Analysis for Betweenness Centrality in Social Network Models

Kazuya Kawamoto^a, Jun Matsukubo^a, Yukio Hayashi^b

^aKitakyushu national college of technology, Shii, Kokuraminami, Kitakyushu, Fukuoka, Japan

^bJapan advanced institute of science and technology, Asahidai, Nomi, Ishikawa, Japan

*Corresponding Author: d30913kk@apps.kct.ac.jp

Abstract

In the real social network, it's often important to identify nodes which have huge influence. For searching such node, Betweenness has been proposed, which is a fraction of shortest paths through a node. However, to obtain Betweenness needs huge cost of calculation in large-scale networks. Therefore in recent years, an approximation for Betweenness has been proposed.

In this study, we calculate Betweenness on network models to imitate real social networks. In addition, we analyze results of Betweenness by both the above approximate method and conventional one.

As a result, Betweenness on the above network models tends to be converged on few nodes in the case that high degree nodes tends to connected each other and precision of the approximation depends on hop number, which limits search ranges.

Keywords: network, Betweenness centrality.

1. Introduction

In actual social networks, such as coauthor relations of scientists, costarring relations of movie stars, acquaintance relations of company managers, SNS, it is often important in deals, negotiations, suggestions of new plans to identify person (node) who have huge influence for business.

In actual network such as the above, each node builds links independently, and whole of network is constructed. In other words, such networks are self-organized. As a common feature of many self-organized networks is that degree distribution follows power law. The degree distribution following power law expresses that enormous nodes with only few degrees is mixed with very few hubs.

Recently, as an index to find important elements in

networks, Betweenness centrality has been proposed, which is fraction of shortest paths through a node. However, to obtain Betweenness, it is necessary to search for shortest paths for all pair of nodes. Therefore we need high calculation cost for Betweenness in large-scale networks [1].

In late years an approximation to obtain Betweenness in large-scale networks is proposed. This approximation estimates Betweenness in the whole network from range-limited Betweenness in a restrictive range under the assumption that the number of nodes possible to search increases exponentially for hop number.

On the other hand, whereas, in social networks, hubs (node with high degree) tends to connect each other, while in technical networks such as the Internet, power grids and so on, hubs and nodes with few links tend to connect.

Because the number of nodes possible to search can change with the same hop number if relation of degrees between adjacent nodes are different, the node number can influence to approximation of Betweenness.

In this report, we analyze a property of Betweenness and approximate one for the relations of the degree between adjacent nodes by computer simulations. As a result, it shows that nodes which have high approximate values of Betweenness tend to have high true one. This indicates we can discover central elements with less calculation cost by our proposed method than one for true values of Betweenness.

The paper is organized as follows: Section 2 introduces generation models of the self-organized network in this study. Section 3 gives a calculation of approximate Betweenness with an extrapolation of $Bl(u)$ concretely, which is necessary for implementation and shows results of both our proposed method and the original Betweenness in numerical experiments. And finally, Section 4 gives an abstract.

2. Generation model of the network

2.1 Generation model of a self-organized network

The degree distribution of a self-organized network obeys power law. As a generation model of such networks, BA model is proposed. The BA model is constructed through two procedures; "addition of node" and "preferential destination search." In this model, nodes with various degrees are connected each other. Now, we explain *configuration model* that can regulate relations of degrees between adjacent nodes in the next section

2.2 The generation method of the configuration model

The degree distribution of the configuration model is determined like BA model. We can regulate probability of connection between nodes with the same degree by a parameter a (It regulate probability to connect nodes each other). Let us denote the number of the nodes by N . The generation method is as follows:

- (1) Determine degrees of each node of network same as BA model (Do not connect nodes each other).
 - (2-1) Select randomly two nodes i, j those have the numbers of connected links less than their degrees.
 - (2-2) Connect them by the next probability if they are not connected
- $$\frac{1}{1 + a\Delta_{ij}} \quad (1)$$
- (Δ_{ij} : Difference of the rank of node's degree. The node with the highest degree is the first place)
- (3-1) If it failure to connect in procedure (2) consecutively N times, select link randomly and let l the link.
 - (3-2) Let u, v terminal of l , respectively. Disconnect u and v .
 - (3-3) Select two nodes that the numbers of connected links are less than their degrees randomly and connect each node to u, v with probability of equation (1).
 - (4) Repeat from the procedure (2-1) until degree distribution of the whole network is satisfied.

3. Approximation of Betweenness

For BA model and configuration model, we performed approximation in the following procedures.

- (1) Calculate the average of node number possible to search: $\langle Z_l \rangle$ with hop number $l = 1, 2, \dots, M$.
- (2) Approximate $\langle Z_l \rangle$ by the next expression and calculate parameters A and λ using the least squares method.

Table 1: Parameters

Parameter	Value
The number of the vertexes N	2000
Exponent	-3
Mean degree	3
The number of the samples	10
Number of the maximum search hops M	5

Table 2: Experiment conditions

Processor	AMD A10-6800K 4.1GHz×4
Memory	15.1GB (DDR3-2133)
OS	Ubuntu 12.04 LTS

$$\langle Z_l \rangle = \{A - (\langle Z_l \rangle - 1)\}^{[1 - e^{\lambda(-l-1)\lambda}]}$$

$$+ (\langle Z_l \rangle - 1) \quad (2)$$

- (3) L^* is the hop number l , when $\langle Z_l \rangle$ is equal to N .
 - (4) Approximate Betweenness $B_l(u)$ by the next expression to calculate parameter $C(u)$ using the least squares method, where $l = 1, 2, \dots, M$.
- $$B_l(u) = \{C(u) - (B_1(u) - 1)\}^{[1 - e^{\lambda(-l-1)\lambda}]}$$
- $$+ (B_1(u) - 1) \quad (3)$$
- (5) Extrapolate $B_l(u)$ and calculate approximate value of Betweenness $B^*(u) = B_{L^*}(u)$.

