Proportional Selection in Networks

Georgios Papasotiropoulos[1], **Oskar Skibski**[1], **Piotr Skowron**[1] and **Tomasz Wąs**[2]
[1]University of Warsaw
[2]University of Oxford
gpapasotiropoulos@gmail.com, {oskar.skibski, p.skowron}@mimuw.edu.pl, tomasz.was@cs.ox.ac.uk

Abstract

We address the problem of selecting k representative nodes from a network, aiming to simultaneously achieve two objectives: identifying the most influential nodes and ensuring that the selection proportionally reflects the diversity within the network. We propose a general approach to accomplish this by combining ideas from network science and computational social choice. Notably, our algorithms depend only on the connections between nodes and do not utilize any additional information that would explicitly identify groups of nodes. We analyze them theoretically, and demonstrate their effectiveness through a series of experiments.

1 Introduction

Consider the *peer-selection problem* of selecting k nodes from a network. Our goal is twofold: to identify the most influential nodes, and to ensure that the selection proportionally represents the structure of the network. For instance, consider a network composed of three densely connected groups, containing 50%, 30%, and 20% of all nodes, respectively, with relatively sparse connections between the groups. If the objective is to select $k = 10$ nodes, a proportional approach would involve selecting five most influential nodes from the first group, three from the second, and two from the third group. In this paper we design selection methods that capture this intuition, yet apply to networks with arbitrary structures.

In recent years, a number of papers proposed methods to select influential nodes in a network in a way that guarantees fair treatment of different communities [Anwar *et al.*, 2021; Stoica *et al.*, 2020; Tsang *et al.*, 2019; Tsioutsiouliklis *et al.*, 2021]. All of these focus on the so-called *feature-aware* fairness, which means that the guarantees are given based on specific attributes shared by a group. This may be problematic in some applications, as the characteristics of the participants are not always available, e.g., due to anonymity [Kilbertus *et al.*, 2018]. At other times, such attributes cannot be used due to ethical, legal, or privacy reasons [Bekkum and Borgesius, 2023], or can be incomplete, unreliable, or noisy [Ashurst and Weller, 2023]. In social-network settings, groups may naturally emerge dynamically or implicitly from structural relations rather than from explicit features. Hence, in this work we focus on *feature-blind* fairness, which gives proportional guarantees to the groups of nodes only based on the topology of the graph. A recent survey on fairness in social network analysis underscores the need for developing methods with such guarantees [Saxena *et al.*, 2024].

In order to achieve this goal, we bridge together two distinct areas of research: *centrality measures* that aim at finding the most influential elements of the network [Koschützki *et al.*, 2005]; and *social choice theory* which focuses, among others, on the selection of proportional outcomes [Lackner and Skowron, 2023]. We propose a framework that transforms a centrality measure into an election to which we apply state-of-the-art methods from social choice that exhibit strong theoretical proportionality guarantees [Peters *et al.*, 2021; Papasotiropoulos *et al.*, 2025a]. While we focus on certain prominent centrality measures, this is for illustration purposes only; our approach is designed to be broad and general enough to be compatible with most centralities, as well as with custom machine learning models for assessing node importance. As a result, our work enables the extension of existing, well-established metrics of importance, to the task of selecting groups of nodes, in a way that maintains the underlying logic of each metric while also ensuring proportional representation.

We analyze our methods axiomatically and characterize their behavior on graph classes, where proportional solutions can be intuitively captured. Our theoretical results suggest that the proposed methods can identify and fairly represent large cohesive groups in any input, without requiring node labels or agent attributes. Our experiments highlight the effectiveness of the proposed methods in promoting proportional outcomes, while still finding highly influential nodes. In the experiments, we rely on structured graphs where proportionality is both interpretable and measurable, and show that our methods achieve fair representation across different communities—even when membership information is not explicitly provided and can only be inferred from the network's structure.

Our framework and approach have broad applicability across various real-world scenarios. Imagine a network where nodes represent political blogs or news websites, and links indicate references between them. In this context, proportional selection ensures a balanced representation of major viewpoints—such as left-wing and right-wing political perspectives, or other widely held opinions that we may not even be aware of—within the selected set of sites. Similarly, one may use our methods to find influential users in social media discussions, while giving a fair treatment to all discussion circles; or select prominent researchers in an area ensuring that all subareas are fairly recognized. In online decision-making platforms where participants may delegate their voting rights to others, such as LiquidFeedback [Behrens *et al.*, 2014] or Google Votes [Hardt and Lopes, 2015], our methods can identify the individuals that proportionally represent the electorate.

Proofs and additional experimental results can be found in the full version of the paper [Papasotiropoulos *et al.*, 2025b].

Further Related Work

Computational social choice distinguishes between three major objectives in the context of selecting subsets of candidates: individual excellence, diversity, and proportionality [Faliszewski *et al.*, 2017; Lackner and Skowron, 2023]. Individual excellence aims to select the best candidates overall. This corresponds to choosing the k nodes with the highest centrality, which can severely neglect minority groups. At the other extreme, diversity ensures that as many individuals as possible enjoy some representation. Group centrality measures that promote diversity have been already explored [Lyu *et al.*, 2021; Mavroforakis *et al.*, 2015; Kempe *et al.*, 2003; Lin *et al.*, 2010]. However, under this objective, large groups may be significantly underrepresented in favor of extremely small groups of disconnected nodes. We take an intermediate approach, aiming to represent different components of a network at a rate proportional to their size.

An active line of research applies the idea of proportional representation beyond traditional voting setting. It was considered in the contexts of online deliberative discourse [Revel *et al.*, 2025], AI-augmented civic participation platforms [Fish *et al.*, 2024], fairness-aware versions of the secretary problem [Papasotiropoulos and Pishbin, 2026], and ranking aggregation settings relevant to online e-commerce platforms [Lederer *et al.*, 2024]. Our work relates to the research on group centrality measures [Everett and Borgatti, 1999; Angriman *et al.*, 2020], but proportionality has not been yet considered in this context. Strategic aspects of node selection in networks have also been explored [Alon *et al.*, 2011; Holzman and Moulin, 2013; Aziz *et al.*, 2016]. Finally, there is a long line of work on community detection in networks, node representation learning, and facility location—indicatively see, respectively, [Chunaev, 2020; Bothorel *et al.*, 2015], [Shi *et al.*, 2024; Köse and Shen, 2021], and [Zhou *et al.*, 2022; Blanco and Gázquez, 2023; Lam *et al.*, 2024]. However, in contrast to what is typically done in that research, our focus is on ensuring that each group of nodes is proportionally represented without relying on additional attributes of the nodes.

