

Structural Group Unfairness: Measurement and Mitigation by Means of the Effective Resistance

Adrian Arnaiz-Rodriguez¹, Georgina Curto Rex², Nuria Oliver¹

¹ELLIS Alicante, Alicante, Spain

²University of Notre Dame, Notre Dame, IN, USA
{adrian, nuria}@ellisalicante.org, gcurtore@nd.edu

Abstract

Social networks contribute to the distribution of social capital, defined as the relationships, norms of trust and reciprocity within a community or society that facilitate cooperation and collective action. Therefore, better positioned members in a social network benefit from faster access to diverse information and higher influence on information dissemination. A variety of methods have been proposed in the literature to measure social capital at an individual level. However, there is a lack of methods to quantify social capital at a group level, which is particularly important when the groups are defined on the grounds of protected attributes. To fill this gap, we propose to measure the social capital of a group of nodes by means of the effective resistance and emphasize the importance of considering the entire network topology. Grounded in spectral graph theory, we introduce three effective resistance-based measures of group social capital, namely *group isolation*, *group diameter* and *group control*, where the groups are defined according to the value of a protected attribute. We denote the social capital disparity among different groups in a network as *structural group unfairness*, and propose to mitigate it by means of a budgeted edge augmentation heuristic that systematically increases the social capital of the most disadvantaged group. In experiments on real-world networks, we uncover significant levels of structural group unfairness when using gender as the protected attribute, with females being the most disadvantaged group in comparison to males. We also illustrate how our proposed edge augmentation approach is able to not only effectively mitigate the structural group unfairness but also increase the social capital of all groups in the network.

1 Introduction

Online social networks play an important role in defining and sustaining the social fabric of human communities. They allow individuals to connect, interact and share information with one another over the internet. They have opened up new opportunities for personal and professional networking, entertainment and learning. However, the formation of social networks—whether through organic growth or recommendations—can create imbalances in network positions which condition the access to resources and information (Burt 2004). These network inequalities have an impact

on the social capital of its members, which exists in the relations among individuals (Coleman 1988). Better positioned network members benefit from faster access to diverse information, higher influence on information dissemination and more control of the information flow (Granovetter 1983; Burt 2000; Gündoğdu et al. 2019). In practical terms, this means that individuals with a strategic position in the network will have more influence over others, and better access to information and opportunities regarding jobs, health, education or finance.

Furthermore, link recommendation algorithms that pervade social media platforms tend to connect similar users, contributing to the homophily and clustering of the network (Su, Sharma, and Goel 2016). These *filter bubbles* limit the access to diverse individuals (Granovetter 1983), exacerbate the isolation and polarization of groups, reduce the opportunities of innovation and aggravate the perpetuation of societal stereotypes (Garimella et al. 2018). In sum, the topology of the network can lead to a vicious cycle where those who are disadvantaged accumulate fewer opportunities to improve their social capital (Fish et al. 2019).

A variety of graph intervention methods have been proposed in the literature to mitigate disparities in social capital at an individual level (Bashardoust et al. 2023). However, there is a lack of methods that consider such disparities at a group level, which is particularly relevant when the groups correspond to socially vulnerable groups, *i.e.*, those defined on the grounds of sex, race, color, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other (United Nations 1966). Focusing on the group level also supports the development of inclusive solutions at scale that benefit entire communities, promoting equity, diversity and the inclusion of disadvantaged groups. We denote the disparity in social capital among different groups in the network as *structural group unfairness*.

We consider a setting where each node in the network is a source of unique information and, therefore, access to all nodes is equally important. In this context, information flow is an integral component of the social capital and a distance metric that quantifies the total information flow in the graph, considering high-order relations that expand beyond the immediate neighbors, is of utmost importance. We propose using the effective resistance to measure the overall

information flow between pairs of nodes (and thus the social capital), since it is a theoretically grounded continuous graph diffusion metric that considers both *local* and *global* properties of the network’s topology (Chung 1997). In Section 3.2, we introduce three measures of group social capital—*group isolation*, *group diameter* and *group control*—based on the effective resistance where the groups are defined according to the value of a protected attribute of interest. Using these measures of social capital, we define three measures of structural group unfairness in Section 3.3, and frame the challenge of mitigating structural group unfairness as a budgeted edge augmentation task in Section 3.4. This section also presents the Effective Resistance Group Link (ERG-Link) algorithm, a greedy edge augmentation algorithm that iteratively adds edges to the graph to increase the social capital of the most disadvantaged group. In experiments on real-world networks, described in Section 4, we uncover significant levels of structural group unfairness when using gender as the protected attribute, with females being the most disadvantaged group in comparison to males. We also illustrate how our approach is able to not only mitigate disparities in group social capital, but also increase the social capital of all the groups in the network.

2 Related Work

Social Capital Social capital is as a multidimensional construct that has been extensively studied in sociology, political science, economics, and more recently, computational social science (Putnam 2015). It is defined as the networks, relationships, and norms of trust and reciprocity within a community or society that facilitate cooperation and collective action (Coleman 1988). In simple terms, social capital is the value derived from connections between people. It can be measured and analyzed both at an individual and collective levels (Borgatti, Jones, and Everett 1998) and it has been characterized according to different criteria. Some authors propose three main dimensions of social capital, namely: structural, emphasizing the relationships among individuals, organizations and communities; cognitive, focusing on the shared values, norms and beliefs that bind members of a group or community; and relational, highlighting the intensity and quality of relationships, including reciprocity, trust and obligations among individuals (Nahapiet and Ghoshal 1998). Others have proposed the distinction between bonding, bridging, and linking social capital (Szreter and Woolcock 2004). Bonding social capital captures the aspects of “inward looking” communities that reinforce exclusive identities and homogeneous groups (Coleman 1988); bridging social capital refers to “outward looking” networks across different groups that do not necessarily share similar identities (Burt 2004; Granovetter 1983); and linking social capital characterizes the trusting relationships and norms of respect across power or authority gradients (Woolcock et al. 2001). The three forms are important for the well-being of individuals and communities: bonding social capital contributes to social cohesion and support; bridging social capital to mutual understanding, solidarity and respect; and linking social capital to mobilize political resources and power.

Computational Models of Social Capital Network analysis offers a robust computational framework to examine and quantify social capital (Coleman 1988). We consider a setting where all the nodes in the network may be sources of relevant information. As a consequence, access to all nodes—not just the sources or seeds of information—is equally important. In this context, *information flow* is an integral component of the social capital, and a variety of methods have been proposed to characterize it, mainly through two concepts: centrality and criticality (Borgatti 2005).

Centrality measures the relative importance or prominence of a node in the network, quantifying its ability to reach the rest of nodes. Different approaches have been proposed in the literature to measure centrality, including the degree centrality, closeness, local clustering, the assortativity coefficient (Newman 2003), Katz centrality (Katz 1953) and PageRank (Page et al. 1998). Criticality reflects the node’s level of influence or vulnerability within the network (Tizghadam and Leon-Garcia 2009). Nodes with high criticality are essential, such that their failure or disruption can have significant consequences, cascading effects or system-wide impact. Measures of criticality include effective size (Burt 2004), redundancy (Borgatti, Jones, and Everett 1998) and shortest path betweenness (Jackson 2019).

However, previously proposed methods are insufficient to accurately quantify the overall information flow in the network for several reasons. First, they model the distance between nodes as the shortest path distance (geodesic distance (Newman 2003)) which overlooks alternative routes and indirect connections that may exist between distant nodes, thereby underestimating the potential pathways for information diffusion, influence propagation (Stephenson and Zelen 1989) or resource exchange. This myopic view can lead to oversimplified representations of network dynamics, ignoring the interplay between weak ties, bridge nodes and overlapping communities that facilitate connectivity and communication across disparate components in the network. Second, most of the proposed approaches only consider first-order—direct and local—relationships between nodes, relying on small neighborhoods of the graph. As a result, they ignore global structural information (Karhadkar, Banerjee, and Montúfar 2023), such as the properties of the network topology and long-range interactions between nodes, which can lead to inaccurate insights on how information flows globally (Rampášek et al. 2022). Third, popular approaches to model information flow in a network, such as the Independent Cascade model (Kempe, Kleinberg, and Tardos 2003) assume homogeneous, deterministic and instantaneous interactions between neighboring nodes which might lead to inaccurate predictions, biased estimations and misrepresentations of actual diffusion patterns observed in complex networks (Tauch, Liu, and Pears 2015).

Conversely, graph diffusion metrics, such as the *effective resistance* (Stephenson and Zelen 1989; Klein and Randić 1993), offer a principled approach to quantifying distances and interactions between nodes within a network, addressing the above limitations. The effective resistance accurately captures not only short-range but also long-range relationships between nodes (Chung 1997) because it considers al-

ternative pathways, including the network dynamics, and quantifies connectivity between distant nodes. Therefore, it constitutes a natural information distance metric between nodes in a graph (Bozzo and Franceschet 2013; Stephenson and Zelen 1989). Previous work has theoretically formulated measures of node centrality and criticality based on effective resistances (Brandes and Fleischer 2005; Tizghadam and Leon-Garcia 2010; Bozzo and Franceschet 2013; Newman 2005), yet we are not aware of any work that has modeled the social capital of a group of nodes by means of the effective resistance. From a practical perspective, the concept of effective resistance has been used to measure polarization in social networks (Hohmann, Devriendt, and Coscia 2023) and to rank user-items relations in recommender systems (Fouss et al. 2007). In this paper, we propose quantifying the social capital of a group of nodes in the network by means of three measures derived from the effective resistance: the group isolation, group diameter and group control, explained in Section 3.2.

Fairness in Graphs Networks are used for a variety of purposes, including decision-making on nodes, link prediction, node embedding learning, clustering and community detection, ranking, and influence maximization. Fairness has been extensively analyzed in these scenarios (Saxena, Fletcher, and Pechenizkiy 2022; Dong et al. 2023, and references therein).

Regarding social capital, social status plays a role in defining the structure of a network (Ball and Newman 2013) and a node’s position in a network is a form of social capital (Burt 2000). Thus, there are structural advantages in information flow depending on the position that a node occupies in the network. Prior work has studied fairness from the perspective of disparities in access to information by differently positioned nodes in the graph, particularly in the case of influence maximization, *i.e.*, when a single piece of information is spread in the network (Wang, Varol, and Eliassi-Rad 2022). However, there is a scarcity of studies that model group fairness considering that all the nodes in the network are sources of information and therefore access to all the nodes is equally important (Bashardoust et al. 2023). In this context and to the best of our knowledge, no previous research has considered fairness in graphs from a group perspective, when the groups are defined according to protected attributes —such as gender, ethnicity, religion or socio-economic status. In this paper, we fill this gap by defining, measuring and mitigating *structural group unfairness*, understood as disparities in social capital between different groups and where social capital is measured by information flow.

Network Interventions to Mitigate Unfairness in Graphs

Network interventions draw upon social network theory and structural analysis to understand and address the underlying mechanisms of unfairness within social networks. Interventions to mitigate structural unfairness in a network may entail redesigning network structures (Santos, Lelkes, and Levin 2021; Hohmann, Devriendt, and Coscia 2023) or altering (adding and/or removing) edges (Tong et al. 2012) to eliminate discriminatory barriers, reduce homophily, and

foster diversity within the networks (Granovetter 1983). These interventions aim to enhance connectivity, promote inclusivity, and facilitate equitable access to resources, opportunities, and support networks (Borgatti 2005).

When aiming to improve the social capital in a network, edge augmentation (*i.e.*, adding edges) is considered to be the natural intervention to mitigate disparities (Bashardoust et al. 2023). Several edge augmentation strategies have been proposed in the literature, such as connecting similar nodes to improve bonding social capital (Zhang, Anderson, and Zhan 2011), linking nodes with the highest product of eigenvector centralities (Tong et al. 2012) or creating edges between the most disadvantaged nodes and the central node (Bashardoust et al. 2023). However, these strategies are defined for individual notions of social capital and they do not consider long-range interactions between nodes.

Contributions Given previous work, the main contributions of our work are:

- (1) We propose three effective resistance-based measures of group social capital in social networks —namely *group isolation*, *group diameter* and *group control*— that consider short- and long-range interactions between nodes;
- (2) We define *structural group unfairness* as a disparity in the values of such measures by different groups in the graph, where the groups are defined according to the values of a protected attribute. This approach is particularly relevant from a social perspective when the disadvantaged group in the network corresponds to a vulnerable social group;
- (3) We propose *ERG-Link*, an effective resistance-based greedy edge augmentation algorithm that iteratively adds edges to the network to maximize the social capital of the most disadvantaged *group*. We approach this objective by adding weak ties (Granovetter 1983) between the disadvantaged group and the rest of the graph;
- (4) In experiments on real-world networks, we uncover significant levels of group structural unfairness when using gender as a protected attribute, with females being the most disadvantaged group in comparison to males. We also illustrate how our approach is the most effective in reducing structural group unfairness when compared to the baselines.

3 Measuring Structural Group Unfairness

First, we define how to measure information flow between nodes in a social network, which forms the theoretical foundation for the proposed measures of group social capital. Next, we introduce three metrics to quantify a group’s social capital within a graph and define structural group unfairness as the disparity in these metrics across different groups. Finally, we propose a greedy graph intervention (edge augmentation) algorithm designed to mitigate structural group unfairness, *i.e.*, disparities in group social capital.

3.1 Preliminaries

Effective Resistance and Social Capital We focus on the structural dimension of social capital, which emphasizes the relationships among individuals or communities (Nahapiet and Ghoshal 1998), and propose to measure it as the information flow of a node in the network. Such a measure is

captured by the *effective resistance* (Doyle and Snell 1984; Klein and Randić 1993) of the node. Given nodes u and v in graph $G = \{\mathcal{V}, \mathcal{E}\}$, where \mathcal{V} is the set of nodes, $\mathcal{E} = \{(u, v) \in \mathcal{V} \times \mathcal{V} : A_{uv} = 1\}$ is the set of edges and \mathbf{A} is the graph’s adjacency matrix, the effective resistance R_{uv} between nodes u and v is a distance metric given by:

$$R_{uv} = (\mathbf{e}_u - \mathbf{e}_v) \mathbf{L}^\dagger (\mathbf{e}_u - \mathbf{e}_v)^\top, \quad (1)$$

where \mathbf{e}_u is the unit vector with a unit value at u -th index and zero elsewhere; $\mathbf{L}^\dagger = \sum_{i>1} \frac{1}{\lambda_i} \phi_i \phi_i^\top$ is the pseudo-inverse of the graph’s Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A} = \Phi \Lambda \Phi^\top$, with \mathbf{D} the graph’s degree matrix, $D_{u,u} = \sum_{j \in \mathcal{V}} A_{u,j}$ and 0 elsewhere; and λ_i the i -th smallest eigenvalue of \mathbf{L} corresponding to the ϕ_i eigenvector. The complete matrix of all pairwise effective resistances in a graph, \mathbf{R} is given by $\mathbf{R} = \mathbf{1} \text{diag}(\mathbf{L}^\dagger)^\top + \text{diag}(\mathbf{L}^\dagger) \mathbf{1}^\top - 2\mathbf{L}^\dagger$.

The effective resistance is a distance *metric* since it satisfies the symmetry, non-negativity and triangle inequality conditions (Ellens et al. 2011). In addition, R_{uv} is proportional to the commute times between u and v , *i.e.*, the expected number of steps in a random walk starting at v to reach node u and come back: $R_{uv} \propto \mathbb{E}_u[v] + \mathbb{E}_v[u]$, where $\mathbb{E}_u[v]$, $\mathbb{E}_v[u]$ are the expected number of steps that a random walker takes to go from u to v and from v to u , respectively (Chandra et al. 1989; Tetali 1991). A high value of R_{uv} means that u and v generally struggle to visit each other in a random walk, *i.e.*, nodes with high effective resistance between them are unlikely to exchange information. R_{uv} can be expressed as

$$R_{uv} = \sum_{i=0}^{\infty} \left(\frac{1}{d_u} (\mathbf{A}^i)_{uu} + \frac{1}{d_v} (\mathbf{A}^i)_{vv} - \frac{1}{\sqrt{d_u d_v}} 2(\mathbf{A}^i)_{uv} \right),$$

being \mathbf{A}^k the matrix that defines the number paths of length k between u and v (Black et al. 2023). Hence, it is able to capture both short- and long-range interactions between nodes in the graph.

