Accelerating computation of distance based centrality measures for spatial networks

Kouzou Ohara[1], Kazumi Saito[2], Masahiro Kimura[3], and Hiroshi Motoda[4,5]

[1] Department of Integrated Information Technology, Aoyama Gakuin University
ohara@it.aoyama.ac.jp
[2] School of Administration and Informatics, University of Shizuoka
k-saito@u-shizuoka-ken.ac.jp
[3] Department of Electronics and Informatics, Ryukoku University
kimura@rins.ryukoku.ac.jp
[4] Institute of Scientific and Industrial Research, Osaka University
motoda@ar.sanken.osaka-u.ac.jp
[5] School of Computing and Information Systems, University of Tasmania

Abstract. In this paper, by focusing on spatial networks embedded in the real space, we first extend the conventional step-based closeness and betweenness centralities by incorporating inter-nodes link distances obtained from the positions of nodes. Then, we propose a method for accelerating computation of these centrality measures by pruning some nodes and links based on the cut links of a given spatial network. In our experiments using spatial networks constructed from urban streets of cities of several types, our proposed method achieved about twice the computational efficiency compared with the baseline method. Actual amount of reduction in computation time depends on network structures. We further experimentally show by examining the highly ranked nodes that the closeness and betweenness centralities have completely different characteristics to each other.

Keywords: closeness centrality, betweenness centrality, spatial network, distance-based centrality, cut link

1 Introduction

Studies of the structure and functions of large complex networks have attracted a great deal of attention in many different fields such as sociology, biology, physics and computer science [15]. Our final goal of this research is to develop methodologies and technologies that are accurate, efficient and scalable to analyze complex networks which are now ubiquitous in almost every field of science and to discover useful knowledge from there. For this we think it is critical to be able to assess and utilize inherent and intrinsic characteristics of such networks.

One common approach to analyze large complex networks is investigating their characteristics through a measure called centrality. Various kinds of centralities are used according to what we want to know. For example, if our goal is to know the topological characteristics of a network, degree, closeness, and betweenness centralities [11] can be used. if it is to know the importance of nodes that constitute a network, HITS [5] and PageRank [3] centralities are often used. Influence degree centrality [12] is another one to measure the importance of nodes. Among these conventional centralities, we focus on the closeness and betweenness centralities in this work because they are closely related to real world problems such as location planning of commercial or evacuation facilities in a wide area. Here, note that closeness and betweenness centralities usually approximate the distance between two distinct nodes by the number of links traversed to get to one node from another. This approximation may not be realistic when analyzing networks such as real traffic networks, one of the real world problems. Thus, as a particular class, we focus on spatial networks embedded in the real space, like urban streets, whose nodes occupy a precise position in two or three-dimensional Euclidean space, and whose links are real physical connections [10]. Analyzing and characterizing the structure of such large spatial networks will play an important role for understanding and improving the usages of these networks, as well as discovering new insights for developing and planning city promotion, trip tours and so on. To facilitate such research work, in this paper, we propose techniques useful to accelerate their computations based on network pruning. This is motivated by the fact that the computation time to calculate the value of such conventional *step-based* centralities for every single node in a network becomes larger as the size of the network gets larger because of the necessity to traverse each link in the network multiple times.

In this paper, we first extend the conventional step-based closeness and betweenness centralities by incorporating inter-nodes link distances obtained from the positions of nodes. Hereafter, we refer to these extended ones as distance-based centrality measures. Then, we propose a method for accelerating computation of these centrality measures by pruning some nodes and links that are related to the cut links of a given spatial network, where each cut link divides the network into two connected components by its removal. Here, we should note that our approach is applicable to a wider range of spatial networks, and potentially to complex networks with inter-nodes link distances, although in this paper, we focus on spatial networks constructed from urban streets by mapping the intersections of streets into nodes and the streets between the nodes into links. In the experiments we evaluate the performance of our proposed acceleration techniques for the spacial networks constructed from cities of several types, and discuss the characteristics of the distance-based centralities.

This paper is organized as follows. After first explaining the related work in Section 2, we describe the details of our proposed methods in Section 3. In Section 4, we evaluate the characteristics of networks obtained by our proposed methods both qualitatively and quantitatively, and give our conclusion in Section 5.

2 Related Work

As mentioned earlier, analyzing and characterizing the structure of large spatial networks like urban streets will play an important role for understanding and improving the usages of these networks embedded in the real space. Thus, the structure and functions of spatial networks have also been studied by many researchers [4, 10, 13, 17, 18, 20]. From structural viewpoints, centrality measures have been widely used to analyze spatial networks [10, 18], especially by extending the conventional notions of centrality measures on simple networks into those of weighted networks based on road usage frequency of urban streets [13, 17]. From functional viewpoints, traffic usage patterns in urban streets have been investigated [4, 20]. In this paper, unlike these previous studies, we focus on extending the conventional centralities by incorporating inter-nodes link distances obtained from the positions of nodes, and propose a method for accelerating the calculation of these centrality measures.