2 Preliminaries

We consider unweighted simple directed graphs. Yet, our approach can be adapted to take into account weighted edges. A *graph* (or a *network*) is a pair, $G = (V, E)$, where V is a set of n nodes and $E \subseteq V \times V$ is a set of m edges; the edges are ordered pairs of nodes. An edge (u, v) is an *outgoing edge* for node u and an *incoming edge* for v. The number of outgoing (resp., incoming) edges of u is called *out-degree* (resp., *in-degree*) of u and denoted by $\deg^+(u)$ (resp., $\deg^-(u)$).

A *walk* is a sequence of nodes $(v_1, \dots, v_k)$ such that every two consecutive nodes are connected by an edge: $(v_i, v_{i+1}) \in E$ for every $i \in \{1, \dots, k-1\}$. The length of such a walk is equal to $k-1$, i.e., the number of edges in the walk. Note that a single node is a walk of length 0. A node u is called a *predecessor* of w if there exists a walk $(v_1, \dots, v_k)$ such that $u = v_1$ and $w = v_k$. A node u is a *successor* of w if w is a predecessor of u. The set of predecessors of node u is denoted by $\mathrm{Pred}(u)$ and its set of successors is denoted by $\mathrm{Succ}(u)$. For a set of nodes $S \subseteq V$ we define $\mathrm{Succ}(S) = \bigcup_{u \in S} \mathrm{Succ}(u)$. The set of all walks in G is denoted by $\Omega(G)$.

For a set of nodes $S \subseteq V$, the subgraph induced by S, denoted by $G[S]$, is the graph $(S, \{(u, v) \in E : u, v \in S\})$, i.e., the graph containing S and the edges between nodes from S. A graph G is *strongly connected* if for every two nodes u, v of G there exists a walk from u to v; G is *weakly connected* if every two nodes are connected by a walk in the underlying undirected graph inferred by G. A *component* of G is a maximal weakly connected subgraph of G. A *clique* is a graph in which, for every two nodes u, v, there exist edges (u, v) and (v, u).

A graph $G = (V, E)$ is called (directed) *bipartite* if all walks in $\Omega(G)$ are of length at most 1, in other words, if its set of nodes V can be divided into two disjoint sets $V = V_1 \cup V_2$ such that every edge in E is an outgoing for a node in V_1 and an incoming for a node in V_2. A graph $G = (V, E)$ is called *functional* if $\deg^+(u) \leq 1$, for all nodes $u \in V$.

Given a network $G(V, E)$ and an integer $k < n$ our goal is to select k nodes from V. A method that performs this selection is referred to as a *group selection rule for networks*, in short, a *rule*, and denoted by $\mathcal{R}$. We will also refer to the outcome of such a rule simply as $\mathcal{R}(G, k)$.

2.1 Centrality Measures

A *centrality measure* F is a function that for each graph $G = (V, E)$ and node $v \in V$ assigns a real value, denoted by $F_G(v)$. The higher the value, the more central the node is considered. Our methods can be combined with any centrality measure; for illustration, we focus on the most popular centrality measures based on walks; PageRank and Katz centrality.

PageRank [Page *et al.*, 1999]: For a given decay factor $\alpha \in (0, 1)$, PageRank of a node v in graph G is defined as:

$$\mathsf{PR}^{\alpha}_{G}(v) = \sum_{(u_1, \dots, u_t, v) \in \Omega(G)} \frac{\alpha^t}{\prod_{i=1}^{t} \deg^+(u_i)}. \tag{1}$$

At a high level, PageRank of v is proportional to the expected number of times that v is being visited by a random walk that starts from a random node and in each step follows an outgoing edge chosen uniformly at random or ends the walk with a probability of $1 - \alpha$. Hence, it mostly depends on the number and the importance of its predecessors. We note that many variants of it appear in the literature [Wąs and Skibski, 2023].

Katz centrality [Katz, 1953]: For a given decay factor $\alpha \in (0, 1/\lambda)$, where λ is the largest eigenvalue of the adjacency matrix of G, Katz centrality of a node v in G is defined as:

$$\mathsf{K}^{\alpha}_{G}(v) = \sum_{(u_1,\dots,u_t,v)\in\Omega(G)} \alpha^t. \qquad (2)$$

It can be considered a parallel version of PageRank, where the importance of a node is not split between its outgoing edges, but transferred through all of these simultaneously.

For notational simplicity, when it is clear from the context, we omit G and α from the scripts.

2.2 Election Rules

A committee election is a triple $(V, C, \mu = \{\mu_i\}_{i\in V})$, where $\mu_i \colon C \to \mathbb{R}_+$ is a function representing the preferences of a voter $i \in V$ over the candidates in C. A higher value of $\mu_i(c)$ indicates a stronger level of support from i towards c. A voting rule is a function that for each election (V, C, μ) and natural number $k \in \mathbb{N}$ returns a subset of k candidates from C. Voters' preferences can be expressed in various ways, e.g., using approval ballots where $\mu_i(c) = 1$ if voter i approves candidate c and $\mu_i(c) = 0$, otherwise, or using general utility functions in which $\mu_i(c)$ is an arbitrary non-negative value indicating the satisfaction of i from electing c. We particularly focus on scenarios where $V \equiv C$, i.e., when the voters aim to select a committee from among themselves.

Two of the simplest election rules are *Approval Voting* (AV) and *Satisfaction Approval Voting* (SAV). Adapting their principles to our setting, we will say that each candidate c receives a score from each voter $i \in V$, which equals $\mu_i(c)$ for AV and $\mu_i(c)/\sum_{c'\in C}\mu_i(c')$ for SAV. In both, the k candidates having the maximum total score form the winning committee. A more involved voting rule that aims to select sets of candidates in a proportional way has been proposed under the name Method of Equal Shares (MES) [Peters and Skowron, 2020; Peters *et al.*, 2021]: Let b_i be a virtual budget of a voter i, initially set to k/n. The rule operates in rounds. We say that a not yet elected candidate c is *ρ-affordable* for $\rho \in \mathbb{R}_+$, if its supporters can cover its (unit) cost in such a way that each of them pays ρ per unit of utility, or all of their remaining funds, i.e., if $\sum_{i\in V}\min(b_i, \mu_i(c)\cdot\rho) \geq 1$. We calculate the minimum value of ρ satisfying $\sum_{i\in V}\min(b_i, \mu_i(c)\cdot\rho) \geq 1$. In a given round the method selects the candidate that is ρ-affordable for the lowest possible value of ρ and updates the budgets accordingly: $b_i := b_i - \min(b_i, \mu_i(c)\cdot\rho)$. The procedure stops if there is no ρ-affordable candidate for any finite ρ. Note that the method may terminate with less than k candidates selected. In this case, in our experiments, we will use the method with the Add1U completion [Faliszewski *et al.*, 2023]. Additionally, the very recently proposed *Method of Equal Shares with Bounded Overspending* (BOS) [Papasotiropoulos *et al.*, 2025a]—a robust variant of MES that balances proportionality and efficiency—will also be explored in our simulations.

