Closeness Centrality for Networks with Overlapping Community Structure

Mateusz K. Tarkowski$^1$ and Piotr Szczepański$^2$

Talal Rahwan$^3$ and Tomasz P. Michalak$^{1,4}$ and Michael Wooldridge$^1$

$^1$Department of Computer Science, University of Oxford, United Kingdom
$^2$Warsaw University of Technology, Poland
$^3$Masdar Institute of Science and Technology, United Arab Emirates
$^4$Institute of Informatics, University of Warsaw, Poland

Abstract

Certain real-life networks have a community structure in which communities overlap. For example, a typical bus network includes bus stops (nodes), which belong to one or more bus lines (communities) that often overlap. Clearly, it is important to take this information into account when measuring the centrality of a bus stop—how important it is to the functioning of the network. For example, if a certain stop becomes inaccessible, the impact will depend in part on the bus lines that visit it. However, existing centrality measures do not take such information into account. Our aim is to bridge this gap. We begin by developing a new game-theoretic solution concept, which we call the Configuration semivalue, in order to have greater flexibility in modelling the community structure compared to previous solution concepts from cooperative game theory. We then use the new concept as a building block to construct the first extension of Closeness centrality to networks with community structure (overlapping or otherwise). Despite the computational complexity inherited from the Configuration semivalue, we show that the corresponding extension of Closeness centrality can be computed in polynomial time. We empirically evaluate this measure and our algorithm that computes it by analysing the Warsaw public transportation network.

Introduction

One of the key problems in network science involves identifying the most important (or central) nodes (Freeman 1979; Dezső and Barabási 2002; Keinan et al. 2004; Page et al. 1999). The four best-known centrality measures are Degree, Betweenness, Closeness and Eigenvector centralities (Bonacich 1972; Freeman 1979), each of which views centrality from a different perspective, focusing on certain traits that make nodes important, or, “central,” to the functioning of a network (Brandes and Erlebach 2005; Koschutski et al. 2005). Our focus in this paper is on Closeness centrality. This measure considers the important nodes to be those that are relatively close to all other nodes in the network: the closer a node is to the others, the higher its centrality. Closeness centrality has many applications, from coauthorship networks (Yan and Ding 2009), through tourism (Shih 2006), to social networks (Barabási 2003; Karinthy 2006).

One aspect of networks that has been largely ignored in the literature on centrality is the fact that certain real-life networks have a predefined community structure. In public transportation networks, for example, bus stops are typically grouped by the bus lines (or routes) that visit them. In coauthorship networks, the various venues where authors publish can be interpreted as communities (Szczepański, Michalak, and Wooldridge 2014). In social networks, individuals grouped by similar interests can be thought of as members of a community. Clearly for such networks, it is desirable to have a centrality measure that accounts for the predefined community structure. Yet, to the best of our knowledge, only one such measure has been developed to date (Szczepański, Michalak, and Wooldridge 2014), which extends Degree centrality to networks with community structure. Despite this recent development, one important aspect of real-life networks remains missing from existing centrality measures: the ability to consider overlapping communities. Take social networks, for example, where such overlaps are widespread due to the various affiliations and interests of the individuals involved (Kelley et al. 2012). Likewise, in our example of transportation networks, a bus stop may be on the route of multiple (i.e., overlapping) bus lines. If such a stop becomes inaccessible, then all the bus lines that visit it would no longer function properly. As such, the importance of a bus stop clearly depends (at least partially) on the importance of the bus lines to which it belongs.

In an attempt to define a centrality measure that accounts for overlapping communities, we focused on game-theoretic centrality measures.$^1$ The inspiration behind this line of research comes from solution concepts in cooperative game theory. In essence, given a set of players, a cooperative solution concept typically defines a payoff for each player by comparing his or her contribution to the various groups of players (more on this in the next section). The rich repository of solution concepts has been extensively refined and expanded over the past decades, making it an ideal toolkit for quantifying the importance of individuals in a setting where those individuals co-exist and operate in groups. In the context of game-theoretic network centrality, the indi-

---

$^1$See www.game-theoretic-centrality.com and www.network-centrality.com for an overview of this line of research.
In the following two subsections, we introduce the relevant game-theoretic concepts, and computed them in polynomial time.

