ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-651

A Shapley Value-based Approach to Determine Gatekeepers in Social Networks with Applications

Ramasuri Narayanam[1] and **Oskar Skibski**[2] and **Hemank Lamba**[3] and **Tomasz Michalak**[4]

Abstract. Inspired by emerging applications of social networks, we introduce in this paper a new centrality measure termed *gatekeeper centrality*. The new centrality is based on the well-known game-theoretic concept of Shapley value and, as we demonstrate, possesses unique qualities compared to the existing metrics. Furthermore, we present a dedicated approximate algorithm, based on the Monte Carlo sampling method, to compute the gatekeeper centrality. We also consider two well known applications in social network analysis, namely community detection and limiting the spread of mis-information; and show the merit of using the proposed framework to solve these two problems in comparison with the respective benchmark algorithms.

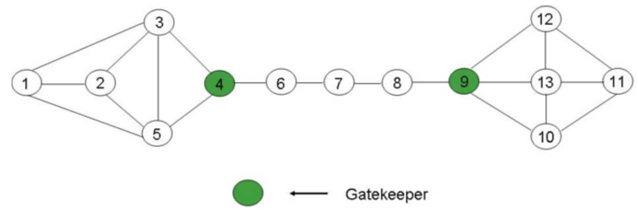

Figure 1. A stylized network

1 Introduction

Social networks are prevalent in several real world scenarios such as online social networks (e.g., Facebook or Flickr), collaboration networks, email networks, trading networks, and R & D networks [8, 3]. Social networks are social structures made up of individuals (or autonomous entities) and connections among these individuals. In the literature, such networks are conveniently represented using graphs, where nodes represent entities in the networks and edges represent the connections among these entities. A significant amount of work on social network analysis in the literature is devoted to understanding the role played by nodes/edges, with respect to either their structural placement in the network or their behavioral influence over others in the network. To this end, it is important to rank nodes/edges in a given network based on either their positional power or their behavioral influence. There exist several well known ranking mechanisms in the literature, ranging from the well known centrality measures from social sciences such as degree centrality, closeness centrality, clustering coefficient, and betweenness centrality [8, 3] to Google PageRank [5].

However, the existing centrality measures in the literature are often inadequate to satisfactorily serve the needs of emerging real-life applications. Let us consider one such scenario as follows: *We want to determine a group of nodes of specific cardinality such that these nodes can disconnect the given network into connected components having cardinalities as close as possible*. Such scenarios arise in applications like community detection in networks and limiting the spread of misinformation over social networks. For instance, consider a stylized graph of a social network as shown in Figure 1 and we want to find a group of two nodes; then choosing node 4 and node 9 is the best solution as the resulting components have cardinalities 4, 3, and 4 respectively. On the other hand, we can also rank nodes in this network using any well-known centrality metric and take the top 2 nodes to address the problem. Table 1 lists the top two nodes using degree, closeness, betweenness, clustering coefficient (CC), eigenvector, and PageRank centrality measures. Strikingly, none of these well-known centrality measures pick node 4 and node 9 as the solution for the problem. In this paper, we refer to such nodes (that is, node 4 and node 9 in Figure 1) as *gatekeeper* nodes. We wish to propose a centrality measure that ranks the nodes based on their ability of being gatekeepers and we refer to such a measure as *gatekeeper centrality*. To the best of our knowledge, none of the existing centrality measures in the literature are adequate to identify the gatekeeper nodes in a given social network.

Centrality Measure	**Rank** 1	**Rank** 2
Degree	9	3,5,10,11,12,13
Closeness	7	6,8
Betweenness	7	6,8
CC	10,11,12,13	9
EigenVector	1,2,10,11,12,13	3,5
PageRank	9	3,5

Table 1. Centrality measures for the nodes in Figure 1

In this paper, we present an efficient algorithm to determine the gatekeeper centrality of a given network. We believe that this new notion of centrality can address certain social network analysis tasks in an advantageous way as compared to that of other algorithms available in the literature. We demonstrate it using the following two social network analysis tasks:

[1] IBM Research, India. Email: ramasurn@in.ibm.com
[2] University of Warsaw, Poland. Email: oskar.skibski@mimuw.edu.pl
[3] IBM Research, India. Email: helamba1@in.ibm.com
[4] University of Oxford, UK, and University of Warsaw, Poland. Email: tomasz.michalak@cs.ox.ac.uk

Community Detection: Finding communities in a given network has found significant attention from the network science research community [13]. There exists several variants of this community detection problem; of these, determining communities (or clusters) of similar size is very important as it finds applications in job scheduling over computing resources and VLSI circuit design. There exist in the literature well known measures to determine the quality of the communities—notably, modularity [13] and coverage [3]. In this paper, we apply the framework of gatekeeper centrality to determine communities in networks and we observe that the modularity and coverage of the communities found using our approach either outperforms or is close to that of certain benchmark community-detection algorithms.

