A Study on Relationship between Modularity and Diffusion Dynamics in Networks from Spectral Analysis Perspective

Kiyotaka Ide, Akira Namatame
Department of Computer Science
National Defense Academy of Japan
Yokosuka, Japan

Loganathan Ponnambalam , Fu Xiuju,
Rick Siow Mong Goh
Computing Science, Institute of High Performance Computing,
A*STAR
Singapore, Singapore

Abstract—**Modular structure is a typical structure that is observed in most of real networks. Diffusion dynamics in network is getting much attention because of dramatic increasing of the data flows via the www. The diffusion dynamics in network have been well analysed as probabilistic process, but the proposed frameworks still shows the difference from the real observations. In this paper, we analysed spectral properties of the networks and diffusion dynamics. Especially, we focus on studying the relationship between modularity and diffusion dynamics. Our analysis as well as simulation results show that the relative influences from the non-largest eigenvalues and the corresponding eigenvectors increase when modularity of network increases. These results have the implication that, although network dynamics have been often analysed with the approximation manner utilizing only the largest eigenvalue, the consideration of the other eigenvalues is necessary for the analysis of the network dynamics on real networks. We also investigated Node-level Eigenvalue Influence Index (NEII) which can quantify the relative influence from each eigenvalues on each node. This investigation indicates that the influence from each eigenvalue is confined within the modular structures in the network. These findings should be made consideration by researchers interested in diffusion dynamics analysis on real networks for deeper analysis.**

Keywords—complex network; modularity; diffusion; SIS model; graph spectra; eigenvalue and eigenvector

I. INTRODUCTION

Diffusion phenomena ongoing on today's well-networked society can be often analyzed as probabilistic diffusion processes in complex networks. And, because the social systems have been keeping glowing up more dynamically and complexly, studying the probabilistic diffusion process in complex networks has been gathering a lot of attentions. Also, the probabilistic diffusion process has been well-applied to various fields, such as information spreads, dissemination of new products, computer virus spread, and epidemics.

Modular structure is a ubiquitous characteristic found in many real networks [e.g. 1]. Identifying hidden modular structure in real networks has been studied by many researchers in the scope of social network analysis [e.g., 2–8]. For instance, Newman's community detecting algorithm using betweenness centrality [4] is the pioneer work that triggered the development of community detecting algorithms. But,

their algorithm have two problems; 1) the number of communities is needed to be estimated in advance, even if the user might want to know the most optimized number of partitioned communities as the result of optimization. 2) computation time is too long. The first problem was solved by the introduction of modularity Q that is an index that can quantify the accuracy of the partitioning [5]. Then the users can identify the most proper partitioning instead of deciding the number of communities in the network in advance. The second problem was solved using the greedy algorithm that the completely separated graphs connect to be higher modularity Q [6]. After that, many researches have proposed various methods and especially apply them to the social network analysis. In addition to that, diffusion properties on modular networks have been studies by many researchers. For instance, Gao et al. [9] investigated the relationship between the number of modules and the properties of percolation on the randomly modularized network, which results that modularized networks are more destructible than a single independent network. In terms of the probabilistic diffusion dynamics on modular networks, it is reported that resonance-like phenomena can be seen in the probabilistic diffusion processes on modular networks [10]. Also, Saumuell-Mendiola [11] studied the SIS diffusion model on a coupled network which consists of two independent networks combined each other, and they reported that epidemics are prone to arise on the interconnected networks comparing to the single independent networks. Furthermore, Sahneh et al. [12] and Wang et al. [13] proposed the theories that the epidemic threshold for a coupled network can be calculated by the adjacency matrix of the coupled network. Also, for the Susceptible-Infected-Recovered (SIR) diffusion model, analyzing the interconnection between communities is also important [14, 15].

Analyzing the diffusion process in networks as Susceptible-Infected-Susceptible (SIS) model has been one of the conventional approaches that can be well-applied to the study of information diffusion as well as epidemics [16-21]. In the SIS model, each node in the network is probable to be assigned to two states, susceptible state and infected state. In the epidemics context, the susceptible state nodes represent the healthy individuals that are probable to be infected. On the other hand, the infected state nodes represent patients that are probable to influence its neighbors at a certain infection rate in

the epidemics context. Then, the infected state nodes are possible to automatically return to the susceptible state at a certain recovery rate. One of the important insights from the studies on the SIS model in complex network is that critical phenomenon, which the steady-state fraction of infected nodes suddenly jumping up at the certain condition, can be observed. Then, finding the critical point (i.e. threshold or tipping point) has been attracting many researchers' interests, and many analytical approaches as well as simulative approaches have been proposed so far [19,22-24]. For instance, Kephart and White [22] firstly analyzed SIS model and formulated time-evolution of the steady-state fraction of infected nodes in homogeneous network. Wang et al. [23] proposed a more advanced analytical framework for general networks from the spectral point of view. They reported that the critical point can be approximately calculated by the inverse of the largest eigenvalue of the adjacency matrix of the underlying network. Mieghem et al. [24] developed "the N-intertwined mean field approximation model". And our work is based on their analytical frameworks. In addition, Mieghem et al. [25] also proposed another approach from spectral analysis perspective.

Although many theories have been proposed, as mentioned above, Pastor-Satorras and Vespignani [26] reported that, in scale-free network, the critical phenomena cannot be observed from their analysis of the empirical survey results of computer virus spread. They also found that the infections are localized within very small areas before the critical point. Furthermore, they reported that the steady-state fraction of infection saturates to very lower than analytically expected. These empirical facts differ from the analytical results introduced above. Recently, this contradiction is elucidated by the localization-delocalization phenomenon reported by Goltsev et al. [27]. The In their paper, the inverse participation ratio (IPR) is applied to the network diffusion analysis from the spectral point of view, and concludes that hubs, edges with large weight and dense sub-graphs in networks are probable to be the centers of localization.

Because of the limitation of computational performance and data availability, previous analysis on the diffusion model had tried to find out better approximation approaches to figure out the dynamics. These analytical results based on the conventional linear algebra based analysis indicate that the largest eigenvalue and the principal eigenvector of the adjacency matrix can approximate the diffusion dynamics on general networks [23, 24]. However, according to the results of our analysis from spectral point of view and numerical simulations, the accuracy of this approximation method varies depending on the modularity of the network. In our previous works [28], we quantified the accuracy of the approximation method utilizing only the largest eigenvalue of the adjacency matrix and found that the accuracy is low in some real networks. In this paper, we insist that the accuracy of the approximation method depends on the modularity of networks, which verifies numerical simulations.

