APSITT2008/Copyright 2008 IEICE 07SB0208

RENC: Recursive Estimation of Node Characteristics using Topological Structure of Complex Networks

Kouhei Sugiyama*, Hiroyuki Ohsaki*, Makoto Imase*
*Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita, Osaka, 565-0871, Japan
E-mail: {k-sugi,oosaki,imase}@ist.osaka-u.ac.jp

***Abstract*—In this paper, we propose a recursive estimation method of node characteristics called RENC (Recursive Estimation of Node Characteristics) using the topological structure of a complex networks. RENC reduces the effect of noise by recursively estimating node characteristics using the topological structure of a network. In this paper, we also propose a network generation model called LRE (Linkage with Relative Evaluation). The network generation model LRE is for simulating a social network, in which every node is likely to make decision based on relative evaluation, so that it can reproduce several characteristics of a social network. In this paper, we evaluate the effectiveness of our recursive estimation method RENC by applying RENC to several networks generated with LRE. Consequently, we show that the estimation accuracy of node characteristics can be improved by using our recursive estimation method RENC.**

I. Introduction

In recent years, attention to topological structure of real complex networks has been increasing [1]. For instance, researches on a mechanism determining topological structure of real complex networks and cause of small-world and scale-free structures in real complex networks have been performed [2], [3].

Among those researches analyzing the topological structure of real complex networks, researches on social networks, which represent communications among people, have been actively performed [4]–[6].

Using network analysis techniques, topological structures of several real complex networks (e.g., AS-level topology in the Internet, the hyperlink structure of Web pages, the trackback structure of weblogs, the costarring relation of movie actors, the citation relation of scientific papers, corporate transaction, and co-occurrence relation of words in natural languages) have been actively analyzed (See [1], [7], [8] and references therein).

Such a topological structure of complex networks could be utilized in various ways. For instance, we believe that information retrieval, information analysis, and data mining can be enhanced by utilizing characteristics of the topological structure of complex networks.

For instance, limitation of conventional keyword-based information retrieval systems has been addressed [9]. It is therefore required that a new information retrieval system by utilizing topological structure of a complex network should be realized.

Once characteristics of a node in a network (*node characteristics*) can be estimated from the topological structure of a complex network, estimated node characteristics can be utilized in various ways. Node characteristic are factors that affect the topological structure of a network. For instance, when corporate transactions are viewed as a network [10], popularity and/or reliability of corporations are factors that affect the structure of such a corporate transaction network. Generally, topological structure of a social network is determined by various social activities. We therefore expect that node characteristics (e.g., characteristics of people and/or organizations) are of great importance.

Several estimation methods of node characteristics from topological structure of a complex network are proposed [10]–[12]. In those studies, however, insufficient investigation on what the network structure tells us and how the network structure can be utilized has been performed.

In this paper, we propose a recursive estimation method of node characteristics called RENC (Recursive Estimation of Node Characteristics) using the topological structure of a complex network. RENC reduces the effect of noise by recursively estimating node characteristics using the topological structure of a network.

In this paper, we also propose a network generation model called LRE (Linkage with Relative Evaluation). The network generation model LRE is for simulating a social network, in which every node is likely to make decision based on relative evaluation, so that it can reproduce several characteristics of a social network. In this paper, we evaluate the effectiveness of our recursive estimation method RENC by applying RENC to several networks generated with LRE. Consequently, we show that the estimation accuracy of node characteristics is improved by using our recursive estimation method RENC.

The organization of this paper is as follows. First, related works are summarized in Section II. In Section III, a recursive estimation method of node characteristics called RENC is explained. In Section IV, a network generation model LRE that can reproduce characteristics of a social network is explained. Section IV evaluates the effectiveness of the recursive estimation method RENC. Finally, Section VI concludes this paper and discusses future works.

II. Related Work

Several researches regarding estimation method of node characteristics from the topological structure of a complex network have been performed. In [11], the hyperlink structure of Web pages is regarded as a network, and an estimation method of node characteristics using network structure is proposed. In [11], node characteristic corresponds to popularity and/or importance of a Web page. An estimation method of node characteristic called *PageRank* using topological structure of a network is proposed.