To find important nodes in a network, several centrality measures have been presented in the field of social and information network analysis. Representative centrality measures include degree, closeness and betweenness centralities [11], HITS (hub and authority) centrality [5], and PageRank centrality [3]. These measures are also closely related to quantifying how influential each node is in the context of information diffusion, and influence degree centrality can be defined by evaluating the influence of each node. Unlike centrality measures derived only from network topology, influence degree centrality exploits a dynamical process on the network as well. An efficient method of simultaneously estimating the influence degrees of all the nodes was presented under the SIR model setting [12]. We note that influence degree centrality can also be employed for identifying super-mediators in the social network [19]. In this paper, as our first step to develop these centrality notions for spatial networks, we focus on closeness and betweenness centralities, which have been widely used in the field of social network analysis. Our proposed method in this paper shares the same basic idea of the previous method presented under the SIR model setting [12] in that redundant nodes and links are pruned for accelerating the computation of the centrality measures,

For some centrality measures such as betweenness centrality, their computation becomes harder as the network size increases, since it needs to take the global network structure into account. Thus, several researchers presented methods of approximating such centralities [1, 16, 6–8]. For instance, a bottom-k sketch [7, 8] is obtained by associating with each node in a network an independent random rank value drawn from a probability distribution. The bottom-k sketch of each node in the network can be quite efficiently calculated by orderly assigning the rank values from the smallest one to those nodes reachable by reversely following links over the network. Based on this framework, a greedy Sketch-based Influence Maximization (SKIM) algorithm has been proposed, and it has been shown that the SKIM algorithm scales to graphs with billions of edges, with one to two orders of magnitude speedup over the best greedy methods [9]. Here in this paper, unlike the above approximation approaches, we rather focus on exactly computing the centrality measures whereby there is no need to worry about the approximation performance. Note that these exact solutions can be used as the ground-truth for evaluating the approximation performance.

3 Proposed Method

Let $G = (\mathcal{V}, \mathcal{E})$ be a spatial network consisting of a single connected component without self-loops, where $\mathcal{V} = \{u, v, w, \cdots\}$ and $\mathcal{E} = \{e, \cdots\} \subset \mathcal{V} \times \mathcal{V}$ are sets of nodes and undirected links, respectively. For each link $e = (u, v)$, we express the distance between nodes u and v by $d(u, v)$, where we can obtain these distances from the positions of nodes in the spatial network. For each pair of nodes $u, w \in \mathcal{V}$ without the direct connection, we define the distance $d(u, w)$ as the geodesic distance over the network, as usual. Then, for each node $u \in \mathcal{V}$, we can define the following distance based closeness centrality measure:

$$DC(u) = \left(\sum_{w \in \mathcal{V}} d(u, w) \right)^{-1}. \tag{1}$$

Note that the distance based closeness centrality $DC(u)$ is a natural extension to the conventional step based closeness centrality $SC(u)$ because $DC(u)$ reduces to $SC(u)$ by setting $d(u, v) = 1$ for each link $(u, v) \in \mathcal{E}$. Similarly, for each node $v \in \mathcal{V}$, we can define the following distance based betweenness centrality measure:

$$DB(v) = \sum_{u \in \mathcal{V} \setminus \{v\}} \sum_{w \in \mathcal{V} \setminus \{u, v\}} \frac{\sigma(u, w; v)}{\sigma(u, w)} \tag{2}$$

where $\sigma(u, w)$ is the total number of the paths with the smallest distance between node u and node w in G and $\sigma(u, w; v)$ is the number of those paths between node u and node w in G that passes through node v. Again, note that the distance based betweenness centrality $DB(v)$ is a natural extension to the conventional step based betweenness centrality $SB(v)$ because $DB(v)$ also reduces to $SB(v)$ by setting $d(u, v) = 1$ for each link $(u, v) \in \mathcal{E}$. By applying the best-first search algorithm starting from each node $u \in \mathcal{V}$ with respect to distance $d(u, w)$, we can calculate these centrality measures, $DC(u)$ and $DB(v)$, for all the nodes in G. As mentioned earlier, when calculating closeness and betweenness centrality measures, their computation becomes harder as the network size increases. Below we propose a method of improving the computational efficiency to calculate these centrality measures.

3.1 Pruning Techniques for Closeness Centrality

We say that a link $e \in \mathcal{E}$ is a cut link if the network G is divided into two connected components by eliminating the link e. For a given cut link $e = (u, v)$, let $\mathcal{V}(u \setminus v)$ and $\mathcal{V}(v \setminus u)$ be the sets of nodes in the two connected components, each of which includes the nodes u and v, respectively, i.e., $u \in \mathcal{V}(u \setminus v)$, $v \in \mathcal{V}(v \setminus u)$, $\mathcal{V}(u \setminus v) \cap \mathcal{V}(v \setminus u) = \emptyset$ and $\mathcal{V}(u \setminus v) \cup \mathcal{V}(v \setminus u) = \mathcal{V}$. Let $\eta(u)$ and $\delta(u)$ be the number of the nodes and the accumulated value of distances obtained by our pruning process relating to the node u, respectively. Then, after initializing $\eta(u) \leftarrow 1$ and $\delta(u) \leftarrow 0$ for each node $u \in \mathcal{V}$, and setting $\mathcal{W} \leftarrow \mathcal{V}$ and $\mathcal{F} \leftarrow \mathcal{E}$, we consider calculating $DC(x)^{-1}$ for each node $x \in \mathcal{W}$ by using the following formula:

$$DC(x)^{-1} = \sum_{w \in \mathcal{W}} (\eta(w) d(x, w) + \delta(w)) \tag{3}$$

Now, let $e = (u, v)$ be a cut link; then, for a node $x \in \mathcal{V}(u \setminus v)$, we can decompose $DC(x)^{-1}$ as follows:

$$\begin{aligned} DC(x)^{-1} &= \sum_{w \in \mathcal{V}(u \setminus v)} d(x, w) + \sum_{w \in \mathcal{V}(v \setminus u)} (d(x, u) + d(u, v) + d(v, w)) \\ &= \sum_{w \in \mathcal{V}(u \setminus v)} d(x, w) + |\mathcal{V}(v \setminus u)| d(x, u) + |\mathcal{V}(v \setminus u)| d(u, v) + \sum_{w \in \mathcal{V}(v \setminus u)} d(v, w). \end{aligned} \tag{4}$$