3 Framework for Election-Based Selection

A natural approach to selecting influential nodes in a network is to choose those with the highest centrality. In the context of liquid democracy, Boldi *et al.* [2011] proposed selecting the k nodes with the highest PageRank—a method we refer to as TOPRANK. This use of PageRank captures the diminishing trust along delegation paths to selected nodes. We define TOPKATZ as the analogous rule based on Katz centrality. As we will show, these methods can critically fail to ensure proportional representation in the network. While proportionality has been highlighted as a key open problem in liquid democracy [Brill, 2018], no proportional selection rules have been proposed to date, to the best of our knowledge. To address this, we introduce a general framework for proportional selection that applies not only to liquid democracy (see the recent survey by Grossi *et al.* [2026] for details on the topic) but also to a broader range of network settings. Our methods take only graphs as input, without any additional information that would explicitly identify groups of nodes within the network. This approach constitutes our main conceptual contribution.

Our proposal leverages proportional committee election rules. At a high level, it transforms the input graph into an election based on nodes' importance. By interpreting nodes as participants, it applies a suitable rule to select representatives. In particular, for any two distinct nodes u and v, we define a utility for node u derived from including node v in the selected set. Such an assessment can be derived from most centrality measures in a natural way, but can also be the result of a link prediction, similarity measures, or other custom machine learning models. For PageRank, with a decay factor α, we define $\mu^{\alpha}_{G}(u, v)$ as the expected number of visits at v of the random walk that starts at node u, based on Equation (1):

$$\mu^{\alpha}_{G}(u, v) := \sum_{(u_1,u_2,\dots,u_k,v)\in\Omega(G):u=u_1} \frac{\alpha^k}{\prod_{i=1}^{k}\deg^+(u_i)}. \qquad (3)$$

According to Equation (3), the utility of u from selecting v is high if node v can be reached with high probability from u. If we interpret edges as votes' delegation, the node to which a vote can be delegated more directly is preferred. Note that $\mathsf{PR}^{\alpha}_{G}(v) = \sum_{u\in V}\mu^{\alpha}_{G}(u, v)$, so TOPRANK is equivalent to AV rule applied to the election (V, V, μ^{α}_{G}). Instead, we mainly use the Method of Equal Shares due to its strong proportionality guarantees, resulting in a rule we refer to as MESRANK. The proposed selection rule is parameterized by a value $\alpha \in (0, 1)$, as it relies on Equation (1), and essentially applies MES to the election where the set of candidates coincides with the set of voters V, and preferences are determined by Equation (3). In short, $\text{MESRANK}^{\alpha}(G, k) := \text{MES}((V, V, \mu^{\alpha}_{G}), k)$. For Katz centrality the definition of a utility function is analogous—but based on Equation (2)—and this gives rise to a rule that we will call MESKATZ.

4 Theoretical Analysis

In this section, we provide a theoretical analysis of MESRANK and MESKATZ, beginning with the structural analysis of their behaviour based on specific graph classes and then turning to establishing axiomatic properties satisfied by the methods.

In particular, we first illustrate our methods on bipartite and functional graphs, as defined in Section 2. In these graph families, proportionality can be intuitively captured, and thus they provide a first way of examining how the proposed methods achieve proportionality and how they differ from one another. Bipartite graphs mimic the scenario of representative democracy, where the set of candidates is separate from the set of voters. Functional graphs can be viewed as elections with 1-approval ballots (where each voter supports at most one candidate). They also reflect the delegation structure used in LiquidFeedback [Behrens *et al.*, 2014]—the most prominent implementation of liquid democracy. Functional graphs were also the exclusive focus of Boldi *et al.* [2011] where TopRank was proposed.

4.1 Characterization for Bipartite Graphs

We begin by looking at directed bipartite graphs that are particularly useful in highlighting the differences between methods based on PageRank and Katz centrality. Moreover, they are suitable for illustrating the characteristic behavior of our proportional selection rules. Without loss of generality we assume that at least k nodes have incoming edges. Since there are only walks of length 0 and 1 in the examined graphs, the Katz centrality and PageRank of each node v can be easily determined. Specifically, if v belongs to V_1, its centrality is minimal, i.e., $\mathsf{PR}^{\alpha}_{G}(v) = \mathsf{K}^{\alpha}_{G}(v) = 1$. Conversely, if v belongs to V_2, then: $\mathsf{PR}^{\alpha}_{G}(v) = 1 + \sum_{(u,v)\in E} \frac{\alpha}{\deg^{+}(u)}$, and $\mathsf{K}^{\alpha}_{G}(v) = 1 + \alpha \cdot \deg^{-}_{G}(v)$. In words, nodes in V_2 get α per incoming edge under Katz centrality, and α divided by the number of outgoing edges per each supporting voter under PageRank. As a result, TopKatz *and* TopRank *applied to a bipartite graph are equivalent to applying AV and SAV, respectively, to the corresponding approval election instance.* Consequently, no proportionality guarantees can be achieved by these two rules.

The approach based on the Method of Equal Shares works differently and as follows: The utility that node u assigns to node v is $\mu^{\alpha}_{G}(u, v) = \alpha/\deg^{+}(u)$ for PageRank and $\mu^{\alpha}_{G}(u, v) = \alpha$ for Katz centrality. Consequently, MesKatz *applied to a bipartite graph is equivalent to applying* MES *to the corresponding approval election instance, while* MesRank *is equivalent to applying a version of* MES *in which the utilities are divided by the number of approvals of each voter.*

For a visualization, consider the bipartite graph depicted in Figure 1. Methods that select nodes based on the highest PageRank or Katz centrality will choose all k nodes from the first group of V_2, ignoring the votes of the remaining 60% of voters. On the other hand, MesRank and MesKatz will select $0.4k$ nodes from the first group and $0.3k$ nodes from each other group of V_2 (up to rounding). This follows from the EJR property of MES [Peters and Skowron, 2020] which says that every large enough group of voters approving the same set of candidates should get a proportional representation in the selected outcome. It is straightforward to generalize this observation, based on EJR, to bipartite graphs with groups of any size.

