



ELSEVIER

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint



Efficient algorithms for game-theoretic betweenness centrality

Piotr L. Szczepański^{a,*}, Tomasz P. Michalak^{b,c}, Talal Rahwan^d^a Institute of Informatics, Warsaw University of Technology, Pl. Polit. 1, 00-661 Warsaw, Poland^b Department of Computer Science, University of Oxford, OX1 3QD, UK^c Institute of Informatics, University of Warsaw, ul. Banacha 2, 02-097 Warsaw, Poland^d Masdar Institute of Science and Technology, Abu Dhabi 54224, United Arab Emirates

ARTICLE INFO

Article history:

Received 11 October 2014

Received in revised form 23 October 2015

Accepted 2 November 2015

Available online 5 November 2015

Keywords:

Betweenness centrality

Shapley value

Semivalue

ABSTRACT

Betweenness centrality measures the ability of different nodes to control the flow of information in a network. In this article, we extend the standard definition of betweenness centrality using *Semivalues*—a family of solution concepts from cooperative game theory that includes, among others, the Shapley value and the Banzhaf power index. Any Semivalue-based betweenness centrality measure (such as, for example, the Shapley value-based betweenness centrality measure) has the advantage of evaluating the importance of individual nodes by considering the roles they each play in different groups of nodes. Our key result is the development of a general polynomial-time algorithm to compute the Semivalue-based betweenness centrality measure, and an even faster algorithm to compute the Shapley value-based betweenness centrality measure, both for weighted and unweighted networks. Interestingly, for the unweighted case, our algorithm for computing the Shapley value-based centrality has the same complexity as the best known algorithm for computing the standard betweenness centrality due to Brandes [15]. We empirically evaluate our measures in a simulated scenario where nodes fail simultaneously. We show that, compared to the standard measure, the ranking obtained by our measures reflects more accurately the influence that different nodes have on the functionality of the network.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Betweenness centrality is a measure introduced independently by Anthonisse [3] and by Freeman [31], based on which the importance (or “centrality”) of a node, v , is assumed to increase with the percentage of shortest paths (from all nodes to all other nodes) that pass through v . Everett and Borgatti [29] extended this measure to *groups* of nodes, whereby the importance of any such group increases with the percentage of shortest paths that pass through that group; this is called the *group Betweenness centrality measure*.

Several applications of betweenness centrality have been considered in the literature. Among others, Puzis et al. [55] studied how group betweenness centrality can guide the process of deploying intrusion-detection devices in a complex network. Dolev et al. [24] developed an extension of the standard measure, called routing betweenness centrality, to identify key routers in a communication network. Girvan and Newman [34] proposed a community-detection algorithm based on edge betweenness centrality. Furthermore, according to the European Central Bank [28], betweenness centrality is arguably the most suitable, simple measure of systemic importance of institutions in the financial system.

* Corresponding author. Tel.: +48 692144559.

E-mail addresses: szczep.pl@gmail.com (P.L. Szczepański), Tomasz.Michalak@cs.ox.ac.uk (T.P. Michalak), trahwan@masdar.ac.ae (T. Rahwan).

Table 1

Summary of the computational results obtained in this article. The complexity results, except for group betweenness centrality, concern computing the entire cardinal ranking of nodes. The complexity of group betweenness centrality concerns computing the centrality of a single set of nodes.

Centrality	Unweighted graphs	Weighted graphs
Standard betweenness	$O(V E)$, Brandes [15]	$O(V E + V ^2 \log(V))$, Brandes [15]
Betweenness of a particular group	$O(V E)$, Brandes [16]	$O(V E + V ^2 \log(V))$, Brandes [16]
Shapley value-based betweenness	$O(V E)$, this article	$O(V ^2 E) + V ^2 \log V $, this article
Semivalued-based betweenness	$O(V ^2 E)$, this article	$O(V ^3 E) + V ^3 \log V $, this article

In this article, we propose a generalization of betweenness centrality based on solution concepts from cooperative game theory. Our work falls under a wider body of research in which networks are analyzed using the combinatorial structure of coalitional games (see the next section for more details). The key advantage of this approach is that nodes are ranked not only based on their individual roles in the network but also based on their contribution to the role played by all possible subsets (or groups) of nodes. This is particularly relevant to settings in which the role played by a group of nodes is not simply the sum of the individual roles played by the group members. Consider, for instance, an epidemiology application, where the aim is to contain the spread of a disease [25]. If we ask the question of whether the vaccination of any individual node is sufficient to stop the spread of the disease then the answer is most probably negative. A much more likely way to contain the disease is to simultaneously vaccinate a group of carefully-selected nodes. In such a scenario, to quantify the importance of a node, one needs to consider the potential gain of vaccinating that node as part of a group, rather than just considering the potential gain of vaccinating it alone.

Such an analysis of groups of nodes in the network directly corresponds to coalitional game theory, where the performance of players is studied in “coalitions” (i.e., subsets of players). By imposing the combinatorial structure of a coalitional game over a network, it becomes possible to analyze the performance of nodes using a plethora of game-theoretic solution concepts, developed over decades to analyze the performance of players in a cooperative game. Note that a cooperative game typically places no assumptions or restrictions on how the groups are evaluated. Thus, when using such a game to model the roles played by nodes in a network, the group-evaluation function can be tailored to best fit the centrality measure at hand. For instance, a group of nodes can be evaluated based on the average degree therein, or based on its closeness to other nodes, etc. (more on this in the next section). As such, several game-theoretic network centralities have been proposed in the literature to date, each based on a different group-evaluation function [62,44,48].

Compared to the standard centrality measures, a potential downside of *game-theoretic centrality measures* is that they are based on solution concepts that are themselves hard to compute. For instance, given a coalitional game defined over a network of $|V|$ nodes, a straight-forward computation of the *Shapley value*—a fundamental game-theoretic solution concept—requires considering all possible $2^{|V|}$ coalitions (i.e., groups) of nodes. This is clearly prohibitive for networks with hundreds, or even tens, of nodes. Indeed, certain game-theoretic centrality measures have been proven impossible to compute in time polynomial in the size of the network [49]. On the other hand, some positive computational results have also been found. In particular, Michalak et al. [48] analyzed various Shapley value-based extensions of degree and closeness centrality and showed that it is possible to compute those in polynomial time. Nevertheless, these positive results were only for certain game-theoretic extensions of degree centralities, which are computationally less challenging than betweenness centrality. In fact, no game-theoretic extension of betweenness centrality has been developed to date.

Given this, our contributions in this article can be summarized as follows:

- We propose a game-theoretic extension of Betweenness centrality, based on a family of solution concepts from cooperative game theory known as *Semivalues* [50]. This family includes two of the most fundamental solution concepts in the literature, namely the *Shapley value* [56] and the *Banzhaf Index* [9]. To put it in the context of measuring the systemic importance of financial institutions, the standard betweenness centrality measure, suggested by the European Central Bank [28] as a simple index of systemic importance, assumes that any crisis is initiated by a single insolvent institution. On the other hand, our Semivalued-based betweenness centrality measure takes into consideration the fact that a crisis in the financial network can be initiated by a group of insolvent institutions, with varying probability depending on the size of the group.
- Our main technical contribution is to propose polynomial time algorithms to compute the Semivalued-based betweenness centrality measure, and an even faster algorithm to compute the Shapley value-based betweenness centrality measure, both for weighted and unweighted networks. Interestingly, as shown in Table 1, for the unweighted case our algorithm for computing the Shapley value-based centrality has the same complexity as the best known algorithm for computing the standard betweenness centrality, due to Brandes [15]. In particular, both algorithms run in $O(|V||E|)$ time, where V is the set of nodes and E is the set of edges in the network.
- Finally, we compare the performance of the new measures with that of the standard betweenness-centrality measure in a scenario of simultaneous node failures (in the spirit of Holme et al. [39]). To this end, we quantify the functionality of the resulting network based on four well-known measures, namely the *Average Inverse Geodesic* [39], the *Average*

Clustering Coefficient [65], the *Largest Component* [1], and the *Fragmentation Ratio* [19].¹ In a number of experiments that involve both real and synthetic networks, we show that the ranking obtained by our measures (compared to the one obtained by the standard measure) reflects more accurately the influence that different nodes have on the network functionality.

The remainder of the article is organized as follows. Section 2 discusses the related work. Section 3 introduces the basic notation and definitions. Section 4 defines our centrality measure and discusses its properties. Section 5 describes our polynomial time algorithms for computing the new centrality measures. Section 6 provides an empirical evaluation of our measure and our algorithms. Section 7 summarizes our results and presents some potential directions for future research. Finally, we provide more details on various game-theoretic centrality measures at: www.game-theoretic-centrality.com, and provide the implementations of all our algorithms at: www.network-centrality.com.

2. Related work

One of the main research directions in network science is the analysis of centrality measures. In his seminal work, Freeman [31] proposed and formalized three different concepts of centrality: (i) *degree centrality* is based on being directly connected to other nodes; (ii) *closeness centrality* is based on being close to all other nodes; and (iii) *betweenness centrality* is based on lying between other nodes.² Many authors have subsequently proposed new centrality measures, either by refining and extending the classical ones or by developing utterly novel approaches [17,54].

Grofman and Owen [38] were the first to analyze the centrality of nodes using techniques from cooperative game theory. In particular, they considered all induced connected subgraphs of a directed graph, and defined a centrality measure whereby the more central a node is, the more often its removal breaks paths in such subgraphs. The subsequent literature on game-theoretic centrality can be divided into two main categories.

- In the first category, the game-theoretic approach to centrality follows the well-known coalitional game model defined on graphs, due to Myerson [52]. Specifically, in this model, cooperation is restricted by an underlying communication graph: nodes are able to cooperate only if they can communicate directly along the edges of the graph or indirectly via intermediary nodes. Without either direct or indirect communication, cooperation is not feasible. In the context of game-theoretic centrality, this means that whether a group of nodes is connected (i.e., whether it induces a connected subgraph) has a significant influence on its value. Consequently, in this line of research, the central nodes are those that play pivotal roles in connecting various groups of nodes. This line includes the works of Gomez et al. [36], del Pozo et al. [23], Amer et al. [2], Narayanam et al. [53], and Skibski et al. [57,58].
- In the second category, to which our article contributes, the value of a group of nodes does not depend directly on whether the subgraph induced by that group is connected or not. The origins of this approach can be traced back to the concept of *group centrality*, whereby centrality is computed not just for individual nodes, but also for groups of nodes [29]. By computing group centrality for every possible group of nodes (not just those that are connected), we obtain a well-defined coalitional game. One can then apply a solution concept, such as the Shapley value, on this game to quantify the role played by individual nodes in all possible groups of nodes (we will explain this approach in more detail in Section 4).

The first work that belongs to the latter category is due to Suri [62] who applied a game-theoretic centrality to study influence propagation in networks. Specifically, the authors constructed a coalitional game by assigning to every coalition (i.e., every group of nodes) a payoff equal to its group degree. Then, the authors used the Shapley value to rank the nodes in the network, and proposed the top k nodes in this ranking as an approximate solution to the *top- k -node* problem, i.e., the problem of identifying the k most influential nodes in the network.

The next step in this line of research was taken by Michalak et al. [48], who mapped onto a network a number of coalitional games; in each such game, a coalition's value (i.e., payoff) was taken to be some function of group degree centrality. For most of these games, Michalak et al. proposed exact algorithms to compute the Shapley value in polynomial time.

Another step was taken by Szczepanski et al. [64], who developed the first measure of centrality that takes into account networks with predefined *community structure* (e.g., a social network in which nodes represent individuals and communities represent nationalities). Their measure is based on a generalization of the Owen value—a solution concept from cooperative game theory that focuses on games with *a priori*-given unions (i.e., cooperative arrangements) of players. The authors developed a polynomial-time algorithm to compute their measure in games where the value (i.e., payoff) of every coalition is equal to the coalition's group degree centrality.

Another line of research that combines network science and cooperative game theory is that of *network flow games* [40, 41] and *threshold network flow games* [4]. In the former games, players are edges, and the value of a coalition, C , is the

¹ See Section 4.1.2 for their exact definitions.

² For an overview and a theoretical foundation of various centrality measures, see [32].

maximum amount of flow that can be sent from the *source*, s , to the *sink*, t , via edges from C . In the latter games, the value of a coalition is 1 if the maximum flow going through it exceeds a certain threshold, and 0 otherwise. For such games, it was possible to identify certain subclasses in which the Shapley value and the Banzhaf index can be computed in polynomial time [5]. An overview of the literature on network flow games can be found in the book by Chalkiadakis et al. [18].

Next, we discuss the literature related to our simulation section, where the performance of our Semivalue-based betweenness centrality measure is compared to the performance of the standard betweenness centrality measure in a scenario of simultaneous node failures. In more detail, we design our experiments in the same spirit as that of Holme et al. [39], who examined how different strategies of attack on a network affect its performance. More specifically, the authors studied how removing the nodes with the highest betweenness centrality affects communication throughout the network. They found that, in random scale-free networks, removing such nodes causes an exponential decay in network performance. While Holme et al. [39] measured the performance of the network as the average distance between nodes, in our simulations we consider three other measures from the literature (see Section 6 for more details).

Naturally, the body of research on network vulnerability analysis reaches far beyond the work of Holme et al. [39]. Consider, for instance, the literature on three-stage Stackelberg games [61]. In a nutshell, two opponents, called the *interdictor* and the *operator*, try to maximize their gains in a game consisting of three stages. In the first stage, the operator fortifies the network by choosing the nodes that should be fully protected. In the second stage, the interdictor performs an attack that involves removing a limited number of unfortified nodes. Finally, in the third stage, the operator acts on the network in order to fulfill its objectives, e.g., he can travel between two nodes, or send some load between nodes. The interdictor can be thought of as a human, a natural disaster, or a random failure of nodes. Typically, the game is studied from the perspective of the operator, where the focus is either on the worst case as per Murphy's Law [59], or on the average case, where the interdictor's attacks are assumed to follow a certain stochastic model [22]. One variation of Stackelberg games is called *Shortest Path Interdiction*, where the *interdictor* attempts to remove k edges in order to maximize the distance between two nodes s and t . Here, although the problem faced by the operator was shown to be NP-hard [8], Malik et al. [45] proposed an effective approximation algorithm, and Corley and Sha [20] solved this problem in polynomial time for the special case where $k = 1$.

Finally, we mention the link to the literature on multiagent systems, where the failure of agents has been studied both in cooperative settings [6] as well as non-cooperative settings [47], and even in network games [7].

