HOKKAIDO UNIVERSITY

Title	Scale-free property of local-world networks and their community structures
Author(s)	Furuya, Shuhei; Yakubo, Kousuke
Citation	Physica A : Statistical Mechanics and its Applications, 389(24), 5878-5886 https://doi.org/10.1016/j.physa.2010.09.009
Issue Date	2010-12-15
Doc URL	https://hdl.handle.net/2115/44448
Type	journal article
File Information	PhyA389-24_5878-5886.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Scale-free property of local-world networks and their community structures

Shuhei Furuya, Kousuke Yakubo

Department of Applied Physics, Hokkaido University, Sapporo 060-8628, Japan

Abstract

The scale-free property and community structures of complex networks formed by local events have been studied theoretically, numerically, and empirically. We showed analytically and numerically that the degree distribution function $P(k)$ of the local-world evolving network exhibits a crossover from an exponential to power-law form by increasing the local-world size M. For M much larger than the crossover local-world size $M_{\rm co}$, the distribution function $P(k)$ has a power-law form for any degree $k(\gg 1)$. Below $M_{\rm co}$, however, $P(k)$ obeys a power law for $1 \ll k \ll k_{\rm co}$ and decays exponentially for $k \gg k_{\rm co}$. The crossover size $M_{\rm co}$ and the crossover degree $k_{\rm co}$ have been also elucidated. In addition, we constructed the drug prescription network (DPN) as a real local-world network, in which the local-world subsets are definitely specified, to reveal how the local-world nature affects properties of real-world networks. We found that the community structure of the DPN strongly correlates with the local worlds.

Keywords: complex networks, scale-free networks, local world, community structure
PACS: 89.75Hc, 89.75.Fb, 89.75.Da, 02.50.-r

1. Introduction

It is widely accepted that many complex networks found in fields as diverse as physics, biology, sociology, ecological science, information engineering, and so on, possess the scale-free property, i.e., the distribution function

Email addresses: s-furuya@eng.hokudai.ac.jp (Shuhei Furuya), yakubo@eng.hokudai.ac.jp (Kousuke Yakubo)

of degree k obeys a power law, $P(k) \propto k^{-\gamma}$ [1, 2]. Barabási et al. [3] explained the origin of the scale-free property by introducing the Barabási-Albert (BA) model with the network growth and the preferential attachment. In this model, a newly added node in the growing process has to know degrees of all existing nodes to connect to the network. This is, however, not always realistic especially when the network size becomes large. In order to settle this problem, several network models have been proposed so far, in which the scale-free property is realized based only on local information on network topology [4, 5, 6, 7, 8, 9, 10, 11]. Li and Cheng [4] have studied the relation between the emergence of the scale-free property in the BA model and the range of local information. They proposed the local-world evolving network (LWEN) model in which a newly added node can connect preferentially to nodes only in a subset (local world) of the existing nodes. It has been predicted that the degree distribution of an LWEN shows a transition from an exponential to a power-law behavior depending on the size of the local world. Although statistical properties of the LWEN and its extensions have been studied so far [5, 8, 9, 10], there still remain open questions on the LWEN model. In particular, the transition point of the local-world size above which the network becomes scale free has never been clarified though the transition was argued by several authors [4, 5, 10]. In addition, it is not well understood how statistical properties of real-world complex networks are affected by the local-world nature in their formation rules. This is due to the difficulty in quantifying local-world sizes of real-world networks. Especially, it is important to elucidate the relation between the local-world property and community structures of networks. Although a huge number of works have been devoted to the investigation of community structures of complex networks [12] and some works have argued how modules are developed by local events in evolving network models [7, 13], no attempt has been made to find a relation between the local-world subsets of a real-world network and its modules.

In this paper, we investigate the transition property of the LWEN model [4] by means of a mean field treatment and numerical calculations. We conclude that the change from an exponential to a power-law (scale-free) behavior of the degree distribution function $P(k)$ by increasing the local-world size is not the transition but a *crossover*. If the local-world size is much larger than the crossover size, the network exhibits the scale-free property over the whole range of k ($\gg 1$). Below the crossover local-world size, however, $P(k)$ obeys a power law for $k \ll k_{\rm co}$ and has an exponential form for $k \gg k_{\rm co}$,

where $k_{\rm co}$ is a crossover degree depending on the local-world size and the average degree. We also examine local-world effects in the community structure of a real-world network. Constructing the drug prescription network (DPN) from real prescription data, we evaluate quantitatively the local-world size of the network and discuss how the modules correlate with the local-world subsets.

This paper is organized as follows. In Section 2, we discuss the crossover behavior of the degree distribution function of the LWEN. In Section 3, the DPN is introduced as a real local-world network formed by a rule with the local-world nature, and several basic quantities characterizing the DPN are presented. In Section 4, the DPN is divided into modules by a decomposition scheme based on metaweights. The correlation between these modules and the local-world subsets in the DPN is quantified by introducing a new statistical quantity. Some concluding remarks are finally given in Section 5.

2. Crossover behavior of the local-world evolving network

To construct the local-world evolving network (LWEN) model [4], we start with a small network of a few nodes (m_0) and edges (e_0). At each time step, a node with m edges is added into the existing network system and a local-world subset S_i is defined for this newly added node i. The local-world subset contains M nodes randomly chosen from the existing network. Each of m edges of the new node i is connected to existing nodes with the probability π_{ij} given by

$$\pi_{ij} = \Delta_{j\in S_i} \frac{k_j}{\sum_{l\in S_i} k_l}, \tag{1}$$

where $\Delta_{j\in S_i} = 1$ if the node j is an element of S_i and $\Delta_{j\in S_i} = 0$ otherwise. Repeating these procedures t times, we obtain a LWEN of $m_0 + t$ nodes and $e_0 + mt$ edges.

