RESEARCH

Multiple perspective centrality measures based on facility location problem under inter-group competitive environment

Takayasu Fushimi[1*] , Seiya Okubo[2] and Kazumi Saito[3,4]

*Correspondence:
takayasu.fushimi@gmail.com
[1] School of Computer
Science, Tokyo University
of Technology, 1404-1
Katakuramachi, Hachioji City,
Tokyo 192-0982, Japan
Full list of author information
is available at the end of the
article

Abstract

In this study, we propose novel centrality measures considering multiple perspectives of nodes or node groups based on the facility location problem on a spatial network. The conventional centrality exclusively quantifies the global properties of each node in a network such as closeness and betweenness, and extracts nodes with high scores as important nodes. In the context of facility placement on a network, it is desirable to place facilities at nodes with high accessibility from residents, that is, nodes with a high score in closeness centrality. It is natural to think that such a property of a node changes when the situation changes. For example, in a situation where there are no existing facilities, it is expected that the demand of residents will be satisfied by opening a new facility at the node with the highest accessibility, however, in a situation where there exist some facilities, it is necessary to open a new facility some distance from the existing facilities. Furthermore, it is natural to consider that the concept of closeness differs depending on the relationship with existing facilities, cooperative relationships and competitive relationships. Therefore, we extend a concept of centrality so as to considers the situation where one or more nodes have already been selected belonging to one of some groups. In this study, we propose two measures based on closeness centrality and betweenness centrality as behavior models of people on a spatial network. From our experimental evaluations using actual urban street network data, we confirm that the proposed method, which introduces the viewpoints of each group, shows that there is a difference in the important nodes of each group viewpoint, and that the new store location can be predicted more accurately.

Keywords: Centrality measure, Multiple perspectives, Facility location problem, Intergroup competition, Spatial network

Introduction

In recent years, networks have been widely observed around us, and the technology of network science has been applied to various real-world problems. As a typical method of social network analysis, there is a centrality measure that extracts important nodes from a large amount of nodes constituting a network. The conventional centrality exclusively quantifies the global properties of each node in a network such as closeness and betweenness, and extracts nodes with high scores as important

nodes. These measures are applicable not only to social networks but also to spatial networks including road networks, where junctions and streets between them are regarded as nodes and links, respectively. In the context of road networks, high closeness nodes and high betweenness nodes have meanings of highly accessible junctions and frequently passed junctions, respectively. Nodes with such promising characteristics might be applicable to decide the effective location of a facility that plans to open. From the standpoint of facility builders, in order to satisfy the demands of many customers and maximize the profits of the facility, it is necessary to be located at a site that is easily accessible for local residents, and for people on the move. In order to satisfy the needs of more customers, it is necessary to place all the facilities in a balanced manner throughout the network. In other words, when opening a new store, it is necessary to consider the location of existing facilities and to find a location that improves accessibility for all customers. The classical closeness and betweenness centralities only quantify global and exclusive properties of each node (candidate location) and do not take into account the current situation such as the location of existing facilities. On the other hand, group centrality, which defines centrality scores for groups of nodes, was introduced (Everett and Borgatti 1999). When adding a new node (new facility) to a group (existing facility) based on the concept of group centrality, it is natural to select the node that can raise the group centrality score most. This is equivalent to the greedy solution method for the combinatorial optimization problem that finds the combination of nodes so as to maximizes the group centrality score. In this way, by combining group centrality and the greedy solution method, it is possible to calculate the centrality score for each node considering the current situation. In addition, generally speaking, in facilities such as convenience stores, gas stations, and supermarkets, there are some classes like brands, chains, affiliates, etc., and facilities in the same class function cooperatively with each other, and facilities in different classes function competitively. Therefore, we extend a concept of centrality so as to considers the situation where one or more nodes have already been selected belonging to one of some groups.

In this study, by considering the above-mentioned context of the facility location problem, we propose novel centrality measures considering multiple perspectives of nodes or node groups. Concretely, we consider the situation that there are some facility groups each of which has competitive relationships with other groups, and each facility belonging to the same group has a cooperative relationship. When opening a new facility of a certain group, the location is desirable that could acquire more customers from the facilities of contending groups rather than those of the same group. When identifying the customers (trading-area) of each facility, our measures take two behavioral models of people on a spatial network (Fushimi and Yazaki 2020). The first one models such behavior as each person living at a node goes to the nearest facility, which is based on the concept of closeness centrality. The second one models such behavior as each person living at a node drops by at a facility that is passed most frequently on the shortest routes to various destinations, which is based on the concept of betweenness centrality. Based on these behavioral models, each resident (node) is assigned one of the facilities, that is customers of each facility are identified. In this manner, our measures quantify the number of customers expected to obtain from other groups and rank the importance of each

Table 1 List of abbreviations

Abbreviations	Meaning
CLC	Classical closeness centrality
SPCLC	Single perspective closeness centrality
MPCLC	Multiple perspective closeness centrality
BWC	Classical betweenness centrality
SPBWC	Single perspective betweenness centrality
MPBWC	Multiple perspective betweenness centrality

node, where the extracted important nodes differ depending on the perspective of each group.

In this paper, we substantially extended our previous study (Fushimi et al. 2019a), by adding new content as follows:

- We proposed a new measure based on betweenness centrality, i.e., in the conference paper, we only proposed and evaluated closeness centrality based measure.
- We added research on facility location in a discrete space (Hotelling 1929; Drezner 1994; Drezner and Drezner 1996; Drezner et al. 1998, 2011) to the references and discuss these related studies in "Related work" section. Through that discussion, we further clarify the originality of our work in the field.
- We provide additional experimental results of prediction accuracy in "Comparison of highly ranked nodes" section.
- We also revised and extended our Introduction and Conclusion according to the above-mentioned additions.

The paper is organized as follows. "Related work" section describes related work. "Problem setting" and "Proposed measure" section give our problem setting and explain our proposed method. In "Experiments" section, we report and discuss experimental results using real-world data. Finally, "Conclusion" section concludes this paper and addresses future work.

Table 1 is a list of abbreviations used in this paper.

Related work

In this section, we briefly review existing work on spatial network analysis using centrality measures and the facility location problem over a network.

Centrality analysis of spatial networks

There are many studies that analyzed road networks by network analysis approach (Crucitti et al. 2006; Montis et al. 2007; Park and Yilmaz 2010; Tabata et al. 2017; Fushimi et al. 2019b). Crucitti et al. analyzed the distribution of four centrality indices in a road network considering distance weights between junctions (Crucitti et al. 2006). The area with a similar road structure is classified by the fitting parameter and Gini coefficient of the centrality distribution. Montis et al. analyzed multiple undirected networks with municipalities as nodes and commuter traffic between municipalities

Fushimi *et al. Appl Netw Sci*   (2020) 5:80

Page 4 of 21

as weighted links (Montis et al. 2007). The relationship between the degree and the clustering coefficient indicates that there is a hierarchy in the municipality and that there is a positive correlation between the centrality index and the population or wealth of residents. Park et al. evaluated the difference in the topological structure of residential areas and downtown areas by applying the centrality index to the road network and calculating its entropy (Park and Yilmaz 2010). In the Tabata et al.'s method, the top node in the closeness centrality ranking, which is equal to the node selected first in the greedy method of k-medoids clustering, is calculated very quickly, and that node can be regarded as a facility location site (Tabata et al. 2017). Fushimi et al. proposed a centrality measure that quantifies the connectivity of each node based on the expected value of reachable nodes in an uncertain graph, modeling the road blockage that occurs stochastically due to a natural disaster as an uncertain graph (Fushimi et al. 2019b). This centrality is a technique that can be applied to select evacuation facility sites that can be reached by more residents under the event of a disaster.

The centrality measure quantifies, in principle, the unique properties of a node with respect to the entire network. Therefore, adjacent nodes tend to have similar scores due to overlapping effects of adjacent nodes. Nodes with high scores are not always suitable for sites that allocate facilities because of the potential for concentration and competition. To overcome this problem, group centrality, which defines centrality scores for groups of nodes, was introduced (Everett and Borgatti 1999). The previous work (Everett and Borgatti 1999) compared the maximum, minimum, and average values as the definition of the distance between a group and a node outside the group in the calculation of group-closeness centrality. In the context of this study, residents go to the nearest facility, that is, select the node with the shortest distance among the nodes in the group, so we consider the group-closeness centrality with the minimum distance. Furthermore, finding a node set that minimizes the objective function of group closeness centrality is equivalent to solving the p-median problem, which is a facility-location problem, based on the graph distance. Our measure can be regarded as a generalization of group centrality, where multiple groups with cooperative intragroup and competitive intergroup relationships are considered. To the best of our knowledge, this is the first study that considers such cooperative and competitive relationships to analyze a road network and applies it to facility location issues.