Having discussed the related work, in the next section we introduce the basic definitions and notation used throughout the article.

3. Basic definitions and notation

Let us consider first the relevant notions from graph theory. A *graph* (or *network*) G is composed of *vertices* (or *nodes*) and *edges*. Their sets will be denoted by $V(G)$ and $E(G)$, respectively, or simply by V and E when there is no risk of confusion. An edge connecting nodes $u, v \in V(G)$ will be denoted by (u, v) . We will also consider *weighted networks* in which a weight (label) is associated with every edge in $E(G)$. A *path* is a sequence of connected edges in the network, whereas a *shortest path* between two nodes, u and v , is a path that ends with these nodes and minimizes the weights of the involved edges (or minimizes the number of edges, in the case of unweighted networks).

Now, we are ready to formally introduce betweenness centrality:

Definition 1. *Betweenness centrality* of a node v is defined as a function $c_b : V \rightarrow \mathbb{R}$ such that:

$$c_b(v) = \sum_{\substack{s \in V \setminus \{v\} \\ t \in V \setminus \{v\}}} \frac{\sigma_{st}(v)}{\sigma_{st}}, \quad (1)$$

where σ_{st} is the number of shortest paths from s to t (if $s = t$ then $\sigma_{st} = 1$), and $\sigma_{st}(v)$ is the number of shortest paths from s to t passing through node v (if $v \in \{s, t\}$ then $\sigma_{st}(v) = 0$).³

Intuitively, the nodes with the highest betweenness centrality can be thought of as nodes that *control* more shortest paths in the network than others. Arguably, then, these are the most important nodes from the perspective of controlling the flow through the network.

Everett and Borgatti [29] introduced group betweenness centrality. Viewed from the perspective of network flow, a group of nodes controls a shortest path if this path includes at least one node from the group.

Definition 2. The *group betweenness centrality* of a set of nodes $S \subseteq V$ is defined as a function $c_{gb} : 2^V \rightarrow \mathbb{R}$ such that:

³ To deal with unconnected graphs, it is assumed that $\frac{0}{0} = 0$.

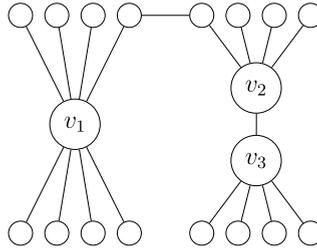


Fig. 1. In the above sample network, both v_1 and v_2 are ranked equally by standard betweenness centrality; $c_b(v_1) = c_b(v_2) = 98$. Indeed, v_1 and v_2 control the same number of shortest paths in this network. However, if we consider the roles that nodes play within various groups—as is the case in our scenario where financial difficulties may affect multiple institutions simultaneously—then v_1 and v_2 should not be evaluated equally. To see why this is the case, let us consider nodes v_2 and v_3 . A significant percentage of the shortest paths controlled by v_2 are also controlled by v_3 . Consequently, the ability to control financial flows by the group of banks $\{v_2, v_3\}$ is smaller than by the group $\{v_1, v_3\}$. As a result, v_1 is more systemically important than v_2 .

$$c_{gb}(S) = \sum_{\substack{s \in V \setminus S \\ t \in V \setminus S}} \frac{\sigma_{st}(S)}{\sigma_{st}}, \tag{2}$$

where $\sigma_{st}(S)$ is the number of shortest paths from s to t passing through at least one node in S (if $s \in S$ or $t \in S$ then $\sigma_{st}(S) = 0$).

We note that the group betweenness centrality of an arbitrary $S \subseteq V$ is not, in general, equal to the sum of betweenness centralities of the nodes in S (see Fig. 1 for an example). We also note that Kolaczyk et al. [42] proposed another way to extend the standard betweenness centrality to groups of nodes. In their centrality, called *co-betweenness*, a path is controlled by a set, S , if that path contains all the members of S . While one could also think of other ways of defining group betweenness centrality, in this article, we build upon Definition 2, since it is arguably the most established in the literature.

Next, we introduce some basic concepts from cooperative game theory. Let us denote by $A = \{1, \dots, |A|\}$ the set of players that participate in a coalitional game. A *characteristic function* $v : 2^A \rightarrow \mathbb{R}$ assigns to every coalition $C \subseteq A$ a numerical value representing its performance; this is referred to as the *value*, or the *payoff*, of the coalition. It is assumed that $v(\emptyset) = 0$. A *coalitional game in characteristic function form* is then a tuple (A, v) .

It is usually assumed that the *grand coalition*, i.e., the coalition of all the players in the game, is formed. This happens, for example, in superadditive games, where the union of any two disjoint coalitions has a value greater than, or equal to, the sum of the values of those two coalitions. One of the fundamental problems in cooperative game theory is then to determine how to divide the payoff of the grand coalition among the players. Whereas, in principle, there are infinitely many solutions to this problem, we are interested in *solution concepts* that meet certain desirable criteria. In this context, Shapley [56] focused on analyzing the *marginal contributions* of players to coalitions, where the marginal contribution of player i to a coalition $C \subseteq A \setminus \{i\}$ is simply the difference in value that player i makes when joining C , i.e., it is equal to $v(C \cup \{i\}) - v(C)$. With this in mind, Shapley [56] proposed that the role of a player, i , is evaluated as a weighted average of the marginal contributions of i to all possible coalitions. Formally:

Definition 3. In a coalitional game (A, v) the *Shapley value* for a player i is:

$$SV_i(A, v) = \frac{1}{|A|!} \sum_{\pi \in \Pi(A)} [v(P_\pi(i) \cup \{i\}) - v(P_\pi(i))], \tag{3}$$

where $\pi \in \Pi(A)$ denotes a permutation of players in A , and $P_\pi(i)$ denotes the coalition made of all predecessors of i in π , e.g., $P_{(2,4,1,3)}(1) = \{2, 4\}$. We will often write SV_i instead of $SV_i(A, v)$ for conciseness.

The intuition behind the above solution concept is as follows. Suppose that the players arrive at a certain meeting point in a random order. Furthermore, suppose that every player i who arrives receives the marginal contribution that his arrival brings to those already at the meeting point. Then, according to the Shapley value, the payoff of player i is his average marginal contribution, taken over all the possible orders of arrival.⁴

The attractiveness of the Shapley value stems from the fact that it is the unique solution concept characterized by certain intuitive axioms. One such axiomatization was proposed by Young [68], who showed that the Shapley value is the only possible solution concept satisfying all of the following desirable properties: (i) *Symmetry*—players that are symmetric receive the same payoff; (ii) *Efficiency*—the entire payoff of the grand coalitions is distributed among the players; and (iii) *Marginality*—if a player contributes to each coalition in game (A, v) no less than what he contributes to the same coalition

⁴ See the book by Maschler et al. [46] for more analysis of the Shapley value, and the work by Moretti and Patrone [51] for an overview of its applications.

in game (A, v') then his payoff in the former game should be no less than in the latter one. Many other axiomatizations of the Shapley value have been proposed to date, see the work by Winter [67] for an overview.

One can alternatively compute the Shapley value using the following formula:

$$SV_i(A, v) = \sum_{S \subseteq A \setminus \{i\}} \frac{|S|!(|A| - |S| - 1)!}{|A|!} [v(S \cup \{i\}) - v(S)]. \quad (4)$$

This is more computationally efficient than Equation (3), as it iterates over all possible coalitions (which are $2^{|A|}$ in total), instead of all permutations (which are $|A|!$ in total). Nevertheless, $2^{|A|}$ is still a prohibitive number for large $|A|$. As such, even with Equation (4), computing the Shapley value is challenging, unless we leverage any additional information that may be available on how coalitions are evaluated.

The Shapley value belongs to a family of solution concepts known as *Semivalues* [26,66] whereby the payoff of a player is a weighted average of that player's marginal contributions to all coalitions. Formally:

Definition 4. For any given game (A, v) , a *Semivalue*⁵ is defined by a vector of weights $(p_0, \dots, p_{n-1}) \in \mathbb{R}^n : \sum_{k=0}^{|A|-1} p_k \binom{|A|-1}{k} = 1$, and is computed for every player $i \in A$ as follows:

$$S_i(A, v) = \sum_{S \subseteq A \setminus \{i\}} p_{|S|} \times [v(S \cup \{i\}) - v(S)]. \quad (5)$$

One of the most prominent Semivalues is the Shapley value, defined by the weights $p_k = 1/(|A| \binom{|A|-1}{k})$, $\forall k \in \{0, \dots, n-1\}$. Another prominent Semivalue is the Banzhaf Index [9], defined by the weights $p_k = 1/(2^{|A|-1})$, $\forall k \in \{0, \dots, n-1\}$. All Semivalues satisfy Symmetry and Marginality. The Shapley value, however, is the unique Semivalue that also satisfies Efficiency.

In the following section, we will show how Semivalues in general, and the Shapley value in particular, can be used to generalize the concept of betweenness centrality.

4. The new centrality and its properties

In this section we extend the standard betweenness centrality measure, and analyze the basic properties of our extension. We start in Section 4.1 with a few motivating scenarios that demonstrate the need for a more involved centrality measure than standard betweenness centrality. Sections 4.2 and 4.3 introduce our Shapley value-based and Semivalue-based betweenness centralities, respectively. Finally, Sections 4.4 and 4.5 analyze the marginal contributions of nodes, in the context of the Shapley value-based and Semivalue-based betweenness centralities, respectively.

4.1. Motivating scenarios

We will consider two motivating scenarios: ranking institutions in a financial system according to their systemic importance (Section 4.1.1), and analyzing network vulnerability (Section 4.1.2).

4.1.1. Measuring systemic importance of financial institutions

The financial crisis of 2007–2008 clearly highlighted the need for a more stringent supervision of financial systems. Consequently, a number of systemic risk monitoring and supervisory bodies have been created worldwide, namely the Financial Stability Oversight Council (USA), the European Systemic Risk Board and the Financial Stability Board. Furthermore, new prudential policies and regulations were laid out. Most notably, the Basel III accords aim at introducing micro- and macro-prudential regulations such as stricter capital reserve ratios, increased audit transparency and more thorough risk management by banks and other financial institutions.

At a discretion of supervising authorities, systemically important institutions may be asked to follow additional macro-prudential policies including a combination of capital surcharges, contingent capital and bail-in debt. As stated in the Financial Stability Board's interim report to G20 leaders, "*Financial institutions should be subject to requirements commensurate with the risks they pose to the financial system*" [30].

In this context, betweenness centrality is considered to be the most suitable simple centrality to measure systemic importance of institutions in the financial system [28].⁶ In more detail, "*the betweenness centrality of a bank A connecting pairs of nodes in the network is a measure of the dependence of these other banks from A to transfer the loans.*" [33]. As a result, the greater the betweenness centrality of an institution, the more attention and scrutiny it should receive.

⁵ We write "a Semivalue" instead of "the Semivalue" because there are infinitely many Semivalues, depending on the chosen weights: (p_0, \dots, p_{n-1}) .

⁶ Note that more complex measures and models of systemic importance have been proposed recently, e.g., Sink Rank [60] and DebtRank [11]; these approaches are, however, much more involved since they rely on various additional data such as banks' liquidity constraints and not solely on network topology as betweenness centrality.

However, imposing additional capital requirements on financial institutions according to their standard betweenness centrality ignores the fact that financial crises are often ignited not by factors pertaining to a single bank but rather by combinations of factors that affect several banks simultaneously. For instance, the infamous financial crises of 2009 started with the collapse of the subprime residential mortgage market in the United States and then spread to the rest of the world, due to the exposure to American real estate assets via a complex array of financial derivatives. As a result, many financial institutions in various countries have asked for urgent state interventions [43].

Against this background, it would be desirable to extend betweenness centrality so that it evaluates individual institutions while taking into account the role that each such institution plays in various groups of institutions, since any such group can be affected simultaneously. We formalize this problem as follows:

Problem 1. Given a graph $G = (V, E)$ and a discrete probability distribution $PD : \{1, \dots, n\} \rightarrow [0, 1]$, identify the cardinal ranking of nodes that reflects the expected marginal contribution of each node to the group betweenness centrality of a random subset $S \subseteq V$ the size of which is drawn from PD and members of which are chosen uniformly at random from V .⁷

Later on in this article, we will propose a polynomial-time exact algorithm to solve Problem 1.⁸

4.1.2. Analysis of network vulnerability

One of the extensive research directions on betweenness centrality involves the analysis of network vulnerability [13,39]. The aim here is to identify (and possibly protect) the most critical nodes—those whose removal degrades the functionality of the network the most. In this context, the four standard measures for assessing the condition of a network are:

- *Average Inverse Geodesic Measure (IGM)*, which is the sum of the inverse distances between any pair of nodes in the network [39,1]. Formally, it is defined as follows:

$$IGM(V, E) = \sum_{v \in V} \sum_{v \neq u \in V} \frac{1}{d(v, u)},$$

where $d(v, u)$ is the distance between nodes v and u .

- *Average Clustering Coefficient (CC)*, which measures the degree to which nodes tend to cluster together [65]. Formally:

$$CC(V, E) = \frac{1}{|V|} \sum_{v \in V} \frac{2|\{(u, w) : (u, w) \in E \text{ and } m, n \in N(v)\}|}{deg(v)(deg(v) - 1)},$$

where $deg(v)$ is the degree of node v and $N(v)$ is the set of neighbors of v .

- *Largest Component (LC)*, which focuses on the size of the largest connected group of nodes in the network [1]. Formally:

$$LC(V, E) = \frac{\text{size of the largest component}}{|V|}.$$

- *Fragmentation Ratio (FR)*, which focuses on the number of connected groups of nodes in the network [19]. Formally:

$$FR(V, E) = \frac{1}{\text{number of components}}.$$

We formalize the problem of network vulnerability as follows:

Problem 2. Given a graph $G = (V, E)$ and a discrete probability distribution $PD : \{1, \dots, n\} \rightarrow [0, 1]$, identify a set of nodes $C \subseteq V$ that maximizes the expected value of the following expression:

$$M(V \setminus S) - M(V \setminus (S \setminus C)),$$

where M is one of the aforementioned four measures of network condition, i.e., $M \in \{IGM, CC, LC, FR\}$, and $S \subseteq V$ is a random subset whose size is drawn from PD and whose members are chosen uniformly at random from V .

In Section 6, we will empirically show that the Semivalue-based betweenness centrality measure achieves consistently better results in approximating Problem 2 than the standard betweenness centrality measure.

