An Extension of the Owen-value interaction index and Its Application to Inter-links Prediction

Piotr L. Szczepański¹ and Tomasz P. Michalak²,³ and Talal Rahwan⁴ and Michael Wooldridge²

Abstract. Link prediction is a key problem in social network analysis: it involves making suggestions about where to add new links in a network, based solely on the structure of the network. We address a special case of this problem, whereby the new links are supposed to connect different communities in the network; we call it the inter-links prediction problem. This is particularly challenging as there are typically very few links between different communities. To solve this problem, we propose a local node-similarity measure, inspired by the Owen-value interaction index—a concept developed in cooperative game theory and fuzzy systems. Although this index requires an exponential number of operations in the general case, we show that our local node-similarity measure is computable in polynomial time. We apply our measure to solve the inter-links prediction problem in a number of real-life networks, and show that it outperforms all other local similarity measures in the literature.

1 INTRODUCTION

Link prediction is one of the key problems in social network analysis [32, 28, 46]. Informally, it involves making recommendations about where to add new links in a network, based solely on the structure of that network. Link prediction has many applications, such as (i) identifying potential customers in online shops [10]; (ii) discovering the interactions between proteins in biological networks [5]; and (iii) finding hidden connections between terrorists [23].

The problem of link prediction is strongly associated with the notion of similarity between nodes in a network [32]: the greater the similarity between two nodes, the greater the likelihood of having a link between them. Broadly speaking, computing the similarity between any two nodes may either involve local or global information about those nodes. Each approach has its relative strengths and weaknesses. In particular, compared to local measures, global ones generally yield better results but are harder to compute, which limits their applicability to small networks (more details can be found in Section 7). In this paper, we restrict our attention to the problem of link prediction based on local information.

Some researchers [44, 51] have suggested exploiting the fact that, in real-life networks, nodes tend to form densely-connected groups, or communities [15], and that nodes from the same community are more likely to be connected.

We address a new problem, whereby we are given a network and a community structure, and want to recommend new links between different communities. We call these “inter-links” (as opposed to intra-links, which connect nodes belonging to the same community). To see why this new problem is significant, consider some applications of the general link-prediction problem:

• One of the most lucrative business applications of link prediction is product recommendation in e-commerce [29], whereby any customer viewing a certain product is presented with a list of similar products. In this context, besides obviously-similar products, it may be worthwhile to also recommend some other products that are different yet relevant. This can be modeled as the problem of recommending inter-links between products belonging to different categories, or “communities”.

• Another promising application of link prediction is to recommend new collaborations in academic networks [53]. While current tools focus on recommending collaborations between members of different communities, e.g., to foster interdisciplinary research and promote the creation of diverse teams.

Our approach to the inter-link prediction problem draws inspiration from the field of cooperative game theory. Concepts from network science may be understood in a cooperative game theoretic setting as follows:

• a node is represented as a player;
• a group centrality [13] is represented as a characteristic function that assigns to each group a real value reflecting its payoff, or power, according to some metric;
• a community (or a subset of nodes) is a coalition (or a subset of players), and the community structure corresponds to a coalition structure.

With this mapping in place, it is possible to measure the similarity between any two nodes using the interaction index [18]—a game-theoretic concept that measures the interaction between two players by analyzing the payoffs of the many possible coalitions in the game. At its core, an interaction index is built around a payoff-division scheme (more on this in Section 2). Among the many schemes that can be used for this purpose, one particularly attractive family of schemes is Semivalues [19]; by using it, we obtain the Semivalue interaction index. This particular index was recently proposed as a local measure of node similarity [46]. Although this measure was shown to be useful for link prediction, it does not take into account the underlying community structure. To overcome this issue, our idea is to use the Owen value [38]—a payoff division scheme inherently designed to handle situations where there is an underlying coalition structure; the resulting node-similarity measure is the Owen-value interaction index. We also propose a family of schemes that generalize

¹ Warsaw University of Technology, Poland
² University of Oxford, UK
³ University of Warsaw, Poland
⁴ Masdar Institute, UAE
the Owen value, namely \textit{Coalitional Semivalues} [48]; the resulting node-similarity measure is the \textit{Coalitional-Semivalue interaction index.} To the best of our knowledge, we are the first to study this latter interaction index.

In summary, the contribution of this paper is as follows:

- We formulate the problem of inter-link prediction in networks with a community structure;
- We introduce an extension of Owen-value interaction index called the \textit{Coalitional-Semivalue interaction index}, and use it to define the new local similarity measure on networks;
- We propose polynomial time algorithms to efficiently compute the Owen-value and Coalitional-Semivalue interaction index on networks;
- We empirically evaluate our measure by applying it to solve the inter-links prediction problem for a number of real-life networks, and show that it outperforms other local node-similarity measures.

The remainder of this paper is organized as follows. In Section 2, we introduce some basic notations and concepts from graph theory and cooperative games theory. We formally define the new node-similarity measure in Section 5, and analyze its computational complexity in Section 4. An efficient algorithm is then presented in Section 5. The experimental results are presented in Section 6. A brief discussion of related bodies of literature is presented in Section 7, before concluding the paper.

2 PRELIMINARIES

In this section we introduce the key definitions and notation used throughout the paper.

\textbf{Network notation:} A graph (or a network) is denoted by $G(V, E)$, where $V = \{v_1, \ldots, v_{|V|}\}$ is the set of nodes and $E$ is the set of edges. We will sometimes write $G$ instead of $G(V, E)$ for brevity. In this paper we consider only undirected and unweighted graphs. We will often use $v$ and $u$ to denote two arbitrary nodes. For any two nodes, $v, u \in V$, the distance (i.e., the length of the shortest path) between them will be denoted by $d(v, u)$. A community in a network is a subset of nodes, whereas a community structure is an exhaustive and disjoint set of communities.

A \textit{centrality index} (or simply a \textit{centrality}) measures the importance of individual nodes. One of the fundamental centrality indices is \textit{degree centrality} [14, 13, 36], simply measures the importance of a node, $v$, based solely on the degree of $v$—the number of nodes within 1 step from $v$. This can be generalized to $k$ steps, resulting in what is known as \textit{k-steps degree centrality}. The notion of centrality can also be generalized to groups of nodes, resulting in what is known as \textit{group centrality} [13]. One such group centrality that we will focus on in this paper is \textit{k-steps group degree centrality}.