In [12], using the structure of an inter-industry relation network, where a node corresponds to industry and a link to relation between industries, an estimation method of node characteristic is proposed. In [12], node characteristic corresponds to structural superiority of industry. Similar to [12], an index called *SSI (Structural Superiority Index)* is proposed for an inter-industry relation network. In the inter-industry relation network, SSI indicates how advantageous a certain industry is as compared with other industries from a network structural viewpoint.

In [10], an estimation method of node characteristic is proposed for a corporate transaction network, where a node corresponds to a corporation and a link corresponds to corporate transaction. In [10], node characteristic corresponds to superiority of corporation. An index called *SSI (Structural Superiority Index)* is proposed for a corporate transaction network. In a corporate transaction network, SSI indicates how advantageous a certain corporation is as compared with other corporations.

Our recursive estimation method RENC is not for substituting any of those estimation methods, but for enhancing those estimation methods. Namely, our recursive estimation methods RENC can be applied to those estimation methods [10]–[12]. Application of our recursive estimation method RENC to the estimation method proposed in [10] will be discussed in Section III-B.

III. RENC (Recursive Estimation of Node Characteristics)

In this section, we explain the basic idea and the algorithm of RENC (Recursive Estimation of Node Characteristic), followed by an example application of the recursive estimation method RENC to SSI (Structural Superiority Index) [10].

A. Algorithm

In Section II, several estimation methods of node characteristic using the topological structure of a network are described [10]–[12]. These methods simply utilize the topological structure of an *entire* network; i.e., node characteristic is estimated from information on *all* nodes and links in a network. Node characteristic might be estimated from nodes and links including irrelevant ones, which are noise for estimating node characteristics. We expect that estimation accuracy of node characteristics can be improved by using only nodes and links, which seem to be relevant for estimating node characteristics, rather than all nodes and links comprising of a network.

The idea of recursive estimation method of node characteristics is that node characteristics are estimated from nodes and links, which are relevant for estimating node characteristics, rather than using all nodes and links. Node characteristics are estimated using a subset of the network, where irrelevant nodes and links are removed for estimating node characteristics.

The problem here is how irrelevant nodes and links for estimating node characteristics are identified. For identifying irrelevant nodes and links, RENC utilizes an existing estimation method of node characteristics. RENC first applies an existing estimation method of node characteristics for identifying irrelevant nodes and links, and removes those nodes and links for making a subset of the network. The estimation method of node characteristics is then recursively applied to the subset of the network obtained in the previous step. Thus, node characteristics are estimated recursively while repeatedly removing nodes and links, which are likely to be irrelevant for estimating node characteristics. Such a recursive algorithm is expected to reduce the effect of noise in estimation of node characteristics.

In summary, the algorithm of the recursive estimation method RENC is as follows.

(1) Estimate node characteristics from the topological structure of a network using a conventional node characteristic estimation method (e.g., an estimation method for SSI [10] or PageRank [11]).

(2) Remove irrelevant nodes, which have small node characteristic value (e.g., small PageRank or SSI), and neighboring nodes connected with those irrelevant nodes.

(3) Return to the step (1).

B. Example Application to SSI (Structural Superiority Index)

In what follows, we explain how the recursive estimation method RENC can be applied to an existing estimation method of node characteristics. As an example application of RENC, we use the estimation method of SSI proposed in [10]. We note that application of RENC to the estimation method for PageRank proposed in [11] is not impossible, but application of RENC to the estimation method for PageRank does not always give good result. When RENC is applied to PageRank, noise may be superimposed in process of calculation [11]. This is because the PageRank algorithm requires normalization of a connectivity matrix [11], which must be noise for estimating node characteristics.

SSI (Structural Superiority Index) is an index that shows how advantageous a certain node is as compared with other nodes from a network structural viewpoint [10]. When a link in a network has direction (i.e., network is a directed graph), by assuming that an upstream node is superior to a downstream node, superiority of each node is estimated.