⁷ In other words, for a given size $k \sim PD$, a subset is drawn uniformly at random from $\{S \subseteq V : |S| = k\}$, implying that every such subset is chosen with probability $1/\binom{|V|}{k}$.

⁸ Note that the probability PD must be known *a priori* to the algorithm.

4.2. The Shapley value-based betweenness centrality measure

In this subsection, we focus on the problem of measuring the importance of individual nodes based on the group centralities of different groups of nodes. Since a similar problem was studied by game theorists in the context of coalitional games, we can directly use their solution concepts, and the Shapley value in particular, to achieve our goal.

First, we cast the network $G(V, E)$ into the combinatorial structure of a coalitional game. That is, we consider all subsets of $V(G)$ and define a function that assigns to every such subset a numerical value; we set this value to be the subset's group betweenness centrality. Formally, borrowing notation from coalitional games, we define a game $(V(G), \nu)$, where the characteristic function $\nu : 2^{V(G)} \rightarrow \mathbb{R}$ assigns to each $S \subseteq V(G)$ the value $c_{gb}(S)$, and $\nu(\emptyset) = 0$. Now since $(V(G), \nu)$ has the same combinatorial structure as that of a coalitional game, we can use the Shapley value to quantify the importance of nodes based on the impact that their membership makes on group betweenness centrality. In other words, we obtain a measure of centrality of individual nodes which aggregates information about their role in group betweenness centrality across various groups. Formally, the Shapley value-based betweenness centrality measure is defined as follows:

Definition 5 (The Shapley value-based betweenness centrality measure). Given a network $G = (V, E)$, the Shapley value-based betweenness centrality of node $v \in V(G)$ is defined as the function $c_{Sh} : V \rightarrow \mathbb{R}$ such that:

$$c_{Sh}(v) = SV_v(V(G), \nu),$$

where $\nu : 2^{V(G)} \rightarrow \mathbb{R}$ is defined for every subset of nodes, $S \subseteq V(G)$, as follows:

$$\nu(S) = \sum_{\substack{s \in V \setminus S \\ t \in V \setminus S}} \frac{\sigma_{st}(S)}{\sigma_{st}} \text{ if } S \neq \emptyset, \text{ and } \nu(S) = 0 \text{ otherwise.} \tag{6}$$

Example 1. Consider again nodes v_1 and v_2 from Fig. 1. Here:

$$c_{Sh}(v_1) = 18.2 \text{ and } c_{Sh}(v_2) = 16.0833.$$

Unlike the standard betweenness centrality, the ranking produced by our Shapley-value based betweenness centrality reflects the fact that node v_2 has many common shortest paths with v_3 , while v_1 does not.

Intuitively, the Shapley value-based betweenness centrality measure represents the load placed on a given node, taking into consideration the role it plays in all possible groups of nodes in the network. Note that this centrality measure has all the properties of the Shapley value, including Symmetry, Efficiency and Marginality (see Section 3); although not all of them (especially Efficiency) seem as appealing in the network context as they are in the game-theoretic context.

Naturally, it is possible to normalize the Shapley value-based betweenness centrality measure such that it ranges from 0 to 1. To do so, it suffices to replace $SV_v(V(G), \nu)$ in Definition 5 with $\frac{SV_v(V(G), \nu)}{2 \max_{v \in V} c_b(v)} + \frac{1}{2}$. This is because $SV_v(V(G), \nu)$ is a weighted average of the marginal contributions of v , and every such marginal contribution is not greater than the betweenness centrality of v . That is, $(\nu(C \cup \{v\}) - \nu(C)) \leq c_b(v)$ for every $v \in V$ and every $C \subseteq V$.

Proposition 1. The Shapley value-based betweenness centrality is the exact solution to Problem 1 when $PD(k) = \frac{1}{|V|}, \forall k \in \{1, \dots, n\}$.

Proof. To prove the proposition, it suffices to rewrite Equation (4) as follows:

$$SV_v(V, \nu) = \frac{1}{|V|} \sum_{1 \leq k \leq |V|} \frac{1}{\binom{|V|-1}{k-1}} \sum_{\substack{S \subseteq V \\ |S|=k \\ v \in S}} (\nu(S) - \nu(S \setminus \{v\})). \tag{7}$$

In more detail, the inner sum in the above equation runs through all subsets of nodes containing v of a given size k , and $\binom{|V|-1}{k-1}$ is the number of such subsets. Now, the fraction $\frac{1}{|V|}$ implies that the size $k \in \{1, \dots, n\}$ is chosen uniformly at random, i.e., each with probability $\frac{1}{|V|}$. On the other hand, the fraction $\frac{1}{\binom{|V|-1}{k-1}}$ implies that out of all subsets of size k , one is chosen uniformly at random. \square

Equation (7) has the following interpretation in the context of measuring systemic importance of financial institutions:

1. Fraction $\frac{1}{|V|}$ means the probability that a financial crisis starts with a group of financial institutions of size $|S| = k$ is the same for any $k \in \{1, \dots, |V|\}$.
2. Fraction $\frac{1}{\binom{|V|-1}{k-1}}$ means that all k -sized subsets of institutions containing v have the same probability of becoming bankrupt.

While the latter implication seems somewhat natural, the former implication seems less so. For example, it seems unnatural to assume that the probability of a single institution becoming bankrupt is exactly the same as the probability of all institutions becoming bankrupt (especially if the network is very large). To relax this assumption, we need to change the fraction $\frac{1}{|V|}$ in Equation (7). This is possible by replacing the Shapley value with Semivalues, as we will show in the following section.

4.3. The semivalue-based betweenness centrality measure

Recall that Semivalues generalize the Shapley value by allowing for subsets of different sizes to have different weights (see the weighted sum in Equation (5), where every subset S is assigned a weight, that depends on $|S|$, i.e., $p_{|S|}$). Building upon Semivalues, we will now introduce a centrality measure that generalizes the Shapley value-based betweenness centrality.

Definition 6 (Semivalue-based betweenness centrality measure). Given a network $G = (V, E)$, the Semivalue-based betweenness centrality of node $v \in V(G)$ is defined as the function $c_S : V \rightarrow \mathbb{R}$ given by:

$$c_S(v) = S_v(V(G), v),$$

where $v : 2^{V(G)} \rightarrow \mathbb{R}$ is the function defined for every $S \subseteq V(G)$ as follows: $v(S) = \sum_{\substack{s \in V \setminus S \\ t \in V \setminus S}} \frac{\sigma_{st}(S)}{\sigma_{st}}$.

Proposition 2. For any instance of Problem 1 defined by a graph $G = (V, E)$ and by a discrete probability distribution $PD(k)$, the exact solution to this problem is the Semivalue-based betweenness centrality defined by $p_{k-1} = \frac{PD(V, k)}{\binom{|V|-1}{k-1}}$ for all $k \in \{0, \dots, |V| - 1\}$.

Proof. To prove the proposition it suffices to transform Equation (5) to:

$$S_v(V, v) = \sum_{1 \leq k \leq |V|} PD(V, k) \frac{1}{\binom{|V|-1}{k-1}} \sum_{\substack{S \subseteq V \\ |S|=k \\ v \in S}} (v(S) - v(S \setminus \{v\})). \tag{8}$$

The above equation can be interpreted as follows. Each size $|S| = k$ is chosen with probability $PD(V, k)$, and each set $S \subseteq V$ such that $v \in S$ and $|S| = k$ is chosen with equal probability, i.e., $PD(V, k) \times \frac{1}{\binom{|V|-1}{k-1}}$. \square

From the above proposition, it follows that the Semivalue-based betweenness centrality measure is a generalization of the standard betweenness centrality measure (to see how this is the case, observe that when $PD(1) = 1$ and $PD(k) = 0, \forall k \in \{2, \dots, n\}$, the new measure coincides with the standard one). The above proposition also shows that the Semivalue-based betweenness centrality measure is much more flexible than the Shapley value-based measure as it allows for an arbitrary probability distribution on the sizes of subsets of nodes. This makes the measure applicable to a much wider spectrum of problems. For instance, in our motivating scenario of measuring systemic importance of financial institutions, with the Semivalue-based betweenness centrality we are able to assume an arbitrary probability distribution on the size of the bankrupting institution. Likewise, in the scenario of distributed network intrusion, we can choose an arbitrary probability distribution on the size of the infected hubs.

We note that the Semivalue-based between centrality can be normalized to the interval $[0, 1]$ just like we did earlier with the Shapley value-based betweenness centrality. Specifically, it suffices to replace $c_S(v)$ from Definition 6 with:

$$\frac{c_S(v)}{2 \max_{v \in V} c_b(v)} + \frac{1}{2}.$$

In the following subsections, we analyze the Shapley value-based betweenness centrality and the more-general Semivalue-based betweenness centrality.

4.4. A look at marginal contribution for the Shapley value

Recall that, as far as our measures are concerned, the centrality of a node is measured by computing a weighted average of the node’s marginal contributions to the group-betweenness centrality of all subsets of nodes. In this subsection, we show how to compute this weighted average efficiently.

To this end, we will start from Definition 3—the definition of the Shapley value. In particular, given a graph G and some node $v \in V(G)$, we would like to compute the expected marginal contribution of this node to the set of nodes $P_\pi(v)$ that precede v in a random permutation π consisting of all nodes of the graph. We will focus in our analysis on three exhaustive cases, where the marginal contribution is positive, negative, or neutral, respectively.

First, let us consider the positive contributions, focusing on some particular shortest path p that contains node v . Recall that σ_{st} denotes the number of shortest paths between nodes s and t , whereas $\sigma_{st}(v)$ denotes the number of such paths

that pass through v , where $t \neq v \neq s$ (each such path is said to be *controlled* by v). We say that node v makes a positive contribution to coalition $P_\pi(v)$ through a path p if and only if that path is controlled by v and not controlled by any member of $P_\pi(v)$; this way, once v joins $P_\pi(v)$, the resulting coalition will be able to control p . This positive contribution equals $\frac{1}{\sigma_{st}}$, because out of all the shortest paths between s and t , one additional path, namely p , will be added to the list of paths controlled by the coalition. Formally, the necessary and sufficient condition for such a positive contribution to occur is to have: $v \in \Psi(p)$ and $\Psi(p) \cap P_\pi(v) = \emptyset$, where $\Psi(p)$ is the set of all nodes lying on the path p , including endpoints.

Now, let us introduce a Bernoulli random variable $B_{v,p}^+$ that indicates whether node v makes a positive contribution to the set $P_\pi(v)$ through shortest path p . Thus, we have:

$$\mathbb{E}\left[\frac{1}{\sigma_{st}} B_{v,p}^+\right] = \frac{1}{\sigma_{st}} P[\Psi(p) \cap P_\pi(v) = \emptyset],$$

where $P[\cdot]$ denotes probability, and $\mathbb{E}[\cdot]$ denotes expected value. In other words, we need to know the probability of v preceding all the members of $\Psi(p) \setminus \{v\}$ in a random permutation of all nodes in the graph.

Theorem 1. Let K be a set of elements such that $|K| = k$. Furthermore, let L and R be two disjoint subsets of K , and let $|L| = l$ and $|R| = r$. Now, given some element $x \in K$, where $x \notin L \cup R$, and given a random permutation $\pi \in \Pi(K)$, the probability of having every element in L before x , and every element in R after x , is:

$$P[\forall e \in L \pi(e) < \pi(x) \wedge \forall e \in R \pi(e) > \pi(x)] = \frac{1}{(l+1) \binom{l+r+1}{r}}. \tag{9}$$

Proof. Let us first count the number of ways in which a permutation $\pi \in \Pi(K)$ can be constructed such that the following condition holds:

$$\forall e \in L \pi(e) < \pi(x) \wedge \forall e \in R \pi(e) > \pi(x).$$

To this end⁹:

- Let us choose $l+r+1$ positions in the permutation. There are $\binom{k}{l+r+1}$ such choices.
- Now, in the first l chosen positions, place all the elements of L . Directly after those, place the element x . Finally, in the remaining r positions, place all the elements of R . The number of such line-ups is: $l!r!$.
- As for the remaining elements (i.e., the elements in $K \setminus (L \cup \{x\} \cup R)$), they can be arranged in the permutation in $(k - (l+r+1))!$ different ways.

Thus, the number of permutations satisfying our condition is:

$$\binom{k}{l+r+1} l!r!(k - (l+r+1))! = \frac{k!}{(l+1) \binom{l+r+1}{r}}.$$

Finally, since we have a total of $k!$ permutations, the probability that one of them satisfies our condition is given by Equation (9). □

Based on Theorem 1, we can obtain the probability of an event in which the node v lies on the path p and precedes all the other nodes from this path in a random permutation of all nodes in the graph G . Now, by setting $K = V(G)$, $L = \emptyset$ and $R = \Psi(p) \setminus \{v\}$, we obtain the desired probability: $\frac{1}{|\Psi(p)|}$. Thus:

$$\mathbb{E}\left[\frac{1}{\sigma_{st}} B_{v,p}^+\right] = \frac{1}{\sigma_{st} |\Psi(p)|}. \tag{10}$$

Having discussed the first case, where marginal contributions are positive, let us now discuss the second case, where the marginal contribution of node v to set $P_\pi(v)$ is negative. This happens when path p ends or starts with v . Without loss of generality, we focus on the case where v is the end point. Specifically, if coalition $P_\pi(v)$ already controls path p (which includes node v), then not only is there no value added from v becoming a member of this coalition, but there is a negative effect of this move. In particular, since the group betweenness centrality assumes that a set of nodes S controls only those paths with both ends not belonging to S , then as soon as v joins $P_\pi(v)$, the coalition will lose the ability to control p . In so doing, v makes a negative contribution that is equal to $-\frac{1}{\sigma_{sv}}$.