As another stream of centrality research, modular centrality has been developed in recent years, which is an extended version of classical centrality measures in terms of two perspectives, local and global (Ghalmane et al. 2019a, b; Cherifi et al. 2019). This measure considers the local influence of each node in its belonging community by calculating the score over a network where inter-community links are removed from an original network, and the global one to the other communities by calculating the score over a network where intra-community links are removed. Experimental results using synthetic and real-world datasets showed influential nodes on information diffusion can be extracted more accurately by the modular centrality than by classical centralities. Although the framework for extracting important nodes from multiple viewpoints is common to our proposed method, it is different from this research, which focuses on viewpoints among multiple competing node-groups.

Fushimi *et al. Appl Netw Sci*   (2020) 5:80

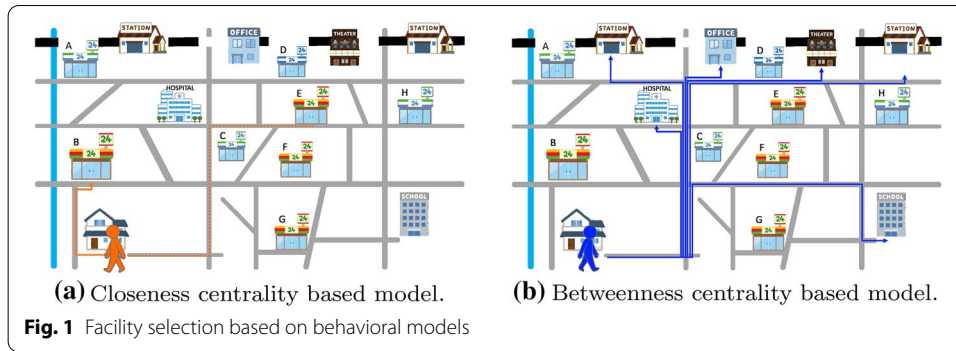Page 5 of 21

Facility location problem on networks

A vast amount of studies about facility location in a network have been conducted and the purpose of which is to determine the best place to build one or more new facilities. Studies on facility location are roughly classified into three streams. The most classical stream of studies includes p-center or p-median problems, where demand points (residents, customers) select the closest facility by physical distance. In this assumption, the most simple way to estimate shares is the proximity approach (Hotelling 1929), the idea of which is equal to our closeness based model. Tabata et al. proposed an approximation algorithm for a 1-median problem as an identification problem of the top node of closeness centrality (Tabata et al. 2017). In our experiments, we compare with classical closeness centrality, which can be interpreted as a p-median problem stated in (Tabata et al. 2017). To solve the p-median and the p-center problems in polynomial or pseudo-polynomial time, lots of algorithms have been developed (Thorup 2004; Rahmaniani and Ghaderi 2013; Agra et al. 2017; Tamir 2001; Jinmei and Kejia 2010; Gimadi 2017; Puerto et al. 2018). But these algorithms need much computation time so the experiments were conducted using small-size graphs with a simple topological structure like tree or line.

The second stream is based on gravity model (Huff 1964), which assumes the consumer selects one facility from the possible candidates in a stochastic manner (Drezner et al. 1998). As a similar approach, the cover-based approach assumes that each facility has a certain radius of influence, and that consumers within that range will choose the facility (Drezner et al. 1998, 2011). In studies of the above-mentioned first stream, customers select one of the facilities according to the distance, while in the gravity-based and the cover-based approaches, customers choose some facilities stochastically.

The last one is based on facility design models, where market share, facility attractiveness, buying power (demand), and distance can be factors to model the competitive environment (Hotelling 1929; Drezner 1994; Aboolian et al. 2007, 2020). In our proposed method, the market share that is the sum of influential area of the existing facility in each group is calculated using two behavioral models, and the location where more customers of the competing group can be obtained is output as the important node (location). To take into account the differences in facility attractiveness, the utility approach calculated the utility of a facility for a consumer, which incorporates the distance between them and the attractiveness of the facility like floor space or price bracket (Drezner 1994; Drezner and Drezner 1996). These methods and our approach are developed based on the same mind, but our method is different in that all the nodes of the network are ranked and the upper node is suggested as a candidate location of a newly opening facility. Models assumed in all of the above-mentioned studies are different from our newly proposed model in which people drop by the facility located along the shortest paths to their destinations.

Problem setting

First, we formally introduce the problem to be tackled in this study. Let $G = (\mathcal{V}, \mathcal{E})$ be the undirected graph structure of a given spatial network, where $\mathcal{V} = \{u, v, w, \ldots\}$ is a set of nodes that correspond to junctions and $\mathcal{E} = \{e = (u, v), \ldots\}$ is a set of links that correspond to the roads between junctions. Let $\mathcal{D} \subset \mathcal{V}$ be a set of junction nodes that already

(a) Closeness centrality based model.   **(b)** Betweenness centrality based model.

Fig. 1 Facility selection based on behavioral models

have facilities (stores). Then, a normal node $\mathcal{V} \setminus \mathcal{D}$, where no facility is built, selects a facility according to certain criteria and is included in the influential area, what-is-called trading area, of one of the facilities.

In this study, we consider two types of movement behaviors (Fushimi and Yazaki 2020): the closeness centrality based model and the betweenness centrality based model. The closeness based model is a behavioral model that assumes people move from the residence to the nearest facility (store). This can be said to be a behavior model whose purpose is to go to the store. In Fig. 1a, we show an example of closeness based model, when the orange person goes to a convenience store, she would select the nearest store, B, rather than distant stores like D, E, and H. Therefore, for each facility node $v \in \mathcal{D}$, we can obtain the following trading-area of v, $\mathcal{C}(v; \mathcal{D})$, such that its nearest facility is v:

$$\mathcal{C}(v; \mathcal{D}) = \{u \in \mathcal{V} \mid d(u, v) < \min_{w \in \mathcal{D} \setminus \{v\}} d(u, w)\},$$

where $d(u, w)$ is a distance between node u and w.

The betweenness based model is a behavioral model that assumes people drop in at facilities (stores) on the route when traveling to destinations such as schools, workplaces, sightseeing spots, and hospitals. This model is different from the closeness based model in which the store is assumed to be the destination in that it is assumed to be a stopover on the way from the starting point to the destination. In this model, residents select a store that most frequently appears over the shortest route toward various destinations. In Fig. 1b, we show an example of betweenness based model, when the blue person goes to each of destinations like the west station, the office, the theater, the east station, the hospital, and the school. When going to these destinations by the shortest route, he passes store C five times, store D twice, and store F once. Therefore, it is most likely to drop in store C that passes most frequently.

The ratio of passing through store v over the shortest path between starting node u and the various destinations t can be defined as

$$\delta_u(v) = \sum_{t \in \mathcal{V} \setminus \{u\}} \frac{\sigma_{u,t}(v)}{\sigma_{u,t}}, \tag{1}$$

where $\sigma_{u,t}$ is the number of shortest paths from the starting node u to the destination node t and $\sigma_{u,t}(v)$ is the number of those paths that pass through the store node v.

Although in Eq. (1), all nodes, $t \in \mathcal{V}$, are treated as destinations, it is also possible to set hospital, school, station, and so on like Fig. 1b as destinations.

Destinations are For each facility node $v \in \mathcal{D}$, we can obtain the following trading-area of v, $\mathcal{B}(v; \mathcal{D})$, such that its most-frequently-passing facility is v:

$$\mathcal{B}(v; \mathcal{D}) = \left\{ u \in \mathcal{V} \,|\, \delta_u(v) > \max_{w \in \mathcal{D} \setminus \{v\}} \delta_u(w) \right\}.$$

In this way, we compute the influential area of each existing facility $v \in \mathcal{D}$. Then, our problem to tackle in this paper is finding an optimal location that a store that plans to open can acquire more customers from these trading areas of existing facilities.

Proposed measure

In this section, we explain our proposed measure that extracts candidate sites for new facilities (stores) so that more residents (customers) can be obtained from competitors and the trading area can be expanded. For each node $v \in \mathcal{V}$, we assume that v has some weight denoted by $n(v)$ that can represent the number of residents around node v in a road network. Then, we can obtain the following weighted sum of nodes, $f(v; \mathcal{D})$, which can be interpreted as the number of residents whose nearest facility is v.

$$f(v; \mathcal{D}) = \sum_{u \in \mathcal{C}(v; \mathcal{D})} n(u) \tag{2}$$

Similarly, by replacing the $\mathcal{C}$ with $\mathcal{B}$ in Eq. 2, we can also obtain the weighted sum of nodes, which can be interpreted as the number of residents whose most-frequently-passing facility is v. Then, as a new node to be added to the set of facility nodes, we can compute the following node, $\hat{x} \in \mathcal{V} \setminus \mathcal{D}$, whose weighted sum becomes the maximum value.

$$\hat{x} = \arg\max_{x \in \mathcal{V} \setminus \mathcal{D}} f(x; \mathcal{D} \cup \{x\}) \tag{3}$$

Namely, we can assume that the node $\hat{x}$ has the largest number of residents as their nearest or most-frequently-passing facility.

