Reducing Access Disparities in Networks using Edge Augmentation*

Ashkan Bashardoust
ashkanb@cs.utah.edu
University of Utah
Salt Lake City, UT, USA

Sorelle A. Friedler
sorelle@cs.haverford.edu
Haverford College
Haverford, PA, USA

Carlos E. Scheidegger
carlos.scheidegger@gmail.com
Posit PBC
Tucson, AZ, USA

Blair D. Sullivan
sullivan@cs.utah.edu
University of Utah
Salt Lake City, UT, USA

Suresh Venkatasubramanian
suresh_venkatasubramanian@brown.edu
Brown University
Providence, RI, USA

ABSTRACT

In social networks, a node's position is, in and of itself, a form of *social capital*. Better-positioned members not only benefit from (faster) access to diverse information, but innately have more potential influence on information spread. Structural biases often arise from network formation, and can lead to significant disparities in information access based on position. Further, processes such as link recommendation can exacerbate this inequality by relying on network structure to augment connectivity.

In this paper, we argue that one can understand and quantify this social capital through the lens of information flow in the network. In contrast to prior work, we consider the setting where all nodes may be sources of distinct information, and a node's (dis)advantage takes into account its ability to access all information available on the network, not just that from a single source. We introduce three new measures of advantage (broadcast, influence, and control), which are quantified in terms of position in the network using *access signatures* – vectors that represent a node's ability to share information with each other node in the network. We then consider the problem of improving equity by making interventions to increase the access of the least-advantaged nodes. Since all nodes are already sources of information in our model, we argue that edge augmentation is most appropriate for mitigating bias in the network structure, and frame a budgeted intervention problem for maximizing broadcast (minimum pairwise access) over the network.

Finally, we propose heuristic strategies for selecting edge augmentations and empirically evaluate their performance on a corpus of real-world social networks. We demonstrate that a small number of interventions can not only significantly increase the broadcast measure of access for the least-advantaged nodes (over 5 times more than random), but also simultaneously improve the minimum

influence. Additional analysis shows that edge augmentations targeted at improving minimum pairwise access can also dramatically shrink the gap in advantage between nodes (over 82%) and reduce disparities between their access signatures.

CCS CONCEPTS

• **Theory of computation → Graph algorithms analysis**; • **Information systems → Social networks**; Social recommendation.

KEYWORDS

algorithmic fairness; information access; social networks; edge interventions

1 INTRODUCTION

One of the promises of a highly-connected world is an impartial spread of opinions driven by free and unbiased sources of information, leading to an equitable exposure of opinion to the wide public. On the contrary, the social network platforms currently governing news diffusion, while offering many seemingly-desired features like search, personalization, and recommendation, are reinforcing the centralization of information spread and the creation of so-called echo chambers and filter bubbles [5]. A person's position within these networks often determines their access to information and opportunities such as jobs, education, and health information [13, 24] and can confer advantage via influence on others [36]. Network position can therefore be viewed as a form of *social capital* [12, 15] – a function of social structure that produces advantage [23].

The dynamics of how social networks are formed (including organic growth and recommendations) can lead to skews in network position based on demographics, gender, or other attributes. Experiments show that introducing even slight demographic bias to network formation processes can exacerbate differences in network structure between groups [64]. This becomes even more problematic when seen in light of boyd, Levy, and Marwick's argument [10] that position in the network is itself a *feature* that can lead to discrimination separately from individual demographic attributes,

and modern social networks might be vehicles for a more direct propagation of (dis)advantage. Social networks' topology can cause better-positioned users to benefit more from the privileges of their position, leading to even better connections. On the other hand, less well-connected individuals – because of demographics, class, wealth, or other factors that drive network position – will find it much harder to improve their network status. As a result, the gap in power between the most and least advantaged users can lead to a cascading cycle where those with more capital have better opportunities for additional improvement, creating increased inequality.

In order to mitigate the differential accumulation of social capital, one could consider intervening in the network to change the spread of information. However, in order to do this in an automated fashion, we need ways to measure social capital based on network position. Fish et al. [31] first introduced the notion of *information access* as a resource and used it to propose a formal description for an individual's access to information. Beilinson et al. [7] subsequently defined an *access signature* to encode the "view" from a node of its access to information sent from other nodes in the network. We build on these approaches to model structural access advantage and formulate appropriate metrics for its evaluation. We design intervention strategies that use these metrics to achieve our main goal of ensuring equitable information access.

Our setup differs from prior work in a significant way. In influence maximization, a single piece of information is being spread in the network, and one can improve access for disadvantaged nodes by augmenting the set of initial sources. In contrast, we consider a setting such as those which occur on LinkedIn, where each node is the source of a unique piece of information, and access to all pieces is equally important. Given this key difference, we argue that instead of trying to select additional seeds for some or all of the pieces to improve dispersal, the natural intervention is adding edges to the network, representing the idea of purposefully strengthening weak ties [36] to mitigate bias in the structure and increase connectivity.

In this work, we have three primary contributions:

(1) Using a normative framework and drawing on prior work, we formulate three measures – *broadcast*, *influence*, and *control* – to model structural advantage with respect to access.

(2) We focus on intervening in the network using budgeted edge augmentation to improve the structural position of least-advantaged nodes, reduce the advantage gap, and ensure that nodes have similar "views" of the network (as measured via their access signature). At the core of our approach is the idea that to mitigate inequality, we should maximize the minimum access of the least-advantaged node – which in turn reduces to maximizing the minimum access between all pairs of nodes in the network.

(3) We introduce heuristic algorithms for selecting edge augmentations and empirically evaluate them on a corpus of social network data. We further show experimentally that while this process directly maximizes the *broadcast* measure of access advantage, it also simultaneously improves *influence* and *control* disparities among nodes, as well as making node access signatures more uniform.

2 RELATED WORK AND PRELIMINARIES

Motivated by the design of viral marketing strategies, Domingos and Richardson [28] introduced an algorithmic problem for social networks in which one wished to convince an initial subset of individuals to adopt a new product or innovation in order to maximize the cascade of further adoptions. This model can be generalized to many types of information spread beyond adoption and was formalized as the discrete optimization problem of *influence maximization* by Kempe et al. [44], leading to an extensive literature on the subject (see the survey [51]), including many applications in public health awareness [71, 76–78].

Structural Advantage. Information propagation in networks has been studied for decades in social and computing sciences [13, 24], and network position is known to dramatically impact a node's access to other network members [36]. It has been repeatedly argued that one's position in a network is itself a form of wealth or social capital [12, 15, 23, 37], enabling better and faster access to circulating information and important individuals. This translates into better access to opportunities (such as jobs) and enables well-positioned people to be more effective brokers, make better decisions, and innovate more efficiently [12]. Further, in public health scenarios, people rarely act on mass-media information unless it is also transmitted through personal ties [43, 60], leading to well-connected nodes having improved outcomes in crises.

Bias in Network Structure. The network itself can act as a transmitter for bias when the structural advantages described above interact with network formation mechanisms that encourage homophily and clustering of demographic groups. Schelling demonstrated how local neighborhood-based decisions could lead to segregation [62], and recent work has explored how bias in localized decisions about new connections can result in networks that have significant skew [42, 48]. Sociologists have extensively studied the role of social status in shaping network structure, showing in small-scale experiments that it significantly influences whether individuals end up in central vs. peripheral network positions [18, 52].More recently, studies in network science have extended these ideas to large-scale networks by developing computational methods for characterizing the structural influence of social status at scale [4, 49]. For example, Clauset et al. quantify the ways in which institutional reputation (and the auxiliary features of demographics and productivity) shapes the structure of faculty hiring networks among academic departments [21, 75] and subsequently the differential spread of ideas [56].