\textbf{Definition 1} Given a network $G$, an integer $k \in \{1, \ldots, |V|\}$, and a community $S \subseteq V$, the \textit{k-steps group degree centrality} of $S$ is:

$$\left\{v \in V : \min_{u \in S} d(u, v) \leq k\right\} \setminus S$$

Some authors [35, 46] interpret the above formula as a \textit{sphere of influence} of the community $S$ in the network. From this perspective, the parameter $k$ can be interpreted as the “diameter” of this sphere.

\textbf{Coalitional games:} A coalitional game in characteristic function form (also called a cooperative game) is comprised of a set of players $N = \{1,2,\ldots,|N|\}$ and a characteristic function $\nu : 2^N \to \mathbb{R}$ which evaluates each coalition $C \subseteq N$ of players, under the assumption that $\nu(\emptyset) = 0$. We often refer to $\nu(C)$ as the value, or payoff, of $C$.

\textbf{Semivalues:} This is a family of payoff-division schemes, or solution concepts, designed to specify how the payoff of any given coalition should be divided among its members [11]. It is centered around the notion of \textit{marginal contribution}; for every player, $i \in N$, and every coalition, $C \subseteq N$, the marginal contribution of $i$ to $C$ is:

$$MC(C, i) = \nu(C \cup \{i\}) - \nu(C).$$

Now, let $\beta : \{0, \ldots, |N| - 1\} \to [0,1]$ be a discrete probability distribution, where $\beta(k)$ is the probability that a coalition of size $k$ is drawn from the set of all possible coalitions whose size is no greater than $|N| - 1$. Then, a Semivalue is defined as follows:

\textbf{Definition 2} Given a game, $(N, \nu)$, and a discrete probability distribution, $\beta : \{0, \ldots, |N| - 1\} \to [0,1]$, $\sum_{0 \leq k < |N|} \beta(k) = 1$, the Semivalue of a player, $i \in N$, is:

$$SEMI_i(N, \nu) = \sum_{0 \leq k < |N|} \beta(k) \mathbb{E}[MC(X^k, i)],$$

where $X^k$ is a coalition of size $k$ drawn uniformly from $\{C : C \subseteq N \setminus \{i\} \land |C| = k\}$, and $\mathbb{E}[\cdot]$ is the expected-value operator.

The first Semivalue to appear in the literature was the \textit{Shapley value} [43], which is now recognized as a fundamental concept in cooperative game theory due to its many desirable properties, see, e.g., [7]. Another prominent Semivalue is the \textit{Banchoff power index} [3], which has also been studied extensively. Those two Semivalues are defined by the following $\beta$-functions:

$$\beta^{\text{Shapley}}(i) = \frac{1}{|N|} \quad \text{and} \quad \beta^{\text{Banchoff}}(i) = \left(\frac{|N|-1}{|N|}\right)^i.$$

\textbf{Interaction index:} One possible way to interpret the synergy (or added value) that results from the interaction between players $i$ and $j$ is as follows: $S(i, j) = \nu(\{i, j\}) - \nu(\{i\}) - \nu(\{j\})$. One can also measure such synergy with respect to any particular coalition, $C \subseteq N$, as follows:

$$S(C, i, j) = MC(C \cup \{i, j\}) - MC(C, i) - MC(C, j),$$

where $MC(C, \{i, j\}) = \nu(C \cup \{i, j\}) - \nu(C)$. The interaction index of $i$ and $j$, denoted by $I_{i,j}(N, \nu)$, is a weighted average of such synergy, taken over all coalitions in the game. The absolute value of $I_{i,j}(N, \nu)$ indicates the intensity of the interaction between the two players; greater values indicate greater intensity. In contrast, the sign of $I_{i,j}(N, \nu)$ reflects the nature of the influence that $i$ and $j$ have on one another: $I_{i,j}(N, \nu) < 0$ means they influence each other negatively; $I_{i,j}(N, \nu) > 0$ means they influence each other positively; $I_{i,j}(N, \nu) = 0$ means they either do not influence each other, or their influences cancel out.

By combining a Semivalue with the interaction index, we obtain a \textit{Semivalue interaction index}, defined as follows:

\textbf{Definition 3} Given a game, $(N, \nu)$, and a discrete probability distribution, $\beta : \{0, \ldots, |N| - 1\} \to [0,1]$, $\sum_{0 \leq k < |N|} \beta(k) = 1$, the Semivalue interaction index of a pair of players, $i, j \in N$, is:

$$I_{i,j}^{SEMI}(N, \nu) = \sum_{0 \leq k < |N| - 2} \beta(k) \mathbb{E}[S(X^k, i, j)],$$

where $X^k$ is a coalition of size $k$ drawn uniformly from $\{C : C \subseteq N \setminus \{i, j\} \land |C| = k\}$, and $\mathbb{E}[\cdot]$ is the expected-value operator.
where $X^k$ is a coalition drawn uniformly at random from: $\{C : C \subseteq N \setminus \{i, j\} \land |C| = k\}$, and $\mathbb{E}[\cdot]$ is the expected-value operator.

The three Semivalue interaction indices that are widely studied in literature are presented in Table 1.

<table>
<thead>
<tr>
<th>Interaction index name</th>
<th>$\beta(l)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapley-value interaction index [38, 17]</td>
<td>$\frac{1}{n}$</td>
</tr>
<tr>
<td>Banzhaf-index interaction index [42]</td>
<td>$\frac{\binom{</td>
</tr>
<tr>
<td>Chaining interaction index [34]</td>
<td>$\frac{2(l + 1)}{(n-1)(n-2)}$</td>
</tr>
</tbody>
</table>

Table 1. Values of $\beta$ for the three main Semivalue interaction indices.

In addition to cooperative game theory, the interaction index has also been studied in various other fields, such as fuzzy systems, aggregation function theory, multi-criteria decision making, statistics and data analysis [33].