Next, we analyze the probability of such a negative contribution of node v . Observe that if the marginal contribution is negative, then p must end with v . However, the opposite does not necessarily hold, i.e., if p ends with v then the marginal contribution of v is not necessarily negative; it could also be neutral. Based on this, in order to compute the probability of

⁹ We thank the anonymous reviewer for suggesting this formulation of the proof.

a negative contribution, we will start by focusing on the complementary event where v makes a neutral contribution to the set $P_\pi(v)$, as this simplifies the analysis. This complementary event happens if and only if either node s —the starting point of path p —belongs to set $P_\pi(v)$, or the path p is not controlled by any of the nodes in $P_\pi(v)$. Formally:

$$s \in P_\pi(v) \vee (\Psi(p) \cap P_\pi(v)) = \emptyset.$$

Based on this, we can introduce a Bernoulli random variable $B_{v,p}^-$ indicating whether node v makes a negative contribution through path p to the set $P_\pi(v)$ as follows:

$$\mathbb{E}\left[-\frac{1}{\sigma_{sv}} B_{v,p}^-\right] = -\frac{1}{\sigma_{sv}} (1 - P[s \in P_\pi(v) \vee (\Psi(p) \cap P_\pi(v)) = \emptyset]).$$

Again, using the combinatorial arguments presented in [Theorem 1](#), one can show that the probability $P[\Psi(p) \cap P_\pi(v) = \emptyset]$ equals $\frac{1}{|\Psi(p)|}$. Furthermore, due to symmetry, we have: $P[s \in P_\pi(v)] = \frac{1}{2}$. Finally, from the disjointness of the aforementioned complementary events, we have:

$$\mathbb{E}\left[-\frac{1}{\sigma_{sv}} B_{v,p}^-\right] = \frac{2 - |\Psi(p)|}{2\sigma_{sv}|\Psi(p)|}. \tag{11}$$

Recall that our analysis was divided into three exhaustive cases, where the marginal contribution is positive, negative, or neutral. Having dealt with the first two cases, it remains to deal with the case of neutral marginal contribution. As a matter of fact, since this case does not affect the expected marginal contribution of v to $P_\pi(v)$, we simply disregard it during the computation.

Next, let us compute the expected marginal contribution, by aggregating the cases of positive and negative contributions. To this end, let ∂_{st} be the set of all shortest paths from s to t , and analogously let $\partial_{st}(v) \subseteq \partial_{st}$ be the set of shortest paths from s to t that pass through node v .¹⁰ Now, using the expected value of the Bernoulli random variables (10) and (11) we are able to compute the Shapley value of node v , which is the expected marginal contribution of v to $P_\pi(v)$, as:

$$\begin{aligned} SV_v(V(G), v) &= \sum_{s \neq v \neq t} \sum_{p \in \partial_{st}(v)} \mathbb{E}\left[\frac{1}{\sigma_{st}} B_{v,p}^+\right] + \sum_{s \neq v} \sum_{p \in \partial_{sv}} \mathbb{E}\left[-\frac{1}{\sigma_{st}} B_{v,p}^-\right] \\ &= \sum_{s \neq v \neq t} \sum_{p \in \partial_{st}(v)} \frac{1}{\sigma_{st}|\Psi(p)|} + \sum_{s \neq v} \sum_{p \in \partial_{sv}} \frac{2 - |\Psi(p)|}{2\sigma_{sv}|\Psi(p)|}. \end{aligned} \tag{12}$$

Interestingly, the above equation provides insight into the Shapley value-based betweenness centrality; it is a combination of two factors: Firstly, the summation on the left resembles the standard betweenness centrality, but scaled by the number of nodes that belong to each shortest path. Secondly, the summation on the right resembles the closeness centrality,¹¹ but with distances measured as the number of nodes on the shortest paths.

4.5. A look at marginal contributions for the semivalue

We start our analysis by rewriting Equation (8) as:

$$S_v(V, v) = \sum_{1 \leq k \leq |V|} PD(V, k) \mathbb{E}[v(S_{k-1} \cup \{v\}) - v(S_{k-1})], \tag{13}$$

where $\mathbb{E}[\cdot]$ is the expected marginal contribution of node v to a random coalition S_{k-1} drawn uniformly at random from the set $\{S \subseteq V \setminus \{v\} : |S| = k - 1\}$. Now, analogously to the previous subsection, we need to analytically compute the probability that the node v contributes to a coalition through a shortest path p . Again, here we only focus on the cases of positive and negative contributions, since the neutral case has no impact on the expected marginal contribution.

Firstly, we consider the positive contributions, focusing on some particular shortest path p that contains v . Through this path, v makes a positive contribution to coalition S_{k-1} if and only if it is not controlled by (i.e., does not pass through) any node from set S_{k-1} . In this case, the positive contribution equals $\frac{1}{\sigma_{st}}$. The necessary and sufficient condition for this to happen is: $\Psi(p) \cap S_{k-1} = \emptyset$, where $\Psi(p)$ is the set of all nodes lying on the path p including endpoints.

Now, let us introduce a Bernoulli random variable $BS_{k-1,v,p}^+$ which indicates whether v makes a positive contribution through path p to coalition S_{k-1} . Note that, if $|V| - \Psi(p) < k - 1$ then the probability of this event equals 0. Otherwise, we have:

¹⁰ Note that $\sigma_{st} = |\partial_{st}|$ and $\sigma_{st}(v) = |\partial_{st}(v)|$.

¹¹ Recall that closeness centrality assigns to each node v the value $\sum_u \frac{1}{d(u,v)}$.

$$\begin{aligned} \mathbb{E}\left[\frac{1}{\sigma_{st}}BS_{k-1,v,p}^+\right] &= \frac{1}{\sigma_{st}}P[\Psi(p) \cap S_{k-1} = \emptyset] = \frac{1}{\sigma_{st}}\frac{\binom{|V|-\Psi(p)}{k-1}}{\binom{|V|-1}{k-1}} \\ &= \frac{(|V| - \Psi(p))!(|V| - k)!}{\sigma_{st}(|V| - \Psi(p) - k + 1)!(|V| - 1)!}, \end{aligned} \tag{14}$$

where $P[\cdot]$ denotes probability, and $\mathbb{E}[\cdot]$ denotes expected value. In other words, we need to know the probability that a random set S_{k-1} contains v and none of the nodes from $\Psi(p) \setminus \{v\}$.

Having analyzed the case of positive contribution, we now move on to the case of negative contribution. This happens when the path p ends with v and is controlled by the set S_{k-1} . As in the previous subsection, we will simplify the analysis by focusing on the complementary event in which the marginal contribution is neutral. Observe that v makes a neutral contribution to S_{k-1} through a path $p \in \partial_{sv}$ if and only if either node s —the end point of path p —belongs to S_{k-1} , or the path is not controlled by S_{k-1} . Formally:

$$s \in S_{k-1} \vee (\Psi(p) \cap S_{k-1}) = \emptyset.$$

Based on this, we can introduce a Bernoulli random variable $BS_{k-1,v,p}^-$ indicating whether node v makes a negative contribution to S_{k-1} through p . To this end, if $|V| - \Psi(p) \geq k - 1$, we obtain the following expression:

$$\begin{aligned} \mathbb{E}\left[-\frac{1}{\sigma_{sv}}BS_{k-1,v,p}^-\right] &= -\frac{1}{\sigma_{sv}}(1 - P[s \in S_{k-1} \vee (\Psi(p) \cap S_{k-1}) = \emptyset]) \\ &= -\frac{1}{\sigma_{sv}}\left(1 - \left(\frac{k-1}{|V|-1} + \frac{(|V| - \Psi(p))!(|V| - k)!}{(|V| - \Psi(p) - k + 1)!(|V| - 1)!}\right)\right) \\ &= \frac{k-1}{\sigma_{sv}(|V|-1)} + \frac{(|V| - \Psi(p))!(|V| - k)!}{\sigma_{sv}(|V| - \Psi(p) - k + 1)!(|V| - 1)!} - \frac{1}{\sigma_{sv}}. \end{aligned} \tag{15}$$

On the other hand, if $|V| - \Psi(p) < k - 1$, we obtain the following expression:

$$\mathbb{E}\left[-\frac{1}{\sigma_{sv}}BS_{k-1,v,p}^-\right] = \frac{k-1}{\sigma_{sv}(|V|-1)} - \frac{1}{\sigma_{sv}}. \tag{16}$$

Now, let us compute the expected marginal contribution, by aggregating the cases of positive and negative contributions. To this end, recall that ∂_{st} is the set of all shortest paths from s to t , whereas $\partial_{st}(v) \subseteq \partial_{st}$ is the set of shortest paths from s to t passing through v . Now, using the expected value of Bernoulli random variables (14), (15) and (16) we are able to compute the Semivalue of node v as per Equation (13) as follows:

$$\begin{aligned} S_v(V(G), v) &= \sum_{1 \leq k \leq |V|} PD(V, k) \left(\sum_{s \neq v \neq t} \sum_{p \in \partial_{st}(v)} \mathbb{E}\left[\frac{1}{\sigma_{st}}BS_{k-1,v,p}^+\right] + \sum_{s \neq v} \sum_{p \in \partial_{sv}} \mathbb{E}\left[-\frac{1}{\sigma_{st}}BS_{k-1,v,p}^-\right] \right) \\ &= \sum_{1 \leq k \leq |V|} PD(V, k) \left(\sum_{s \neq v \neq t} \sum_{\substack{p \in \partial_{st}(v) \\ |V| - \Psi(p) \geq k - 1}} \frac{(|V| - \Psi(p))!(|V| - k)!}{\sigma_{st}(|V| - \Psi(p) - k + 1)!(|V| - 1)!} \right. \\ &\quad \left. + \sum_{s \neq v} \left(\frac{k-|V|}{|V|-1} + \sum_{\substack{p \in \partial_{sv} \\ |V| - \Psi(p) \geq k - 1}} \frac{(|V| - \Psi(p))!(|V| - k)!}{\sigma_{sv}(|V| - \Psi(p) - k + 1)!(|V| - 1)!} \right) \right). \end{aligned} \tag{17}$$

Although equations (12) and (17) already allow us to efficiently compute in polynomial time the Shapley value-based and Semivalue-based betweenness centralities, respectively, in the following section we introduce algorithms that perform this computation even more efficiently.

5. Algorithms to compute the Shapley value and semivalue based betweenness centrality measures

Generally speaking, given a graph G , computing the Shapley value takes $O(2^{|V(G)|})$ time (to consider all subsets of $V(G)$). To circumvent this major obstacle, we proposed equations (12) and (17) to compute in polynomial time the Shapley Value and the Semivalue, respectively. To speed up the computation even further, we propose in this section four algorithms:

1. Algorithm **SVB** computes the *Shapley value*-based betweenness centrality measure given an *unweighted* graph;
2. Algorithm **SB** computes the *Semivalue*-based betweenness centrality measure given an *unweighted* graph;
3. Algorithm **WSVB** computes the *Shapley value*-based betweenness centrality measure given a *weighted* graph;
4. Algorithm **WSB** computes the *Semivalue*-based betweenness centrality measure given a *weighted* graph.

We also show in this section how the above algorithms can be easily adapted to work on directed graphs. More important, we show that Algorithm **SVB** has the same complexity as the best known algorithm to compute the standard betweenness centrality measure (due to Brandes [15]). Our algorithms are based on a general framework proposed in this article, which generalizes the framework of Brandes [15].

This section is structured as follows. Section 5.1 introduces our generalization of Brandes' framework for unweighted graphs. Building upon it, Sections 5.2 and 5.3 introduce Algorithm **SVB** and Algorithm **SB**, respectively. After that, in Section 5.4, we introduce our generalization of Brandes' framework for weighted graphs. Finally, Sections 5.5 and 5.6 introduce Algorithm **WSVB** and Algorithm **WSVB**, respectively.

5.1. The framework for unweighted graphs

In order to compute the standard betweenness centrality measure for all nodes, a naive algorithm would first compute the number of shortest paths between all pairs of nodes, and then for each node v sum up all pair-dependencies, which are defined as $\frac{\sigma_{st}(v)}{\sigma_{st}}$. This process takes $O(|V|^3)$ time. Brandes [15] proposed an algorithm to improve this complexity by using some recursive relation. This algorithm runs in $O(|V| \cdot |E|)$ time, and requires $O(|V| + |E|)$ space.

Next, we propose a framework that generalizes Brandes' algorithm. We start by defining pair-dependency as:

$$\delta_{s,t}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}} f_{\delta}(d(s, t)), \tag{18}$$

which is the positive contribution that all shortest paths between s and t make to the assessment of node v . The value of the function f_{δ} depends solely on the distance between s and t . For instance, if we set $f_{\delta} = \frac{1}{d(s,t)}$ we obtain the distance-scaled betweenness centrality measure introduced by Borgatti and Everett [14].

Next, we define one-side dependency as:

$$\delta_{s,\cdot}(v) = \sum_{t \in V} \delta_{s,t}(v), \tag{19}$$

which is the positive contribution that all shortest paths starting with node s make to the assessment of node v . Now, building upon the recursive equation proposed by Brandes [15], we obtain the following:

$$\delta_{s,\cdot}(v) = \sum_{\substack{w: (v,w) \in E \\ d(s,w)=d(s,v)+1}} \frac{\sigma_{sv}}{\sigma_{sw}} \left(f_{\delta}(d(s, w)) + \delta_{s,\cdot}(w) \right). \tag{20}$$

Using the above equation, we can compute what we call the *parameterized betweenness centrality measure* c_{fg} for a node v by iterating over all other nodes and summing their contributions as follows:

$$c_{fg}(v) = \sum_{s \neq v} \left(\delta_{s,\cdot}(v) + g_{\delta}(d(s, v)) \right), \tag{21}$$

where the value of the function g_{δ} depends solely on the distance between s and v .

Now, we are ready to introduce our generalized framework, **PBC** (see Algorithm 1) for computing the *parameterized betweenness centrality* in unweighted graphs. More specifically, this framework modifies Brandes' approach and computes the betweenness centrality measure parameterized by two functions: f_{δ} and g_{δ} . All lines that differ from Brandes' original algorithm have been highlighted. In our algorithm we use two basic data structures: queue \mathcal{Q} and stack \mathcal{S} . The operations on these collections of nodes are as follows: *push* adds an element to the top of the stack, *pull* removes the element at the top of the stack, *enqueue* adds an element at the end of the queue, and *dequeue* removes the element at the beginning of the queue.

Firstly, in lines 5–11, the algorithm calculates both the distance and the number of shortest paths from a source s to each node. These lines are exactly the same as in the original algorithm introduced by Brandes. While executing these lines, for each node v , all directly preceding nodes occurring on shortest paths from s to v are stored in memory. This process uses Breadth-First Search [21] which takes $O(|V| + |E|)$ time and $O(|V|)$ space. In the second step (lines 15 and 17), the algorithm uses Equation (21) to calculate the contribution of the source s to the value of our betweenness centrality for each node that is reachable from the source. This step also takes $O(|V|)$ time and $O(|V| + |E|)$ space. Only lines 15 and 17 and additionally line 18 differ from Brandes' original algorithm.