The effective resistance has been characterized as the *information distance* in a network (Stephenson and Zelen 1989; Bozzo and Franceschet 2013) as it quantifies the amount of effort (distance) required to transmit information between the nodes. The total effective resistance R_{tot} of a graph (Ellens et al. 2011) – defined as the sum of all R_{uv} ($R_{\text{tot}} = \mathbf{1} \mathbf{R} \mathbf{1}^\top$) – is therefore inversely proportional to the expected ease of information flow in the graph.

The *total effective resistance of node u* , $R_{\text{tot}}(u)$, is given by $R_{\text{tot}}(u) = \sum_{v \in \mathcal{V}} R_{uv}$, *i.e.*, the sum of all the effective resistances between node u and the rest of nodes in the network. The smaller the total effective resistance of a node, the larger its information flow. In other words, the effective resistance allows to identify which nodes in a graph have limited information flow (*i.e.*, high effective resistance) and thus low social capital. Equivalent terms to denote the effective resistance in the literature include the current-flow closeness centrality (Brandes and Fleischer 2005) and the information centrality (Stephenson and Zelen 1989).

From a computational perspective, calculating R_{uv} does not require hyper-parameter tuning and can be efficiently

calculated, mitigating two significant drawbacks of other diffusion or learnable graph distances (Perozzi, Al-Rfou, and Skiena 2014). An overview of the theoretical properties of R_{uv} are provided in Appendix A.

3.2 Effective Resistance-based Measures of Group Social Capital

Based on the definition of effective resistance above, we propose three metrics that characterize the social capital of a group of nodes in a graph. In the following, a graph $G = \{\mathcal{V}, \mathcal{E}\}$ is composed of a set of nodes \mathcal{V} and edges \mathcal{E} ; and S_i is a group of nodes defined as a subset of \mathcal{V} , *i.e.*, $S_i \subseteq \mathcal{V}$ with $|S_i|$ nodes.

1. Group Isolation The isolation of a group S_i , $R_{\text{tot}}(S_i)$, is given by the average of the total effective resistances of all the nodes in group. $R_{\text{tot}}(S_i)$ is proportional to the expected information distance when sampling one node from group S_i and another node at random. It can be interpreted as a proxy for the marginalization of a group from the perspective of information flow, such that the lower the $R_{\text{tot}}(S_i)$, the more the information flows and thus the less isolated the group S_i is in the network. Therefore, reducing this measure for group S_i would yield an increase in its social capital. It is given by:

$$R_{\text{tot}}(S_i) = \mathbb{E}_{u \sim S_i} [R_{\text{tot}}(u)] = |\mathcal{V}| \mathbb{E}_{u \sim S_i, v \sim \mathcal{V}} [R_{uv}] \quad (2)$$

where $R_{\text{tot}}(u) = \sum_{v \in \mathcal{V}} R_{uv}$ is the total effective resistance of node u . Computing the expectation enables comparing groups of different sizes.

Note that adding links between nodes with the highest R_{uv} —irrespective of which group they belong to— has been found to reduce the total effective resistance of a graph (Black et al. 2023) and hence the isolation of all the graph’s nodes.

2. Group Diameter The group diameter, $\mathcal{R}_{\text{diam}}(S_i)$, measures the average of the maximum distance between any node in group S_i and any node in the graph. A larger group diameter suggests that the nodes in group S_i are distant from the rest of the graph, indicating potential challenges in information exchange with the nodes outside of S_i , and hence it can be interpreted as another measure of social capital. Low diameter is important to ensure diverse information dissemination (*e.g.*, job announcements reaching a diverse pool of candidates) which depends on the entire network and not just on the average distance as *group isolation*. This measure is based on $\mathcal{R}_{\text{diam}}(G)$, which is the maximum effective resistance of the graph (Chandra et al. 1989) and closely related to the cover time. We define *Group Diameter* as:

$$\mathcal{R}_{\text{diam}}(S_i) = \mathbb{E}_{u \sim S_i} [\mathcal{R}_{\text{diam}}(u)] = \mathbb{E}_{u \sim S_i} \left[\max_{v \in \mathcal{V}} R_{uv} \right] \quad (3)$$

where $\mathcal{R}_{\text{diam}}(u) = \max_{v \in \mathcal{V}} R_{uv}$ is the diameter of node u , *i.e.*, the maximum R_{uv} from u to any other node in the graph. $\mathcal{R}_{\text{diam}}(S_i)$ gives an indication of the information flow gap between the group S_i and the rest of the network (Fish et al. 2019). Therefore, the larger the $\mathcal{R}_{\text{diam}}(S_i)$, the lower the social capital of group S_i .

3. Group Control The aforementioned concepts measure the amount of information flow through a group of nodes in the graph. Another relevant variable to assess is the *criticality* of a node for the diffusion of information in the graph, which in the literature has been measured as betweenness (Freeman 1977), redundancy (Borgatti, Jones, and Everett 1998) or effective size (Burt 2004). Nodes with high levels of control serve as important connectors in the network, facilitating the flow of information and enabling communication between otherwise disconnected groups of nodes (Burt 2004).

The control of a node can be computed by restricting the summation of a node’s total effective resistance to its direct neighbors. Thus, it is expressed as $B_R(u) = \sum_{v \in \mathcal{N}(u)} R_{uv}$, where $\mathcal{N}(u) = \{v : (u, v) \in \mathcal{E}\}$ are the neighbors of u . The larger the $B_R(u)$, the more control a node has in the network’s information flow and hence the larger its social capital. The node control of a node is bounded by $1 \leq B_R(u) \leq d_u$, being d_u the number of neighbors of node u (see Theorem B.1). $B_R(u)$ is theoretically related to the current-flow betweenness (Newman 2005), the node’s information bottleneck (Arnaiz-Rodríguez et al. 2022), and the curvature of the node (Devriendt and Lambiotte 2022).

We define the group control or group betweenness $B_R(S_i)$ as the average of the controls of all the nodes in S_i , i.e.:

$$B_R(S_i) = \mathbb{E}_{u \sim S_i}[B_R(u)] = \frac{1}{|S_i|} \sum_{u \in S_i} B_R(u), \quad (4)$$

and is bounded by $1 \leq B_R(S_i) \leq \text{vol}(S_i)/|S_i|$, being $\text{vol}(S_i)/|S_i|$ the average degree of the group S_i (see Theorem B.2). Note that the sum of all R_{uv} for all nodes in a graph is constant at $|V| - 1$ and it is independent of the number of edges (Klein and Randić 1993). If an edge is added, removed or modified in the graph, all R_{uv} are updated accordingly such that their sum remains constant. The sum and the average control of all nodes in the graph are also constant with values $\sum_{u \in V} B_R(u) = 2|V| - 2$ and $\mathbb{E}_{u \sim V}[B_R(u)] = 2 - \frac{2}{|V|}$, respectively, independently of the number of edges (see Appendix B.1 for more details). Consequently, the control of a node or group of nodes is distributed among the nodes in the network and cannot be optimized for every node/group in the graph by adding more edges: if a node or group of nodes increase their control over the information flow in the graph, they must do so at the cost of reducing the control of other nodes. Figure 1 illustrates the three proposed measures of individual and group social capital.

3.3 Structural Group Unfairness

To study disparities in the distribution of social capital in the network, we define groups of nodes in the network S_i according to the values of a sensitive attribute $i \in SA = \{sa_1, sa_2, \dots, sa_{|SA|}\}$, which is a categorical variable with $|SA|$ possible values referring to a socially relevant concept, such as sex, age, gender, religion or race. We denote the value of the sensitive attribute of a node v as $SA(v)$. For instance, if SA is sex with three possible values, $SA = \{\text{male}, \text{female}, \text{non-binary}\}$, the groups S_{male} , S_{female} and $S_{\text{non-binary}}$

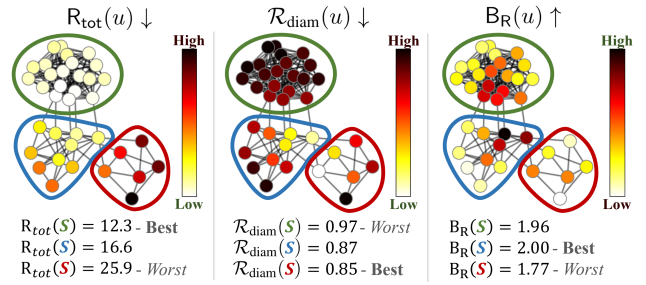


Figure 1: Illustration of the three proposed group social capital metrics on the same graph. The color of the nodes corresponds to $R_{\text{tot}}(u)$, $\mathcal{R}_{\text{diam}}(u)$ and $B_R(u)$, respectively. The nodes are grouped according to three different values of the protected attribute S indicated as green, blue and red.

are the set of nodes whose sex is labeled as male, female and non-binary, respectively.

We define the *structural group unfairness* in a network as the disparity in information flow between the nodes belonging to different groups in the network. Since we have defined the groups in terms of protected attributes, the structural group unfairness is socially relevant as it informs about potential disparities in information flow (and hence social capital) between a vulnerable group and the rest of the network. We present here three metrics to characterize the structural group unfairness, namely *isolation disparity*, *diameter disparity* and *control disparity*.

1. Group Isolation Disparity Ideally, every group in the network should have the same levels of information flow and hence the same —and low— levels of group isolation, namely:

$$R_{\text{tot}}(S_i) = R_{\text{tot}}(S_j), \quad \forall i, j \in SA. \quad (5)$$

Deviations from equality lead to isolation disparity ΔR_{tot} , which is defined as the maximum over all groups in the graph of the differences in group isolation: $\Delta R_{\text{tot}} = \max_{i, j \in SA} |R_{\text{tot}}(S_i) - R_{\text{tot}}(S_j)|$.

Reducing the isolation disparity contributes to increasing the social capital of the most disadvantaged group and equalizes the information flow between the groups in the network.

2. Group Diameter Disparity Ideally, every social group in the network should have the same — and low — group diameter:

$$\mathcal{R}_{\text{diam}}(S_i) = \mathcal{R}_{\text{diam}}(S_j), \quad \forall i, j \in SA. \quad (6)$$

Any deviations from equality lead to diameter disparity, $\Delta \mathcal{R}_{\text{diam}}$, defined as the maximum over all groups in the graph of the differences in group diameter: $\Delta \mathcal{R}_{\text{diam}} = \max_{i, j \in SA} |\mathcal{R}_{\text{diam}}(S_i) - \mathcal{R}_{\text{diam}}(S_j)|$.

Achieving equal diameter entails equalizing the worst-case scenario in information flowing to the entire network from the perspective of any group of nodes in the graph. By promoting equal group diameter, we generate a fairer information-sharing environment.

3. Group Control Disparity By striving for equalized control in all groups in the network, no particular group would dominate or be marginalized from the perspective of their control of information flow in the network:

$$B_R(S_i) = B_R(S_j) = 2 - \frac{2}{|\mathcal{V}|}, \forall i, j \in SA. \quad (7)$$

Then, control disparity, ΔB_R , is defined as the maximum over all groups in the graph of the differences in group control: $\Delta B_R = \max_{i,j \in SA} |B_R(S_i) - B_R(S_j)|$.

Control is a bounded resource to be distributed among the groups of nodes in the graph with an expected value of $2 - \frac{2}{|\mathcal{V}|}$. Hence, reducing the control disparity entails a redistribution of the control in all the groups in the graph converging to $B_R(S_i) = 2 - \frac{2}{|\mathcal{V}|}, \forall i \in SA$, and hence leading to a more equitable allocation of the control that different groups play regarding the information flow in the network.

3.4 Structural Group Unfairness Mitigation

Edge Augmentation Edge augmentation has been proposed as the natural intervention to mitigate information flow disparities in a network where all the nodes are sources of unique pieces of information (Bashardoust et al. 2023). According to Rayleigh’s monotonicity principle (Doyle and Snell 1984), adding edges to a graph always improves information flow, which is the aim of our intervention. Socially, edge addition enhances the nodes’ social capital since information reaches a larger audience. Moreover, note that edge deletion is not a suitable intervention as it could lead to a disruption of social dynamics by breaking existing connections between individuals, which is undesirable (Jackson 2019).

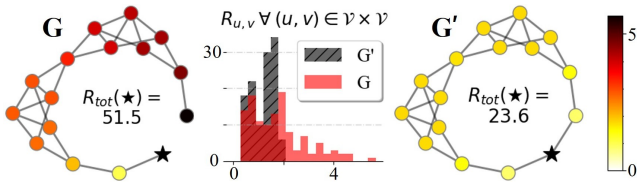


Figure 2: Illustration of the impact of adding one edge on the information flow of G . Note how all the effective resistances between the star node and the rest of nodes in the network ($R_{\star,v}$) decrease in G' even if there is no change in the geodesic distance between them.

Regarding which structural group unfairness measure to optimize, we argue that we should primarily focus on improving the isolation disparity (ΔR_{tot}) of the most isolated group in the graph. Note that mitigating isolation will also yield an improvement in the diameter and control disparities, as illustrated in our experiments. The reduction of ΔR_{tot} entails creating edges between distant nodes, *i.e.*, fostering the creation of weak ties. Granoveter’s work (Granovetter 1983) provides evidence that information spreads more effectively through weak ties than through strong ties because weak ties give peripheral nodes more visibility in the network, which leads to a decrease in group isolation and diameter. Adding

weak ties reduces discontinuities in the information flow, increases redundancies in the paths between nodes and improves the control of peripheral nodes while reducing the control of dominant ones (Burt 2004).

Previous work has proposed connecting peripheral isolated nodes (with high isolation and low control) to salient nodes (with low isolation and high control) (Tong et al. 2012). However, these solutions lead to a *rich-get-richer* phenomenon that benefits the best connected nodes and potentially increases disparities in information access and control (Burt 1999). Therefore, we advocate creating edges between the most distant nodes in the network without necessarily connecting them to a central node. Note that adding edges between the nodes with maximum R_{uv} theoretically leads to a minimization of R_{tot} and R_{diam} for all nodes in the graph while balancing B_R . Hence, the choice of $R_{tot}(S)$.

Problem Definition We consider a budgeted edge augmentation intervention: given a maximum number B of allowed new connections to be created in the graph, we aim to identify the B new edges \mathcal{E}' to be added to the graph G that would maximally reduce the group isolation disparity of the most disadvantaged group in the graph. This leads to a new graph G' with lower levels of structural group unfairness:

$$G' = \arg \min_{G'=(\mathcal{V},\mathcal{E}')} \mathbb{E}_{u,v \sim \mathcal{V} \times \mathcal{V}} [R_{uv}] \quad (8)$$

s.t. $|\mathcal{E}' \setminus \mathcal{E}| = B$ and $\mathcal{E} \subset \mathcal{E}'$

Algorithm To tackle the problem above, we introduce ERG-Link, a greedy algorithm that adds edges between the nodes with the largest effective resistance between them, where at least one of the nodes belongs to the most isolated group as per Sec. 3.2, and groups in the graph are defined on the grounds of a protected attribute. Note that this strategy also reduces the isolation (total effective resistance) of the entire graph (Ghosh, Boyd, and Saberi 2008).

Algorithm 1 outlines the main steps of ERG-Link. Given a graph $G = (V, E)$, a protected attribute S and a total budget B of new edges to add, the group isolation $R_{tot}(S)$ is computed for each group according to S . The most disadvantaged group S_d is identified as the group with the largest $R_{tot}(S)$. Then, $R_{uv} \forall (u, v) \in \mathcal{V} \times \mathcal{V}$ is computed, and a ranking of all potential new edges in the graph is created from the highest to the lowest values of effective resistance. In each iteration, ERG-Link adds the new edge to the graph that yields the largest improvement in the information flow of S_d , *i.e.*, the edge that connects the two nodes with largest effective resistance between them where at least one of the nodes belongs to the disadvantaged group S_d . See Ghosh, Boyd, and Saberi (2008); Black et al. (2023) for a proof that such an edge is the one that maximally improves the information flow in the graph.