Algorithmic Fairness in Information Propagation. In the setting of information access, natural questions of fairness arise in the problem of ensuring similar allocation among demographic groups, which are often represented as disjoint subsets of nodes. Inspired by the literature on social position initiated by Granovetter's strength of weak ties [36] and framed in the context of online social networks by boyd, Levy, and Marwick [10], there has been a rash of recent work on computational questions around fairness in access on social networks [1, 6, 29–31, 38, 39, 45, 53, 57, 63, 69, 74]. The key underlying idea is that *information access is a resource*, and Fish et al. [31] argued that access based on network position is a form of privilege, which they used to define a notion of individual fairness.

Much of the work on defining and applying fairness has been undertaken in the influence maximization framework. One important thrust has been improving equity among demographic groups

within a network, typically defined based on protected classes (e.g., race, gender) [1, 39, 57, 63, 69]. They develop metrics and algorithms to ensure that roughly equal amounts of information reach each demographic group while optimizing influence maximization. In all cases, a single piece of information is being spread in the network. Intervention takes the form of augmenting the seed set in most work, though [39, 66] consider adding edges instead of seeds. In other recent work on edge augmentation to maximize the influence of a given group [6, 26], advantage is inherently defined to be access to (a small) seed set.

Several other recent papers in the space consider variants of the basic access problem. Becker et al. [5] consider μ sources of diverse information in a network and maximize the expected number of nodes receiving at least v types of information. In our setting $\mu = n$ (we consider each user as a potential source of information), and our objective function is different. Specifically, Becker et al. [5] are not focused on fairness, but rather maximizing the number of users with diverse information. In contrast, we want to improve the access of the most disadvantaged node. Ramachandran et al.[58] consider another related but distinct problem, using a diffusion model of mobility dynamics and try to achieve equity in group-level access in the facility location problem.

Graph Neural Networks (GNN). There has been recent work on analyzing the information access problem in theoretical GNNs [2, 3, 35, 68], and in solving it using link recommendation [3, 8, 41, 68]. Several of these papers suggest using random-walk spectral metrics to characterize information flow [3, 35, 72].

2.1 Preliminaries

As in the discrete optimization setting of [44], we use a stochastic *information flow model* describing how information might transmit from one node to another along the edges of G (for example, Independent Cascade, Linear Threshold, or an infection flow model from epidemiology [44]). These models all work by assuming that at time zero, an initial *seed set* of nodes that possess the information to be spread. For each node v_j as the seed set, there is then a (potentially hard to compute) probability p_{ij} – which we call *access proximity* – that node $v_i \in V$ possesses v_j's information once the spread process has terminated. Similarly, their *access distance* can be defined as $1 - p_{ij}$. Inversely, p_{ji} is called the *reach* of v_i with respect to v_j. We restrict our attention to the undirected setting (where social network links require mutual consent and typically create a giant connected component – Facebook's has 99.9% of users [70]), $p_{ij} = p_{ji}$ and we use access proximity and reach interchangeably.

Independent Cascade Model. In this work, we utilize the standard probabilistic model of influence propagation, Independent Cascade (IC) [44] with a uniform transmission probability α. In this model, a node exists in one of three states: *ready to receive*, *ready to transmit*, or *dormant*. Initially (at time zero), all nodes are ready to receive information, while the *seed* nodes also possess the information and are ready to transmit. At each time step, a node that is ready to transmit sends its information along each incident edge independently with probability α. All such transmissions are imagined to happen simultaneously, after which the transmitting node goes dormant. Computing the access proximities for Independent Cascade is #P-hard [20], so we use standard *Monte Carlo* simulations to estimate them when needed.

Access signatures. Since we view a piece of information as being uniquely identified by its originator, describing the access of a node requires a vector of $n - 1$ probabilities, which is standardized to length n to facilitate easy indexing and comparison across nodes, and $p_{ii} := 1$. These vectors are called *information access signatures*, and were introduced by Beilinson et al. [7], who argued that nodes that have similar "status" based on network position receive similar information. The signature encodes the "view" from a node of its access to information sent from the other nodes in G; people who are likely to receive information from the same part(s) of the network will have similar signatures.

Definition 1 (Access Signature [7]). The *access signature* $a_\alpha^G \colon V \to \mathbb{R}^n$ of a node $v_i \in V$ in graph G on n nodes is:

$$s_\alpha^G(i) = (p_{i1}, ..., p_{ij}, ...p_{in})$$

3 STRUCTURAL ADVANTAGE

How does network position impact access and influence? In social networks, structural advantage can manifest in many ways. Inspired by prior work, we formalize three distinct notions of advantage arising from network position and propose measures for quantifying each.

3.1 Access-based Definitions

We begin by defining analogues of graph-theoretic proximity, diameter, and betweenness centrality, highlighting when the access-based variants diverge from their traditional counterparts.

Access Proximity. In graph theory, the *proximity* of nodes v_i and v_j is the number of edges in a shortest v_iv_j-path. To adapt this to an information flow setting, we let the *access proximity* be

$$\text{dist}^*(v_i, v_j) = p_{ij},$$

the probability that v_i receives v_j's information after the completion of Independent Cascade. We observe that these measures can diverge in even simple networks. Consider two nodes connected by an edge; they have distance 1 and access proximity α. If instead, these nodes were connected by t disjoint paths of length 2 they would have distance 2, but access proximity $1 - (1 - \alpha^2)^t$. Assume $t > \frac{\log(1-\alpha)}{\log(1-\alpha^2)}$. While the nodes are graph-theoretically closer in the first scenario, in the information access setting they are closer in the second.

Access Diameter. For large networks, we often rely on summary statistics as indicators of network structure. One such metric is the *diameter*, defined to be the maximum distance between any two nodes (equivalently, the length of a longest shortest path). The analogous notion in the information access setting is then then the smallest access proximity between two nodes (equivalently, the lowest probability of pairwise information transmission). We call this the *access diameter*:

$$\text{diam}^*_G = \min_{v_i, v_j \in V} p_{ij}.$$

Access Centrality. Finally, since we are interested in assessing influence or control with respect to information flow, we consider the *betweenness centrality*, which measures how often a node appears on the shortest paths between others. Specifically, if we let σ_{jk} be the number of shortest $v_j v_k$-paths, and $\sigma_{jk}(i)$ the number of shortest $v_j v_k$-paths passing through vertex v_i, we can define the betweenness centrality of v_i as

$$g(v_i) = \sum_{v_j \neq v_k \neq v_i \in V} \frac{\sigma_{jk}(i)}{\sigma_{jk}}.$$

One can think of this as measuring the brokerage ability of a node in a world where information flows along the shortest paths. To adapt to the Independent Cascade model, we want to measure the fraction of other nodes' pairwise access that depends on v_i. In other words, the *access centrality* v_i is

$$\text{cent}^*(v_i) = \sum_{v_i \neq v_j \neq v_k \in V} \frac{p_{jk}(i)}{p_{jk}},$$

where $p_{jk}(i) = p_{jk} - p'_{jk}$ can be computed using the access proximity p' in $G' = G \setminus v_i$. We note this is computationally expensive, as you must re-estimate access proximity in $G \setminus v$ for each vertex v.

To see where these two notions diverge, consider nodes a, b connected with a path of length two through node c. The betweenness and access centrality of c are both 1. Now augment this graph by adding t disjoint ab-paths of length 3; the betweenness centrality of c remains 1, but the access centrality tends to 0 as t increases, as the fraction of information passing through c becomes insignificant.

3.2 Measures of Advantage

We now formalize three different notions of structural advantage, arising from various perspectives on fairness and information flow.

3.2.1 Broadcast Advantage: From a fairness point of view, Fish et al.[31] argued that the performance of a source should be measured by how effectively it reaches least-advantaged nodes. In this vein, we propose our first advantage function, *broadcast*, to measure how difficult it is for a node to disseminate its information to *all* others in the network.