Regarding k_i and t as continuous variables for their large values, we can evaluate the degree at time t by a mean-field approximation. The time development of the degree of the node i is then described by the rate equation,

$$\frac{\partial k_i}{\partial t} = \frac{mM}{m_0 + t} \Delta_{i\in S(t)} \frac{k_i}{\sum_{l\in S(t)} k_l}, \tag{2}$$

where $S(t)$ is the local-world subset defined for the newly added node at time t. Here, we replace the denominator $\sum_{l\in S(t)} k_l$ in Eq. (2) by its mean value.

The validity of this replacement is evaluated by the central limit theorem. According to this theorem, the distribution of the average degree over a local-world (*sample mean*), $\langle k \rangle_L \equiv \sum_{l \in S(t)} k_l / M$, becomes the normal distribution with the global average (*population mean*) $\langle k \rangle_G$ and the standard deviation $\sigma_{\langle k \rangle_L}$, where

$$\langle k \rangle_G = \frac{2(e_0 + mt)}{m_0 + t}, \tag{3}$$

$$\sigma_{\langle k \rangle_L} = \frac{\sigma_k}{\sqrt{M}}, \tag{4}$$

and σ_k is the standard deviation of the degree k of the entire node set. When $\sigma_{\langle k \rangle_L}$ is much smaller than the global mean $\langle k \rangle_G$, we can approximate $\langle k \rangle_L$ by $\langle k \rangle_G$, and the denominator $\sum_{l \in S(t)} k_l$ in Eq. (2) can be replaced by $\langle k \rangle_G M$. Hereafter, $\sigma_{\langle k \rangle_L}$ is assumed to be smaller than $\langle k \rangle_G$, namely, $M \gg (\sigma_k / \langle k \rangle_G)^2$. It should be noted that the numerator k_i can exceed $\langle k \rangle_G M$ and the quantity $k_i / (\langle k \rangle_G M)$ cannot be regarded as a probability. In such a case, the node i is very likely connected to the newly added node at time t if the node i is selected as a member of the local-world subset $S(t)$. Thus, we can approximate Eq. (2) by

$$\frac{\partial k_i}{\partial t} = \frac{mM}{m_0 + t} \Pi_L \left(\frac{k_i}{\langle k \rangle_G M} \right), \tag{5}$$

where the function $\Pi_L(x)$ is given by

$$\Pi_L(x) = \begin{cases} x & \text{if } x < 1, \\ 1 & \text{if } x \geq 1. \end{cases} \tag{6}$$

If $\langle k \rangle_G M$ is larger than k_i, from Eqs. (3) and (5), we have

$$\frac{\partial k_i}{\partial t} \approx \frac{k_i}{2t}. \tag{7}$$

This equation is the same with the rate equation for the BA model [3]. Thus, the degree distribution function $P(k)$ is proportional to k^{-3} in a large k region. On the other hand, if $\langle k \rangle_G M < k_i$, Eq. (5) becomes

$$\frac{\partial k_i}{\partial t} = \frac{mM}{m_0 + t}. \tag{8}$$

It is easy to show that Eq. (8) leads the exponential degree distribution function, $P(k) \sim \exp(-k/mM)$.

The above argument elucidates how the degree distribution $P(k)$ of the LWEN of size $N = (m_0 + t)$ changes its profile as varying the local-world size M. When M is larger than

$$M_{\rm co} = \frac{k_{\rm max}}{\langle k \rangle_G}, \tag{9}$$

$P(k)$ is proportional to $k^{-\gamma}$ with $\gamma = 3$ over the whole range of k, where $k_{\rm max}$ is the maximum degree of the network with $M = N$ (i.e., the BA network of size N). If $M < M_{\rm co}$, however, the power-law relation $P(k) \propto k^{-\gamma}$ holds only for $k < k_{\rm co}$ but $P(k) \propto \exp(-k/mM)$ for $k > k_{\rm co}$, where $k_{\rm co}$ is given by

$$k_{\rm co} = \langle k \rangle_G M. \tag{10}$$

Thus, the form of $P(k)$ is summarized as

$$P(k) \sim \begin{cases} k^{-3} & , \text{for } N \geq M > M_{\rm co} \\ W(k) & , \text{for } (\sigma_k/\langle k \rangle_G)^2 < M < M_{\rm co} \ , \end{cases} \tag{11}$$

where

$$W(k) \sim \begin{cases} k^{-3} & , \text{for } k \ll k_{\rm co} \\ \exp\left(-\dfrac{k}{mM}\right) & , \text{for } k \gg k_{\rm co} \ . \end{cases} \tag{12}$$

Although our mean-field analysis is not valid for $M < (\sigma_k/\langle k \rangle_G)^2$, $P(k)$ of the LWEN with $M < (\sigma_k/\langle k \rangle_G)^2$ is supposed to have a similar degree distribution function to that of a random graph, because the quantity $(\sigma_k/\langle k \rangle_G)^2$ is of the order of unity and the LWEN with $M = 1$ has the same structural feature with that of a random graph. The profile change of $P(k)$ at $k = k_{\rm co}$ is expected to be a smooth crossover but not the abrupt transition because of a finite standard deviation $\sigma_{\langle k \rangle_L}$ given by Eq. (4). If Eq. (5) is rigorous, the distribution function $P(k)$ changes its profile abruptly at $k = k_{\rm co}$ (and then at $M = M_{\rm co}$). Actually, due to fluctuations in the quantity $\sum_{l \in S(t)} k_l$ in Eq. (2) caused by the finite $\sigma_{\langle k \rangle_L}$, Eq. (5) is just an approximation of Eq. (2), and these fluctuations round off the transition at $k = k_{\rm co}$ (and also at $M = M_{\rm co}$). The analytical form of $W(k)$ near $k = k_{\rm co}$ is, however, difficult to obtain. The crossover size $M_{\rm co}$ is an increasing function of N, because $k_{\rm max}$ increases