Also, our proposed measure, Node-level Eigenvector Influence Index (NEII) [28] which can quantify and visualize the influences from an arbitrary eigenvalue to each node, captures the insight that only considering the largest

eigenvalue cannot appropriately approximate the dynamics on the highly modularized networks.

In the second section, we review some existing analytical frameworks and provide our proposed analytical frameworks that will be fundamentals for the later discussions. In the third section, we examine the spectral properties of some artificial complex networks and real networks, which indicates that the importance of non-largest eigenvalue for the diffusion properties on the modular networks and verify the hypothesis in the previous section by numerical simulations. In the fourh section, we develop the parameterized modular network formation algorithm to verify the hypothesis. In the fifth section, we introduce the Node-level Eigenvector Influence Index (NEII). Then the investigation of real modular networks conducted in the sixth section. Finally, we conclude this paper in the seventh section.

II. ANALYSIS OF PROBABILISTIC DIFFUSION ON NETWORKS

A. N-Intertwined mean field approximation model

Mieghem et al. [24] developed the N-intertwined mean field approximation model, then an important results of their analysis is the following Markov differential equation,

$$\frac{d\mathbf{V}(t)}{dt} = \beta\mathbf{A}\mathbf{V}(t) - \text{diag}\big(v_i(t)\big)(\beta\mathbf{A}\mathbf{V}(t) + \delta\mathbf{e})$$
$$= (\beta\mathbf{A} - \delta\mathbf{I})\mathbf{V}(t) - \beta\text{diag}\big(v_i(t)\big)\mathbf{A}\mathbf{V}(t),$$

(1)

where $v_i(t)$ denotes the probability that the node i is infected at time t, β is infection rate, δ is recovery rate, $\mathbf{V}(t) = \big(v_1(t), v_2(t), v_3(t), \cdots, v_N(t)\big)^T$, $\mathbf{e}$ is the all-one vector, and $\text{diag}\big(v_i(t)\big)$ is the diagonal matrix where the diagonal elements consists of $v_1(t), v_2(t), v_3(t), \cdots, v_N(t)$. According to the comparison results with the numerical simulation results in small networks, the accuracy of this model is good enough except the region around threshold. In the studies of Susceptible-Infected-Susceptible (SIS) model [16-21], researchers have been making efforts to identify the threshold τ_c of the effective infection ratio $\tau \equiv \frac{\beta}{\delta}$. Then, several approaches have been taken to analytically calculate the threshold [22-26]. One of the most prominent achievements is that the threshold can be derived by the inverse of the largest eigenvalue of the adjacency matrix as follows,

$$\tau_c = \frac{1}{\lambda_1(\mathbf{A})},$$

(2)

where $\lambda_1(\mathbf{A})$ denotes the largest eigenvalue of the adjacency matrix $\mathbf{A}$.

B. Spectral analysis

To solve the differential equation (1), we assumes that the fraction of infection on each node $v_i(t)$ is sufficiently small and ignoring the term, the equation (1) can be solved as the expression (3) using the eigenvalue decomposition,

$$\mathbf{V}(t) = \mathbf{U}\text{diag}(\exp((\beta\lambda_k - \delta))t)\mathbf{U}^{\mathrm{T}}\mathbf{V}(0), \qquad (3)$$

where $\lambda_k(\mathbf{A})$ is kth eigenvalue of the adjacency matrix $\mathbf{A}$ and $\mathbf{U}$ denotes the orthonormal matrix which the kth column consists of the eigenvector of the kth eigenvalue. Then, the equation (3) can be rewritten as,

$$\mathbf{V}(t) = \sum_k \exp\big((\beta\lambda_k(\mathbf{A}) - \delta)t\big)\mathbf{u}_k\mathbf{u}_k^{\mathrm{T}}\mathbf{V}(0), \qquad (4)$$

where $\mathbf{u}_k$ is eigenvector of the kth eigenvalue of the adjacency matrix $\mathbf{A}$.

Assuming that the initial infection is randomly assigned to each node i at the probability $v_i(0) = 1/N$, the probability of infection on the node i at time t can be obtained as below,

$$v_i(t) = \frac{1}{N}\big(\exp((\beta\lambda_1 - \delta)t)u_{1i}\|\mathbf{u}_1\| \\ + \exp((\beta\lambda_2 - \delta)t)u_{2i}\|\mathbf{u}_2\| + \cdots \\ + \exp((\beta\lambda_n - \delta)t)u_{ni}\|\mathbf{u}_n\|), \qquad (5)$$

where the norm $\|\mathbf{u}_k\|$ stands for the sum of all elements of the eigenvector corresponding kth eigenvalue, that is $\|\mathbf{u}_k\| = u_{k1} + u_{k2} + u_{k3} + \cdots + u_{kn}$. Then, each term in the formula (5) implies that the influence from the kth eigenvalue λ_k toward a given node i is governed by the product of the ith component u_{ik} and the norm $\|\mathbf{u}_k\|$. Furthermore, the fraction of infected nodes over the whole network $y(t)$ can be calculated by taking the average of $v_i(t)$ as follows,

$$y(t) = \frac{1}{N}\sum_{i=1}^{N} v_i(t) \\ = \frac{1}{N^2}\sum_{k=1}^{N}\exp\big((\beta\lambda_k - \delta)t\big)\|\mathbf{u}_k\|^2. \qquad (6)$$

III. SPECTRAL PROPERTY AND DIFFUSION

A. Influence from the Non-Largest Eigenvalues

In the previous literatures, accuracy of the approximation method only utilizing the largest eigenvalue have not been discussed extensively and believed that is applicable to any types of networks. However, our analytical framework from the spectral point of view shows that influences from not only the largest eigenvalue of the adjacency matrix but also the other non-largest eigenvalues are important to express diffusion processes more accurately, which is validated by numerical simulation.

Then our investigation of the real networks shows that the modular networks with high modularity tend to show the property that the influences from the non-largest eigenvalues and the corresponding eigenvectors are significant.