As visible from Equation (21), in an undirected graph, each path is considered twice. Thus, at the end of the algorithm, in line 18, we halve the accumulated result. In line 17, we multiply the influence of the function g_{δ} by two, because the influence of each source even in undirected graphs is considered only once. Finally, we note that it is very easy to adopt Algorithm 1 to directed graphs. To this end, we remove the loop from line 18 and halve the contribution of the node s from line 17, which now should look as follows:

$$17: c_{Sh}(w) \leftarrow c_{Sh}(w) + \delta_{s,\cdot}(w) + g_{\delta}(d(s, w));$$

So, for directed and undirected graphs, the algorithm runs in $O(|V| \cdot |E|)$ time, and requires $O(|V| + |E|)$ space.

Algorithm 1: PBC A general framework to compute the *parameterized betweenness centrality* in *unweighted graphs*.

Input: Graph $G = (V, E)$, functions: f_δ and g_δ
Data: queue Q , stack S for each $v \in V$ and some source s
 $d(s, v)$: distance from v to the source s
 $Pred_s(v)$: list of predecessors of v on the shortest paths from source s
 σ_{sv} : the number of shortest paths from s to v
 $\delta_{s,\cdot}(v)$: the one-side dependency of s on v
Output: $c_{fg}(v)$ parameterized betweenness centrality for each node $v \in V$

```

1 foreach  $s \in V$  do
2   foreach  $v \in V$  do
3      $Pred_s(v) \leftarrow$  empty list;  $d(s, v) \leftarrow \infty$ ;  $\sigma_{sv} \leftarrow 0$ ;
4    $d(s, s) \leftarrow 1$ ;  $\sigma_{ss} \leftarrow 1$ ;  $Q.enqueue(s)$ ;
5   while  $Q$  is not empty do
6      $v \leftarrow Q.dequeue()$ ;  $S.push(v)$ ;
7     foreach  $w$  such that  $(v, w) \in E$  do
8       if  $d(s, w) = \infty$  then
9          $d(s, w) \leftarrow d(s, v) + 1$ ;  $Q.enqueue(w)$ ;
10      if  $d(s, w) = d(s, v) + 1$  then
11         $\sigma_{sw} \leftarrow \sigma_{sw} + \sigma_{sv}$ ;  $Pred_s(w).append(v)$ ;
12  foreach  $v \in V$  do  $\delta_{s,\cdot}(v) \leftarrow 0$ ; while  $S$  is not empty do
13     $w \leftarrow S.pop()$ ;
14    foreach  $v \in Pred_s(w)$  do
15       $\delta_{s,\cdot}(v) \leftarrow \delta_{s,\cdot}(v) + \frac{\sigma_{sv}}{\sigma_{sw}}(f_\delta(d(s, w)) + \delta_{s,\cdot}(w))$ ;
16    if  $w \neq s$  then
17       $c_{fg}(w) \leftarrow c_{fg}(w) + \delta_{s,\cdot}(w) + 2g_\delta(d(s, w))$ ;
18 foreach  $v \in V$  do  $c_{fg}(v) = \frac{c_{fg}(v)}{2}$ ;

```

5.2. The algorithm for the Shapley value-based betweenness centrality measure for unweighted graphs

In this subsection we will construct an efficient algorithm for computing the Shapley value-based betweenness centrality measure for unweighted graphs. Specifically, in such graphs, the number of nodes on the shortest path between s and t is equal to the distance between s and t , denoted by $d(s, t)$.¹² In other words, we have: $|\Psi(p)| = d(s, t)$. Based on this, it is possible to simplify (12) as follows:

$$\begin{aligned}
 SV_v(V(G), v) &= \sum_{s \neq v \neq t} \sum_{p \in \partial_{st}(v)} \frac{1}{\sigma_{st} d(s, t)} + \sum_{s \neq v} \sum_{p \in \partial_{sv}} \frac{2 - d(s, v)}{2\sigma_{sv} d(s, v)} \\
 &= \sum_{s \neq v} \left(\sum_{t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st} d(s, t)} + \frac{2 - d(s, v)}{2d(s, v)} \right). \tag{22}
 \end{aligned}$$

The above equation provides the same insight as Equation (12) but for specific unweighted graphs. By transforming the second element of the inner sum $\frac{2-d(s,v)}{2d(s,v)} = \frac{1}{d(s,v)} - \frac{1}{2}$ we find the following: In unweighted graphs, the Shapley value of the corresponding game (whose characteristic function assigns to each coalition its group betweenness centrality) is in fact the sum of the *distanced-scaled betweenness centrality measures* (introduced by Borgatti and Everett [14]) and the *closeness centrality*, shifted by half. Additionally, the above equation allows us to compute the Shapley value-based betweenness centrality measure in $O(|V|^3)$, but this result will be improved further.

Now, we adopt the framework presented in the previous subsection in order to accommodate Equation (22). We simply need to set $f_\delta = \frac{1}{d(s,t)}$ and set $g_\delta = \frac{2-d(s,v)}{2d(s,v)}$. This way, we can use our general framework—**PBC**, see Algorithm 1—to compute the Shapley value-based betweenness centrality measure in $O(|V||E|)$. This method is presented in Algorithm 2.

5.3. The algorithm for the semivalue-based betweenness centrality measure for unweighted graphs

We will follow the same reasoning as in the previous subsection. Specifically, since in unweighted graphs it holds that $|\Psi(p)| = d(s, t)$, we can transform Equation (17) into:

¹² For notational convenience, we assume the distance between two nodes to be the number of nodes on the shortest path between them (not the number of edges), e.g., $d(s, s) = 1$.

Algorithm 2: SVB The Shapley value-based betweenness centrality.

Input: Graph $G = (V, E)$
Output: $c_{Sh}(v)$ the Shapley value-based betweenness centrality for each node $v \in V$
1 $f_\delta(d) \leftarrow \frac{1}{d}$;
2 $g_\delta(d) \leftarrow \frac{2-d}{2d}$;
3 **PBC**(G, f_δ, g_δ)

Algorithm 3: SB Semivalue-based betweenness centrality.

Input: Graph $G = (V, E)$
Output: $c_S(v)$ Semivalue-based betweenness centrality for each node $v \in V$
1 $f_\delta(d) \leftarrow$ **if** $|V| - d \geq k - 1$ **then** $\frac{(|V|-d)! (|V|-k)!}{(|V|-d-k+1)! (|V|-1)!}$ **else** 0 ;
2 $g_\delta(d) \leftarrow f_\delta(d) + \frac{k-|V|}{|V|-1}$;
3 **for** $k \leftarrow 1$ **to** $|V|$ **do**
4 **PBC**(G, f_δ, g_δ) ;
5 **foreach** $v \in V$ **do**
6 $c_S(v) \leftarrow c_S(v) + PD(V, k)c_{f_g}(v)$;

$$\begin{aligned}
 S_v(V(G), v) &= \sum_{1 \leq k \leq |V|} PD(V, k) \left(\sum_{s \neq v \neq t} \sum_{\substack{p \in \partial_{st}(v) \\ |V|-d(s,t) \geq k-1}} \frac{(|V|-d(s,t))!(|V|-k)!}{\sigma_{st}(|V|-d(s,t)-k+1)! (|V|-1)!} \right. \\
 &\quad \left. + \sum_{s \neq v} \left(\frac{k-|V|}{|V|-1} + \sum_{\substack{p \in \partial_{sv} \\ |V|-d(s,v) \geq k-1}} \frac{(|V|-d(s,v))!(|V|-k)!}{\sigma_{sv}(|V|-d(s,v)-k+1)! (|V|-1)!} \right) \right) \\
 &= \sum_{1 \leq k \leq |V|} PD(V, k) \left(\sum_{\substack{s \neq v \neq t \\ |V|-d(s,t) \geq k-1}} \frac{\sigma_{st}(v)(|V|-d(s,t))!(|V|-k)!}{\sigma_{st}(|V|-d(s,t)-k+1)! (|V|-1)!} \right. \\
 &\quad \left. + \sum_{\substack{s \neq v \\ |V|-d(s,v) \geq k-1}} \left(\frac{(|V|-d(s,v))!(|V|-k)!}{(|V|-d(s,v)-k+1)! (|V|-1)!} \right) + k - |V| \right). \tag{23}
 \end{aligned}$$

Our framework can be easily adopted to deal with Semivalues; all we need to do is to set:

$$f_\delta = \begin{cases} \frac{(|V|-d)! (|V|-k)!}{(|V|-d-k+1)! (|V|-1)!} & \text{if } |V| - d(s, t) \geq k - 1 \\ 0 & \text{if } |V| - d(s, t) < k - 1 \end{cases}$$

and set $g_\delta = f_\delta + \frac{k-|V|}{|V|-1}$. This way, we can use Equation (23) along with our general framework, **PBC**, to compute the Semivalue-based betweenness centrality measure for unweighted graphs in $O(|V|^2|E|)$ time. The pseudo-code is provided in **Algorithm 3**.

5.4. The framework for weighted graphs

While the focus of the previous subsections was on unweighted graphs, in this subsection we show how to generalize our framework to deal with weighted graphs. In particular, we consider one of the most popular semantics of weighted graphs, where the weight $\lambda(v, u)$ of the edge between v and u is interpreted as the distance between v and u . Thus, unlike the case with unweighted graphs, where $|\Psi(p)| = d(s, t)$, in the case of weighted graphs this equality does not necessarily hold.

To introduce our generalized framework, we need additional notation. Let $T_{st}[i] : i \in \{1, \dots, |V|\}$ be the number of shortest paths between s and t that contain exactly i nodes. The array T_{st} uniquely determines the polynomial W_{st} with terms $T_{st}[i]x^i$. We define the following operations on T_{st} :

- **Shifting:** T_{st}^{\rightarrow} and T_{st}^{\leftarrow} increase and decrease the indices of all values of the array by one, respectively. This takes $O(|V|)$ time.
- **Adding:** $T_{sv} \oplus T_{su}$ is the operation of adding two polynomials W_{sv} and W_{su} ; it takes $O(|V|)$ time. We will denote by \oplus the sum of a series of polynomials.
- **Multiplying:** $T_{sv} \otimes T_{vt}$ is the operation of multiplying two polynomials W_{sv} and W_{vt} . This takes $O(|V| \log |V|)$ time using the polynomial multiplying algorithm from Cormen [21].
- **Resetting:** $T_{sv} \leftarrow 0$ is an operation that assigns 0 to each cell in T_{sv} .

Next, we will show how the above operations allow us to tackle two algorithmic problems: (i) how to count all shortest paths and (ii) how to derive the recursive relation in order to compute one-side dependency. We start by considering the first of those problems, i.e., counting the shortest paths. Here, we will use the following relation:

$$T_{sv} = \bigoplus_{u: d(s,u)+\lambda(u,v)=d(s,v)} T_{su}^{\rightarrow}. \tag{24}$$

Using Dijkstra’s algorithm [21], as well as Equation (24), we can compute T_{st} for every pair of nodes s and t . If node u immediately precedes node v on some shortest path from source s , all shortest paths stored in T_{su} extended by node v are part of the set of shortest paths stored in T_{sv} . This procedure takes $O(|V|^2|E| + |V|^2 \log |V|)$ time.

In order to count all paths passing through a given node v , we will introduce the following relationship, where $T_{st}(v)$ is an array defined just like T_{st} except that it only counts the shortest paths between nodes s and t that pass through v :

$$T_{st}(v) = (T_{sv} \otimes T_{vt})_{st}^{\leftarrow} = T_{sv} \otimes T_{vt}^{\leftarrow}. \tag{25}$$

The above equation can be interpreted as follows. Every path stored in the array T_{sv} can be extended by every path stored in the array T_{vt} . The outcome of this operation, which is in fact the multiplication of the two polynomials W_{sv} and W_{vt} , gives us information about all shortest paths from s to t passing through v . Note that the node v is counted twice. To avoid such duplicate counting, we shift the result of multiplication to the left, which effectively reduces the number of nodes in each path by one.

In weighted graphs, the pair-dependency limited to the shortest paths consisting of i nodes is defined as follows:

$$\delta_{s,t}^*(v)[i] = \frac{T_{st}(v)[i]}{\sigma_{st}} f_{\delta^*}(i). \tag{26}$$

This limited pair-dependency measures the influence of all shortest paths between s and t consisting of i nodes on the evaluation of node v . Note that the value of the function f_{δ^*} depends solely on the number of nodes lying on the shortest paths between s and t .

Taking all possible values of i into consideration, we obtain the pair-dependency in weighted graphs:

$$\delta_{s,t}^*(v) = \sum_{i=1}^{|V|} \delta_{s,t}^*(v)[i], \tag{27}$$

which is the positive contribution that all shortest paths between s and t make to the assessment of node v .

The definition of one-side dependency is similar to the one presented earlier in Subsection 5.1, except that we replace δ with δ^* :

$$\delta_{s,\cdot}^*(v) = \sum_{t \in V} \delta_{s,t}^*(v). \tag{28}$$

This is the positive contribution that all shortest paths starting with node s make to the evaluation of node v . The version of Equation (28) that only considers the shortest paths containing exactly i nodes each is:

$$\delta_{s,\cdot}^*(v)[i] = \sum_{t \in V} \delta_{s,t}^*(v)[i]. \tag{29}$$

Now, we are able to infer the recursive relation for weighted graphs:

$$\delta_{s,\cdot}^*(v)[i] = \sum_{\substack{w: (v,w) \in E \\ d(s,w)=d(s,v)+1}} \frac{T_{sv}[i]}{\sigma_{sw}} \left(f_{\delta^*}(i+1) + \delta_{s,\cdot}^*(w)[i+1] \right). \tag{30}$$

Using the above equation, we are able to compute a parameterized betweenness centrality for weighted graphs c_{fg}^* . Specifically, for a node v , we compute $c_{fg}^*(v)$ by iterating over all other nodes and summing their contributions. More formally:

$$c_{fg}^*(v) = \sum_{s \neq v} \sum_{i=1}^{|V|} \left(\delta_{s,\cdot}^*(v)[i] + \frac{T_{sv}[i]}{\sigma_{sv}} g_{\delta^*}(i) \right), \tag{31}$$

where the value of the function g_{δ^*} depends solely on the number of nodes lying on the shortest paths between s and v .

Now, we are ready to introduce our general framework, namely **WPBC** (see Algorithm 4) to compute the *parameterized betweenness centrality measure* in weighted graphs. All lines that differ from the original algorithm of Brandes have been highlighted. In lines 4–13, our algorithm uses Dijkstra’s algorithm [21] to traverse the graph and count all shortest paths from the source s . In so doing, our algorithm differs from the algorithm of Brandes, because ours uses polynomial arithmetic

Algorithm 4: WPBC A general framework to compute the *parameterized betweenness centrality measure* in *weighted graphs*.