Definition 2 (Broadcast Advantage). The *broadcast advantage* of a node is the worst-case probability that its information is received – equivalently, the minimum entry in its access signature:

$$\text{broadcast}(v_i) = \min_{p_{ij} \in s(i)} p_{ij}.$$

In some sense, this represents how "loud" the node is – a larger broadcast means a better probability that everyone else in the network will receive your information. Consider the case of recruiters using a network like LinkedIn, wanting to spread information about a job opportunity. In order to ensure a diverse candidate pool and broad reach, the employer wants a high probability the ad will reach all suitable nodes in the network. Since well-connected users receive many such ads, the measure of recruiting effectiveness will depend on how well they can disseminate the information to the least-advantaged members of the network. Better-positioned recruiters will have higher broadcast.

Further, social media is often used in public health *epidemiologic monitoring and surveillance* for early detection of disease outbreaks.

Staff responsible for dispelling misinformation and identifying high-risk or affected groups need access not only to the majority of people, but especially to those who are poorly-connected (and thus at risk of being neglected in treatment [33, 61]), motivating us to improve their broadcast.

From another perspective, the broadcast is a lower bound on the probability that v_i will get information from v_j, *regardless of which v_j is selected*! Increasing broadcast(v_i) necessarily improves information flow to/from the parts of the network that are currently least accessible from v_i, increasing the novelty and diversity of its information. Novel information often represents a resource or opportunity due to local scarcity, and users with access to it enjoy social and economic advantages, including more success in wages, promotion, job placement, and creativity [12, 36].

3.2.2 Influence Advantage: Network prominence has been studied as a type of advantage [11, 46]. A central or well-connected node is more likely to have high visibility, which Jackson's *friendship paradox* argues can lead to over-representation and increased influence [37]. This type of advantage does not require the ability to reach all nodes in the network, just many of them.

Being able to disseminate information to a large set of other members enables a user to build their social reputation, express and diffuse their opinion, and discover novel content and information [26], which can be viewed as *media power* or celebrity capital. This may also lead to opportunities for revenue from advertisement [17]. Consider the example of collaborations in a scientific community. If someone can reach more people to share her research, she gets more recognition, and feedback which enables improvement, collaboration opportunities, and directions or ideas for future work [25, 65]. We propose *influence* advantage as a measure of this form of structural advantage, drawing on influence maximization [44] in choosing a quantification.

Definition 3 (Influence Advantage). The *influence advantage* of a node is the average probability that its information is received – equivalently, the mean of the entries in its access signature:

$$\text{influence}(v_i) = \frac{1}{n} \sum_{p_{ij} \in s(i)} p_{ij} = \frac{1}{n} \sum_{v_j \in V} \text{dist}^*(v_i, v_j)$$

3.2.3 Control Advantage: Burt [12] introduced the idea of brokerage advantage. Individuals in networks with many "structural holes" may derive information and control benefits from the lack of external connectivity among people they can reach. Burt introduced this form of social capital as an information benefit or vision advantage that improves performance by providing early access to diverse and novel perspectives, ideas, and information. Hence, a person's reach is a form of power as it enables her to broker favors and consolidate strength by being uniquely positioned to coordinate the actions of others. We call this type of structural advantage *control*.

While Burt proposed several ways to measure structural holes, including bridge count [16], and network constraint/redundancy [12], in more recent work Jackson [37] used betweenness centrality [32] to measure brokerage advantage. This generic measure of importance in a network captures a node's ability to act as an intermediary to coordinate others, where nodes rely on it in order to reach other users along shortest paths. Higher centrality corresponds to more

control over information flow in the network. In turn, we use access centrality to measure the control advantage.

Definition 4 (Control Advantage). The *control advantage* of a node is given by its access centrality:

$$\text{control}(v_i) = \text{cent}^*(v_i) = \sum_{v_i \neq v_j \neq v_k \in V} \frac{p_{jk}(i)}{p_{jk}}$$

We observe that control can be rewritten as a nested sum over nodes, revealing a useful finer-grained notion of advantage. For example, suppose the node v_i has one neighbor v_j, which is a leaf, and another neighbor which is a member of a large clique. Clearly, v_i has a large degree of control over v_j, as it is an intermediary to all access to the clique, yet $\text{control}(v_i)$ might remain small, as v_i plays little role in access between clique nodes. We use control_j^i to denote the brokerage v_i has over information reaching node v_j, where

$$\text{control}_i^j = \sum_{v_j \in V} \frac{p_{ij}(c)}{p_{ij}}.$$

Our measure can then be written as $\text{control}(v_i) = \sum_j \text{control}_j^i$.

When trying to mitigate inequity in access, we would like to see the control values decrease for better-positioned nodes. Additionally, we argue that in an ideal network, no node has a monopoly over others' access to information, and we would like to prevent situations where control_j^i is close to 1 for any pair (i, j).

4 EDGE INTERVENTION & WELFARE

In contrast to the standard framework of influence maximization, we argue that when considering information flow in a network, it is important to have access to information from all individuals, not just a seed set. Further, given this shift in objective, adjustments to the model of intervention are warranted, and we propose edge augmentation as the natural candidate. We support our argument from three perspectives: *variety*, *structure*, and *voice*.

Variety. Since ideas travel a variety of paths from many sources [34, 54, 73], access to more diverse information and a greater number of individuals is important [36] and can provide a vision advantage that translates into social capital [12]. Key functionalities of social networks like LinkedIn rely on the fact that important information is frequently being disseminated from a multitude of constantly-changing sources. Traditional influence maximization is insufficient for assessing access and proposing equity-improving interventions in this setting, as we no longer know the seed set, nor can we afford to try and augment sources for each new announcement.

Structure. Granovetter introduced the idea of network manipulation to achieve specific goals [36]. Since network position is a critical form of social capital in information access, and positional disparities arise from biases in the network structure, we argue that interventions which change the underlying connectivity of the network are necessary. The natural candidate is to increase access through edge augmentation. This approach is further supported when one thinks of these edges as representing the addition of weak ties to the transmission network, as research shows that information can traverse greater social distance and reach more people when diffused along weak ties instead of strong ones [36].

Voice. While it is easy to focus on improving access for poorly-positioned nodes, it is also important to consider the effect of interventions on already-advantaged users. Specifically, node interventions increase the reach (and thus influence) of selected individuals [37], essentially amplifying their information within the network. To give voice to all participants, we argue that edge augmentation improves fairness by increasing the reach of all nodes.

Now that we have argued for using edge augmentation to intervene in the network, we turn to the question of which structural measure of advantage to optimize. We use a normative framework to select one of broadcast, influence, and control, and draw on the Rawlsian Maximin argument [59] in proposing that we should maximize the advantage of the least advantaged node(s). Rawlsian principle has been used in other algorithmic fairness methods [19, 31, 40].

To choose a notion of advantage, we begin by observing that optimizing influence encourages the formation of edges to well-positioned nodes. Therefore, nodes with better connections become more attractive to connect to [37], leading to a rich-get-richer phenomenon and potentially increasing the advantage gap instead of equalizing access [14]. These peripheral-central connections also increase the control of central nodes over others, especially the disadvantaged. On the other hand, using broadcast as the objective prioritizes connectivity for the most disadvantaged nodes. As John Stuarts Mills noted, "it is hardly possible to overrate the value . . . of placing human beings in contact with persons dissimilar to themselves and with modes of thought and action unlike those with which are familiar . . . Such communication has always been and is peculiarly in the present age, one of the primary sources of progress" [55]. Optimizing for control, on the other hand, prioritizes the brokerage ability of nodes over their access to diverse information, which could lead to polarization and centralized information distribution. We argue that increasing broadcast, which tends to also reduce the control of other nodes, is preferable since depending on powerful information-brokers reduces one's chance of unbiased access to diverse opinions.