Cooperative games with coalition structure: A cooperative game can be viewed and analyzed from the ex ante perspective, where the agents have not yet decided on which coalitions to form. Conversely, the game may be analyzed from the a priori perspective, where the agents have already formed a specific coalition structure, $CS = \{C_1, \ldots, C_m\}$. From this perspective, a cooperative game with a coalition structure is a tuple, $(N, CS, \nu)$. Arguably, the most established extension of the Shapley value to such games is the Owen value [39]. Before explaining how it works, we need to first introduce the notion of a quotient game. In particular, given a cooperative game with a coalition structure, $(N, CS, \nu)$, the corresponding quotient game, $(CS, \nu^\beta)$, is a game whose set of players is $CS$ (i.e., every coalition in $CS$ is considered to be a single player), and whose characteristic function is defined as follows:

$$\nu^\beta(R) = \nu\left( \bigcup_{x \in R} C_x \right)$$

for all $R \subseteq M$, where $M = \{1, \ldots, m\}$ is the set of coalition numbers. For every $R \subseteq M$, we will use $Q_R$ to denote $\bigcup_{x \in R} C_x$. For example, if $CS = \{C_1, C_2, C_3\}$: $C_1 = \{1, 2\}$, $C_2 = \{3, 4\}$, $C_3 = \{5\}$, then $Q_{\{1, 2, 3\}} = \{1, 2, 5\}$, and $\nu^{\beta}\left(\{1, 2, 3\}\right) = \nu\{1, 2, 5\}$.

Having presented the quotient game, we are now ready to define the Owen value as follows:

**Definition 4** Given a cooperative game with a coalition structure, $(N, CS, \nu)$, the Owen value of a player $i \in C_x \in CS$ is:

$$OV_i(N, CS, \nu) = \sum_{R \subseteq M \setminus \{x\}} \frac{1}{|M| - |R|} \sum_{c \subseteq C_x \setminus \{i\}} \frac{1}{|C_x| - |c|} MC(Q_R \cup C, i). \quad (4)$$

One generalization of the Owen value that was recently introduced in the literature is the Coalitional Semivalues [48], defined as follows:

**Definition 5** Given a game, $(N, CS, \nu)$, and a discrete probability distribution, $\beta : \{0, \ldots, |M| - 1\} \rightarrow [0, 1]$, $\sum_{0 \leq k \leq |M|} \beta(k) = 1$,

the Coalitional Semivalue of a player $i \in C_x \in CS$ is:

$$CSEMI_i(N, CS, \nu) = \sum_{0 \leq k \leq |M| - 1} \beta(k) \sum_{0 \leq l \leq |C_x|} \alpha(l)\mathbb{E}[MC(Q_k \cup X^l, i)]. \quad (5)$$

where $T^k$ is a subset drawn from $\{R : R \subseteq M \setminus \{x\} \land |R| = k\}$ uniformly at random; $X^l$ is a subset of size $l$ drawn from $\{C : C \subseteq C_x \setminus \{i\} \land |C| = l\}$ uniformly at random; $\mathbb{E}[\cdot]$ is the expected-value operator; and $\alpha : \{0, \ldots, |C_x| - 1\} \rightarrow [0, 1]$, $\sum_{l=0}^{C_x-1} \alpha(l) = 1$.

As shown in Table 2, by adopting the appropriate probability distributions, we obtain the Owen value [39] or any of its modifications proposed in the literature to date, namely: Owen-Banzhaf value [40], symmetric coalitional Banzhaf value [2], and symmetric coalitional p-binomial Semivalues [6].

<table>
<thead>
<tr>
<th>Solution name</th>
<th>$\beta(k)$</th>
<th>$\alpha(l)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Owen value [39]</td>
<td>$\frac{</td>
<td>M</td>
</tr>
<tr>
<td>Owen-Banzhaf value [40]</td>
<td>$\frac{</td>
<td>M</td>
</tr>
<tr>
<td>symmetric coalitional Banzhaf value [2]</td>
<td>$\frac{</td>
<td>M</td>
</tr>
<tr>
<td>symmetric coalitional p-binomial Semivalue 6</td>
<td>$p^k(1-p)^{</td>
<td>M</td>
</tr>
</tbody>
</table>

Table 2. Values of $\alpha$ and $\beta$ for the Owen value and its various extensions.

**Interaction index on games with coalition structure**: Finally we are ready to define the Owen-value interaction index for cooperative games with a coalition structure. Here, our definition is for the interaction between nodes belonging to different coalitions [54].

**Definition 6** Given a cooperative game with a coalition structure, $(N, CS, \nu)$, and two players, $i \in C_x \in CS$ and $j \in C_y \in CS$ such that $C_x \neq C_y$, the Owen-value interaction index between $i$ and $j$ is:

$$I^{OV}_{i,j}(N, CS, \nu) = \sum_{R \subseteq M \setminus \{x, y\}} \frac{1}{(|M| - 1)(|M| - 2)} \sum_{c \subseteq (C_x \cup C_y) \setminus \{i, j\}} S(Q_R \cup C_I, i, j). \quad (6)$$

3 **A NEW INTERACTION INDEX FOR NETWORKS**

Inspired by the inter-links prediction problem, we construct a new node-similarity measure using three building blocks. The first block is the interaction index (to analyze pairs of nodes), the second block is the Coalitional Semivalue (to analyze nodes given a community structure), and the third block is the $k$-steps group degree centrality (to quantify the importance of subsets of nodes). To put the three pieces together, our first step is to introduce the following game.

**Definition 7** A cooperative game with a coalition structure (played on a graph is a tuple, $(G, CS, \nu_G)$, where $G$ is a graph, $CS$ is a
community structure, and \( \nu_G : 2^{|V|} \to \mathbb{R} \) is a characteristic function defined over the graph \( G \).

We use one such game, where the characteristic function is the k-steps group degree centrality, defined for all \( k \in \{1, \ldots, |V|\} \) and all \( S \subseteq V \) as follows:

\[
\nu^k(S) = \left\{ v \in V : \min_{u \in \bar{S}} d(u, v) \leq k \right\} \setminus S .
\]

The second step is to combine the Coalitional Semivalue with the interaction index, as shown below:

**Definition 8** Given a cooperative game with a coalition structure, \((N, CS, \nu)\), a discrete probability distribution, \( \beta : \{0, \ldots, |N| - 2\} \to [0, 1]\), \( \sum_{k \in \{0, \ldots, |N| - 2\}} \beta(k) = 1 \), and two players, \( i \in C_z \in CS \) and \( j \in C_y \in CS \) such that \( C_z \neq C_y \), the Coalitional-Semivalue interaction index between \( i \) and \( j \) is:

\[
I^C_{i,j} (N, CS, \nu, \beta) = \sum_{0 \leq k \leq |N| - 1} \beta(k) \sum_{0 \leq l \leq |C_z|-1} \alpha(l) E[S(Q_{k,l} \cup X^I, i, j)], \tag{7}
\]

where \( T^k \) is a subset of size \( k \) drawn from \( \{R : R \subseteq M \setminus \{x, y\} \setminus |R| = k\} \) uniformly at random; \( X^I \) is a subset of size \( I \) drawn from \( \{C : C \subseteq C_z \cup C_y \setminus \{i, j\} \setminus |C| = I\} \) uniformly at random; \( E[l] \) is the expected value operator; \( \alpha : \{0, \ldots, |C_z \cup C_y \setminus \{i, j\}\} \to [0, 1] \); \( \sum_{l=0}^{\infty} E[l] \alpha(l) = 1 \).