ERG-Link leverages *Rayleigh’s monotonicity principle* (Doyle and Snell 1984; Ellens et al. 2011), according to which the total effective resistance of a graph can only decrease when new edges are added to it, as illustrated in Figure 2. Therefore, creating an edge between nodes with maximum R_{uv} where one of the nodes belongs to the disadvantaged group not only improves the information flow

Algorithm 1: ERG-Link

Data: Graph $G = (\mathcal{V}, \mathcal{E})$, a protected attribute SA , budget B of total number of edges to add

Result: New Graph $G' = (\mathcal{V}', \mathcal{E}')$ with B new edges

```

1  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ ;
2  $S_d = \operatorname{argmax}_{S_i \forall i \in SA} R_{\text{tot}}(S_i)$ 
3 Repeat
4    $\mathbf{L}^\dagger = \sum_{i>0} \frac{1}{\lambda_i} \phi_i \phi_i^\top = \left( \mathbf{L} + \frac{\mathbf{1}\mathbf{1}^\top}{n} \right)^{-1} - \frac{\mathbf{1}\mathbf{1}^\top}{n}$ 
5    $\mathbf{R} = \mathbf{1} \operatorname{diag}(\mathbf{L}^\dagger)^\top + \operatorname{diag}(\mathbf{L}^\dagger) \mathbf{1}^\top - 2\mathbf{L}^\dagger$ 
6    $C = \{(u, v) \mid u \in S_d \text{ or } v \in S_d, (u, v) \notin \mathcal{E}'\}$ 
7    $\mathcal{E}' = \mathcal{E}' \cup \operatorname{argmax}_{(u,v) \in C} R_{uv}$ 
8    $\mathbf{L} = \mathbf{L} + (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^\top$ 
9 Until  $|\mathcal{E}' \setminus \mathcal{E}| = B$ ;
10 return  $G'$ ;

```

between the two nodes (increasing their social capital) but it also improves the information flow of the entire graph.

Note that the addition of each new edge changes all the pairwise information distances between nodes in the graph, requiring the re-computation of all distances (effective resistances) in each iteration. Therefore, it is not feasible to perform this type of edge augmentation by means of Independent Cascade distance estimation (Bashardoust et al. 2023; Kempe, Kleinberg, and Tardos 2003), random-walk embeddings (Perozzi, Al-Rfou, and Skiena 2014) or GNNs (Wu et al. 2022) since these methods require training expensive neural networks or running complex simulations for the estimation of the distances in each iteration. In addition, they only capture short-range interactions between nodes. Conversely, the effective resistance captures both short and long-range interactions between nodes in the graph and it is efficient to update. While it requires the computation of the Laplacian pseudo-inverse, Woodbury’s formula (Black et al. 2023) can be used to avoid recomputing \mathbf{L}^\dagger in line 3 of Alg. 1, as reflected in Alg. 2.

4 Experiments

4.1 Datasets and Set-up

To empirically evaluate ERG-Link, we tackle the challenge of mitigating group social capital disparities in three real-world networks (school and online social networks), where the nodes are users and the edges correspond to connections between them, *i.e.*, friendships. The three datasets are commonly used in the graph fairness literature, namely:

- (1) The Facebook dataset (Leskovec and Mcauley 2012), a dense graph of 1,034 Facebook users ($|\mathcal{V}|$) and 26,749 edges ($|\mathcal{E}|$). It corresponds to a large ego-network where nodes are connected if they are friends in the social network;
- (2) The UNC28 dataset (Red et al. 2011), consisting of a 2005 snapshot from the Facebook network of the university of North Carolina ($|\mathcal{V}|=3985, |\mathcal{E}|=65287$);
- (3) The Google+ dataset (Leskovec and Mcauley 2012), an ego-network of G+, the social network developed by Google, with 3,508 nodes ($|\mathcal{V}|$) and 253,930 edges ($|\mathcal{E}|$).

Gender is the protected attribute in all networks with two possible values $SA = \{\text{male, female}\}$. We select the largest connected component for all the datasets. The original values of group social capital per gender are depicted in Table 1. As seen in the Table, our study unveils that the disadvantaged group according to the three defined measures corresponds to females in the three datasets. We provide a detailed breakdown of the statistics and characteristics of the datasets in Appendix C. The code is publicly available¹.

G	$R_{\text{tot}} \downarrow$	$\mathcal{R}_{\text{diam}} \downarrow$	$B_R \uparrow$
Facebook (female)	221.4	2.29	1.93
Facebook (male)	179.8	2.25	2.03
UNC28 (female)	608.6	2.11	1.99
UNC28 (male)	586.3	2.11	2.00
Google+ (female)	564.1	1.31	1.81
Google+ (male)	287.7	1.24	2.32

Table 1: Group social capital metrics in the original graphs for each of the groups.

4.2 Baselines

We compare edge augmentation by means of ERG-Link with five baselines:

- (1) *Random*, which adds edges at random to the graph.
- (2) *DW*, which adds edges with the lowest dot product similarity of *DeepWalk* embeddings (Perozzi, Al-Rfou, and Skiena 2014).
- (3) *Cos*, a greedy algorithm that adds edges with the lowest cosine similarity of the rows of the adjacency matrix (Saxena, Fletcher, and Pechenizkiy 2022). This is an example of a classic method based on neighborhood similarity.
- (4) *SDRF*, an edge augmentation method originally proposed to mitigate over-squashing (Topping et al. 2022). It identifies the edge with the minimum Ricci curvature and adds the edges that maximally improve the Ricci curvature of that edge. The Ricci curvature of an edge is related to R_{uv} and $B_R(u)$ as per Eq. 25.
- (5) *FOSR*, a graph rewiring algorithm to increase the spectral gap (λ_2) and hence avoid over-squashing (Karhadkar, Banerjee, and Montúfar 2023).

Note that *DW* and *Cos* correspond to an algorithm similar to Alg. 1 (lines 2 and 6 remain the same) with one difference: instead of using the effective resistances to quantify the distances between nodes, they consider *DW* or cosine distances, respectively. As previously noted, we do not include any Independent Cascade distance estimation, random-walk embedding or GNN-based method as baselines because they require training a neural network or a simulation to estimate the pairwise distances in the graph, which is computationally unfeasible in our task as it would entail retraining the neural network every time a new edge is added (Wu et al. 2022).

¹<https://github.com/ellisalicante/StructuralGroupUnfairness>

(a) Facebook ($B=50$)				(b) UNC28 ($B=5000$)				(c) Google+ ($B=5000$)			
	ΔR_{tot}	$\Delta \mathcal{R}_{\text{diam}}$	ΔB_R		ΔR_{tot}	$\Delta \mathcal{R}_{\text{diam}}$	ΔB_R		ΔR_{tot}	$\Delta \mathcal{R}_{\text{diam}}$	ΔB_R
G (original)	41.62	0.042	0.107	G (original)	22.4	0.006	0.009	G (original)	276.4	0.078	0.51
Random	38.7	0.039	0.108	Random	19.8	0.005	0.014	Random	129.4	0.037	0.47
SDRF	41.6	0.042	0.106	SDRF	22.2	0.006	0.007	SDRF	276.1	0.079	0.52
FOSR	34.5	0.027	0.109	FOSR	19.7	0.005	0.017	FOSR	240.7	0.068	0.50
DW	36.3	0.031	0.104	DW	22.2	0.006	0.004	DW	274.1	0.078	0.51
Cos	28.7	0.029	0.120	Cos	19.1	0.005	0.102	Cos	86.8	0.025	0.47
ERG	10.3	0.009	0.098	ERG	8.8	0.002	0.003	ERG	37.1	0.011	0.29
S-ERG	41.6	0.042	0.107	S-ERG	22.3	0.006	0.004	S-ERG	276.4	0.079	0.52

Table 2: Structural group unfairness, i.e. differences in group social capital between males and females, before and after the graph interventions. Best result highlighted in bold.

4.3 Experimental Methodology

We set a budget B of 5,000 new edges to be added to the UNC28 and Google+ datasets, which corresponds to approximately 0.05% of the number of all potential edges in the graph. We also run experiments with a maximum of 50 new edges for the Facebook dataset to evaluate the performance of the algorithms with extremely low budgets. We compute the social capital for each group (male and female) and the structural group unfairness on both the original and the augmented graphs (after all edges have been added) based on the defined measures. Moreover, we compute them at each step of edge addition to shed light on the evolution of the structural group unfairness as new edges are added.

4.4 Structural Group Unfairness Mitigation

Table 2 depicts the three structural group unfairness measures on the original graph G and after adding 50 edges to the Facebook dataset and 5,000 edges to the UNC28 and Google+ datasets. The groups are defined based on gender (male, female) and the disadvantaged group are females according to the three structural group unfairness measures, as depicted in Table 1 and top row of Table 2. The disparities in social capital between groups are particularly large in the Google+ network with a ΔR_{tot} of 276.4, meaning that females have significantly lower levels of information flow than males in this network. ΔB_R also shows a difference of 0.51 on the control of the network, which is a large difference given that the values of B_R are in the range $[0, 2 - 2/|\mathcal{V}|]$.

Regarding the results of the graph intervention algorithms, we observe how edge augmentation via ERG-Link outperforms all the baseline methods on the three datasets in terms of reducing disparities in group social capital. Interestingly, the larger the unfairness in the original graph, the larger the improvement after the intervention with ERG-Link. For example, the isolation disparity, ΔR_{tot} , improves by **75%**, **61%** and **87%** after ERG-Link’s intervention in the Facebook, UNC28 and Google+ networks, respectively. A similar behavior is observed for the other structural group unfairness measures.

For illustration purposes, we also report the results of applying Algorithm 1 but where the added edges connect the

nodes with the smallest —instead of the largest— R_{uv} , similar to how link recommendation algorithms work. We refer to this method as the “Strong” version and denote it with an “S-”. As expected, strong edges do not improve the structural group unfairness since such methods are not designed to improve the information flow in the most isolated nodes in the graph, but to connect nodes that are already structurally close to each other (foster strong ties).

Note how the edge augmentation when using DW and Cos distances or via the curvature (SDRF) and λ_2 (FOSR) methods yields graphs with significant levels of structural group unfairness. Furthermore, in the case of using DW distances, the improvement in performance worsens as the graph gets larger (Google+). While using Cos distance for edge augmentation improves ΔR_{tot} and $\Delta \mathcal{R}_{\text{diam}}$, it is unable to always improve ΔB_R due to the inherently more intricate nature of control disparity optimization. Unlike ΔR_{tot} and $\Delta \mathcal{R}_{\text{diam}}$, minimizing ΔB_R requires the precise identification of network gaps which are difficult to detect using a cosine similarity distance.

Over-squashing oriented methods (SDRF and FOSR) aim to improve the information flow by focusing only on the information bottlenecks of the network, which leads to a small effect on improving the overall information flow and reducing disparities. SDRF identifies where to add the edges partly based on the control $B_R(u)$ (Eq. 25) rather than based on the diameter $\mathcal{R}_{\text{diam}}$ as ERG-Link does. FOSR’s goal is based solely on increasing the spectral bottleneck,—i.e. λ_2 —, while ERG-Link considers the entire spectrum, leading to a better characterization and improvement of the information flow.

For completeness, the social capital metrics for each group in the three data sets are shown in Tab. 4 in Appendix D.1. Finally, edge augmentation with ERG-Link is not only more effective in mitigating the group structural unfairness in the graph, but also more computationally efficient than most of the baselines as shown in Appendix E.

4.5 Overall Social Capital Improvement

Figures 3 and 4 illustrate how ERG-Link is more effective than the baseline methods not only to reduce the group unfairness in social capital (ΔR_{tot} and $\Delta \mathcal{R}_{\text{diam}}$), but also to improve the overall social capital measures of group iso-

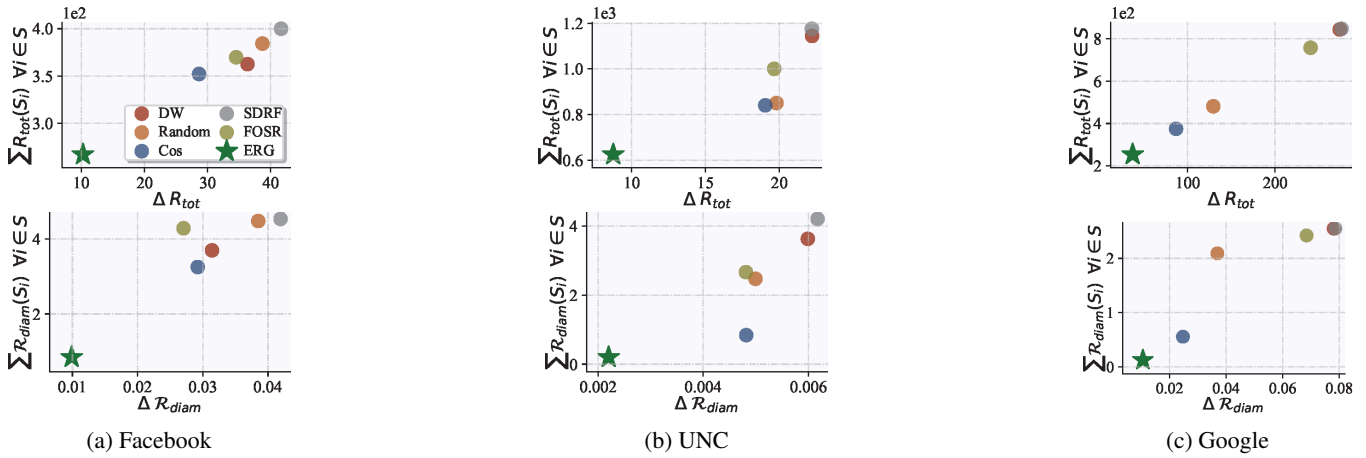


Figure 3: Pareto front of the structural group unfairness (X-axis) vs the sum of the group’s isolation of all the groups (Y-axis) using R_{tot} and R_{diam} (denoted by X-axis’ label). Best results correspond to the bottom left corner of the graph.

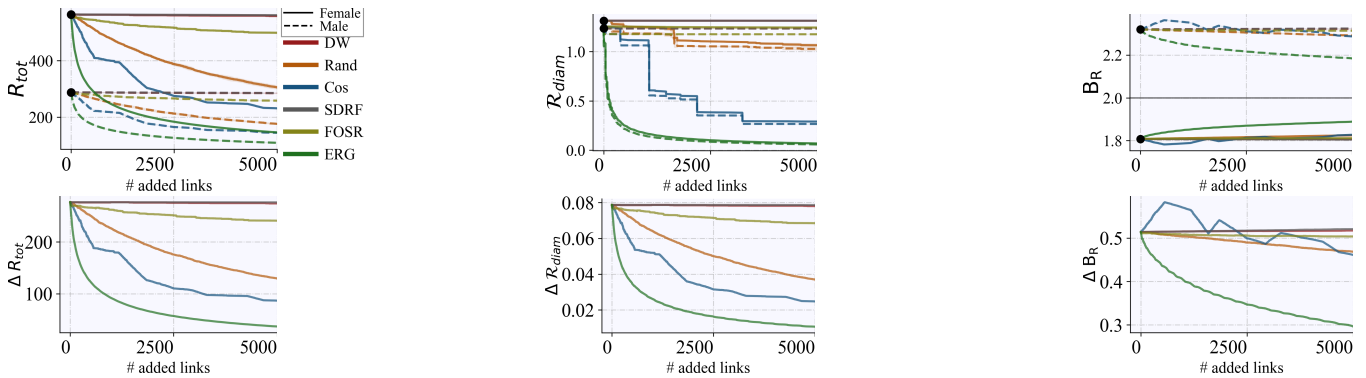


Figure 4: Evolution of the group metrics (top-row) and structural group unfairness metrics (bottom-row) as the number of added edges increases on the Google+ dataset with a total budget of 5,000 new edges.

lation $R_{tot}(S_i)$ and group diameter $R_{diam}(S_i)$ for all the groups in the graph. For each dataset, the structural group unfairness metric (ΔR_{tot} or ΔR_{diam}) is shown on the x-axis of the graphs and the global improvement (sum of group isolation for all groups or sum of diameters for all groups) on the y-axis. In both axes, the lower the values, the better. Edge augmentation via ERG-Link clearly outperforms any other graph intervention strategy, providing evidence that it not only reduces inequalities in social capital between groups, but also improves the social capital of all groups.

4.6 Evolution of Structural Group Unfairness Throughout the Interventions

Figure 4 illustrates the evolution of the structural group unfairness metrics for the Google+ dataset as new edges are added to the network with a total budget of 5,000 edges. As seen in the Figure, edge augmentation via ERG-Link quickly mitigates the group isolation and diameter disparities, even after the addition of a small number of edges. Furthermore, edge augmentation via ERG-Link exhibits a smoother and more consistent reduction in control disparity (ΔB_R), in contrast to the stair-step behavior observed when

adding edges using the baseline methods.

Note that B_R is a finite resource to be allocated among the groups and cannot be globally maximized. Hence, the right-most graphs show how ΔB_R is improved by decreasing B_R of the group with the highest initial B_R and increasing it otherwise (top-right). The goal is to converge both to the optimal bound of $2 - 2/|\mathcal{V}|$ as indicated by the black line.