In our previous work [28], we investigated the values of $\|\mathbf{u}_k\|^2$ in equation (6) in several artificial complex networks and real networks. The equation (6) can be expanded,

$$y(t) = \frac{\exp((\beta\lambda_1 - \delta)t)\|\mathbf{u}_1\|^2}{N^2}\bigg(1 \\ + \exp(\lambda_2 - \lambda_1)\frac{\|\mathbf{u}_2\|^2}{\|\mathbf{u}_1\|^2} + \cdots \qquad (7) \\ + \exp(\lambda_n - \lambda_1)\frac{\|\mathbf{u}_n\|^2}{\|\mathbf{u}_1\|^2}\bigg).$$

As the equation (7) indicates, when the absolute value of $\big|\lambda_1 - \lambda_{k,\{k\neq1\}}\big|$ is large and the dominance index ρ, which we defined as $\|\mathbf{u}_k\|^2/\|\mathbf{u}_1\|^2$, is small, $y(t)$ can be significantly governed by the largest eigenvalue and its corresponding eigenvector. In contrast, when $\big|\lambda_1 - \lambda_{k,\{k\neq1\}}\big|$ is small and the ρ is large, the term including the kth non-largest eigenvalue should be considered, that is, the approximation method only using the largest eigenvalue and corresponding eigenvectors, such as equation (2), is not applicable in this case.

Billen et al. [29] show that, as the number of triangles (i.e. clusters or three cycles) in a network increases, the spectrum of the network is positively skewed. This insight indicates that the value of $\big|\lambda_1 - \lambda_{k,\{k\neq1\}}\big|$ increases when the clustering coefficient of the network increases, which implies $y(t)$ tends to be governed by the largest eigenvalue when the clustering coefficient of the network is large. However, several studies on real data analysis and numerical simulation results [20] that scale-free network which has comparably larger clustering coefficient shows small steady-state fraction of infections, $y(\infty)$, which indicates that consideration of the absolute value of $\big|\lambda_1 - \lambda_{k,\{k\neq1\}}\big|$ is not important, and consideration of the value of ρ is more inevitable to measure the importance from the eigenvalues in each term of the equation (7).

Then, we investigated distribution of ρ in some artificial complex networks and real networks. Figure 1 shows the comparison of the distribution of ρ among the several networks, such as Barabasi-Albert scale-free network (BA), Erdos-Renyi random network (RND), random regular network (RR), Co-authorship Network of Network Scientists (CNNS) [30, 31] and UK members of parliament on Twitter network (UKMPTN) [32, 33]. European road network (EuroRoad) [22, 25], dolphin (Dolphin) network [31, 34], Email network (Email) [31, 35], and Jazz musicians' collaborating network (JazzNet) [36, 37]. Table 1 provides with the detailed network information including the optimal modularity Q that is an index to quantify the goodness of the partitioning and explained in the next section.

As can be seen in the figure, in RND and RR, the relative influence from the largest eigenvalue is apparently dominant and the relative influences from the non-largest eigenvalues are almost negligible. However, in the real networks, the relative importance from the non-largest eigenvalues increases. Especially CNNS and EuroRoad which show apparently high modularity and are apparently influenced from the non-largest eigenvalue. Especially, the 4th eigenvalue for CNNS and the 2nd eigenvalue for EuroRoad, is more dominant than those of their largest eigenvalues. This fact implies that the non-largest eigenvalues are more influential in the networks with the higher optimal modularity Q and the approximation only utilizing the largest eigenvalue and primary eigenvector is not appropriate to analyse these networks.

Network	N	Average Degree	Clustering Coefficient	The Largest Eigenvalue	Optimal # of Community	Optimal Modularity Q
BA	500	3.99	0.112	12.76	14	0.4911
RND	500	4	0.002	5.19	13	0.4901
RR	500	4	0.005	4	15	0.5103
CNNS	379	4.82	0.741	10.38	19	0.8386
UKMPTN	419	9.09	0.281	13.83	5	0.6270
Dolphin	62	5.13	0.259	7.19	4	0.4955
Email	1133	9.62	0.398	20.75	12	0.5070
EuroRoad	1039	2.51	0.0189	4.01	24	0.8649
JazzNet	198	27.70	0.617	40.03	4	0.4389

TABLE I. BASIC INFORMATION FOR THE INVESTIGATED NETWORKS
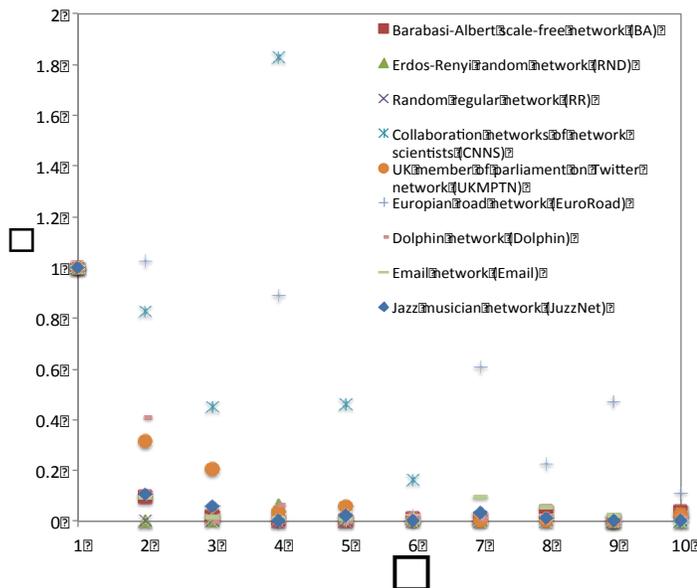
Fig. 1. Comparison of distribution of ρ, which we define at the dominance index of the principal vector, of Barabasi-Albert scale-free network (BA), Erdos-Renyi random network (RND), random regular network (RR), Co-authorship Network of Network Scientists (CNNS), UK member of parliament on Twitter network (UKMPTN), Euro road network (EuroRoad), dolphin (Dolphin) network, Email network (Email), and Jazz musicians' collaboration network (JazzNet).

