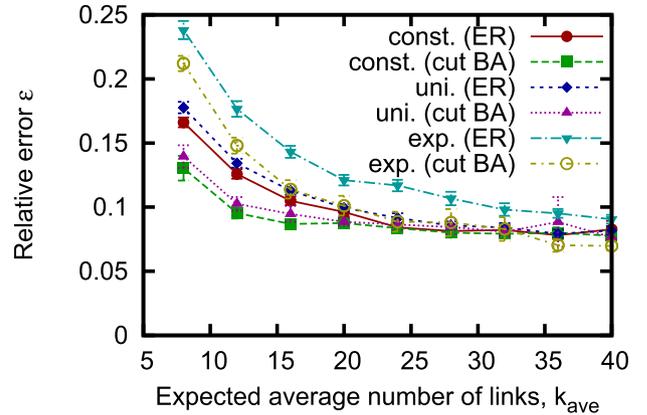


Fig. 5 Relative error ϵ of eigenvalue frequency distribution $f_N(\lambda)$ for randomly-weighted ER and cut BA networks.

condition $k_{min}^2 \gg k_{ave}$ is fulfilled, the eigenvalue frequency distribution of normalized Laplacian matrix N for randomly-weighted ER and cut BA networks satisfies the Wigner semicircle law given by Eq. (7), which is the semicircle distribution determined by only the second smallest eigenvalue λ_2 or the maximum eigenvalue λ_n of N . Hence, λ_2 and λ_n are important to understand the social network property in the a case fulfilling the degree condition.

4. Analysis of the Information Propagation Speed

In this section, we analyze the speed of the information propagation on social networks fulfilling the degree condition $k_{min}^2 \gg k_{ave}$. If the degree condition is not fulfilled in a social network, there are many persons with too small number of friends. However, such persons would be minority in an actual social network, and contribute less to the information propagation on the entire social network. Therefore, we ignore such persons, and focus on social networks fulfilling the degree condition.

4.1 Metric of the Information Propagation Speed

The information propagation in a social network is involved with the chain of word-of-mouth communications (e.g., face-to-face offline conversation, and online communication via SNS) between persons. In such a communication chain, the information is more likely to be propagated to the persons that have many friends. Random walks on a network have similar characteristic that the probability of the random walker arriving at a node is proportional to its node degree [14], which corresponds to the number of friends in a social network. Hence, we model the information propagation on social networks as a random walk. However, the information propagation modeled with the random walk may be too slower than the information propagation in an actual social network since the random walker arrives at the same node multiple times. Therefore, our analysis focuses on the worst-case situation for the information propagation in social networks.

Random walk on network G is formulated by normalized Laplacian matrix N . When node i selects node j with the probability w_{ij}/d_i , arrival probability $x_{a:i}(t)$ of the random walker starting from node a to node i at time t is given by

$$x_{a:i}(t) = \sum_{j \in \partial i} \frac{w_{ji}}{d_j} x_{a:j}(t-1), \quad (10)$$

where $\partial i = \{j \mid A_{ij} > 0, 1 \leq j \leq n\}$. With arrival probability vector $\mathbf{x}_a(t) = (x_{a:i}(t))_{1 \leq i \leq n}$, we obtain

$$\begin{aligned} \mathbf{x}_a(t) &= \mathbf{x}_a(t-1) \mathbf{D}^{-1} \mathbf{A} \\ \mathbf{x}_a(t) \mathbf{D}^{-1/2} &= \mathbf{x}_a(t-1) \mathbf{D}^{-1} \mathbf{A} \mathbf{D}^{-1/2} \\ \mathbf{y}_a(t) &= \mathbf{y}_a(t-1) (\mathbf{I} - \mathbf{N}), \end{aligned} \quad (11)$$

where $\mathbf{y}_a(t) = \mathbf{x}_a(t) \mathbf{D}^{-1/2}$.

