ECAI 2016
G.A. Kaminka et al. (Eds.)

doi:10.3233/978-1-61499-672-9-90

An Extension of the Owen-Value Interaction Index and Its Application to Inter-Links Prediction

Piotr L. Szczepański[1] and **Tomasz P. Michalak**[2,3] and **Talal Rahwan**[4] and **Michael Wooldridge**[2]

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, 45]. 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 [43, 50] 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 [52]. While current tools focus on recommending collaborations between academics from the same field of study, such tools can benefit from identifying any promising 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 [45]. 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* [37]—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 *Coalitional Semivalues* [47]; the resulting node-similarity measure is the *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 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.

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 *centrality index* (or simply a *centrality*) measures the importance of individual nodes. One of the fundamental centrality indices is *degree centrality* [14, 13, 36], which 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 know as `k`-*steps degree centrality*. The notion of centrality can also be generalized to groups of nodes, resulting in what is known as *group centrality* [13]. One such group centrality that we will focus on in this paper is `k`-*steps group degree centrality*.

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

$$\left| \left\{ v \in V : \min_{u \in S} d(u, v) \leq k \right\} \setminus S \right| \tag{1}$$

Some authors [35, 45] interpret the above formula as a *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.

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.

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 *marginal contribution*; for every player, $i \in N$, and every coalition, $C \subseteq N$, the marginal contribution of i to C is:

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

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 more than $|N| - 1$. Then, a Semivalue is defined as follows:

Definition 2 *Given a game, (N, ν), 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:*

$$SEMI_i(N, \nu) = \sum_{0 \leq k < |N|} \beta(k) \mathbb{E}[\mathrm{MC}(X^k, i)], \tag{2}$$

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

The first Semivalue to appear in the literature was the *Shapley value* [42], 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 *Banzhaf power index* [3], which has also been studied extensively. Those two Semivalues are defined by the following β-functions:

$$\beta^{Shapley}(i) = \frac{1}{|N|} \quad \text{and} \quad \beta^{Banzhaf}(i) = \frac{\binom{|N|-1}{i}}{2^{|N|-1}}.$$

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: $\mathrm{S}(i, j) = \nu(\{i, j\}) - \nu(\{i\}) - \nu(\{j\})$. One can also measure such synergy with respect to any particular coalition, $C \subseteq N$, as follows:

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

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

By combining a Semivalue with the interaction index, we obtain a *Semivalue interaction index*, defined as follows:

Definition 3 *Given a game, (N, ν), 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:*

$$I^{SEMI}_{i,j}(N, \nu) = \sum_{0 \leq k \leq |N|-2} \beta(k) \mathbb{E}[S(X^k, i, j)], \tag{3}$$

where X^k is a coalition drawn uniformly at random from: $\{C : C \subseteq N \setminus \{i,j\} \wedge |C| = k\}$, 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.

interaction index name	$\beta(l)$
Shapley-value interaction index [37, 17]	$\frac{1}{\lvert V\rvert-1}$
Banzhaf-index interaction index [41]	$\frac{\binom{\lvert V\rvert-2}{l}}{2^{\lvert V\rvert-2}}$
Chaining interaction index [34]	$\frac{2(l+1)}{(n-1)(n-2)}$

Table 1. Values of β 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, ν). Arguably, the most established extension of the Shapley value to such games is the *Owen value* [38]. 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, ν), the corresponding quotient game, (CS, ν^Q), 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:

$$\nu^Q(R) = \nu\Big(\bigcup_{r\in R} C_r\Big) \text{ for all } R \subseteq M,$$

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

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, ν), the Owen value of a player $i \in C_x \in CS$ is:*

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

One generalization of the Owen value that was recently introduced in the literature is *Coalitional Semivalues* [47], defined as follows:

Definition 5 *Given a game, (N, CS, ν), and a discrete probability distribution, $\beta : \{0, \ldots, |M|-1\} \to [0,1], \sum_{0\le k<|M|} \beta(k) = 1$, the Coalitional Semivalue of a player $i \in C_x \in CS$ is:*

$$CSEMI_i(N, CS, \nu) = \sum_{0\le k\le |M\setminus\{x\}|} \beta(k) \sum_{0\le l<|C_x|} \alpha(l)\mathbb{E}[\text{MC}(Q_{T^k} \cup X^l, i)], \quad (5)$$

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

As shown in Table 2, by adopting the appropriate probability distributions, we obtain the Owen value [38] or any of its modifications proposed in the literature to date, namely: *Owen-Banzhaf value* [39], *symmetric coalitional Banzhaf value* [2], and *symmetric coalitional p-binomial Semivalues* [6].