B. Verification Simulation

Based on the results in the previous section, we hypothesize that, when we analyze the diffusion dynamics in

real networks, we must consider not only the largest eigenvalue and the principal eigenvector of the adjacency matrix, but also the other eigenvalues and their corresponding eigenvectors.

As indicated in formula (2), the critical point is approximately calculated as the inverse of the largest eigenvalue of the underlying network's adjacency matrix. To verify if this approximation method, utilizing only the largest eigenvalue, is appropriate for every network, we simply compare analytically derived approximated thresholds $\tau_{c,AD}$ with numerically calculated thresholds $\tau_{c,Sim}$ in the networks introduced in the previous section. In this series of numerical simulations, we change the effective infection ratio by 0.001 (recovery rate δ is a constant = 1), and the fraction of infected nodes at 100 time-steps is assumed to equal the steady-state fraction of infected nodes, y_∞. At the constant effective infection rate, the simulations repeated 100 trials and the obtained results were averaged. 2% of the nodes were randomly selected as initial infected nodes in each trial. In Figure 2, blue plots display the evolution of y_∞ as the function of τ normalized by $\tau_{c,AD}$ of each network. If the difference between $\tau_{c,AD}$ and $\tau_{c,Sim}$ is minimal, the blue plots begin to increase around 1 on the horizontal axis. Conversely, if the difference between $\tau_{c,AD}$ and $\tau_{c,Sim}$ is significant, the blue plots begin to increase much farther along the horizontal axis. As displayed in this figure, the difference between $\tau_{c,AD}$ and $\tau_{c,Sim}$ in RND and RR, in which the largest eigenvalue is prominently dominant, are almost negligible. In contrast, the differences are significant in CNNS and EuroRoad, which possess a comparatively large ρ for the non-largest eigenvalues as displayed in Figure 1 and large modularity Q as displayed in Table 1. These results demonstrate that an approximation method only considering the largest eigenvalue and the principal eigenvector is not appropriate for a network having comparatively large ρ_k values for its non-largest eigenvalues and large modularity Q.
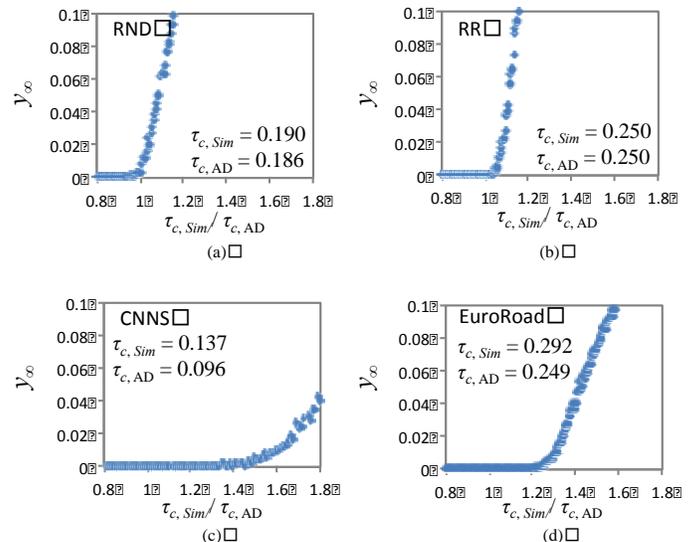
Fig. 2. Simulation results around the threshold. The steady-state fractions of infected nodes, y_∞, for (a) the Erdos-Renyi random network, (b) the random regular network (RR), (c) the Co-author Networks of Network Scientists

(CNNS) and (d) Euro road network (EuroRoad) are plotted as the function of effective infection ratio normalized by the analytically derived threshold of each network. The simulated threshold $\tau_{c,Sim}$ for each network corresponds to the tipping point of blue plots in each figure.

IV. RELATIONSHIP BETWEEN MODULARITY AND IMPORTANCE OF THE NON-LARGEST EIGENVALUE

According to the results in the previous section, it can be hypothesized that not only the largest eigenvalue but also the other non-largest eigenvalues are also important in networks with the large optimal modularity Q. The comparison results in the previous section show that the difference between $\tau_{c,AD}$ and $\tau_{c,Sim}$ in RND and RR, in which the largest eigenvalue is prominently dominant, are almost negligible. In contrast, the differences are significant in the real networks that show high optimal modularity Q, such as CNNS (Q = 0.84) and EuroRoad (Q = 0.86). Also, in our previous work [28], we proposed an index "the diffusion power" that can quantify the ease of diffusion on an arbitrarily network. An investigation of the diffusion power indicates that the networks with high optimal modularity Q shows that there are significant differences between the diffusion power when considering the all eigenvalues and eigenvectors and the diffusion power only considering the largest eigenvalue and principal eigenvector. Therefore, we hypothesized that the modularity of networks relates the importance of the non-largest eigenvalues and eigenvectors for the analysis of their diffusion dynamics. Therefore, we investigate the relationship between the optimal modularity Q and the importance of the non-largest

$$\text{Index} \equiv \frac{\max(\|\mathbf{u}_{2\sim10}\|^2)}{\|\mathbf{u}_1\|^2}. \tag{8}$$

eigenvalues and eigenvectors.

To show the relationship between the optimal modularity Q and the importance of the non-largest eigenvalues, we firstly develop the network formation algorithm that can change optimal modularity Q of the network by changing the network modularity control parameter (NMCP), p. The NMCP determines the ratio of the number of total links connecting the nodes inside the modules to the number of total links interconnect between modules. The step-by-step procedures for this parameterized network is as follows,

Step 1: Determine the number of module, the size of the modules (i.e. the number of nodes in each module), and the number of total links, L, for entire network in advance.

Step 2: Determine the NMCP, then caluculate the number of links interconnecting between modules.

Step 3: Calculate the number of links connecting the nodes inside each module, which is calculated by multiplying the NMCP and the number of total links (i.e. $p*L$).

Step 4: Randomly connect the links within each module by the links calculated in Step 3.

Step 5: Calculate the number of links inter-connecting each module, which is calculated by $(1-p)L$.

Step 6: Randomly connect each module by the links calculated in Step 5.