Table 1: The table outlines the papers that used various solution concepts to extend Degree, Closeness, or Betweenness centralities, and computed them in polynomial time.

<table>
<thead>
<tr>
<th>Solution Concept</th>
<th>Degree</th>
<th>Closeness</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapley value</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Semivalues</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Owen value</td>
<td>5</td>
<td>This Paper</td>
<td>Open</td>
</tr>
<tr>
<td>Coalesional semivalues</td>
<td>5</td>
<td>This Paper</td>
<td>Open</td>
</tr>
<tr>
<td>Configuration value</td>
<td>Open</td>
<td>This Paper</td>
<td>Open</td>
</tr>
<tr>
<td>Configuration semivalues</td>
<td>Open</td>
<td>This Paper</td>
<td>Open</td>
</tr>
</tbody>
</table>

Game-Theoretic Concepts

A cooperative game consists of a set of players $N = \{1, 2, \ldots, n\}$ and a characteristic function $\nu : 2^N \to \mathbb{R}$ such that $\nu(\emptyset) = 0$. This function assigns to each coalition of players its payoff (i.e., an indication of its performance). We will henceforth refer to a game simply by $\nu$. A coalition structure, $CS = \{Q_1, Q_2, \ldots, Q_m\}$, is a partition of $N$ into disjoint coalitions. One of the key questions in cooperative game theory is the following: Given a game $\nu$ and a coalition structure $CS$ that the players have formed, how do we divide the payoff of each coalition among its members? In this context, assuming that $CS = \{N\}$, i.e., assuming that the players have formed the grand coalition, Shapley (1953) proposed a solution concept—now known as the Shapley value—to fairly divide the payoff from cooperation among the players. Banzhaf (1965) proposed another solution concept—now known as the Banzhaf index— which is similar to the Shapley value except for a subtle difference in the way contributions are weighed. To generalize the aforementioned solution concepts, Weber (1979) proposed semivalues—a family of solution concepts that includes both the Shapley value and Banzhaf index. Formally, let $MC(C, i) = \nu(C \cup \{i\}) - \nu(C)$ be the marginal contribution of player $i$ to coalition $C \subseteq N \setminus \{i\}$. Denoting by $\beta(k)$ the probability that any player makes a marginal contribution to a coalition of size $k$, the semivalue of $i$ is:

$$\psi_i(\nu) = \sum_{k=0}^{\lfloor N\rfloor - 1} \beta(k) \mathbb{E}\left[MC(C^k, i)\right],$$

where $C^k$ is a random variable over subsets of size $k$ chosen from the set $N \setminus \{i\}$ with uniform probability, and $\mathbb{E}$ is the expected value operator for this variable. The Shapley value and Banzhaf index are two semivalues, defined by $\beta(k) = 1/|N|$ and $\beta(k) = \binom{|N| - 1}{k}/2^{|N| - 1}$, respectively.

Importantly, all semivalues assume that $CS = \{N\}$, i.e., that the grand coalition is formed. To relax this assumption, Owen (1977) introduced a solution concept—now known as the Owen value—that divides the payoff of any a priori coalition structure $CS$. Now, when $CS = \{N\}$ or $CS = \{\{i\}\}_{i \in N}$, the Owen value is equivalent to the Shapley value. As such, the Owen value is a generalization of the Shapley value; one that does not generalize the $\beta$ function (as semivalues do), but rather generalizes the assumed coalition structure $CS$. Another step in this line of research was taken by Szczepański, Michalak, and Wooldridge (2014), who proposed a generalization combining both the Owen value and semivalues; they called it coalesional semivalues. Formally, given a coalition structure, $CS$, and discrete probability distributions: $\beta : \{0, \ldots, |CS| - 1\} \to [0, 1]$ and $\alpha_j : \{0, \ldots, |Q_j| - 1\} \to [0, 1]$ for all $j \in \{1, \ldots, |CS|\}$, coalesional semivalues are defined by:

$$\gamma_i(\nu, CS) = \sum_{k=0}^{|CS| - 1} \beta(k) \sum_{l=0}^{|Q_j| - 1} \alpha_j(l) \mathbb{E}\left[MC\left((\bigcup\limits^{T^k} \cup C^i, i)\right)\right]$$

where $Q_j$ is the coalition in $CS$ that player $i$ belongs to, $T^k$ is a random variable over subsets of size $k$ chosen from
CS \ {Q_j} \) with uniform probability. \( C^l \) is a random variable over subsets of size \( l \) chosen from \( Q_j \ \{ i \} \) with uniform probability, and \( \mathbb{E} \) is the expected value operator. The coaxial semivalue is equivalent to the Owen value when \( \beta(k) = 1/|CS| \) and \( \alpha_j(l) = 1/|Q_j|, \forall j \in \{ 1, \ldots, |CS| \} \).

None of the solution concepts discussed thus far considers overlapping coalitions. To address this issue, Albizuri, Aurrecoechea, and Zarzuelo (2006) generalised the Owen value to situations where the \textit{a priori} coalition structure \( CS \) contains overlapping coalitions; they called this generalisation the \textit{Configuration value}. Formally, it is defined as follows, where \( T_i = \{ j : j \in \mathbb{N} \text{ and } Q_j \in CS \text{ and } i \in Q_j \} \):

\[
\chi_i(\nu, CS) = \sum_{T_i \subseteq CS} \sum_{j \in T_i} \sum_{C \subseteq Q_j} \lambda MC \left( \left( \bigcup_{t \in T_i} C, i \right) \right),
\]

where \( \lambda = \frac{|T|!|\{CS| - |T| - 1|!|C|!(|Q_j| - |C| - 1|)!}{|Q_j|!} \). \( (3) \)

**Graph-Theoretic Concepts**

A network is a graph, \( G = (V, E) \), comprised of a set of nodes \( V = \{v_0, v_1, \ldots, v_{n-1}\} \) and a set of edges \( E \subseteq V \times V \). A path is simply a chain of connected nodes. The distance between two nodes, \( s \) and \( t \), denoted by \( \text{dist}(s, t) \) is the length of the shortest path between the two (we assume that \( \text{dist}(v, v) = 0 \)). Given a node \( v \in V \) and a set of nodes \( C \subseteq V \), we say that \( \text{dist}(C, v) \) is equal to the minimum distance between any node \( u \in C \) and \( v \) (this implies that if \( v \in C \) then \( \text{dist}(C, v) = 0 \)).

**Closeness centrality** quantifies the importance of nodes based on their average distance to other nodes (Freeman 1979). In its most general form, it is formulated as follows:

\[
\text{closeness}(v) = \sum_{u \in V} f(\text{dist}(v, u)),
\]

where the function \( f : \mathbb{N} \rightarrow \mathbb{R} \) determines how the distance influences the centrality. When \( f(k) = k \), we obtain the classical Closeness centrality, where the \textit{smaller} the value, the more central the node. In this paper, we focus on an alternative formulation, where \( f(k) = 1/k \) and throughout the paper \( \frac{1}{k} = 0 \). The resulting centrality is known as \textit{harmonic} centrality (Boldi and Vigna 2013). With this modification, the \textit{greater} the value, the more central the node (which is in line with most centrality measures).

Everett and Borgatti (1999) introduced \textit{group Closeness centrality}, which extends the notion of Closeness to groups of nodes as follows:

\[
\nu_C(C) = \sum_{u \in (V \backslash C)} f(\text{dist}(C, u)). \quad (4)
\]