In our problem setting, we also assume that each facility belongs to one of K classes (chains), functioning cooperatively with facilities of the same class but competitively among different classes. Hereafter, we express the class of each facility $v \in \mathcal{D}$ as an integer denoted by $z(v) \in \{1, \ldots, K\}$, and define the set of facility belonging to the class k as $\mathcal{D}_k = \{v \in \mathcal{D} | z(v) = k\}$. Then, we can obtain the following set of nodes $\mathcal{V}_k$ such that the class of their nearest facilities is k.

$$\mathcal{V}_k = \bigcup_{v \in \mathcal{D}_k} \mathcal{C}(v; \mathcal{D})$$

Similarly, by replacing the $\mathcal{C}$ with $\mathcal{B}$, we can obtain the set of nodes such that the class of their most-frequently-passing facilities is k. Furthermore, we can obtain the following partial weighted sum of nodes, $f_k(x; \mathcal{D} \cup \{x\})$, which can be interpreted as the number of residents whose nearest facility is x, and the classes of their former nearest facilities without x are not k.

$$f_k(x; \mathcal{D} \cup \{x\}) = \sum_{u \in \mathcal{C}(x; \mathcal{D} \cup \{x\}) \backslash \mathcal{V}_k} n(u) \tag{4}$$

Similarly, by replacing the $\mathcal{C}$ with $\mathcal{B}$ in Eq. 4, we can obtain the number of residents whose most-frequently-passing facility is x, and the classes of their former most-frequently-passing facilities without x are not k.

Then, as a new node to be added to the set of facility nodes for each class k, we can compute the following node, $\hat{x}_k \in \mathcal{V} \backslash \mathcal{D}$, whose partial weighted sum becomes the maximum value.

$$\hat{x}_k = \underset{x \in \mathcal{V} \backslash \mathcal{D}}{\arg \max} f_k(x; \mathcal{D} \cup \{x\}) \tag{5}$$

Namely, we can assume that the node $\hat{x}_k$ has the largest number of residents such that the classes of their former nearest or most-frequently-passing facilities without x are not k. Hereafter, we refer to the methods that extract a node according to Eq. 3 based on closeness and betweenness models as Single Perspective CLoseness Centrality (SPCLC) and Single Perspective BetWeenness Centrality (SPBWC), respectively. Similarly, we call the methods that extract a node according to Eq. 5 based on closeness and betweenness models to Multiple Perspective CLoseness Centrality (MPCLC) and Multiple Perspective BetWeenness Centrality (MPBWC), respectively.
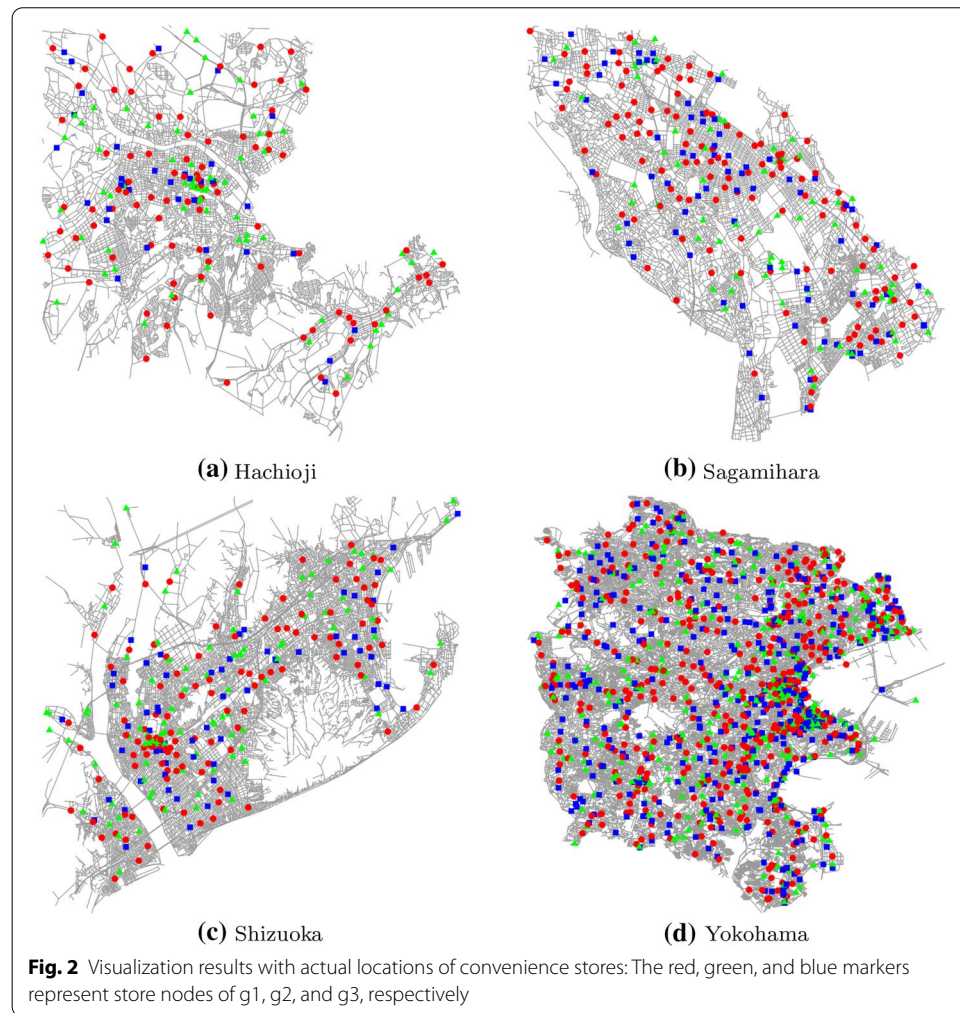
Experiments

Dataset

In our experiments, we selected the following five cities, Hachioji (Tokyo), Sagamihara (Kanagawa), Shizuoka (Shizuoka), Yokohama (Kanagawa) and Naha (Okinawa) as target areas, each of whose road structure was collected from OpenStreetMap[1] and extracted all junctions and roads. We then constructed a spatial network with the junctions as the nodes and the roads between the junctions as the links. For each city, we collected the actual location information of convenience stores that belong to the three major chains in Japan from the navigation service site NAVITIME.[2] In our experiment, the junction closest to the actual location of the store is approximately treated as the store node. If there are multiple stores very close to each other, they are assigned to one node. In such a case, the proposed measure can be calculated by apportioning the number of customers among the stores assigned to the same node. We used Hachioji, Sagamihara, Shizuoka, and Yokohama for the confirmation of the difference between the proposed method and compared method, and Naha for the evaluation of prediction accuracy. In this study, we regard that the convenience stores are located at some of the nodes and the residents lived at other nodes. Table 2 shows the number of nodes where convenience stores of three chains (g1: 7-Eleven, g2: FamilyMart and g3: Lawson) exist and the number of normal nodes where no store exists. In Fig. 2, the map of the former four cities are shown, where red circles, green triangles and blue squares are the actual locations of convenience stores of g1, g2, and g3, respectively.

Fushimi *et al. Appl Netw Sci* (2020) 5:80

Page 9 of 21

Table 2 The numbers of nodes where stores belonging to each group (g1: 7-Eleven, g2: FamilyMart and g3: Lawson) exist and normal nodes

City	g1	g2	g3	Normal nodes	Total
Hachioji	95	76	36	12,117	12,324
Sagamihara	132	77	79	18,822	19,110
Shizuoka	118	106	65	30,752	31,041
Yokohama	492	417	428	123,379	124,716
Naha	12	72	58	4408	4550

(a) Hachioji **(b)** Sagamihara

(c) Shizuoka **(d)** Yokohama

Fig. 2 Visualization results with actual locations of convenience stores: The red, green, and blue markers represent store nodes of g1, g2, and g3, respectively

In our experiments, we employ the geodesic distance for each pair of nodes, i.e., the distance between directly connected nodes is approximately calculated by the Hubeny formula and between nodes that are not directly connected is computed according to the Dijkstra's algorithm. In addition, we assume that residents live equally at each node, $n(u) = 1$, regardless of residential area, urban area, or mountainous area. However, a more realistic analysis is possible by assigning the number of inhabitants at each node to $n(u)$ obtained from the population density data.

Fushimi *et al. Appl Netw Sci* (2020) 5:80

Page 10 of 21

Measures used for comparison

Each of our proposed methods is based on closeness centrality or betweenness centrality, so we compare to results by these centrality measures. Let $d(u, v)$ be the distance between node u and v, and the closeness of node u is defined as the inverse of harmonic average of distances to other nodes like

$$\text{clc(u)} = \frac{1}{|\mathcal{V}| - 1} \sum_{v \in \mathcal{V}, v \neq u} d(u, v)^{-1} n(v).$$

We can extract highly-accessible nodes by $\arg\max_{x \in \mathcal{V} \setminus \mathcal{D}} \text{clc}(x)$, and we refer to this method as CLC. The reason why we employed harmonic centrality is its notion corresponds to our closeness model where residents go to stores in the neighborhood. That is, it is intended to extract nodes that are close to other nodes in a local sense rather than in a global sense. In addition, population weight $n(v)$ can be easily and naturally introduced to harmonic centrality.