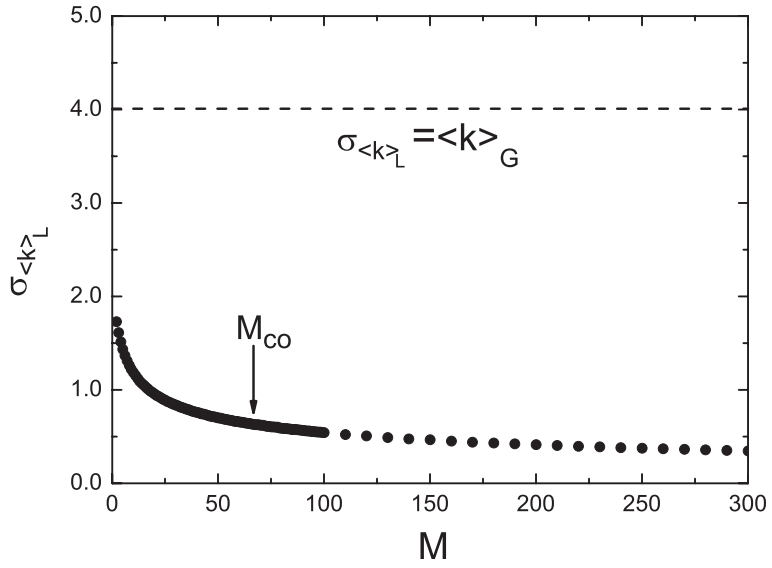

Figure 1: Standard deviation of the averaged degree in a local world of the LWEN as a function of the local-world size. All standard deviations are less than the average degree of the LWEN.

with increasing N while $\langle k\rangle_G$ keeps to be a constant ($= 2m$) for a large N. Since the maximum degree $k_{\max}$ is estimated by $N\int_{k_{\max}}^{\infty} P(k)dk = 1$, we have $M_{\rm co} \propto N^{1/(\gamma-1)} \propto \sqrt{t}$. Therefore, after a long time of the network evolution, the crossover at $M = M_{\rm co}$ becomes sharp because $\sigma_{\langle k\rangle_L}$ for this local-world size $M_{\rm co}$ decreases with time t as $\sigma_{\langle k\rangle_L} \propto t^{-1/4}$.

In order to confirm the above predictions, we calculate the degree distribution function $P(k)$ for the LWEN by varying the local-world size M. In actual calculations, the network size N, the numbers of nodes (m_0) and edges (e_0) in the initial network, and the number of edges (m) of a newly added node are chosen as $N = 10,000$, $m_0 = 3$, $e_0 = 3$, and $m = 2$, respectively. Results are obtained by averaging over 100 samples. The average degree $\langle k\rangle_G$ is 4 in our systems, and the numerically obtained maximum degree $k_{\max}$ is 266 which is the averaged value over 100 samples. Thus, the crossover local-world size given by Eq. (9) is estimated as $M_{\rm co} = 67$. Before discussing the degree distribution, we check the condition for Eq. (5), i.e., $\sigma_{\langle k\rangle_L} \ll \langle k\rangle_G$. The local-world size dependence of $\sigma_{\langle k\rangle_L}$ is shown in Fig. 1. The fact that $\sigma_{\langle k\rangle_L}$ is less than $\langle k\rangle_G = 4$ for any value of M implies the validity of our approximation. Calculated degree distribution functions are

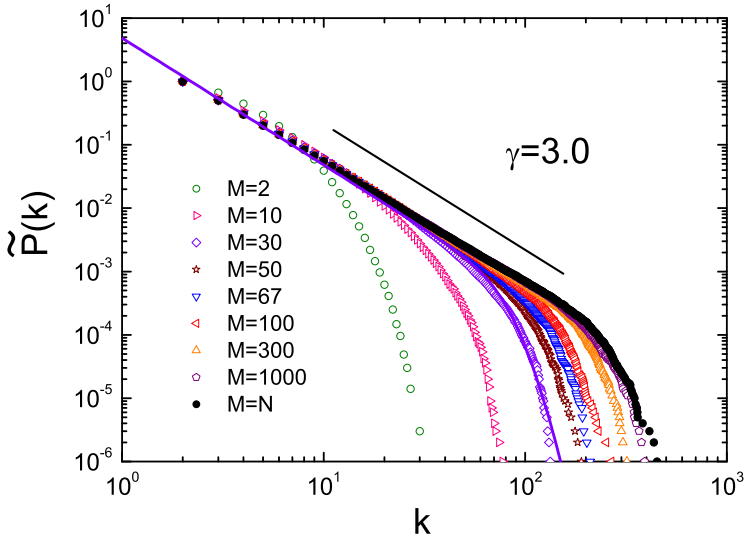

Figure 2: Cumulative degree distributions of LWENs with different local-world sizes. These results are obtained by averaging over 10^2 networks with 10^4 nodes. Solid straight line represents a guide to the eye with the slope corresponding to the non-cumulative degree distribution with $\gamma = 3$. Solid line through the data for $M = 30$ shows the curve fitted by $ak^{-2}/[1 + e^{b(k-c)}]$ which leads the asymptotic forms of $W(k)$ given by Eq. (12), where $a = 4.87$, $b = 0.07$, and $c = 73.5$ are the fitting parameters. Among several candidates of functional forms leading Eq. (12), this functional form gives the best fit for any M.

presented in Fig. 2. These results clearly show that $P(k)$'s for $M \gg M_{\rm co}$ are not largely different from the degree distribution function of the BA model (filled circle) and crossover to exponential forms with decreasing M from $M_{\rm co}$. The reason why $P(k)$ starts to change its profile even above $M_{\rm co}$ is due to a non-negligible $\sigma_{\langle k\rangle_L}$ at $M = M_{\rm co}$ ($\sigma_{\langle k\rangle_L}/\langle k\rangle_G = 0.16$). We see that $P(k)$ for $M < M_{\rm co}$ exhibits the crossover from the scale-free behavior in a small k region to the exponential one in a large k region. The crossover degree $k_{\rm co}$ seems to decrease linearly with decreasing the local-world size M. All these numerical results support our predictions.