Limiting the Speed of Misinformation Over Social Networks: Given a constant k, consider the problem of which k nodes should be removed so that the speed of misinformation over a network is minimized. This scenario is important in the context of controlling the spread of a virus in a network and limiting the spread of negative opinions about a product in a social network. The goal is to remove some nodes in the network such that the virus/misinformation is not allowed to spread rapidly to the entire network. As an application of the proposed gatekeeper centrality to this context, we consider the top-k nodes with high gatekeeper centrality and this choice turns out to be very attractive in comparison with the benchmark algorithms.

1.1 Why a Game-Theoretic Approach is Essential

The common feature of all the standard centrality measures in the literature [4, 8] is that they evaluate the importance of a node (or an edge) by focusing only on the role played within the network by this node by itself. However, such an approach does not take into account the fact that in many cases there occur joint effects if the functioning of nodes is considered together in groups. For instance, removing any single node might not be enough to stop communication between two communities, but removing a group of nodes could accomplish this goal. Standard centrality metrics completely ignore such synergies. To address this issue the notion of *group centrality* was developed [9]—it works in a virtually the same way as standard centrality measures but it evaluates the functioning of a given group of nodes, rather than individual nodes.

The concept of group centrality allows only for computing the centrality of a group of nodes. However, even if one computes all such groups, it is not clear how to construct a consistent ranking of individual nodes using the group results. Some nodes can play a decisive role in certain groups and can be completely irrelevant in others.

Fortunately, the issue of assessing individual entities given their participation in various groups has been extensively studied in the field of coalitional game theory. In particular, given a game where coalitions are allowed to form, one of the fundamental questions is how to distribute the surplus achieved by cooperation among the participating players. To answer this question, Shapley [28] proposed to remunerate players with payoffs that are a function of their individual marginal contributions to the game. For a given player, an individual marginal contribution is measured as the weighted average marginal increase in the payoff of any coalition caused by the entrance of the player. Importantly, the division scheme proposed by Shapley, called the *Shapley value*, is the only scheme that meets certain desirable normative properties (see Section 3 for more details). Given this, the fundamental idea of the game theoretic approach to centrality[5] is to define an appropriate cooperative game over the social network in which players are the nodes, coalitions are the groups of nodes, and payoffs of coalitions are defined so as to meet our requirements.

[5] See the next section for an overview of the literature on this topic.

In the context of the gatekeeper centrality as well, it is required to first assess an appropriate score gained by each possible group of nodes based on their ability to disconnect the network into components of similar size and then we have to derive a fair ranking of the individual nodes based on these group scores. We propose to use the Shapley value approach to perform this task. Towards this end, we have to first define an appropriate cooperative game that captures the notion of gatekeeper centrality and then we need to compute the Shapley value of the nodes of this cooperative game to generate a consistent ranking of these individual nodes.

1.2 Our Contributions

In this paper, we formally model the notion of gatekeeper centrality and present an algorithm to approximate it. In particular:

- We propose an appropriate cooperative game that formally models the notion of gatekeeper centrality.
- We propose to work with the Shapley value of the proposed cooperative game. And, as we show, the nodes with high Shapley values are the nodes with high ability of being gatekeepers.
- We present an efficient algorithm to approximate the Shapley values of the nodes.
- We finally apply the proposed notion of gatekeeper centrality to solve two popular social network analysis tasks, namely community detection and limiting the spread of misinformation. It turns out that our proposed approach solves these two tasks in an impressive manner as compared to the respective benchmark algorithms.

Organization of the Paper: We present the relevant work in Section 2 and preliminary definitions and notation in Section 3. We then formally present our model in Section 4. We next present the algorithm to compute the gatekeeper centrality in Section 5. In Section 6, we conduct thorough experimentation of our proposed approach. We conclude the paper in Section 7 by pointing out a few important directions to future work.

