Modified Betweenness Centrality to Identify Relay Nodes in Data Networks

Masaaki Miyashita and Norihiko Shinomiya,
Graduate School of Engineering,
Soka University,
Tokyo, Japan
Email: shinomi@ieee.org

***Abstract*—Several types of data networks require relay nodes to transmit data because the nodes would impact on services with the networks. Betweenness centrality is one of the measures that reveal important nodes for a network topology. However, the measure does not specialized in relay nodes of survivable networks. This paper proposes a evaluation method to designate the relay nodes in the network modifying betweenness centrality for relay nodes and survivability. In our simulation, we used two routing algorithms for the survivability of the network and compared the modified measure applying each algorithm with original betweenness centrality. The simulation results show that our approach estimates features different from the original does. The different implies that our method is effective to identify relay nodes.**

Keywords–Betweenness centrality; Graph theory; Data networks

I. INTRODUCTION

Several types of data networks including a wireless ad-hoc network require relay nodes to transmit data because the nodes would impact on services with the networks. One of the measures to reveal important nodes like the relay nodes is centrality which belongs to graph theory.

Centrality is known as a measure to capture characteristics of network topology in fields of social network, computer science, physics and biology [1], which, in detail, analyzes the importance of each node from some points of view. The importance is rated by a real-valued function on a node. Above all, betweenness centrality is applied to the network in which something such as packets or messages of the Internet flows between other nodes. For each node, betweenness centrality counts the number of shortest paths that the node lies on. Betweenness centrality assumes that every pair of nodes interchanges a message with equal probability in equal time intervals [1]. However, this assumption can be unsuitable for several types of networks. Therefore, Freeman et al. [2] suggests other betweenness centrality with not shortest paths but max flow.

Though a relay node is considered to be located closer to the middle of a route along which two nodes exchange data each other, suggested betweenness centrality may not be able to recognize such. In addition, the original betweenness centrality does not take consideration on survivability because of single path connecting two nodes.

Consequently, our study proposes the modified betweenness centrality based on two node disjoint paths to identify relay nodes in the data network. The measure is expected to give a ranking on the possibility that each node is to be a relay one of a network.

Section II describes the definition of betweenness centrality. Section III presents the modification of betweenness centrality. Simulation results are presented in Section IV. Section V concludes this paper.

II. BETWEENNESS CENTRALITY

According to Newman [1], betweenness centrality is defined as follows.

In a graph denoted as $G = (V, E)$, let σ_{st} be the number of shortest paths between two vertices $s, t \in V$, and let $\sigma_{st}(v)$ indicate the number of the paths through a vertex $v \in V$. Betweenness centrality C_B on v is defined by

$$C_B(v) = \sum_{\substack{s,t \in V \setminus \{v\} \\ s \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}}. \tag{1}$$

This formula supposes that if there exists multiple shortest paths connecting s and t, i.e. $\sigma_{st} \geq 2$, then one of the paths is chosen equally, so that $\sigma_{st}(v)$ is divided by σ_{st}.

III. MODIFICATION OF BETWEENNESS CENTRALITY

This section shows the modification of betweenness centrality. A network are described as an undirected edge-weighted graph $G = (V, E)$. Let $w(e) \in \mathbb{R}^+$ be the weight of an edge $e = (u, v) \in E$. Denote $v \in p$ if a vertex $v \in V$ is included in a path p.

Because network flow relayed on a node could be regarded as several paths across the node which some data flows along, a path set is modeled as the flow. Considering endpoints and survivability. Since the graph G is undirected, it is unnecessary to distinguish the source and the destination of a path. Additionally, Not all nodes can send and receive data depending on the network. As a result, let a subset $V' \subseteq V$ include all endpoints contained in the graph.

On the other hand, to guarantee survivability of the network, the graph G is required to be 2-vertex connected. Thus, endpoints are certainly connected by two vertex disjoint paths. Denote $\psi = (p, p')$ as a pair of two paths p and p' that are composed of the same endpoints. Depending on the way to find a path pair of a couple of endpoints and the topology of the graph G, no pair may be found. Therefore, let Ψ_S be a set of ψ able to be discovered between two endpoints s and t, where a set $S = \{s, t\}$ $(s, t \in V', s \neq t)$.

Let $l(p) \in \mathbb{R}^+$ be the length of a path p which is the sum of the weights on edges in the path. Let $d_p(v)$ indicate the distance from either of endpoints of a path to a vertex v. Transmission distance of the network is limited to the upper bound $L \in \mathbb{R}^+$. If $l(p)$ holds $L < l(p) \leq 2L$, a relay vertex should belong to the set $\{v \in p \mid l(p) - L \leq d_p(v) \leq L\}$ to intercommunicate between the endpoints of the path. If $l(p) \leq L$, then p does not need the vertex. If $l(p) > 2L$, then at least

 ISBN: 978-1-61208-382-7

two relay vertices are necessary, so this paper assumes that $L \geq l(p)/2$, for all p for simplicity.