The crossover behavior argued in this section can be straightforwardly extended to the scale-free evolving network with *weights*. Pan et al. [5] proposed the generalized local-world evolving network model based on Barrat-Barthélemy-Vespignani (BBV) model [14], and predicted transitions from scale-free to exponential behaviors in the strength and degree distributions.

Considering the rate equations for the weight w_{ij} and the strength $s_i(\equiv \sum_{j\in \mathrm{n.n.}(i)} w_{ij})$ and taking into account that strength s_i can be larger than the average $\langle s\rangle_G M$ of the local-world summation $\sum_{l\in S(t)} s_l$, a similar argument leads crossover behaviors in both the strength and degree distribution functions. It is worthwhile to argue the degree distribution function of a non-evolving network with a local-world property. The fitness model proposed by Caldarelli et al. [15] is known to be a representative non-evolving scale-free network model. We can incorporate the local-world nature into the fitness model by restricting a node to connect other nodes only in a randomly chosen local-world subset of size M. Estimating the degree of this model by a mean-field treatment, it can be shown that the degree distribution function is truncated at $k = 2M$ and obeys a power-law form up to this truncation [16]. Thus, there is no crossover behavior from a scale-free to a small-world regime in the local-world fitness model. We see that the effect of the local-world nature depends on the network formation algorithm.

3. Drug prescription network

Many real-world complex networks are formed by mechanisms based on local information, i.e., local-world mechanisms. It is quite important to clarify the relation between properties of real networks and the local-world nature. As far as we know, however, there is no empirical study on this problem. This is because identification of local-world subsets and quantification of their sizes are difficult in real-world networks. In order to elucidate how the local-world nature affects properties of real-world complex networks, especially community structures, by identifying local-world subsets, we propose the drug prescription network (DPN). In this network, nodes represent sorts of ethical drugs used in a large-scale hospital and two nodes are connected if these are prescribed in one prescription. We construct the DPN based on the prescription data presented by Hokkaido University Hospital in Japan. The data consist of 9,139,617 prescription records from August 1st, 2001 to May 31st, 2008. Each record contains information about the prescription code, the drug code (designated by the Ministry of Health, Labour and Welfare of Japan), the administration period, the dosage, the patient code, the medical department code, and the attending physician code. The drug code distinguishes drug names, actions, components, and shapes of drugs. In the DPN, each drug code corresponds to one node. There are totally 727 nodes and $53,402$ edges in our DPN. The DPN is regarded as a bipartite network as

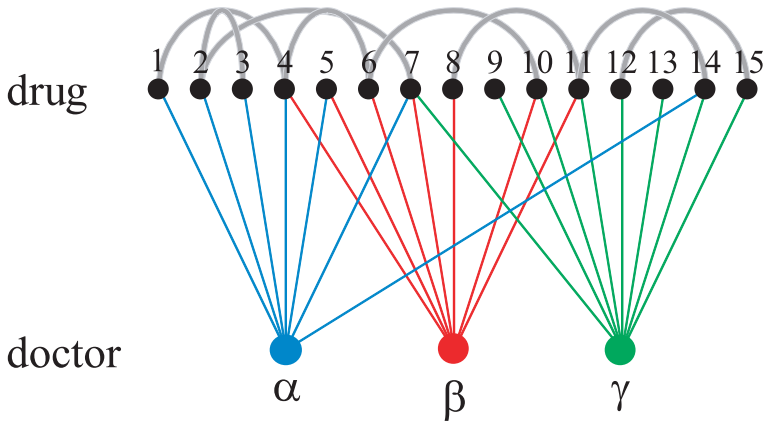

Figure 3: Construction of the DPN. The network is a bipartite network and each doctor provides a local world.

shown schematically in Fig. 3. A simultaneous prescription of the drugs i and j can occur if a prescribing doctor is familiar with both drugs i and j. Each thin line in Fig. 3 represents the relation between a doctor and drugs that the doctor has ever prescribed. For simplicity, we do not distinguish whether drugs are prescribed simultaneously. The drug 1, for example, can connect to the drugs 2, 3, 4, 5, 7, and 14, but never to other drugs. In this sense, a set of drugs that one doctor prescribes comprises a local-world subset in the DPN formation. From the prescription data, we see that the number of doctors $N_{\rm loc}$ is $3,118$ and the average number of drugs $\langle M\rangle$ prescribed by a doctor is 93.9. These quantities $N_{\rm loc}$ and $\langle M\rangle$ are considered as the number of local-world subsets and the average local-world size, respectively.

In the rest of this section, we present fundamental properties of the DPN. The cumulative degree distribution function $\tilde{P}(k)$ is presented in Fig. 4. The distribution $\tilde{P}(k)$ shown in Fig. 4 and its inset seems to have a power-law form in a small k region and decrease exponentially for large k's. In order to find the functional form of the cumulative distribution $\tilde{P}(k)$, we have examined several modeling functions possessing the form of a power law with exponential cut-off. Here, we considered

$$\tilde{P}(k) = \frac{ak^{-(\gamma-1)}}{1+e^{b(k-k_{\rm c})}}\ , \tag{13}$$

and three other similar functions, namely, $\tilde{P}(k) = a\int_k^\infty x^{-\gamma}/[1+e^{b(x-k_{\rm c})}]dx$,