By using the above decomposition shown in Equation (4), we can derive a recursive TP (Top-down Pruning) technique for closeness centrality. More specifically, we consider updating $\eta(u)$ and $\delta(u)$ by using the following formulae:

$$\begin{aligned}\eta(u) &\leftarrow 1 + |\mathcal{V}(v \setminus u)|,\\ \delta(u) &\leftarrow |\mathcal{V}(v \setminus u)| d(u,v) + \sum_{w \in \mathcal{V}(v \setminus u)} d(v,w),\end{aligned} \tag{5}$$

Then, after pruning the node set $\mathcal{V}(v \setminus u)$ and the cut link $e = (u,v)$ from the network $G = (\mathcal{V},\mathcal{E})$ as $\mathcal{W} \leftarrow \mathcal{V} \setminus \mathcal{V}(v \setminus u)$ and $\mathcal{F} \leftarrow \mathcal{E} \setminus \{e\}$, we can equivalently calculate $DC(x)^{-1}$ by using Equation (3) for each node $x \in \mathcal{V}(u \setminus v)$ as obtained by Equation (1). Clearly, we can apply the same arguments to the case of $x \in \mathcal{V}(v \setminus u)$. Therefore, by generalizing the update formulae of Equation (5) as follows:

$$\begin{aligned}\eta(u) &\leftarrow \eta(u) + \sum_{w \in \mathcal{W}(v \setminus u)} \eta(w),\\ \delta(u) &\leftarrow \delta(u) + d(u,v) \sum_{w \in \mathcal{W}(v \setminus u)} \eta(w) + \sum_{w \in \mathcal{W}(v \setminus u)} \eta(w) d(v,w),\end{aligned} \tag{6}$$

for all the node $x \in \mathcal{V}$, we can exactly calculate the closeness centrality measure $DC(x)$ by using the above recursive TP update shown in Equation (6), where $\mathcal{W}(u \setminus v)$ and $\mathcal{W}(v \setminus u)$ denote the sets of nodes in the two connected components, respectively, which are obtained by the removal of the cut link $e = (u,v)$, just like those defined for $\mathcal{V}$.

As a special case, by focusing on the fact that each link of degree-one node is a cut link, we derive a recursive BP (Bottom-up Pruning) technique algorithm for closeness centrality. More specifically, for a degree-one node v and its cut link $e = (u,v)$, we consider updating $\eta(u)$ and $\delta(u)$ as follows:

$$\begin{aligned}\eta(u) &\leftarrow \eta(u) + \eta(v),\\ \delta(u) &\leftarrow \delta(u) + \eta(v) d(u,v) + \delta(v)\end{aligned} \tag{7}$$

Then, after pruning the node v and its cut link $e = (u,v)$ from the network $G = (\mathcal{W},\mathcal{F})$ as $\mathcal{W} \leftarrow \mathcal{W} \setminus \{v\}$ and $\mathcal{F} \leftarrow \mathcal{F} \setminus \{e\}$, and updating $\eta(u)$ and $\delta(u)$ by Equation (7), we can also equivalently calculate $DC(x)^{-1}$ by using Equation (3) for each node $x \in \mathcal{W}$ as obtained by Equation (1). Here note that after obtaining the value of $DC(u)^{-1}$, we can calculate $DC(v)^{-1}$ as follows:

$$DC(v)^{-1} \leftarrow DC(u)^{-1} + (|\mathcal{V}| - 2\eta(v)) d(u,v)$$

Therefore, for all the node $x \in \mathcal{V}$, we can exactly calculate the closeness centrality measure $DC(x)$ by using the above recursive TP and BP techniques.

3.2 Pruning Techniques for Betweenness Centrality

We derive the recursive pruning techniques for calculating the betweenness centrality measure $DB(x)$. Let $\phi(u)$ be the number of node pairs obtained by our pruning process relating to the node u. Again, after initializing $\eta(u) \leftarrow 1$ and $\phi(u) \leftarrow 0$ for each node

$u \in \mathcal{V}$, and setting $\mathcal{W} \leftarrow \mathcal{V}$ and $\mathcal{F} \leftarrow \mathcal{E}$, we consider calculating $DB(x)$ for each node $x \in \mathcal{W}$ by using the following formula:

$$DB(x) = \phi(x) + \sum_{w \in \mathcal{W} \setminus \{x\}} \eta(w) \xi(x; w). \tag{8}$$

Here, $\xi(x; w)$ stands for the number of descendant nodes of x obtained by the best first search starting from the node w, i.e.,

$$\xi(x; w) = \eta(x) - 1 + \sum_{z \in C(x;w)} (\xi(z; w) + 1) \frac{\sigma(w, x)}{\sigma(w, z)},$$

where $C(x; w)$ means the set of direct child nodes of x for the search from node w. Now, let $e = (u, v) \in \mathcal{E}$ be a cut link; then, for a node $w \in \mathcal{V}(u \setminus v)$, we can calculate $\xi(u; w)$ as follows:

$$\xi(u; w) = |\mathcal{V}(v \setminus u)| + \sum_{z \in C(u;w) \setminus \{v\}} (\xi(z; w) + 1) \frac{\sigma(w, u)}{\sigma(w, z)}, \tag{9}$$

and $\xi(u; w) = |\mathcal{V}(u \setminus v)| - 1$ for $w \in \mathcal{V}(v \setminus u)$. By using the above calculation shown in Equation (9) and noting that the node u mediates any node pair (w, y) starting from $w \in \mathcal{V}(v \setminus u)$ and ending at $y \in \mathcal{V}(u \setminus v) \setminus \{u\}$, but no pair starting from $w \in \mathcal{V}(v \setminus u)$ and ending at $y \in \mathcal{V}(v \setminus u)$, we can derive a recursive TP technique for betweenness centrality. More specifically, we consider updating $\eta(u)$ and $\phi(u)$ by using the following formulae:

$$\begin{aligned} \eta(u) &\leftarrow 1 + |\mathcal{V}(v \setminus u)|, \\ \phi(u) &\leftarrow \mathcal{V}(v \setminus u)|(|\mathcal{V}(u \setminus v)| - 1)|, \end{aligned} \tag{10}$$