Solution name	$\beta(k)$	$\alpha(l)$
Owen value [38]	$\frac{1}{\lvert M\rvert-1}$	$\frac{1}{\lvert C_x\rvert-1}$
Owen-Banzhaf value [39]	$\frac{\binom{\lvert M\rvert-1}{k}}{2^{\lvert M\rvert-1}}$	$\frac{\binom{\lvert C_x\rvert-1}{l}}{2^{\lvert C_x\rvert-1}}$
symmetric coalitional Banzhaf value [2]	$\frac{\binom{\lvert M\rvert-1}{k}}{2^{\lvert M\rvert-1}}$	$\frac{1}{\lvert C_x\rvert}$
symmetric coalitional p-binomial Semivalue [6]	$p^k(1-p)^{\lvert M\rvert-1-k}$ $p \in [0,1]$	$\frac{1}{\lvert C_x\rvert}$

Table 2. Values of α and β 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 [53].

Definition 6 *Given a cooperative game with a coalition structure, (N, CS, ν), 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:*

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

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, ν_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:

$$\nu_D^k(S) = \Big| \big\{ v \in V : \min_{u \in S} d(u, v) \leq k \big\} \setminus S \Big|.$$

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, ν)*, a discrete probability distribution,* $\beta : \{0, \ldots, |N| - 2\} \to [0, 1], \sum_{0 \leq k \leq |N|-2} \beta(k) = 1$*, and two players,* $i \in C_x \in CS$ *and* $j \in C_y \in CS$ *such that* $C_x \neq C_y$*, the Coalitional-Semivalue interaction index between* i *and* j *is:*

$$I_{i,j}^{CSEMI}(N, CS, \nu) = \sum_{0 \leq k \leq |M \setminus \{C_x, C_y\}|} \beta(k) \sum_{0 \leq l \leq |C_x \cup C_y \setminus \{i,j\}|} \alpha(l) \mathbb{E}[S(Q_{T^k} \cup X^l, i, j)], \quad (7)$$

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

This is a natural extension of Owen-value interaction index that is in line with the definition of Coalitional Semivalues. For instance, by setting $\beta(k) = \frac{1}{|M|-1}$ and $\alpha(l) = \frac{1}{|C_x|-1}$, 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* $v \in C_x$ *and* $u \in C_y$ *in graph* $G(V, E)$ *with community structure* CS *is defined as:*

$$I_{u,v}^{CSEMI}(V, CS, \nu_D^k).$$

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 allows 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-Semivalues 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 CSDEGREEII denote the problem of calculating $\mathrm{I}_{u,v}^{CSEMI}(V, CS, \nu_D^k)$, where CS is a community structure, ν_D^k is the k-steps degree centrality, and $u, v \in V$. The main theoretical result in this paper is as follows:

Theorem 1 CSDEGREEII *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 $deg_k(v)$ denote the number of such nodes. More formally, we have: $N_k(v) = \{u \in V : d(v, u) \leq k \wedge v \neq u\}$ and $deg_k(v) = |N_k(v)|$. We extend this notation to sets of nodes. That is, $N_k(C) = \bigcup_{v \in C} N_k(v) \setminus C$ and $deg_k(C) = |N_k(C)|$. Moreover, for any given node, $v \in C_x \in CS$, we denote the set of adjacent communities as $N_k^{CS}(v) = \{C_y \in CS \setminus C_x : C_y \cap N_k(v) \neq \emptyset\}$, the inter-community degree as $deg_k^{CS}(v) = |N_k^{CS}(v)|$, 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 $deg_k^y(v) = |N_k^y(v)|$. These can be extended to two communities as follows: $N_k^{y,z}(v) = N_k(v) \cap (C_y \cup C_z)$ and $deg_k^{y,z}(v) = |N_k^{y,z}(v)|$.