Several other normative reasons underlie our preference for broadcast to measure structural advantage, when one considers outcomes in a network containing several (mostly-disjoint) minority groups. First, while these groups may have common interests, they will not individually have enough influence to accomplish them. Connecting disadvantaged nodes directly (instead of through a central node) will enable them to support one another and access important information, while countering the ever-increasing power of the majority. In support of this argument, we note that Kogan et al. show that geographically vulnerable (disadvantaged) users propagate more information during disasters, and are more likely to propagate tweets from other geographically vulnerable users [47]. A final argument arises from work on mitigating polarization in social networks by increasing the similarity of users' exposure to a broad diversity of news and ideas. Since minimizing diameter can speed up communication [27] and increase the uniformity of exposure times, we argue that optimizing broadcast is the natural analogue in the information access setting.

To formalize a discrete optimization problem, we must now transform our advantage measure into an objective function. Following

the Rawlsian *Maximin Principle* that one should maximize the welfare of the worst-off person [59], we seek to maximize broadcast for the least-advantaged nodes, and formalize this as the *welfare*.

Definition 5. The *welfare* of a graph $G = (V, E)$ is

$$\mu(G) = \min_{v_i \in V} \text{broadcast}(v_i) = \min_{v_i, v_j \in V} \text{dist}^*(v_i, v_j)$$

Our central problem is to find a *budgeted intervention* optimizing welfare.

MAXWELFARE-AUGMENTATION

Input: A graph $G = (V, E)$ and an integer $k \in \mathbb{N}$.
Problem: Find a set $E^+ \subseteq V \times V$ of size at most k so that $\mu(H)$ is maximized, where $H := (V, E \cup E^+)$.

---

5 HEURISTICS

In this section, we introduce several heuristics for MAXWELFARE-AUGMENTATION which greedily select new edges using advantage-based criteria. We employ two basic strategies – connecting disadvantaged nodes to a *central* one, and adding a *chord* between two peripheral nodes. We will compare these with a baseline (rand) which chooses both ends of each new edge uniformly at random. It is also natural to try a greedy approach baseline, by adding the edge that maximizes the objective function in each iteration. The idea has been used in different settings [26, 31]. This greedy approach in each round, for each non-existing edge computes the broadcast and picks the edge that maximize the broadcast. Since the greedy baseline is computationally very expensive, we consider a naïve variation of the greedy baseline (greedy-bc) which, instead of proceeding in rounds, in the first round picks the top k edges that maximize the broadcast.

We begin by defining the *center* of the network to be the node with maximum broadcast. In our algorithms, we select this node in the un-augmented network and fix it for the duration of the edge selection process. As we iteratively make interventions, it is possible that a new central node emerges (one with higher broadcast than the selected center). While we could update at every step, this incurs a high computational cost. In order to evaluate the likelihood and impact of a shifting center, we re-ran the experiments on the three smallest networks and recorded how often the maximum broadcast increased, along with the L_1 norm of the access signature difference between initial and new centers. The initial center node remained central more than 99% of the time, and the signature difference was less than 0.01 in the other 1% of cases. Based on this and the significant computational cost, we choose to fix a center node based on the initial network.

Before proceeding to the heuristics, we need two additional observations. First, computing the access proximities is known to be #P–hard [20]; as such, whenever our strategies use p_{ij}, we rely on simulation to estimate the access proximities using Reverse Influence Sampling (RIS) [9, 67]. Second, greedy heuristics may select a pair of vertices to connect which already have an edge in the graph. When this happens, we select an alternative augmentation in one of two ways: (1) if the heuristic was trying to connect a node u to the *center*, we instead connect u to the node with second-highest broadcast, continuing down the broadcast order as needed until

we find a non-neighbor of u; (2) if the heuristic was adding a *chord* or random edge, we "randomly replace an endpoint." We can now define our strategies for reducing the access diameter of a network.

Broadcast-based Strategies. To reduce the access diameter of the network we must affect at least one node with minimum broadcast. If v_i, v_j is a pair of nodes so that p_{ij} is minimum, we call them *diameter-defining*. Our first heuristic bc-chord finds a diameter-defining pair and adds the edge between them. A natural alternative strategy is to connect one or both of the pair to the *center*; we do this in bc-both and bc-one, respectively. Note that bc-both adds pairs of edges, and runs for only $\frac{k}{2}$ steps; we constrain k to even values in experiments to ensure fair comparisons.

Influence-based Strategies. Another reasonable approach to improving access in the network is to equalize influence. Similar to broadcast, we connect the node with minimum influence to the center, and call this heuristic infl.

Diameter-based Strategies. Finally, we consider a measure that can be computed without simulation, the *diameter* of the underlying network. While the shortest-path distances and access proximities may diverge, they are not independent, and creating short paths between nodes will improve their pairwise access. Similar to bc-chord, diam-chord adds an edge between a pair of nodes with maximum $d(u, v)$;

6 EXPERIMENTS

We implemented the heuristics from Section 5 in C++ and compiled with gcc 8.1.0; all experiments were run on identical hardware equipped with 40 CPUs (Intel Xeon Gold 6230 @ 2.10GHz) and 190 GB of memory, running CentOS Linux release 7.9.2009. To evaluate the effectiveness of our intervention strategies, we used a corpus of real-world networks sourced from the SNAP [50] and ICON [22] repositories, as described in Table 1. We treated all data as undirected, and used the largest connected component for each.

As briefly mentioned in Section 5, we use Reverse Influence Sampling (RIS) [9] to estimate access proximities; we generate $R = 10,000$ instances per simulation. To evaluate the accuracy, we ran each estimation 10 times and measured the fluctuations in access proximities. In all cases, pairwise accesses varied by less than 0.03 (3% of the range), and the average difference was at most 0.004 (0.4% of the range). The heuristics bc-chord, bc-both, bc-one, and infl use RIS, requiring $O(Rm + Rnk)$ time and $O(n^2 + Rn)$ space. The greedy-bc heuristic, adds $O(m)$ to the complexity for estimating the access proximities for each missing edge, and drops $O(k)$ since it is done in one round; which makes it run in $O(Rn^3)$ time and $O(n^2 + Rn)$ space. Since even this variation of greedy-bc is still computationally expensive, we only run it on the three smaller datasets.