4.2 Group Selection on Functional Graphs

We now move to functional graphs. PageRank and Katz centralities are identical on such graphs, hence, for ease of exposition, our analysis will focus on rules based on PageRank. For simplicity, we assume that the decay factor α approaches 1. Then the PageRank of nodes that do not lie on any cycle is nearly equal to its number of predecessors.

Consider an arbitrary connected functional graph. Such a graph consists of one cycle or a root node and in-trees attached to it. Nodes from the cycle or root clearly have the maximal PageRank and obtain non-zero (close to 1) utility from all the

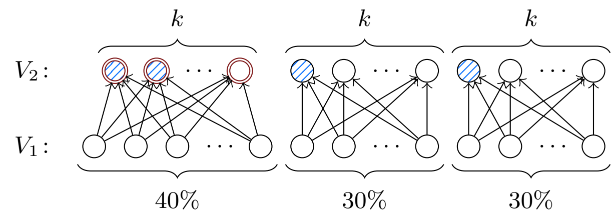

Figure 1: Selecting k nodes from a bipartite graph consisting of three fully connected groups. TopRank (red double lines) selects all k nodes from the first group of V_2. MesRank (blue pattern) selects $0.4k$ of nodes from the first group and $0.3k$ from each other group.

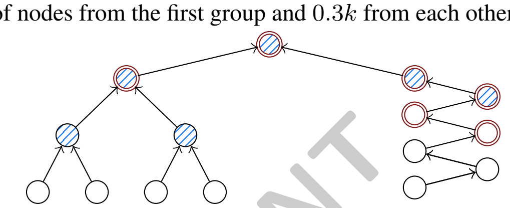

Figure 2: Selecting $k = 6$ nodes from a directed in-tree with two equally sized branches. TopRank (red double lines) selects 4 nodes from the right branch and 1 node from the left, even though the latter has about half the nodes of the network. MesRank (blue pattern) achieves a more balanced selection.

nodes. Thus, if the cycle contains at least k nodes, then both TopRank and MesRank would select only nodes from the cycle. If not, both methods select all nodes from the cycle (or the root) and then some nodes from the attached in-trees. This is, however, where the two methods begin to differ; we refer to Figure 2 for a specific example. TopRank selects nodes with the highest number of predecessors and may pick all or most nodes from the same in-tree. In contrast, MesRank selects nodes from each in-tree proportionally to its size, while preferring nodes supported by more predecessors.

4.3 Group Selection on General Graphs

We now introduce formal axiomatic properties that capture the intuition that a sufficiently large and cohesive group of nodes deserves a proportional number of representatives in the selected outcome. Our axioms are inspired by the literature on multiwinner election rules [Aziz *et al.*, 2017; Lackner and Skowron, 2023]. The idea is that each node should influence the selection of roughly k/n fraction of the selected nodes. Consequently, a cohesive group S is entitled to at least $\lfloor k \cdot |S|/n \rfloor$ representatives. The key question, then, is: in the context of our study, which groups of nodes can be considered cohesive, and which nodes qualify as proper representatives of these groups? Roughly speaking, we view a group of nodes as cohesive if all nodes are connected to each other, either directly or indirectly.

The groups of nodes that are most cohesive are the ones forming a component that is a clique. Our first axiom states that each such a component is entitled to a representation proportional to its size.

Clique-Entitlement: A rule $\mathcal{R}$ satisfies Clique-Entitlement if for every graph $G = (V, E)$, and every component $S \subseteq V$ that is a clique, it holds that $|S \cap \mathcal{R}(G, k)| \geq \lfloor k \cdot |S|/n \rfloor$.

TopRank *and* TopKatz *do not satisfy even this basic axiom.* Both may entirely overlook nodes from a sufficiently large clique when a larger component is present (regardless of whether it is densely connected or not) selecting only nodes from the latter. This again highlights that they are not suitable for proportional representation. In contrast, *this minimal guarantee is satisfied by* MesRank *and* MesKatz.

Next, one could generalize Clique-Entitlement and require the component only being strongly connected, without necessarily being a clique. We go a step beyond and define an even stronger property. Consider a strongly connected subgraph within a larger component of a graph. Since this subgraph is not entirely separate from other nodes, its delegated votes may flow outside the group. However, these votes will always flow to their successors—which are shared by all nodes in the (strongly connected) subgraph. Under clique-entitlement, we assumed that a group deserving representation would be represented by its own members; we now allow for representation through successors, as those are also supported by the members of the considered group.

Subgraph-Entitlement: A rule $\mathcal{R}$ satisfies Subgraph-Entitlement if for every graph $G = (V, E)$, and every subset of nodes $S \subseteq V$ such that $G[S]$ is strongly connected, it holds that $|(S \cup \mathrm{Succ}(S)) \cap \mathcal{R}(G, k)| \geq \lfloor k \cdot {}^{|S|}/_{n} \rfloor$.

It is straightforward to observe that Subgraph-Entitlement implies Clique-Entitlement, and, consequently, Subgraph-Entitlement is violated by TopRank and TopKatz. Before turning to MesRank and MesKatz, we take one more step beyond Subgraph-Entitlement, aiming to relax the requirement of strong connectivity among the vertices in $G[S]$. A prominent guarantee in social choice for fairness considerations is the Proportional Justified Representation (PJR) axiom [Sánchez-Fernández *et al.*, 2026]. It ensures that any group of voters comprising at least an $\ell \cdot {}^{n}/_{k}$ fraction of the population, who all agree on at least ℓ common candidates, are collectively guaranteed at least ℓ representatives. The strengthening of Subgraph-Entitlement that we define below closely resembles PJR —this formulation is also conceptually noteworthy, as it repurposes a standard fairness notion from social choice to another domain.

Group-Entitlement: A rule $\mathcal{R}$ satisfies Group-Entitlement if for every graph $G = (V, E)$, every subset of nodes $S \subseteq V$, and every $\ell \in \mathbb{N}$ such that $|S| \geq \ell \cdot {}^{n}/_{k}$ and $|\bigcap_{u \in S} \mathrm{Succ}(u)| \geq \ell$, it holds that $|\mathrm{Succ}(S) \cap \mathcal{R}(G, k)| \geq \ell$.

MesRank *and* MesKatz *satisfy Group-Entitlement, and, as a byproduct, also Clique- and Subgraph-Entitlement.* This provides a solid proportionality guarantee: any sufficiently large group of nodes with substantial overlap in their successors is entitled to proportional representation in the outcome.