As shown in Figure 3, when the value of p increases, densities inside each module are increase while the connections between each module become sparse, then the optimal modularity Q also increases.
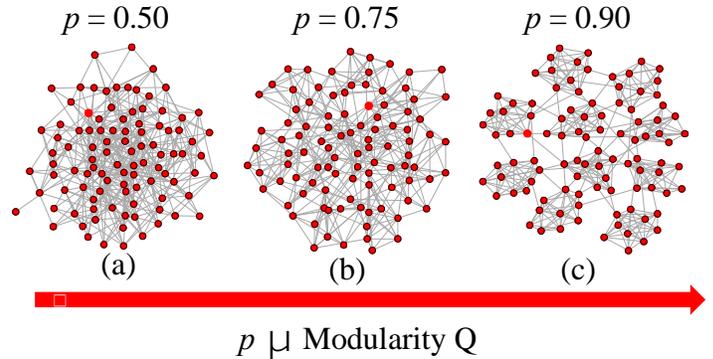
Fig. 3. Examples of modular network created by the parameterized network formation algorithm. The parameterized modular network formation algorithm can change optimal modularity Q of the network by changing network modularity control parameter (NMCP), p. For these three modular networks, the number of modules is 10, the size of each node is 10 nodes and the total number of links is 400. (a) Modular network when p=0.50, 200 links for inner-module links and 200 links for the inter-module links. (b) Modular network when p=0.75, 300 links for inner-module links, and 100 links for the inter-module links. (c) Modular network when p=0.90, 360 links for inner-module links and 40 links for the inter-module links.

According from the equation (7), an index to quantify the importance from the non-largest eigenvalues is defined as follows,

Then, using the proposed network formation method, we constructed modular networks as changing the value of p from 0.5 to 0.9 by 0.05. 10,000 modular networks for a given p are constructed and calculated the optimal modularity Q and the proposed index. Then, the average values of these values are plotted in figure 4.

(a)
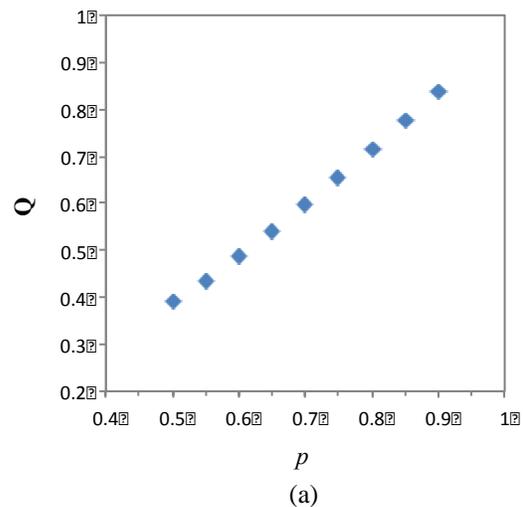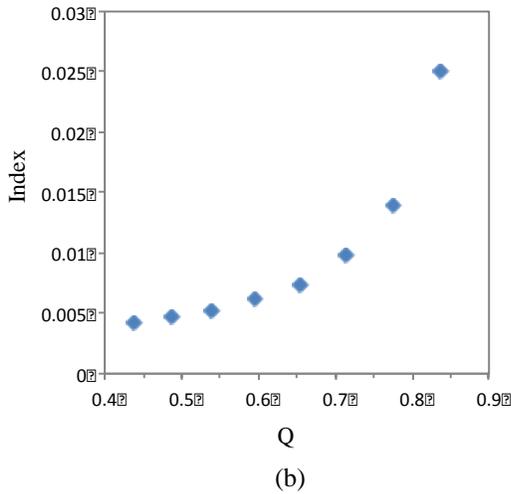
Fig. 4. (a) Relationships between the parameter p in the proposed network formation algorithm and the average value of optimal modularity Q, (b) Relationships between averaged optimal modularity Q and the average value of Index

The figure 4 shows results for the modular networks with 400 nodes and 2,000 links created by the proposed network formation method. Figure 4-(a) shows the relationship between the parameter p and the averaged value of the optimal modularity Q. Also, figure 4-(b) shows the relationship between the averaged optimal modularity Q and the average value of the index. As can be seen these figure, optimal modularity Q linearly increase as the value of p in the proposed network formation method increase. Also, the average value of the Index exponentially increases as the average value of optimal modularity Q increase, which verifies the fact that the importance from the non-largest eigenvalues in diffusion dynamics increase as the optimal modularity Q of the network increase.

V. NODE-LEVEL EIGENVALUE INFLUENCE INDEX

In this section, we investigate how does modular structure affects the spectral properties in the networks. Our proposed an index the Node-level Eigenvalue Influence Index (NEII), ω, that can quantify the influences from an arbitrary eigenvalue, λ_k, to the dynamical process on each node.

The equation (5) indicates that the significance of the contributions from an arbitrary kth eigenvalue to increase the infection probability on a node i is governed by the value of $u_{ki}\|\mathbf{u}_k\|$. Therefore, we defined the Node-level Eigenvalue Influence Index (NEII) $\omega_{ki} \equiv u_{ki}\|\mathbf{u}_k\|$ and investigate the ω_{ki} on each node. In the previous literatures [28], the localized-delocalized phenomenon of eigenvalues is measured by the inverse participation ratio (IPR). If IPR is large, the infections diffuse only within the small confined area, and vice versa. However, the IPR does not distinguish positive or negative, so that the IPR is only applicable for the largest eigenvalue. On the other hand, NEII can distinguish positive or negative influence from all eigenvalues and can be applied to the analysis of the impacts from the all eigenvalues. According to Perron-Frobenius theory, the only eigenvalue in which all elements in the corresponding eigenvector are non-negative is

the largest eigenvalue λ_1. In other words, the other eigenvectors corresponding to the other non-largest eigenvalues have negative elements, which means that the corresponding ω_{ki} contributes to decrease the probability of infection on the node i if ω_{ki} is negative.

Fig. 5 shows the shows the distribution of ω_{ki} on each node in the benchmark toy network. The benchmark toy network consists of four different size star networks connecting each other via the four-nodes complete graph at the center, as shown in Fig. 6 As can be seen in Fig. 5 and Fig. 6, ω_{1i} for the largest eigenvalue is always positive because of the Perron-Frobenius theory. Also, the value of ω_{1i} for the largest eigenvalue and its corresponding eigenvector are positively maximized on node #61 that is the hub node in the largest star graph. The influences from the second largest eigenvalue and the corresponding eigenvector are the positively maximized on the node #31 the hub node in the second largest star graph, but, as can be seen in Fig. 5-(b) that is the enlarged view of Fig. 5-(a), the second largest eigenvalue negatively affect to the node #61 that is the hub node in the largest star graph. In addition to that the third largest eigenvalue positively influence on the node #11 that is the hub of third largest star graph, but negatively affect the hubs of the largest and the second largest star graphs.