Input: weighted graph $G = (V, E)$, with weight function $\lambda : E \rightarrow \mathbb{R}^+$
Data: priority queue \mathcal{Q} with key $d()$, stack \mathcal{S}
 $d(s, v)$: the distance from s to v
 $Pred_s(v)$: the list of predecessors of v on the shortest paths from source s
 σ_{sv} : the number of shortest paths from s to v
 $\delta_{s,\cdot}^*(v)[i]$: one-side dependency of s on v
 $T[i]_{sv}$: the number of shortest paths from s to v containing i nodes
Output: $c_{fg}^*(v)$ parameterized betweenness centrality

```

1 foreach  $s \in V$  do
2   foreach  $v \in V$  do  $Pred_s(v) \leftarrow$  empty list;  $d(s, v) \leftarrow \infty$ ;  $\sigma_{sv} \leftarrow 0$ ;
3    $d(s, s) \leftarrow 1$ ;  $\sigma_{ss} \leftarrow 1$ ;  $T_{ss}[1] \leftarrow 1$ ;  $\mathcal{Q}.enqueue(s)$ ;
4   while  $\mathcal{Q}$  is not empty do
5      $v \leftarrow \mathcal{Q}.extract()$ ; with minimal  $d(s, v)$ ;  $\mathcal{S}.push(v)$ ;
6     foreach  $w$  such that  $(v, w) \in E$  do
7       if  $d(s, w) > d(s, v) + \lambda(v, w)$  then
8          $d(s, w) \leftarrow d(s, v) + \lambda(v, w)$ ;
9          $\mathcal{Q}.insert/update(w)$  with  $d(s, w)$ ;
10         $\sigma_{sw} \leftarrow 0$ ;  $T_{sw} \leftarrow 0$ ;  $Pred_s(w) \leftarrow$  empty list;
11        if  $d(s, w) = d(s, v) + \lambda(v, w)$  then
12           $\sigma_{sw} \leftarrow \sigma_{sw} + \sigma_{sv}$ ;  $T_{sw} = T_{sw} \oplus T_{sv}^*$ ;
13           $Pred_s(w).append(v)$ ;
14        foreach  $v \in V$  do  $\delta_{s,\cdot}^*(v) \leftarrow 0$ ; while  $\mathcal{S}$  is not empty do
15           $w \leftarrow \mathcal{S}.pop()$ ;
16          foreach  $v \in Pred_s(w)$  do
17            for  $i \leftarrow 1$  to  $|V| - 1$  do
18               $\delta_{s,\cdot}^*(v)[i] \leftarrow \delta_{s,\cdot}^*(v)[i] + \frac{T_{sv}[i]}{\sigma_{sw}} \left( f_{\delta^*}(i+1) + \delta_{s,\cdot}^*(w)[i+1] \right)$ ;
19            if  $w \neq s$  then for  $i \leftarrow 1$  to  $|V|$  do
20               $c_{fg}^*(w) \leftarrow c_{fg}^*(w) + \delta_{s,\cdot}^*(w)[i] + 2 \frac{T_{sw}[i]}{\sigma_{sw}} g_{\delta^*}^*(i)$ ;
21 foreach  $v \in V$  do  $c_{fg}^*(v) = \frac{c_{fg}^*(v)}{2}$ ;

```

to count paths and count the number of nodes lying on those paths. Next, in lines 18 and 20, which also differ from Brandes' framework, we implemented the recursive formula (30).

Algorithm 4 runs in $O(|V|^2|E| + |V|^2 \log |V|)$ time and requires $O(|V|^2)$ space. Furthermore, this algorithm (just like Algorithm 1) can be easily adapted to directed graphs.

5.5. The algorithm for Shapley value-based betweenness centrality measure in weighted graphs

In this subsection we will construct an efficient algorithm for computing the Shapley Value-based betweenness centrality measure for weighted graphs. We will use the notation introduced in the previous subsection, and transform Equation (12) into:

$$SV_v(V(G), v) = \sum_{s \neq v} \left(\sum_{t \neq v} \sum_{i=1}^{|V|} \frac{T_{st}(v)[i]}{\sigma_{st} i} + \sum_{i=1}^{|V|} \frac{T_{sv}[i](2-i)}{2\sigma_{sv} i} \right). \tag{32}$$

The framework for weighted graphs (introduced in Section 5.4) can be used to compute (32). All we need is to define $f_{\delta^*}^*(i) = \frac{1}{i}$ and $g_{\delta^*}^*(i) = \frac{2-i}{2i}$. By doing so, we can use our general framework—WPBC, see Algorithm 4—to compute the weighted Shapley value-based betweenness centrality measure for all nodes in $O(|V|^2|E| + |V|^2 \log |V|)$. This method is presented in Algorithm 5.

5.6. The algorithm for semivalue-based betweenness centrality measure in weighted graphs

In this subsection we use our framework for weighted graphs, namely WPBC (see Algorithm 4), to compute the Semivalue-based betweenness centrality measure for weighted graphs. To this end, we can transform Equation (17) into:

Algorithm 5: WSVB The Shapley value-based betweenness centrality for weighted graphs.

Input: weighted graph $G = (V, E)$, with weight function $\lambda : E \rightarrow \mathbb{R}^+$
Output: $c_{Sh}^*(v)$ Shapley value-based betweenness centrality for each $v \in V$

- 1 $f_\delta^*(i) \leftarrow \frac{1}{i}$;
- 2 $g_\delta^*(i) \leftarrow \frac{2-i}{2i}$;
- 3 **WPBC**($G, \lambda, f_\delta^*, g_\delta^*$)

Algorithm 6: WSB Weighted semivalue-based betweenness centrality.

Input: weighted graph $G = (V, E)$, with weight function $\lambda : E \rightarrow \mathbb{R}^+$
Output: $c_S^*(v)$ weighted Semivalue-based betweenness centrality for $v \in V$

- 1 $f_\delta^*(i) \leftarrow$ **if** $|V| - i \geq k - 1$ **then** $\frac{(|V|-i)!(|V|-k)!}{(|V|-i-k+1)!(|V|-1)!}$ **else** 0 ;
- 2 $g_\delta^*(i) \leftarrow f_\delta^*(i) + \frac{k-|V|}{|V|-1}$;
- 3 **for** $k \leftarrow 1$ **to** $|V|$ **do**
- 4 **WPBC**($G, \lambda, f_\delta^*, g_\delta^*$) ;
- 5 **foreach** $v \in V$ **do**
- 6 $c_S^*(v) \leftarrow c_S^*(v) + PD(V, k)c_{fg}^*(v)$;

$$S_v(V(G), v) = \sum_{1 \leq k \leq |V|} PD(V, k) \left(\sum_{s \neq v \neq t}^{|V|-i \geq k-1} \frac{T_{st}(v)[i](|V|-i)!(|V|-k)!}{\sigma_{st}(|V|-i-k+1)!(|V|-1)!} \right. \\ \left. + \sum_{s \neq v} \left(\frac{k-|V|}{|V|-1} + \sum_{i=1}^{|V|-i \geq k-1} \frac{T_{sv}[i](|V|-i)!(|V|-k)!}{\sigma_{sv}(|V|-i-k+1)!(|V|-1)!} \right) \right). \quad (33)$$

Our framework can be easily adopted to deal with Semivalues; all we need is to set:

$$f_\delta^*(i) = \begin{cases} \frac{(|V|-i)!(|V|-k)!}{(|V|-i-k+1)!(|V|-1)!} & \text{if } |V| - i \geq k - 1 \\ 0 & \text{if } |V| - i < k - 1 \end{cases}$$

and set $g_\delta^* = f_\delta^* + \frac{k-|V|}{|V|-1}$. This way, we can use our general framework for weighted graphs, namely **WPBC** (see [Algorithm 4](#)) to compute the Semivalue-based betweenness centrality in $O(|V|^3|E| + |V|^3 \log |V|)$ time. The pseudo-code is presented in [Algorithm 6](#).

6. Empirical evaluation

So far, we argued that our extensions of betweenness centrality are better tools of network analysis than the standard betweenness centrality measure in situations in which one needs to consider centrality of various groups of nodes and not only individual nodes. Furthermore, we presented polynomial time algorithms to compute our centrality measures in weighted and unweighted graphs. This section provides an empirical evaluation of these contributions.

In more detail, we study the resilience of a network against random simultaneous node failures (see [Problem 2](#) in [Section 4.1.2](#)), where the level of protection of each node is assumed to be determined by the node's ranking (i.e., nodes with greater ranking received a greater level of protection). In our simulations, we consider all four measures of network performance, i.e., IGM, CC, FR, and LC, outlined in [Section 4.1.2](#). For each measure, we compare the node ranking obtained by our Semivalue-based betweenness centrality measure against the ranking obtained by the standard betweenness centrality measure. In so doing, we also implicitly evaluate our Shapley value-based betweenness centrality measure, since it is a special case of the Semivalue-based one.

We start by describing the simulation setup in [Section 6.1](#). After that, in [Section 6.2](#), we experiment with four unweighted real-life networks: (i) the karate-club network in [\[69\]](#), (ii) the proteins network in [\[12\]](#), (iii) the American football teams network in [\[34\]](#), and (iv) the jazz musicians network in [\[35\]](#). In [Section 6.3](#), we consider a subnetwork of the financial institution network studied by Soramaki and Cook [\[60\]](#). Finally, in [Section 6.4](#), we consider two types of random graphs: (i) unweighted random scale-free graphs, generated using the *preferential attachment* mechanism of Barabasi and Albert [\[10\]](#), and (ii) binomial random graphs generated using the model of Erdos and Renyi [\[27\]](#).

6.1. Simulation setup

Each experiment consists of two phases. The first phase involves the following steps:

1. Upload a real-life network, or generate an unweighted random scale-free network, or a binomial random network.

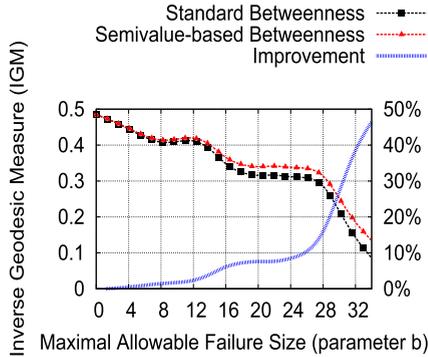


Fig. 2. The difference in performance in terms of IGM on the Zachary karate club network.

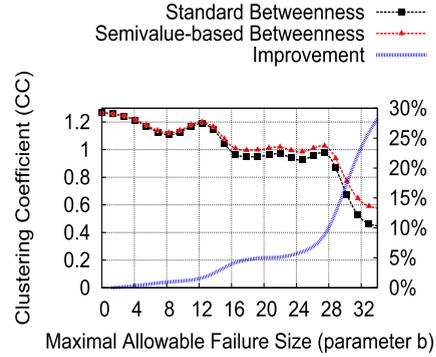


Fig. 3. The difference in performance in terms of CC on the Zachary karate club network.

2. Set the interval $[a, b)$; this means that the number of nodes to be simultaneously exposed to failure is between a (inclusive) and b (exclusive).
3. For every node $v \in V$, compute its normalized centrality, denoted $c(v)$.¹³ Depending on the experiment, this normalized centrality is computed either using the standard betweenness centrality measure, or using our Semivalued-based betweenness centrality measure with the following probability distribution:

$$PD(V, k) = \begin{cases} \frac{1}{b-a} & \text{if } k \in [a, b) \\ 0 & \text{otherwise.} \end{cases}$$

The normalized centralities of all nodes can be interpreted as a ranking of nodes, where the greater the normalized centrality of a node, the greater its ranking.

The second phase of the experiment is divided into the following steps:

1. Protect the nodes in the network. Here, for each node, the level of protection is determined based on its ranking; the greater the ranking, the higher the probability of protecting the node. More specifically, node v is protected with probability $\min(c(v) \times pr, 1)$, where pr is a constant that either increases or decreases the level of protection.¹⁴ In all our experiments, we assume that $pr = 10\% \times |V|$, which intuitively means we have limited resources that are sufficient to protect only 10% of the nodes. Note that we test the cardinal (instead of just the ordinal) ranking of nodes.
2. Choose randomly a set of nodes $S \subseteq V$, which will be exposed to failure.¹⁵ In all our experiments, we choose 10,000 such random sets. The size of each such set is chosen uniformly at random from the range $[a, b)$, where we set $a = 1$ and $b = 1, 2, \dots, |V|$. The members of the set are also chosen uniformly at random.
3. For every exposed node $v \in S$, the probability of failure is assumed to be: $\min(c(v) \times pr, 1)$.

In the next section, we begin our experiments by considering four real-life networks.

6.2. Real-life networks

The first experiment is performed on the well-known Zachary club network which involves 34 members of a karate club along with their social relations [69]. Each of the four measures of network performance, i.e., IGM, CC, FR, and LC, is presented on a separate figure (see Section 4.1.2 for a description of these measures). More specifically, Fig. 2 presents IGM, Fig. 3 presents CC, Fig. 4 presents FR, and Fig. 5 presents LC. In every such figure¹⁶:

- The black line depicts the network-performance measure when the network protection is based on the standard betweenness centrality measure (we denote this by M_B);
- The red line depicts the network-performance measure when the network protection is based on the Semivalued-based centrality measure (we denote this by M_S);
- The blue line depicts the relative improvement, computed as: $(M_S - M_B)/M_B$.

¹³ We normalize the centralities to ensure that: $\forall v \in V \ c(v) \in [0, 1]$ and $\sum_{v \in V} c(v) = 1$.

¹⁴ Note that the normalized centralities, $c(v) : v \in V$, on average become smaller as the number of nodes increases. This is offset by pr .

¹⁵ In our setting, being exposed to failure does not necessarily imply that the node will fail; this depends on the level of protection that the node has against failure.

¹⁶ For interpretation of the colors in these figures, the reader is referred to the web version of this article.

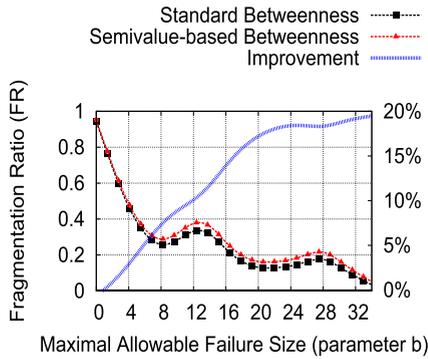


Fig. 4. The difference in performance in terms of FR on the Zachary karate club network.