This is a natural extension of Owen-value interaction index that is in line with the definition of Coalitional Semivalues. For instance, by setting \( \beta(k) = \frac{1}{2^{k+1}} \) and \( \alpha(l) = \frac{1}{2^{k+1}} \), we obtain the Owen-value interaction index.

Now, we are ready to introduce our new node-similarity measure:

**Definition 9** The Coalitional-Semivalue similarity measure between \( v \in C_z \) and \( u \in C_y \) in graph \( G(V, E) \) with community structure \( CS \) is defined as:

\[
I^C_{u,v} (V, CS, \nu, \beta) .
\]

Many standard measures evaluate the similarity between two nodes by quantifying the intersection of their spheres of influence. In contrast, the main advantage of our measure is that the intersection is evaluated in the context of the exponential number of subsets of communities and nodes in the network, which may allow us to compute similarity more accurately. One potential drawback of our approach is its potentially-high computational complexity, due to the exponential number of subsets. However, in the following section we develop the closed-form formula for the k-steps Coalitional-Semivalue similarity measure which allows us to compute it in polynomial time.

**4 COMPUTATIONAL ANALYSIS**

In this section, we circumvent the main potential obstacle that may hamper the application of the Coalitional-Semivalue interaction index—the computational complexity. In more detail, Equation (7) requires iterating over an exponential number of subsets of \( V \). However, building upon a combinatorial and probabilistic analysis, we will develop two polynomial algorithms: one for the Coalitional-Semivalue interaction index, which runs in \( O(|V|^3) \) time, and the other is for a special case of this index, namely the Owen-value interaction index, which runs in just \( O(|V|) \) time.

To this end, let \( \nu^k_{\text{CSD}} (V, CS, \nu^k_\beta) \), where \( CS \) is a community structure, \( \nu^k_\beta \) is the k-steps degree centrality, and \( u, v \in V \). The main theoretical result in this paper is as follows:

**Theorem 1** \( \text{CSDEGREE}_\text{II} \) is in \( P \).

We note that the above theorem fills a gap in the literature, as highlighted in Table 7 (see Section 7). Before presenting the proof, we first need some additional notation. For every node \( v \in V \), let \( N_k(v) \) denote the set of “neighbors” reachable from \( v \) with at most \( k \) steps, and let \( d_{eu}(u) \) denote the number of such nodes. More formally, we have: \( N_k(v) = \{ u \in V : d(u, v) \leq k \land v \neq u \} \) and \( d_{eu}(v) = |N_k(v)| \). We extend this notation to sets of nodes. That is, \( N_k(C) = \bigcup_{v \in C} N_k(v) \setminus C \) and \( d_{eu}(C) = |N_k(C)| \). Moreover, for any given node, \( v \in C_z \in CS \), we denote the set of adjacent communities as \( N^k_{CS}(v) = \{ C_y \in CS \setminus C_z : C_y \cap N_k(v) \neq \emptyset \} \), the inter-community degree as \( d_{CS}(v) = |N^k_{CS}(v)| \), the set of neighbors within some community \( C_y \in CS \) as \( N^k_y(v) = N_k(v) \cap C_y \), and the corresponding intra-community degree as \( d^k_{CS}(v) = |N^k_y(v)| \).

These can be extended to two communities as follows: \( N^k_{CS}(v) = N_k(v) \cap (C_y \cup C_z) \) and \( d_{CS}(v) = |N^k_{CS}(v)| \).

In our proof we follow the line of our earlier work [46], where we developed an algorithm to computed the Shapley value-based interaction index was proposed. In this work, we will extend the proof from [46] to take into consideration both the community structure the Owen value-based interaction index, which is much more complex than its Shapley value-based counterpart.

**Proof:** First of all, let us focus on Equation (7). More specifically, for each pair of nodes \( u, v \in V \) such that: \( v \in C_z \in CS \) and \( u \in C_y \in CS \neq C_z \), we will show how to compute \( E[S(Q_{k,l} \cup X^I, u, v)] \) as the expected value of their synergy with respect to the random set \( Q_{k,l} \cup X^I \). Recall that \( T^k \) is drawn uniformly from the set \( \{R : R \subseteq M \setminus \{i, j\} \setminus |R| = k\} \), and \( X^I \) is drawn uniformly from the set \( \{C : C \subseteq C_z \cup C_y \setminus \{i, j\} \setminus |C| = I\} \). Also recall that \( Q_{k,l} \subseteq \bigcup_{i \in R} C_i \). Now if we denote \( R_{k,l} = Q_{k,l} \cup X^I \), then:

\[
E[S(R_{k,l}, u, v)] = E[MC(R_{k,l}, \{u, v\})] - E[MC(R_{k,l}, u)] - E[MC(R_{k,l}, v)] = E[MC(R_{k,l}, u)] + E[MC(R_{k,l}, v)] - E[MC(R_{k,l}, u \cap v)] - E[MC(R_{k,l}, u)] - E[MC(R_{k,l}, v)] = -E[MC(R_{k,l}, u \cap v)],
\]

where \( MC(R_{k,l}, u \cap v) \) is what we call “common” contribution of two nodes \( u \) and \( v \), which is illustrated in Figure 1, and defined as follows:

\[
E[MC(R_{k,l}, u \cap v)] = E[MC(R_{k,l}, \{u, v\})] - E[MC(R_{k,l}, u)] - E[MC(R_{k,l}, v)].
\]

Now given the function \( \rho_{k,l}^2 \), the pair of nodes \( u, v \in V \) can make a positive common contribution to the set of nodes \( R_{k,l} \) only through some node from the intersection of their neighborhoods, i.e., some node \( n \in N_k(u) \cap N_k(v) \). Intuitively, this happens when such a node \( n \) is not under the influence of the set \( R_{k,l} \) but is under the influence of \( R_{k,l} \cup \{u, v\} \).