In this research, we also tested the method to calculate $\lambda(u)$ about each node u in procedure(2) without taking average about all nodes in procedure(1).

4. Analyze of our approximation

In this chapter, we analyze true and approximate value of Betweenness in BA model and configuration model. We show parameters of network models in Tab.1 and a computer specification which we use for simulations in Tab.2.

As a measurement to indicate relation of degrees between adjacent nodes, *assortativity* [3] is proposed. When assortativity is a positive value, it is frequently that nodes with same degree connect with each other. On the contrary, when assortativity is a negative value, it is frequently that nodes with different degree connect. In BA model and configuration model, assortativity was 0 and 0.45 respectively. It is revealed that in the BA model nodes with various degrees connect with each other and in configuration model nodes with same degrees connected.

Table 3: Indispensability calculation time and accuracy

Model	$t_a[s]$			$t[s]$	$\langle (B^*(u) - B(u)) / B(u) \rangle$
	calculation time of $Z_l(u)$	calculation time of $B^*(u)$	total time		
BA	211	26	237	212	0.0560
BA specific λ	202	31	233	203	0.0962
configuration	59	3	62	190	0.8351
configuration specific λ	57	5	62	185	0.8659

t_a Time to calculate with our approximation of 2.2 section

t Time to calculate with Brandes' algorithm

$\langle |(B^*(u) - B(u)) / B(u)| \rangle$ Average of relative error about all node u

Table 4: Average of the number of the vertexes within the l hop

l		1	2	3	4	5
$\langle Z_l \rangle$	BA	6	71	545	1633	1993
	configuration	6	35	114	258	457

Fig. 1. Comparison of $B(u) - B^*(u)$ (BA model) Fig. 2. Comparison of $B(u) - B^*(u)$ (configuration model)

BA ($L^* = 5.41$)

configuration ($L^* = 69.4$)

Fig. 5. curves of $\langle Z_l \rangle$ to l

BA ($L^* = 5.41$)

BA λ for individual node ($L^* = 5.41$)

Fig. 6. Curves of $B_l(u)$ to l (BA model)

configuration ($L^* = 69.4$)

configuration λ for individual node ($L^* = 69.4$)

Fig. 7. Curves of $B_l(u)$ to l (configuration model)

4.1. Time necessary to calculate and accuracy of approximation

We show time necessary to calculate and accuracy of approximation in Tab.3. In BA model, our approximation takes more time than Brandes' algorithm, however accuracy of approximation seems to increase. On the other hand, in configuration model, time for approximation shorten, however the accuracy decreases, because in configuration model, there is less number of nodes possible to search in M hop than in BA model. Tab.4 also shows this fact.

4.2. Comparison between approximate and true value of betweenness

We show results of comparison between approximate value of betweenness, $B^*(u)$ and true one, $B(u)$ of each model. $B^*(u)$ in BA model tend to be slightly bigger than $B(u)$ with common λ and smaller with individual λ as in Fig.1 and 2. However, in the both cases, there seems to be little replacement of rank to nodes with high true value of Betweenness.

Furthermore, when we compare both, betweenness of node with high rank in configuration model is twice as large

as one in BA model. In the other words, it becomes more frequently that shortest paths through specific nodes in configuration model than in BA model. Thus, it can be thought that link structure of configuration model includes coarseness and minuteness more clearly than BA model.

4.3. Comparison between approximate curve and true value

We show curves of $\langle Z_l \rangle$ to l in Fig.5 and curves of $B_l(u)$ to l in Fig.6 and 7.

According to Fig.5, a curve to approximate $\langle Z_l \rangle$ is almost consistent with original $\langle Z_l \rangle$. However according to Fig.7 and 8, approximate curves of $B_l(u)$ are slightly different from original $B_l(u)$.

Meanwhile, in Ref. [1], using a monotone increasing function, an approximation for $B_l(u)$ is consistent with original $B_l(u)$; nevertheless $\langle Z_l \rangle$ and $B_l(u)$ converge on a certain value, respectively.

The fact indicates that there are important elements for accurate approximation of $B_l(u)$ except accommodating a form of approximation function with both of original $\langle Z_l \rangle$ and $B_l(u)$. So, we can give a task to improve a function of approximate curve as a future work.

5. Conclusions

In this report, we analyzed properties with both of true Betweenness and approximate one for a relation of degrees between adjacent nodes by computer simulations. As results, about BA model, we can calculate approximate Betweenness with high accuracy by our method. Moreover, the highest Betweenness in configuration model is higher than that in BA model. In addition, high ranks of approximate Betweenness coincidence mostly high ranks of true one.

References

-
- [1] Mária Ercsey-Ravasz, Ryan Lichtenwalter, Nithsh V. Chawla and Zoltán Toroczkai : “Range-limited Centrality Measures in Complex Networks”, PHYSICAL REVIEW E, Vol. 85, No.066103, 2012
 - [2] Réka Albert and Albert-László Barabási : “Statistical mechanics of complex networks”, REVIEWS OF MODERN PHYSICS, Vol. 74, pp. 47-97, 2002
 - [3] M. E. J. Newman : “Mixing patterns in networks”, PHYSICAL REVIEW E, Vol.67, No.026126, 2003