Then, after pruning the node set $\mathcal{V}(v \setminus u)$ and the cut link $e = (u, v)$ as $\mathcal{W} \leftarrow \mathcal{V} \setminus \mathcal{V}(v \setminus u)$ and $\mathcal{F} \leftarrow \mathcal{E} \setminus \{e\}$, we can equivalently calculate $DB(x)$ by using Equation (8) for each node $x \in \mathcal{V}(u \setminus v)$ as obtained by Equation (2). Again, we can apply the same arguments to the case of $x \in \mathcal{V}(v \setminus u)$. Therefore, by generalizing the update formulae of Equation (10) as follows:

$$\begin{aligned} \eta(u) &\leftarrow \eta(u) + \sum_{w \in \mathcal{W}(v \setminus u)} \eta(w), \\ \phi(u) &\leftarrow \phi(u) + \sum_{w \in \mathcal{W}(v \setminus u)} \eta(w) \left(\sum_{w \in \mathcal{W}(u \setminus v)} \eta(w) - 1 \right), \end{aligned} \tag{11}$$

for all the node $x \in \mathcal{V}$, we can also exactly calculate the betweenness centrality measure $DB(x)$ by using the above recursive TP update shown in Equation (11).

Similarly to the case of the closeness centrality, by focusing on the fact that each link of degree-one node is a cut link, we derive a recursive BP technique algorithm for betweenness centrality. More specifically, for a degree-one node v and its cut link $e = (u, v)$, we consider updating $\eta(u)$, $\phi(u)$ and $\phi(v)$ as follows:

$$\begin{aligned} \eta(u) &\leftarrow \eta(u) + \eta(v), \\ \phi(u) &\leftarrow \phi(u) + \eta(v)(|\mathcal{V}| - \eta(v) - 1), \\ \phi(v) &\leftarrow \phi(v) + (\eta(v) - 1)(|\mathcal{V}| - \eta(v)). \end{aligned} \tag{12}$$

Then, after pruning the node v and its cut link $e = (u, v)$ from the network $G = (\mathcal{W}, \mathcal{F})$ as $\mathcal{W} \leftarrow \mathcal{W} \setminus \{v\}$ and $\mathcal{F} \leftarrow \mathcal{F} \setminus \{e\}$, and updating $\eta(u)$, $\phi(u)$ and $\phi(v)$ by Equation (12), we can equivalently calculate $DB(x)$ by using Equation (8) for each node $x \in \mathcal{W}$ as obtained by Equation (2). Here note that for any pruned nodes v, we calculate its measure by $DB(v) \leftarrow \phi(v)$. Therefore, for all the node $x \in \mathcal{V}$, we can exactly calculate the closeness centrality measure $DB(x)$ by using the above recursive TP and BP techniques.

3.3 Summary of Proposed Method

Hereafter, we consider simultaneously calculating both the closeness and betweenness centrality measures for all the node $v \in \mathcal{V}$. In our proposed method, the BP technique is applied before the TP techniques, because we can easily know the degree-one nodes in our network G. Although we can individually incorporate these techniques into the baseline method that do not employ our proposed pruning techniques, we only consider the proposed method without the TP technique, which is referred to as the BP method. Since it is difficult to analytically examine the effectiveness of these techniques, we empirically evaluate the computational efficiency of these three methods in comparison to the baseline method without the proposed pruning techniques, which is referred to as the BL method.

4 Experiments

In this section, after explaining the dataset used in our experiments, we evaluate the performance of our proposed acceleration algorithm, and discuss the characteristics of the distance based centralities.

4.1 Dataset

We used OSM (OpenStreetMap) data of eight cities in our experiments, i.e., Barcelona (Spain, Europe), Bologna (Italy, Europe), Brasilia (Brazil, South America), Cairo (Egypt, Africa), Washington D.C. (United States, North America), New Delhi (India, Asia), Richmond (United States, North America), and San Francisco (United States, North America). These are a subset of cities studied in [10]. In August, 2015, we obtained the OSM data of these eight cities from Metro Extracts[1]. Here note that in our experiments, the area of each city is more than 100 times larger than those of the previous study [10].

From the OSM data of each city, we extracted all highways and all nodes appearing in them, and constructed each spatial network by mapping the ends, intersections and curve-fitting-points of streets into nodes and the streets between the nodes into links. Then, based on GRS80 [14], we calculated each inter-node link distance from the positions of the nodes, each of which is described by a pair of latitude and longitude. Table 1 shows the basic statistics of the networks for the eight selected cities, where *deg* means the average degree of nodes calculated by $deg = 2|\mathcal{E}|/|\mathcal{V}|$, and *avg*, *s.d.* and *max* stands for the average, standard deviation and maximum of inter-node link distances, respectively. From this table, we can see that although the area and the numbers of nodes and

[1] https://mapzen.com/data/metro-extracts

Table 1. Basic statistics as network.