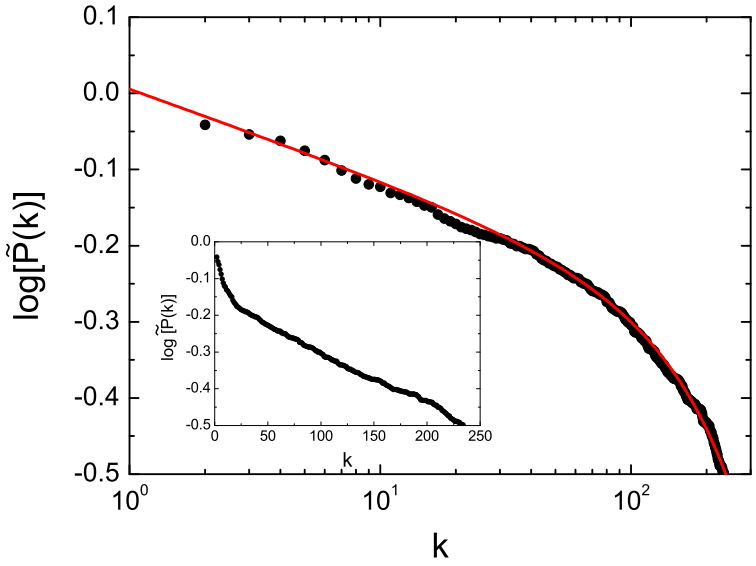

Figure 4: Cumulative degree distribution of the DPN in a log-log scale (main panel) and in a semi-log scale (inset). Solid line represents the curve fitted by Eq. (13) with $\gamma = 1.117$, $k_c = 207.9$, $a = 1.32$, and $b = 0.0056$. The maximum degree of the DPN is 497. For $k \gtrsim 250$, however, the degree distribution is not reliable because of the insufficient number of data due to the finite size effect. we show $\tilde{P}(k)$ only for $k \lesssim 250$.

$\tilde{P}(k) = ak^{-(\gamma-1)}e^{-k/k_c}$, and $\tilde{P}(k) = a\int_k^\infty x^{-\gamma}e^{-x/k_c}dx$, where a, b, k_c, and γ are the fitting parameters. All these functions lead the non-cumulative degree distribution $P(k)$ proportional to $k^{-\gamma}$ for $k \ll k_c$ and having an exponential form for $k \gg k_c$. The maximum likelihood estimation reveals that $\tilde{P}(k)$ given by Eq. (13) provides the best fitting among four models, and leads $\gamma = 1.117 \pm 0.003$ and $k_c = 207.9 \pm 14.9$. These results seem to assert that the degree distribution shows the crossover at $k_c = 207.9$ from the scale-free to the exponential behavior. The crossover region is, however, quite broad due to the small b ($= 0.0056$). Thus, the crossover degree $k_c = 207.9$ is not very meaningful. The inset of Fig. 4 suggests that the crossover actually occurs near $k \simeq 30$. The degree exponent $\gamma \simeq 1.12$ is remarkably small comparing to typical values of γ for other real-world scale-free networks [17]. The scale-free property of the DPN implies the existence of hub nodes. Drugs corresponding to hub nodes include, for example, gastrointestinal medicines and antipsychotics. These drugs are prescribed with many other medications.

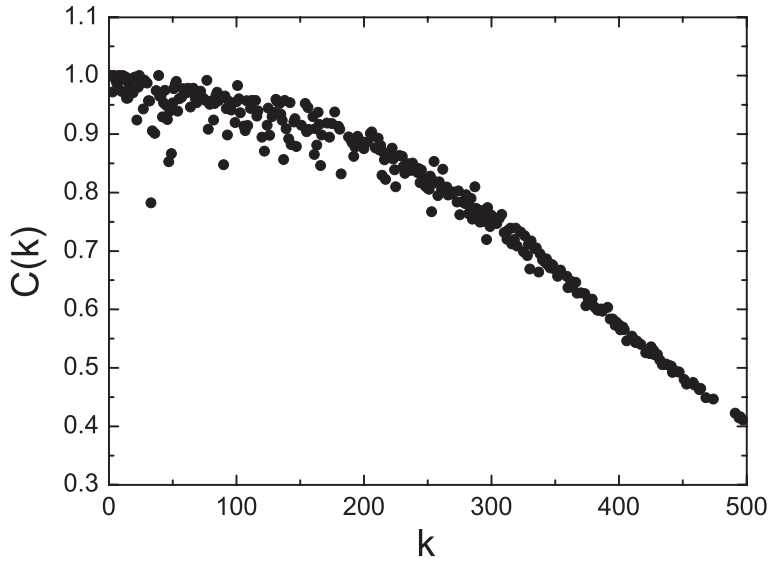

Figure 5: Clustering coefficient of the DPN as a function of the degree k.

The largest degree node (a gastrointestinal medicine) is connected to more than half of the entire nodes. We should note that the crossover behavior of $\tilde{P}(k)$ at $k = k_{\rm c}$ cannot be identified a priori to the crossover $k_{\rm co}$ argued in the previous section because the DPN is not an evolving network. The formation mechanism of the DPN is rather dominated by a *fitness* of each drug. Such a crossover to an exponential degree distribution can be a consequence of a finite-size effect. The origin of the crossover in the DPN must be considered carefully by investigating the formation mechanism of the DPN and its relation to the degree distribution function.

The average node-pair distance $\langle l \rangle$ of the largest cluster of the DPN is only 1.84 and the network diameter $l_{\rm D}$ (the largest node-pair distance) is 4. The clustering coefficient C of this DPN is calculated as $C = 0.78$. The very short $\langle l \rangle$ and the large C indicate that the DPN is a small-world network [18]. Figure 5 represents the clustering coefficient as a function of the degree k. From this result, we see that $C(k)$ is close to unity for $k \lesssim 100$, which implies that low degree nodes form clique structures, and decreases linearly as increasing k for $k \gtrsim 200$. The small $C(k)$ at a large k will be explained soon later in connection with an interpretation of the nearest-neighbor degree correlation.