Figure D.3 depicts the evolution of the structural group unfairness and group social capital metrics when the budget B is very small (50 edges) and large (5,000 edges) to study the efficacy of the interventions. Figure 8 in Appendix D.3 shows the distribution of the nodes’ social capital metrics in each of the groups in the original graph and the resulting graphs after edge augmentation. The results are consistent: for a fixed budget, edge augmentation via ERG-Link yields the best results both for the mitigation of the structural group unfairness and the increase in the overall social capital of all the groups in the graph. Additionally, for a fixed structural group unfairness mitigation goal, edge augmentation via ERG-Link achieves it with a significantly smaller budget of added edges than any of the baseline methods. Finally, Appendix D.4 contains examples of synthetic graphs.

5 Discussion

The proposed `ERG-Link` algorithm may be seen as an *algorithmic reparative* social tool (Davis, Williams, and Yang 2021), *i.e.*, a computational approach designed to address and repair social inequalities, injustices or biases, in our case, in the context of social networks. Our method contributes to the use of social networks, seen as socio-technical systems (Baxter and Sommerville 2011), for social good purposes. While an extensive body of work has been dedicated to the identification and mitigation of algorithmic bias (Danks and London 2017; Barocas and Selbst 2016), this paper aims to trigger a deeper reflection and further discussion about the principles of algorithmic fairness and justice in relation social networks, their ultimate goals and their intricate business models of data trading among platform owners, data brokers, service and advertising companies (Srnicek 2016). There is not a universal interpretation of fairness (Curto and Comim 2023), yet the distributive paradigm of fairness (Rawls 1971) dominates the political and philosophical discussions of social justice today. As a result, most discussions and implementations of fairness focus on how outcomes or benefits are distributed across different socially relevant groups. Nevertheless, there are alternative notions of fairness that remain unexplored in social networks. For example, a prioritarian conception of fairness (Coeckelbergh 2022) would give priority to the individuals belonging to a vulnerable social group. Moreover, prominent voices in social philosophy have long defended that fairness can be understood not only as the redistribution of individual social capital, but also as the social recognition of identity groups (Fraser and Honneth 2003). In this paper, we focus on group fairness and combine both a prioritarian and a Rawlsian approach since the proposed intervention aims to mitigate the social capital gap for the most disadvantaged group while increasing the social capital of all groups. Given the importance of social networks in the definition of the social fabric, the ultimate goal of our work is to spur a reflection on the potential use of social networks as *reparative* tools for social inequality (Davis, Williams, and Yang 2021).

While one could argue that defining groups in terms of protected attributes can be considered a form of discrimination, our approach does not aim to systematically increase the social capital of a pre-defined vulnerable social group, but to detect the social group or groups that is structurally disadvantaged in the network and implement a mechanism at scale that benefits an entire community. This type of mitigation of structural unfairness is particularly relevant since the potential disadvantages suffered by groups in social networks can add up to already existing social conditions that contribute to systemic injustice (Zalta and Nodelman 2011).

The implications of our work go beyond the mitigation of disparities in group social capital and information silos within social networks. The proposed metrics and graph intervention algorithm can also promote innovation and social mobility, since new connections between individuals across different groups are created while prioritizing the information flow gain to/from the most structurally disadvantaged group, and thus strengthening the influence of individuals that suffer from systemic injustice. From a practical perspec-

ive, we envision our proposal as a complement to existing graph interventions in a hybrid setup that combines the addition of edges connecting nodes with both small (strong ties) and large (weak ties) effective resistances. `ERG-Link` provides incentives to users in terms of increasing their social capital (Bashardoust et al. 2023) and opportunities for innovation by connecting distant nodes (Rajkumar et al. 2022). Therefore the deployment of `ERG-Link` implies a more *socially responsible* way to manage connections. Note that the aim of this proposal is not to present a universal solution to group social capital disparities in social networks, but a contribution to spur further reflections regarding the societal impact of algorithms in relation to their purpose and use.

Finally, `ERG-Link`'s objectives are well aligned with the Digital Services Act (DSA)² in Europe, which aims to regulate online platforms and services, enhancing the transparency, accountability and safety of social platforms. The DSA imposes strict responsibilities to very large online platforms (those with more than 45 million EU users), given their significant societal impact. These platforms must conduct regular risk assessments, reduce systemic risks –such as the spread of harmful content or the manipulation of democratic processes– and submit to independent audits. In particular, Article 34 of the DSA states that “providers of very large online platforms need to assess their risks to society and adapt their algorithms if necessary” and Article 38 demands that “providers of very large online platforms and of very large online search engines that use recommender systems shall provide at least one option for each of their recommender systems which is not based on profiling”. The solution offered by `ERG-Link`, where new edges are added with the goal of mitigating disparities in social capital across groups while increasing everyone’s social capital, could be of great interest to providers of very large platforms in the context of this regulation.

6 Conclusion and Future Work

In this paper, we have presented a novel method, based on the effective resistance, to measure and mitigate group social capital disparities within a social network, where the groups are defined according to the values of a protected attribute. Grounded in spectral graph theory, we have introduced three measures of group social capital based on the effective resistance and we have proposed to mitigate disparities in group social capital by means of `ERG-Link`, a budgeted edge augmentation algorithm that systematically increases the social capital of the most disadvantaged group in the network and hence reduces the disparities in group social capital. In extensive experiments with three benchmark graph datasets, we illustrate how `ERG-Link` is the most effective method to decrease disparities in group social capital when compared to five baselines.

In future work, we plan to explore alternative non-greedy edge augmentation algorithms to mitigate structural group unfairness using our effective resistance-based measures and we would like to incorporate additional node features corre-

²https://www.eu-digital-services-act.com/Digital_Services_Act_Article_34.html

sponding to the characteristics of the individuals in the network. We also maintain ongoing discussions with several community organizations to define a project with participants that arrive in another country as refugees, for whom information access and connection with the local population of the hosting country are of paramount importance.

Acknowledgements

The ELLIS Unit Alicante Foundation acknowledges support from by Intel corporation, a nominal grant received at the ELLIS Unit Alicante Foundation from the Regional Government of Valencia in Spain (Convenio Singular signed with Generalitat Valenciana, Conselleria de Innovaci3n, Industria, Comercio y Turismo, Direcci3n General de Innovaci3n) and a grant by the Banc Sabadell Foundation. In addition, it is funded by the European Union EU - HE ELIAS – Grant Agreement 101120237. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA).

References

- Arnaiz-Rodr3guez, A.; Begga, A.; Escolano, F.; and Oliver, N. 2022. DiffWire: Inductive Graph Rewiring via the Lov3sz Bound. In *Proceedings of the First Learning on Graphs Conference*, volume 198 of *Proceedings of Machine Learning Research*, 15:1–15:27. PMLR.
- Ball, B.; and Newman, M. E. 2013. Friendship networks and social status. *Network Science*, 1(1): 16–30.
- Barab3si, A.-L. 2013. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987): 20120375.
- Barocas, S.; and Selbst, A. D. 2016. Big Data’s Disparate Impact. *SSRN Electronic Journal*.
- Bashardoust, A.; Friedler, S. A.; Scheidegger, C. E.; Sullivan, B. D.; and Venkatasubramanian, S. 2023. Reducing Access Disparities in Networks using Edge Augmentation. In *Proceedings of the 2023 FAccT*. ACM.
- Baxter, G.; and Sommerville, I. 2011. Socio-technical systems: From design methods to systems engineering. *Interacting with computers*, 23(1): 4–17.
- Black, M.; Wan, Z.; Nayyeri, A.; and Wang, Y. 2023. Understanding Oversquashing in GNNs through the Lens of Effective Resistance. In *Proceedings of the 40th ICML*.
- Borgatti, S. P. 2005. Centrality and network flow. *Social networks*, 27(1): 55–71.
- Borgatti, S. P.; Jones, C.; and Everett, M. G. 1998. Network measures of social capital. *Connections*, 21(2): 27–36.
- Bozzo, E.; and Franceschet, M. 2013. Resistance distance, closeness, and betweenness. *Social Networks*, 35(3): 460–469.
- Brandes, U.; and Fleischer, D. 2005. Centrality measures based on current flow. In *Annual symposium on theoretical aspects of computer science*, 533–544. Springer.
- Burt, R. S. 1999. The social capital of opinion leaders. *The Annals of the American Academy of Political and Social Science*, 566(1): 37–54.
- Burt, R. S. 2000. The network structure of social capital. *Research in organizational behavior*, 22: 345–423.
- Burt, R. S. 2004. Structural holes and good ideas. *American journal of sociology*, 110(2): 349–399.
- Chandra, A. K.; Raghavan, P.; Ruzzo, W. L.; and Smolensky, R. 1989. The electrical resistance of a graph captures its commute and cover times. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, 574–586.
- Chung, F. R. 1997. *Spectral graph theory*, volume 92. American Mathematical Soc.
- Coeckelbergh, M. 2022. *The Political Philosophy of AI*. Polity.
- Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; and Zucker, S. W. 2005. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the NAS of the United States of America*, 102(21): 7426–7431.
- Coleman, J. S. 1988. Social capital in the creation of human capital. *American journal of sociology*, 94: S95–S120.
- Curto, G.; and Comim, F. 2023. SAF: Stakeholder’s Agreement on Fairness in the Practice of Machine Learning Development. *Science and Engineering Ethics*, 29.
- Danks, D.; and London, A. J. 2017. Algorithmic Bias in Autonomous Systems. In *Ijcai*, volume 17, 4691–4697.
- Davis, J.; Williams, A.; and Yang, M. 2021. Algorithmic reparation. *Big Data and Society*, 8(2).
- Devriendt, K.; and Lambiotte, R. 2022. Discrete curvature on graphs from the effective resistance. *Journal of Physics: Complexity*, 3(2): 025008.
- Di Giovanni, F.; Giusti, L.; Barbero, F.; Luise, G.; Lio, P.; and Bronstein, M. 2023. On Over-Squashing in Message Passing Neural Networks: The Impact of Width, Depth, and Topology. In *In Proceedings of the 40th ICML*.
- Dong, Y.; Ma, J.; Wang, S.; Chen, C.; and Li, J. 2023. Fairness in graph mining: A survey. *IEEE Transactions on Knowledge and Data Engineering*.
- Doyle, P. G.; and Snell, J. L. 1984. *Random walks and electric networks*, volume 22. American Mathematical Soc.
- Ellens, W.; Spieksma, F. M.; Van Mieghem, P.; Jamakovic, A.; and Kooij, R. E. 2011. Effective graph resistance. *Linear algebra and its applications*, 435(10): 2491–2506.
- Fish, B.; Bashardoust, A.; Boyd, D.; Friedler, S.; Scheidegger, C.; and Venkatasubramanian, S. 2019. Gaps in information access in social networks? In *The World Wide Web Conference*, 480–490.
- Fouss, F.; Pirotte, A.; Renders, J.-M.; and Saerens, M. 2007. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on knowledge and data engineering*, 19(3): 355–369.
- Fraser, N.; and Honneth, A. 2003. *Redistribution or recognition? A political-philosophical exchange*. Verso Books.
- Freeman, L. C. 1977. A set of measures of centrality based on betweenness. *Sociometry*, 35–41.

- Garimella, K.; De Francisci Morales, G.; Gionis, A.; and Mathioudakis, M. 2018. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 WWW conference*, 913–922.
- Ghosh, A.; Boyd, S.; and Saberi, A. 2008. Minimizing effective resistance of a graph. *SIAM review*, 50(1): 37–66.
- Granovetter, M. 1983. The strength of weak ties: A network theory revisited. *Sociological theory*, 201–233.
- Gündoğdu, D.; Panzarasa, P.; Oliver, N.; and Lepri, B. 2019. The bridging and bonding structures of place-centric networks: Evidence from a developing country. *PloS one*, 14(9): e0221148.
- Hohmann, M.; Devriendt, K.; and Coscia, M. 2023. Quantifying ideological polarization on a network using generalized Euclidean distance. *Science Advances*, 9(9): eabq2044.
- Jackson, M. O. 2019. *The human network: How your social position determines your power, beliefs, and behaviors*. Vintage.
- Karhadkar, K.; Banerjee, P. K.; and Montúfar, G. 2023. FoSR: First-order spectral rewiring for addressing over-squashing in GNNs. In *The Eleventh International Conference on Learning Representations*.
- Katz, L. 1953. A new status index derived from sociometric analysis. *Psychometrika*, 18(1): 39–43.
- Kempe, D.; Kleinberg, J.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD*, 137–146.
- Klein, D. J.; and Randić, M. 1993. Resistance distance. *Journal of mathematical chemistry*, 12: 81–95.
- Leskovec, J.; and McAuley, J. 2012. Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25.
- Lipman, Y.; Rustamov, R. M.; and Funkhouser, T. A. 2010. Biharmonic distance. *ACM Transactions on Graphics (TOG)*, 29(3): 1–11.
- Lovász, L. 1993. Random walks on graphs. *Combinatorics, Paul erdos is eighty*, 2(1-46): 4.
- Nahapiet, J.; and Ghoshal, S. 1998. Social capital, intellectual capital, and the organizational advantage. *Academy of management review*, 23(2): 242–266.
- Newman, M. E. 2003. Mixing patterns in networks. *Physical review E*, 67(2): 026126.
- Newman, M. E. 2005. A measure of betweenness centrality based on random walks. *Social networks*, 27(1): 39–54.
- Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1998. The pagerank citation ranking: Bring order to the web. Technical report, technical report, Stanford University.
- Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD conference*, 701–710.
- Putnam, R. D. 2015. Bowling alone: America’s declining social capital. In *The city reader*, 188–196. Routledge.
- Qiu, H.; and Hancock, E. R. 2007. Clustering and embedding using commute times. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11): 1873–1890.
- Rajkumar, K.; Saint-Jacques, G.; Bojinov, I.; Brynjolfsson, E.; and Aral, S. 2022. A causal test of the strength of weak ties. *Science (American Association for the Advancement of Science)*, 377(6612): 1304–1310.
- Rampášek, L.; Galkin, M.; Dwivedi, V. P.; Luu, A. T.; Wolf, G.; and Beaini, D. 2022. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems*, 35: 14501–14515.
- Rawls, J. 1971. *A Theory of Justice*. Oxford: Oxford University Press.
- Red, V.; Kelsic, E. D.; Mucha, P. J.; and Porter, M. A. 2011. Comparing community structure to characteristics in online collegiate social networks. *SIAM review*, 53(3): 526–543.
- Santos, F. P.; Lelkes, Y.; and Levin, S. A. 2021. Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences*, 118(50): e2102141118.
- Saxena, A.; Fletcher, G.; and Pechenizkiy, M. 2022. Fairness: Algorithmic fairness in social network analysis. *arXiv preprint arXiv:2209.01678*.
- Srnicek, N. 2016. *Platform capitalism*. Wiley.
- Stephenson, K.; and Zelen, M. 1989. Rethinking centrality: Methods and examples. *Social networks*, 11(1): 1–37.
- Su, J.; Sharma, A.; and Goel, S. 2016. The effect of recommendations on network structure. In *Proceedings of the International WWW Conference*, 1157–1167.
- Szreter, S.; and Woolcock, M. 2004. Health by association? Social capital, social theory, and the political economy of public health. *International journal of epidemiology*, 33(4): 650–667.
- Tauch, S.; Liu, W.; and Pears, R. 2015. Measuring cascade effects in interdependent networks by using effective graph resistance. In *2015 INFOCOM Workshops*, 683–688. IEEE.
- Tetali, P. 1991. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4: 101–109.
- Tizghadam, A.; and Leon-Garcia, A. 2009. A graph theoretical approach to traffic engineering and network control problem. In *2009 21st International Teletraffic Congress*, 1–8.
- Tizghadam, A.; and Leon-Garcia, A. 2010. Betweenness centrality and resistance distance in communication networks. *IEEE Network*, 24(6): 10–16.
- Tong, H.; Prakash, B. A.; Eliassi-Rad, T.; Faloutsos, M.; and Faloutsos, C. 2012. Gelling, and melting, large graphs by edge manipulation. In *Proceedings of the 21st ACM CIKM*, 245–254.
- Topping, J.; Giovanni, F. D.; Chamberlain, B. P.; Dong, X.; and Bronstein, M. M. 2022. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations*.
- United Nations. 1966. International Covenant on Civil and Political Rights.
- Wang, X.; Varol, O.; and Eliassi-Rad, T. 2022. Information access equality on generative models of complex networks. *Applied Network Science*, 7(1): 1–20.