No.	Name	Area	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	*deg*	*avg*	*s.d.*	*max*
1	Barcelona	45×30km	344,095	378,074	2.2	27.0	34.7	1,396.4
2	Bologna	60×45km	262,839	281,294	2.1	34.2	59.3	4,341.1
3	Brasilia	120×104km	197,829	240,546	2.4	78.6	135.6	8,369.9
4	Cairo	87×86km	195,228	225,049	2.3	69.8	123.2	8,078.4
5	Washington D.C.	23×18km	114,758	128,746	2.2	25.3	33.7	1,058.5
6	New Delhi	109×75km	284,964	336,183	2.4	66.2	99.7	10,968.3
7	Richmond	54×35km	371,174	394,161	2.1	28.5	45.5	15,359.4
8	San Francisco	90×50km	492,266	541,162	2.1	33.9	48.3	7,241.0

links, $|\mathcal{V}|$ and $|\mathcal{E}|$, are substantially different, the average degrees are quite similar as common characteristics of these spatial networks. On the other hand, it seems that each city has its own characteristics about the statistics of link distances.

In order to more closely study the characteristics of link distances, we examined the frequency of link distances obtained by

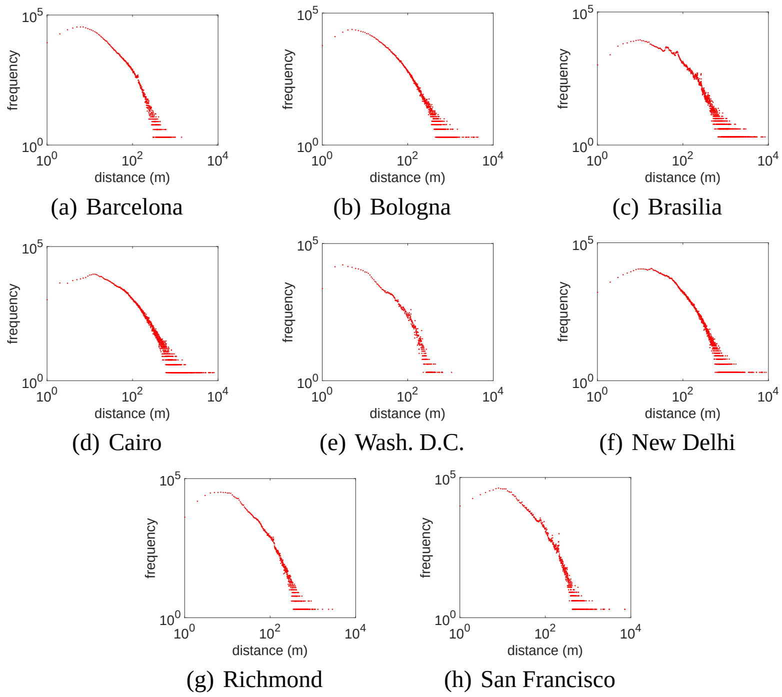
$$f(k) = |\{e = (u, v) \in \mathcal{E} \ : \ \epsilon_k \leq d(u, v) < \epsilon_{k+1}\}|, \tag{13}$$

where k is a non-negative integer, and ϵ_k is set to k meter. Figure 1 shows our analysis results of the eight cities shown in Tab. 1, where the horizontal and vertical axes stand for the distance and frequency, respectively. From these results plotted as log-log graphs, we can observe that each frequency curve is reasonably approximated by a power-law-like distribution. This suggests that since some link distances have relatively quite larger values, the distance-based centrality measures may have significantly different characteristics in comparison to those of the step-based centrality measures. In this paper, we experimentally evaluate these differences by focusing on the closeness and betweenness centralities.

4.2 Computational Efficiency

As described earlier, we evaluated the efficiency of the proposed method which simultaneously calculates $DC(v)$ and $DB(v)$ for each node $v \in \mathcal{V}$, by comparing the computation time of the baseline (BL), only bottom-up pruning (BP), and the proposed (PR) methods. We implemented the BL method based on Brandes's algorithm [2] known as the standard and efficient technique for computing the betweenness centrality of each node in a network. Figure 2 shows the computation time of each method, where the dataset numbers shown in the horizontal axis are identical to those of Tab. 1. Figure 2(a) compares the actual processing time of these methods, where our programs implemented in C were executed on a computer system equipped with two Xeon X5690 3.47GHz CPUs and 192GB main memory with a single thread within the memory capacity. Figure 2(b) compares the reduction rates of computation time for these methods from the BL method.

From Figs. 2(a) and 2(b), we can see that for all the networks, the BP method steadily improves the computational efficiency of the BL method, and the PR method

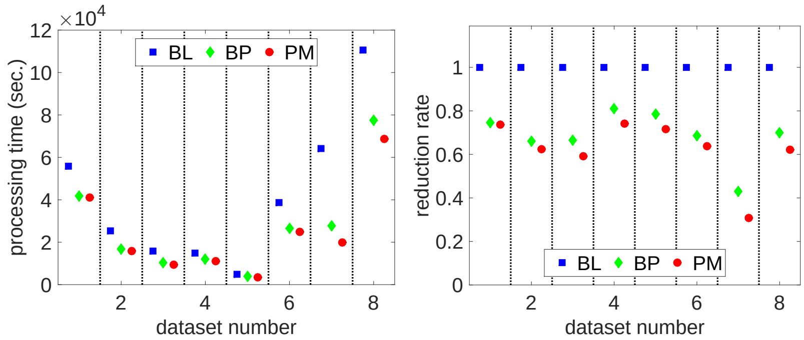


(a) Barcelona (b) Bologna (c) Brasilia (d) Cairo (e) Wash. D.C. (f) New Delhi (g) Richmond (h) San Francisco

Fig. 1. Frequency of link distances.

slightly improve that of the BP method. These results demonstrate the effectiveness of the proposed techniques. More specifically, as expected, from Fig. 2(a) and Tab. 1, we can see that the processing time of the BL method is almost proportional to the size of network. In contrast, from Fig. 2(b), we can see that the reduction rates of the BP and PR methods depend on the network, i.e., around from 0.4 to 0.8 for the BP method and around from 0.3 to 0.7 for the PR method. These results indicate that the effects of our pruning techniques depend on the networks.