We highlight that our positive axiomatic results are not specific to PageRank and Katz; they hold for any rule that combines the Method of Equal Shares with any measure of nodes' importance, as long as a basic consistency condition is met: the utility of a node u from a node v, as per the function μ, is non-zero if and only if there exists a walk from u to v, i.e., u is connected to v, at least indirectly.

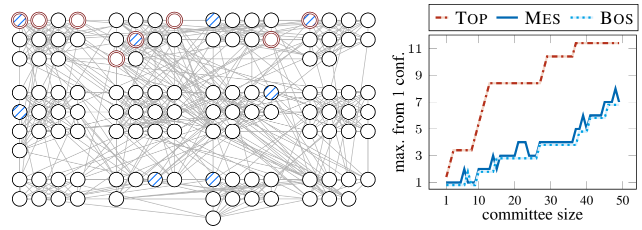

Figure 3: The College Football Network and the outcomes of TopKatz (red double lines) and MesKatz (blue pattern) for $k = 8$ (Left) and the maximum number of nodes selected from a single conference for a given committee size (Right).

5 Experimental Evaluation

In this section, we analyze the discussed rules empirically, on both real-life and synthetic data. Due to space constraints, we present here only the results on Katz-based rules, the rest are included in the full version of the paper [Papasotiropoulos *et al.*, 2025b]. Additionally, we consider BosKatz defined similarly to MesKatz, but using the Method of Equal Shares with Bounded Overspending (BOS) [Papasotiropoulos *et al.*, 2025a], i.e., the fine-tuned variant of MES which is argued to better handle data with high variance in candidate utilities, a common characteristic of our model. We fix $\alpha = 0.85/\lambda$ for Katz-based rules (as an analogue to $\alpha = 0.85$ for PageRank used, e.g., by Brin and Page [1998]).

First, we analyze the proportionality aspects of our rules based on two network datasets often used as benchmarks for community detection algorithms [Jin *et al.*, 2021]. Then we extend our study by analyzing larger real-world data as well as synthetic datasets. Whenever undirected graphs are used, we convert them into directed graphs by replacing each undirected edge with two directed edges in opposite directions. Our experiments demonstrate that the proposed methods can proportionally represent different communities of the network and that relying on the network structure is enough to enable fair selection that reflects the underlying structure even with respect to groups that are unknown or not explicitly defined.

5.1 College Football Network

The first dataset [Girvan and Newman, 2002] is a graph of 115 nodes representing U.S. college football teams. Each undirected edge is a game played in Division IA during the 2000 Fall season. The teams are split into 11 conferences and a group of independents; 64% of games occur within the conferences. Refer to Figure 3 (Left) for an illustration, in which each group of nodes represents a conference or a set of independent teams. It also shows the outcomes of TopKatz and MesKatz for $k = 8$. There are two conferences from which TopKatz selects three nodes, and hence it covers only four conferences in total. In contrast, MesKatz distributes its selection more evenly across conferences, selecting at most one team per conference. To generalize this analysis for other values of k, we plot the maximum number of teams selected from a single conference for each rule and for $k \in \{1, \ldots, 50\}$ (see Figure 3 (Right)). We also include BosKatz there. The result shows that as we increase k, the difference between TopKatz and the other two rules becomes even more pronounced.

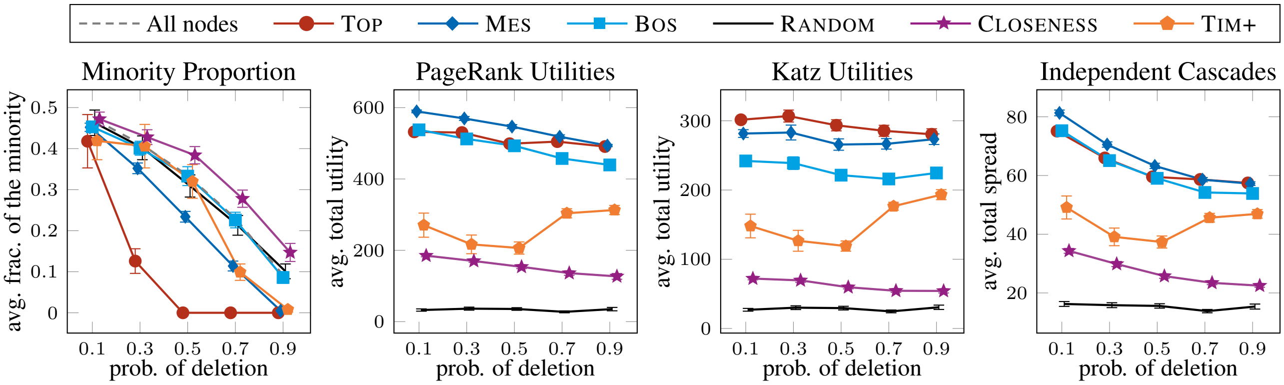

Figure 4: Results of our experiments on Political Blogs Network for $k = 10$, for TOPKATZ, MESKATZ, and BOSKATZ. Each plot presents the values of a different statistic depending on the probability of removing blogs from one side (x-axis). In the leftmost plot, we show the average fraction of minority-labeled nodes in the outputs of the rules (in this plot we also present the fraction of minority-labeled nodes among all nodes in the network—grey dashed line). In the two middle plots, we present the average summed values of PageRank and Katz centrality, respectively, among the selected nodes. In the rightmost plot, we give the average number of infected nodes in the independent cascades model with 2% infection probability starting from the output of a rule as the seed set. Vertical bars denote 95% confidence intervals.

5.2 Political Blogs Network

The second network we analyze [Adamic and Glance, 2005] consists of 1,490 nodes representing political blogs that were active during the 2004 U.S. presidential election. An edge from blog A to blog B indicates a front-page link from A to B. Blogs are labeled as *liberal* or *conservative*, with a roughly balanced composition: 758 liberal and 732 conservative.

To assess how our rules perform on unbalanced data, we delete each node of a given fixed label with probability p and evaluate how well the label distribution in the outcomes reflects that in the modified graph. We also check the effectiveness of the selection using three metrics: summed PageRanks of the selected nodes; summed Katz centralities; and an average spread of influence under the *independent cascades* model [Kempe *et al.*, 2003] (refer to the full version for details) with the selected nodes as seed and uniform transmission probability of 2% (a low value is used since the network is very dense). As baselines we include a random selection of nodes, the local search group-closeness centrality [Angriman *et al.*, 2021] that aims at diversity, and a standard influence maximization algorithm TIM+ [Tang *et al.*, 2014] (we also considered other influence maximization algorithms like CELF++ [Goyal *et al.*, 2011], but their running times were prohibitive for our experiment). For each $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, we generated 100 graphs: 50 graphs with liberal nodes removed, and 50 with conservative. Figure 4 shows the results of our analysis for $k = 10$ (results for $k = 20$ are similar, and can be found in the full version).