In our proof we follow the line of our earlier work [45], where we developed an algorithm to computed the Shapley value-based interaction index was proposed. In this work, we will extend the proof from [45] 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 $v, u \in V$ such that: $v \in C_i \in CS$ and $u \in C_j \in CS$ and $C_i \neq C_j$, we will show how to compute $\mathbb{E}[\mathrm{S}(Q_{T^k} \cup X^l, u, v)]$—the expected value of their synergy with respect to the random set $Q_{T^k} \cup X^l$. Recall that T^k is drawn uniformly from the set $\{R : R \subseteq M \setminus \{i, j\} \wedge |R| = k\}$, and X^l is drawn uniformly from the set $\{C : C \subseteq C_i \cup C_j \setminus \{u, v\} \wedge |C| = l\}$. Also recall that Q_R denotes $\bigcup_{r \in R} C_r$. Now if we denote $R^{k,l} = Q_{T^k} \cup X^l$, then:

$$\begin{aligned}
\mathbb{E}[\mathrm{S}(R^{k,l}, u, v)] &= \mathbb{E}[\mathrm{MC}(R^{k,l}, \{u, v\})] - \mathbb{E}[\mathrm{MC}(R^{k,l}, u)] - \mathbb{E}[\mathrm{MC}(R^{k,l}, v)] \\
&= \mathbb{E}[\mathrm{MC}(R^{k,l}, u)] + \mathbb{E}[\mathrm{MC}(R^{k,l}, v)] - \mathbb{E}[\mathrm{MC}(R^{k,l}, u \cap v)] \\
&\quad - \mathbb{E}[\mathrm{MC}(R^{k,l}, u)] - \mathbb{E}[\mathrm{MC}(R^{k,l}, v)] \\
&= -\mathbb{E}[\mathrm{MC}(R^{k,l}, u \cap v)],
\end{aligned}$$

where $\mathrm{MC}(R^{k,l}, u \cap v)$ is what we call the "*common*" contribution of two nodes v and u, which is illustrated in Figure 1, and defined as follows:[5]

$$\begin{aligned}
\mathbb{E}[\mathrm{MC}(R^{k,l}, u \cap v) &= \mathbb{E}[\mathrm{MC}(R^{k,l}, u)] + \mathbb{E}[\mathrm{MC}(R^{k,l}, v)] \\
&\quad - \mathbb{E}[\mathrm{MC}(R^{k,l} \cup \{u\}, v)] - \mathbb{E}[\mathrm{MC}(R^{k,l} \cup \{v\}, u)].
\end{aligned}$$

Now given the function ν_D^k, the pair of nodes $v, u \in V$ 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 $n \in N_k(v) \cap N_k(u)$. 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 $u \cap v$ when referring to the intersection between the contributions of u and v.

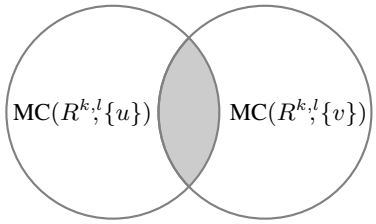

Figure 1. An illustration of MC($R^{k,l}, u \cap v$).

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 $n \in N_k(v) \cap N_k(u)$:

$$\mathbb{E}[B^{+}_{k,l,v,u,n}] = P[(N_k(n) \cup \{n\}) \cap R^{k,l} = \emptyset]. \tag{8}$$

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 $R^{k,l} \cup \{v, u\}$.[6] 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 v or u is defined as:

$$\mathbb{E}[B^{-}_{k,l,v}] = P[(N_k(v)) \cap R^{k,l} \neq \emptyset], \tag{9}$$

$$\mathbb{E}[B^{-}_{k,l,u}] = P[(N_k(u)) \cap R^{k,l} \neq \emptyset]. \tag{10}$$

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 Ω, where $|\Omega| = \binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}$. Having this in mind, we denote $P[(N_k(n) \cup \{n\}) \cap R^{k,l} = \emptyset]$ by P^+ and we get:

$$P^{+} = \begin{cases} \frac{\binom{|M|-1-deg^{CS}_k(n)}{k}\binom{|C_i|+|C_j|-deg^{i,j}_k(n)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}} & \text{if } n \notin C_i \cup C_j \\ \frac{\binom{|M|-1-deg^{CS}_k(u)}{k}\binom{|C_i|+|C_j|-1-deg^{i,j}_k(n)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}} & \text{if } n \in C_i \cup C_j \end{cases} \tag{11}$$

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 $|M| - 2 - deg^{CS}_k(n)$ 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 this into account, the final number of communities is: $|M| - 1 - deg^{CS}_k(n)$. Thus, the probability of choosing a set $R^{k,l}$ satisfying our condition $N_k(v) \cap T^k \neq \emptyset$ is exactly: $\binom{|M|-1-deg^{CS}_k(n)}{k} / \binom{|M|-2}{k}$.