It is considered that the closeness of a vertex to the center of a path corresponds to the appropriateness of the vertex for the relay one. Thus, Define the function $f : [0, 1] \to [0, 1]$ that satisfies the following:

- $f\left(\frac{1}{2}\right) = 1$
- $f(x) = f(1 - x)$
- $f(x) < f(x'), \quad \text{if } x < x' \leq \frac{1}{2} \text{ or } \frac{1}{2} < x' < x.$

The degree of the capability that a vertex $v \in V$ acts as the relay vertex on either path in a pair ψ is designated as the function $\delta_\psi : V \to [0, 1]$ on v with the function f

$$\delta_\psi(v) = \begin{cases} f\left(\dfrac{d_p(v)}{l(p)}\right), & \text{if } \exists p \in \psi \text{ s.t. } v \in p, \\ & l(p) - L \leq d_p(v) \leq L \\ 0, & \text{otherwise} \end{cases}.$$

From the above, modified betweenness centrality C_M as a function on a vertex v is defined by

$$C_M(v) = \sum_{\substack{S \in \mathcal{P}(V' \setminus \{v\}) \\ |S| = 2}} \frac{\sum_{\psi \in \Psi_S} \delta_\psi(v)}{|\Psi_S|}. \qquad (2)$$

IV. Experiments and Results

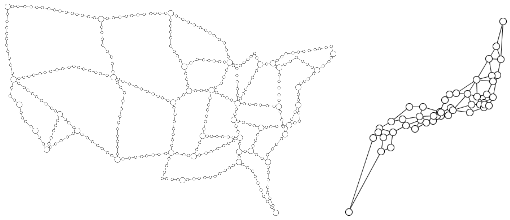
Two network models were analyzed by betweenness centrality and modified one. The results from each measures were compared.

Modified betweenness centrality is made with Constrained Shortest Path First (CSPF) and Vertex Disjoint Shortest Pair (VDSP) for the pair routing algorithm. Simplified algorithm of CSPF is given as follows:

1. Find a shortest path by Dijkstra algorithm in a graph.
2. Modify the graph to remove vertices in the path without endpoints.
3. Search the other shortest path between the same endpoints in the modified graph.
4. Obtain the pair of first path and second path.

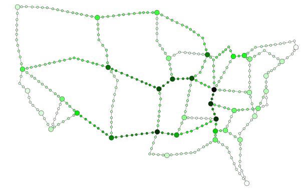
VDSP proposed by Bhandari [3] is an algorithm for finding two paths that have minimum total length. It enables to get the vertex disjoint pair in any 2-vertex connected graph.

Figure 1 shows network models for the evaluation of analysis methods. Model A in the figure limits endpoints sets, and model B does not. Model A possess 365 vertices. The

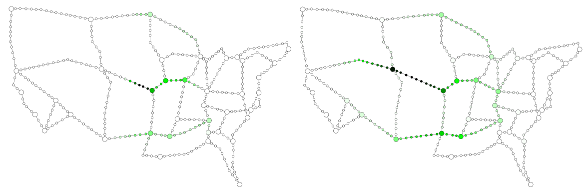


(a) Model A (b) Model B

Figure 1. Two types of graph for experiments.

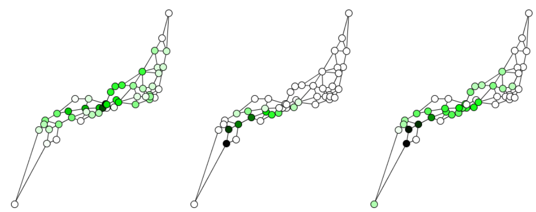


(a) Original betweenness centrality



(b) Modified with CSPF (c) Modified with VDSP

Figure 2. Visualized evaluates on model A.



(a) Original (b) with CSPF (c) with VDSP

Figure 3. Visualized evaluates on model B.

endpoints set V' is composed of big circles illustrated in the figure, of which the size is 39. Model B contains 49 vertices. Its endpoints set is the same set as own vertex set.

Figures 2 and 3 show visualization of evaluation results for the models by coloring each vertex the color of which brightness corresponds to each evaluate value of the vertex (palest = 0, darkest = max). These figures describe that locations of high value vertices in both of the modified measures are different from original measures on both model A and B. From this result, each modified centralities can reveal characteristics which existent betweenness centrality can not.

V. Conclusion

This paper proposed a method to identify relay nodes based on modified betweenness centrality. This method is constructed of a set of pairs of vertex disjoint paths, and the degree of closeness to the center of each path. In our simulation, the two types of modified measures by two algorithms CSPF and VDSP to determine a set of pairs were compared with original betweenness centrality in two different network models. The simulation results show that our approach indicates attributes the original does not. Thus, our method may be capable of designating relay nodes.

References

[1] M. Newman, Networks: an introduction. Oxford University Press, 2010.

[2] L. C. Freeman, S. P. Borgatti, and D. R. White, "Centrality in valued graphs: A measure of betweenness based on network flow," Social networks, vol. 13, no. 2, 1991, pp. 141–154.

[3] R. Bhandari, Survivable networks: algorithms for diverse routing. Springer, 1999.