On the other hand, we note that the improvement rates of the PR method over the BP method are modest, i.e., the reduction rates by the TP technique are not so remarkably effective. This must be partly because the TP technique requires additional computation costs for detecting cut links. Overall, we can conjecture that the proposed method combining both the BP and TP techniques is more reliable than the other two methods in terms of computation time because it produced the best performance for all of the eight networks. In short, reduction of computation time depends on network structures, but overall we can say that use of both techniques can increase the computational efficiency by nearly twice of the BL method.



(a) Processing time (b) Reduction rate

Fig. 2. Computation time comparison.

4.3 Comparison with Conventional Centralities

As noted earlier, we experimentally evaluate the characteristics of the distance- and the step-based measures by focusing on the closeness and betweenness centralities. In order to make a fair comparison of these different types of centralities, we consider the normalized measures that are divided by their maximum values, i.e., normalized distance-based closeness and betweenness centralities are calculated by

$$\begin{aligned} nDC(v) &= DC(v)/\max_{x\in\mathcal{V}}\{DC(x)\}, \\ nDB(v) &= DB(v)/\max_{x\in\mathcal{V}}\{DB(x)\}, \end{aligned} \tag{14}$$

respectively. Here, let $v(k_{DC})$ be the k-th node which has the k-th largest value in the distance-based closeness centrality values $\{DC(v) \ : \ v \in \mathcal{V}\}$. Similarly, we can define the k-th node for the other centrality measures as $v(k_{SC})$, $v(k_{DB})$ and $v(k_{SB})$, respectively. Recall that we denote the step-based closeness and betweenness centrality measures as $SC(v)$ and $SB(v)$, respectively. Then, we consider characterizing the top-k nodes of these measures by plotting the following quartets:

$$\{(k, nDC(v(k_{DC}))), (k, nDC(v(k_{SC}))), (k, nDB(v(k_{DB}))), (k, nDB(v(k_{SB})))\}. \tag{15}$$

Namely, we evaluate the top-k nodes of the distance- and the step-based closeness centralities by the normalized measures of the distance-based closeness centrality and those of the distance- and the step-based betweenness centralities by the normalized measures of the distance-based betweenness centrality.

Figure 3 shows our experimental results, where the horizontal and vertical axes stand for the rank k up to top-100 and the normalized distance-based closeness or betweenness centrality measure, respectively. We can see that the closeness and betweenness centralities have completely different characteristics. In fact, the top-100 values of the distance- and the step-based closeness centralities (DC= $nDC(v(k_{DC}))$ and SC= $nDC(v(k_{SC}))$) are almost the same, while the situation is completely different for the two betweenness centralities. The top-100 values of the distance-based

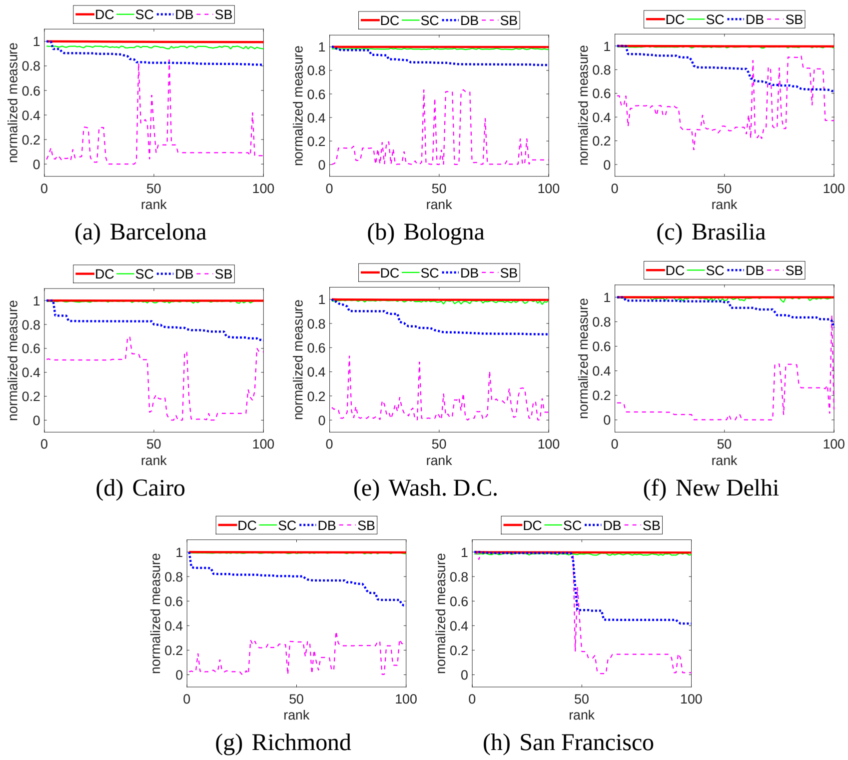
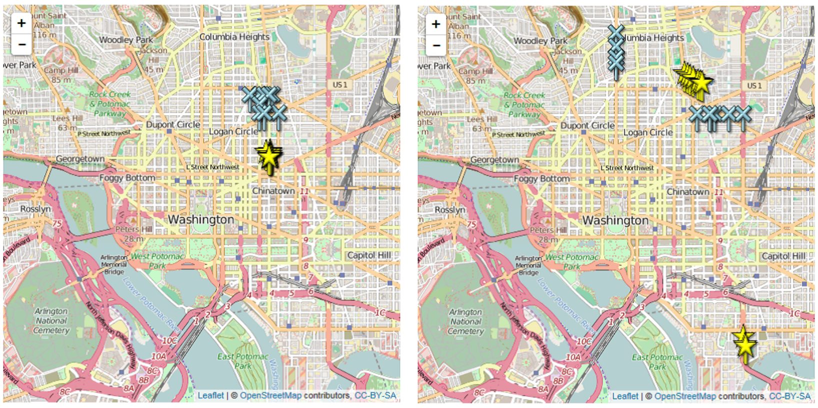