A fundamental metric to evaluate of the information propagation speed in a social network is a *first arrival time*, which is the time required until the information is first propagated to a person. Such a first arrival time of the information corresponds to the time until the random walker first arrives at a node in the random walk. The existing study [14] derived first arrival time $f_{a:i}$ of the random walker starting from node a to node i as

$$f_{a:i} = 2|E| \sum_{l=2}^n \frac{1}{\lambda_l} \left(\frac{q_l^2(i)}{d_i} - \frac{q_l(a) q_l(i)}{\sqrt{d_a d_i}} \right), \quad (12)$$

where $q_l(i)$ is the i -th element of eigenvector \mathbf{q}_l of normalized Laplacian matrix N , and $|E| = 1/2 \sum_{i=1}^n d_i$. Then, we define the metric of the information propagation speed by using first arrival time $f_{a:i}$. For different nodes i and j , $f_{a:i}$ is different from $f_{a:j}$. Hence, we use expected time m^\dagger of first arrival time $f_{a:i}$ in order to define the metric for the entire network. In the random walk starting from node a , the steady-state probability arriving at node i is given by $d_i/(2|E|)$. We assume that the first arrival event of node i occurs with the probability $d_i/(2|E|)$, and expected time m of first arrival time $f_{a:i}$ is given by

$$\begin{aligned} m &= \sum_{i=1}^n \frac{d_i}{2|E|} f_{a:i} \\ &= \sum_{i=1}^n \sum_{l=2}^n \frac{1}{\lambda_l} \left(q_l^2(i) - \frac{q_l(a) q_l(i)}{\sqrt{d_a}} \sqrt{d_i} \right) \\ &= \sum_{l=2}^n \frac{1}{\lambda_l} \left(\sum_{i=1}^n q_l^2(i) - \frac{q_l(a)}{\sqrt{d_a}} \sum_{i=1}^n q_l(i) \sqrt{d_i} \right) \\ &= \sum_{l=2}^n \frac{1}{\lambda_l} \left(\sum_{i=1}^n q_l^2(i) - \frac{\sqrt{2|E|} q_l(a)}{\sqrt{d_a}} \sum_{i=1}^n q_l(i) q_1(i) \right) \\ &= \sum_{l=2}^n \frac{1}{\lambda_l}. \end{aligned} \quad (13)$$

[†]Note that ‘‘expected value’’ does not mean ‘‘arithmetically average’’. The expected value of a variable is defined by the weighted average using the probability of its occurrence as the weight.

When we derived the above equation, we used $q_1(i) = \sqrt{d_i/(2|E|)}$ and the property that \mathbf{q}_l is the orthonormal basis. From the above equation, we find that m can be calculated with only eigenvalue λ_l of normalized Laplacian matrix N , and does not depend on starting node a . In our analysis, we use m as the information propagation metric of the social network.

Equation (13) is also derived as Eq. (3.3) in [14]. However, Eq. (3.3) is slightly different from Eq. (13) because λ_k in [14] is defined as the k -th eigenvalue of $\mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$. In this paper, λ_k is the k -th eigenvalue of $\mathbf{N} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$. Hence, λ_k in this paper corresponds to $1 - \lambda_k$ in [14]. According to the difference between these definitions, Eq. (13) is equivalent to Eq. (3.3) in [14].

4.2 The Information Propagation Speed with the Universality of Normalized Laplacian Matrix N

Using the universality shown in Sect. 3, we analyze the information propagation speed. According to the universality, eigenvalue frequency distribution $f_N(\lambda)$ of N for randomly-weighted networks is approximated by $\rho_N(\lambda)$, which given by Eq. (7). We first derive \tilde{m} that is the approximated value of m with the assumption that $f_N(\lambda) \simeq \rho_N(\lambda)$. Then, we discuss the information propagation speed on the basis of \tilde{m} .