Specifically, SSI of node i in a network, S_i, is defined as

$$S_i \equiv \sum_{j \neq i} \zeta_{ji} (\tilde{S}_j + 1), \tag{1}$$

where $\tilde{S}_j$ is an approximation of node j's SSI, and is defined as

$$\tilde{S}_j \equiv \sum_{k \neq j} \zeta_{kj} (k_k^I + 1), \tag{2}$$

where k_i^I is the input degree of node i. ζ_{ij} is 1 if a link exists from node i to node j, and 0 otherwise. Equation (1) indicates that when k_j^I is large (i.e., node j connected with node i is connected by many nodes), SSI takes a large value.

We then explain how the recursive estimation method RENC can be applied to the estimation method for SSI. In what follows, the number of repetitions of the RENC algorithm is denoted as n, and the estimated SSI of node i at the n-th repetition is denoted as S_i^n.

When RENC is applied to the estimation method for SSI, node characteristics are estimated recursively as follows.

(1) Initialize the number of repetitions n as $n \leftarrow 1$.
(2) Calculate SSI S_i^n from the topological structure of a network.
(3) Remove irrelevant nodes, which have small SSI values S_i^n, and neighboring nodes connected with those irrelevant nodes.
(4) Increment the number of repetitions n as $n \leftarrow n + 1$ and return to the step (2).

IV. Network Generation Model LRE

LRE is a network generation model, which aims to reproduce characteristics of various social networks. LRE utilizes the fact that *people generally build relationship with others based on relative evaluation, rather than absolute evaluation*.

In the literature, several models for generating networks have been proposed [2], [13]. For instance, BA (Barabási Albert) model [13] is one of representative models for generating a scale-free network. Notable features of BA model are network growth and link preferential attachment. In BA model, a connected network with a small number of nodes is created, and nodes are incrementally added to the network. A link is created between a new node and existing one, which is randomly chosen from all existing nodes with a probability proportional to the degree of those nodes.

A network generation model for reproducing characteristics of a social network is proposed in [2]. In this model, similarly to BA model, nodes are incrementally added to the network. A link is created between a new node and existing one with a probability proportional to the attribute of those nodes. Link creation algorithm of those network generation models can be viewed as an algorithm based on *absolute evaluation*; i.e., a newly added node determines importance of each node using global knowledge.

However, in a social network, it is natural to assume that relations among people are built not with absolute evaluation but with *relative evaluation*.

The network generation model LRE mimics: (1) a new node creates a link to an existing node randomly chosen with a probability proportional to its importance, and (2) importance of an existing node is determined not by absolute evaluation but by relative evaluation. With these features, we expect that LRE can reproduce several characteristics of a social network.

In what follows, the number of nodes in a network is N, and the average degree is $\overline{k}$. Characteristic value of node i $(1 \leq i \leq N)$ is U_i. Note that U_i corresponds to a hidden variable of a node in [14].

The network generation model LRE generates a network as follows.

(1) Generate N disjoint nodes.
(2) Randomly choose nodes i and j from N nodes.
(3) Create a link from node i to node j with a probability p_{ij} defined by

$$p_{ij} = \alpha \left(\frac{U_j}{U_i + U_j} \right), \tag{3}$$

where α is a constant.
(4) If the number of links in a network reaches $\overline{k}N/2$, terminate the algorithm. Otherwise, return to the step (2).

Example networks generated with the network generation model LRE are shown in Figs. 1 and 2. Figure 1 shows an example of network generated with LRE for $N = 1,000$, $\overline{k} = 3$, and U_i following the normal distribution $N(100, 900)$. Figure 2 shows an example of network generated with LRE for $N = 1,000$, $\overline{k} = 4$, and U_i following the normal distribution $N(100, 900)$.

V. Evaluation

In this section, we evaluate the effectiveness of RENC by applying RENC to networks generated with LRE. Specifically, we evaluate the effectiveness of our recursive estimation method RENC in terms of estimation accuracy of node characteristics.