---

5 Here, we do not mean to take the intersection of nodes \( u \) and \( v \), as this would be incorrect. Instead, for notation convenience, we write \( u \cup v \) when referring to the intersection between the contributions of \( u \) and \( v \).
We formalize the above observation by introducing the Bernoulli random variable indicating whether nodes $v, u$ make a positive common contribution to the set $R^{k,l}$ through a node $n \in N_k(v) \cap N_k(u)$:

$$E[B^+_k,l,v,u,n] = P[(N_k(n) \cup \{n\}) \cap R^{k,l} = \emptyset].$$

(8)

On the other hand, the pair $v, u \in V$ can make a negative common contribution to the set of nodes $R^{k,l}$ through $v$ or $u$. This happens when either of those two nodes is under the influence of $R^{k,l}$ but not under the influence of $R^{k,l} \cup \{v, u\}$. Note that when analyzing the common contribution, we only consider nodes in $N_k(v) \cap N_k(u)$. As such, from the negative-contribution perspective, we only consider cases where $v \in N_k(u)$ or $u \in N_k(v)$.

We formalize the above observation by introducing two Bernoulli random variables indicating whether nodes $v, u$ make a negative common contribution to $R^{k,l}$ through the node $n$ or $v$ is defined as:

$$E[B^-_{k,l,v,u,n}] = P[(N_k(n) \cap \{v\}) \cap R^{k,l} \neq \emptyset],$$

(9)

$$E[B^-_{k,l,u,v,n}] = P[(N_k(n) \cap \{u\}) \cap R^{k,l} \neq \emptyset].$$

(10)

Now, we will develop an exact formula for the equations (8), (9) and (10). We start with a positive contribution. The important observation is that the set $R^{k,l}$ is drawn from the sample space $\Omega$, where $|\Omega| = \binom{|M|-2}{|C_i|+|C_j|-2}$. Having this in mind, we denote $P[(N_k(n) \cup \{n\}) \cap R^{k,l} = \emptyset]$ by $P^+$ and we get:

$$P^+ = \begin{cases} \frac{\binom{|M|-1-\deg_k CS(n)}{k}}{\binom{|M|-2}{k}} & \text{if } n \notin C_i \cup C_j \\ \frac{\binom{|M|-1-\deg_k CS(n)}{k}C_i|+|C_j|-\deg_k CS(n)}{\binom{|M|-2}{k}} & \text{if } n \in C_i \cup C_j \end{cases}$$

(11)

Now if $n \notin C_i \cup C_j$, the above formula has the following combinatorial interpretation. The random set $R^{k,l}$ can contain any community from $CS$ except for $C_i$ and $C_j$; there are $|M|-2$ such communities, where $M = \{1, \ldots, m\}$ is the set of community numbers. In order to satisfy the condition $N_k(v) \cap T^k \neq \emptyset$, from all communities in $CS \setminus \{C_i, C_j\}$ we can draw all those that are not in the scope of $n$. There are $|M|-2 - \deg_k CS(n)$ such communities. However, two additional facts also play an important role: the fact that the community containing $n$ should not be in $R^{k,l}$, and the fact that $R^{k,l}$ can contain any community from $CS$ except for $C_i$ and $C_j$. Taking into account, the final number of communities is: $|M|-1 - \deg_k CS(n)$. Thus, the probability of choosing a set $R^{k,l}$ satisfying our condition $N_k(v) \cap T^k \neq \emptyset$ is exactly: $\frac{\binom{|M|-1-\deg_k CS(n)}{k}}{\binom{|M|-2}{k}}$.

Next, we show how to satisfy the condition that $N_k(v) \cap X^l \neq \emptyset$. To this end, from the set $(C_i \cup C_j) \setminus \{v, u\}$ we need to exclude those nodes that are in the scope of $n$. There are $(C_i \cap C_j) - 2 - \deg_k CS(n)$ such nodes. However, taking into account that $v, u \in N_k^CS(n)$, the probability of choosing a set $R^{k,l}$ satisfying the condition $N_k(v) \cap X^l \neq \emptyset$ is exactly: $\frac{\binom{|C_i|+|C_j|-\deg_k CS(n)}{k}}{\binom{|M|-2}{k}}$.

In order to compute the negative contribution, we consider the complementary event: $P^- = 1 - P[(N_k(v) \cap R^{k,l} = \emptyset]$. Using the same combinatorial argument as the one used to compute $P^+$, we get:

$$P^- = 1 - \left(\frac{\binom{|M|-1-\deg_k CS(n)}{k}}{\binom{|M|-2}{k}}\right)^2,$$

(12)

The formula combining equations (8) and (9) and its analytic form given in equations (11) and (12) is:

$$E[S(R^{k,l}, u, v)] = E[B^-_{k,l,v,u}] + E[B^-_{k,l,u,v}] - \sum_{n \in N_k(v) \cap N_k(u)} E[B^+_{k,l,v,u,n}],$$

(13)

Notice that the sets $N_k^CS(n), N_k^CS(u)$ and $N_k(u)$ are easily computable in polynomial time. Based on this, the closed-form formula (13) together with Equation (7) prove that CSDEGREEII is solvable in polynomial time, i.e., it belongs to the class P.

Finally, note that in our proof we omitted the case when $v$ and $u$ belong to the same community. Although, in this case, the reasoning slightly differs from the above, it results in almost the same formula. Nevertheless, such a case is not interesting from the perspective of inter-links recommendation.

Now, let us denote by OVDGEEII the problem of calculating $\nu^D_{k}(V, x, y, \nu_D^k)$, where $CS$ is a community structure, $\nu_D^k$ is the $k$-steps degree centrality, and $u, v \in V$. Then, the following corollary immediately follows from Theorem 1:

Corollary 1 OVDGEEII is in P.

Building upon the above theoretical results, we will propose in the next section two algorithms; one solves CSDEGREEII in $O(|V|^3)$ time; the other solves OVDGEEII in just $O(|V|^2)$ time, after some preprocessing stage that requires $O(|V|^2)$ time.