If $f_N(\lambda) \simeq \rho_N(\lambda)$, expected value m of first arrival time $f_{a:i}$ is approximated by

$$\begin{aligned} m &= \sum_{l=2}^n \frac{1}{\lambda_l} \simeq (n-1) \sum_{i=1}^{n_h} \frac{1}{\theta_i} f_N(\theta_i) \\ &\simeq (n-1) \int_{\lambda_2}^{\lambda_n} \frac{1}{\lambda} \rho_N(\lambda) d\lambda \\ &= \frac{2(n-1)}{\pi r^2} \int_{\lambda_2}^{\lambda_n} \frac{1}{\lambda} \sqrt{r^2 - (\lambda-1)^2} d\lambda \\ &= \frac{2(n-1)}{r^2} \left(1 - \sqrt{1-r^2} \right) = \tilde{m}, \end{aligned} \quad (14)$$

where $\theta_i = \lambda_2 + (i-1/2)h_b$ and $h_b = (\lambda_n - \lambda_2)/2$. The detailed deviation process of Eq. (14) is provided in the appendix. According to Eq. (14), \tilde{m} is only determined by the number n of nodes, and radius r of the semicircle distribution.

On the basis of Eq. (14), we analyze the worst-case speed of the information propagation in a social network. Figure 6 shows \tilde{m} given by Eq. (14) as the function of radius r . Note that r is within the range $(0, 1)$ because $0 < \lambda_2, \lambda_n < 2$. From this figure, \tilde{m} is the monotonically increasing function of r because of $d\tilde{m}/dr > 0$ for $0 < r < 1$. Hence, if m is able to be approximated by \tilde{m} , the worst-case speed of the information propagation in a social network becomes slower as r increases. Moreover, since the range of r is $0 < r < 1$, the lower bound and upper bound of \tilde{m} are given by $n-1$ and $2(n-1)$, respectively. Hence, if the structure (i.e., relationship among people) of a social network changes, the worst-case speed of the information propagation changes up to twice.

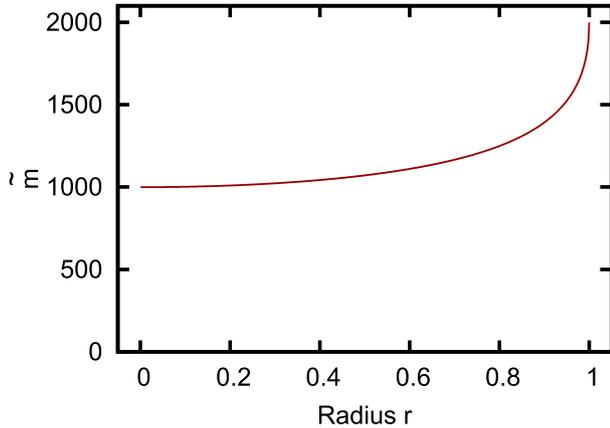


Fig. 6 Radius r vs. \tilde{m} .

4.3 The Validity of Our Analysis

Our analysis is valid if the difference between m and \tilde{m} is sufficiently small. To investigate the difference, we use relative error ϵ_m , which is defined by

$$\epsilon_m = \frac{|\tilde{m} - m|}{m}. \quad (15)$$

In the investigation, we repeat the generation of N 100 times, and calculate the average of related error ϵ_m . Like Sect. 3, we use the the parameter configuration shown in Table 1 as default, and randomly-weighted ER or cut BA networks. These networks have no clusters that are the set of densely-connected nodes, and are often observed in an actual social network. A network with no clusters can be obtained from a network with clusters by replacing clusters to nodes. Hence, our investigation of the information propagation using ER and cut BA networks corresponds to the investigation ignoring the information propagation time in each cluster. In general, the information propagation in a cluster is very fast. Hence, when investigating the information propagation in a network with clusters, it is important to analyze the information propagation times among the clusters. Therefore, our investigation also helps to understand the information propagation in networks with clusters.

Figure 7 shows relative error ϵ_m for randomly-weighted ER and cut BA networks. According to this figure, ϵ_m approaches to 0 as k_{ave} increases regardless of the link weight distribution and network topology. Although the discretization error due to n and n_h affects relative error ϵ of eigenvalue frequent distribution $f_N(\lambda)$, it does not affect ϵ_m . This characteristic is useful for the information propagation analysis.

By the comparison between Figs. 2 and 7, we conclude that our analysis based on Eq. (14) is valid for social networks if the degree condition $k_{\text{min}}^2 \gg k_{\text{ave}}$ is fulfilled.