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 $|C_i \cap C_j| - 2 - deg^{i,j}_k(n)$ such nodes. However, taking into account that $v, u \in N^{i,j}_k(n)$, the probability of choosing a set $R^{k,l}$ satisfying the condition $N_k(v) \cap X^l \neq \emptyset$ is exactly: $\binom{|C_i|+|C_j|-deg^{i,j}_k(n)}{l} / \binom{|C_i|+|C_j|-2}{l}$.

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:

$$P^{-} = 1 - \frac{\binom{|M|-1-deg^{CS}_k(v)}{k}\binom{|C_i|+|C_j|-1-deg^{i,j}_k(v)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}}, \tag{12}$$

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

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

$$= - \sum_{n \in (N_k(v) \cap N_k(u)) \setminus (C_i \cup C_j)} \left(\frac{\binom{|M|-1-deg^{CS}_k(n)}{k}\binom{|C_i|+|C_j|-deg^{i,j}_k(n)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}} \right)$$

$$- \sum_{n \in (N_k(v) \cap N_k(u)) \cap (C_i \cup C_j)} \left(\frac{\binom{|M|-1-deg^{CS}_k(n)}{k}\binom{|C_i|+|C_j|-1-deg^{i,j}_k(n)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}} \right)$$

$$+ \sum_{n \in (\{v\} \cap N_k(u)) \cup (\{u\} \cap N_k(v))} \left(1 - \frac{\binom{|M|-1-deg^{CS}_k(n)}{k}\binom{|C_i|+|C_j|-1-deg^{i,j}_k(n)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}} \right) \tag{13}$$

Notice that the sets $N^{CS}_k(u)$, $N^{i,j}_k(u)$ 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 u and v 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 OVDEGREEII the problem of calculating $\mathrm{I}^{OV}_{u,v}(V, CS, \nu^k_D,)$ where CS is a community structure, ν^k_D is the k-steps degree centrality, and $u, v \in V$. Then, the following corollary immediately follows from Theorem 1:

Corollary 1 OVDEGREEII *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 OVDEGREEII in just $O(|V|)$ 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 $\mathbb{E}[\mathrm{S}(R^{k,l}, u, v)]$ is computed using Equation (13).

[6] In some work the definition of $\nu^k_D(C)$ also counts the nodes from C [35]. However, in this paper we follow the convention that the influence of C affects only nodes from outside C.

Algorithm 1: Computing the k-steps Coalitional-Semivalue interaction index

Input: Graph $G(V, E)$, functions β and α, 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

$N_k^{CS}(u)$—the set of adjacent communities

$N_k^{i,j}(u)$—the set of adjacent nodes within $C_i \cup C_j$

Output: $\mathrm{I}_{u,v}^{CSEMI}$ k-steps coalitional-Semivalue interaction index

$\mathrm{I}_{u,v}^{CSEMI} \leftarrow 0$;

for $k \leftarrow 0$ **to** $|M| - 2$ **do**

  for $l \leftarrow 0$ **to** $|C_i \cup C_j| - 2$ **do**

    $\mathrm{S} \leftarrow 0$;

    foreach $n \in \big(N_k(v) \cap N_k(u)\big) \setminus \big(C_i \cup C_j\big)$ **do**

      $\mathrm{S} \leftarrow \mathrm{S} - \frac{\binom{|M|-1-deg_k^{CS}(n)}{k}\binom{|C_i|+|C_j|-deg_k^{i,j}(n)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}}$

    foreach $n \in \big(N_k(v) \cap N_k(u)\big) \cap \big(C_i \cup C_j\big)$ **do**

      $\mathrm{S} \leftarrow \mathrm{S} - \frac{\binom{|M|-1-deg_k^{CS}(n)}{k}\binom{|C_i|+|C_j|-1-deg_k^{i,j}(n)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}}$

    if $v \in N_k(u)$ **then**

      $\mathrm{S} \leftarrow \mathrm{S} + 1 - \frac{\binom{|M|-1-deg_k^{CS}(v)}{k}\binom{|C_i|+|C_j|-1-deg_k^{i,j}(v)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}}$

      $\mathrm{S} \leftarrow \mathrm{S} + 1 - \frac{\binom{|M|-1-deg_k^{CS}(u)}{k}\binom{|C_i|+|C_j|-1-deg_k^{i,j}(u)}{l}}{\binom{|M|-2}{k}\binom{|C_i|+|C_j|-2}{l}}$