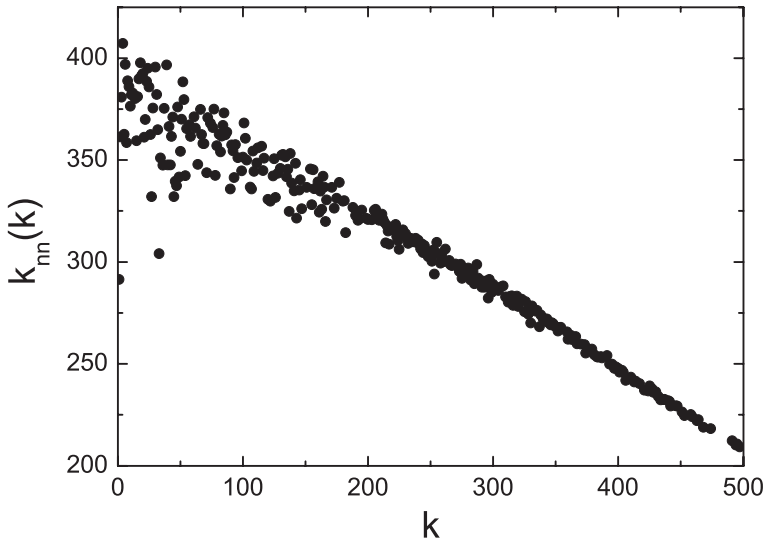

Figure 6: Averaged nearest-neighbor degree of the DPN as a function of the the degree k. The decreasing function $k_{\rm nn}(k)$ indicates the disassortative mixing in the DPN.

The nearest-neighbor degree correlation is characterized by the average degree $k_{\rm nn}(k)$ of neighboring nodes of nodes with degree k [19]. Figure 6 shows that $k_{\rm nn}(k)$ is a decreasing function of k, which indicates the disassortative mixing (repulsive correlations between hubs) in the DPN. This property is related to the fact that prescribed drugs are roughly classified into two classes. One is a set of drugs prescribed for curing diseases or injuries, and the other consists of drugs preventing or reducing side effects of the former ones. As in the case of gastrointestinal or antipsychotic drugs, many of hub nodes represent drugs for preventing side effects, and curative medicines form neighboring nodes of hubs. Therefore, the DPN exhibits the disassortative mixing. In addition, a huge number of curative medicines connected to a hub drug are usually used for independent diseases, and they are not prescribed at the same time. Thus, the clustering coefficient of a hub node is small.

4. Community structure of the DPN and the local-world property

Many complex networks are constructed by relatively weakly connected modules (or communities) in which nodes are tightly connected. It is crucial

to know how a network is composed of such modules and how clearly the community structure is embedded in the network. Thus, the problem of community structures has been one of the most active field in the network science [12]. It is plausible that the community structure of a network is closely related to the local-world property. In fact, it has been argued how modules are developed by local events in evolving network models [7, 13]. However, the relation between the local-world subsets of a real-world network and its community structure has been less understood. Here, we study how strongly the local-world property of the DPN correlates with its community structure.

First, we decompose the DPN into modules. As shown in the previous section, the DPN is a very dense network with large clustering coefficient $C = 0.78$. Therefore, it is difficult to extract modules from the topological structure of the DPN. In this section, we regard the DPN as a weighted network. The weight w_{ij} is defined as the number of simultaneous prescriptions of the drugs i and j. A module in this weighted DPN means a group of drugs that are frequently prescribed simultaneously. In order to find the community structure of the weighted DPN, we employed a metaweighted extremal optimization algorithm. In this method, modules are extracted by the weighted extremal optimization algorithm [20] with replacing the weight w_{ij} by the metaweight w_{ij}^q, where q is a real parameter [21]. Using the metaweight, it becomes possible to extract modules with varying systematically the balance of the importance of topology and weights. Once a module decomposition is done for a specific value of q, we calculate the modularity [22],

$$Q(q) = \frac{1}{2W} \sum_{i,j} \left\{ w_{ij} - \frac{s_i s_j}{2W} \right\} a_{ij} \delta_{m(i),m(j)}, \tag{14}$$

where a_{ij} is the adjacency matrix element, $s_i = \sum_j a_{ij} w_{ij}$ is the strength of the node i, W is the total weight given by $W = \sum_i s_i/2$, and $m(i)$ is the index of the module including the node i. Since the module decomposition depends on q, the Kronecker delta $\delta_{m(i),m(j)}$ (thus, the modularity Q) depends on q. The set of modules giving the maximum $Q(q)$ with respect to q provides finally the community structure of the weighted network. It should be noted that the result provides the high modularity community structure for the original weighted network ($q = 1$) because the modularity $Q(q)$ is defined for the weighted network with $q = 1$ but not for the metaweighted network ($q \neq 1$). For the DPN, the maximum modularity $Q_{\max} = 0.250$ is obtained

at $q = 2.0$, and 32 modules are extracted. The community structure of the DPN is depicted in Fig. 7. Roughly speaking, each module corresponds to a set of drugs used mainly in one medical department.

The community structure of the DPN has been detected here by the metaweighted extremal optimization algorithm based on the modularity maximization. The modularity $Q_{\max}$ of the obtained community structure is, however, not very high. In addition, modularity-based decomposition schemes have problems of the resolution limit and degenerate community structures near the optimized modularity [23, 24]. In order to verify the validity of the community structure shown in Fig. 7, we have decomposed the DPN into modules by an alternative method not based on the modularity, namely, the infomap method proposed by Rosvall and Bergstrom [25], and calculated the *similarity* s between two community structures by the different methods [26]. The large similarity we obtained implies that the community structure is reasonably detected by the metaweighted extremal optimization algorithm.

In order to quantify the correlation between the local-world subsets and the modules in the DPN, we introduce a quantity measuring the strength of the correlation between two families of a general set S. Assume that the importance of each element of S is not equal each other. The weight of the element i is denoted by ω_i. Two families $\Sigma_{\rm a}$ and $\Sigma_{\rm b}$ contain their elements $\{S^{\rm a}_1, S^{\rm a}_2, \cdots, S^{\rm a}_{N_{\rm a}}\}$ and $\{S^{\rm b}_1, S^{\rm b}_2, \cdots, S^{\rm b}_{N_{\rm b}}\}$, respectively, where $S^{\rm a}_\alpha (\alpha = 1, 2, \cdots, N_{\rm a})$ and $S^{\rm b}_\beta (\beta = 1, 2, \cdots, N_{\rm b})$ are subsets of S and intersections of these subsets are generally not null sets. We assume $\cup^{N_{\rm a}}_\alpha S^{\rm a}_\alpha = \cup^{N_{\rm b}}_\beta S^{\rm b}_\beta = S$. At first, we define the fraction of the intersection $S^{\rm a}_\alpha \cap S^{\rm b}_\beta$ in the union $\cup^{N_{\rm b}}_{\beta=1}(S^{\rm a}_\alpha \cap S^{\rm b}_\beta)$, namely,