Fig.6 visualizes the significance of the value of ω_{ki} for the largest eigenvalue to the fourth eigenvalue ($k = 1$ to 4) by color gradient on each node on network. In these figures, the maximum positive value of ω_{ki} is coloured by the deepest red and gradually changes to green as the relative significance approaches to zero. The minimum negative value of ω_{ki} is coloured by the deepest blue and gradually changes to green as the relative significance approaches to zero. The size of each node is proportional to the absolute value of ω_{ki}.

(a)

(b)
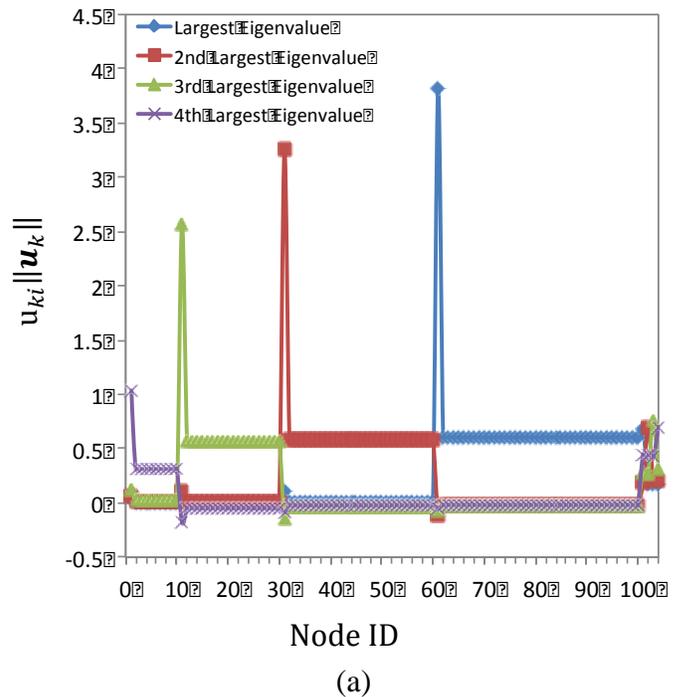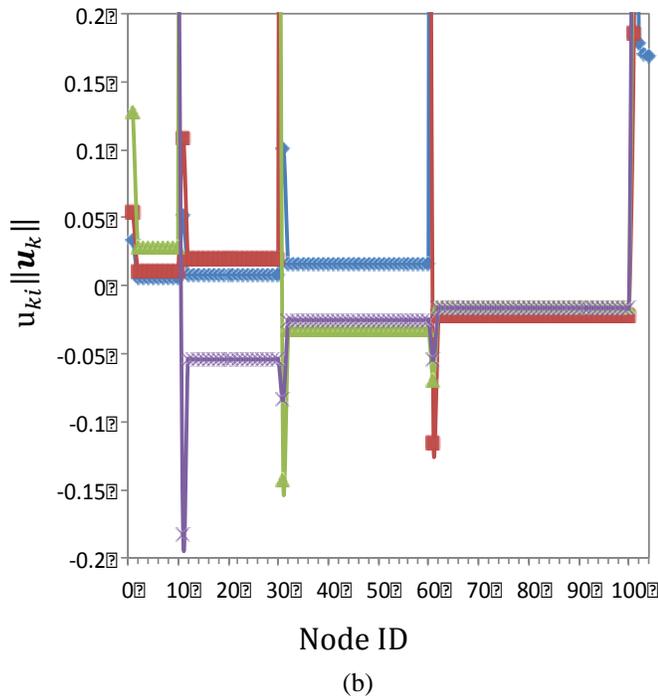
Fig. 5. (a) Distribution of ω_{ki} on each node in the benchmark toy network. (b) Enlarged view
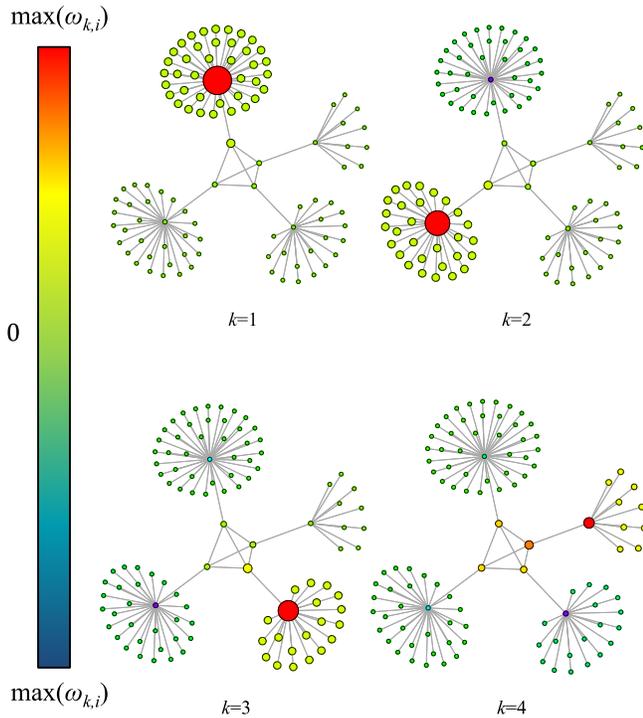
Fig. 6. Visualization the significance of the value of the ω_{ki} by color gradient on each node on network
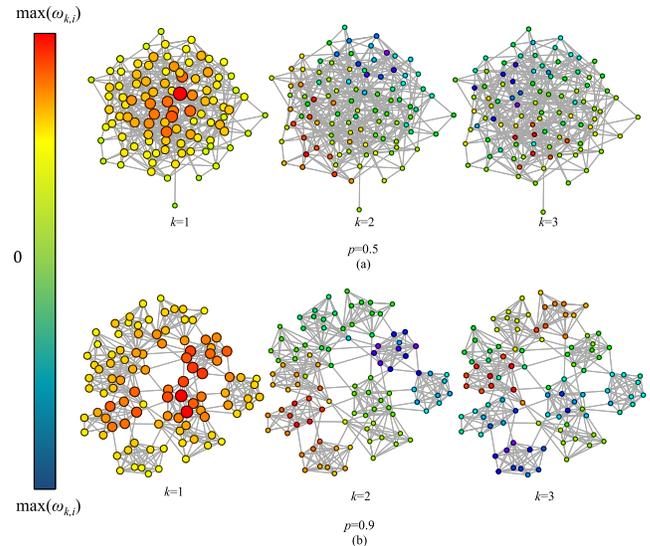
Fig. 7. Distribution and significance of the value of ω_{ki} when p for the proposed parameterized modular network formation algorithm is changed (a) $p = 0.5$ and (b) $p = 0.9$.