5 ALGORITHMS

In this section, we use Equation (13) to develop a polynomial time algorithm for computing the $k$-steps Coalitional-Semivalue interaction index. In particular, Algorithm 1 computes the $k$-steps Coalitional-Semivalue interaction index for a given pair of nodes $u, v \in V$ in the graph $G$. This algorithm is basically an implementation of Equation (7), whereby the expected value operator $E[S(R^{k,l}, u, v)]$ is computed using Equation (13).
Algorithm 1: Computing the k-steps Coalitional-Semivalue interaction index

**Input:** Graph $G(V, E)$, functions $\beta$ and $\alpha$, community structure $CS$, nodes $v, u \in V$ where $v \in C_i \in CS, u \in C_j \in CS$

**Data:** for $u \in V:$
- $N_k(u)$—the set of k-neighbors
- $N^CS_k(u)$—the set of adjacent communities
- $N^L_j(u)$—the set of adjacent nodes within $C_i \cup C_j$

**Output:** $I^{CS} u,v$ k-steps coalitional-Semivalue interaction index

1. $I^{CS} u,v \leftarrow 0$
2. for $k \leftarrow 0$ to $|M| - 2$
3. for $l \leftarrow 0$ to $|C_i \cup C_j| - 2$
4. $S \leftarrow 0$
5. foreach $n \in (N_k(n) \cap N_k(u)) \setminus (C_i \cup C_j)$
   - $S \leftarrow S - \frac{(I_m), n \setminus (C_i \cup C_j) - 1 - deg^L_j(n)}{(I_m), n \setminus (C_i \cup C_j) - 1 - deg^L_j(n)}$
6. foreach $n \in (N_k(n) \cap N_k(u) \setminus (C_i \cup C_j)$
   - $S \leftarrow S - \frac{(I_m), n \setminus (C_i \cup C_j) - 1 - deg^L_j(n)}{(I_m), n \setminus (C_i \cup C_j) - 1 - deg^L_j(n)}$
7. if $u \in N_k(u)$ then
   - $S \leftarrow S - \frac{(I_m), n \setminus (C_i \cup C_j) - 1 - deg^L_j(n)}{(I_m), n \setminus (C_i \cup C_j) - 1 - deg^L_j(n)}$
8. $I^{CS} u,v \leftarrow I^{CS} u,v + \beta(k)\alpha(l)\alpha(k)$

It is easy to see that Algorithm 1 runs in $O(|V|^3)$ time. Note that the algorithm requires a preprocessing stage in which the sets $N^CS_k(u)$, $N^L_j(u)$ and $N_k(u)$ are computed. Let us analyze the time required to perform this preprocessing stage. For each node $n \in V$, the sets $N^CS_k(u)$ and $N_k(u)$ can be computed using breadth-first search in $O(|V||V| + |E|)$ time. Furthermore, we can store all coalition values, i.e., store $\nu(C), \forall C \subseteq N$, using $O(|V|^2)$ space. Next, for the pair of communities $C_i \cup C_j$ we can compute the set $N^L_j(n)$ in $O(|V|^3)$ time. As can be seen, compared to the time required to run Algorithm 1, the preprocessing stage takes negligible time.

Although $O(V^3)$—the time required to run Algorithm 1—is very fast compared to a naive (exponential-time) algorithm, it is still not fast enough to be applied for link prediction in large networks. With this in mind, we now present an even faster algorithm to compute k-steps Owen-value interaction index; see Algorithm 2. Specifically, this algorithm runs in $O(|V|^3)$ time, and requires the aforementioned preprocessing stage to compute the sets $N^CS_k(u)$, $N^L_j(u)$ and $N_k(u)$. This improvement allows us to compute the similarity between each pair of nodes (not just a single pair) in $O(|V|^3)$ time. Thus, the entire procedure of link prediction also requires $O(|V|^3)$ time.

6 Empirical Evaluation

In this section, we empirically demonstrate the effectiveness of our node-similarity measure in detecting links across communities. Specifically, in our experiments we use the Owen value-based variant of our measure, as it can be computed efficiently using Algorithm 2.

We compare our measure against six local similarities measures outlined in Table 3; these are arguably the most efficient solutions to the local link-prediction problem in the literature [32, 46].

<table>
<thead>
<tr>
<th>Similarity Name</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Neighbors (CN)</td>
<td>$S^{CN}_{u,v} = N_k(u) \cap N_k(v)$</td>
</tr>
<tr>
<td>Salton Index (SI)</td>
<td>$S^{SI}_{u,v} = \frac{N_k(u) \cap N_k(v)}{\sqrt{</td>
</tr>
<tr>
<td>Jaccard Index (JI)</td>
<td>$S^{JI}_{u,v} = \frac{</td>
</tr>
<tr>
<td>Adamic-Adar Index (AA)</td>
<td>$S^{AA}<em>{u,v} = \sum</em>{n \in N_k(u) \cap N_k(v)} \frac{1}{deg_S(n)}$</td>
</tr>
<tr>
<td>Resource Allocation (RA)</td>
<td>$S^{RA}<em>{u,v} = \sum</em>{n \in N_k(u) \cap N_k(v)} \frac{1}{N_k(v)}$</td>
</tr>
<tr>
<td>Shapley-value interaction index (SV)</td>
<td>$S^{SV}<em>{u,v} = I^{SV}(V, \nu</em>{D})$</td>
</tr>
</tbody>
</table>

Table 3. The six local node-similarity measures used in our experiments.

We evaluate the effectiveness of each node-similarity measure using a standard procedure from the literature on link-prediction. In particular, we compute the similarity of each pair of disconnected nodes, $v, u \in V : (v, u) \notin E$, belonging to two different communities, $C_i, C_j \subseteq CS$ where $v \in C_i, u \in C_j$. After that, links are proposed between the most similar such pairs of nodes. Since in this paper we focus on finding the most accurate predictions based only on local information, we set $k = 1$ in all our experiments. Each such experiment is conducted as follows: given a graph $G = (V, E)$ with a community structure $CS$, we create a new graph $G' = (V, E')$, which is similar to $G$ and with the same $CS$ but where 10% or 20% of inter-edges are removed at random. Then, we compute the similarity of each disconnected pair of nodes from different communities.

We also conducted experiments in which 30% and 40% of inter-edges were removed. The effectiveness of all measures was reduced proportionally.
in $G'$. In order to evaluate the results, we use the Area Under the Curve (AUC) measure [32]. The whole process is repeated 100 times and the average AUC is taken.