In the following results, our recursive estimation method RENC is applied to the estimation method for SSI. As a metric for evaluating estimation accuracy of node characteristics, we introduce an index defined as

$$S_k \equiv \sum_{l=1}^{k} U_{v(l)}, \tag{4}$$

where $v(l)$ represents a node whose rank is l (i.e., the node characteristic value of node $v(l)$ is l-th largest among

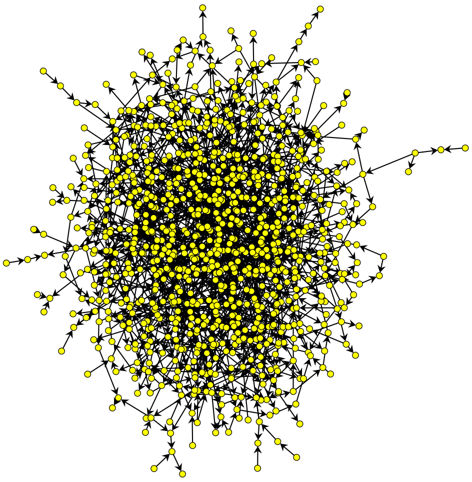

Figure 1: Example network generated with LRE for $N = 1,000$, $\overline{k} = 3$, and U_i following the normal distribution $N(100, 900)$

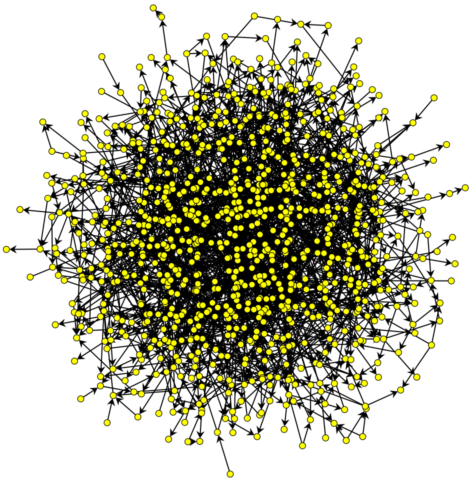

Figure 2: Example network generated with LRE for $N = 1,000$, $\overline{k} = 4$, and U_i following the normal distribution $N(100, 900)$

those of all nodes). S_k takes a larger value when estimation accuracy of node characteristics is higher.

Two networks are generated with LRE for $N = 1000$, and $\overline{k} = 3$ or 5, and U_i following the normal distribution $N(100, 900)$.

Numerical results for a network with $\overline{k} = 3$ are shown in Figs. 3 through 7. These figure shows S_k (labeled as *SSI w/ RENC*), in which $v(l)$ is determined by the estimated node characteristic value, as a function of k. Our recursive estimation method RENC is repeated five times, and S_k at each repetition is shown in Figs. 3 through 7, respectively. Numerical results for a network with $\overline{k} = 5$ are shown in Figs. 8 through 12. For comparison purposes, S_k obtained with the estimation method [10] (labeled as *SSI*) and the ideal values of S_k are plotted in all figures. Note that these values are not affected by the number of repetitions of the RENC algorithm.

One can find from these figures that values of S_k with RENC are larger than those of S_k without RENC when the number of repetitions n is large. This indicates that estimation accuracy of node characteristics with RENC is higher than that without RENC when the number of repetitions n is large. Such high accuracy is resulted from the effect of reducing noise with our recursive estimation method RENC. One can also find that the estimation accuracy is well improved at the fourth repetition (i.e., $n = 4$), which demonstrates the effectiveness of recursive estimation. It should be noted that although results are not included here, we observed similar results when the node characteristic value follows the Pareto distribution instead of the normal distribution.

VI. Conclusion and Future Works

In this paper, we have proposed a recursive estimation method of node characteristics called RENC (Recursive Estimation of Node Characteristics) using the topological structure of a complex network. We have also proposed a network generation model called LRE (Linkage with Relative Evaluation) which aimed to reproduce several characteristics of various social network. We have evaluated the effectiveness of our recursive estimation method RENC by applying RENC to several networks generated with LRE. Consequently, we have shown that the estimation accuracy of node characteristics could be improved by using our recursive estimation method RENC.