2 Relevant Work

The fundamental notion of centrality in networks [11] determines the relative importance of nodes in the network, for instance how influential an individual is within a social network. Several classical measures of centrality in networks have been proposed in the literature [4] such as degree centrality, closeness centrality, and betweenness centrality. Further, game theoretic approaches have been employed either to offer new centrality measures or to enrich the existing well-known centrality measures to complement the literature on the theory of centrality in networks [16, 33, 31, 32, 14, 30]. For instance, Grofman and Owen [16] were the first to present a game theoretic centrality measure to offer a new definition of degree centrality. Szczepanski *et al.* [30] proposed the Shapley-value-based betweenness centrality measure to enrich the classical betweenness centrality [4]. Brink *et al.* [32] presented a Shapley value-based approach to define a new network centrality metric, namely β-measure. Gomez *et al.* [15] proposed a new Shapley value-based network centrality measure for the class of graph-restricted games [22] (where each feasible coalition is induced by a subgraph of the given graph).

Game theoretic approaches have also been used to work with central (or influential) nodes in the network in order to solve certain important problems associated with social network analytics. For instance, Hendrickx *et al.* [17] proposed a Shapley value-based approach to identify key nodes to optimally allocate resources over the network. Alon *et al.* [1] proposed a game theoretic approach to determine k most popular or trusted users in the context of directed social networks. Ramasuri and Narahari [26] proposed a Shapley value-based approach to measure the influential capabilities of individual nodes in the context of viral marketing and then presented Monte Carlo simulation based heuristic to determine top k influential nodes for effective viral marketing over social networks. There also exists work in the literature that tackled the issue of computing the Shapley value-based centrality [30, 20, 21, 29].

3 Preliminary Definitions and Notation

Let $G = (N, E)$ be a directed and unweighted graph that models the given social network where N is the set of nodes corresponding to the individuals in the social network and E is the set of edges that captures the connections between the individuals in the social network. A path in G is an alternating sequence of nodes and edges, beginning at a node and ending at another node, and which does not visit any node more than once. Consider a graph $H = (A, B)$ where A is the set of nodes and B is the set of edges among nodes in A. We call that H is a subgraph of G if $A \subseteq N$ and $B \subseteq E$. Note that a connected component in G is a subgraph in which any two vertices are connected to each other by paths. Consider any subset $S \subseteq N$. Let $G(N \setminus S, E(N \setminus S))$ be the graph that is obtained by removing all nodes in S and all the edges incident to the nodes in S from G. Also let $\Phi(S)$ be the set of connected components in $G(N \setminus S, E(N \setminus S))$.

Let us now formalize the notions of a coalitional game and the Shapley value. To this end, we denote by $A = \{a_1, \ldots, a_{|A|}\}$ the set of players of a coalitional game. A *characteristic function* $v :\to \mathbb{R}$ assigns to every coalition $C \subseteq A$ a real number representing payoff attainable by this coalition. By convention, it is assumed that $v(\emptyset) = 0$. A *characteristic function game* is then a tuple (A, v).

It is usually assumed that the grand coalition, i.e., the coalition of all the agents in the game, forms. Given this, one of the fundamental questions of coalitional game theory is how to distribute the payoff of the grand coalition among the players. Among many different answers, Shapley [27] proposed to evaluate the role of each player in the game by considering his marginal contributions to all coalitions this player could possibly belong to. A certain weighted sum of such marginal contributions constitutes a player's payoff from the coalition game and is called the Shapley value. Importantly, Shapley proved that his payoff division scheme is the only one that meets, at the same time, the following four desirable criteria:

(i) *efficiency* — all the payoff of the grand coalition is distributed among players;
(ii) *symmetry* — if two agents play the same role in any coalition they belong to (i.e. they are symmetric) then their payoff should also be symmetric;
(iii) *null player* — agents with no marginal contributions to any coalitions whatsoever should receive no payoff from the grand coalition; and
(iv) *additivity* — values of two uncorrelated games sum up to the value computed for the sum of both games.