Similarly, the betweenness of node u is defined as arithmetic average of ratio of the number of shortest paths that pass through node u, $\sigma_{s,t}(u)$ to the total number of shortest paths from s to t, $\sigma_{s,t}$ like

$$\text{bwc(u)} = \frac{1}{|\mathcal{V}| - 1} \frac{1}{|\mathcal{V}| - 2} \sum_{s \in \mathcal{V} \setminus \{u\}} \sum_{t \in \mathcal{V} \setminus \{u,s\}} \frac{\sigma_{s,t}(u)}{\sigma_{s,t}} n(s).$$

We can extract heavily-passing nodes by $\arg\max_{x \in \mathcal{V} \setminus \mathcal{D}} \text{bwc}(x)$, and we refer to this method as BWC.

Comparison of extracted trade-area

First, in order to show how many customers the location extracted by each method will capture from the competitive group, we validate the expected number of covered nodes (acquired customers), $f(x; \mathcal{D})$, which can be regarded as the size of trade-area of stores or groups, when a new store opens at the extracted top node x. Here we recall that, in our experiments, residents are assumed to live equally at each node, i.e., we set $n(u) = 1$ for all the node $u \in \mathcal{V}$.

For a given normal node x, we introduce its breakdown vector as $\mathbf{g}(x)$ whose k-th element is defined by $g(x)_k = |\mathcal{C}(x; \mathcal{D} \cup \{x\}) \cap \mathcal{V}_k|$. We extract the candidate node using some methods and plot the expected number of covered nodes $f(\cdot; \mathcal{D})$ and their breakdown of each group:

- MPCLC/MPBWC for class k: $\hat{x}_k = \arg\max_{x \in \mathcal{V} \setminus \mathcal{D}} f_k(x; \mathcal{D} \cup \{x\})$;
- SPCLC/SPBWC for class k: $\hat{y}_k = \arg\max_{y \in \mathcal{V} \setminus \mathcal{D}} f(y; \mathcal{D}_k \cup \{y\})$;
- SPCLC/SPBWC for all: $\hat{y} = \arg\max_{y \in \mathcal{V} \setminus \mathcal{D}} f(y; \mathcal{D} \cup \{y\})$;
- CLC/BWC: $\hat{z} = \arg\max_{z \in \mathcal{V} \setminus \mathcal{D}} \text{clc}(z)$ or $\hat{z} = \arg\max_{z \in \mathcal{V} \setminus \mathcal{D}} \text{bwc}(z)$;
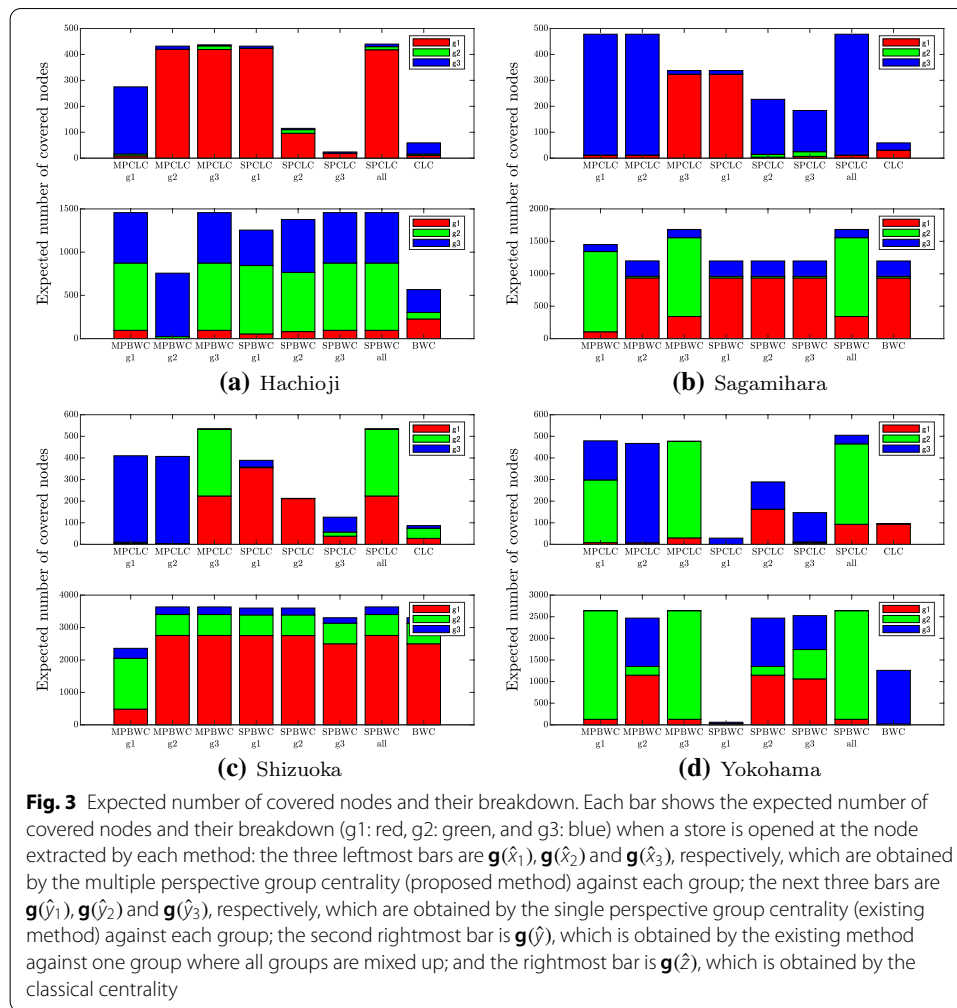
Fig. 3 Expected number of covered nodes and their breakdown. Each bar shows the expected number of covered nodes and their breakdown (g1: red, g2: green, and g3: blue) when a store is opened at the node extracted by each method: the three leftmost bars are $\mathbf{g}(\hat{x}_1)$, $\mathbf{g}(\hat{x}_2)$ and $\mathbf{g}(\hat{x}_3)$, respectively, which are obtained by the multiple perspective group centrality (proposed method) against each group; the next three bars are $\mathbf{g}(\hat{y}_1)$, $\mathbf{g}(\hat{y}_2)$ and $\mathbf{g}(\hat{y}_3)$, respectively, which are obtained by the single perspective group centrality (existing method) against each group; the second rightmost bar is $\mathbf{g}(\hat{y})$, which is obtained by the existing method against one group where all groups are mixed up; and the rightmost bar is $\mathbf{g}(\hat{z})$, which is obtained by the classical centrality

Figure 3 represents the expected number of covered nodes and their breakdown of each group when opening a new store at the top node ranked by each method, where red, green, and blue bars mean group g1, g2, and g3, respectively. From Fig. 3, we can confirm the following observations.

1. By the multiple perspective methods, the breakdown of covered nodes differs according to the viewpoint, g1, g2, and g3. For example, in the results of Hachioji, the node extracted by MPCLC from the viewpoint of group g1 covers more nodes belonging to group g3 (blue) rather than its own group g1 (red), and from the viewpoint of g2 covers more nodes belonging to g1.

2. The expected numbers of nodes covered by the nodes extracted by the betweenness based methods are larger than those by the closeness based methods. Especially for Shizuoka, the difference is about 7 times.

3. By the single perspective methods against each class, the breakdown of covered nodes is not adequate. For example, in the results of Hachioji, the node extracted by SPCLC against g1 covers quite many nodes belonging to group g1 (red) rather than the other groups g2 (green) and g3 (blue), i.e., internecine strife has occurred.

4 By SPCLC and SPBWC against all classes, the total amount of covered nodes is the most.

5 The total amount of covered nodes by proposed methods are much larger than those of classical centrality measures, CLC and BWC, except for the BWC method in Shizuoka.

By the multiple perspective methods, MPCLC and MPBWC, we can obtain the location where nodes of other groups can be covered, according to the different viewpoints, but in some cases, even if the viewpoint is changed, the nodes in each group are covered at a similar proportion such as MPCLC for g2 and g3, and MPBWC for g1 and g3 in Hachioji. This is probably because the store opening strategies of these groups in the area are similar. The sizes of covered nodes of MPCLC and MPBWC are remarkably different because the closeness based methods assume the local movement such as from residence to the nearest facility, while the betweenness based methods assume the global movement such as from residence to hospital, school, station, and so on. In addition, the betweenness based methods for Shizuoka, the expected number of covered nodes are relatively large on the whole. From this, it can be predicted that existing stores in Shizuoka are not located in places with high betweenness centrality, and by locating new stores in such places, more customers can be captured.

Against each group, the single perspective methods, SPCLC and SPBWC, cover nodes at undesirable proportions. These results were obtained by calculating the centrality score based on Eq. (3), independently for each group, g1, g2, and g3. In this way, the nodes far from the existing store turn out to be included in the trading-area of the new store, and the total utility of the inhabitants increases. On the other hand, since the locations of existing stores in other groups are not considered at all, even if a new store is opened at the extracted node, there is no guarantee that customers in other groups will be acquired. On the contrary, like SPCLC for g1 in Hachioji, there may be cases where more customers of the same group are captured.