Building upon this formula, the first game-theoretic extension of Closeness centrality was introduced by (Michalak et al. 2013b). In particular, the authors defined a game in which the players are the nodes of the network, and the characteristic function is \( \nu_C \). The centrality of each node was then determined using the Shapley value. The resulting game-theoretic centrality measure is called the \textit{Shapley value-based Closeness centrality}. Roughly speaking, the harmonic Closeness centrality evaluates how close a node is to others, whereas the Shapley value-based variant evaluates the role that a node plays in \textit{bringing other nodes closer together}. To illustrate this difference, consider networks (a) and (b) from Figure 1. Here, according to harmonic Closeness, \( v_1 \) is relatively more important in (b) than in (a) since it is close to more nodes in (b) than in (a). In contrast, according to the Shapley-value based Closeness, \( v_1 \) is actually more important in (a), since the removal of \( v_1 \) from (a) has a greater impact on the distances between the other nodes, compared to the removal of \( v_1 \) from (b).

We present in Table 2 the harmonic and Shapley value closeness rankings for Figure 1 (d).

The Configuration value Closeness ranking is as follows: \( v_9, v_7, v_1, v_6, v_5, v_8, v_10, v_3, v_2 \). The configuration value makes use of community information, promoting nodes \( v_9, v_1, v_6 \) and \( v_3 \). This ranking is also more fine-grained (i.e., there are no ties), because it draws upon community information, which is different for most nodes in the example.

**Our Centrality Measures**

As stated earlier, no centrality measures to date can readily be applied to networks with overlapping community structures. An example is depicted in networks (c) and (d) from Figure 1. Specifically, in network (c), nodes \( v_5 \) and \( v_4 \) are symmetric except that \( v_5 \) belongs to a seemingly-important community; one that connects the two parts of the network. Arguably, when taking this additional information into consideration, \( v_3 \) should be considered more important than \( v_4 \). Moving on to network (d), node \( v_1 \) belongs to more communities than \( v_2 \), and the communities of \( v_1 \) seem to be equally important, if not more important, than those of \( v_2 \). This could mean, for example, that more bus, tram or train routes visit (and rely on) the bus stop \( v_1 \) than the bus stop \( v_2 \), implying that \( v_1 \) should be more central once the underlying community structure is taken into consideration.

**Configuration Semivalues:** We now propose a family of solution concepts, which we believe to be the most general

![Sample networks](image)

**Figure 1:** Sample networks. In (c) and (d), communities are highlighted by same-coloured edges.
of its kind to date, as it not only allows for an arbitrary $\beta$, $\alpha$ and $CS$, but also allows for overlapping coalitions. We call it Configuration semivalues, and define it as follows, where $T^k$ is a random variable over subsets of size $k$ of $CS \setminus T_i$ and $E$ is the expected value operator:

$$\phi_\nu(\nu, CS) = \beta(k) \sum_{j \in T_i} \sum_{l=0}^{Q_j - 1} \alpha_j(l) E \left[ MC \left( \bigcup_{T^k} \cup C^l \right) \right]$$

(5)

In particular, given a community structure $CS$ in which communities do not overlap, this family of solution concepts is equivalent to coalitional semivalues. Given $CS = \{N\}$, it is equivalent to semivalues. Further restrictions on $\beta$ and $\alpha_j$ (as discussed in the preliminaries) lead to the Shapley value, the Banzhaf index, or the Owen value. Compared to the configuration value, our configuration semivalues offer greater control over the contributions of players, due to a probability distribution over the number of communities to which a player contributes and a distribution over the number of nodes from his own community that he can contribute to. In applications such as counterterrorism (Lindelauf, Hamers, and Husslage 2013; Michalak et al. 2013a), this can represent the expected size of an attack on a network or the number of targeted communities.

**Configuration Semivalue Community Index**: Whereas to measure the importance of a community using the Owen value it suffices to sum up the power of the nodes comprising it, this is not the case for the Configuration value. In particular, the power of a node may be the result of its membership to many communities. For this reason, the distinction must be made as to which of the player’s marginal contributions are made because of which community. To this end, we propose the following measure of community strength:

$$CP_\nu(\nu, CS) = \sum_{j \in Q_j} \sum_{k=0}^{CS-1} \beta(k) \sum_{l=0}^{Q_j - 1} \alpha_j(l) E \left[ MC \left( \bigcup_{T^k} \cup C^l \right) \right],$$

where $C^l \subseteq Q_j \setminus \{i\}$ and $T^k \subseteq CS \setminus Q_j$ are random variables. Although an axiomatic characterisation of this community index is outside the scope of this paper, we mention that in the case of the Configuration value, the index is efficient, and in the case of the Owen value, it is equal to the sum of the Owen values of the members of a community.