Woolcock, M.; et al. 2001. The place of social capital in understanding social and economic outcomes. *Canadian journal of policy research*, 2(1): 11–17.

Wu, S.; Sun, F.; Zhang, W.; Xie, X.; and Cui, B. 2022. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5): 1–37.

Zalta, E. N.; and Nodelman, U. 2011. Stanford Encyclopedia of Philosophy.

Zhang, S.; Anderson, S. G.; and Zhan, M. 2011. The differentiated impact of bridging and bonding social capital on economic well-being: An individual level perspective. *J. Soc. & Soc. Welfare*, 38: 119.

Zhu, J.; Yan, Y.; Zhao, L.; Heimann, M.; Akoglu, L.; and Koutra, D. 2020. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems*, 33: 7793–7804.

Ethics Checklist

1. For most authors...

- (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **Yes. The goal of the paper is mitigate to avoid discrimination.**
- (b) Do your main claims in the abstract and introduction accurately reflect the paper’s contributions and scope? **Yes**
- (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes**
- (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes, we evaluate the social capital of the groups in the different dataset**
- (e) Did you describe the limitations of your work? **Yes**
- (f) Did you discuss any potential negative societal impacts of your work? **Yes, we discuss the future work of evaluation in the real world.**
- (g) Did you discuss any potential misuse of your work? **NA, it is aimed to reduce information access disparities.**
- (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **NA**
- (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes**

2. Additionally, if your study involves hypotheses testing...

- (a) Did you clearly state the assumptions underlying all theoretical results? **Yes**
- (b) Have you provided justifications for all theoretical results? **Yes**

- (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **Yes, we state why approaches not based on edge augmentation are not recommended socially.**
- (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **Does not apply.**
- (e) Did you address potential biases or limitations in your theoretical framework? **Yes**
- (f) Have you related your theoretical results to the existing literature in social science? **Yes**
- (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? **Yes**

3. Additionally, if you are including theoretical proofs...

- (a) Did you state the full set of assumptions of all theoretical results? **Yes**
- (b) Did you include complete proofs of all theoretical results? **Yes**

4. Additionally, if you ran machine learning experiments...

- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **Yes, all datasets are publicly available and we have referenced the original dataset. The code will be made public upon acceptance.**
- (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **Yes**
- (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **Yes, for the stochastic methods (random and DW). Variability is very small in the Figure 4 but its significant in Figure 7 and 8.**
- (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **Yes**
- (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? **Yes**
- (f) Do you discuss what is “the cost” of misclassification and fault (in)tolerance? **NA**

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

- (a) If your work uses existing assets, did you cite the creators? **Yes**
- (b) Did you mention the license of the assets? **Yes, we reference the original source.**
- (c) Did you include any new assets in the supplemental material or as a URL? **No**
- (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? **NA, they are public datasets.**
- (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **NA, they are public datasets.**

- (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? NA
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...NA
- (a) Did you include the full text of instructions given to participants and screenshots? NA
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
 - (d) Did you discuss how data is stored, shared, and de-identified? NA

A Effective Resistance and information flow

A.1 Graph Diffusion Measures

Discrete Information Propagation Information propagation in networks has been widely studied (Kempe, Kleinberg, and Tardos 2003), prominently by means of graph diffusion and Random Walks methods. A Random Walk (RW) on a graph is a Markov chain that starts at a given node i , and moves randomly to another node from its neighborhood with probability $1/D_{i,i}$. The RW transition probability matrix is given by $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$ and defines the discrete probability of a random walker to move from node u to node v . \mathbf{P}^k is the k -th power of the transition matrix P : the entry $(P^k)_{ij}$ denotes the probability of transitioning from node i to node j in exactly k steps. The graph’s diffusion matrix is defined as $\mathbf{T} = \sum_{k=0}^{\infty} \theta_k \mathbf{P}^k$, and it represents the cumulative effect of multiple steps of a random walk on the graph. Each entry T_{ij} of \mathbf{T} corresponds to the probability of transitioning from node i to node j over an infinite number of steps. θ_k is known as the teleport probability at step k in the random walk. It quantifies the likelihood that, at each step, the random walker will teleport to a random node instead of following an edge. Thus, the sequence $\{\theta_k\}$ is a series of teleport probabilities over the steps. The resulting T captures the cumulative probabilities of transitioning between nodes over an infinite number of steps in the random walk, such that the probability of co-occurrence of two nodes on a random walk corresponds to the probability of information flowing between these two nodes.

However, this approach to assess information flow between nodes in a graph has several limitations. First, it requires considering all the potential paths in a graph, which might not be computationally feasible for large graphs. To overcome this issue, a value of k is typically chosen, which limits the power of the method. Second, the teleport probabilities, θ_k , need to be defined for each k -hop. Several methods have studied how to approximate it, such as Independent Cascade (Kempe, Kleinberg, and Tardos 2003), Katz (Katz 1953), SIR or PageRank (Page et al. 1998). Independent

Cascade or SIR methods (Kempe, Kleinberg, and Tardos 2003) are based on infection models, where they sample guided random walks and, thus, usually rely on expensive Monte Carlo simulations leading to a sub-optimal probability of transition, unable to consider the topology of the entire graph.

Graph Continuous Diffusion Metrics Graph continuous diffusion metrics —such as the Heat kernel distance, (Coifman et al. 2005), effective resistance (or *commute times* distance) (Klein and Randić 1993; Fouss et al. 2007) or the bi-harmonic distance (Lipman, Rustamov, and Funkhouser 2010)— arise as a generalization of random walk metrics. Their mathematical foundations allow for a better characterization of the information flow and an intuitive interpretation of the diffusion processes in a network.

Diffusion metrics define distances based on fine-grained nuances of the topology of the graph that are not captured by simple geodesic distances. When two nodes can be reached by many paths, they should be *closer* than when they can be reached only by few paths of equal length. When two nodes can be reached by a set of edge-independent paths, they are *closer* than when they are reached by redundant paths. Similarly, when two nodes are separated by a shorter path, they are *closer* than when they are separated by a longer path (Bozzo and Franceschet 2013).

In addition, these metrics provide a node embedding, *i.e.*, a numerical representation of each node in the graph that reflects its importance in the process of information diffusion. These embeddings capture the global structure of the network because they incorporate both the local and global geometry of the graph.

The continuous diffusion metrics can be computed using the pseudo-inverse (or Green’s function) of the combinatorial graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$, or the normalized Laplacian $\mathcal{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ (Fouss et al. 2007). The pseudo-inverse, denoted as \mathbf{L}^+ is computed using the spectral decomposition: $\mathbf{L}^+ = \sum_{i \geq 0} \lambda_i^{-1} \phi \phi^\top$ where $\mathbf{L} = \Phi \Lambda \Phi^\top$, where λ_i is the i -smallest eigenvalue of the Laplacian corresponding to the ϕ_i eigenvector.

In this paper, we use the effective resistance, which is a continuous diffusion metric.

A.2 Effective Resistance and Commute Times

The Commute Time (CT) (Lovász 1993), $CT(u, v)$, is the expected number of steps that a random walker needs to go from node u to v and come back to u . The Effective Resistance, R_{uv} , is the Commute Time divided by the volume of the graph (Klein and Randić 1993). In addition to providing a distance for all pairs of nodes —whether connected or not— R_{uv} may be viewed as an indicator of the criticality or importance of the edges in the flow of information throughout the network (Tizghadam and Leon-Garcia 2009).

Intuitively, this distance captures how structurally similar and connected are two nodes in a graph. If two nodes are structurally similar to each other, then the effective resistance between them will be small. Conversely, if two nodes are weakly or not connected, then their effective resistance will be large. In addition, we can define a commute time

embedding (CTE, $\mathbf{Z} = \sqrt{\text{vol}(G)}\mathbf{\Lambda}^{-1/2}\mathbf{\Phi}^\top$) of the nodes in the graph—similar to the idea of the *node’s access signature* in Bashardoust et al. (2023)—, where the Euclidean distance in such an embedding corresponds exactly to the commute times $\text{CT}(u, v) = \|Z_{u,\cdot} - Z_{v,\cdot}\|^2 = \mathbb{E}_u[v] + \mathbb{E}_v[u] = 2|\mathcal{E}|R_{uv}$. This distance is upper bounded by the geodesic distance, with equality in the case of the graph being a tree.

Note that the effective resistance does not rely on any parameter and it is an accurate metric to measure the graph’s information flow (Chandra et al. 1989; Ghosh, Boyd, and Saberi 2008; Arnaiz-Rodríguez et al. 2022), as explained next. Finally, note that R_{uv} can be computed in a spectral manner:

$$R_{uv} = \sum_{i>0} \frac{1}{\lambda_i} (\phi_i(u) - \phi_i(v))^2 \quad (9)$$

where it becomes explicit that R_{uv} depends on all the eigenvectors of the Laplacian, leading to a better characterization of the information flow.

Information Flow in a Graph The graph’s information flow is given by the graph’s conductance, which is measured leveraging the Cheeger Constant, h_G , of a graph (Chung 1997):

$$h_G = \min_{H \subseteq V} \frac{|\{e = (u, v) : u \in S, v \in \bar{H}\}|}{\min(\text{vol}(H), \text{vol}(\bar{H}))} \quad (10)$$

The larger h_G , the harder it is to disconnect the graph into separate communities. Therefore, to increase the information flow in the network, one could add edges to the original graph G , creating a new graph G' , such that $h_{G'} > h_G$. In addition, by virtue of the Cheeger Inequality, h_G is bounded by the smallest non-zero eigenvalue of \mathbf{L} defined as λ_2 :

$$2h_G \leq \lambda_2 < \frac{h_G^2}{2} \quad (11)$$

Finally, the CT is bounded by λ_2 as per the Lovász Bound (Lovász 1993)

$$\left| \frac{\text{CT}(u, v)}{\text{vol}(G)} - \left(\frac{1}{d_u} + \frac{1}{d_v} \right) \right| \leq \frac{1}{\lambda_2} \frac{2}{d_{\min}} \quad (12)$$

where $\text{vol}(G)$ is the volume of G , *i.e.*, the sum of the degrees of all nodes in the graph; d_u, d_v are the degrees of nodes u and v , respectively; and d_{\min} is the minimum degree in the graph.

Therefore, a graph’s information flow is bounded by λ_2 which is bounded by h_G . The intuition is that graphs with large $\lambda_2 \propto h_G$ have short CT distances and thus they have better information flow. Edge augmentation in a graph would lead to a new graph G' where $h_{G'} > h_G$, with smaller CT distances and therefore better information flow.

The effective resistance is also related to other ways of computing the information flow between two nodes in a graph, such as the Jacobian (Topping et al. 2022; Di Giovanni et al. 2023; Black et al. 2023).

A.3 Measures Derived from Effective Resistance

The group social capital measures that we propose in this paper (Sections 3.2 and 3.3) are grounded in previous metrics from the literature.

Total Effective Resistance R_{tot} (Ellens et al. 2011) is the sum of all effective distances in the graph. A lower value of R_{tot} indicates ease of signal propagation across the entire network and hence larger information flow. R_{tot} is given by:

$$R_{\text{tot}} = R_{\text{tot}}(\mathcal{V}) = \frac{1}{2} \mathbf{1}^\top \mathbf{R} \mathbf{1} = \frac{1}{2} \sum_{(v,u) \in \mathcal{V}} R_{uv} \quad (13)$$

$$= n \sum_2^n \frac{1}{\lambda_n} = n \text{Tr}(\mathbf{L}^\dagger) \quad (14)$$

The minimum $R_{\text{tot}} = |V| - 1 = n - 1$ is achieved in a fully connected graph. Conversely, the maximum R_{tot} is achieved on a path graph (or linear graph) and $R_{\text{tot}} = \sum_i^{n-1} i = \frac{1}{2}(n(n-1))$. Therefore, R_{tot} is—for connected graphs—in the range $[n-1, \frac{n(n-1)}{2}]$

Additionally, since the distance between u and v is the Euclidean distance in the embedding \mathbf{Z} , R_{tot} can be obtained as follows (Ghosh, Boyd, and Saberi 2008):

$$R_{\text{tot}} = \sum_{(v,u) \in \mathcal{V}} \|Z_u - Z_v\|^2 = n \sum_{u \in V} \|Z_u\|^2 \quad (15)$$

Similarly to R_{uv} , R_{tot} is theoretically related to the connectivity of the graph defined by its smallest non-zero eigenvalue. Ellens et al. (2011) demonstrated the relation between R_{tot} and λ_2 :

$$\frac{n}{\lambda_2} < R_{\text{tot}} \leq \frac{n(n-1)}{\lambda_2}. \quad (16)$$

Resistance Diameter The proposed group resistance diameter is based on the resistance diameter of a graph $\mathcal{R}_{\text{diam}}$, which is the maximum effective resistance on the graph (Chandra et al. 1989; Qiu and Hancock 2007):

$$\mathcal{R}_{\text{diam}} = \max_{u,v \in V} R_{uv} \quad (17)$$

$\mathcal{R}_{\text{diam}} \propto \lambda_2$ (Chung 1997; Chandra et al. 1989), since

$$\frac{1}{n\lambda_2} \leq \mathcal{R}_{\text{diam}} \leq \frac{2}{\lambda_2} \quad (18)$$

and specifically (Qiu and Hancock 2007; Arnaiz-Rodríguez et al. 2022):

$$h_G \leq \frac{\alpha^\epsilon}{\sqrt{\mathcal{R}_{\text{diam}} \cdot \epsilon}} \text{vol}(S)^{\epsilon-1/2}, \quad (19)$$

By (Qiu and Hancock 2007) we know that

$$\mathcal{R}_{\text{diam}} \leq \frac{1}{\lambda_2} \text{ and } \mathcal{R}_{\text{diam}} \leq \frac{1}{h_G^2} \quad (20)$$

In addition, $\mathcal{R}_{\text{diam}}$ is related to the *cover time* of the graph, which is the expected time required for a random walk to visit every node at least once, *i.e.*, the expected time for a piece of information to reach the entire network. $\mathcal{R}_{\text{diam}}$ can be used to estimate the cover time of the graph, as per (Chandra et al. 1989):

$$m \mathcal{R}_{\text{diam}} \leq \text{cover time} \leq O(m \mathcal{R}_{\text{diam}} \log n) \quad (21)$$

Resistance Betweenness and Curvature As the effective resistance is an information distance, this metric can be used to propose alternative betweenness or criticality metrics to the shortest path betweenness (Newman 2005). In the literature, several effective resistance-based measures have been proposed to determine a node’s criticality, such as: the current flow betweenness (Tizghadam and Leon-Garcia 2009, 2010; Bozzo and Franceschet 2013; Brandes and Fleischer 2005; Newman 2005), the resistance curvature of a node (Topping et al. 2022; Devriendt and Lambiotte 2022), and the information bottleneck property of a node (Arnaiz-Rodríguez et al. 2022; Karhadkar, Banerjee, and Montúfar 2023; Black et al. 2023). Note that the last two definitions are mathematically connected (Devriendt and Lambiotte 2022).

In this work, we focus on the information bottleneck which has a close connection with the resistance curvature of a node. The resistance curvature of a node (Devriendt and Lambiotte 2022) is expressed as

$$p_u = 1 - \frac{1}{2} \sum_{v \in \mathcal{N}(u)} R_{uv}. \quad (22)$$

Therefore, it fulfills the following equality:

$$\begin{aligned} p_u &= 1 - \frac{1}{2} \sum_{v \in \mathcal{N}(u)} R_{uv} = 1 - \frac{1}{2} B_R(u) \\ &\rightarrow B_R(u) = -2(p_u - 1). \end{aligned} \quad (23)$$

Although the definition of a node’s resistance curvature involves computing the sum of the effective resistances between the node and its neighboring nodes, the overall structure of the graph affects all R_{uv} ’s and, consequently, the value of p_i and $B_R(u)$. The curvature of a node is bounded by $1 - d_u/2 \leq p_u \leq 1/2$.