Against one group where all groups are mixed up, SPCLC and SPBWC cover the largest number of nodes among the compared methods because they extract the nodes that can acquire more customers regardless of the belonging group, which is equal to the no restriction about acquiring from competing groups. Although classical centralities do not consider the existing facilities, so the expected number of covered nodes is relatively small, in Shizuoka, since the location of the node extracted by BWC was coincidentally a good location for g2 and g3, the expected number of acquired customers from g1 (red) got to be high.

From these results, we conclude that our methods can extract effective nodes that are expected to cover more nodes (acquire more customers) from other groups.

Prediction accuracy

Next, we evaluate our proposed methods in terms of prediction accuracy in order to show which method extracted the location more realistically in the scenario of new store opening. In these experiments, we utilized the road network of Naha because in Naha city, g1 (red) chain stores opened for the first time on July 11, 2019, and before that, only g2 (green) and g3 (blue) exist. Therefore, by using our proposed

methods, MPCLC and MPBWC, and compared methods, SPCLC and SPBWC, we predicted the location of newly opened g1 stores one by one, under the situation that only the stores of g2 and g3 are exist. Now, we define $\mathcal{D}_k^{(j)}$ as the set of store nodes belonging to the group k at the j-th node prediction step, i.e., at the initial state, $\mathcal{D}_1^{(0)} \leftarrow \emptyset$ for g1, and $\mathcal{D}_2^{(0)}$ and $\mathcal{D}_3^{(0)}$ are fixed to the actual locations of stores belonging to g2 and g3. Then, we predict the node where the g1 chain will open a new store in the following greedy manner:

$$\hat{x}_1^{(j)} \leftarrow \underset{x \in \mathcal{V} \setminus \mathcal{D}^{(j-1)}}{\arg \max} f_1(x; \mathcal{D}^{(j-1)} \cup \{x\}), \tag{6}$$

where f_1 is a function defined in Eq. (2) for group g1, and the j-th predicted node $\hat{x}_1^{(j)}$ are merged into the set of stores as $\mathcal{D}_1^{(j)} \leftarrow \mathcal{D}_1^{(j-1)} \cup \{\hat{x}_1^{(j)}\}$, and $\mathcal{D}^{(j)} \leftarrow \bigcup_{k \in \{1,2,3\}} \mathcal{D}_k^{(j)}$. Similarly, by using the compared methods, we predict the node by changing the function f_1 to f of Eq. (2) in Eq. (6).

Figure 4 plots the actual locations of newly opened 12 stores of the g1 chain and the locations predicted by each method as red nodes. The numbers in the red nodes indicate the order j in which the stores were opened. From Fig. 4, we can confirm the following observations.

1 Some nodes predicted by SPCLC and SPBWC are located more closely with each other than those by MPCLC and MPBWC.
2 Nodes by the closeness based methods are distributed more well-balanced throughout the network than those by the betweenness based methods.
3 Some of the nodes predicted by MPBWC, specifically 3, 5, 7, and 10, are located in the downtown area of Naha city where many actual stores are located.

From these observations, we can see that the nodes near the actual locations can be predicted by MPBWC compared to other methods.

In order to quantitatively evaluate the predictive accuracy, the following evaluation measure is introduced. Let $\mathcal{S}^{(j)} = \{s_1, \ldots, s_j\}$ be the set of nodes corresponding to the actual stores, and $\mathcal{R}_X^{(j)}$ be the set of store nodes predicted by a method $X \in \{\text{SPCLC}, \text{SPBWC}, \text{MPCLC}, \text{MPBWC}\}$. By using the distance $d(s, r)$ between nodes s and r, we define the average of minimum matching distances (MMD) as

$$\mathrm{MMD}(\mathcal{S}^{(j)}, \mathcal{R}_X^{(j)}) = \frac{1}{2} \left(\frac{1}{j} \sum_{i=1}^{j} \min_{1 \le h \le j} d(s_i, r_h) + \frac{1}{j} \sum_{i=1}^{j} \min_{1 \le h \le j} d(r_i, s_h) \right). \tag{7}$$

The first term in parentheses in the Eq. (7) represents the average distance between the i-th actual store node s_i and the nearest predicted nodes r_h. The second term represents the average distance between the i-th predicted location r_i and the nearest actual store s_h. The smaller the value of this measure, the closer the positional relationship between the predicted location and the actual store, and it can be said that the method can predict more accurately. As the distance, the graph distance, the Euclidean distance when connecting the nodes with a straight line, and the geodesic distance are used.
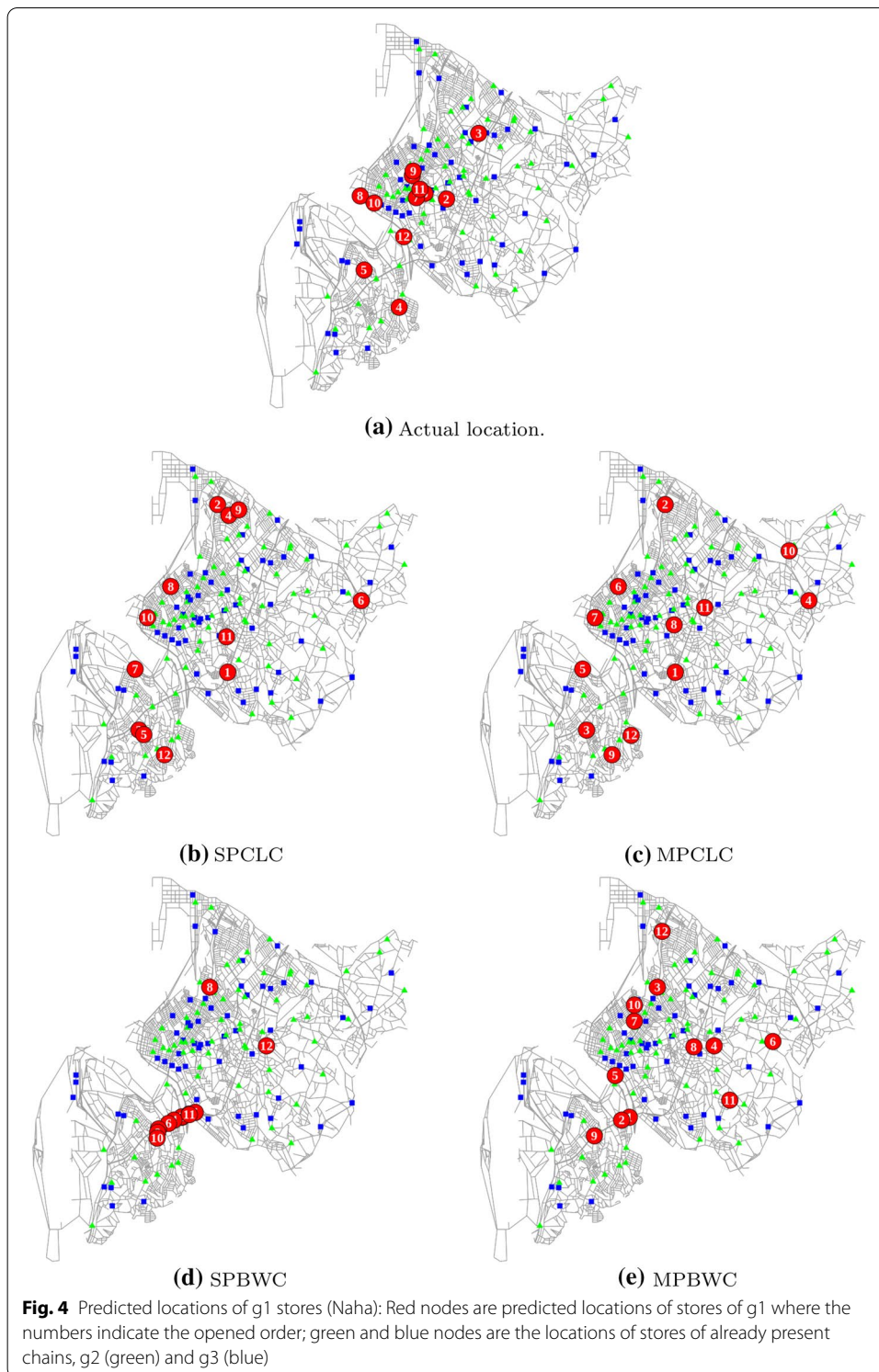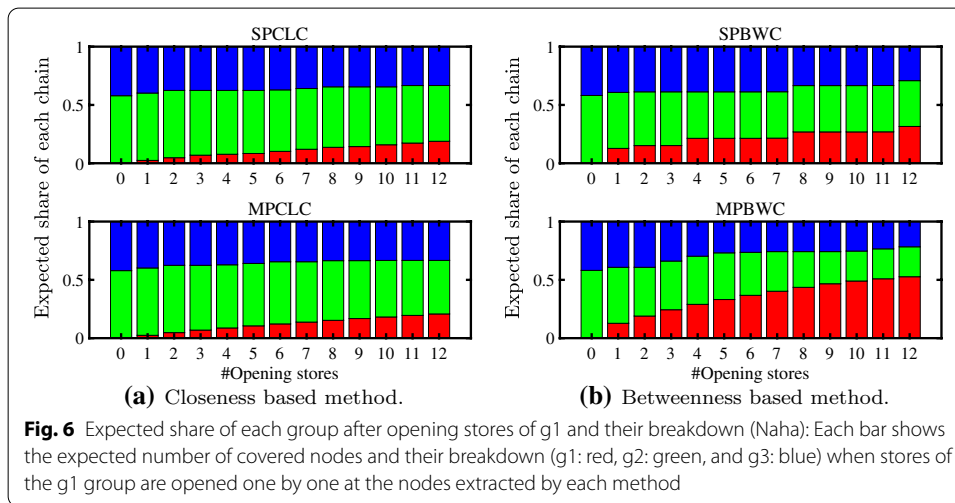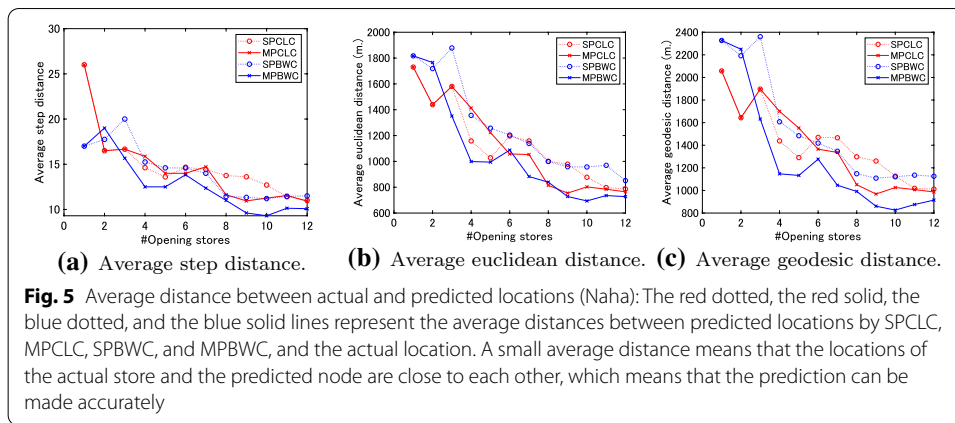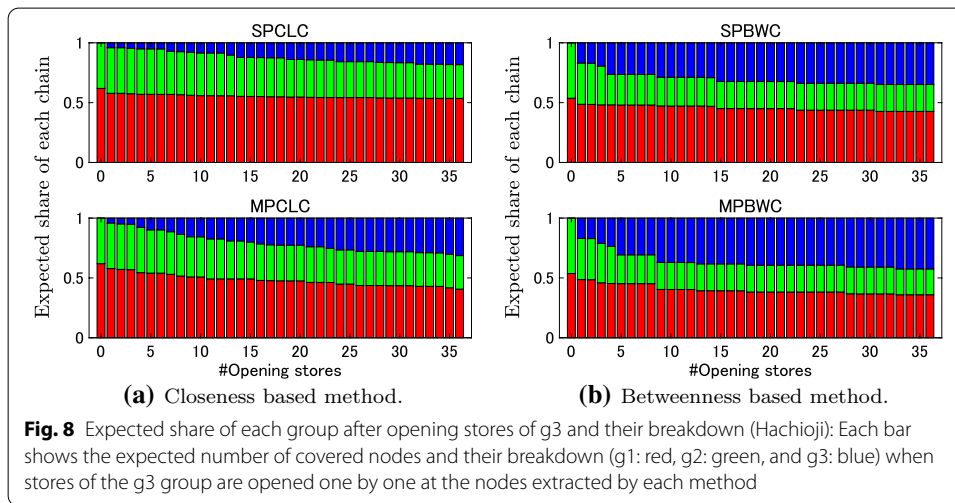
(a) Actual location.