**Configuration Semivalue Closeness Centrality**: Our extension of closeness centrality (which accounts for overlapping and non-overlapping community structures) involves using our Configuration semivalue (see Equation 5) with the characteristic function for group Closeness centrality (see Equation 4). More formally, it is: $\phi_\nu(\nu, CS)$.

**Proof of Theorem 1**: Starting from Equation (5), which defines the configuration semivalue, let us first replace the arbitrary characteristic function therein, i.e., $\nu$, with that of group Closeness centrality (defined in Equation 4). We get:

$$\phi_\nu(\nu, CS) = \sum_{j \in T_i} \sum_{k=0}^{CS-1} \frac{|Q_j|}{|CS|} \sum_{l=0}^{Q_j - 1} \alpha_j(l) \beta(k)$$

where $\beta(k) = f(\text{dist}(\bigcup_{T^k} \cup C^l \cup \{v\}), u) - f(\text{dist}(\bigcup_{T^k} \cup C^l, u)).$

Although this may seem inconsistent, our next step is to rearrange the summation over $u$ and bring it to the forefront:

$$\phi_\nu(\nu, CS) = \sum_{u \in V} \sum_{j \in T_i} \sum_{k=0}^{CS-1} \frac{|Q_j|}{|CS|} \sum_{l=0}^{Q_j - 1} \alpha_j(l) \beta(k)$$

Next, for $v \in Q_j$ we will split the equation into:

$$MC_j^+(v, u, j) = \sum_{k=0}^{CS-1} \frac{|Q_j|}{|CS|} \sum_{l=0}^{Q_j - 1} \alpha_j(l) \beta(k)$$

$$f(\text{dist}(\bigcup_{T^k} \cup C^l \cup \{v\}, u))$$

and

$$MC_j^-(v, u, j) = \sum_{k=0}^{CS-1} \frac{|Q_j|}{|CS|} \sum_{l=0}^{Q_j - 1} \alpha_j(l) \beta(k)$$

$$f(\text{dist}(\bigcup_{T^k} \cup C^l, u))$$

with the additional constraint on $T^k$ and $C^l$ such that:

$$\text{dist}(\bigcup_{T^k} \cup C^l, u) \neq \text{dist}(\bigcup_{T^k} \cup C^l \cup \{v\}, u).$$

We can now state the following:

$$\phi_\nu(\nu, CS) = \sum_{u \in V} \sum_{j \in T_i} \sum_{k=0}^{CS-1} \frac{|Q_j|}{|CS|} \sum_{l=0}^{Q_j - 1} \alpha_j(l) \beta(k)$$

$$f(\text{dist}(\bigcup_{T^k} \cup C^l \cup \{v\}, u)) - f(\text{dist}(\bigcup_{T^k} \cup C^l, u))$$

The constraint in Equation (8) simply allows us to avoid redundant computations (the contribution in the opposite case is trivially zero, since by entering such a coalition, $v$ does not change the distance to $u$). The remainder of the proof will focus on computing Equations (6) and (7). We will first focus on Equation (6).

**Algorithms**

We show that any configuration semivalue of group Closeness centrality for all nodes in a weighted graph $G = (V, E, \omega)$ can be computed in $O(|V|^4|CS|)$ time.