5. Conclusion and Future Work

Spectral graph theory cannot be simply applied to social

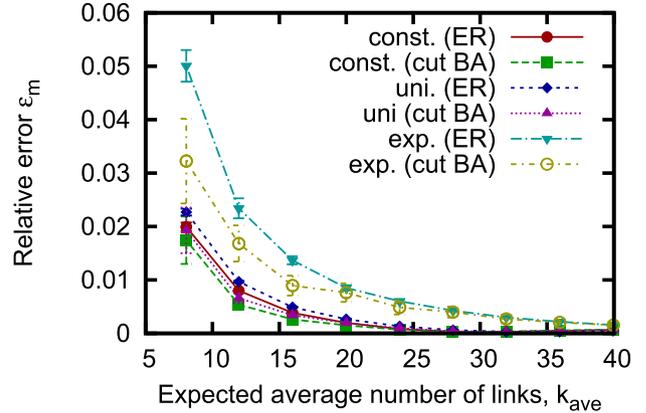


Fig. 7 Relative error ϵ_m of expected value of first-arrival time $f_{a:i}$ for randomly-weighted ER and cut BA networks.

network analysis because the matrix elements used in the theory cannot be given exactly to represent the structure of a social network. For this reason, we first discussed the universality of random matrices with the feature of social networks. As such a random matrix, we used the normalized Laplacian matrix N for a network where link weights are randomly given. We clarified that the universality (i.e., the Wigner semicircle law given by Eq. (7)) of normalized Laplacian matrix N appears regardless of the link weight distribution in ER networks and cut BA networks, which are the modified version of the BA networks [5]. According to the universality, eigenvalue frequency distribution $f_N(\lambda)$ of N is determined by only the number n of nodes and semicircle radius $r = 1 - \lambda_2$ or $\lambda_n - 1$ where λ_2 and λ_n are the second minimum eigenvalue and the maximum eigenvalue of N , respectively. Then, we analyzed the information propagation speed in a social network on the basis of the spectral graph theory and the clarified universality. In this analysis, we modeled the information propagation by a random walk in light of the resemblance between their characteristics, and investigated expected value m of first arrival times of the random walker for each node. Our analysis showed that the worst-case speed of the information propagation changes up to twice if the structure (i.e., relationship among people) of a social network changes.

As future work, we will investigate the relationship between topological property (e.g., scale-free property) and radius r since r determines the information propagation speed in a social network. Next, we plan to design a social media for effective information propagation on the basis of the finding by our work.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 15K00431.

References

- [1] F. Chung, Spectral Graph Theory, American Mathematical Society, 1997.

- [2] T. Guhr, A. Müller-Groeling, and H.A. Weidenmüller, "Random-matrix theories in quantum physics: common concepts," *Physics Reports*, vol.299, no.4, pp.189–425, June 1998.
- [3] E.P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Annals of Mathematics*, vol.67, no.2, pp.325–327, 1958.
- [4] P. Erdős and A. Rényi, "On random graphs," *Mathematica*, vol.6, no.26, pp.290–297, 1959.
- [5] A.L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol.286, no.5439, pp.509–512, Oct. 1999.
- [6] C. Zhan, G. Chen, and L.F. Yeung, "On the distributions of laplacian eigenvalues versus node degrees in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol.389, no.8, pp.1779–1788, April 2010.
- [7] S. Jalan and J.N. Bandyopadhyay, "Random matrix analysis of network laplacians," *Physica A: Statistical Mechanics and its Applications*, vol.387, no.2, pp.667–674, Jan. 2008.
- [8] I.J. Farkas, I. Derényi, A.L. Barabási, and T. Vicsek, "Spectra of "real-world" graphs: Beyond the semicircle law," *Phys. Rev. E*, vol.64, no.2, p.026704, July 2001.
- [9] Y. Sakumoto and H. Ohsaki, "Fluid-based analysis for understanding tcp performance on scale-free structure," *J. Inf. Process.*, vol.24, no.4, pp.660–668, July 2016.
- [10] F. Chung, L. Lu, and V. Vu, "Spectra of random graphs with given expected degrees," *Proc. National Academy of Sciences*, vol.100, no.11, pp.6313–6318, May 2003.
- [11] A.L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol.286, no.5439, pp.509–512, Oct. 1999.
- [12] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, "Efficiency of scale-free networks: Error and attack tolerance," *Physica A: Statistical Mechanics and its Applications*, vol.320, pp.622–642, March 2003.
- [13] B. Viswanath, A. Post, K.P. Gummadi, and A. Mislove, "An analysis of social network-based sybil defenses," *ACM SIGCOMM Comput. Commun. Rev.*, vol.40, no.4, pp.363–374, Oct. 2010.
- [14] L. Lovász, "Random walks on graphs: a survey," *Combinatorics*, Paul Erdős is eighty, vol.2, pp.353–398, 1996.