In each experiment, we used even values of k from 0 to 200, aiming for a practical intervention size relative to the network (less than a tenth of a percent of $|E|$). In the Independent Cascade model, the spread of information depends on the input parameter α (the probability of transmission along an edge in a time step). For each network in our corpus, we computed the distribution of access proximities for varied α and selected four (network-specific) values: one each to represent poorly-spreading and well-spreading
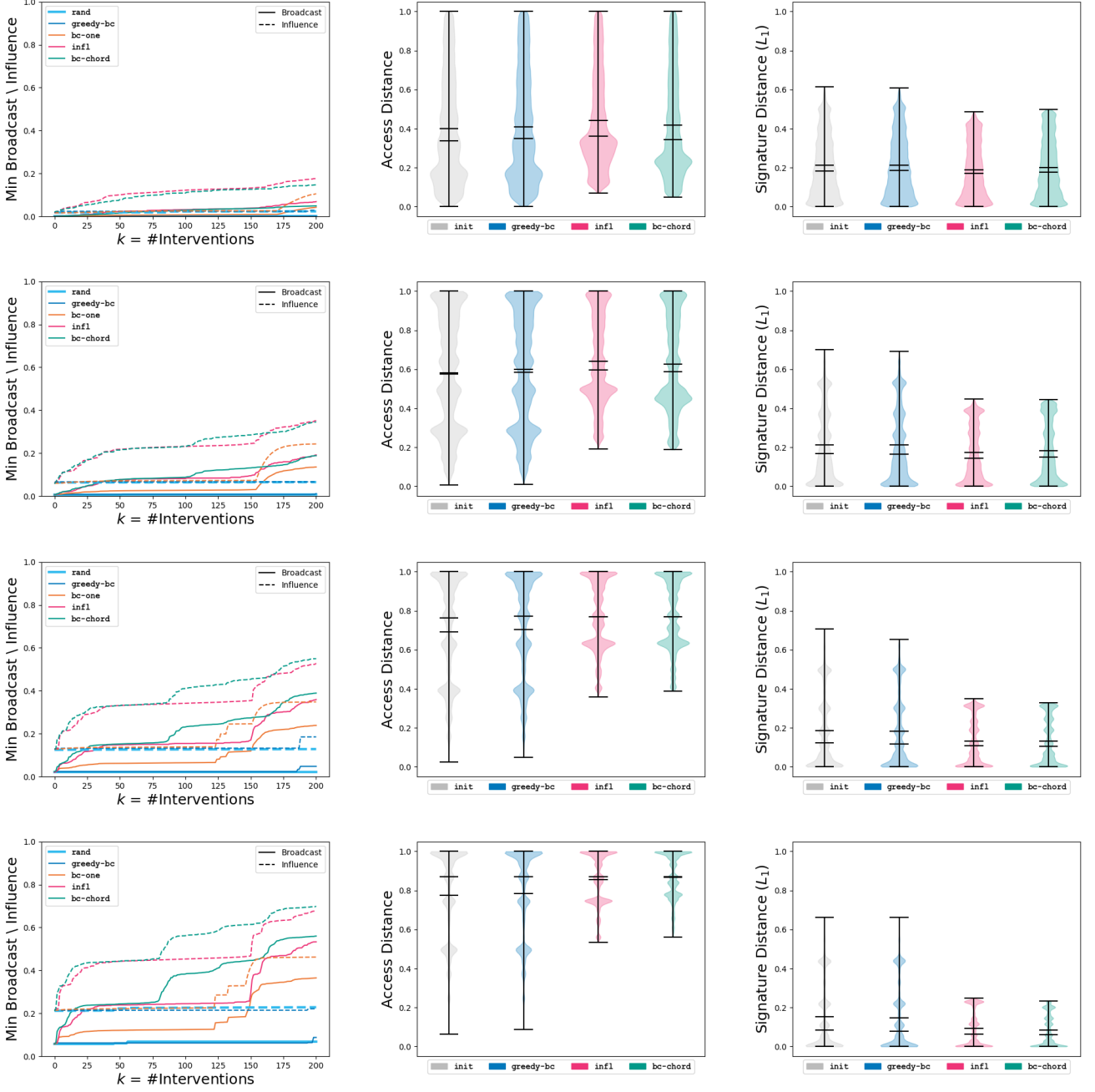
Figure 1: Results for Email-Arenas with $\alpha = \{0.2, 0.3, 0.4, 0.5\}$ (top to bottom). We plot improvement in minimum broadcast and influence (left); the violin plots show the distribution of pairwise access proximities (middle) and L_1 signature distances (right).

scenarios, and two in the critical region of moderate spread (for lower values of α, it is not possible to significantly increase the broadcast within the limited budget of interventions).

6.1 Summary of Experimental Results

The primary objective of this work is to intervene in a network to improve access for the most disadvantaged nodes and reduce disparities in advantage by making access signatures more similar. To assess whether our strategies achieve these goals, we employ several methods for evaluating the outcome of interventions. First,

we directly measure the improvement in the minimum values of broadcast and influence realized in the network. Next, we shift our attention to the access signatures, where we evaluate whether our interventions have increased the similarity among nodes' views of the network using Manhattan distance. Finally, we consider whether our approaches improve disparity by reducing the advantage gap between the most- and least-privileged nodes.

Table 1: Summary of Datasets

Name	Nodes	Edges	Max Deg.	Diam.
Email-EU	803	24729	338	5
Email-Arenas	1133	5451	71	8
Irvine	1294	19026	231	7
Facebook	4039	88234	1045	8
ca-GrQc	4158	13428	81	17
ca-HepTh	8638	24827	65	18

In Figure 1, we present a comprehensive view of all three evaluations for a single network across its four transmission probabilities. From the first column, we observe that the heuristics `bc-chord`, `bc-both` and `infl` are most effective at improving broadcast and influence, with the latter two performing almost identically. Further, `bc-chord` surpasses the other approaches as information spreads more easily. These results are qualitatively replicated by the other networks in our corpus (see Appendix A). Given this, we restrict our attention to the `bc-chord` and `infl` approaches in subsequent figures, with `infl` favored over `bc-both` to increase the diversity among our strategies. Further, we note that the behavior with respect to α remained consistent across all networks, and is well-represented by considering only the low-moderate-spread and well-spreading values of α (2nd and 4th columns). Due to space constraints, plots for the entire corpus (Figures 2, 3 – 7), only show these two transmission probabilities. In the second column of Figure 1, we use violin plots to show the distribution of access proximities for all pairs before (`init`) and after (`greedy-bc`, `infl`, `bc-chord`) intervention. We observe that while `greedy-bc` augmentation has little effect, both heuristics significantly reduce the maximum pairwise access proximity, with `bc-chord` again out-performing `infl` as α increases. While the distributions for other networks vary in initial shape, the pattern of improvement was consistent (see Section 6.3). Finally, the third column of Figure 1 illustrates our success in increasing the uniformity among each node's view of the network as measured by reducing the maximum distance between access signatures. Results for other networks are summarized in Section 6.3.

To round out our evaluation, we also computed how our interventions affected the advantage gaps for broadcast, influence, and control, as discussed in Section 6.4. For the network featured in Figure 1, these results are in the second row of Tables 2 and 3. One surprising result was that while the absolute broadcast gap increased, the relative one decreased. We believe this is caused by interventions increasing access by a larger additive amount for central nodes than peripheral ones. Over the entire corpus, `bc-chord` shrank the broadcast / influence gaps by over 85% / 82%, respectively.

Overall, we observe that our interventions are most effective when the network is better-connected – whether because α is

higher, or the underlying graph is denser (e.g. in EU and Fb). Additionally, our analysis showed that `bc-both` and `infl` perform almost identically (Figures 2 and 3), suggesting that the nodes with minimum broadcast and influence may have similar access signatures. To further investigate this phenomenon, we measured the signature difference between the nodes selected by each of these heuristics at each intervention step and found them to be consistently in the bottom 10% of all pairs, with the average falling in the bottom 1%. This leads us to hypothesize that the set of least-advantaged nodes with respect to broadcast and influence are almost identical.

6.2 Improving Minimum Broadcast / Influence

The broadcast and influence measures quantify a node's structural advantage as a function of its signature. Here we evaluate whether edge interventions can improve these measures for the most disadvantaged nodes in the network. Figures 2 and 3 plot the trajectory of the minimum broadcast and influence as the number of interventions k increases with low-moderate- and well-spreading α for each network in the corpus. We observe that `infl` and `bc-both` consistently show the most improvement for both advantage measures. Further, `rand` and `greedy-bc` both have similarly poor performance. Most social networks contain multiple distance minority groups (consider the friendship network in a high school). Since `greedy-bc` does not update the access proximities, it may get stuck using all its intervention on one disadvantaged group, while ignoring the rest.

6.3 Making Nodes & Signatures Closer

One goal of intervention is to increase access for nodes that have the lowest probability of receiving some types of information. In Figures 4 and 5 (and the second column of Figure 1 for Email-Arenas), we plot the distribution of pairwise access proximities before and after intervention; we again consider two transmission probabilities (low-moderate-spread and well-spreading) for each of the 6 networks in the corpus. We observe that while the median value does not move significantly, the lower tail of the distribution gets much shorter and thinner. The amount of improvement increases with α, and is more pronounced in the denser networks (EU, Irvine, and Fb). In some cases, with only 200 interventions, we are able to increase the minimum pairwise access proximity by 0.7, more doubling the probability of information transmission!