**Theorem 1** Any configuration semivalue of group Closeness centrality for all nodes in a weighted graph $G = (V, E, \omega)$ can be computed in $O(|V|^4|CS|)$ time.
• $Com_{>d}(u) = \{ Q : Q \in M$ and $dist(Q,u) \sim d \}$;
• $Nod_{>d}(u) = \{ s : s \in C_j$ and $dist(s,u) \sim d \}$,
where $\sim$ will be one of $<,\geq,\leq,\geq$ or $\equiv$. These sets are fairly simple to precompute, and will allow us to count the number of required coalitions. In this particular case, we will use $Com_{>dist(u,v)}$ to count the number of communities farther from $u$ than $v$. Similarly, $Nod_{>dist(u,v)}$ counts the number of nodes in the community $Q_j$ that are farther from $u$ than the distance from $u$ to $v$. Altogether, there are $(\binom{Com_{>dist(u,v)}}{k})$ coalitions of communities and $(\binom{Nod_{>dist(u,v)}}{l})$ coalitions from $Q_j$ satisfying the requirements. Let $d = dist(u,v)$. Finally:

$$MC^{+}_{k,l}(v,u,j) = \binom{\binom{Com_{>d}(u)}{k}}{l} \binom{\binom{Com_{>d}(u)}{l}}{k} f(dist(v,u)).$$

As for Equation (7), we divide computations as follows:

$$MC^{−}_{k,l}(v,u,j) = \sum_{d} MC^{−}_{k,l}(v,u,j,d),$$

where

$$MC^{−}_{k,l}(v,u,j,d) = \sum_{T^k \subseteq CS \setminus v} \sum_{C^l \subseteq \{Q_j \setminus \{v\} \cup \{u\}} \alpha_j(l) \beta(k)$$

$$\frac{f(dist(\bigcup_{T^k \cup C^l} u)))}{(\binom{|CS|−1}{k} \binom{|Q_j|−1}{l})},$$

keeping in mind the constraint from Equation (8) and adding the constraint on $C^l$ and $T^k$ such that

$$dist(\bigcup_{\bigcup T^k \cup C^l} u) = d. \quad (10)$$

Now, the computation of $MC^{−}_{k,l}(v,u,j,d)$ reduces to computing the number of coalitions $C^l$ and coalitions of communities $T^k$ that satisfy both constraints. By using the inclusion-exclusion principle, and assuming that $dist(v,u) > d$, we have the following:

$$MC^{−}_{k,d}(v,u,j,d) = \sum_{d} MC^{−}_{k,d}(v,u,j,d),$$

where

$$MC^{−}_{k,d}(v,u,j,d) = \binom{Com_{>d}(u)}{k} \binom{Nod_{>d}(u)}{l} - \binom{Com_{>d}(u)}{k} \binom{Nod_{>d}(u)}{l}.$$

To explain this, we first compute the number of coalitions of communities of size $k$ that are at distance $d$ or farther from $u$. We do the same for coalitions of nodes from $Q_j$. However, we must take away the number of occurrences when no nodes from $C^l$ and communities from $T^k$ are at distance $d$ from $u$, in order to satisfy the constraint from Equation (10).

Computing all $MC^{−}_{k,l}(v,u,j)$ variables is the most time-consuming, as it takes $O(\binom{|V|}{d} |CS|)$ time. This may be counter-intuitive, since computing all $MC^{−}_{k,l}(v,u,j,d)$ variables would imply $O(\binom{|V|}{d} |CS|)$, but dynamic programming eliminates this need. Precomputations are implemented in Algorithm 1 and marginal contributions in Algorithm 2.

**Theorem 2** The configuration value of group Closeness centrality for all nodes in a weighted graph $G = (V,E,\omega)$ can be computed in $O(|V|^2 \log(|V|) + |CS|) + |V||E|$ time.

**Sketch of Proof of Theorem 2:** The key idea, is to use an alternate formula for the configuration value:

$$\phi_c(v, CS) = \sum_{j \in T} \sum_{\Pi \in \Pi(Q_j)} \sum_{\Pi \in \Pi(Q_j)} f(dist(\Pi \cup \{v\}, u)) - f(dist(\Pi \cup \{v\}, u)),$$

where $\Pi(X)$ refers to a permutation of the set $X$, and $\Pi_J$ is the set of elements preceding $v$ in the permutation $J$. Next, we group computations for both $k$ and $l$ as follows: $MC^+(v,u,j) = \sum_{k,l} MC^+_{k,l}(v,u,j)$ and $MC^−(v,u,j,d) = \sum_{k,l} MC^−_{k,l}(v,u,j,d)$. By counting the permutations that satisfy the constraints from Theorem 1, we reach the following:

$$MC^+(v,u,j) = \frac{f(d)(|CS|)!(|Q_j|)!}{\binom{MC^+(v,u,j,d)}{d}}$$

$$MC^−(v,u,j,d) = \frac{MC^−(v,u,j,d)}{\binom{MC^−(v,u,j,d)}{d}}.$$

Precomputations are presented in Algorithm 1, and computation of marginal contributions is presented in Algorithm 2. Algorithm 3 computes the Configuration value with better time complexity. In Algorithm 1, lines 1 to 20 compute node distances using Johnson’s algorithm (Johnson 1977) and sort them in a descending order, allowing us to use dynamic programming to avoid redundant computation. As an example, if $d$ directly follows $d$ in such a list, then $COM_{>d}(u) = COM_{>d}(u) + COM_{>d}(u)$. The last step removes duplicate distances from the community distances list, which also helps avoid redundant computation.

Algorithm 2 computes marginal contributions by moving backwards (largest to smallest) through the possible distances, and also uses dynamic programming. The data computed thus-far is held in the variable prev.val and accumulated in $\phi_c$. s_merge(a,b) is a function that takes two sorted sets (descending order) and returns a sorted set, such that any items in $b$ that are smaller than the smallest item in $a$ are not included. Finally, the array CP must be mentioned, which collects the power of communities as a whole. If a marginal contribution is made by a node through community $Q_j$, then $CP_j$ is incremented by the value of the contribution.

We conclude by noting that the configuration value is a generalisation of the Shapley value. Algorithm 3 can compute the Shapley value-based Closeness centrality (Michalak 2004).
Warsaw Public Transportation Network

Algorithms 1, 2 and 3 were implemented in Java. The Configuration-based Closeness centrality (i.e., CV-based Closeness centrality) was used to analyse the Warsaw public transportation network.7 Edge-weights were defined as the average travel times between nodes. In total, the weighted network consists of 1425 nodes, 2135 edges and 380 communities formed by bus lines, trams, and underground and suburbia trains. We note that the ranking of nodes according to CV-based Closeness differs significantly from the Harmonic Closeness and Shapley value-based (i.e., SV-based) Closeness rankings. We present the four most prominent nodes according to these centralities in Table 3.

Configuration-based Closeness ranks Centrum (or “city center”) as most important. This result is intuitive and makes sense, since this stop is a large hub in downtown Warsaw, where passengers can choose from 68 bus, tram and train connections, and one metro line (see Table 4). The top ranked nodes according to the other centrality measures are also near the city center, but provide less connections that are not as important as those in the city center.

The most surprising rankings are for PKP Falenica and PKP Radość, which are railway stations far from downtown Warsaw. They are important because—for certain source-destination pairs—it is almost impossible to find routes that omit these stations. Additionally, trains play an important role in shortening the travel time between distant nodes, since they are the fastest means of transportation.

Interestingly, the fifth node according to CV-based Closeness, Płowiecka, is not in the top ten according to the other measures. Harmonic Closeness misses the fact that this stop is not easily replaced. Shapley value-based Closeness misses

---

7The data, experiment results and programs can be downloaded from https://github.com/szczep/gtna.
that the bus lines that stop at Płowiecka are very important. Even if the stop is omitted, a traveller must often still use bus lines that visit Płowiecka for further travel. As for communities—unsurprisingly—long routes that bring in commuters from all of Warsaw (including the metro line M1) are ranked as most important.

Harmonic Closeness centrality prioritises the topologically most central nodes. The SV-based centrality considers a new dimension, promoting irreplaceable nodes. We go one step further and provide a new CV-based measure that promotes hubs with powerful connections, while taking into account the previous two considerations. Importantly, only our measure is able to detect the most central and popular hub in Warsaw.