$$f(\alpha, \beta) = \frac{T(\alpha, \beta)}{\sum^{N_{\rm b}}_{\beta=1} T(\alpha, \beta)}, \tag{15}$$

where $T(\alpha, \beta) = \sum_{i \in S^{\rm a}_\alpha \cap S^{\rm b}_\beta} \omega_i$. It should be noted that the fraction $f(\alpha, \beta)$ is generally not equal to $f(\beta, \alpha)$ while $T(\alpha, \beta) = T(\beta, \alpha)$. Then, we introduce the quantity defined by

$$F^2(\alpha) = \frac{1}{2(N_{\rm b} - 1)} \sum^{N_{\rm b}}_{\beta, \beta' = 1} [f(\alpha, \beta) - f(\alpha, \beta')]^2. \tag{16}$$

The quantity $F(\beta)$ is similarly defined by replacing α, β, β', and $N_{\rm b}$ in the right-hand side of Eq. (16) by β, α, α', and $N_{\rm a}$, respectively. From the above

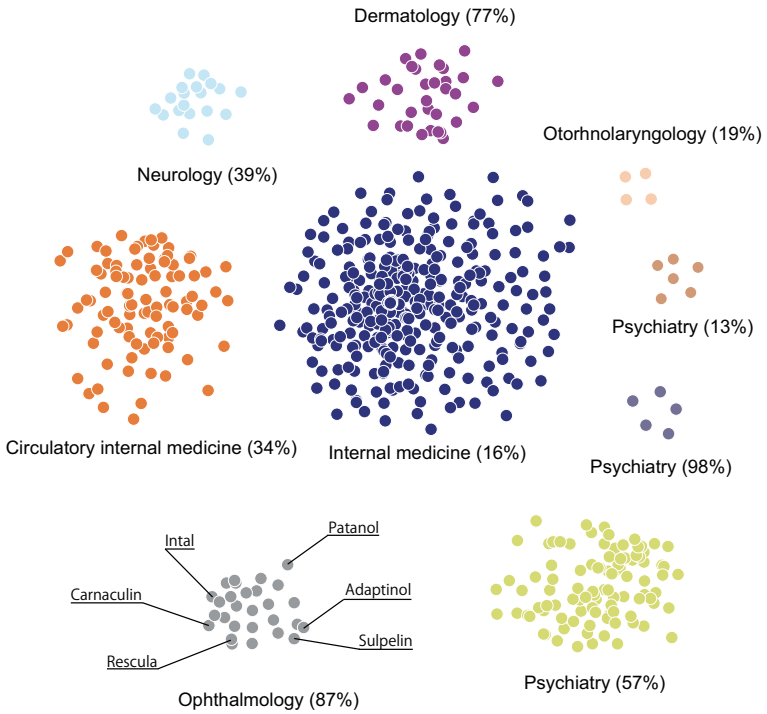

Figure 7: Community structure of the DPN. Different colors represent different modules. Although there exist 32 modules in the DPN, only modules containing more than four nodes are depicted and all edges are removed for clarity. Each module roughly corresponds to a set of drugs used mainly in one medical department as labeled. The percentage in the parenthesis indicates the ratio of the number of prescription times of drugs in the module used in the corresponding department to the total number of prescription times of all drugs in the module.

definition, $F(\alpha)$ takes a large value if the subset $S_\alpha^{\rm a}$ has large intersections with specific subsets of the family $\Sigma_{\rm b}$, and vice versa. The quantity $F(\alpha)$ becomes unity when the subset $S_\alpha^{\rm a}$ is equal to one of the elements of $\Sigma_{\rm b}$ (strong correlation), and $F(\alpha) = 0$ if intersections between $S_\alpha^{\rm a}$ and all of subsets in $\Sigma_{\rm b}$ are the same, namely, $f(\alpha, \beta) = f(\alpha, \beta')$ for any pair of β and β' (no correlation). Averaging this quantity $F(\alpha)$ over the subsets in $\Sigma_{\rm a}$ and over the contributions from $\Sigma_{\rm a}$ to $\Sigma_{\rm b}$ and from $\Sigma_{\rm b}$ to $\Sigma_{\rm a}$, we finally define

$$F = \frac{1}{2N_{\rm a}} \sum_{\alpha=1}^{N_{\rm a}} F(\alpha) + \frac{1}{2N_{\rm b}} \sum_{\beta=1}^{N_{\rm b}} F(\beta). \tag{17}$$

This quantity ($0 \leq F \leq 1$) represents the strength of the correlation between two families $\Sigma_{\rm a}$ and $\Sigma_{\rm b}$. Although the similarity s proposed by [26] also measures how two families of subsets in the whole set S are similar each other, these two quantities, F and s, have different meanings. In particular, F is largely different from s when $S^{\rm a}_\alpha \cap S^{\rm a}_{\alpha'} \neq \emptyset$ for many pairs of α and α'. In an extreme case, for example, $s = (N-1)/N \simeq 1$ while $F = 0$, if $S^{\rm a}_\alpha$ contains all elements except for αth element of the set S with N elements (thus, $N_{\rm a} = N$) and $S^{\rm b}_\beta = S$ ($N_{\rm b} \neq 1$) for any β.