Appling this visualization technique to the artificial modular networks crated by the parameterized modular network formation algorithm proposed in the previous section, we observed how the distribution and significance of the value of ω_{ki} varies as modularity of networks increase. Fig. 7 shows the 100-nodes modular networks (the number of modules is 10 and each module size is 10 nodes) created by the proposed network formation algorithm in which each node is colored by the significance of the value of ω_{ki} for $k = 1$ to 4. Figure 7-(a) and (b) correspond with the modular network for $P = 0.5$ and 0.9 respectively. As can be seen in the figures for $k=1$, the color of all nodes is always red because of the positive values of the elements of principal eigenvector due to the Perron-Frobenius theorem. Also, it can be observed that, when the modularity of the network increases, the influences from the non-largest eigenvalues are localized within some modules whether the influences are positive or negative.

VI. REAL MODULAR NETWORKS

In this section, we investigate the NEII, ω_{ki}, in real modular networks that show comparatively high optimal modality Q, such as CNNS and EuroRoad. The results highlights that the needs of consideration of non-largest eigenvalues and corresponds eigenvectors when analyze the diffusion process on the real modular networks. Fig. 8 indicates the colored network by the significance of ω_{ki} for CNNS of which the optimal modularity Q is about 0.84. As shown, the maximum impact is provided by the fourth largest eigenvalue, which fit with the insight in the Fig. 1. Also, in Fig. 9 for EuroRoad of which the optimal modularity Q is about 0.86, the impacts from the second and fourth eigenvalues are significant, which fit with the insight in the figure 1, too.

These results indicate that we need to consider the non-largest eigenvalues and eigenvectors to capture the diffusion dynamics and the well-used approximation method only utilizing the largest eigenvalue is not applicable for the real networks with high optimal modularity.
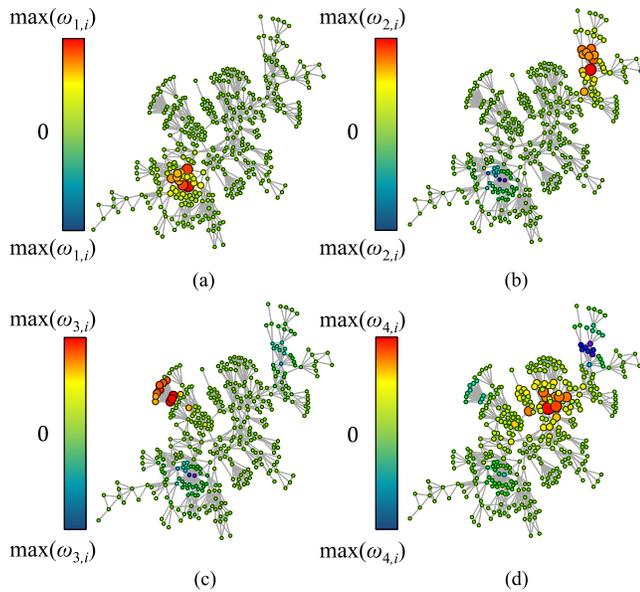
Fig. 8. Visualization of ω_{ki} on CNNS The color gradient and the size of each node indicate the relative significance of the value of ω_{ki} for (a) the largest eigenvalue λ_1, (b) the second largest eigenvalue λ_2, (c) the third eigenvalue λ_3, and (d) the fourth largest eigenvalue λ_4.
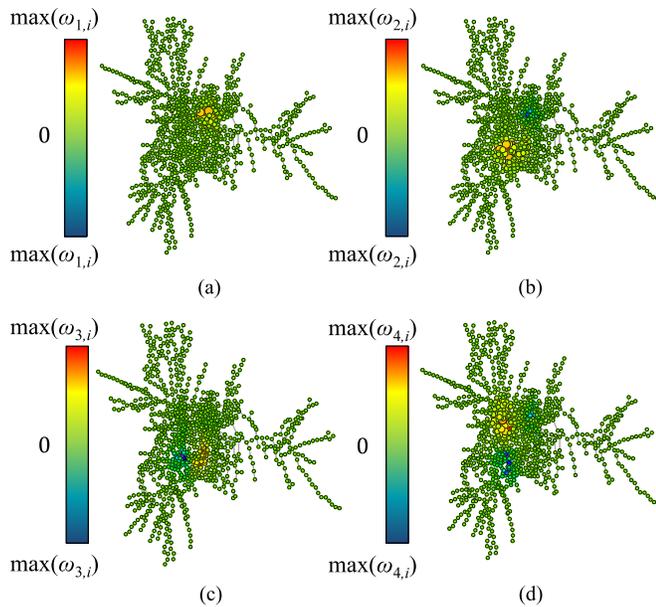
Fig. 9. Visualization of ω_{ki} on EuroRoad The color gradient and the size of each node indicate the relative significance of the value of ω_{ki} for (a) the largest eigenvalue λ_1, (b) the second largest eigenvalue λ_2, (c) the third eigenvalue λ_3, and (d) the fourth largest eigenvalue λ_4.