In addition to the node resistance curvature, the link resistance curvature is defined in Devriendt and Lambiotte (2022) as

$$\kappa_{uv} = \frac{2(p_u + p_j)}{R_{uv}} \quad (24)$$

and writing it in terms of the proposed metrics in this work it will translate in:

$$\begin{aligned} \kappa_{uv} &= \frac{2(p_u + p_j)}{R_{uv}} \\ &= \frac{2 \left((1 - \frac{1}{2} B_R(u)) + (1 - \frac{1}{2} B_R(v)) \right)}{R_{uv}} \\ &= \frac{4 - B_R(u) - B_R(v)}{R_{uv}} \end{aligned} \quad (25)$$

Additionally, this is the value that the *SDRF* algorithm (Topping et al. 2022) will use to identify the link to add links around. *SDRF* identifies the link with *lowest* κ_{uv} and adds an edge between a pair the neighbors of the endpoints of that edge that mostly improves the information bottleneck.

B Group Social Capital Metrics and Edge Augmentation Algorithm

B.1 Group Social Capital Metrics

Group Isolation and Isolation Disparity Group Isolation is based on the previously explained notion of total effective resistance of a graph and its close connection with the current flow closeness centrality (Newman 2005).

We propose to define the isolation of a node as its total effective resistance $R_{\text{tot}}(u) = \sum_{v \in \mathcal{V}} R_{uv}$. This centrality metric based on all R_{uv} considers all the eigenvector for its computation, since R_{uv} can be defined using the whole spectrum of the graph. We proceed to derive $R_{\text{tot}}(u)$ in terms of the pseudo-inverse of the graph:

$$\begin{aligned} R_{\text{tot}}(i) &= \sum_v R_{uv} = \sum_u L_{uu}^+ + L_{vv}^+ - 2L_{uv}^+ \\ &= NL_{uu}^+ + \text{Tr}(\mathbf{L}^+) - 2 \sum_v L_{uv}^+ \\ &= NL_{uu}^+ + \text{Tr}(\mathbf{L}^+) \end{aligned} \quad (26)$$

using Ellens et al. (2011, Theorem 4.2) and Bozzo and Franceschet (2013, Eq. 15). Here, L_{uu}^+ indicates how close node u is on expectancy to all the nodes, and a global component, $\text{Tr}(\mathbf{L}^+)$, proportional to the average pairwise distance in the network, which is denotes of how large is the network overall, independently of the particular node u .

Based on the introduced $R_{\text{tot}}(u)$, we propose group isolation which is obtained as the expectation in $R_{\text{tot}}(u)$ for all the nodes in the group:

$$\begin{aligned} R_{\text{tot}}(S_i) &= \mathbb{E}_{u \sim S_i} [R_{\text{tot}}(u)] = |\mathcal{V}| \mathbb{E}_{u \sim S_i} \left[\frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} R_{uv} \right] \\ &= |\mathcal{V}| \mathbb{E}_{u \sim S_i} [\mathbb{E}_{v \sim \mathcal{V}} [R_{uv}]] \\ &= |\mathcal{V}| \mathbb{E}_{u \sim S_i, v \sim \mathcal{V}} [R_{uv}] \end{aligned} \quad (27)$$

$R_{\text{tot}}(S_i)$ is computed as the expectation of $R_{\text{tot}}(u)$ for all nodes in group S_i :

$$\begin{aligned} R_{\text{tot}}(S_i) &= |\mathcal{V}| \mathbb{E}_{u \sim S_i, v \sim \mathcal{V}} [R_{uv}] \\ &= |\mathcal{V}| \frac{1}{|S_i|} \sum_{u \in S_i} \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} R_{uv} \\ &= \frac{1}{|S_i|} \sum_{u \in S_i} \sum_{v \in \mathcal{V}} R_{uv} = \frac{1}{|S_i|} \sum_{u \in S_i} R_{\text{tot}}(u) \\ &= \mathbb{E}_{u \sim S_i} [R_{\text{tot}}(u)] \end{aligned} \quad (28)$$

Finally, the cumulative group isolation across all groups in the graph fulfills the following equality with the total effective resistance of the graph.

$$\begin{aligned} \frac{1}{2} \sum_{i \in SA} |S_i| R_{\text{tot}}(S_i) &= \frac{1}{2} \sum_{i \in SA} |S_i| \frac{1}{|S_i|} \sum_{u \in S_i} \sum_{v \in \mathcal{V}} R_{uv} \\ &= \frac{1}{2} \sum_{v \in \mathcal{V}} \sum_{u \in \mathcal{V}} R_{uv} = R_{\text{tot}} = n \text{Tr}(\mathbf{L}^\dagger) \end{aligned} \quad (29)$$

Therefore, since R_{tot} is related to the information flow, $R_{\text{tot}}(S_i)$ measures the ease of information flow through group S_i in the graph.

The optimal — yet extreme — scenario of maximum information flow in a graph is such where all nodes in the graph are connected. Hence, G will be a fully connected graph. In this scenario, the total effective resistance of the graph reaches its minimum $n - 1$, where $n = |\mathcal{V}|$. The total effective resistance of all existing edges in the graph is $n - 1$, and given that we have all connections, R_{tot} also sums up to $n - 1$. Therefore, every pair of nodes will be separated by $R_{uv} = 2/n$. In this scenario, $R_{\text{tot}}(u) = (2/n)(n - 1) = 2 - 2/n$ for all nodes in the graph. It leads to an group isolation of $R_{\text{tot}}(S_i) = 2 - 2/n$ for all the different groups in the graph and therefore a Isolation disparity $\Delta R_{\text{tot}} = 0$.

Regarding group isolation disparity, Equation 5 can be re-defined using Equation 28. It is equivalent to the equality for all groups of the mean R_{uv} between of the nodes in the group and all nodes in the graph:

$$\begin{aligned} R_{\text{tot}}(S_i) &= R_{\text{tot}}(S_j), \forall i, j \in SA & (30) \\ \mathbb{E}_{u \sim S_i, v \sim \mathcal{V}} [R_{uv}] &= \mathbb{E}_{u \sim S_j, v \sim \mathcal{V}} [R_{uv}], \forall i, j \in SA \times SA \end{aligned}$$

Group Control and Control Disparity We provide the proof for the bounds associated with the proposed control metrics.

Bounds of $B_R(u)$ and $B_R(S_i)$ We show the proof of the bounds of the node and group control.

Theorem B.1. *The control of a node is bounded by $1 \leq B_R(u) \leq d_u$, being d_u the degree of node u . Equality on the upper bound holds when all the edges are cut edges, i.e., edges that if removed the graph would become disconnected.*

Proof. The resistance curvature of a node is known to be bounded by $1 - d_u/2 \leq p_u \leq 1/2$ (Devriendt and Lambiotte 2022), and the relation between the curvature and group control is given by the equality $p_u = 1 - \frac{1}{2} B_R(u)$. Therefore, we obtain the bounds as

$$\begin{aligned} 1 - \frac{d_u}{2} &\leq p_u \leq \frac{1}{2} \rightarrow \\ 1 - \frac{d_u}{2} &\leq 1 - \frac{1}{2} B_R(u) \leq \frac{1}{2} \rightarrow \\ -d_u &\leq -B_R(u) \leq -1 \rightarrow \\ 1 &\leq B_R(u) \leq d_u \end{aligned}$$

□

Theorem B.2. *The control of a group is bounded by $1 \leq B_R(S_i) \leq \frac{\text{vol}(S_i)}{|S_i|}$, being $\text{vol}(S_i)$ the sum of the degrees of node u . Thus, $\text{vol}(S_i)/|S_i|$ is the average degree of all the nodes in S_i . Equality on the upper bound holds when the subgraph with all nodes of S_i and their neighbors is a tree graph, i.e., a connected acyclic undirected graph.*

Proof. The control of a group is defined as $B_R(S_i) = \mathbb{E}_{u \sim S_i} [B_R(u)]$ and using Theorem B.1, we derive the bounds

of $B_R(S_i)$ as follows:

$$\begin{aligned} 1 &\leq B_R(u) \leq d_u \rightarrow \\ 1 &\leq \mathbb{E}_{u \sim S_i} [B_R(u)] \leq \mathbb{E}_{u \sim S_i} [d_u] \\ 1 &\leq B_R(S_i) \leq \frac{\text{vol}(S_i)}{|S_i|} \end{aligned} \tag{31}$$

□

Control as an Allocation Problem Group control is ($B_R(S_i)$) a limited resource to be distributed in the network. The sum of R_{uv} for every edge always equals to $|\mathcal{V}| - 1$ (Eilens et al. 2011):

$$\sum_{(u,v) \in \mathcal{E}} R_{uv} = |\mathcal{V}| - 1.$$

Therefore, the sum of the node controls in the graph is defined as

$$\sum_{u \in \mathcal{V}} B_R(u) = \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{N}(u)} R_{uv} = 2 \times \sum_{(u,v) \in \mathcal{E}} R_{uv} = 2|\mathcal{V}| - 2,$$

and the expectation as:

$$\mathbb{E}_{u \sim \mathcal{V}} [B_R(u)] = 2 - \frac{2}{|\mathcal{V}|}$$

independently of the number of edges (density) of the graph.

As a consequence of the definition of group control (Equation 4), the weighted sum of group control for all groups in the graph and the weighted mean also remain constant at:

$$\sum_{S_i \in \mathcal{V}} |S_i| \times B_R(S_i) = 2|\mathcal{V}| - 2$$

and

$$\frac{1}{|\mathcal{V}|} \sum_{S_i \in \mathcal{V}} |S_i| \times B_R(S_i) = 2 - \frac{2}{|\mathcal{V}|}.$$

B.2 Group Isolation vs Group Information Share

The group isolation $R_{\text{tot}}(S_i)$ is given by Equation 2 and it is proportional to the expected information distance when sampling one node from group S_i and another node at random. Thus, $R_{\text{tot}}(S_i)$ quantifies the information access of a group of nodes in the network.

Conversely, the amount of information that a group of nodes S_i shares with the rest of the network is given by the expected information distance between a random node v from the network and a node in group S_i . We define this metric as the *Group Information Share* ($R_{\text{From-}S_i}(V)$).

Mathematically, both concepts are equivalent as $R_{\text{tot}}(S_i)$ not only quantifies how isolated a group is (in terms of information access), but also how much it contributes to the information flow that reaches other nodes in the network. Therefore, the disparity in information access between groups and the disparity in information sharing between groups on a randomly selected node in the network are also equivalent.

Mathematical Equivalence The Group Isolation is defined as the expected total effective resistance (or information distance, R_{uv}) between a randomly chosen node in group S_i to all other nodes in the network:

$$R_{\text{tot}}(S_i) = \mathbb{E}_{u \sim S_i} [R_{\text{tot}}(u)] = |\mathcal{V}| \mathbb{E}_{u \sim S_i} [\mathbb{E}_{v \sim \mathcal{V}} [R_{uv}]], \quad (32)$$

where $R_{\text{tot}}(u)$ represents the total effective resistance from node u to all other nodes.

Similarly, the Group Information Share measures the expected information distance from a randomly selected node in the network to a group S_i :

$$R_{\text{From-}S_i}(V) = \mathbb{E}_{v \sim \mathcal{V}} [R_{\text{From-}S_i}(v)] = |\mathcal{V}| \mathbb{E}_{v \sim \mathcal{V}} [\mathbb{E}_{u \sim S_i} [R_{uv}]] \quad (33)$$

Thus, the expected information distance when a random node is sampled from group S_i is the same as when a random node is sampled from the network and connected to group S_i :

$$|\mathcal{V}| \mathbb{E}_{u \sim S_i} [\mathbb{E}_{v \sim \mathcal{V}} [R_{uv}]] = |\mathcal{V}| \mathbb{E}_{v \sim \mathcal{V}} [\mathbb{E}_{u \sim S_i} [R_{uv}]] \quad (34)$$

$$R_{\text{tot}}(S_i) = R_{\text{From-}S_i}(V) \quad (35)$$

Group Isolation Disparity In terms of disparity between groups, the Group Isolation Disparity is defined as the difference in isolation between two groups S_i and S_j :

$$\Delta R_{\text{tot}} = R_{\text{tot}}(S_i) - R_{\text{tot}}(S_j) \quad (36)$$

Similarly, the Group Information Share Disparity measures the disparity in information contribution between groups:

$$\Delta R_{\text{From-}S}(V) = R_{\text{From-}S_i}(V) - R_{\text{From-}S_j}(V) \quad (37)$$

Since $R_{\text{tot}}(S_i) = R_{\text{From-}S_i}(V)$, it follows that:

$$\Delta R_{\text{tot}} = \Delta R_{\text{From-}S_i}(V) \quad (38)$$

Thus, the disparities in Group Isolation and in Group Information Share are the same, highlighting the dual role of the proposed metric in capturing both access to information from the network and information sharing with the rest of the network.

B.3 Efficient Version of ERG

Algorithm 2 shows an efficient manner to update the pseudo-inverse of the Laplacian after adding one edge to the graph. Therefore, we avoid the computation of \mathbf{L}^\dagger after every edge addition. \mathbf{L}^\dagger is easily updated using the Woodbury’s formula which is based on the values of \mathbf{L}^\dagger (see Black et al. (2023) for a proof).

C Dataset Statistics

Number of male and female nodes Table 3 shows a breakdown of the number of nodes, edges, graph density and number of males and females in each of the studied datasets. The Facebook and UNC28 datasets have a majority of males whereas the Google+ dataset has a majority of females. Interestingly, the Google+ dataset has the highest levels of disparity in group social capital (see Table 2) which means that, although there are fewer males in the network, they are better positioned than females with respect to their access to information and thus social capital.

Algorithm 2: ERG-Link

Data: Graph $G = (\mathcal{V}, \mathcal{E})$, a protected attribute SA , budget B of total number of edges to add

Result: Rewired Graph $G' = (\mathcal{V}', \mathcal{E}')$

- 1 $\mathbf{L} = \mathbf{D} - \mathbf{A}$;
 - 2 $S_d = \text{argmax}_{S_i, v_i \in SA} R_{\text{tot}}(S_i)$;
 - 3 $\mathbf{L}^\dagger = \sum_{i>0} \frac{1}{\lambda_i} \phi_i \phi_i^\top = \left(\mathbf{L} + \frac{\mathbf{1}\mathbf{1}^\top}{n} \right)^{-1} - \frac{\mathbf{1}\mathbf{1}^\top}{n}$;
 - 4 **Repeat**
 - 5 $\mathbf{R} = \mathbf{1} \text{diag}(\mathbf{L}^\dagger)^\top + \text{diag}(\mathbf{L}^\dagger) \mathbf{1}^\top - 2\mathbf{L}^\dagger$;
 - 6 $C = \{(u, v) \mid u \in S_d \text{ or } v \in S_d, (u, v) \notin E\}$;
 - 7 // Select edge candidates
 $\mathcal{E}' = \mathcal{E}' \cup \text{argmax}_{(u,v) \in C} R_{uv}$;
 - 8 // Fast update of \mathbf{L} and \mathbf{L}^\dagger
 $\mathbf{L} = \mathbf{L} + (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^\top$;
 - 9 $\mathbf{L}^\dagger = \mathbf{L}^\dagger - \frac{1}{1+R_{uv}} \times (\mathbf{L}_{u,:}^\dagger - \mathbf{L}_{v,:}^\dagger) \otimes (\mathbf{L}_{u,:}^\dagger - \mathbf{L}_{v,:}^\dagger)$;
 - 10 // updated by Woodbury
 - 10 **Until** $|\mathcal{E}' \setminus \mathcal{E}| = B$;
 - 11 **return** G' ;
-

Dataset	Facebook	UNC28	Google+
$ V $	1,034	3,985	3,508
$ \mathcal{E} $	26,749	65,287	253,930
Density	0.05	0.008	0.04
# Males	685	2307	1312
# Females	349	1678	2196
h_{edge}	0.58	0.55	0.51

Table 3: Dataset Statistics

Edge homophily We also report the edge homophily (Zhu et al. 2020) for the three datasets. Edge homophily, denoted as $h_{edge} \in [0, 1]$, represents the percentage of edges connecting nodes with the same protected attribute. The larger the group homophily, the more internally connected the groups are, with fewer connections between different groups. The obtained values of edge homophily suggest that the groups are well interconnected, and thus, there is no significant community polarization between males and females. In consequence, the differences between groups in access and control of information flow, as shown in Table 2, stem from disparities in roles and network positions, rather than from a polarized social structure.

Degree Distribution Figure 5 shows the degree distribution for the three different datasets grouped by protected attribute. As seen in the Figure, the number of high degree nodes is larger for males than females in the Facebook and Google+ datasets. Note that although there are fewer males than females in the Google+ dataset, the majority of nodes with high degree correspond to males, which explains their larger group social capital.