Appendix:

We describe the detailed deviation process of Eq. (14). If eigenvalue frequency distribution $f_N(\lambda)$ is given by the Wigner semicircle law, expected value m of the first-arrival time is approximated by \tilde{m} . \tilde{m} is given by

$$\begin{aligned} \tilde{m} &= (n-1) \int_{\lambda_2}^{\lambda_n} \frac{1}{\lambda} \rho_N(\lambda) d\lambda \\ &= \frac{2(n-1)}{\pi r^2} \int_{\lambda_2}^{\lambda_n} \frac{1}{\lambda} \sqrt{r^2 - (\lambda-1)^2} d\lambda. \end{aligned} \quad (\text{A} \cdot 1)$$

By substituting $\lambda = r \cos \theta + 1$ into Eq. (A·1), we obtain

$$\begin{aligned} \tilde{m} &= \frac{2(n-1)}{\pi r^2} \int_0^\pi \frac{r^2 \sin^2 \theta}{r \cos \theta + 1} d\theta \\ &= \frac{2(n-1)}{\pi r^2} \int_0^\pi \frac{r^2 - 1 + 1 - r^2 \cos^2 \theta}{r \cos \theta + 1} d\theta \\ &= \frac{2(n-1)}{\pi r^2} \int_0^\pi \left(\frac{r^2 - 1}{r \cos \theta + 1} + 1 - r \cos \theta \right) d\theta \\ &= I_1(r) + I_2(r), \end{aligned} \quad (\text{A} \cdot 2)$$

where $I_1(r)$ and $I_2(r)$ are given by

$$I_1(r) = \frac{2(n-1)}{\pi r^2} \int_0^\pi \frac{r^2 - 1}{r \cos \theta + 1} d\theta, \quad (\text{A} \cdot 3)$$

$$I_2(r) = \frac{2(n-1)}{\pi r^2} \int_0^\pi (1 - r \cos \theta) d\theta, \quad (\text{A} \cdot 4)$$

respectively. When we use the half-angle formula of $\cos \theta$, $I_1(r)$ is given by

$$\begin{aligned} I_1(r) &= \frac{2(n-1)}{\pi r^2} \\ &\times \int_0^\pi \frac{r^2 - 1}{r \left(\cos^2 \frac{\theta}{2} - \sin^2 \frac{\theta}{2} \right) + \left(\cos^2 \frac{\theta}{2} + \sin^2 \frac{\theta}{2} \right)} d\theta \\ &= \frac{2(n-1)}{\pi r^2} \int_0^\pi \frac{r^2 - 1}{(1+r) \cos^2 \frac{\theta}{2} + (1-r) \sin^2 \frac{\theta}{2}} d\theta \\ &= \frac{2(n-1)}{\pi r^2} \int_0^\pi \frac{r^2 - 1}{(1+r) + (1-r) \tan^2 \frac{\theta}{2}} \frac{1}{\cos^2 \frac{\theta}{2}} d\theta \\ &= \frac{2(n-1)}{\pi r^2} \int_0^\pi \frac{r-1}{1 - i^2 \frac{1-r}{1+r} \tan^2 \frac{\theta}{2}} \frac{1}{\cos^2 \frac{\theta}{2}} d\theta, \end{aligned} \quad (\text{A} \cdot 5)$$

where i is the imaginary unit. By substituting $x = \sqrt{\frac{1-r}{1+r}} \tan \frac{\theta}{2}$ into Eq. (A·5), $I_1(r)$ is derived as