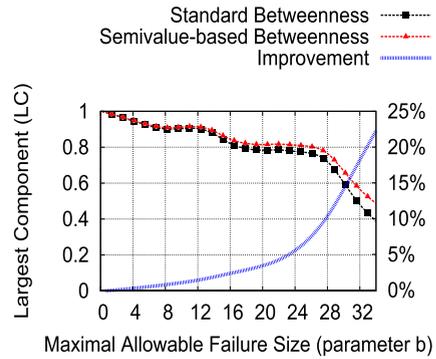


Fig. 5. The difference in performance in terms of LC on the Zachary karate club network.

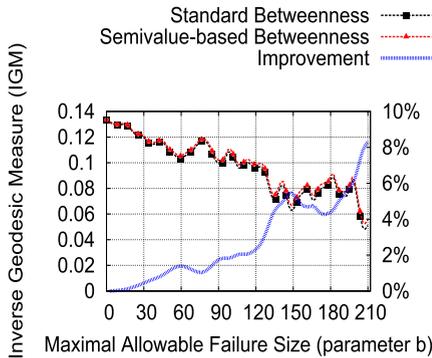


Fig. 6. The difference in performance in terms of IGM on the protein network.

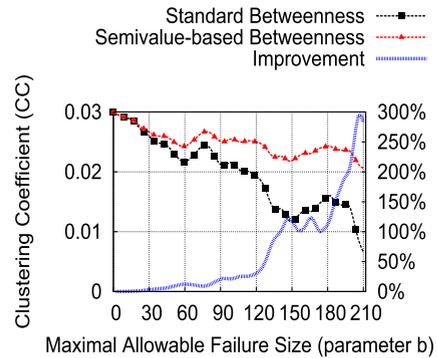


Fig. 7. The difference in performance in terms of CC on the protein network.

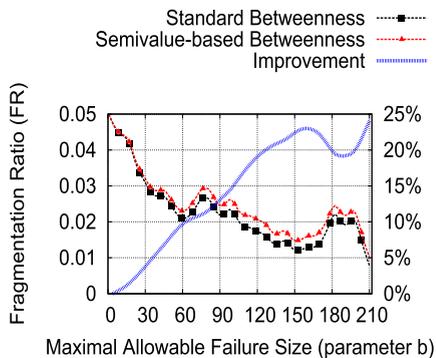


Fig. 8. The difference in performance in terms of FR on the protein network.

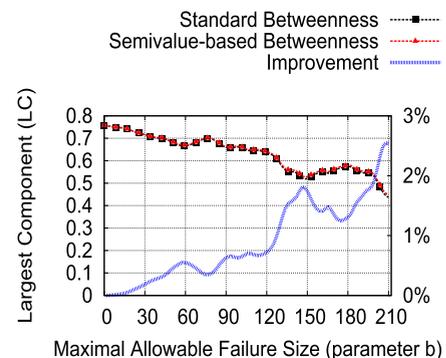


Fig. 9. The difference in performance in terms of LC on the protein network.

Note that the left vertical axis concerns both M_B and M_S , whereas the right one concerns the relative improvement.

As can be seen, when b is small, the performance of the Semivalued-based betweenness centrality measure is almost identical to the standard betweenness centrality. This is expected since our centrality measure, given the failure interval $[1, 2)$, is exactly the same as the standard measure. However, as we increase b , more nodes can fail simultaneously, and so the difference between our measure and the standard one becomes more evident. As can be seen, in such cases, the new measure outperforms the standard one in terms of network protection under all four network performance measures, with the relative improvement reaching up to 45%, depending on the network-performance measure.

The second experiment is carried out on the proteins networks [12] (see Figs. 6–9). This is a sparse network of 212 proteins and 244 relations between them. Again, our measure outperforms the standard betweenness centrality measure in terms of network protection, with the relative improvement reaching up to 300%.

The last two experiments in this subsection are performed on the American football teams network [34] and the jazz musicians network [35]. The first one consists of 115 nodes (football teams), where each edge represents a regular season

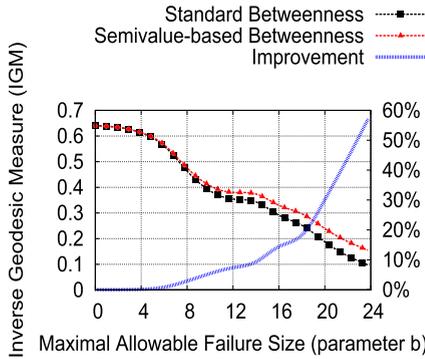


Fig. 10. The difference in performance in terms of IGM on the bank network.

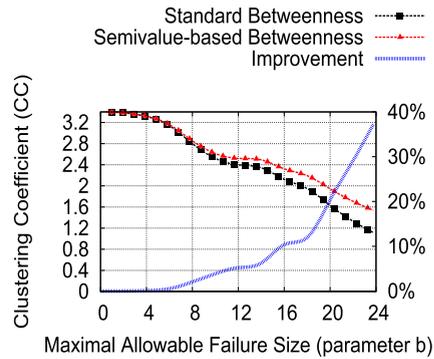


Fig. 11. The difference in performance in terms of CC on the bank network.

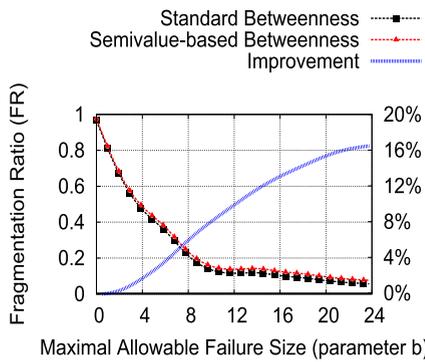


Fig. 12. The difference in performance in terms of FR on the bank network.

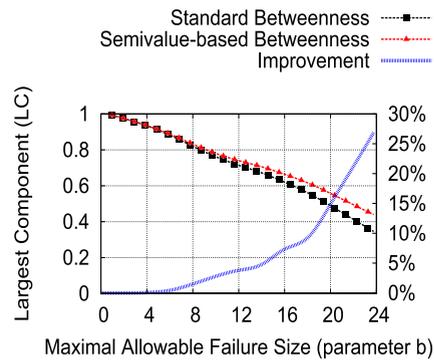


Fig. 13. The difference in performance in terms of LC on the bank network.

game. In the second network, each of the 198 nodes corresponds to a jazz musician, and two artists are linked if they have played in the same band. For the American football network we achieve the following relative improvements¹⁷: 3.6% for the CC measure, 11.9% for the FR measure, 4.5% for the IGM measure, and 2.5% for the LC measure. As for the jazz musicians network, we reach 3.1% for CC, 43.5% for FR, 3.8% for IGM, and 1.9% for LC.

Finally, note that in all our experiments we compared the performance of the Semivalued-based betweenness centrality with the standard betweenness centrality and showed that the former measure outperforms the latter one for all four performance measures. Naturally, one can try to develop dedicated methods to optimize each performance measure separately. Such an analysis is however well beyond the scope of this article.

6.3. A simulated network of inter-bank exposures

Perhaps the most important challenge facing a supervisory authority such as the Financial Stability Board is to maintain post-crisis connectives of the financial system. With this in mind, the experiment in this subsection involves the 25-bank subnetwork of a simulated financial network from the work by Soramaki and Cook [60].¹⁸ This network was generated using a model similar to the preferential attachment model of Barabasi and Albert [10] except for some modifications that reproduce the main statistical properties of a real financial network, namely the Fedwire network in the USA. The resulting network is an undirected one, where a link between two banks indicates that the two are connected by significant capital flows.

The results of our experiments are depicted in Figs. 10 to 13. As can be seen, the Semivalued-based betweenness centrality measure outperforms the standard betweenness centrality measure by about 16%, 25%, 55% and 35% in terms of the FR, LC, IGM, and CC network-performance measures, respectively.

¹⁷ Note that the corresponding figures have been omitted due to space constraints.

¹⁸ We thank Sadoghi Amirhossein from the Frankfurt School of Finance & Management for this dataset.

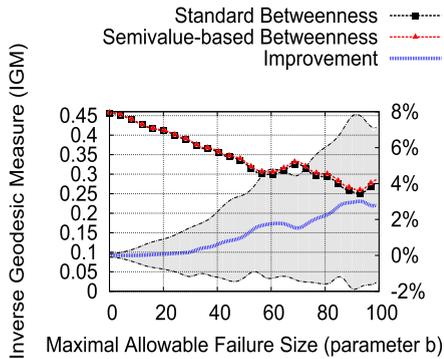


Fig. 14. The difference in performance of IGM on random scale-free graphs ($|V| = 100$).

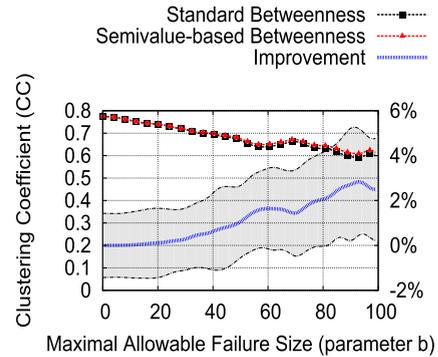


Fig. 15. The difference in performance of CC on random scale-free graphs ($|V| = 100$).

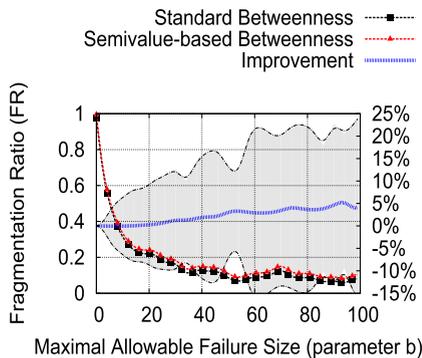


Fig. 16. The difference in performance of FR on random scale-free graphs ($|V| = 100$).

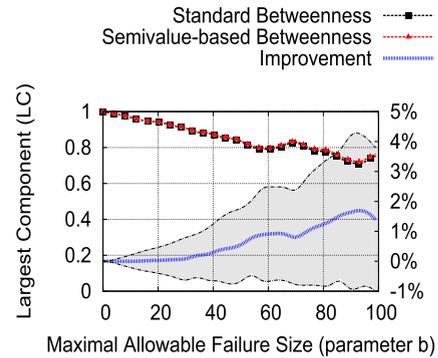


Fig. 17. The difference in performance of LC on random scale-free graphs ($|V| = 100$).

6.4. Random networks

The final experiment of this section is on the random scale-free networks of Barabasi and Albert [10] and the binomial networks of Erdos and Renyi [27]. The former ones were generated with an average node degree of $q = \sqrt{|V|}$, whereas the latter ones were generated with every possible edge having a probability $p = 0.3$ of being formed.¹⁹

Figs. 14, 15, 16, and 17 present the average results from 30 random scale-free networks, with the confidence intervals corresponding to 75%. As can be seen, the relative improvement is smaller than the improvement obtained with the real-life networks or the simulated financial network from the previous subsections. Specifically, in terms of IGM or CC, the improvement barely exceeds 2%, while for FR it reaches 5% and for LC it reaches about 1.5%.

The same trend is observed when the random networks are generated using the Erdős–Rényi model. In particular, the relative improvement only reaches 0.5% for the FR, LC, and CC measures, and only up to 1.5% for the IGM measure (see Figs. 18 to 21).

7. Summary and future work

In this article, we proposed an extension of betweenness centrality, which is based on Semivalues—a wide and flexible family of solution concepts from coalitional game theory. Our measure provides a ranking of nodes by considering the roles they each play in different groups of nodes, where a group is evaluated in terms of its group-betweenness centrality. Although Semivalues are, in general, computationally challenging, we showed that the new measure can be computed efficiently. More specifically, we proposed polynomial-time algorithms to compute all Semivalued-based betweenness centralities for weighted graphs as well as unweighted graphs. These include both the Shapley value-based and the Banzhaf Index-based betweenness centralities. Interestingly, our algorithm for computing the Shapley value-based centrality for unweighted networks has the same time complexity as the best known algorithm due to Brandes [15] for computing the standard betweenness centrality.

To compare our Semivalued-based betweenness centrality measure against the standard one, we considered scenarios in which simultaneous node failures occur. Here, the level of protection (against failure) for each node was set to be propor-

¹⁹ We experimented with different values of q and p and obtained qualitatively-similar results.

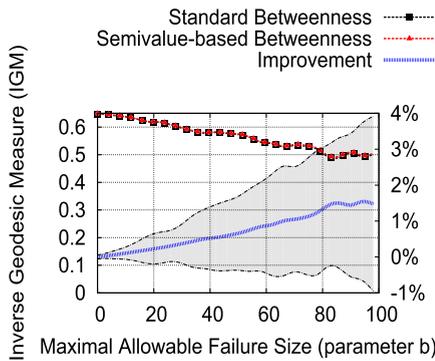


Fig. 18. The difference in performance of IGM on random binomial graphs ($|V| = 100$).

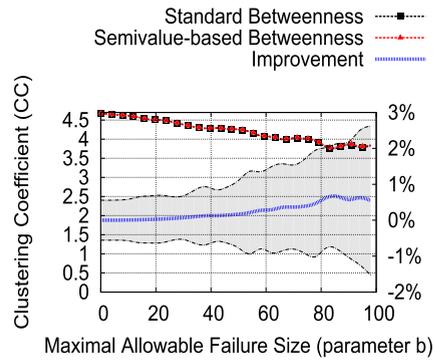


Fig. 19. The difference in performance of CC on random binomial graphs ($|V| = 100$).

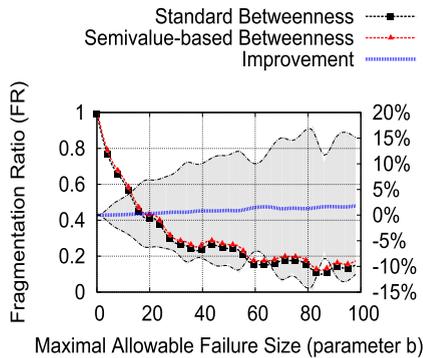


Fig. 20. The difference in performance of FR on random binomial graphs ($|V| = 100$).