    $\mathrm{I}_{u,v}^{CSEMI} \leftarrow \mathrm{I}_{u,v}^{CSEMI} + \beta(k)\alpha(l)S$

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 $N_k^{CS}(u)$, $N_k^{i,j}(u)$ 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 $N_k^{CS}(u)$ 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 and C_j we can compute the set $N_k^{i,j}(n)$ in $O(|V|^2)$ 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 $O(|V|)$ time, and requires the aforementioned preprocessing stage to compute the sets $N_k^{CS}(u)$, $N_k^{i,j}(u)$ 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 Emiprical 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

Algorithm 2: Computing the k-steps Owen-value interaction index

Input: Graph $G(V, E)$, 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

$N_k^{CS}(u)$—the set of adjacent communities

$N_k^{i,j}(u)$—the set of adjacent nodes within $C_i \cup C_j$

Output: $\mathrm{I}_{u,v}^{OV}$ k-steps coalitional-Semivalue interaction index

$\mathrm{I}_{u,v}^{OV} \leftarrow 0$;

foreach $n \in \big(N_k(v) \cap N_k(u)\big) \setminus \big(C_i \cup C_j\big)$ **do**

  $\mathrm{I}_{u,v}^{OV} \leftarrow \mathrm{I}_{u,v}^{OV} - \frac{1}{(deg_k^{CS}(n))(deg_k^{i,j}(n))}$

foreach $n \in \big(N_k(v) \cap N_k(u)\big) \cap \big(C_i \cup C_j\big)$ **do**

  $\mathrm{I}_{u,v}^{OV} \leftarrow \mathrm{I}_{u,v}^{OV} - \frac{1}{(deg_k^{CS}(n))(deg_k^{i,j}(n)-1)}$

if $v \in N_k(u)$ **then**

  $\mathrm{I}_{u,v}^{OV} \leftarrow \mathrm{I}_{u,v}^{OV} + 1 - \frac{1}{(deg_k^{CS}(v))(deg_k^{i,j}(v)-1)}$

  $\mathrm{I}_{u,v}^{OV} \leftarrow \mathrm{I}_{u,v}^{OV} + 1 - \frac{1}{(deg_k^{CS}(u))(deg_k^{i,j}(u)-1)}$

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, 45].

Similarity Name	Measure
Common Neighbors **(CN)**	$S_{u,v}^{CN} = N_k(u) \cap N_k(v)$
Salton Index **(SI)**	$S_{u,v}^{SI} = \frac{N_k(u) \cap N_k(v)}{\sqrt{N_k(u) \times N_k(v)}}$
Jaccard Index **(JI)**	$S_{u,v}^{JI} = \frac{N_k(u) \cap N_k(v)}{N_k(u) \cup N_k(v)}$
Adamic-Adar Index **(AA)**	$S_{u,v}^{AA} = \sum_{n \in N_k(u) \cap N_k(v)} \frac{1}{\log N_k(n)}$
Resource Allocation **(RA)**	$S_{u,v}^{RA} = \sum_{n \in N_k(u) \cap N_k(v)} \frac{1}{N_k(n)}$
Shapley-value interaction index **(SV)**	$S_{u,v}^{SV} = \mathrm{I}_{i,j}^{SV}(V, \nu_D^k)$

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, $v, u \in V : (v, u) \notin V$, belonging to two different communities, $C_i, C_j \in CS$ 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.[7] Then, we compute the similarity of each disconnected pair of nodes from different communities

[7] 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 $\mathbf{n} = |M||N|$ be the number of all comparisons between the missing links and the non-existing links. Furthermore, let $\mathbf{n}'$ be the number of such comparisons in which the missing link is ranked higher than the non-existing link. Finally, let $\mathbf{n}''$ be the number of such comparisons in which both links are ranked the same. Then, the AUC is computed as follows:

$$AUC = \frac{\mathbf{n}' + \frac{\mathbf{n}''}{2}}{\mathbf{n}}.$$

With this forumla, if AUC equals 1 then all missing links are ranked higher than the non-existing links; this is the best possible ranking, where $\mathbf{n}' = \mathbf{n}$ and $\mathbf{n}'' = 0$. 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 $\mathbf{n}' = 0$ and $\mathbf{n}'' = 0$. 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 [54], 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.[8]. For each of them, we report the results for the community structure identified by the multilevel community-detection algorithm [51]. We also experimented with other community-detection algorithms, such as *Walktrap* [40], *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 the other local link-prediction measures from Table 3. As can be seen, in the experiments where 10% of inter-links were removed (i.e., Table 5), our measure outperforms all the other measures. As for the experiments where 20% of inter-links were removed (i.e., Table 6), our measure also outperforms the other alternatives for all networks except for the *Football* network.