In more detail, the AUC is computed using the Mann-Whitney $U$ test [22]. To this end, let $R$ be the set of all disconnected pairs of nodes from different communities in $G'$. This set can be divided into two disjoint sets: the set of “missing” links, denoted by $M$ (i.e., $M = E \setminus E'$), and the set of “non-existing” links, denoted by $N$ (i.e., $N = (V \times V) \setminus E$). Let $n = |M||N|$ be the number of all comparisons between the missing links and the non-existing links. Furthermore, let $n'$ be the number of such comparisons in which the missing link is ranked higher than the non-existing link. Finally, let $n''$ be the number of such comparisons in which both links are ranked the same. Then, the AUC is computed as follows:

$$AUC = \frac{n' + n''}{n}.$$  

With this formula, if AUC equals 1 then all missing links are ranked higher than the non-existing links; this is the best possible ranking, where $n' = n$ and $n'' = 0$. On the other extreme, if AUC equals 0 then none of the missing links is ranked higher than, or even the same as, any of the non-existing links; this is the worst possible ranking, where $n' = 0$ and $n'' = 0$. A completely random ranking falls between the two extremes, with an expected AUC of 0.5.

We study 8 widely-used real-life networks: Tribes [12], Taro [21], Zachary [55], Terrorists [24], Surfers [30], Polbooks [1], Football [15] and Jazz [16]. Table 4 specifies the sizes of these eight networks, as well as the sizes of the community structures therein. For each of them, we report the results for the community structure identified by the multilevel community-detection algorithm [52]. We also experimented with other community-detection algorithms, such as Walktrap [41], Fastgreedy [9], and Girvan-Newman [15]; they produced almost the same community structures as the multilevel algorithm. As such, the choice of the community-detection algorithm had a negligible impact on our results.

| Network  | $|V|$ | $|E|$ | $|CS|$ | Network  | $|V|$ | $|E|$ | $|CS|$ |
|----------|------|------|------|----------|------|------|------|
| Tribes   | 16   | 58   | 3    | Surfers  | 43   | 336  | 2    |
| Taro     | 22   | 39   | 5    | Polbooks| 105  | 441  | 4    |
| Zachary  | 34   | 78   | 4    | Football| 115  | 613  | 10   |
| Terrorists|64 | 243  | 5    | Jazz    | 198  | 2742 | 4    |

Table 4. The sizes of the networks and their community structures used in the experiments.

Tables 5 and 6 present the results for our Owen value-based measure, as well as the six measures from Table 3, given 8 real-life networks in which 10% of inter-links were removed.

<table>
<thead>
<tr>
<th>Network</th>
<th>OV</th>
<th>SV</th>
<th>RA</th>
<th>CN</th>
<th>AA</th>
<th>JI</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tribes</td>
<td>0.754</td>
<td>0.438</td>
<td>0.579</td>
<td>0.502</td>
<td>0.581</td>
<td>0.633</td>
<td>0.658</td>
</tr>
<tr>
<td>Taro</td>
<td>0.573</td>
<td>0.483</td>
<td>0.394</td>
<td>0.516</td>
<td>0.550</td>
<td>0.571</td>
<td>0.483</td>
</tr>
<tr>
<td>Zachary</td>
<td>0.703</td>
<td>0.628</td>
<td>0.684</td>
<td>0.533</td>
<td>0.657</td>
<td>0.573</td>
<td>0.617</td>
</tr>
<tr>
<td>Terrorists</td>
<td>0.834</td>
<td>0.774</td>
<td>0.778</td>
<td>0.744</td>
<td>0.795</td>
<td>0.782</td>
<td>0.748</td>
</tr>
<tr>
<td>Surfers</td>
<td>0.881</td>
<td>0.731</td>
<td>0.770</td>
<td>0.746</td>
<td>0.765</td>
<td>0.797</td>
<td>0.791</td>
</tr>
<tr>
<td>Polbooks</td>
<td>0.806</td>
<td>0.783</td>
<td>0.789</td>
<td>0.743</td>
<td>0.785</td>
<td>0.755</td>
<td>0.772</td>
</tr>
<tr>
<td>Football</td>
<td>0.403</td>
<td>0.346</td>
<td>0.365</td>
<td>0.314</td>
<td>0.336</td>
<td>0.373</td>
<td>0.388</td>
</tr>
<tr>
<td>Jazz</td>
<td>0.962</td>
<td>0.918</td>
<td>0.928</td>
<td>0.927</td>
<td>0.902</td>
<td>0.919</td>
<td>0.918</td>
</tr>
</tbody>
</table>

Table 5. The area under curve (AUC) for our measure (OV) as well as the six measures from Table 3, given 8 real-life networks in which 10% of inter-links were removed.

<table>
<thead>
<tr>
<th>Network</th>
<th>OV</th>
<th>SV</th>
<th>RA</th>
<th>CN</th>
<th>AA</th>
<th>JI</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tribes</td>
<td>0.729</td>
<td>0.421</td>
<td>0.584</td>
<td>0.493</td>
<td>0.569</td>
<td>0.622</td>
<td>0.641</td>
</tr>
<tr>
<td>Taro</td>
<td>0.525</td>
<td>0.431</td>
<td>0.439</td>
<td>0.358</td>
<td>0.441</td>
<td>0.505</td>
<td>0.510</td>
</tr>
<tr>
<td>Zachary</td>
<td>0.644</td>
<td>0.584</td>
<td>0.622</td>
<td>0.519</td>
<td>0.616</td>
<td>0.511</td>
<td>0.548</td>
</tr>
<tr>
<td>Terrorists</td>
<td>0.819</td>
<td>0.771</td>
<td>0.782</td>
<td>0.746</td>
<td>0.785</td>
<td>0.759</td>
<td>0.715</td>
</tr>
<tr>
<td>Surfers</td>
<td>0.867</td>
<td>0.735</td>
<td>0.766</td>
<td>0.730</td>
<td>0.762</td>
<td>0.766</td>
<td>0.778</td>
</tr>
<tr>
<td>Polbooks</td>
<td>0.780</td>
<td>0.739</td>
<td>0.748</td>
<td>0.746</td>
<td>0.747</td>
<td>0.752</td>
<td>0.748</td>
</tr>
<tr>
<td>Football</td>
<td>0.351</td>
<td>0.313</td>
<td>0.334</td>
<td>0.300</td>
<td>0.331</td>
<td>0.353</td>
<td>0.365</td>
</tr>
<tr>
<td>Jazz</td>
<td>0.966</td>
<td>0.919</td>
<td>0.928</td>
<td>0.905</td>
<td>0.918</td>
<td>0.921</td>
<td>0.924</td>
</tr>
</tbody>
</table>

Table 6. The area under curve (AUC) for our measure (OV) as well as the six measures from Table 3, given 8 real-life networks in which 20% of inter-links were removed.