Formally, let $\pi \in \Pi(A)$ denote a permutation of agents in A, and let $C_\pi(i)$ denote the coalition made of all predecessors of agent a_i in π (if we denote by $\pi(j)$ the location of a_j in π, then: $C_\pi(i) = \{a_j \in \pi : \pi(j) < \pi(i)\}$). Then the Shapley value is defined as follows [27]:

$$SV_i(v) = \frac{1}{|A|!} \sum_{\pi \in \Pi} [\nu(C_\pi(i) \cup \{a_i\}) - \nu(C_\pi(i))], \tag{1}$$

i.e., the payoff assigned to a_i in a coalitional game is the average marginal contribution of a_i to coalition $C_\pi(i)$ over all $\pi \in \Pi$. It is easy to show that the above formula can be rewritten as:

$$SV_i(v) = \sum_{C \subseteq A \setminus \{a_i\}} \frac{|C|!(|A| - |C| - 1)!}{|A|!} [\nu(C \cup \{a_i\}) - \nu(C)]. \tag{2}$$

In our context, we will define a coalitional game over a network G. In this game the players will be nodes in the network, i.e., $A = V(G)$ and a characteristic function v will depend in a certain way on G. Thus the coalitional game will be formally tuple $\langle V(G), v \rangle$.

4 The Proposed Game Theoretic Model

In this section, we present the coalitional game that is the cornerstone of the gatekeeper centrality. Note that the nodes in N is the set of players in the coalitional game. In what follows, we define two variants of the characteristic function. The intuition behind the characteristic function is as follows. Consider any group of nodes, call it $S \subseteq N$. The more close is the sizes of the connected components of the graph after removing the nodes in S, the higher the value of S should be. This objective is accomplished by defining the characteristic function as a function of inverse of the cardinalities of these connected components.

- Version 1: We define the first variant of the characteristic function $v_1(.)$ as follows: $\forall S \subseteq N$,

$$v_1(S) = \frac{1}{\sum_{i \in \Phi(S)} |C_i|^2}, \tag{3}$$

  where $\Phi(S) = \{1, 2, \ldots, t\}$ is the set of indices for the t connected components (i.e. $C_1, C_2, \ldots, C_t$) in $G(N \setminus S, E(N \setminus S))$.
- Version 2: For each $S \subseteq N$, we define the second variant of the characteristic function $v_2(.)$ as follows: $\forall S \subseteq N$,

$$v_2(S) = \frac{t}{|C_1| + |C_2| + \ldots + |C_t|} \tag{4}$$

  where $\Phi(S) = \{1, 2, \ldots, t\}$ is the set of indices for the t connected components (i.e. $C_1, C_2, \ldots, C_t$) in $G(N \setminus S, E(N \setminus S))$.

We now consider the following example to illustrate the two different versions of the characteristic functions defined above.

Example 1 *Let $N = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and consider the graph G as shown in Figure 2. Let $S = \{3, 4\}$. By removing the nodes and the edges incident to the nodes in S from G, we get 4 connected components as shown in Figure 2(ii). That is $\Phi(S) = \{C_1, C_2, C_3, C_4\}$ where $C_1 = \{1, 2\}$, $C_2 = \{5, 6\}$, $C_3 = \{7, 8, 9\}$, and $C_4 = \{10\}$.*

Note that $|C_1| = 2$, $|C_2| = 2$, $|C_3| = 3$, and $|C_4| = 1$. The two different versions of the characteristic function are as follows:

- $v_1(S) = \frac{1}{4+4+9+1} = \frac{1}{18}$, *and*
- $v_2(S) = \frac{4}{2+2+1+3} = \frac{1}{2}$.

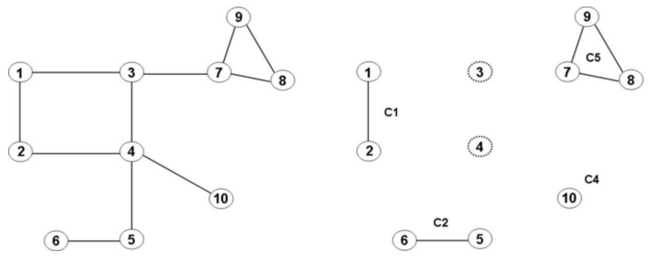

Figure 2. A stylized example

Algorithm 1: Approximation algorithm for the New Centrality Measures

Input: Graph $G = (N, E)$ and a function $v : 2^N \to \mathbb{R}$
Output: Shapley value, Sh_i, of each node $i \in N$

1 **for all** $i \in N$ **do** $Sh_i \leftarrow 0$;
2 **for** $k = 1$ **to** $numberOfSamples$ **do**
3     $\pi \leftarrow$ *random permutation of* N;
4     $SGC.createEmpty()$;
5     $valueOfSGC \leftarrow v(SGC)$;
6   **foreach** $i \in \pi$ **do**
7         $Sh_i = Sh_i + valueOfSGC$;
8         $SGC.addNode(i)$;
9   **foreach** $j \in neighbours(i)$ **do**
10   **if** $SGC.exist(j)$ **then**
11                 $SGC.addEdge(i, j)$;
12         $valueOfSGC \leftarrow v(SGC)$;
13         $Sh_i = Sh_i - valueOfSGC$;
14 **for all** $i \in N$ **do** $Sh_i \leftarrow Sh_i / numberOfSamples$;