Fig. 3. Comparison of top-100 nodes.

betweenness centrality substantially decrease (DB= $nDB(v(k_{DB}))$), while the top-100 values of the step-based betweenness centrality have no regularity and violently change (SB= $nDB(v(k_{SB}))$). Namely, these results indicate that the ranking by the distance- and the step-based closeness centralities is almost the same, while the ranking by the distance-based betweenness centrality is substantially different from the ranking by the step-based betweenness centrality. These experimental results suggest that for arbitrary pairs of nodes, most of the paths with the minimum distances are different between the distance- and the step-based centralities, while the distances of these paths are relatively close to each other. On the other hand, we can see that the curves for the top-100 nodes are somewhat different from each other across different network datasets, although their global behaviors are generally quite similar regardless of any pairs of the centrality measures and the networks. Thus, it is expected that the normalized measure curves shown in Fig. 3 may be able to uncover some characteristics of these networks, e.g., the nodes within the top-50 in San Francisco are quite similar for both the closeness and betweenness centralities,

Next, we investigate how differently the top ranked nodes in each centrality measure are distributed on a spatial map. To this end, we plotted the top-10 nodes in each centrality measure on the street map for the Washington D.C. dataset as shown in Fig. 4,

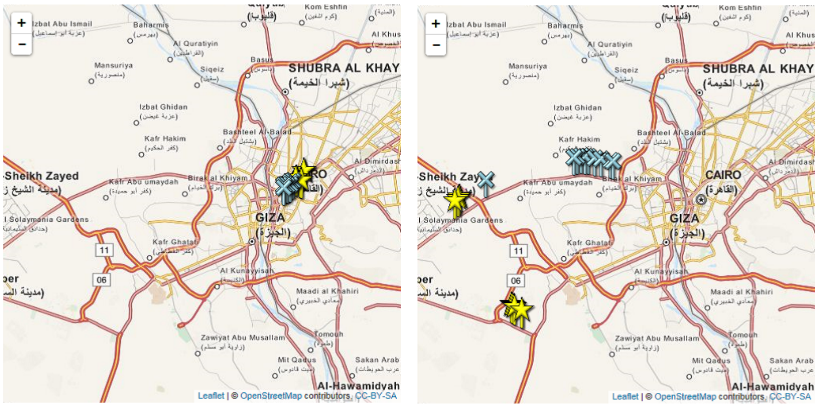


(a) Closeness centrality (b) Betweenness centrality

Fig. 4. Top-10 points in the distance-based (star-shaped markers) and step-based (cross-shaped markers) centralities for Washington D.C.

where the top-10 nodes in the distance-based centrality are denoted by a star-shaped marker, and those in the step-based centrality are denoted by a cross-shaped marker. From Fig. 4(a), it is found that the top-10 locations in the closeness centrality are concentrated in a small area for both the distance-based and the step-based centralities. It seems natural that close neighboring nodes have similar values for the closeness centrality because the distances to another node could be similar regardless of the criterion whether it be distance-based or step-based. Here, remember that the rankings by the distance- and the step-based closeness centralities tend to be almost the same in Fig. 3. But, the actual areas in which the top-10 nodes are located are different in the distance- and the step-based centralities. Especially, the area is more limited and closer to the central area of the city for the distance-based centrality. This characteristic of the distance-based closeness centrality might be helpful for determining the point to build facilities such as a delivery center or a disaster shelter.

On the other hand, Fig. 4(b) in which the top-10 nodes in the distance- and the step-based betweenness centralities are plotted in the map exhibits a bit different tendency from Fig. 4(a). In Fig. 4(b), whether the measure is distance-based or step-based, the top-10 points are split into two groups, each of which is located on different streets that are apart from each other. The top-10 nodes in the distance-based betweenness centrality have a higher degree of concentration than those in the step-based one, similarly to the case of closeness centrality in Fig. 4(a). Further, the streets along which the top-10 points in the distance-based betweenness centrality are plotted are away from the streets along which the top-10 points in the step-based betweenness centrality are plotted. This tendency coincides with what we observed in Fig. 3 that the ranking by the distance-based betweenness centrality is substantially different from the ranking by the step-based betweenness centrality. Interestingly, the street that is close to the bottom-right corner and has three star-shaped markers is connecting to a freeway via a bridge, which is different from the other streets with markers. Indeed, one of the three points



(a) Closeness centrality (b) Betweenness centrality

Fig. 5. Top-10 points in the distance-based (star-shaped markers) and step-based (cross-shaped markers) centralities for Cairo

corresponds to the top node in the distance-based betweenness centrality. These points might be more critical in this urban traffic network than the others because it seems difficult to find alternative routes having a similar distance-based betweenness centrality value from the nearby street. If the street is closed and the traffic is blocked, we would have to go a very long way around. The distance-based betweenness centrality might be helpful to find such critical points in a traffic network.

We further plotted the top-10 nodes in each centrality measure on the street map for the Cairo dataset in the same fashion as in Fig. 4. The resulting maps are shown in Fig. 5. It is found that the findings we obtained above from Fig. 4 also hold in Fig. 5 although the two cities have different distributions of link distances as shown in Tab. 1. Note that the zoom level in Fig. 5 is different from the one in Fig. 4. It shows a wider area than Fig. 4 does. Thus, the groups of the top-10 nodes found in Fig. 5(b) are more distant from each other than those in Fig. 4(b). This is due to the difference between Washington D.C. and Cairo in the distribution of link distances.