Another of our objectives is increasing similarity among access signatures so that all nodes have a similar "view" of the network. We use the Manhattan distance (L_1 metric) to measure the distance between two signatures[1]. Figures 6 and 7 (and the third column of Figure 1 for Email-Arenas), show violin plots of the distribution of these distances. The maximum signature difference was consistently reduced (at least 43% for well-spreading α), and while the median was relatively stable, the tail of the distributions shifted noticeably downward.

6.4 Measuring the Gap

The final central premise of this work is that improving equity requires reducing access disparities between nodes. To evaluate

---

[1]using Euclidean distance (L_2) results in similar trends and no qualitative differences

Table 2: Absolute/Relative Advantage Gaps

(Network, α)	Gap	Heuristic			
		init	rand	infl	bc-chord
(Email-EU, 0.3)	bc	0.21/2.49	0.21/2.48	0.14/0.21	**0.08/0.10**
	infl	0.67/2.40	0.68/2.40	0.17/0.20	**0.09/0.10**
(Email-Arenas, 0.4)	bc	**0.13**/5.30	0.13/5.95	0.24/0.68	0.24/**0.61**
	infl	0.71/5.62	0.72/5.64	0.35/0.66	**0.33/0.60**
(Irvine, 0.4)	bc	**0.06** /4.83	0.08 /5.21	0.23 /0.56	0.17 /**0.29**
	infl	0.85 /13.8	0.86 /14.4	0.33 /0.54	**0.21** /**0.29**
(Facebook, 0.3)	bc	**0.14** /5.97	0.20 /2.63	0.26 /0.97	0.25 /**0.79**
	infl	0.79 /4.55	0.68 /2.61	0.47 /1.01	**0.42** /**0.80**
(ca-GrQc, 0.6)	bc	**0.07** /11.6	0.07/7.63	0.25/0.93	0.25/**0.76**
	infl	0.76/12.4	0.78/12.4	0.41/0.90	**0.36/0.74**
(ca-HepTh, 0.6)	bc	**0.09**/9.56	0.10/8.33	0.25/1.32	0.25/**1.03**
	infl	0.75/9.11	0.75/8.14	0.48/1.27	**0.43/1.01**

this, we measure the advantage gap for broadcast and influence, as well as the maximum amount of control achieved in the network (which can be viewed as a gap, since there are always nodes on the periphery with control value essentially zero).

Broadcast/Influence Gaps. We begin by calculating both the absolute (max − min) and relative ($\frac{\max - \min}{\min}$) advantage gaps for broadcast and influence on networks in the corpus; Table 2 shows these when α is well-spreading. As mentioned in Section 6.1, the absolute broadcast gap often increases with intervention, while the influence gap is typically reduced. However, the relative advantage gap behaves quite differently, consistently decreasing significantly with bc-chord, yet increasing in most cases for infl. This supports our argument that infl may contribute to a rich-get-richer phenomenon by increasing advantage for central nodes, and is an important distinction between the two heuristics.

Reducing Control. Finally, we consider how our interventions affect control. In Table 3, we report the maximum values of not only the primary control measure of cent* but also the finer-grained pairwise control (control$_i^c$). Here, we must restrict our analysis to the three smallest networks in our corpus due to the exceptionally high cost of computing control for all nodes (which requires removing each node from the network and re-estimating access proximities); we use the same well-spreading α values as in our gap analysis. The results are encouraging, as they show that intervention can increase the independence of nodes in the network when accessing information and prevent better-positioned nodes from having a monopoly over others. It is noteworthy that bc-chord not only uniformly achieves more than 53% reduction in pairwise control, it never increases the control (whereas infl can cause a 10-fold jump).

Table 3: Maximum Control Values

(Network, α)	Measure	Heuristic			
		init	rand	infl	bc-chord
(Email-EU, 0.3)	cent*	0.009	0.007	0.014	**0.002**
	control$_i^c$	1.000	1.000	0.107	**0.056**
(Email-Arenas, 0.4)	cent*	0.008	**0.006**	0.112	0.008
	control$_i^c$	1.000	1.000	0.476	**0.464**
(Irvine, 0.4)	cent*	0.008	0.007	0.050	**0.006**
	control$_i^c$	1.000	1.000	0.573	**0.217**

7 CONCLUSION

In this work, we propose a novel method for quantifying social capital through the lens of information flow in a network when all nodes have unique, equally-important information to disseminate. We introduce three new measures of structural advantage quantified in terms of network position, argue for intervening through edge augmentation to reduce bias in network structure, and formalize the budgeted intervention problem of MaxWelfare-Augmentation for mitigating structural inequity in information access. Finally, we propose heuristic strategies that improve access for the least-advantaged nodes, reduce advantage disparities, and increase the similarity in access signatures. We perform a case study on a corpus of social networks and demonstrate that our bc-chord heuristic improves the minimum broadcast *and* influence, dramatically shrinks advantage gaps, and reduces variance among access signatures. Although, chord-heuristics connections may not be practical depending on the setting and kinds of entities in the network. It is worth mentioning these links may cause unintended social consequences.

Our work is inherently limited by our use of a uniform transmission probability in the Independent Cascade model, and by ignoring the time at which information is received (as we know that early access plays an important role in social capital). Although, our measures and strategies are independent of the IC model, and can be applied to any probabilistic models of information flow, and may improve many existing diameter-based approaches. Further, the quantification of control is computationally infeasible for large networks, limiting our empirical evaluation.

We leave open many directions for future work, including the adaptation of these ideas to directed networks where access and reach may differ ($p_{ij} \neq p_{ji}$), and optimizing for one may lead to trade-offs for the other (note that most heuristics can be modified to work in the directed setting). It would also be interesting to adapt this problem to the group fairness setting by defining and optimizing advantage measures on groups. Finally, we note that our assumption is not true for misinformation, and information overflow.

REFERENCES

[1] Junaid Ali, Mahmoudreza Babaei, Abhijnan Chakraborty, Baharan Mirzasoleiman, Krishna P. Gummadi, and Adish Singla. 2019. On the Fairness of Time-Critical Influence Maximization in Social Networks.

[2] Uri Alon and Eran Yahav. 2021. On the Bottleneck of Graph Neural Networks and its Practical Implications. In *International Conference on Learning Representations.* https://openreview.net/forum?id=i80OPhOCVH2

[3] Adrian Arnaiz-Rodriguez, Ahmed Begga, Francisco Escolano, and Nuria Oliver. 2022. DiffWire: Inductive Graph Rewiring via the Lovász Bound. arXiv:2206.07369 [cs.LG]

[4] Brian Ball and M.E.J. Newman. 2013. Friendship networks and social status. *Network Science* 1, 1 (2013).

[5] Ruben Becker, Federico Corò, Gianlorenzo D'Angelo, and Hugo Gilbert. 2020. Balancing Spreads of Influence in a Social Network. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 01 (Apr. 2020), 3–10.

[6] Ruben Becker, Gianlorenzo D'Angelo, Sajjad Ghobadi, and Hugo Gilbert. 2021. Fairness in Influence Maximization through Randomization. *Proceedings of the AAAI Conference on Artificial Intelligence* 17 (May 2021), 14684–14692.

[7] Hannah C. Beilinson, Nasanbayar Ulzii-Orshikh, Ashkan Bashardoust, Sorelle A. Friedler, Carlos E. Scheidegger, and Suresh Venkatasubramanian. 2020. Clustering via Information Access in a Network. *arXiv* abs/2010.12611 (2020).

[8] Wendong Bi, Lun Du, Qiang Fu, Yanlin Wang, Shi Han, and Dongmei Zhang. 2022. Make Heterophily Graphs Better Fit GNN: A Graph Rewiring Approach. arXiv:2209.08264 [cs.LG]

[9] Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan Lucier. 2014. Maximizing Social Influence in Nearly Optimal Time. In *Proceedings of the 2014 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA).* 946–957.

[10] Danah Boyd, Karen Levy, and Alice Marwick. 2014. The networked nature of algorithmic discrimination. *Data and Discrimination: Collected Essays. Open Technology Institute* (2014).