5 Approximate Algorithm for Computing Gatekeeper Centrality

A few methods for approximating the Shapley value for various types of coalitional games have been recently discussed in the literature. [10] studied weighted voting games and [2] focused on a broader class of simple coalitional games in which the characteristic function is binary. A method to approximate the Shapley value in any game in characteristic function form was studied by Castro *et al.* [6]. In this method, based on Monte Carlo sampling, a chosen permutations of the set of all players are generated iteratively. For any given permutation, the marginal contribution of each player to the coalition made of all his predecessors is computed. The approximated Shapley value is then given by the average marginal contribution to all sampled permutations. Castro *et al.* showed that the precision of the solution increases (statistically) with every new permutation analysed.

With a growing number of agents in the game, computing any reasonable approximation of the Shapley value may require sampling millions of permutations. Consequently, the time efficiency of Monte Carlo approach hinges upon the way in which $|N|$ marginal contributions are calculated in every permutation.

Given this, in this section we present a Monte Carlo algorithm that is dedicated to dealing efficiently with gatekeeper games. Let us start our analysis with the following fundamental observation: in gatekeeper games, the value of the coalition S depends only on the structure formed by the outside players $N \setminus S$. Thus, we can traverse the permutation backward and, as we sequentially add players, assign the changes in the value of $N \setminus S$ (i.e., players' marginal contributions) to adequate players.

For this purpose, we propose a dedicated structure to store subgraph components (*SGC*) based on the idea of *FindUnion*, a disjoint-set data structure [12]. The main concept here is to store separate components of the graph as trees. Whenever we add a new edge between different components, we attach the root of one tree as a child of the second one. It is important that we do not store all graph edges, but maintain multiple statistics that allow us to calculate the value of the subgraph without traversing the whole structure.

SGC-structure allows for the following operations:

- *createEmpty()* - initializes the structure;
- *addNode(i)* - adds a new component ($parent[i] = i$) and updates statistics;
- *addEdge(i, j)* - finds the roots of the components of i and j (with *path compression*[6]); if roots differ, attach a root of the smaller tree to the bigger one (if $rank[i] < rank[j]$ then $parent[i] = j$; this technique is called *union rank*) and updates statistics; otherwise, only updates statistics if needed;
- *exist(i)* - return *true* if $parent[i]$ is set. Return *false* otherwise

This representation, based on the two improving techniques *union by rank* and the *path compression*, allows us to perform $|E|$ *addEdge()* and $|N|$ *addNode()* operations in time $O(|E| \cdot log^*(|N|))$ where $log^*(x)$ denotes the iterated logarithm and $log^*(x) \leq 5$ for $x \leq 2^{65536}$ [18].

Finally, let us address the statistics that we have to collect in order to calculate the value of the structure. To compute $v_2(S)$ we need to store the number of nodes (variable increased in *addNode()*) and number of components (variable increased in *addNode()* and decreased in *addEdge()* if edge links different components). The formula for $v_1(S)$ is based on the sum of squares of components' sizes (to this end, we store the size with every component, initialize it in *addNode()* and update it in *addEdge()*; in addition, we store the global sum of squares in $O(1)$ and update it whenever the size of a component changes).

The pseudocode is presented in Algorithm 1. In our procedure we aggregate agents' marginal contributions in variables Sh_i, initialized to zero (line 1) and divided at the end by the number of samples considered (line 14). In the main loop (lines 2-13) after the initialization we traverse the random permutation π (lines 6-13) and sequentially add nodes and edges to the SGC-structure (lines 8-11). Based on the value of the structure before and after the addition of a given agent, we calculate its marginal contribution (line 7 and 13).

The time complexity of the algorithm depends on the number of samples chosen to calculate the Shapley value (and that depends on our target precision). Let us then comment on the complexity of single sample, i.e. the calculations needed to update Shapley value based on a randomly chosen permutation. Firstly, the selection of a permutation (line 3) is performed in a linear time using Knuth shuffle. Next, calculating value of a SGC-structure (lines 5 and 12) is done in a constant time. Finally, the loop over the permutation π (lines 6-13) performs $|N|$ operations $addNode()$, $|E|$ operations $addEdge()$ and $2|E|$ operations $exist()$. To summarize, the calculation of a single sample takes $O(|E| \cdot log^*(|N|))$. In other words, this is the time complexity of single iteration of the main loop.