The results in Tables 5 and 6 demonstrate how detecting inter-links locally can be a rather challenging task. For instance, given the *Football* network, all local node-similarity measures are biased; they produce results that are worse than even a completely random classifier whose AUC is expected to be 0.5. Likewise, given the *Taro* network, all measures are either worse, or slightly better than, the random one. Nevertheless, for the remaining networks, all measures outperform the random one with very few exceptions.

[8] The datasets for the eight networks were downloaded from the link: `http://www-personal.umich.edu/~mejn/netdata/` as well as the link: `http://konect.uni-koblenz.de/networks/`.

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

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.

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

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.

Finally, we show in Figure 2 how the runtime of our Algorithm 2 grows with the size of the network. Specifically, the dotted line represents the time needed for the preprocessing stage, whereas the solid line represents the time needed for computing the 1-step Owen Value-based interaction index between each pair of nodes in the network. As can be seen, even for 5000 nodes, our algorithm takes less than one minute.

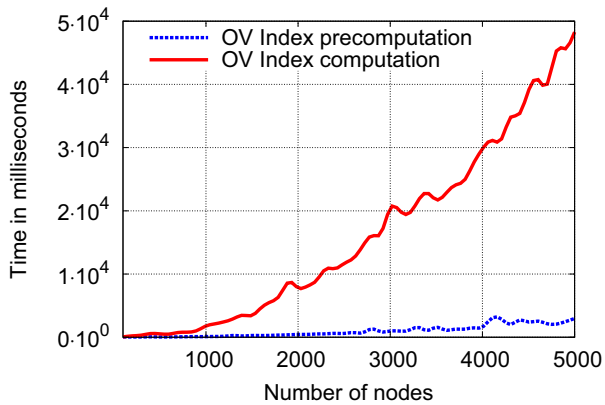

Figure 2. The performance of 1-step Owen value interaction index.

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 $deg_k^{CS}(n)$. 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.

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* centrailty, 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, 46]. Moreover, it was proven that the problem of computing any Semivalues—parametrized by any polynomial time computable discrete probability distribution—belongs to the class P for degree and closeness centralities [44], as well as betweenness centrality [46].

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 [47, 48]. However, it is still an open question whether the same holds of r *betweenness*.

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 [45].

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.

Computational result	Algorithm complexity
SVDEGREE is in P	$O(\|V\| + \|E\|)$ [35]
SDEGREE is in P	$O(\|V\|^2)$ [44]
OVDEGREE is in P	$O(\|V\| + \|E\|)$ [47]
CSDEGREE is in P	$O(\|V\|^3)\|$ [47]
SVDEGREEII is in P	$O(\|V\|)^*$ [45]
SDEGREEII is in P	$O(\|V\|^2)$ [45]
OVDEGREEII is in P	$O(\|V\|)^*$ **[this paper]**
CSDEGREEII is in P	$O(\|V\|^3)$ **[this paper]**

$(*)$ 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 distinguish between three groups of link prediction algorithm: *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* [49], 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 *Superposed 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 [55], which is inspired by the resource allocation dynamics on complex networks; and (ii) the *Shapley-value interaction index* [45] 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 [43, 50]. 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 $O(|V|)$ time, after a preprocessing stage that requires $O(|V||E|)$ 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 OVBETWEENNESSII 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 [48]. 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”).

REFERENCES

[1] L. A. Adamic and N. Glance, 'The political blogosphere and the 2004 u.s. election: Divided they blog', in *Proceedings of the 3rd International Workshop on Link Discovery*, pp. 36–43, (2005).

[2] R. Amer, F. Carreras, and J. M. Giménez, 'The modified banzhaf value for games with coalition structure: an axiomatic characterization.', *Mathematical Social Sciences*, **43**(1), 45–54, (2002).

[3] J. F. Banzhaf, 'Weighted Voting Doesn't Work: A Mathematical Analysis', *Rutgers Law Review*, **19**, 317–343, (1965).

[4] S. Brin and L. Page, 'The anatomy of a large-scale hypertextual web search engine', *Comput. Netw. ISDN Syst.*, **30**(1–7), 107–117, (1998).