VII. CONCLUSION

Several studies reported that there exist modular structures in real networks. In this paper, we investigate spectral property of several networks. Also, diffusion phenomena in society have been studied as probabilistic diffusion dynamics on networks. So far probabilistic diffusion dynamics have been analysed in approximated manner only using the largest eigenvalue of the adjacency matrix. But, our investigation of spectral property of modular networks shows that that not only the largest eigenvalue but also the other eigenvalues are critical when analyse the network with high modularity, which verifies by the parameterized modular network formation method and numerical simulations. Furthermore, we investigated the node-level eigenvalue influence index that measures the relative dominance from each eigenvalues and their corresponding eigenvectors on each node, which indicates that the influences from each eigenvalue and corresponding eigenvectors to diffusion dynamics are localized within the modular structures. For our future works, we will extend our spectral analysis to the investigation of the relationship between Laplacian matrix of the networks and probabilistic diffusion dynamics.

REFERENCE

[1] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney, "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters", Internet Math, Vol. 6, No. 1, pp.1-123, 2009

[2] Fortunato, S. (2010), Community detection in graphs, Physics Reports, Vol.486, No.3-5:75-174

[3] Reichardt, J. (2009) Structure in complex networks, Lecture Notes in Physics 766, Springer

[4] Girvan, M ., and M.E.J. Newman (2002) Community structure in social and biological networks, Proceeding of the National Academy of Sciences of the United States of America, Vol99 :7821-7826

[5] Newman, M.E.J, and M. Girvan (2004) Finding and evaluating community structure in networks, Physical Review E, Vol.69, No. 026113

[6] Newman, M.E.J (2004) Fast algorithm for detecting community structure in networks, Physical Review E, Vol.69, No. 066133

[7] Newman, M.E.J. (2006) Finding community structure in networks using the eigenvectors of matrices, Physical Review E, Vol.74, No.3 :036104

[8] Newman, M.E.J. (2006) Modularity and community structure in networks, Proceeding of the National Academy of Sciences of the United States of America, Vol103 :8577-8582

[9] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, "Networks formed from interdependent networks," Nature Physics, vol. 8, no. 1, pp. 40–48, January 2012.

[10] L. Huang, K. Park, and Y-C Lai, "Information propagation on modular networks", Physical Review E, Vol.73, 035103, 2006

[11] A. Saumell-Mendiola, M. A. Serrano, and M. Boguna, "Epidemic spread on interconnected networks", Physical Review E, Vol. 86, 036106, 2012

[12] F. D. Sahneh, C. Scoglio, and F. N. Chowdhury, "Effect of coupling on the Epidemic threshold in intercoonected complex networks: A spectral analysis", arXiv:1212.4194 [physics.soc-ph], 2012

[13] H. Wang et al., "Effect of interconnection network structure on the epidemic threshold", Physical Review E, Vol.88, 022801, 2013

[14] J. Hindes, S. Singh, C. R. Myers, and D. J. Schneider "Epidemic fronts in complex networks with metapopulation structure", arXiv:1304.4310 [physics.soc-ph]

[15] M. Dickison, S. Havlin, and H. E. Stanley "Epidemics on interconnected networks", Phys. Rev. E Vol.85, 066109, 2012

[16] R. Albert, A. L. Barabasi, "*Statistical mechanics of complex networks*". Reviews of Modern Physics, Vol.74, pp.47-94, 2002.

[17] S. N. Dorogovtsev and J. F. F. Mendes, "*Evolution of networks*", Advances in Physics, Vol.51, pp.1079-1187, 2002.

[18] M. E. J. Newman, "*The structure and function of complex networks*", SIAM Review, Vol.45, pp.167-256, 2003.

[19] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "*Critical phenomena in complex networks*", Reviews of Modern Physics, Vol.80, pp.1275- , 2008.

[20] R. Pastor-Satorras and A. Vespignani, "*Epidemic dynamics and endemic states in complex networks*", Physical Reviews E, Vol.63, 066117, 2001.

[21] M. Bogu ñ a , R. Pastor-Satorras, and A. Vespignani, "*Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations*", Vol. 90, 028701, 2003.

[22] J. O. Kephart, S. R. White, "*Directed-graph epidemiological models of computer viruses*", Proceedings of Research in Security and Privacy, pp.343-359 1991.

[23] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "*Epidemic spreading in real networks: An eigenvalue viewpoint*", Proceeding of International Symposium on Reliable Distributed Systems (SRDS'03) 2003, pp.25-34, 2003.

[24] P. V. Mieghem, J. Omic, and R. Kooij, "*Virus spread in networks*", IEEE/ACM Trans. Netw. Vol 17, No. 1, pp.1-14, 2009.

[25] P. V. Mieghem, "*Epidemic phase transition of the SIS type in networks*", Europysics Letters, Vol.97, 48004, 2012.

[26] R. Pastor-Satorras and A. Vespignani, "*Epidemic Spreading in Scale-Free Networks*", Physical Review Letters, Vol.86, 3200-3203, 2001.

[27] A.V. Goltsev, et al., "*Localization and Spreading of Diseases in Complex Networks*", Physical Review Letters. Vol.109, 128702, 2012.

[28] K. Ide, A. Namatame, L. Ponnambalam, F. Xiuju, and R. S. M. Goh, "Spectral Approach for Information Diffusion", Proceeding of ASE SocialCom 2014, In press

[29] J. Billen, M. Wilson, and A. Baljon, "*Eigenvalue spectra of spatial-dependent networks*", Physical Review E, Vol.80, 046116, 2009

[30] M. E. J. Newman, "*The structure of scientific collaboration networks*", Proc. Natl. Acad. Sci. USA, Vol.98, pp.404-409, 2001.

[31] Open dataset is available at: http://www-personal.umich.edu/~mejn/netdata/

[32] D. Greene, and P. Cunningham, "*Producing a Unified Graph Representation from Multiple Social Network Views*", arXiv preprint arXiv:1301.5809, 2013.

[33] Open dataset is available at: http://mlg.ucd.ie/networks/politics-uk.html

[34] L. Šubelj and M. Bajec, "*Robust network community detection using balanced propagation*", European Physical Journal B,, Vol.81, No.3, pp.353-362, 2011.

[35] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson,"*Behavioral Ecology and Sociobiology*", Vol.54, pp.396-405, 2003.

[36] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt and A. Arenas, "*Self-similar community structure in a network of human interactions*", Physical Review E , vol. 68, 065103(R), 2003.

[37] P.Gleiser and L. Danon , Adv. Complex Syst.6, 565, 2003.