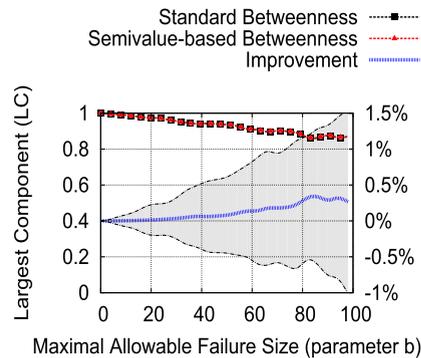


Fig. 21. The difference in performance of LC on random binomial graphs ($|V| = 100$).

tional to its centrality. This way, each centrality measure can be evaluated based on the quality of the corresponding node protection, which in turn is evaluated based on its ability to preserve as much as possible of the network’s performance after the failure has occurred. This revealed that our measure outperforms the standard one in several real-life networks.

There are several potential directions for future work. Firstly, it would be interesting to extend the algorithms proposed in this article to streaming graphs [37]. The challenge here is to efficiently re-compute the centralities and update the ranking of nodes after each change in the structure of the network, i.e., after adding/removing some nodes and/or edges. Secondly, we are keen to develop polynomial-time algorithms for other versions of betweenness centrality, such as routing betweenness centrality [24]. Finally, we would like to identify various axiomatic systems that each uniquely characterize our centrality measure. Such an axiomatic approach is common in the literature on cooperative game theory, and we believe it would be interesting to follow a similar approach to analyze centrality measures in general, and ours in particular.

Acknowledgements

The authors would like to thank the anonymous reviewers and editors for many constructive comments which greatly benefited the article. Piotr L. Szczepański was supported by the Polish National Science Centre based on the project with decision DEC-2013/09/N/ST6/04095. Tomasz P. Michalak was supported by the European Research Council under Advanced Grant 291528 (“RACE”). This work was supported by the Strategic Resilience of Networks project (Homing Plus/2013-7/4) realized within the Homing Plus programme of the Foundation for Polish Science, cofinanced by the European Union from the Regional Development Fund within Operational Programme Innovative Economy (“Grants for Innovation”).

A preliminary version of Section 4 and, to some extent, Sections 5.1 and 5.2 (i.e., the results only concerning the Shapley value-based betweenness centrality) were presented at the 11th International Joint Conference on Autonomous Agents and Multi-Agent Systems and published in the proceedings [63].

References

- [1] R. Albert, H. Jeong, A. Barabási, Error and attack tolerance of complex networks, *Nature* 406 (6794) (2000) 378–382.
- [2] R. Amer, J.G. Miguel, A. Magaña, Accessibility in oriented networks, *Eur. J. Oper. Res.* 180 (2) (2007) 700–712.
- [3] J. Anthonisse, The rush in a directed graph, Tech. rep. BN 9/71, Stichting Mathematisch Centrum, 2e Boerhaavestraat 49, Amsterdam, 1971, <http://oai.cwi.nl/oai/asset/9791/9791A.pdf>.

- [4] Y. Bachrach, The least-core of threshold network flow games, in: *Mathematical Foundations of Computer Science 2011*, in: *Lecture Notes in Computer Science*, vol. 6907, 2011, pp. 36–47.
- [5] Y. Bachrach, J. Rosenschein, Power in threshold network flow games, *Auton. Agents Multi-Agent Syst.* 18 (1) (2009) 106–132.
- [6] Y. Bachrach, R. Savani, N. Shah, Cooperative max games and agent failures, in: *AAMAS '14: Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems*, 2014, pp. 29–36.
- [7] Y. Bachrach, M. Zuckerman, J. Rosenschein, Effort games and the price of myopia, *Math. Log. Q.* 55 (4) (2009) 377–396.
- [8] M.O. Ball, B.L. Golden, R.V. Vohra, Finding the most vital arcs in a network, *Oper. Res. Lett.* 8 (2) (1989) 73–76.
- [9] J.F. Banzhaf, Weighted voting doesn't work: a mathematical analysis, *Rutgers Law Rev.* 19 (1965) 317–343.
- [10] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, *Science* 286 (5439) (1999) 509–512.
- [11] S. Battiston, M. Puliga, R. Kaushik, P. Tasca, G. Caldarelli, DebtRank: too central to fail? Financial networks, the FED and systemic risk, *Sci. Rep.* 2 (541) (2012), 2012/08/02 online.
- [12] T. Beumung, L. Skrabanek, M.Y. Niv, P. Mukherjee, H. Weinstein, PDZBase: a protein–protein interaction database for PDZ-domains, *Bioinformatics* 21 (6) (2005) 827–828.
- [13] E. Bompard, D. Wu, F. Xue, The concept of betweenness in the analysis of power grid vulnerability, in: *COMPENG '10: Proceedings of the Complexity in Engineering Conference*, 2010, pp. 52–54.
- [14] S.P. Borgatti, M.G. Everett, A graph-theoretic perspective on centrality, *Soc. Netw.* 28 (4) (2006) 466–484.
- [15] U. Brandes, A faster algorithm for betweenness centrality, *J. Math. Sociol.* 25 (2) (2001) 163–177.
- [16] U. Brandes, On variants of shortest-path betweenness centrality and their generic computation, *Soc. Netw.* 30 (2) (2008) 136–145.
- [17] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, *Comput. Netw. ISDN Syst.* 30 (1–7) (1998) 107–117.
- [18] G. Chalkiadakis, E. Elkind, M. Wooldridge, *Computational Aspects of Cooperative Game Theory*, *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Springer, 2011.
- [19] S. Colak, H. Lucs, A.R. Atılgan, Vulnerability of networks against critical link failures, *CoRR arXiv:1012.5961*, 2010.
- [20] H. Corley, D.Y. Sha, Most vital links and nodes in weighted networks, *Oper. Res. Lett.* 1 (4) (1982) 157–160.
- [21] T. Cormen, *Introduction to Algorithms*, MIT Press, 2001.
- [22] K.J. Cormican, D.P. Morton, R.K. Wood, Stochastic network interdiction, *Oper. Res.* 46 (2) (1998) 184–197.
- [23] M. del Pozo, C. Manuel, E. González-Arangüena, G. Owen, Centrality in directed social networks. A game theoretic approach, *Soc. Netw.* 33 (3) (2011) 191–200.
- [24] S. Dolev, Y. Elovici, R. Puzis, Routing betweenness centrality, *J. ACM* 57 (4) (May 2010) 25:1–25:27.
- [25] K. Drewniak, J. Helsing, A.R. Mikler, A method for reducing the severity of epidemics by allocating vaccines according to centrality, *CoRR arXiv:1407.7288*, 2014.
- [26] P. Dubey, A. Neyman, R.J. Weber, Value theory without efficiency, Cowles Foundation discussion paper 513, Cowles Foundation for Research in Economics, Yale University, 1979.
- [27] P. Erdős, A. Rényi, On random graphs I, *Publ. Math. (Debr.)* 6 (1959) 290–297.
- [28] European Central Bank, Recent advances in modelling systemic risk using network analysis, <https://www.ecb.europa.eu/pub/pdf/other/modellingsystemicrisk012010en.pdf>, 2010.
- [29] M.G. Everett, S.P. Borgatti, The centrality of groups and classes, *J. Math. Sociol.* 23 (3) (1999) 181–201.
- [30] Financial Stability Board, Reducing the moral hazard posed by systemically important financial institutions, http://www.financialstabilityboard.org/wp-content/uploads/r_101111a.pdf?page_moved=1, 2010.
- [31] L. Freeman, Centrality in social networks: conceptual clarification, *Soc. Netw.* 1 (3) (1979) 215–239.
- [32] N. Friedkin, Theoretical foundations for centrality measures, *Am. J. Sociol.* 96 (6) (1991) 1478–1504.
- [33] S. Gabrieli, Too-interconnected versus too-big-to-fail: banks' network centrality and overnight interest rates, Working paper series 1801390, Social Science Research Network, Feb. 2011, <http://ssrn.com/abstract=1801390>.
- [34] M. Girvan, M.E. Newman, Community structure in social and biological networks, *Proc. Natl. Acad. Sci. USA* 99 (12) (Jun. 2002) 7821–7826.
- [35] P. Gleiser, L. Danon, Community structure in jazz, *Adv. Complex Syst.* 6 (2003) 565.
- [36] D. Gómez, E. González-Arangüena, C. Manuel, G. Owen, M. Del Pozo, J. Tejada, Centrality and power in social networks: a game theoretic approach, *Math. Soc. Sci.* 46 (1) (2003) 27–54.
- [37] O. Green, R. McColl, D.A. Bader, A fast algorithm for streaming betweenness centrality, in: *SocialCom/PASSAT '12: IEEE Third International Conference on Privacy, Security, Risk and Trust/IEEE Third International Conference on Social Computing*, 2012, pp. 11–20.
- [38] B. Grofman, G. Owen, A game-theoretic approach to measuring centrality in social networks, *Soc. Netw.* 4 (1982) 213–224.
- [39] P. Holme, B.J. Kim, C.N. Yoon, S.K. Han, Attack vulnerability of complex networks, *Phys. Rev. E* 65 (5) (2002) 056109.
- [40] E. Kalai, E. Zemel, Generalized network problems yielding totally balanced games, *Oper. Res.* 30 (5) (1982) 998–1008.
- [41] E. Kalai, E. Zemel, Totally balanced games and games of flow, *Math. Oper. Res.* 7 (3) (1982) 476–478.
- [42] E.D. Kolaczyk, D.B. Chua, M. Barthélemy, Group betweenness and co-betweenness: inter-related notions of coalition centrality, *Soc. Netw.* 31 (3) (2009) 190–203.
- [43] L. Laeven, F. Valencia, Resolution of banking crises: the good, the bad, and the ugly, IMF working paper 10/146, International Monetary Fund, Jun. 2010, <https://ideas.repec.org/p/imf/imfwpa/10-146.html>.
- [44] R. Lindelauf, H. Hamers, B. Husslage, Cooperative game theoretic centrality analysis of terrorist networks: the cases of Jemaah Islamiyah and Al Qaeda, *Eur. J. Oper. Res.* 229 (1) (2013) 230–238.
- [45] K. Malik, A.K. Mittal, S.K. Gupta, The k most vital arcs in the shortest path problem, *Oper. Res. Lett.* 8 (4) (1989) 223–227.
- [46] M. Maschler, E. Solan, S. Zamir, *Game Theory*, Cambridge University Press, 2013.
- [47] R. Meir, M. Tennenholz, Y. Bacharach, P. Key, Congestion games with agent failures, in: *AAAI '12: Proceedings of the Twenty-Sixth Conference on Artificial Intelligence*, 2012, pp. 1401–1407.
- [48] T.P. Michalak, K.V. Aaditha, P.L. Szczepański, B. Ravindran, N.R. Jennings, Efficient computation of the Shapley value for game-theoretic network centrality, *J. Artif. Intell. Res.* 46 (2013) 607–650.
- [49] T.P. Michalak, T. Rahwan, P.L. Szczepański, O. Skibski, R. Narayanam, M.J. Wooldridge, N.R. Jennings, Computational analysis of connectivity games with applications to the investigation of terrorist networks, in: *IJCAI'13: Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2013, pp. 293–301.
- [50] D. Monderer, D. Samet, Variations on the Shapley value, in: R. Aumann, S. Hart (Eds.), *Handbook of Game Theory with Economic Applications*, vol. 3, 1st edition, Elsevier, 2002, pp. 2055–2076, Ch. 54.
- [51] S. Moretti, F. Patrone, Transversality of the Shapley value, *Top* 16 (1) (2008) 1–41.
- [52] R. Myerson, Graphs and cooperation in games, *Math. Oper. Res.* 2 (3) (1977) 225–229.
- [53] R. Narayanam, O. Skibski, H. Lamba, T. Michalak, A Shapley value-based approach to determine gatekeepers in social networks with applications, in: *ECAI'14: Proceedings of the 21st European Conference on Artificial Intelligence*, 2014.
- [54] M.E.J. Newman, A measure of betweenness centrality based on random walks, *Soc. Netw.* 13 (2) (2005) 141–154.

- [55] R. Puzis, Y. Elovici, S. Dolev, Finding the most prominent group in complex networks, *AI Commun.* 20 (4) (Dec. 2007) 287–296.
- [56] L.S. Shapley, A value for n -person games, in: H. Kuhn, A. Tucker (Eds.), *Contributions to the Theory of Games*, vol. II, Princeton University Press, 1953, pp. 307–317.
- [57] O. Skibski, T. Michalak, T. Rahwan, M. Wooldridge, Algorithms for the Shapley and Myerson values in graph-restricted games, in: *AAMAS'14: Proceedings of the 13th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 2014, pp. 197–204.
- [58] O. Skibski, T. Michalak, Y. Sakurai, M. Yokoo, A pseudo-polynomial algorithm for computing power indices in graph-restricted weighted voting games, in: *IJCAI'15: Proceedings of the 24th International Conference on Artificial Intelligence*, AAAI Press, 2015, pp. 631–637.
- [59] J. Smith, C. Lim, Algorithms for network interdiction and fortification games, in: *Pareto Optimality, Game Theory and Equilibria*, in: *Springer Optimization and Its Applications*, vol. 17, 2008, pp. 609–644.
- [60] K. Soramäki, S. Cook, Sinkrank: an algorithm for identifying systemically important banks in payment systems, *Econ., Open-Access Open-Assess.* E-J. 7 (28) (2013).
- [61] H. von Stackelberg, *The Theory of the Market Economy*, William Hodge and Co., 1952.
- [62] N. Suri, Y. Narahari, A Shapley value based approach to discover influential nodes in social networks, *IEEE Trans. Autom. Sci. Eng.* 99 (2010) 1–18.
- [63] P. Szczepański, T. Michalak, T. Rahwan, A new approach to betweenness centrality based on the Shapley value, in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, vol. 1, AAMAS '12, 2012, pp. 239–246.
- [64] P. Szczepański, T. Michalak, M. Wooldridge, A centrality measure for networks with community structure based on a generalization of the Owen value, in: *ECAI'14: Proceedings of the 21st European Conference on Artificial Intelligence*, 2014, pp. 867–872.
- [65] D. Watts, S. Strogatz, Collective dynamics of small-world networks, *Nature* 393 (6684) (1998) 440–442.
- [66] R. Weber, Probabilistic values for games, in: A. Roth (Ed.), *The Shapley Value. Essays in Honor of Lloyd S. Shapley*, Cambridge University Press, 1988, pp. 101–120.
- [67] E. Winter, The Shapley value, in: *Handbook of Game Theory with Economic Applications*, vol. 3, 2002, pp. 2025–2054.
- [68] H.P. Young, Monotonic solutions of cooperative games, *Int. J. Game Theory* 14 (2) (1985) 65–72.
- [69] W. Zachary, An information flow model for conflict and fission in small groups, *J. Anthropol. Res.* 33 (1977) 452–473.