As a future work, we need more investigation on the effectiveness of our recursive estimation method RENC. In particular, it would be of importance to apply RENC to several real complex networks for estimating their node characteristics.

Acknowledgements

The authors would like to thank Mr. Junichi Murayama, Mr. Takeshi Yagi of NTT Corporation, Japan for their kind support and valuable discussions.

References

[1] R. Albert and A.-L. Barabasi, “Statistical mechanics of complex networks,” *Reviews of Modern Physics*, vol. 74, no. 47, Jun. 2002.

[2] M. Jin, M. Girvan, and M. E. J. Newman, “The structure of growing social networks,” *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 64, no. 026118, pp. 381–399, Oct. 2001.

[3] D. Liben-Nowell and J. Kleinverg, “The link prediction problem for social networks,” in *Proceedings of the 12th ACM International Conference on Information and Knowledge Management (CIKM’03)*, Jun. 2003, pp. 556–559.

[4] G. F. Davis, M. Yoo, and W. E. Baker, “The small world of the american corporate elite, 1982-2001,” *Strategic Organization*, vol. 1, no. 3, pp. 301–326, Aug. 2003.

[5] E. Holger, M. Lutz-Ingo, and B. Stefan, “Scale-free topology of e-mail networks,” *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 66, no. 035103, pp. 1–4, Sep. 2002.

[6] M. Granovetter, “The strength of weak ties: a network theory revisited,” *Sociological Theory*, vol. 1, pp. 201–233, Sep. 1983.

[7] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature (London)*, vol. 393, pp. 440–442, Oct. 1998.

[8] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Viesek, “Evolution of the social network of scientific collaborations,” *Physica A*, vol. 311, pp. 590–614, Jan. 2002.

[9] H. A. Kautz, B. Selman, and M. A. Shah, “The hidden Web,” *AI Magazine*, vol. 18, no. 2, pp. 27–36, 1997.

[10] K. Sugiyama, O. Honda, H. Ohsaki, and M. Imase, “Application of network analysis techniques for Japanese corporate transaction network,” in *Proceedings of 6th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT 2005)*, no. 9–10, Nov. 2005, pp. 387–392.

[11] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the Web,” Stanford Digital Library Technologies Project, Tech. Rep., 1998.

[12] Y. Yasuda, *Network analysis — What decides activity*. Sinyosha, Feb. 1992, (in Japanese).

[13] A. L. Barabsi and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, pp. 509–512, Oct. 1999.

[14] A. Fronczak, P. Fronczak, and J. A. Holyst, “Average path length in random networks,” *Physical Review E*, vol. 70, Nov. 2004.

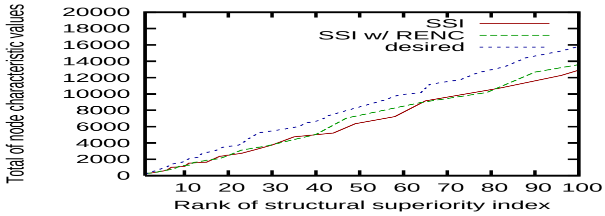

Figure 3: Relation between rank of node and total of node characteristic value ($n = 1$)

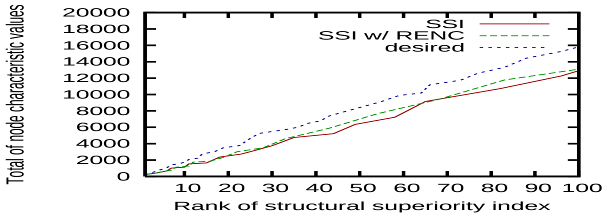

Figure 4: Relation between rank of node and total of node characteristic value ($n = 2$)