For the representativeness of the selection, the differences between the rules are stark. BOSKATZ closely follows the proportions in the input graphs, MESKATZ performs a bit worse, while TOPKATZ completely excludes minority nodes even for moderate values of p. Random selection captures the correct proportions of nodes, while TIM+ underrepresents minority for larger values of p. Group-closeness is skewed in the other direction: it overrepresents the minority.

Next, we look at the total PageRank and Katz centrality of the selected nodes, which serve as metrics for selection of the most influential nodes. The pattern is the same for both metrics: TOPKATZ and MESKATZ perform similarly well, while BOSKATZ exhibits a minimally worse performance. All of our baselines are significantly far behind, with TIM+ performing reasonably well for larger values of p.

For the average spread under the independent cascades model we see a similar pattern to that for PageRank and Katz centrality. This is quite surprising given that TIM+ is an algorithm designed for that purpose (and it was given the exact parameters of the model, including transmission probability, while the remaining methods were blind to it). However, imperfect performance of this algorithm in certain settings has been noted before [Arora *et al.*, 2017].

Only our election-based methods are consistently performing well across all of the examined criteria. The Katz-based rules perform well even when evaluated under total achieved PageRank centrality, showing the robustness of this approach. Thus, BOSKATZ gives almost perfectly proportional representation of the groups in the network, while performing very well under all three efficiency metrics.

5.3 Further Real-World Data

We analyze two additional networks, both relatively large and with nodes split into categories. The first is a network of Facebook pages, where nodes represent official pages and edges denote mutual likes [Rozemberczki *et al.*, 2021]. It includes pages from four categories: politicians, governmental organizations, TV shows, and companies. The second is a CiteSeer network, where nodes represent articles labeled with academic categories [Rossi and Ahmed, 2015]. For both datasets, we report the ℓ_1 distance between the vector of frequencies of different labels in the outputs of our rules and the corresponding vector for the entire network. We report results for $k = 10, 20$ and 50 in Table 1. The pattern is clear: BOSKATZ consistently achieves the minimum distances, while MESKATZ always performs significantly better than TOPKATZ.

	$k = 10$			$k = 20$			$k = 50$		
network	TOP	MES	BOS	TOP	MES	BOS	TOP	MES	BOS
Facebook	1.39	0.99	0.59	1.39	0.99	0.49	1.35	0.59	0.39
CiteSeer	1.59	1.23	0.41	1.59	0.83	0.28	1.23	0.67	0.17

Table 1: The ℓ_1 distance between the vector of frequencies of different labels in the outputs of the Katz-based rules and the corresponding vector for the entire network (best highlighted).

5.4 Euclidean Data

Our final set of experiments is inspired by applications in social choice. As already highlighted, our methods can be applied to elect a committee among a group of voters who vote amongst themselves. In the social choice literature, voters and candidates are often modeled as points in a two-dimensional Euclidean space, typically representing an ideological spectrum [Elkind *et al.*, 2017]. Similarly, we assume that the nodes (representing both voters and candidates) correspond to points. These are sampled from a specific distribution, which we describe later. Edges between nodes are introduced based on the distances between them, using one of four strategies. The first two strategies align with conventional assumptions in social choice, where voters are more likely to prefer candidates closer to them in the ideological space: Under **E-radius**, each node connects (with a fixed probability) to those within a specified radius. Under **E-appr**, each node connects to a fixed number of its closest neighbors. To introduce some noise, we assume that each neighbor can be omitted with a fixed probability.

We also propose a novel Euclidean model, which is well aligned with the idea inspired by liquid democracy: the voters tend to vote for their close friends (i.e., close points), yet in their preferences, exhibit a bias toward candidates with higher competence. Each candidate is assigned an objective value representing their competence (y-coordinate). A node located at point (x, y) connects to the nodes closest to $(x, y+b)$, where b is a constant representing the competence bias. Variants of this model are denoted as **B-radius** and **B-appr**, depending on whether the voters approve candidates within a certain radius of acceptability or a certain number of candidates.

For each of these models, the points are drawn from two Gaussian distributions, with the points divided between the two groups in a $1 : 3$ or $2 : 3$ ratio. We refer to the full version for the full set of results. From each setting, we sample 1000 instances, each consists of $n = 1000$ points. For each instance, we construct a graph with n nodes and select $k = 10$ of them. We identify the points corresponding to the selected nodes and mark them (in green). The points from the 1000 experiments are combined into a single plot, forming a histogram that represents the distribution of selected points.

The histograms are presented in Figure 5. The first column shows the distribution of the points, while the subsequent illustrate the histograms generated for our rules; below each, we provide the proportion of points selected from each half of the plot. First, we observe that BOSKATZ most accurately reflects the original proportions of the points in the two Gaussians; MESKATZ performs slightly worse. TOPKATZ often fails to provide any form of proportionality. Second, we confirm that our methods recover competence bias from the graph, selecting nodes corresponding to high competence.

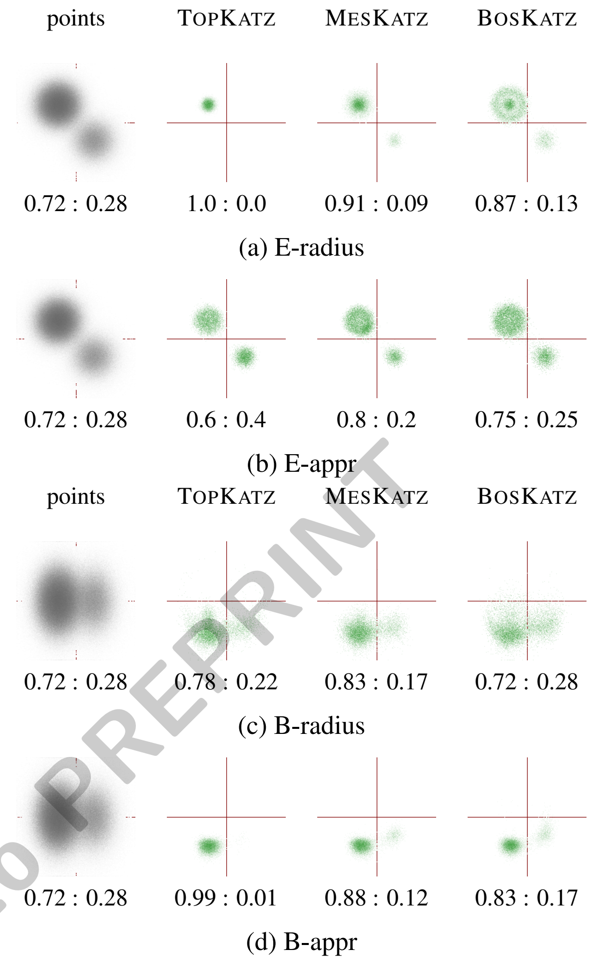

Figure 5: Selection methods for four Euclidean settings: E-radius, E-appr, B-radius, and B-appr. The distributions are presented at their original scale, with histograms at 150% zoom for improved clarity.