As mentioned in Sec. 3, the number of local-world subsets of the DPN with 727 nodes is 3,118 and the average number of nodes in a local world is 93.9, while the number of modules is 32. This implies that intersections between local-world subsets are huge in the DPN. Thus, the quantity F is suitable to measure the correlation of the local-world subsets and the modules of the DPN. In the analysis of the DPN, two families $\Sigma_{\rm a}$ and $\Sigma_{\rm b}$ in the above general argument are regarded as the set of the local-world subsets and the set of the modules, respectively. In this correspondence, $N_{\rm a} = 3,118$ and $N_{\rm b} = 32$. The element weight ω_i is taken as the number of prescriptions of the drug i. The calculated value of F for the DPN is $F = 0.523$. In order to judge whether this value is high enough to admit the correlation between the local-world property and the community structure, we compare F with that calculated under randomized local-world subsets. The randomized local-world subsets are constructed by randomizing nodes contained in each of local-world subsets with keeping the number of local worlds, the number of nodes in each local-world subset (local-world size), and the weight of nodes ω_i. The community structure is intact. The quantity F for the randomized local worlds is 0.335 ± 0.0007, which is apparently smaller than $F = 0.523$. Therefore, we can conclude that the community structure of the DPN distinctly correlates with the local-world property. This specific example of complex networks with the local-world property suggests that community structures of many real-world networks formed by local events are deeply related to their local-world character.

5. Conclusions

We have studied theoretically, numerically, and empirically the relation between statistical properties of complex networks formed by local events and their local-world nature. First, we examined the local-world evolving network (LWEN) model. It has been reported in the literature that the de-

gree distribution function $P(k)$ of the LWEN shows the transition from an exponential to a power-law behavior by increasing the local-world size M. Our analytical argument, however, shows that the change from the exponential to the scale-free behavior is not the transition but a crossover. Below the crossover local-world size $M_{\rm co}$, the distribution function $P(k)$ obeys a power law for $1 \ll k \ll k_{\rm co}$ and decays exponentially for $k \gg k_{\rm co}$. On the other hand, for $M \gg M_{\rm co}$, $P(k)$ has a power-law form for any degree $k(\gg 1)$. The crossover size $M_{\rm co}$ and the crossover degree $k_{\rm co}$ have been also elucidated. These theoretical predictions have been numerically confirmed. We also discussed the scale-free property of a weighted extension of the LWEN model and a non-evolving network with the local-world property. In order to reveal how the local-world nature affects properties of real-world networks, we constructed the drug prescription network (DPN) from real prescription data, in which the local-world subsets are definitely specified. We found that the community structure of the DPN strongly correlates with the local worlds. Although there are several works discussing how communities are generated by local events in algorithmic network models [7, 13], as far as we know, this work is the first attempt to find a relation between the local-world subsets of a real-world network and its communities.

Although we mainly concentrated on the scale-free property and the community structure of local-world networks in this paper, it is important to study the influence of the local-world nature to other network properties, such as the average node-pair distance, the clustering coefficient, the degree correlation, and/or fractality. It is also crucial to specify and quantify local worlds in many real-world networks formed by local events. These future studies will enable us to understand how the local-world nature affects statistical properties of complex networks in the real world.

Acknowledgements

We thank Hideki Ishii and Akira Endoh (Department of Medical Informatics, Hokkaido University Hospital) for providing the valuable prescription data. We also acknowledge helpful and stimulating discussions with Keita Suzuki (Department of Applied Physics, Hokkaido University). This work was supported by Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (Grant Nos. 1956001 and 19360042) and by the Foundation for the Fusion of Science and Technology. Numerical calculations were performed partially on the HITACHI SR-11000 of Supercomputer

Center, Institute for Solid State Physics, University of Tokyo.

References

[1] R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47.

[2] M. E. J. Newman, SIAM Rev. 45 (2003) 167.

[3] A.-L. Barabási, R. Albert, Science 286 (1999) 509.

[4] X. Li, G. Chen, Physica A 328 (2003) 274.

[5] Z. Pan, X. Li, X. Wang, Phys. Rev. E 73 (2006) 056109.

[6] Q. Sen, D. G.-Zhong, Chinese Phys. B 18 (2009) 383.

[7] X.-J. Xu, X. Zhang, J. F. F. Mendes, Physica A 388 (2009) 1273.

[8] Y. J. Cao, G. Z. Wang, Q. Y. Jiang, Z. X. Han, Physics Lett. A 349 (2006) 462.

[9] Z. Zhang, L. Rong, B. Wang, S. Zhou, J. Guan, Physica A 380 (2007) 639.

[10] L.-N. Wang, J.-L. Guo, H.-X. Yang, T. Zhou, Physica A 388 (2009) 1713.

[11] S. Fortunato, A. Flammini, F. Menczer, Phys. Rev. Lett. 96 (2006) 218701.

[12] S. Fortunato, Phys. Rep. 486 (2010) 75, and references therein.

[13] Z. Xie, X. Li, X. Wang, Physica A 384 (2007) 725.

[14] A. Barrat, M. Barthélemy, A. Vespignani, Phys. Rev. Lett. 92 (2004) 228701.

[15] G. Caldarelli, A. Capocci, P. De Los Rios, M. A. Muñoz, Phys. Rev. Lett. 89 (2002) 258702.

[16] Details of the derivation will be discussed elsewhere.

[17] The small degree exponent is due to the speciality of the DPN. In the DPN, drugs with the same components and the same action but different names or shapes are distinguished by different drug codes. Thus, a drug representing a hub node makes a large group of drugs having the same components but different drug codes, which leads the small degree exponent.

[18] D. J. Watts, S. H. Strogatz, Nature 393 (1998) 440.

[19] R. P.-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. Lett. 87 (2001) 258701.

[20] Y. Fan, M. Li, P. Zhang, J. Wu, Z. Di, Physica A 377 (2007) 363.

[21] S. Furuya, K. Yakubo, Phys. Rev. E 78 (2008) 066104.

[22] M. E. J. Newman, Phys. Rev. E 70 (2004) 056131.

[23] S. Fortunato, M. Barthélemy, Proc. Nat. Acad. Sci. USA 104 (2007) 36.

[24] B. H. Good, Y.-A. de Montjoye, A. Clauset, Phys. Rev. E 81 (2010) 046106.

[25] M. Rosvall, C. T. Bergstrom, Proc. Natl. Acad. Sci. USA 105 (2008) 1118.

[26] P. Zhang, M. Li, J. Wu, Z. Di, Y. Fan, Physica A 367 (2006) 577.