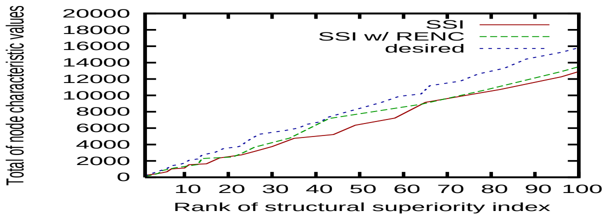

Figure 5: Relation between rank of node and total of node characteristic value ($n = 3$)

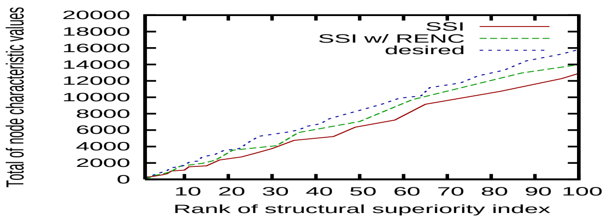

Figure 6: Relation between rank of node and total of node characteristic value ($n = 4$)

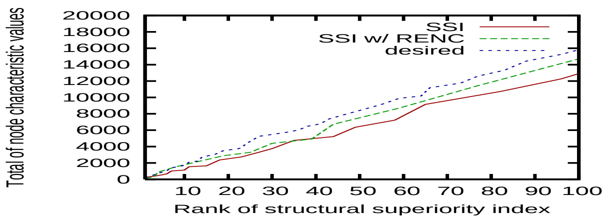

Figure 7: Relation between rank of node and total of node characteristic value ($n = 5$)

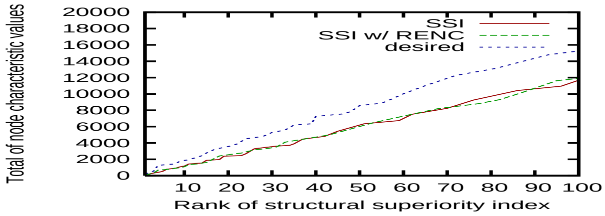

Figure 8: Relation between rank of node and total of node characteristic value ($n = 1$)

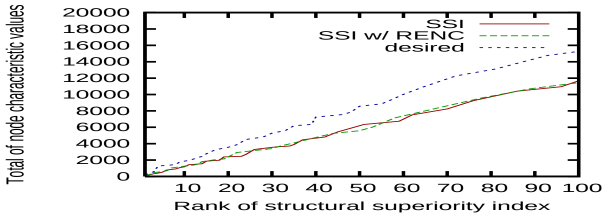

Figure 9: Relation between rank of node and total of node characteristic value ($n = 2$)

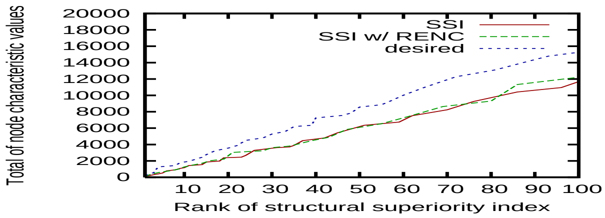

Figure 10: Relation between rank of node and total of node characteristic value ($n = 3$)

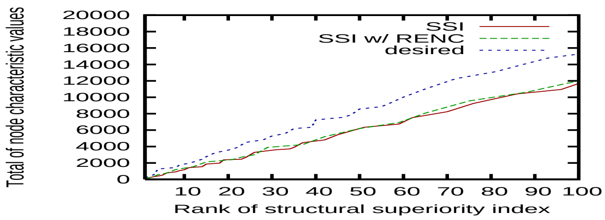

Figure 11: Relation between rank of node and total of node characteristic value ($n = 4$)

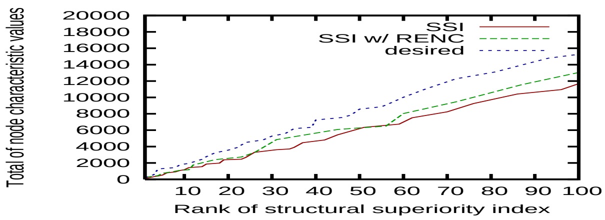

Figure 12: Relation between rank of node and total of node characteristic value ($n = 5$)