6 Concluding Discussion

We introduced the problem of selecting k influential nodes from a network while ensuring proportional representation of different groups implicitly present in it. We proposed a general technique of extending a given centrality measure to a proportional node-selection method. We observe that all of our proposed solutions enable a significantly more representative selection of nodes compared to the existing TOP approach. The difference is particularly pronounced for Katz centrality, but even for PageRank we see noticeable improvements across different datasets. This effect is especially evident in Euclidean graphs (see also the full version) and in the College Football Network (Section 5.1), and reinforced by the theoretical analysis. Among the studied rules, those based on the Method of Equal Shares with Bounded Overspending yield the most representative committees, regardless of whether they are combined with PageRank or Katz centrality. The universality of the BOS approach suggests that it is also a preferable method to be combined with centrality measures beyond the ones studied in this paper as well as machine learning models.

Acknowledgements

Tomasz Wąs was supported by UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/X038548/1. The rest of the authors were partially supported by the European Union (ERC, PRO-DEMOCRATIC, 101076570). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. Oskar Skibski was supported by the National Science Centre under Grant No. 2023/50/E/ST6/00665.

European Research Council
Established by the European Commission

References

[Adamic and Glance, 2005] L. A. Adamic and N. Glance. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the Workshop on the Weblogging Ecosystem*, pages 36–43, 2005.

[Alon *et al.*, 2011] N. Alon, F. Fischer, A. Procaccia, and M. Tennenholtz. Sum of us: Strategyproof selection from the selectors. In *Proceedings of the Conference on Theoretical Aspects of Rationality and Knowledge*, pages 101–110, 2011.

[Angriman *et al.*, 2020] E. Angriman, A. van der Grinten, A. Bojchevski, D. Zügner, S. Günnemann, and H. Meyerhenke. Group centrality maximization for large-scale graphs. In *Proceedings of the Workshop on Algorithm Engineering and Experiments*, pages 56–69, 2020.

[Angriman *et al.*, 2021] E. Angriman, R. Becker, G. d'Angelo, H. Gilbert, A. van Der Grinten, and H. Meyerhenke. Group-harmonic and group-closeness maximization—approximation and engineering. In *Proceedings of the Workshop on Algorithm Engineering and Experiments*, pages 154–168, 2021.

[Anwar *et al.*, 2021] M. S. Anwar, M. Saveski, and D. Roy. Balanced influence maximization in the presence of homophily. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, pages 175–183, 2021.

[Arora *et al.*, 2017] A. Arora, S. Galhotra, and S. Ranu. Debunking the myths of influence maximization: An in-depth benchmarking study. In *Proceedings of the ACM International Conference on Management of Data*, pages 651–666, 2017.

[Ashurst and Weller, 2023] C. Ashurst and A. Weller. Fairness without demographic data: A survey of approaches. In *Proceedings of the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–12, 2023.

[Aziz *et al.*, 2016] H. Aziz, O. Lev, N. Mattei, J. Rosenschein, and T. Walsh. Strategyproof peer selection: Mechanisms, analyses, and experiments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 390–396, 2016.

[Aziz *et al.*, 2017] H. Aziz, M. Brill, V. Conitzer, E. Elkind, R. Freeman, and T. Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48(2):461–485, 2017.

[Behrens *et al.*, 2014] J. Behrens, A. Kistner, A. Nitsche, and B. Swierczek. *The principles of LiquidFeedback*. Interacktive Demokratie, 2014.

[Bekkum and Borgesius, 2023] M. Van Bekkum and F. Z. Borgesius. Using sensitive data to prevent discrimination by artificial intelligence: Does the gdpr need a new exception? *Computer Law & Security Review*, 48:105770, 2023.

[Blanco and Gázquez, 2023] V. Blanco and R. Gázquez. Fairness in maximal covering location problems. *Computers & Operations Research*, 157:106287, 2023.

[Boldi *et al.*, 2011] P. Boldi, F. Bonchi, C. Castillo, and S. Vigna. Viscous democracy for social networks. *Communications of the ACM*, 54(6):129–137, 2011.

[Bothorel *et al.*, 2015] C. Bothorel, J. D. Cruz, M. Magnani, and B. Micenkova. Clustering attributed graphs: models, measures and methods. *Network Science*, 3(3):408–444, 2015.

[Brill, 2018] M. Brill. Interactive democracy. In *Proceedings of the International Conference on Autonomous Agents and MultiAgent Systems*, pages 1183–1187, 2018.

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

[Chunaev, 2020] P. Chunaev. Community detection in node-attributed social networks: a survey. *Computer Science Review*, 37:100286, 2020.

[Elkind *et al.*, 2017] E. Elkind, P. Faliszewski, J. Laslier, P. Skowron, A. Slinko, and N. Talmon. What do multiwinner voting rules do? An experiment over the two-dimensional euclidean domain. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 494–501, 2017.

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

[Faliszewski *et al.*, 2017] P. Faliszewski, P. Skowron, A. Slinko, and N. Talmon. Multiwinner voting: A new challenge for social choice theory. In *Trends in Computational Social Choice*. AI Access, 2017.

[Faliszewski *et al.*, 2023] P. Faliszewski, J. Flis, D. Peters, G. Pierczynski, P. Skowron, D. Stolicki, S. Szufa, and N. Talmon. Participatory budgeting: Data, tools and analysis. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2667–2674, 2023.

[Fish *et al.*, 2024] S. Fish, P. Gölz, D. C. Parkes, A. D. Procaccia, G. Rusak, I. Shapira, and M. Wüthrich. Generative social choice. In *Proceedings of the ACM Conference on Economics and Computation*, page 985, 2024.

[Girvan and Newman, 2002] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[Goyal *et al.*, 2011] A. Goyal, W. Lu, and L. Lakshmanan. Celf++ optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the International Conference Companion on World Wide Web*, pages 47–48, 2011.