[5] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi, 'Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding.', *Bioinformatics*, **29**(13), 199–209, (2013).

[6] F. Carreras and M. Puente, 'Symmetric coalitional binomial semivalues', *Group Decision and Negotiation*, **21**(5), 637–662, (2012).

[7] G. Chalkiadakis, E. Elkind, and M. Wooldridge, *Computational aspects of cooperative game theory*, Morgan & Claypool Publishers, 2011.

[8] A. Clauset, C. Moore, and M. E.J. Newman, 'Hierarchical structure and the prediction of missing links in networks', *Nature*, **453**(7191), 98–101, (2008).

[9] M. E. J. Clauset, A.and Newman and C. Moore, 'Finding community structure in very large networks', *Phys. Rev. E*, **70**, 066111, (2004).

[10] S.F. Crone and D. Soopramanien , 'Predicting customer online shopping adoption-an evaluation of data mining and market modelling approaches.', in *Proceedings of the 2005 International Conference on Data Mining*, pp. 215–221, (2005).

[11] P. Dubey, A. Neyman, and R. J. Weber, ' Value Theory Without Efficiency.', *Mathematics of Operations Research*, **6**, 122–128, (1981).

[12] Kenneth E., 'Cultures of the central highlands.', *Southwestern J. of Anthropology*, **10**(1), 1–43, (1954).

[13] M. G. Everett and S. P. Borgatti, 'The centrality of groups and classes', *Journal of Mathematical Sociology*, **23**(3), 181–201, (1999).

[14] L.C. Freeman, 'Centrality in social networks: Conceptual clarification', *Social Networks*, **1**(3), 215–239, (1979).

[15] M. Girvan and M. E. J. Newman, 'Community structure in social and biological networks', *in Proc. of NAS*, **99**(12), 7821–7826, (2002).

[16] P. Gleiser and L. Danon, 'Community structure in jazz', *Advances in Complex Systems*, **6**, 565, (2003).

[17] M. Grabisch, 'k-order additive discrete fuzzy measures and their representation', *Fuzzy Sets and Systems*, **92**(2), 167 – 189, (1997).

[18] M. Grabisch and M. Roubens, 'An axiomatic approach to the concept of interaction among players in cooperative games', *International Journal of Game Theory*, **28**, 547–565, (1999).

[19] M. Grabisch and M. Roubens, *Probabilistic interactions among players of a cooperative game*, 2000.

[20] R. Guimerà and M. Sales-Pardo, 'Missing and spurious interactions and the reconstruction of complex networks', *Proceedings of the National Academy of Sciences*, **106**(52), 22073–22078, (2009).

[21] P. Hage and F. Harary, *Structural Models in Anthropology.*, Cambridge University Press.

[22] J.A. Hanley and B.J. McNeil, 'The meaning and use of the area under a receiver operating characteristic (roc) curve.', *Radiology*, **143**(1), 29–36, (1982).

[23] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, 'Link prediction using supervised learning', in *In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security*, (2006).

[24] B. Hayes, 'Connecting the dots: Can the tools of graph theory and social-network studies unravel the next big plot?', *American Scientist*, **5**(94), 400–404, (2006).

[25] D. Heckerman, D. Geiger, and D.M. Chickering, 'Learning bayesian networks: The combination of knowledge and statistical data', *Machine Learning*, **20**, 197–243, (1995).

[26] D. Heckerman, C. Meek, and D. Koller, 'Probabilistic Models for Relational Data', Technical report, Microsoft Research, (2004).

[27] A. Lancichinetti and S. Fortunato, 'Community detection algorithms: a comparative analysis', *Physical review E*, **80**(5), 056117, (2009).

[28] D. Liben-Nowell and J. Kleinberg, 'The link-prediction problem for social networks', *J. Am. Soc. Inf. Sci. Technol.*, **58**(7), 1019–1031, (2007).

[29] G. Linden, B. Smith, and J. York, 'Amazon.com recommendations: Item-to-item collaborative filtering', *IEEE Internet Computing*, **7**(1), 76–80, (2003).

[30] F.C. Linton, F.C. Sue, , and G. M. Alaina, 'On human social intelligence', *J. of Social and Biological Structures*, **4**(11), 415–425, (1988).

[31] W. Liu and L. Lü, 'Link prediction based on local random walk', *EPL (Europhysics Letters)*, **89**(5), 58007, (2010).

[32] L. Lü and T. Zhou, 'Link prediction in complex networks: A survey', *Physica A*, **390**(6), 11501170, (2011).