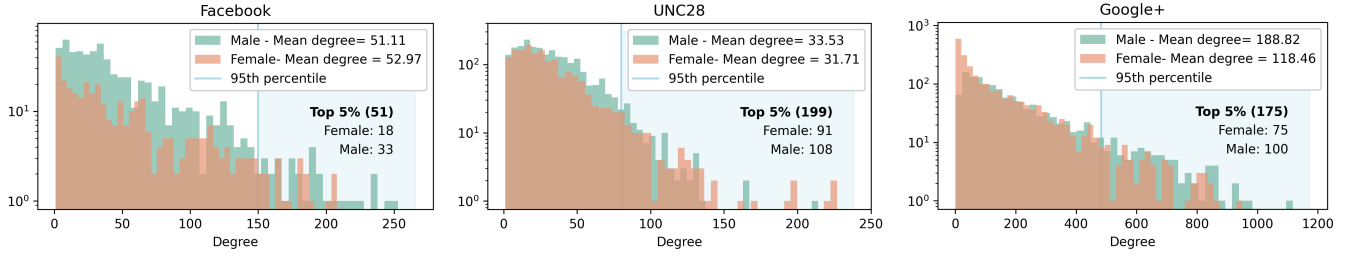


Figure 5: Original degree distribution per dataset and gender. We also include the average degree per group and the composition of the top 5% of nodes.

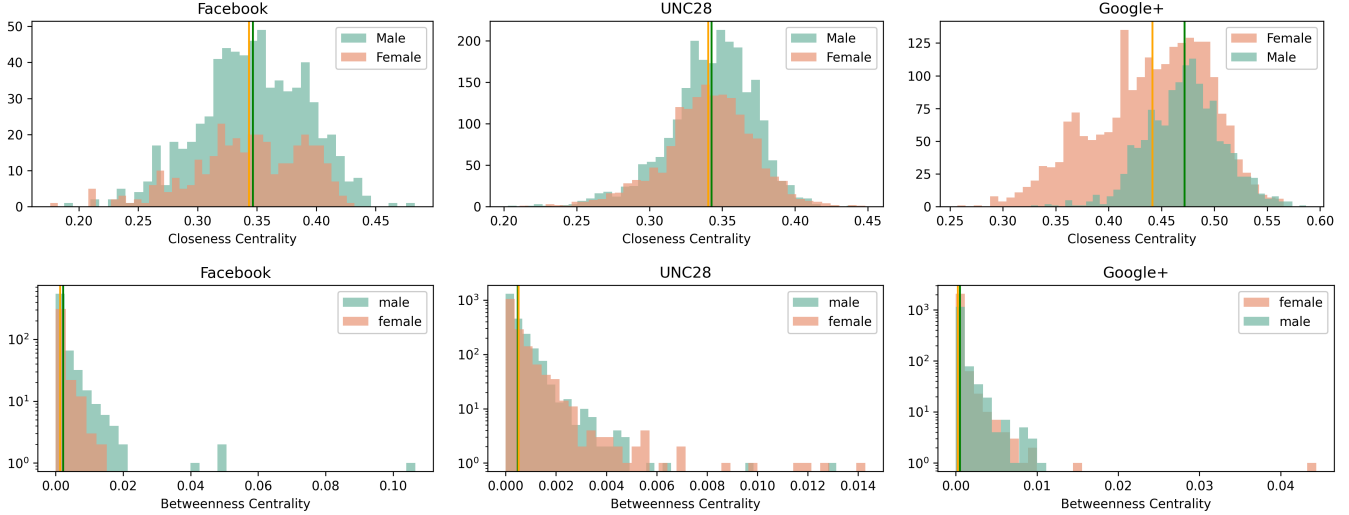


Figure 6: Per-group centrality metrics in the original datasets.

Centrality Metrics Figure 6 depicts the per-group closeness and betweenness centrality metrics (Barabási 2013) for the three datasets, providing an insight into the structural properties of the networks. However these classical metrics are based on geodesic distances and are therefore limited in their ability to capture the entire network topology. For example, the geodesic distance is a poor estimator of long-range connections. In contrast, the proposed metrics, which are built on random walks and consider the entire network, offer a more complete view of information flow within the network (see 2). Nevertheless, even with classical metrics, we observe notable differences between groups, particularly in the Google+ dataset, where males exhibit larger closeness centrality despite being a minority.

D Additional Experiments

All the experiments were done using a workstation with 16GB RAM; and processor 11th Gen Intel(R) Core(TM) i7, 3.00GHz, 2995 Mhz, 4 Core(s), 8 Logical Processor(s).

D.1 Group Social Capital Metrics

The experiments presented in the main paper illustrate how edge augmentation via ERG-Link is able to (1) significantly mitigate the structural group unfairness (Table 2); and

(2) increase the social capital for all the groups in the graph (Figure 3). Table 4 depicts the group social capital for each group after each intervention. Note that the optimal value of B_R is to get every group’s control equal to $2 - 2/|\mathcal{V}|$.

D.2 Distribution of Social Capital by Group

For completeness, Figure 8 illustrates the distributions of all effective resistances R_{uv} and the node’s social capital metrics $-R_{\text{tot}}(u)$, $\mathcal{R}_{\text{diam}}(u)$ and $B_R(u)$ —in the graph before and after the edge augmentation interventions for each of the groups.

The plots correspond to two different edge augmentation experiments on the Facebook dataset, with budgets of $B=50$ and $B=5,000$ new edges. Edge augmentation via ERG-Link is able to drastically reduce all effective resistances of the graph, unlike the other methods and even for the small budget. Note how there is still a long tail in the distribution of effective resistances after edge augmentation with the baseline methods, which illustrates that there are still nodes that struggle to exchange information even after the intervention.

Regarding the node social capital metrics, edge augmentation via ERG-Link reduces all $R_{\text{tot}}(u)$ and $\mathcal{R}_{\text{diam}}(u)$, which explains why $R_{\text{tot}}(S_i)$ and $\mathcal{R}_{\text{tot}}(S_i)$ is improved for all

(a) Facebook (50) Female				(b) UNC28 (5,000) Female				(c) Google+ (5,000) Female			
G	$R_{tot} \downarrow$	$\mathcal{R}_{diam} \downarrow$	B_R	G	$R_{tot} \downarrow$	$\mathcal{R}_{diam} \downarrow$	B_R	G	$R_{tot} \downarrow$	$\mathcal{R}_{diam} \downarrow$	B_R
Random	211.5	2.26	1.927	Random	435.0	1.23	1.992	Random	305.1	1.07	1.824
SDRF	220.7	2.28	1.928	SDRF	599.1	2.11	1.996	SDRF	561.1	1.31	1.805
FOSR	202.1	2.15	1.926	FOSR	509.9	1.33	1.990	FOSR	498.9	1.25	1.811
DW	199.4	1.87	1.929	DW	583.2	1.81	1.997	DW	558.7	1.31	1.806
Cos	190.4	1.64	1.918	Cos	429.6	0.42	1.940	Cos	230.9	0.29	1.827
ERG	138.7	0.43	1.933	ERG	316.8	0.10	2.001	ERG	145.5	0.07	1.889

(d) Facebook (50) Male				(e) UNC28 (5,000) Male				(f) Google+ (5,000) Male			
G	$R_{tot} \downarrow$	$\mathcal{R}_{diam} \downarrow$	B_R	G	$R_{tot} \downarrow$	$\mathcal{R}_{diam} \downarrow$	B_R	G	$R_{tot} \downarrow$	$\mathcal{R}_{diam} \downarrow$	B_R
Random	172.8	2.22	2.035	Random	415.2	1.23	2.005	Random	175.7	1.03	2.293
SDRF	179.1	2.24	2.034	SDRF	576.9	2.10	2.002	SDRF	285.2	1.23	2.325
FOSR	167.6	2.13	2.035	FOSR	490.2	1.33	2.007	FOSR	258.1	1.17	2.315
DW	163.1	1.83	2.033	DW	561.0	1.81	2.001	DW	284.6	1.24	2.323
Cos	161.7	1.61	2.039	Cos	410.6	0.41	2.043	Cos	144.1	0.27	2.288
ERG	128.5	0.41	2.031	ERG	308.0	0.09	1.998	ERG	108.5	0.06	2.185

Table 4: Group Social Capital after Edge Augmentation. The best values are highlighted in bold. Note how edge augmentation via ERG-Link is able to not only increase the social capital of the disadvantaged group (females) but also of the rest of the groups (males).

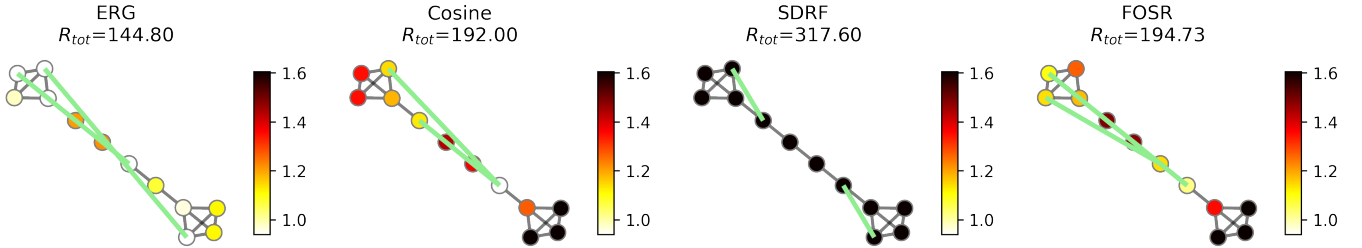


Figure 7: Example of the added links by the different algorithms on a Barbell synthetic graph. Nodes are colored by $R_{tot}(u)$. More examples in Fig 12.

groups in the graph, and particularly for the disadvantaged group. Thus, ΔR_{tot} and $\Delta \mathcal{R}_{diam}$ are significantly reduced.

D.3 Evolution of Group Social Capital During the Interventions

We provide additional results regarding the evolution of group social capital metrics during the graph’s intervention, as initially shown in Figure 4. The goal is to analyze the effectiveness of the edge augmentation methods on both small ($B = 50$ edges) and large ($B = 5,000$ edges) budget scenarios. Figures 9 and 10 show the evolution of both the group social capital metrics for each group and the structural group unfairness metrics on the Facebook dataset.

We observe in Figure 9 how edge augmentation via ERG-Link is able to significantly improve the group social capital for all groups and reduce all structural unfairness metrics. In contrast, the baselines fail to do so.

Figure 10 depicts the evolution of group social capital and disparity metrics when adding 5,000 edges to the Facebook dataset. The more edges we add to a graph, the denser the graph becomes and therefore the better the information

flow. Thus, one can expect that after a large number of added edges, all methods behave similarly.

However, we observe some differences. First, the convergence to minimal isolation and diameter disparities is significantly faster when adding edges via ERG-Link than any other method. Second, the decrease in $R_{tot}(S_i)$ is very significant for both groups (males and females) even after just adding a small number of edges by means of ERG-Link. Third, regarding group control and control disparity, edge augmentation via ERG-Link systematically reduces ΔB_R while converging each group control to the optimal $B_R(S_i) = 2 - 2/|\mathcal{V}|$.

We also show in Figure 11 the evolution of $B_R(S_i)$ and ΔB_R on the UNC28 dataset with 1,000 edge additions along with the distribution of node’s control $B_R(u)$ after the intervention to showcase an scenario where ERG-Link is able to reach the optimal control disparity in the graph, allocating the same amount of control for each group, $B_R(S_i) \approx 2 - 2/|\mathcal{V}| \forall i \in SA \rightarrow \Delta B_R \approx 0$.

Last but not least, the different baselines do not reach a better behavior than the random method, neither on the

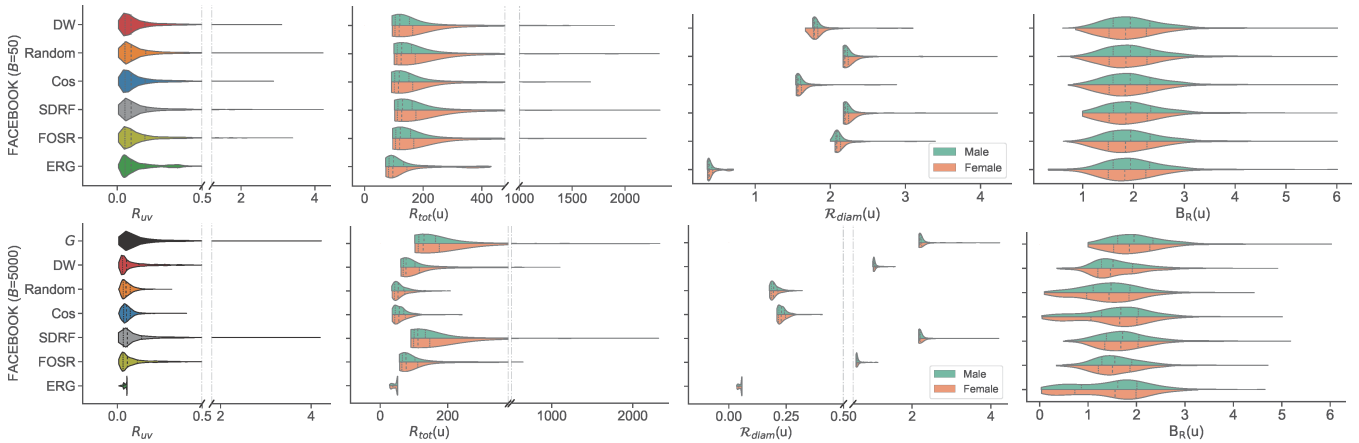


Figure 8: Distribution of R_{uv} and proposed social capital metrics on the Facebook dataset and after different graph interventions, 50 links in the top row and 5,000 in the bottom row. Columns show (left) distribution of all R_{uv} distances, (center-left) distribution of all $R_{\text{tot}}(u)$, (center-right) distribution of all $R_{\text{diam}}(u)$, (right) distribution of all $B_R(u)$. The distributions for the node metrics are shown for the two groups according to the protected attribute (gender).

group social capital metrics nor in terms of structural unfairness. After 5,000 edge additions, the random method significantly reduces the unfairness metrics ΔR_{tot} and ΔR_{diam} while also achieving decent rated of $R_{\text{tot}}(S_i)$ and $R_{\text{diam}}(S_i)$ since the budget is high enough to improve the information flow with no strategy. However, all baselines struggle to optimize $\Delta B_R(S_i)$ and $B_R(S_i)$. None of them is able to improve, and the cosine similarity approach even increased the Control Disparity.

D.4 Additional Edge Augmentation Examples

For completeness, we qualitatively illustrate on Figures 7 and 12 the behavior of four edge augmentation methods on five synthetic graphs. The Figures depict the synthetic graphs and the first two edges added by each algorithm. They also show quantitatively how the methods are able to improve the overall flow of information in the graph, defined as $R_{\text{tot}} = \sum_{(u,v) \in \mathcal{E} R_{uv}}$. This analysis not only provides insight into the performance of the edge augmentation methods, but also demonstrates their adaptability to different graph topologies.

The Figures correspond to graphs from five synthetic datasets, each designed to represent different graph characteristics: (1) a Barbell graph with Barbell size of 4 and path size of 5; (2) a power-law clustered graph with 50 nodes, 2 random new edges per each new node, and a 0.95 probability of closing a triangle; (3) the same graph, but with a 0.8 probability of closing a triangle, meaning that the bottleneck and clustered structures are weaker; (4) an SBM model with intra-cluster probability of 0.4 and inter-cluster probability of 0.01; and (5) a hand-crafted path graph with communities in the middle to simulate a different type of bottleneck. The Barbell graph is shown in Figure 7 and the rest of the graphs in Figure 12.

The Figures illustrate the results for all algorithms described in Section 4 except for DeepWalk because its behavior closely resembles randomness. This similarity arises

from the impracticality of performing extensive training after each edge addition, given the computational limitations of the proposed task. Performing extensive training after each edge addition would be computationally infeasible within the scope of our study.

Observe how ERG-Link always reduces R_{tot} the most by adding the edges with the maximum distance. Other approaches, such as FOSR, do not exhibit a consistent performance across different graph topologies. In fact, graphs with complex structures and/or less defined bottlenecks exhibit a higher degree of difficulty for this method. Cosine similarity suffers from the same disadvantage. SDRF adds edges around the graph’s bottleneck to avoid over-squashing, but this strategy does not necessarily translate into an improvement of the graph’s information flow.