6 As we traverse up the tree to the root we attach all passed nodes directly to the root to flatten the structure:
find(i) {*if (parent[i]* $\neq$ *i) return parent[i] = find(parent[i]);*}.

5.1 Illustration of Algorithm 1

Let us first consider the stylized example shown in Figure 1. The following are the Shapley values of the nodes computed using Algorithm 1 (in non-increasing order): $Sh_9 = 0.135$, $Sh_4 = 0.102$, $Sh_5 = 0.085$, $Sh_3 = 0.085$, $Sh_7 = 0.071$, $Sh_8 = 0.071$, $Sh_6 = 0.071$, $Sh_{10} = 0.068$, $Sh_{12} = 0.068$, $Sh_{11} = 0.068$, $Sh_{13} = 0.068$, $Sh_1 = 0.054$, $Sh_2 = 0.054$.

Next consider the network shown in Figure 2. The following are the Shapley values of the nodes computed using Algorithm 1 (in non-increasing order): $Sh_4 = 0.249$, $Sh_3 = 0.163$, $Sh_7 = 0.149$, $Sh_5 = 0.11$, $Sh_1 = 0.078$, $Sh_2 = 0.078$, $Sh_8 = 0.063$, $Sh_9 = 0.068$, $Sh_{10} = 0.024$, $Sh_6 = 0.024$.

6 Experimental Results

In this section, we show the efficacy of the proposed gatekeeper centrality by applying it to solve two social network analysis tasks: community detection and limiting the speed of misinformation.

6.1 Community Detection using Gatekeeper Centrality

Here we outline the steps of the algorithm that we follow to determine the community structure for a given network using gatekeeper centrality. We first arrange the nodes of the network in non-increasing order of their gatekeeper centrality values. Then we keep on removing the nodes in that order and compute the modularity at each point. We continue to do this process until there are no nodes left to remove. We report the community structure pertaining to the best modularity value. Here we consider *modularity* [13] and coverage [3] as the measures of performance for any clustering algorithm in our experiments. Informally, *coverage* measures the fraction of intra-cluster edges and *modularity* measures internal (and not external) connectivity, but it does so with reference to a randomized null model.

We compare the modularity and coverage of our approach with that of three benchmark algorithms for finding communities in social networks, namely (i) the greedy algorithm proposed by [23] and hereafter we refer to this as *Greedy Algorithm*, (ii) a spectral optimization approach due to [24] and hereafter we refer to this as *Spectral Algorithm*, and (iii) a randomized and game theoretic algorithm due to [7] and hereafter we refer to this as *RGT Algorithm*. The implementations were carried out using Java. Also, all our experiments were run on a Windows based Intel PC with two 2.00 GHz processors and 2GB of RAM.

Table 2 describes four network data sets that are well known in network science community. Table 3 shows the modularity and the coverage obtained using (i) our approach using version 1, (ii) our approach using version 2, (iii) Greedy Algorithm [23], (iv) Spectral Algorithm [24], and (v) RGT Algorithm [7] on these four network data sets. From these results in Table 3, our proposed approach clearly outperforms the benchmark algorithms in terms of coverage; and the modularity of the community structures obtained using our approach are comparable to that of the benchmark algorithms.

6.2 Limiting the Speed of Misinformation over Networks

We consider the problem of controlling the spread of a virus or misinformation in a network. The goal is to remove some nodes in the network such that the virus/misinformation is not allowed to spread rapidly in the entire network. We propose the use of nodes with high gatekeeper centrality to remove in order to decrease the speed of misinformation spread. It has been proven in the literature that the largest eigenvalue of the adjacency matrix of the given network accurately captures the speed of information diffusion in the network [25]. Therefore we use the largest eigenvalue of the adjacency matrix of the underlying social network to measure how the speed of diffusion has changed by removing certain nodes from the network. We present the results in Fig 3 and Fig 4 using two example network data sets. The X-axis denotes the number of nodes removed and Y-axis denotes the value of the leading eigenvalue (smaller is better). We compare our results with three standard centrality metrics (i.e.