[11] Daniel J. Brass. 1992. Power in Organizations: A Social Network Perspective. *Research in Politics and Society* 4 (1992), 295–323. Issue 1.

[12] Ronald S. Burt. 2004. Structural Holes and Good Ideas. *Amer. J. Sociology* 110, 2 (2004), 349–399.

[13] Ronald S. Burt. 1987. Social Contagion and Innovation: Cohesion versus Structural Equivalence. *Amer. J. Sociology* 92, 6 (1987), 1287–1335.

[14] Ronald S. Burt. 1999. The Social Capital of Opinion Leaders. *The Annals of the American Academy of Political and Social Science* 566 (1999), 37–54.

[15] Ronald S. Burt. 2000. The Network Structure Of Social Capital. *Research in Organizational Behavior* 22 (2000), 345–423.

[16] Ronald S. Burt, Robin M. Hogarth, and Claude Michaud. 2000. The Social Capital of French and American Managers. *Organization Science* 11 (2000), 123–147.

[17] Ronald S. Burt, Martin Kilduff, and Stefano Tasselli. 2013. Social Network Analysis: Foundations and Frontiers on Advantage. *Annual Review of Psychology* 64, 1 (2013), 527–547.

[18] Karen E. Campbell, Peter V. Marsden, and Jeanne S. Hurlbert. 1986. Social Resources and Socioeconomic Status. *Social Networks* 8, 1 (1986), 97–117.

[19] Alycia N. Carey and Xintao Wu. 2023. The statistical fairness field guide: perspectives from social and formal sciences. *AI and Ethics* 3 (2023), 1–23. Issue 1. https://doi.org/10.1007/s43681-022-00183-3

[20] Wei Chen, Chi Wang, and Yajun Wang. 2010. Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 1029–1038.

[21] Aaron Clauset, Samuel Arbesman, and Daniel B. Larremore. 2015. Systematic inequality and hierarchy in faculty hiring networks. *Science Advances* 1, 1 (2015), e1400005.

[22] Aaron Clauset, Ellen Tucker, and Matthias Sainz. 2016. The Colorado Index of Complex Networks. https://icon.colorado.edu/

[23] James S. Coleman. 1988. Social Capital in the Creation of Human Capital. *Amer. J. Sociology* 94 (1988), S95–S120.

[24] James S. Coleman, Elihu Katz, and Herbert Menzel. 1966. *Medical Innovation: A diffusion study.* Bobbs-Merril, New York.

[25] danah boyd. 2021. Knitting a Healthy Social Fabric. https://zephoria.medium.com/knitting-a-healthy-social-fabric-86105cb92c1c

[26] Gianlorenzo D'Angelo, Lorenzo Severini, and Yllka Velaj. 2019. Recommending links through influence maximization. *Theoretical Computer Science* 764 (2019), 30–41.

[27] Erik Demaine and Morteza Zadimoghaddam. 2010. Minimizing the Diameter of a Network Using Shortcut Edges. In *Algorithm Theory - SWAT 2010.* Springer Berlin Heidelberg, 420–431.

[28] Pedro Domingos and Matt Richardson. 2001. Mining the Network Value of Customers. *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining* (2001), 57–66.

[29] Yushun Dong, Ninghao Liu, Brian Jalaian, and Jundong Li. 2022. EDITS: Modeling and Mitigating Data Bias for Graph Neural Networks. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France). Association for Computing Machinery, New York, NY, USA, 1259–1269. https://doi.org/10.1145/3485447.3512173

[30] Yushun Dong, Song Wang, Yu Wang, Tyler Derr, and Jundong Li. 2022. On Structural Explanation of Bias in Graph Neural Networks. In *Proceedings of*

the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (Washington DC, USA) *(KDD '22).* Association for Computing Machinery, New York, NY, USA, 316–326. https://doi.org/10.1145/3534678.3539319

[31] Benjamin Fish, Ashkan Bashardoust, danah boyd, Sorelle Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. 2019. Gaps in Information Access in Social Networks?. In *The World Wide Web Conference.* Association for Computing Machinery, New York, NY, USA, 480–490.

[32] Linton C. Freeman. 1977. A Set of Measures of Centrality Based on Betweenness. *Sociometry* 40, 1 (1977), 35–41.

[33] I. C. Fung, Z. T. Tse, and K. W. Fu. 2015. The use of social media in public health surveillance. *Western Pac Surveill Response Journal* (6 2015).

[34] Paul A. Geroski and Mariana Mazzucato. 2002. Learning and the sources of corporate growth. *Industrial and Corporate Change* 11, 4 (2002), 623–644.

[35] Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Lio', and Michael Bronstein. 2023. On Over-Squashing in Message Passing Neural Networks: The Impact of Width, Depth, and Topology. arXiv:2302.02941 [cs.LG]

[36] Mark S Granovetter. 1973. The strength of weak ties. *The American Journal of Sociology* 78, 6 (1973), 1360–1380.

[37] Matthew Jackson. 2019. *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors.* Knopf Doubleday Publishing Group.

[38] Zeinab S. Jalali, Qilan Chen, Shwetha M. Srikanta, Weixiang Wang, Myunghwan Kim, Hema Raghavan, and Sucheta Soundarajan. 2022. Fairness of Information Flow in Social Networks. *ACM Trans. Knowl. Discov. Data* (dec 2022). https://doi.org/10.1145/3578268 Just Accepted.

[39] Zeinab S. Jalali, Weixiang Wang, Myunghwan Kim, Hema Raghavan, and Sucheta Soundarajan. 2020. On the information unfairness of social networks. In *Proceedings of the 2020 SIAM International Conference on Data Mining (SDM).* 613–521.

[40] Jian Kang, Yan Zhu, Yinglong Xia, Jiebo Luo, and Hanghang Tong. 2022. RawlsGCN: Towards Rawlsian Difference Principle on Graph Convolutional Network. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22).* Association for Computing Machinery, New York, NY, USA, 1214–1225. https://doi.org/10.1145/3485447.3512169

[41] Kedar Karhadkar, Pradeep Kr. Banerjee, and Guido Montufar. 2023. FoSR: First-order spectral rewiring for addressing oversquashing in GNNs. In *The Eleventh International Conference on Learning Representations.* https://openreview.net/forum?id=3YjQfCLdrzz

[42] Fariba Karimi, Mathieu Génois, Claudia Wagner, Philipp Singer, and Markus Strohmaier. 2018. Homophily influences ranking of minorities in social networks. *Scientific Reports* 8 (07 2018).

[43] Elihu Katz and Paul F. Lazarsfeld. 1966. *Personal Influence: the Part Played by People in the Flow of Mass Communications.* Free Press.

[44] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the Spread of Influence through a Social Network. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2003), 137–146.

[45] Moein Khajehnejad, Ahmad Asgharian Rezaei, Mahmoudreza Babaei, Jessica Hoffmann, Mahdi Jalili, and Adrian Weller. 2021. Adversarial Graph Embeddings for Fair Influence Maximization over Social Networks *(IJCAI'20).* Article 594, 7 pages.

[46] David Knoke and Ronald S Burt. 1983. Prominence. *Applied Network Analysis* (1983), 195–222.

[47] Marina Kogan, Leysia Palen, and Kenneth Mark Anderson. 2015. Think Local, Retweet Global: Retweeting by the Geographically-Vulnerable during Hurricane Sandy. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (2015).

[48] Jure Leskovec, Lars Backstrom, Ravi Kumar, and Andrew Tomkins. 2008. Microscopic evolution of social networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 462–470.

[49] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. 2010. Kronecker Graphs: An Approach to Modeling Networks. *J. Mach. Learn. Res.* 11 (2010), 985–1042.

[50] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

[51] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. 2018. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering* 30, 10 (2018), 1852–1872.