(b) SPCLC

(c) MPCLC

(d) SPBWC

(e) MPBWC

Fig. 4 Predicted locations of g1 stores (Naha): Red nodes are predicted locations of stores of g1 where the numbers indicate the opened order; green and blue nodes are the locations of stores of already present chains, g2 (green) and g3 (blue)
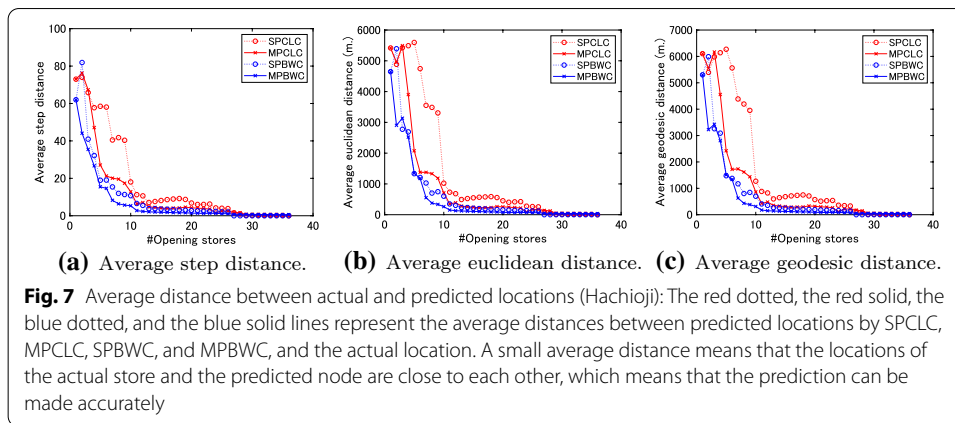
Figure 5 indicates the average distance with respect to the number of opening stores j, where the red dashed and solid lines are SPCLC and MPCLC, the blue dashed and solid lines are SPBWC and MPBWC, respectively. From Fig. 5, we can see that nodes

Fig. 5 Average distance between actual and predicted locations (Naha): The red dotted, the red solid, the blue dotted, and the blue solid lines represent the average distances between predicted locations by SPCLC, MPCLC, SPBWC, and MPBWC, and the actual location. A small average distance means that the locations of the actual store and the predicted node are close to each other, which means that the prediction can be made accurately

Fig. 6 Expected share of each group after opening stores of g1 and their breakdown (Naha): Each bar shows the expected number of covered nodes and their breakdown (g1: red, g2: green, and g3: blue) when stores of the g1 group are opened one by one at the nodes extracted by each method

predicted by MPCLC and MPBWC, which are multiple perspective methods, are closer to actual stores than those by SPCLC and SPBWC, which are single perspective methods, regardless of the distance measures.

Figure 6 depicts the expected share of each chain if the new g1 store opens at the predicted node, where red, green, and blue are g1, g2, and g3, respectively. In Fig. 6, the difference in SPBWC and MPBWC is remarkable. On the other hand, the difference in SPCLC and MPCLC does not seem to be so large compared to that of SPBWC and MPBWC. Thus, from Fig. 6, we can conclude that by opening a new g1 store at the node predicted by MPCLC, the g1 chain is expected to obtain a somewhat larger share from competing groups than SPCLC; and MPBWC can be expected to cover much larger nodes than SPBWC.

Next, by deleting all the 36 stores belonging to the g3 group from Hachioji, we artificially created the situation like the g1 group in Naha and predicted the location of a g3 store based on the above-mentioned greedy algorithm. Figures 7 and 8 show the average distance (Eq. (7)) with respect to the number of opening stores j, and the expected share of each chain if the new g3 store opens at the predicted node, respectively. From Fig. 7, we can confirm that nodes predicted by MPCLC and MPBWC are closer to actual stores

Fig. 7 Average distance between actual and predicted locations (Hachioji): The red dotted, the red solid, the blue dotted, and the blue solid lines represent the average distances between predicted locations by SPCLC, MPCLC, SPBWC, and MPBWC, and the actual location. A small average distance means that the locations of the actual store and the predicted node are close to each other, which means that the prediction can be made accurately

Fig. 8 Expected share of each group after opening stores of g3 and their breakdown (Hachioji): Each bar shows the expected number of covered nodes and their breakdown (g1: red, g2: green, and g3: blue) when stores of the g3 group are opened one by one at the nodes extracted by each method

than those by SPCLC and SPBWC regardless of the distance measures. From Fig. 8, we can conclude that by opening a new g3 store at the node predicted by MPCLC and MPBWC, the g3 chain is expected to obtain a larger share from competing g1 (red) and g2 (green) compared to SPCLC and SPBWC.

Comparison of highly ranked nodes

Finally, we confirm the top node extracted by the proposed and the compared methods in order to show that our measures are uncorrelated to existing ones. Figure 9 shows the visualization results of Hachioji, Sagamihara, Shizuoka and Yokohama with the top 100 nodes extracted by the proposed and the compared methods. In each figure, the red, green, and blue dots represent top nodes by proposed methods, MPCLC and MPBWC, for g1, g2 and g3, the black dots represent top nodes by compared methods, SPCLC and SPBWC, and the yellow dots represent top nodes by methods, CLC and BWC.

From Fig. 9, for all the cities, we can see the following observations.