[33] J.-L. Marichal and P. Mathonet, 'Approximations of lovsz extensions and their induced interaction index.', *DAM*, **156**(1), 11–24, (2008).

[34] J. L. Marichal and M. Roubens, 'The chaining interaction index among players in cooperative games', in *Adv. in Decision Analysis*, volume 4 of *Math.Mod.-TA*, 69–85, (1999).

[35] T. P. Michalak, K. V. Aadithya, P. L. Szczepański, B. Ravindran, and N. R. Jennings, 'Efficient computation of the shapley value for game-theoretic network centrality.', *J. Artif. Intell. Res.*, **46**, 607–650, (2013).

[36] M. E. J. Newman, 'Analysis of weighted networks', *Phys. Rev. E*, **70**(5), 056131, (2004).

[37] G. Owen, 'Multilinear extensions of games', *Management Science*, **18**(5-part-2), 64–79, (1972).

[38] G. Owen, 'Values of games with a priori unions', in *Mathematical economics and game theory: Essays in honor of Oskar Morgenstern*, volume 141 of *LNEMS*, 76–88, Berlin, (1977).

[39] G. Owen, 'Modification of the Banzhaf-Coleman index for games with a priori unions', in *Power, Voting and Voting Power*, 232–238, (1982).

[40] Latapy M. Pons, P., 'Computing communities in large networks using random walks.', *Journal of Graph Algorithms and Applications*, **10**(2), 191–218, (2006).

[41] M. Roubens, 'Interaction between criteria and definition of weights in mcdavproblems', in *44th meet. of the Eur. working gr. MCDA*, (1996).

[42] L. S. Shapley, 'A value for n-person games', in *In Contributions to the Theory of Games, volume II*, 307–317, PUP, (1953).

[43] S. Soundarajan and J. Hopcroft, 'Using community information to improve the precision of link prediction methods', in *PWWW '12*, pp. 607–608, (2012).

[44] P. L. Szczepański, M. K. Tarkowski, T. P. Michalak, P. Harrenstein, and M. Wooldridge, 'Efficient computation of semivalues for game-theoretic network centrality', in *AAAI'15*, pp. 461–469, (2015).

[45] P.L. Szczepański, A.S. Barcz, T.P. Michalak, and T. Rahwan, 'The game-theoretic interaction index on social networks with applications to link prediction and community detection', in *IJCAI'15*, pp. 638–644, (2015).

[46] P.L. Szczepański, T.P. Michalak, and T. Rahwan, 'Efficient algorithms for a game-theoretic betweenness centrality', *Artificial Intelligence*, **2016**, 39–63, (2016).

[47] P.L. Szczepański, T.P. Michalak, and M. Wooldridge, 'A centrality measure for networks with community structure based on a generalization of the owen value', in *ECAI'14*, pp. 867–872, (2014).

[48] M. Tarkowski, P.L. Szczepański, T. Rahwan, T.P. Michalak, and M. Wooldridge, 'Closeness centrality for networks with overlapping community structure', *AAAI '16: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, forthcoming, (2016).

[49] H. Tong, C. Faloutsos, and J.-Y. Pan, 'Fast random walk with restart and its applications', in *Proceedings of the Sixth International Conference on Data Mining*, ICDM '06, pp. 613–622, (2006).

[50] J. Valverde-Rebaza and A de Andrade Lopes, 'Structural link prediction using community information on twitter', in *Computational aspects of social networks (CASoN), 2012 Fourth International Conference on*, pp. 132–137. IEEE, (2012).

[51] Blondel V.D., Guillaume J.-L., Lambiotte R., and Lefebvre E., 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*, **2008**(10), P10008, (2008).

[52] X. Wang and G. Sukthankar, *Social Network Analysis - Community Detection and Evolution*, chapter Link Prediction in Heterogeneous Collaboration Networks, 165–192, 2014.

[53] L. Xiaoning, L. Shujin, and L. Feng, 'The influence between two players in game with graph restricted communication and a priori unions', in *Control and Decision Conference*, pp. 55–59, (2011).

[54] W.W. Zachary, 'An information flow model for conflict and fission in small groups', *J. of Anthropological Research*, **33**, 452–473, (1977).

[55] T. Zhou, L. Lü, and Y.C. Zhang, 'Predicting missing links via local information', *The European Physical Journal B-Condensed Matter and Complex Systems*, **71**(4), 623–630, (2009).