We conclude this section with some negative results. Specifically, when using our measure for quasi-local (rather than local) link prediction (i.e., when $k > 1$) the performance of our algorithm drops considerably in terms of AUC for all the networks in our experiments. We believe this comes from the expression $\deg^2(n)$. In particular, even for a small $k$, every node in the network can reach many different communities, which can negatively influence the performance. This is because in such a case the differences between the nodes diminish, and nodes become indistinguishable by our measure.

![Figure 2](http://example.com/figure2.png)
7 RELATED WORK

Our contribution falls at the intersection of (i) positive computational results in cooperative game theory and (ii) efficient link prediction in graph theory. In this section, we briefly discuss both areas of research.

Starting with cooperative game theory, most solution concepts are NP-hard [7]. However, for cooperative games described over networks there is a growing body of literature with various positive results. In particular, when either group degree centrality, group closeness centrality, or group betweenness centrality [13] is used as a characteristic function, it was shown that the Shapley value can be computed in time polynomial in the network size [35, 47]. Moreover, it was proven that the problem of computing any Semivalues—parameterized by any polynomial time computable discrete probability distribution—belongs to the class P for degree and closeness centralities [45], as well as betweenness centrality [47].

Regarding games with a coalition structure, there are positive results about degree and closeness centralities. In more detail, for both of them we can compute the Owen value and the Coalitional Semivalue in polynomial time [48, 49]. However, it is still an open question whether the same holds for betweenness centrality.

Other positive results can be found on the computation of the interaction index on graphs. More specifically, it was shown that we can compute efficiently the Shapley-value and Semivalue interaction indices with degree centrality [46].

The results that are most closely related to our work are presented in Table 7. The abbreviations SVDegree, SDegree, OVDegree and CSDegree stand for computing the Shapley value-, the Semivalue-, the Owen value- and the Coalitional value-based degree centrality, respectively. Furthermore, SVDegreeII, SDegreeII, OVDegreeII and CSDegreeII are analogous problems related to computing the interaction indices.

<table>
<thead>
<tr>
<th>Computational result</th>
<th>Algorithm complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVDegree is in P</td>
<td>O(</td>
</tr>
<tr>
<td>SDegree is in P</td>
<td>O(</td>
</tr>
<tr>
<td>OVDegree is in P</td>
<td>O(</td>
</tr>
<tr>
<td>CSDegree is in P</td>
<td>O(</td>
</tr>
<tr>
<td>SVDegreeII is in P</td>
<td>O(</td>
</tr>
<tr>
<td>SDegreeII is in P</td>
<td>O(</td>
</tr>
<tr>
<td>OVDegreeII is in P</td>
<td>O(</td>
</tr>
<tr>
<td>CSDegreeII is in P</td>
<td>O(</td>
</tr>
</tbody>
</table>

*(1) some precomputation is required.*

Table 7. Computational complexity results for degree centrality and coalitional games played on graphs.

The second body of literature that is strongly related to our work is that on link prediction. Here, we focus on methods based on node-similarity measures [32]. Generally, one can distinguish between three groups of link prediction algorithms: local, quasi-local and global. In more detail, global algorithms consider the entire network, which is prohibitive in for large networks. To date, the most efficient algorithm in this group is Random Walk with Restart [50], which is based on PageRank [4]. In contrast, quasi-local algorithms try to strike a balance between prediction runtime and efficiency. The most effective algorithms here are Local Random Walk and Superseded Random Walk [31]. Finally, local algorithms predict a link between any pair of nodes based solely on the direct neighborhood of those nodes (see Table 3). In practice, these are the only algorithms that can be applied to large networks, e.g., with millions of nodes.

To the best of our knowledge, the two best approaches in this group are: (i) the Resource Allocation approach [56], which is inspired by the resource allocation dynamics on complex networks; and (ii) the Shapley-value interaction index [46] approach, which is rooted in cooperative game theory.

Some authors have already tried to increase the accuracy of link prediction by taking advantage of the community structure of a network [44, 51]. While they managed to enhance the prediction performance by adding an extra score to nodes from the same community, such an approach method seems to have little value in our application as we are only interested in predicting connections between different communities.

In addition to the methods that are based on node similarity, link prediction can also be carried out based on maximizing likelihood [8, 20], or based on probabilistic models [26, 25]. These methods are computationally complex and are out of scope of this paper.

8 SUMMARY AND FUTURE WORK

In this paper, we proposed a new local node-similarity measure for networks with a community structure. We empirically demonstrated its effectiveness as a solution to the problem of detecting links between (rather that within) communities. Our measure outperforms other local node-similarity measure from the literature, since it is the first one designed specifically to detect links between heterogeneous nodes, rather than homogeneous ones as is the case with the other measures. Importantly, the Owen value-based variant of our measure can be computed very efficiently; it requires O(|V|^2) time, after a preprocessing stage that requires O(|V|^2|E|) time. Interestingly, despite the inherent complexity of our measure (which comes from the complexity of the Owen value), link prediction using our algorithm takes the same time as the fastest alternative from the literature.

There are several directions for future work. Firstly, while we showed in this paper that the problem OVDegreeII is in P, it would be interesting to verify whether the problems OVClosenessII and OVBetweeIInessII are also in P. It would also be interesting to study the similarity measures that correspond to the aforementioned problems, and to evaluate their effectiveness as node-similarity measures for inter-link prediction.

Secondly, since our measure is only restricted to non-overlapping communities, another interesting direction would be to extend our measure to graphs with overlapping communities [27]. Recently, an approach to measure the power of individual nodes in such networks was proposed [49]. In more detail, the authors defined a cooperative game with overlapping coalitions on a graph, and used a game-theoretic concept called the Configuration value to compute the power of an individual node. It is an open question whether CVDegree and CVDegreeII are in P, where CVDegree stands for Configuration value-based degree centrality, and CVDegreeII for Configuration-value interaction index.

Finally, it would be worthwhile to introduce a graph-related axiomatization of our similarity measure, following a similar approach to that with which the interaction indices was axiomatized based on concepts from cooperative games.

ACKNOWLEDGEMENTS

Piotr Szczepański was funded by the Polish National Science Centre based on the decision DEC-2013/09/N/ST6/04095. Tomasz Michalak and Michael Wooldridge were supported by the European Research Council under Advanced Grant 291528 (“RACE”).