1   In all the results, top nodes by each method are different to some extent.
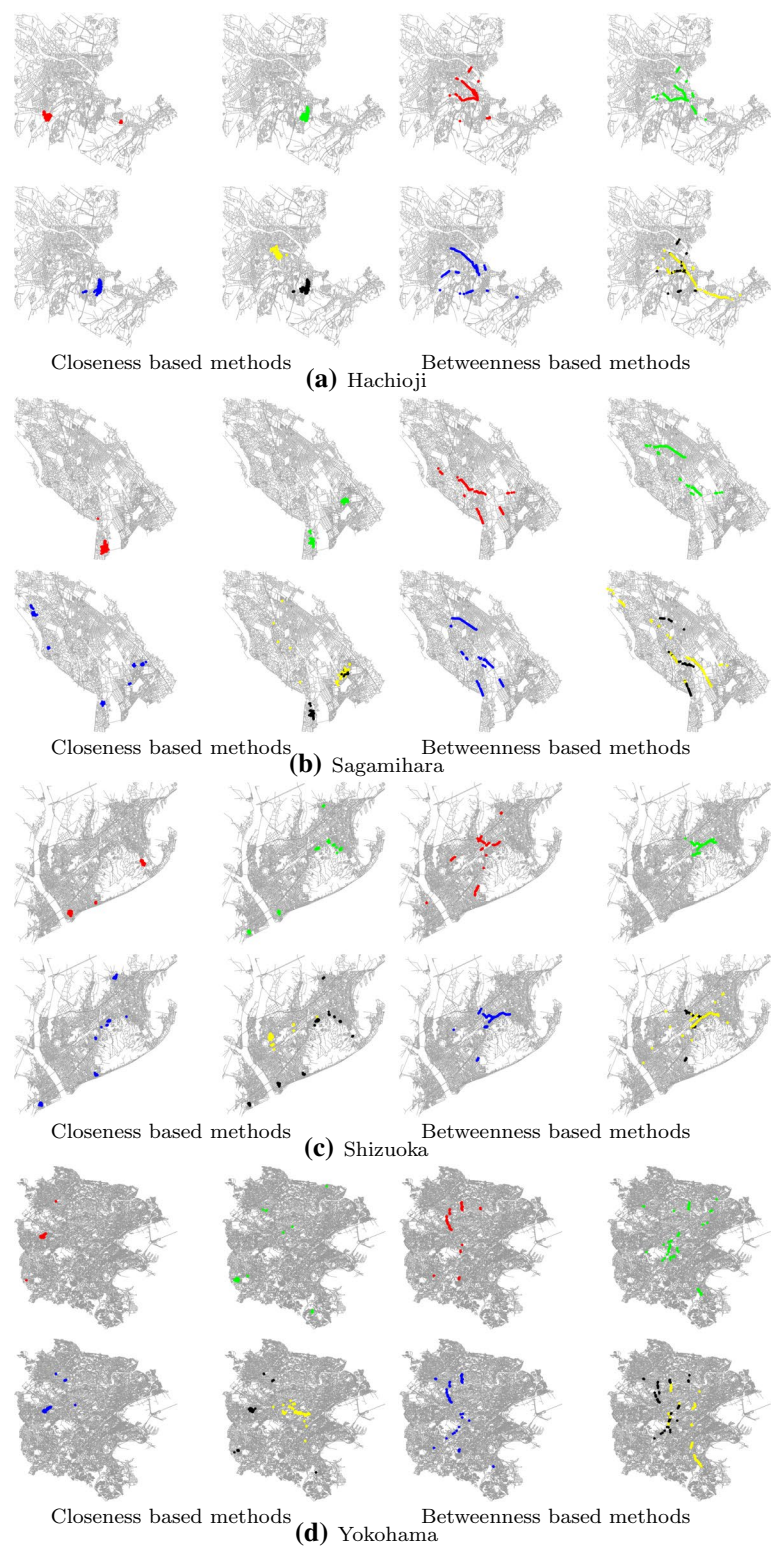
Fig. 9 Extracted top 100 nodes: The red, green, and blue dots represent top nodes by proposed methods, MPCLC and MPBWC, for g1, g2, and g3, respectively. The black dots represent top nodes by compared methods, SPCLC and SPBWC. The yellow dots represent top nodes by methods, CLC and BWC

2 By multiple perspective methods, for some groups, similar nodes in terms of locations are extracted.

3 Many black nodes are located at in similar places with red, green and blue nodes.

4 The top nodes by closeness based methods are located like areas, on the other hand, by betweenness based methods are like lines.

From these observations, we can give the following inferences. Yellow nodes by CLC tend to gather in the center of the city, and yellow nodes by BWC tend to gather along highways. Although many stores have already opened in these areas or along these roads, the classical methods, CLC and BWC do not consider it. In the results of closeness based methods, two of three groups such as g1 (red) and g2 (green) or g2 (green) and g3 (blue) tend to be located near each other. Similarly, in the results of betweenness based methods, two of three groups tend to be located along similar roads. It can be interpreted that such groups expect to obtain more customers from the other group. In fact, in such areas, the stores of one out of three groups dominantly exist as exclusive territories (See Fig. 2). Black nodes by SPCLC and SPBWC are located in areas similar to any of the g1 (red), g2 (green), and g3 (blue) nodes. Therefore, the SPCLC and the SPBWC methods are somewhat similar to the MPCLC and the MPBWC methods, respectively. Figure 11 shows the results of a quantitative examination of this point.

To reveal the detail of similarity, we calculate F-value between sets of top nodes as follows:

$$F(\mathcal{A}(r), \mathcal{B}(r)) = \frac{2|\mathcal{A}(r) \cap \mathcal{B}(r)|}{|\mathcal{A}(r)| + |\mathcal{B}(r)|} = \frac{|\mathcal{A}(r) \cap \mathcal{B}(r)|}{r},$$

where $\mathcal{A}(r)$ is a set of r nodes ranked by the method, A. The higher the value, the more similar the upper node sets by the two methods are. Figure 10 shows the F-value between each of the closeness based methods and the corresponding method of the betweenness based ones, where the horizontal axis represents the rank r on a logarithmic scale. From Fig. 10, in all the cities, the closeness based and the betweenness based methods are not correlated can be confirmed. Especially for the top 100 nodes, there is almost no overlap between them. This is probably because the nodes extracted by our MPCLC and MPBWC are very different, just as the classical CLC and BWC are based on different behavior models. Figure 11 depicts the F-value between the SPCLC or the SPBWC methods and other methods. From Fig. 11, we can see the following observations.

1 The F-value between the SPCLC and the CLC methods, drawn with a yellow line, indicates a small value.

2 The F-value between the SPBWC and the BWC methods also indicates a small value, but not as small as that of closeness based method.

3 The F-values between the SPCLC and the MPCLC methods, drawn with red, green and blue, differ in each city. For example, in Hachioji, for g2 (green) and g3 (blue), the F-values are relatively high compared to g1 (red), on the other hand, in Sagamihara, for g2 (green), the F-value is higher than for g1 (red) and g3 (blue).

4 The F-values for the betweenness based methods also have similar tendencies to those for the closeness based methods.

**Fig. 10** Ranking-similarity between closeness based methods and betweenness based methods
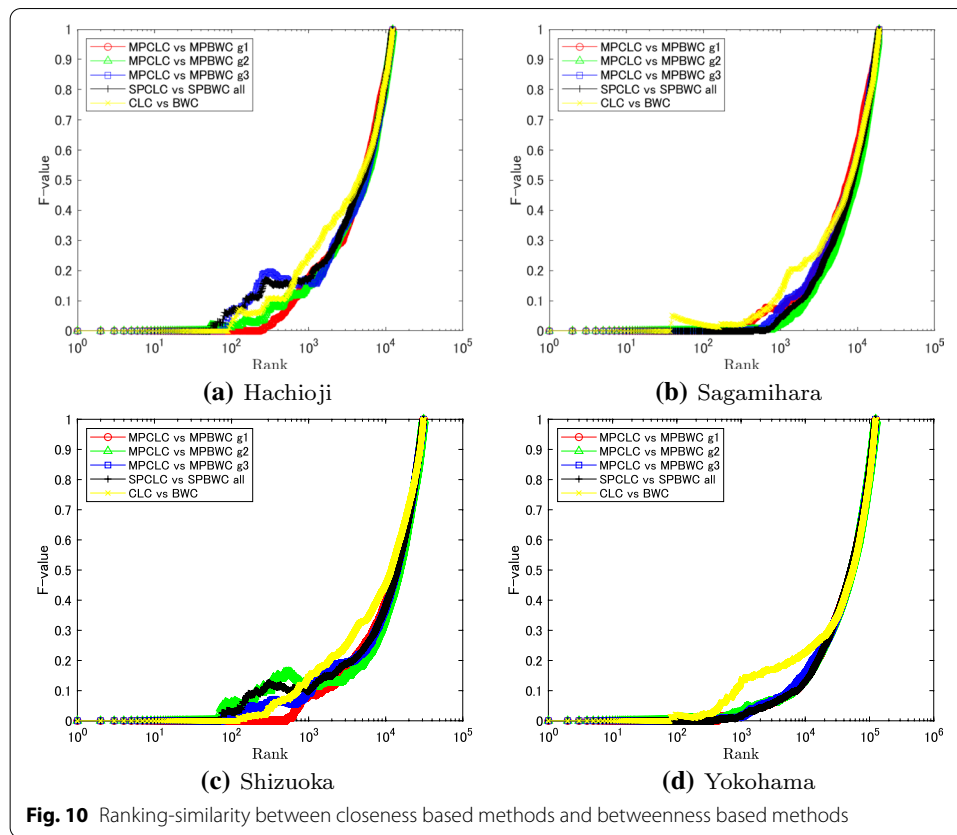
As mentioned above, the classical centrality measures shown in yellow do not consider the position of the existing stores at all, so the F-value is low. On the other hand, the multiple perspective measures indicate high F-value, for some, but not all, groups depending on the city. It means that single perspective measures may extract important nodes that maximize the expected number of customers for some, but not all groups. Therefore it is necessary to extract important nodes from each group perspective by switching the viewpoints.