$$\begin{aligned} I_1(r) &= \frac{2(n-1)}{\pi r^2} \int_0^\infty \frac{-2\sqrt{1-r^2}}{1 - (ix)^2} dx \\ &= \frac{2(n-1)}{\pi r^2} \\ &\times \int_0^\infty \left(-\sqrt{1-r^2} \right) \left(\frac{1}{1+ix} + \frac{1}{1-ix} \right) dx \\ &= \frac{2(n-1)}{\pi r^2} \left(-\sqrt{1-r^2} \right) \\ &\times \left[\frac{1}{i} \log(1+ix) - \frac{1}{i} \log(1-ix) \right]_0^\infty \\ &= \frac{2(n-1)}{\pi r^2} \left(i\sqrt{1-r^2} \right) \\ &\times \lim_{x \rightarrow \infty} \left[\log(1+ix) - \log(1-ix) \right]. \end{aligned} \quad (\text{A} \cdot 6)$$

When we use $\log z = \log |z| + i \arg(z)$, $I_1(r)$ is derived as

$$\begin{aligned} I_1(r) &= \frac{2(n-1)}{\pi r^2} \left(i\sqrt{1-r^2} \right) \lim_{x \rightarrow \infty} 2i \tan^{-1} x \\ &= -\frac{2(n-1)}{r^2} \sqrt{1-r^2}. \end{aligned} \quad (\text{A} \cdot 7)$$

Then, $I_2(r)$ is derived as

$$\begin{aligned} I_2(r) &= \frac{2(n-1)}{\pi r^2} \int_0^\pi (1 - r \cos \theta) d\theta \\ &= \frac{2(n-1)}{\pi r^2} \left[\theta - r \sin \theta \right]_0^\pi \\ &= \frac{2(n-1)}{r^2}. \end{aligned} \quad (\text{A} \cdot 8)$$

By substituting Eqs. (A·7) and (A·8) into Eq. (A·2), \tilde{m} is given as

$$\begin{aligned}\tilde{m} &= I_1(r) + I_2(r) \\ &= \frac{2(n-1)}{r^2} \left(1 - \sqrt{1-r^2}\right).\end{aligned}\quad (\text{A}\cdot 9)$$



Yusuke Sakumoto received M.E. and Ph.D. degrees in the Information and Computer Sciences from Osaka University in 2008 and 2010, respectively. He is currently an assistant professor at Graduate School of Systems Design, Tokyo Metropolitan University, Japan. His research work is in the area of the networking for social, computer, electricity device. He is a member of IEEE, IPSJ and IEICE.



Tsukasa Kameyama received B.E. and M.E. degrees from Tokyo Metropolitan University in 2016 and 2018, respectively. He was a student at Graduate School of System Design, Tokyo Metropolitan University, Japan from 2016 to 2018. He currently works at Nochu Information System Co., Ltd. His research interest is in the area of social network.



Chisa Takano received the B.E. degree in Telecommunication Engineering from Osaka University, Japan, in 2000, and received the Ph.D. in Telecommunications Engineering from Tokyo Metropolitan University, Japan, in 2008. In 2000 she joined the Traffic Research Center, NTT Advanced Technology Corporation (NTT-AT). Since April 2008, she has been an Associate Professor of the Graduate School of Information Sciences, Hiroshima City University. Her research interests are in the area of computer networks and distributed systems. She received the IEICE's Young Researchers' Award in 2003.

compute networks and distributed systems. She received the IEICE's Young Researchers' Award in 2003.



Masaki Aida received the B.S. degree in Physics and M.S. degree in Atomic Physics from St. Paul's University, Tokyo, Japan, in 1987 and 1989, respectively, and received the Ph.D. in Telecommunications Engineering from the University of Tokyo, Japan, in 1999. After joining NTT Laboratories in April 1989, he has been engaged in research on traffic issues in computer communication networks. From April 2005 to March 2007, he was an Associate Professor at the Faculty of System Design, Tokyo Metropolitan

University. He has been a Professor of the Graduate School of Systems Design, Tokyo Metropolitan University since April 2007. He received the Best Tutorial Paper Award of IEICE Communications Society in 2013. Prof. Aida is a member of the IEEE, IEICE, and the Operations Research Society of Japan.