[Grossi *et al.*, 2026] D. Grossi, A. Nitsche, and G. Papasotiropoulos. Modeling liquid democracy: A survey of the (computational) social choice literature. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2026. forthcoming.

[Hardt and Lopes, 2015] S. Hardt and L. C. Lopes. Google Votes: A liquid democracy experiment on a corporate social network. Technical Disclosure Commons, 2015.

[Holzman and Moulin, 2013] R. Holzman and H. Moulin. Impartial nominations for a prize. *Econometrica*, 81(1):173–196, 2013.

[Jin *et al.*, 2021] D. Jin, Z. Yu, P. Jiao, S. Pan, D. He, J. Wu, S. Y. Philip, and W. Zhang. A survey of community detection approaches: From statistical modeling to deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(2):1149–1170, 2021.

[Katz, 1953] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[Kempe *et al.*, 2003] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

[Kilbertus *et al.*, 2018] N. Kilbertus, A. Gascón, M. Kusner, M. Veale, K. Gummadi, and A. Weller. Blind justice: Fairness with encrypted sensitive attributes. In *Proceedings of the International Conference on Machine Learning*, pages 2630–2639, 2018.

[Koschützki *et al.*, 2005] D. Koschützki, K. A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski. Centrality indices. In *Network Analysis: Methodological Foundations*, pages 16–61. 2005.

[Köse and Shen, 2021] O. D. Köse and Y. Shen. Fairness-aware node representation learning. arXiv:2106.05391, 2021.

[Lackner and Skowron, 2023] M. Lackner and P. Skowron. *Multi-winner voting with approval preferences*. Springer Briefs in Intelligent Systems. Springer, 2023.

[Lam *et al.*, 2024] A. Lam, H. Aziz, B. Li, F. Ramezani, and T. Walsh. Proportional fairness in obnoxious facility location. In *Proceedings of the International Conference on Autonomous Agents and MultiAgent Systems*, pages 1075–1083, 2024.

[Lederer *et al.*, 2024] P. Lederer, D. Peters, and T. Wąs. The squared kemeny rule for averaging rankings. In *Proceedings of the ACM Conference on Economics and Computation*, page 755, 2024.

[Lin *et al.*, 2010] G.-L. Lin, H. Peng, Q.-L. Ma, J. Wei, and J.-W. Qin. Improving diversity in web search results re-ranking using absorbing random walks. In *International Conference on Machine Learning and Cybernetics*, pages 2416–2421, 2010.

[Lyu *et al.*, 2021] L. Lyu, B. Fain, K. Munagala, and K. Wang. Centrality with diversity. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, pages 644–652, 2021.

[Mavroforakis *et al.*, 2015] C. Mavroforakis, M. Mathioudakis, and A. Gionis. Absorbing random-walk centrality: Theory and algorithms. In *Proceedings of the IEEE International Conference on Data Mining*, pages 901–906, 2015.

[Page *et al.*, 1999] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford Infolab, 1999.

[Papasotiropoulos and Pishbin, 2026] G. Papasotiropoulos and Z. Pishbin. Fairness in the multi-secretary problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 17188–17196, 2026.

[Papasotiropoulos *et al.*, 2025a] G. Papasotiropoulos, Z. Pishbin, O. Skibski, P. Skowron, and T. Wąs. Method of equal shares with bounded overspending. In *Proceedings of the ACM Conference on Economics and Computation*, pages 841–868. ACM, 2025.

[Papasotiropoulos *et al.*, 2025b] G. Papasotiropoulos, O. Skibski, P. Skowron, and T. Wąs. Proportional selection in networks. arXiv:2502.03545v2, 2025.

[Peters and Skowron, 2020] D. Peters and P. Skowron. Proportionality and the limits of welfarism. In *Proceedings of the ACM Conference on Economics and Computation*, pages 793–794, 2020.

[Peters *et al.*, 2021] D. Peters, G. Pierczyński, and P. Skowron. Proportional participatory budgeting with additive utilities. *Advances in Neural Information Processing Systems*, 34:12726–12737, 2021.

[Revel *et al.*, 2025] M. Revel, S. Milli, T. Lu, J. Watson-Daniels, and M. Nickel. Representative ranking for deliberation in the public sphere. In *Proceedings of the International Conference on Machine Learning*, 2025.

[Rossi and Ahmed, 2015] R. A. Rossi and N. K. Ahmed. The network data repository with interactive graph analytics and visualization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4292–4293, 2015.

[Rozemberczki *et al.*, 2021] B. Rozemberczki, C. Allen, and R. Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2):cnab014, 2021.

[Sánchez-Fernández *et al.*, 2026] L. Sánchez-Fernández, E. Elkind, M. Lackner, N. Fernández, J. Fisteus, P. B. Val, and P. Skowron. Proportional justified representation. *Artificial Intelligence*, page 104503, 2026.

[Saxena *et al.*, 2024] A. Saxena, G. Fletcher, and M. Pechenizkiy. Fairsna: Algorithmic fairness in social network analysis. *ACM Computing Surveys*, 56(8):1–45, 2024.

[Shi *et al.*, 2024] Z. Shi, J. Wang, F. Lu, H. Chen, D. Lian, Z. Wang, J. Ye, and F. Wu. Label deconvolution for node representation learning on large-scale attributed graphs against learning bias. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[Stoica *et al.*, 2020] A.-A. Stoica, J. X. Han, and A. Chaintreau. Seeding network influence in biased networks and the benefits of diversity. In *Proceedings of the Web Conference*, pages 2089–2098, 2020.

[Tang *et al.*, 2014] Y. Tang, X. Xiao, and Y. Shi. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the ACM International Conference on Management of Data*, pages 75–86, 2014.

[Tsang *et al.*, 2019] A. Tsang, B. Wilder, E. Rice, M. Tambe, and Y. Zick. Group-fairness in influence maximization. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 5997–6005, 2019.

[Tsioutsiouliklis *et al.*, 2021] S. Tsioutsiouliklis, E. Pitoura, P. Tsaparas, I. Kleftakis, and N. Mamoulis. Fairness-aware PageRank. In *Proceedings of the Web Conference*, pages 3815–3826, 2021.

[Wąs and Skibski, 2023] T. Wąs and O. Skibski. Axiomatic characterization of PageRank. *Artificial Intelligence*, 2023. 103900.

[Zhou *et al.*, 2022] H. Zhou, M. Li, and H. Chan. Strategyproof mechanisms for group-fair facility location problems. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 613–619, 2022.