From these results, we can confirm that in our proposed methods, MPCLC and MPBWC, from different viewpoints such as groups g1, g2, and g3, different nodes are extracted as important nodes, and our methods can be regarded as different centrality measures from classical ones, CLC and BWC, also different from single perspective centrality, SPCLC and SPBWC.

Conclusion

In this study, we proposed centrality measures to find the best location for the store that plans to open from the perspective of each group under a competitive environment. Our measures quantify the number of customers expected to obtain from competing groups and extract the node that maximizes the number of customers of own group. From our experimental evaluations using actual urban street network data, we confirmed that the proposed methods, MPCLC and MPBWC can extract a better

Fig. 11 Ranking-similarity to the single perspective centrality
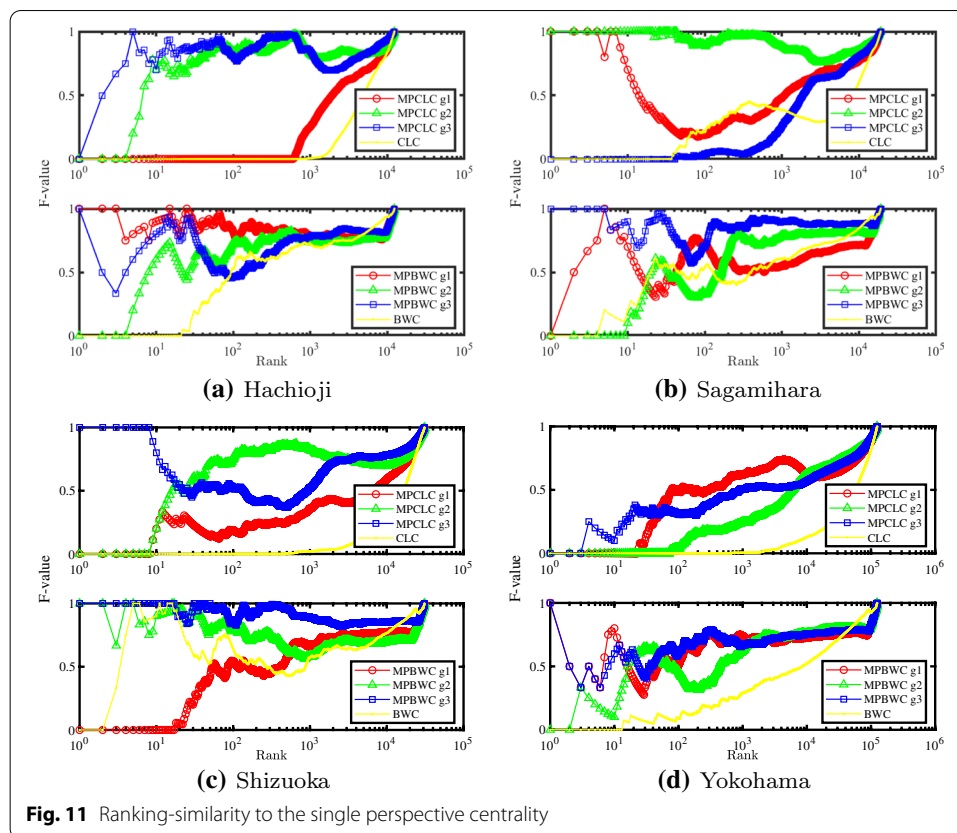
location than the compared methods, SPCLC, SPBWC, CLC, and BWC, in terms of the breakdown of the number of acquired customers and predictive accuracy for newly opening stores.

Our current models are based on the proximity approach, where residents select one nearest or one most-frequently-passing facility. In the future, it will be necessary to generalize the model to stochastically select facilities based on a gravity model or a cover model. Furthermore, we plan to introduce the dynamics where some stores can go out of business if too many stores exist in the near place, by taking game-theory-based approaches.

Author details
[1] School of Computer Science, Tokyo University of Technology, 1404-1 Katakuramachi, Hachioji City, Tokyo 192-0982, Japan. [2] School of Management and Information, University of Shizuoka, 52-1 Yada, Shizuoka City, Shizuoka 422-8526, Japan. [3] Faculty of Science, Kanagawa University, 2946 Tsuchiya, Hiratsuka City, Kanagawa 259-1293, Japan. [4] Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan.

References
Aboolian R, Berman O, Krass D (2007) Competitive facility location and design problem. Eur J Oper Res 182:40–62
Aboolian R, Berman O, Krass D (2020) Optimizing facility location and design. Eur J Oper Res. https://doi.org/10.1016/j.ejor.2020.06.044
Agra A, Cerdeira JO, Requejo C (2017) A decomposition approach for the p-median problem on disconnected graphs. Comput Oper Res 86:79–85
Cherifi H, Palla G, Szymanski B, Lu X (2019) On community structure in complex networks: challenges and opportunities. Appl Netw Sci. https://doi.org/10.1007/s41109-019-0238-9
Crucitti P, Latora V, Porta S (2006) Centrality measures in spatial networks of urban streets. Phys Rev E 73(3):036125+
Drezner T (1994) Locating a single new facility among existing unequally attractive facilities. J Reg Sci 34(2):237–252
Drezner T, Drezner Z (1996) Competitive facilities: market share and location with random utility. J Reg Sci 36(1):1–15
Drezner Z, Wesolowsky GO, Drezner T (1998) On the logit approach to competitive facility location. J Reg Sci 38(2):313–327
Drezner T, Drezner Z, Kalczynski P (2011) A cover-based competitive location model. J Oper Res Soc 62(1):100–113
Everett MG, Borgatti SP (1999) The centrality of groups and classes. J Math Sociol 23(3):181–201
Fushimi T, Saito K, Ikeda T, Kazama K (2019a) Estimating node connectedness in spatial network under stochastic link disconnection based on efficient sampling. Appl Netw Sci 4(66):1–24
Fushimi T, Okubo S, Saito K (2019b) Facility location problem on network based on group centrality measure considering cooperation and competition. In: Proceedings of the 8th international conference on complex networks and their applications, pp 64–76
Fushimi T, Yazaki M (2020) Comparative analysis of store opening strategy based on movement behavior model over urban street networks. In: Proceedings of the 11th international conference on complex networks (CompleNet2020), pp 245–256
Ghalmane Z, Hassouni ME, Cherifi C, Cherifi H (2019a) Centrality in modular networks. EPJ Data Sci. https://doi.org/10.1140/epjds/s13688-019-0195-7
Ghalmane Z, Hassouni ME, Cherifi C, Hassouni ME (2019b) Centrality in Complex networks with overlapping community structure. Sci Rep. https://doi.org/10.1038/s41598-019-46507-y
Gimadi EK (2017) On exact solvability of the restricted capacitated facility location problem. In: Proceedings of the OPTIMA-2017 conference, pp 209–216
Hotelling H (1929) Stability in competition. Econ J 39(153):41–57
Huff DL (1964) Defining and estimating a trading area. J Mark 28(3):34–38
Jinmei W, Kejia Z (2010) Study of facility location and allocation problem based on fuzzy graph theory. In: 2010 international conference on management and service science, pp 1–5
Montis DA, Barthelemy M, Chessa A, Vespignani A (2007) The structure of interurban traffic: a weighted network analysis. Environ Plan B Plan Des 34(5):905–924
Park K, Yilmaz A (2010) A social network analysis approach to analyze road networks. In: Proceedings of the ASPRS annual conference 2010
Puerto J, Ricca F, Scozzari A (2018) Extensive facility location problems on networks: an updated review. TOP 26(2):187–226. https://doi.org/10.1007/s11750-018-0476-5
Rahmaniani R, Ghaderi A (2013) A combined facility location and network design problem with multi-type of capacitated links. Appl Math Model 37(9):6400–6414
Tabata K, Nakamura A, Kudo M (2017) An efficient approximate algorithm for the 1-median problem on a graph. IEICE Trans Inf Syst E100.D(5):994–1002. https://doi.org/10.1587/transinf.2016EDP7398
Tamir A (2001) The k-centrum multi-facility location problem. Discrete Appl Math 109:293–307
Thorup M (2004) Quick k-median, k-center, and facility location for sparse graphs. SIAM J Comput 34(2):405–432

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.