Data Set	Nodes	Edges	Triangle Count
Karate	34	78	45
Dolphins	62	318	95
Political Books	105	882	560
FootBall	115	1226	810

Table 2. Description of network data sets

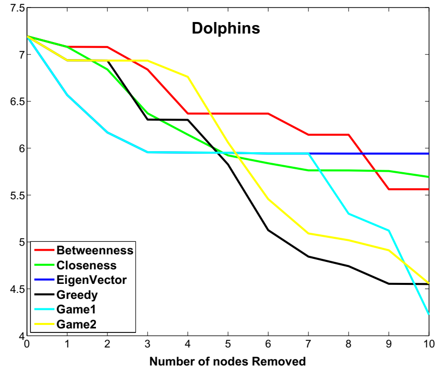

Figure 3. Results for decreasing the speed of misinformation diffusion for the Dolphins data set

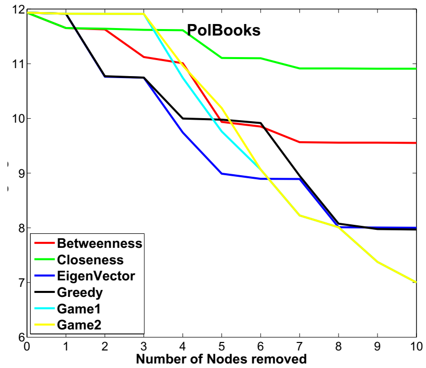

Figure 4. Results for decreasing the speed of misinformation diffusion for the PolBooks data set

Data Set	Game 1		Game 2		Greedy [23]		Spectral [24]		RGT [7]	
	Mod	Cov	Mod	Cov	Mod	Cov	Mod	Cov	Mod	Cov
Karate	0.3092	65.38	0.3092	65.38	0.380	30.76	0.393	25.24	0.392	68.52
Dolphins	0.4833	81.76	0.4606	74.84	0.495	22.01	0.491	22.64	0.502	69.43
Political Books	0.4371	91.83	0.4338	88.2	0.509	59.63	0.469	45.57	0.493	74.45
Football	0.5172	71.9	0.556	73.4	0.566	16.15	0.539	12.39	0.581	67.92

Table 3. Comparison of modularity and coverage due to our proposed approach with that of three benchmark algorithms. In this table, *Mod* means Modularity and *Cov* means Coverage

betweenness, closness, eigenvector) and the well known greedy algorithm [19] for information diffusion. Our approach turns out to be very effective in limiting the speed of misinformation over networks.

7 Conclusions

In this paper, we introduced a new centrality metric for social networks and we referred to this as *gatekeeper centrality*. We proposed an appropriate cooperative game and then presented efficient approximate algorithm to compute Shapley value of this game in order to rank the nodes based on the gatekeeper centrality.

Acknowledgements

Oskar Skibski & Tomasz Michalak were supported the Polish National Science grant DEC-2013/09/D/ST6/03920. Oskar Skibski acknowledges his research was co-financed with the European Union funds by the European Social Fund. Tomasz Michalak was supported by the European Research Council under Advanced Grant 291528 ("RACE").

REFERENCES

[1] N. Alon, F. Fischer, A. Procaccia, and M. Tennenholtz, 'Sum of us: strategyproof selection from the selectors', in *Proceedings of TARK*, pp. 101–110, (2011).

[2] Yoram Bachrach, Evangelos Markakis, Ariel D. Procaccia, Jeffrey S. Rosenschein, and Amin Saberi, 'Approximating power indices', in *In Proceedings of AAMAS*, pp. 943–950, (2008).

[3] U. Brandes and T. Erlebach, *Network Analysis: Methodological Foundations*, Lecture Notes in Computer Science, Volume 3418, Springer-Verlag, 2005.

[4] U. Brandes and E. Thomas, *Network Analysis: Methodological Foundations*, Lecture Notes in Computer Science, Springer, 2005.

[5] S. Brin and L. Page, 'The anatomy of a large-scale hypertextual web search engine', *Computer Networks*, **30**(1-7), 107–117, (1998).

[6] J. Castro, D. Gomez, and J. Tejada, 'Polynomial calculation of the shapley value based on sampling', *C&OR*, **36**(5), (2009).

[7] W. Chen, Z. Liu, X. Sun, and Y. Wang, 'A gametheoretic framework to identify overlapping communities in social networks', *Data Mining and Knowledge Discovery*, **21**, 224–240, (2010).

[8] D. Easley and J.M. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, Cambridge, U.K., 2010.