5 Conclusion

In this paper, we first extended the conventional step-based closeness and betweenness centralities to analyze spatial networks. Unlike these conventional centralities that adopt the number of links to be traversed to reach one node from another as the distance between them, the extended distance-based closeness and betweenness centralities take into account the inter-nodes link distances obtained from the positions of nodes. They are natural extensions of the conventional centralities and general enough to include their definitions as a special case. Second, we have proposed two novel techniques to improve the computational efficiency to compute the distance-based centralities. Both are based on graph cut and recursively applied to a network in order to reduce scanning size needed for computing the centralities. The TP (Top-down pruning) technique recursively decomposes a network into two disjoint sub-networks that are connected by a cut link, while the BP (Bottom-up pruning) technique recursively removes a degree-one node by eliminating a cut link adjacent to it before the TP technique is applied. Note that it is straightforward to extend these techniques so that they can deal with directed networks.

We conducted extensive experiments using real-world road networks of eight different cities that have different distributions of link distances. Major findings we obtained through the experiments are that 1) applying both techniques can improve the computation efficiency to calculate the closeness and betweenness centralities about twice as much of the baseline method, 2) the actual improvement rates depend on networks, and 3) the BP technique is more effective than the TP technique due to the difference in the cost of finding cut links to remove. We further found that the node ranking by the distance-based closeness centrality is almost the same as the ranking by the step-based closeness centrality, while the ranking by the distance-based betweenness centrality is substantially different from the ranking by the step-based betweenness centrality. Locating the top-10 nodes in each centrality measure on an actual street map brought us further insights into their characteristics. The top-10 nodes in the distance-based centralities tend to be concentrated in a narrower area than those in the step-based ones do. Those in the distance-based closeness centrality are more likely to be located in an area closer to the center of a city than those in the step-based one are so. On the other hand, the top-10 nodes in the betweenness centralities tend to be split into two groups whether or not the measure is distance-based or step-based, and these two groups are geographically apart from each other. The distance-based closeness centrality seems more helpful to find locations to build facilities easy to access in a given area, while the distance-based betweenness centrality seems more useful to detect critical points in a traffic network. Finding these points in a network in an efficient way is one of our future directions. In addition, since centrality values of individual nodes can be computed independently, it is possible to parallelize those computations. This synergistically work with the proposed methods, leading to further improvement in efficiency and scalability. We are planning to empirically confirm this effect, too.

Acknowledgments

This material is based upon work supported by the Air Force Office of Scientific Research, Asian Office of Aerospace Research and Development (AOARD) under award number FA2386-16-1-4032, and JSPS Grant-in-Aid for Scientific Research (C) (No. 26330261).

References

1. Boldi, P., Vigna, S.: In-core computation of geometric centralities with hyperball: A hunderd billion nodes and beyond. In: Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW'13). pp. 621–628 (2013)
2. Brandes, U.: A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25, 163–177 (2001)

3. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30, 107–117 (1998)
4. Burckhart, K., Martin, O.J.: An interpretation of the recent evolution of the city of barcelona through the traffic maps. Journal of Geographic Information System 4(4), 298–311 (2012)
5. Chakrabarti, S., Dom, B., Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D., Kleinberg, J.: Mining the web's link structure. IEEE Computer 32, 60–67 (1999)
6. Chierichetti, F., Epasto, A., Kumar, R., Lattanzi, S., Mirrokni, V.: Efficient algorithms for public-private social networks. In: Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15). pp. 139–148 (2015)
7. Cohen, E.: Size-estimation framework with applications to transitive closure and reachability. Journal of Computer and System Sciences 55, 441–453 (1997)
8. Cohen, E.: All-distances sketches, revisited: Hip estimators for massive graphs analysis. In: Proceedings of the 33rd ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. pp. 88–99 (2015)
9. Cohen, E., Delling, D., Pajor, T., Werneck, R.F.: Sketch-based influence maximization and computation: Scaling up with guarantees. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. pp. 629–638 (2014)
10. Crucitti, P., Latora, V., Porta, S.: Centrality measures in spatial networks of urban streets. Physical Review E 73(3), 036125 (2006)
11. Freeman, L.: Centrality in social networks: Conceptual clarification. Social Networks 1, 215–239 (1979)
12. Kimura, M., Saito, K., Ohara, K., Motoda, H.: Speeding-up node influence computation for huge social networks. International Journal of Data Science and Analytics 1, 1–14 (2016)
13. Montis, D.A., Barthelemy, M., Chessa, A., Vespignani, A.: The structure of interurban traffic: A weighted network analysis. Environment and Planning B: Planning and Design 34(5), 905–924 (2007)
14. Moritz, H.: Geodetic reference system 1980. Journal of Geodesy 74(1), 128–133 (2000)
15. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167–256 (2003)
16. Ohara, K., Saito, K., Kimura, M., Motoda, H.: Resampling-based framework for estimating node centrality of large social network. In: Proceedings of the 17th International Conference on Discovery Science (DS'14). pp. 228–239. LNAI 8777 (2014)
17. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks 32(3), 245–251 (2010)
18. Park, K., Yilmaz, A.: A social network analysis approach to analyze road networks. In: Proceedings of the ASPRS Annual Conference 2010 (2010)
19. Saito, K., Kimura, M., Ohara, K., Motoda, H.: Super mediator - a new centrality measure of node importance for information diffusion over social network. Information Sciences 329, 985–1000 (2016)
20. Wang, P., Hunter, T., Bayen, A.M., Schechtner, K., Gonzalez, M.C.: Understanding Road Usage Patterns in Urban Areas. Scientific Reports 2 (2012)