E Computation Time

Table 5 depicts the number of edge additions per second that each algorithm is able to perform. For instance, in the conducted experiment on the Facebook dataset with $B = 5,000$ edges, edge augmentation via ERG-Link required approximately 3 minutes and 30 seconds; via Cosine Similarity (Cos), approximately 6 minutes; and by means of DeepWalk (DW), over 3 hours. In the case of the Google+ dataset, edge augmentation via ERG-Link required 36 minutes; via Cosine Similarity (Cos), 72 minutes; and by means of DeepWalk (DW), over 17 hours.

Dataset	Facebook	Google+	UNC28
SDRF	1.00	0.25	0.46
FOSR	166.6	125.0	83.3
DeepWalk	0.39	0.08	0.07
Cosine	13.9	1.15	1.39
ERG	23.8	2.31	1.66

Table 5: Added edges per second on the three datasets

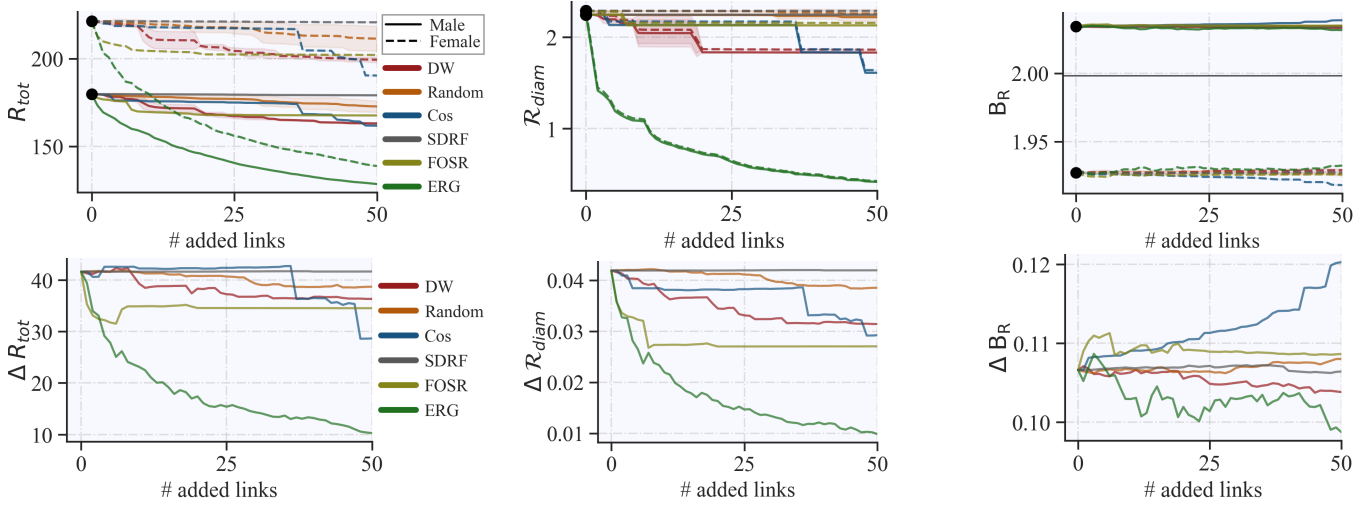


Figure 9: Evolution of group social capital and disparity metrics for both groups as the number of added edges increases on the Facebook dataset with a budget of 50 new edges.

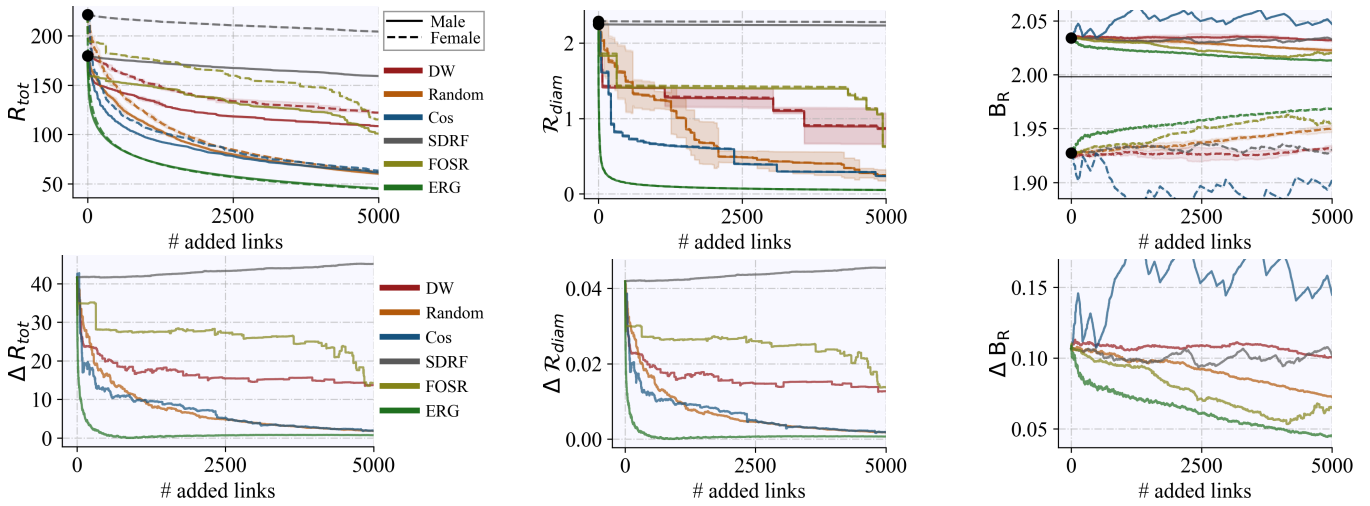


Figure 10: Evolution of group social capital metrics for both groups and fairness metrics as the number of added links increases, on Facebook dataset after adding 5,000 edges. Edge augmentation via ERG-Link exhibits a faster rate of convergence to the optimal scenario than the baselines.

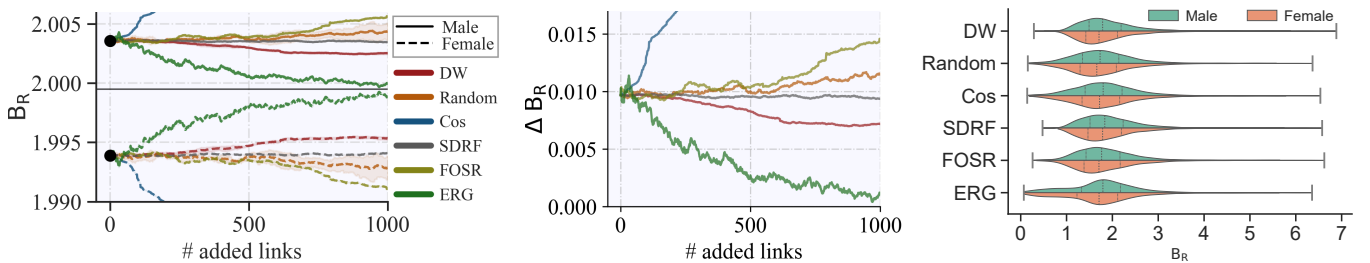


Figure 11: Behavior of node, group control and control disparity ($B_R(u)$, $B_R(S_i)$, ΔB_R). Illustration of the evolution of the group control (left, $B_R(S_i)$) and control disparity (middle, ΔB_R) through an edge augmentation with $B = 1,000$ edges on the UNC dataset. Right-most figure: distribution of the nodes' control ($B_R(u)$) for each group. Note how edge augmentation via ERG-Link yields a control for all groups approaching their optimal value of $2 - 2/|\mathcal{V}|$ by reducing the control of the privileged group (males) while increasing the control of the disadvantaged group (females).

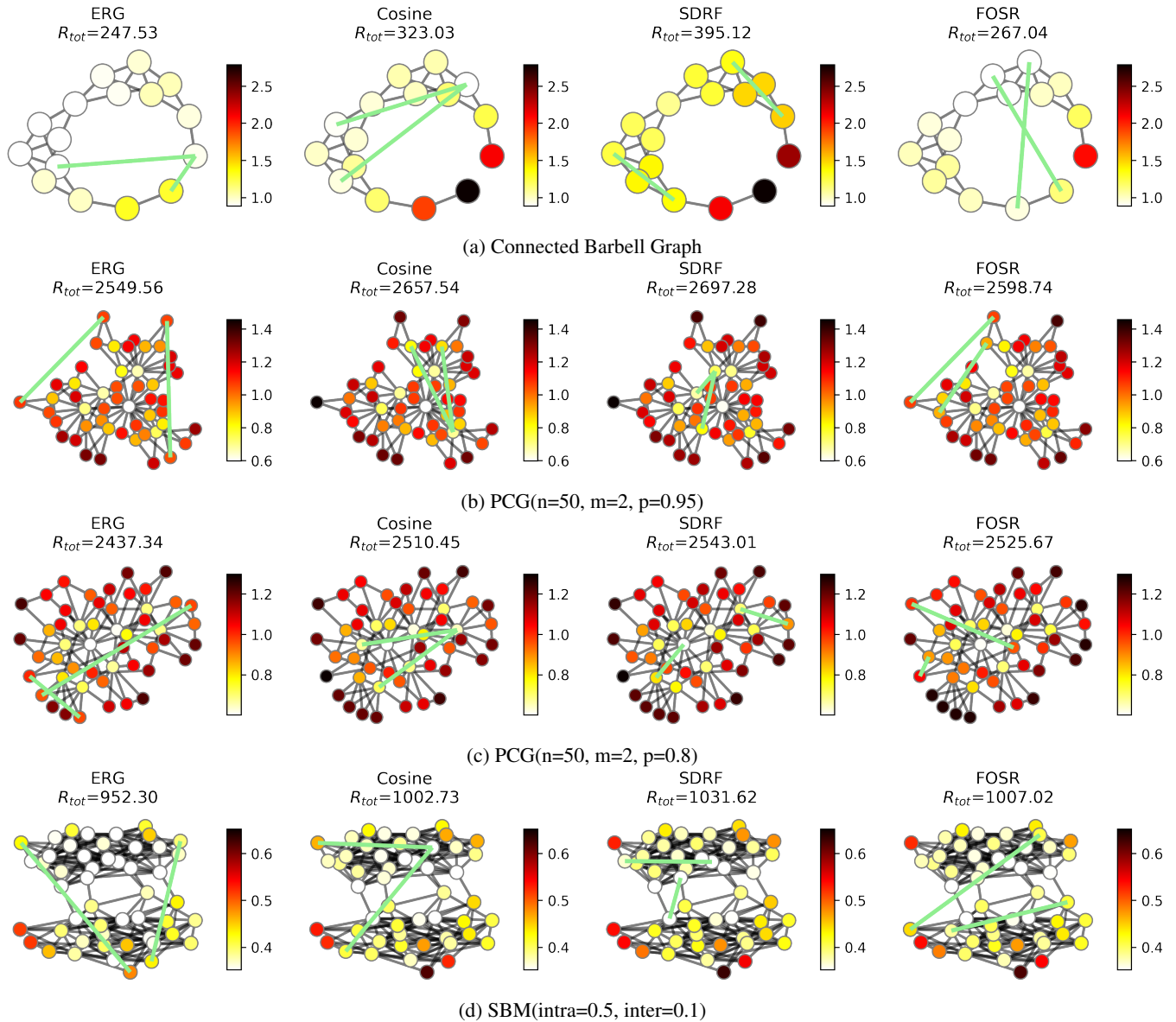


Figure 12: Examples of the added links by the ERG-Link, Cosine, SDRF and FOSR algorithms on synthetic graphs. Nodes are colored by $R_{tot}(u)$. Edge augmentation via ERG-Link consistently yields the largest improvement in information flow in the graph and hence the smallest values of R_{tot} .

Symbol	Description	Definition
$G = (\mathcal{V}, \mathcal{E})$	Graph = (Nodes, Edges)	
$n = \mathcal{V} $	Number of nodes	
\mathbf{A}	Adjacency matrix: $\mathbf{A} \in \mathbb{R}^{n \times n}$	$A_{u,v} = 1$ if $(u, v) \in \mathcal{E}$ and 0 o/w
v or u	Node $v \in \mathcal{V}$ or $u \in \mathcal{V}$	
e	Edge $e \in \mathcal{E}$	
d_v	Degree of node, i.e., number of neighbors v	$d_v = \sum_{u \in \mathcal{V}} A_{v,u}$
\mathbf{D}	Degree diagonal matrix where d_v in D_{vv}	$\mathbf{D} = \text{diag}(d_0, \dots, d_{\mathcal{V}})$
$\text{vol}(G)$	Sum of the degrees of the graph	$\text{vol}(G) = \sum_{u \in \mathcal{V}} d_u = 2 \mathcal{E} = \text{Tr}[\mathbf{D}]$
$\mathcal{N}(u)$	Neighbors of u	$\mathcal{N}(u) = \{v : (u, v) \in \mathcal{E}\}$
S_i	Subset of nodes	$S \subseteq V$
SA	Set of sensitive attributes	$SA = \{sa_1, sa_2, \dots, sa_{ SA }\}$
$SA(v)$	Value of the sensitive attribute of the node v	
sa_i	Specific value of a sensitive attribute, e.g., $sa_i = \text{female}$.	
S_i	Set of nodes defined by their sensitive attribute	$S_i = \{v \in V SA(v) = sa_i\}$
S_d	Set of nodes defined by their sensitive attribute with the highest level of isolation $R_{\text{tot}}(S_i)$	
\mathbf{L}	Graph Laplacian	$\mathbf{L} = \mathbf{D} - \mathbf{A} = \Phi \Lambda \Phi^\top$
Λ	Eigenvalue matrix of \mathbf{L}	
Φ	Matrix of eigenvectors of \mathbf{L}	
λ_i	The i -th smallest eigenvalue of \mathbf{L}	
\mathbf{f}_i	Eigenvector associated with the i -th smallest eigenvalue of \mathbf{L}	
\mathbf{L}^+	The pseudo-inverse of \mathbf{L}	$\mathbf{L}^+ = \sum_{i>1} \lambda_i^{-1} \phi \phi^\top$
h_G	Cheeger constant	Eq. 10
\mathbf{e}_u	Unit vector with unit value at u and 0 elsewhere	
R_{uv}	Effective resistance between nodes u and v	$R_{uv} = (\mathbf{e}_u - \mathbf{e}_v) \mathbf{L}^+ (\mathbf{e}_u - \mathbf{e}_v)$
\mathbf{R}	Effective resistance matrix where the i, j entry corresponds to R_{ij}	$\mathbf{R} = \mathbf{1} \text{diag}(\mathbf{L}^+)^\top + \text{diag}(\mathbf{L}^+) \mathbf{1}^\top - 2\mathbf{L}^+$
\mathbf{Z}	Commute Time Embedding matrix	$\mathbf{Z} = \sqrt{\text{vol}(G)} \Lambda^{-1/2} \Phi^\top$
\mathbf{z}_u	Commute times embedding of node Z_u :	
$\text{CT}(u, v)$	Commute time	$\text{CT}(u, v) = \text{vol}(G) R_{u,v}$
R_{tot}	Total Effective Resistance of G	$R_{\text{tot}} = \frac{1}{2} \mathbf{1}^\top \mathbf{R} \mathbf{1}$
$\mathcal{R}_{\text{diam}}$	Resistance Diameter of G	$\mathcal{R}_{\text{diam}} = \max_{u,v \in \mathcal{V}} R_{u,v}$
$R_{\text{tot}}(u)$	Node Isolation or Total Effective Resistance	$R_{\text{tot}}(u) = \sum_{v \in \mathcal{V}} R_{uv}$
$\mathcal{R}_{\text{diam}}(u)$	Node Resistance Diameter	$\mathcal{R}_{\text{diam}}(u) = \max_{v \in \mathcal{V}} R_{uv}$
$\mathbf{B}_R(u)$	Node Control or Resistance Betweenness	$\mathbf{B}_R(u) = \sum_{v \in \mathcal{N}(u)} R_{uv}$
$R_{\text{tot}}(S_i)$	Group Isolation or Total Effective Resistance	$R_{\text{tot}}(S_i) = S ^{-1} \sum_{u \in S} R_{\text{tot}}(u)$
$\mathcal{R}_{\text{diam}}(S_i)$	Group Resistance Diameter	$\mathcal{R}_{\text{diam}}(S) = S ^{-1} \sum_{u \in S} \mathcal{R}_{\text{diam}}(u)$
$\mathbf{B}_R(S_i)$	Group Control or average Betweenness	$\mathbf{B}_R(S) = S ^{-1} \sum_{u \in S} \mathbf{B}_R(u)$

Table 6: Table of Notation