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

[10] Shaheen S. Fatima, Michael Wooldridge, and Nicholas R. Jennings, 'A linear approximation method for the shapley value', *Artificial Intellilgence*, **172**(14), 1673–1699, (September 2008).

[11] N. Friedkin, 'Theoretical foundations for centrality measures', *American Journal of Sociology*, **96**(6), 1478–1504, (1991).

[12] Bernard A. Galler and Michael J. Fisher, 'An improved equivalence algorithm', *Commun. ACM*, **7**(5), 301–303, (1964).

[13] M. Girvan and M.E.J. Newman, 'Community structure in social and biological networks', *Proc. Natl. Acad. Sci.*, **99**, (2002).

[14] D. Gomez, E. Gonzalez-Arangena, C. Manuel, G. Owen, M. del Pozo, and J. Tejad, 'Centrality and power in social networks: a game theoretic approach', *Mathematical Social Sciences*, **46**, 27–54, (2003).

[15] D. Gómez, E. González-Arangüena, C. Manuel, G. Owen, M. Del Pozo, and J. Tejada, 'Centrality and power in social networks: A game theoretic approach', *Mathematical Social Sciences*, **46**(1), 27–54, (2003).

[16] B. Grofman and G. Owen, 'A game theoretic approach to measuring degree of centrality in social networks', *Social Networks*, **4**, 213–224, (1982).

[17] R. Hendrickx, P. Borm, R. Vandenbrink, and G. Owen, 'The vl control measure for symmetric networks', *Social Networks*, **31**, 85–91, (2009).

[18] J.E. Hopcroft and J.D. Ullman, 'Set merging algorithms', *SIAM Journal on Computing*, **2**(4), 294–303, (1973).

[19] D. Kempe, J. Kleinberg, and É. Tardos, 'Maximizing the spread of influence through a social network', in *KDD '03: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146, (2003).

[20] T. Michalak, K.V. Aadithyaa, P.L. Szczepanski, B. Ravindran, and N.R. Jennings, 'Efficient computation of the shapley value for game-theoretic network centrality', *Journal of Artificial Intelligence Research (JAIR)*, **46**, 607–650, (2013).

[21] T. P. Michalak, T. Rahwan, P. L. Szczepanski, O. Skibski, R. Narayanam, M. J. Wooldridge, and N. R. Jennings, 'Computational analysis of connectivity games with applications to the investigation of terrorist networks', *IJCAI*, (2013).

[22] R.B. Myerson, 'Graphs and cooperation in games', *Mathematics of Operations Research*, **2**(3), 225–229, (1077).

[23] M.E.J. Newman, 'Fast algorithm for detecting community structure in networks', *Physical Review E*, **69, 066133**, (2004).

[24] M.E.J. Newman, 'Modularity and community structure in networks', *Proceedings of National Academy of Science*, **103**, (2006).

[25] B.A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos, 'Threshold conditions for arbitrary cascade models on arbitrary networks', in *Proceedings of (ICDM*, (2011).

[26] N. Ramasuri and Y. Narahari, 'Determining the top-k nodes in social networks using the shapley value', in *Proceedings of AAMAS*, pp. 1509–1512, (2008).

[27] L. S. Shapley, 'A value for n-person games', in *In Contributions to the Theory of Games, volume II*, eds., H.W. Kuhn and A.W. Tucker, 307–317, Princeton University Press, (1953).

[28] L.S. Shapley, *A Value of n-person Games*, Contributions to the Theory of Games, volume II, by H.W. Kuhn and A.W. Tucker, editors. Annals of Mathematical Studies, pp. 307317, Princeton University Press, 1953.

[29] O. Skibski, T. Michalak, T. Rahwan, and M. Wooldridge, 'Algorithms for the myerson and shapley values in graph-restricted games', in *Proceedings of AAMAS*, pp. 197–204, (2014).

[30] P.L. Szczepanski, T.P. Michalak, and T. Rahwan, 'A new approach to betweenness centrality based on the shapley value', in *Proceedings of AAMAS*, pp. 239–246, (2012).

[31] R. Van den Brink, 'The apex power measure for directed networks', *Social Choice and Welfare*, **19**, 845–867, (2002).

[32] R. Van den Brink, P. Borm, R. Hendrickx, and G. Owen, 'Characterization of the β and the degree network power measure', *Theory and Decision*, **64**, 519–536, (2007).

[33] R. Van den Brink and R. Gilles, 'Measuring domination in directed networks', *Social Networks*, **22**, 141–157, (2000).