### Information Diffusion in Social Networks

Borgatti (2006) advocated the use of group closeness centrality for information diffusion. However, this approach does not yield a comprehensive ranking of nodes, is computationally intractable and does not account for communities. Recently, Lin et al. (2015) noted that communities are important for information diffusion, since intra-community diffusion is much faster than inter-community diffusion.

#### Algorithm 2: Efficient Algorithm for Configuration Semivalue Closeness (continued from Algorithm 1).

```plaintext
for u ∈ V do
    for Qj ∈ CS, k ∈ [0, |CS|], l ∈ [0, |Qj|) do
        prev_d ← −1; prev_val ← 0; MCl ← 0;
        for (x, d) ∈ s_merge(dists(j)[u], c_dists[u]) do
            if x ∈ V and prev_d ≠ −1 then
                Com≠k, l+1(u) ← Com>prev_d(u);
                Com>prev_d(u) ← Com≠k, l+1(u);
                if prev_d ≠ d then
                    MCl ← prev_val; prev_d ← d;
                    prev_val ← prev_val + (Com≠k, l+1(u)) (Nod≠k, l+1(u)) (f(d))
                end
                if x ∈ V then
                    MC+ ← f(d) (Com≠k, l+1(u)) (Nod≠k, l+1(u));
                    φl ← φl + β(k)αj(l) (MCk+MC−)/(k−1) (Nod≠k, l+1(u));
                    CPj ← CPj + β(k)αj(l) (MCk+MC−)/(k−1) (Nod≠k, l+1(u));
                    end
        end
    end
end
```

#### Algorithm 3: Efficient Algorithm for Configuration Value Closeness (continued from Algorithm 1).

```plaintext
for u ∈ V do
    for Qj ∈ CS do
        prev_d ← −1; prev_val ← 0; MCl ← 0;
        for (x, d) ∈ s_merge(dists(j)[u], c_dists[u]) do
            if x ∈ V and prev_d ≠ −1 then
                Com≠k, l+1(u) ← Com>prev_d(u);
                Com>prev_d(u) ← Com≠k, l+1(u);
                if prev_d ≠ d then
                    MCl ← prev_val; prev_d ← d;
                    prev_val ← prev_val + (Nod≠k, l+1(u)) (f(d))
                end
                if x ∈ V then
                    MC+ ← (Nod≠k, l+1(u)) (Com≠k, l+1(u));
                    φl ← φl + MC+ − MC−;
                    CPj ← CPj + MC+ − MC−;
                    end
        end
    end
end
```

We conducted an experiment on a YouTube social network with ground-truth communities (Mislove et al. 2007) in order to see the impact of overlapping communities on centrality. We chose the 80 first communities from a list of the 5000 most important ones and studied the sub-network consisting of these communities. Figure 6 shows the four most important nodes. Harmonic centrality indicates that node 5274 is topologically most central, whereas SV-based centrality indicates that it also brings other nodes closer together. The configuration value ranking promotes node 5, since it belongs to important communities, and it is important for bringing nodes within these communities closer together. Moreover, it brings other communities closer together, which is imperative for inter-community information transfer.

### Conclusions and Future Work

We have developed a general solution concept, namely the Configuration semivalue, that encompasses both coalitional semivalues (Szczepański, Michalak, and Wooldridge 2014) and the configuration value (Albizuri, Aurrecoechea, and Zarzuelo 2006). We have used this value in order to develop the first network centrality measure that accounts for an overlapping community structure. We based our centrality on the notion of Closeness, developed polynomial-time algorithms for its computation and used it to analyse the Warsaw public transportation network. This research also fills a gap in the complexity analysis of game-theoretic centrality measures. An interesting direction for future work is to complete the missing entries in Table 1. Finally, although the configuration value has been axiomatised, configuration semivalues in general and community indices need further study.

---

8 Available at https://snap.stanford.edu/data/com-Youtube.html
Acknowledgements
Tomasz Michalak and Michael Wooldridge were supported by the European Research Council under Advanced Grant 291528 (“RACE”). This work was also supported by the Polish National Science Centre grant DEC-2013/09/D/ST6/03920.

References