[52] Nan Lin. 1999. Social Networks and Status Attainment. *Annual Review of Sociology* 25, 1 (1999), 467–487.

[53] Anay Mehrotra, Jeff Sachs, and L. Elisa Celis. 2022. Revisiting Group Fairness Metrics: The Effect of Networks. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 375 (nov 2022), 29 pages. https://doi.org/10.1145/3555100

[54] Tanya Menon and Jeffrey Pfeffer. 2003. Valuing Internal vs. External Knowledge: Explaining the Preference for Outsiders. *Manag. Sci.* 49 (2003), 497–513.

[55] John Stuart Mills. 1848. *Principles of Political Economy.* Augustus M. Kelley, Fairchild, N.J.

[56] Allison C. Morgan, Dimitrios Economou, Samuel F. Way, and Aaron Clauset. 2018. Prestige drives epistemic inequality in the diffusion of scientific ideas. *EPJ Data Science* 7 (2018), 40.

[57] Aida Rahmattalabi, Shahin Jabbari, Himabindu Lakkaraju, Phebe Vayanos, Max Izenberg, Ryan Brown, Eric Rice, and Milind Tambe. 2021. Fair Influence Maximization: A Welfare Optimization Approach. *Proceedings of the AAAI Conference on Artificial Intelligence* (2021).

[58] Govardana Sachithanandam Ramachandran, Ivan Brugere, Lav R. Varshney, and Caiming Xiong. 2021. GAEA: Graph Augmentation for Equitable Access via Reinforcement Learning. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery, New York, NY, USA, 884–894.

[59] John Rawls. 2009. *A Theory of Justice.* Harvard University Press.

[60] Everett M. Rogers. 1962. *Diffusion of Innovations.* Free Press of Glencoe.

[61] Rasmussen SA and Goodman RA. 2019. *The CDC Field Epidemiology Manual.* New York: Oxford University Press.

[62] Thomas C. Schelling. 1971. Dynamic models of segregation. *The Journal of Mathematical Sociology* 1, 2 (1971), 143–186.

[63] Ana-Andreea Stoica and Augustin Chaintreau. 2019. Fairness in Social Influence Maximization. In *Companion Proceedings of The 2019 World Wide Web Conference*. Association for Computing Machinery, 569–574.

[64] Ana-Andreea Stoica and Christopher J. Riederer. 2018. Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity. *WWW '18: Proceedings of the 2018 World Wide Web Conference*, 923–932.

[65] Cassidy R. Sugimoto. 2021. Scientific success by numbers. https://www.nature.com/articles/d41586-021-01169-7

[66] Ian P. Swift, Sana Ebrahimi, Azade Nova, and Abolfazl Asudeh. 2022. Maximizing Fair Content Spread via Edge Suggestion in Social Networks. *Proc. VLDB Endow.* 15, 11 (sep 2022), 2692–2705. https://doi.org/10.14778/3551793.3551824

[67] Youze Tang, Xiaokui Xiao, and Yanchen Shi. 2014. Influence Maximization: Near-Optimal Time Complexity Meets Practical Efficiency. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*. Association for Computing Machinery, 75–86.

[68] Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. 2022. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations*. https://openreview.net/forum?id=7UmjRGzp-A

[69] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. 2019. Group-fairness in influence maximization. In *Proc. of the Int'l Joint Conf. on Artificial Intelligence*. AAAI Press, 5997–6005.

[70] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow. 2011. The Anatomy of the Facebook Social Graph. *arXiv* abs/1111.4503 (11 2011).

[71] Thomas W. Valente and Patchareeya Pumpuang. 2007. Identifying Opinion Leaders to Promote Behavior Change. *Health Education & Behavior* 34, 6 (2007), 881–896.

[72] Ameya Velingker, Ali Kemal Sinop, Ira Ktena, Petar Veličković, and Sreenivas Gollapudi. 2022. Affinity-Aware Graph Networks. arXiv:2206.11941 [cs.LG]

[73] Eric A. von Hippel. 1988. *The Sources of Innovation.* Oxford University Press.

[74] Xindi Wang, Onur Varol, and Tina Eliassi-Rad. 2021. Information Access Equality on Network Generative Models. *ArXiv* abs/2107.02263 (2021).

[75] Samuel F. Way, Daniel B. Larremore, and Aaron Clauset. 2016. Gender, Productivity, and Prestige in Computer Science Faculty Hiring Networks. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences, 1169–1179.

[76] Bryan Wilder, Laura Onasch-Vera, Graham Diguiseppi, Robin Petering, Chyna Hill, Amulya Yadav, Eric Rice, and Milind Tambe. 2020. Clinical trial of an AI-augmented intervention for HIV prevention in youth experiencing homelessness. *arXiv* abs/2009.09559 (2020).

[77] Bryan Wilder, Laura Onasch-Vera, Juliana Hudson, Jose Luna, Nicole Wilson, Robin Petering, Darlene Woo, Milind Tambe, and Eric Rice. 2018. End-to-End Influence Maximization in the Field. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS-18)*.

[78] Amulya Yadav, Bryan Wilder, Eric Rice, Robin Petering, Jaih Craddock, Amanda Yoshioka-Maxwell, Mary Hemler, Laura Onasch-Vera, Milind Tambe, and Darlene Woo. 2018. Bridging the gap between theory and practice in influence maximization: Raising awareness about HIV among homeless youth. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence, 5399–5403.
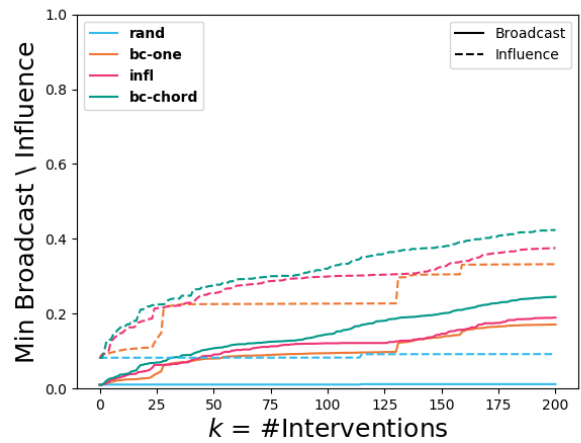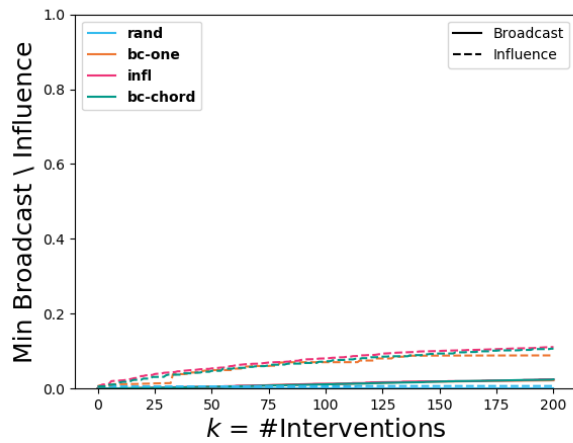
A APPENDIX
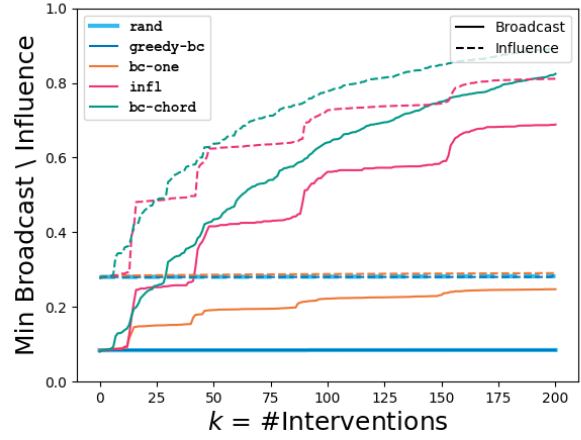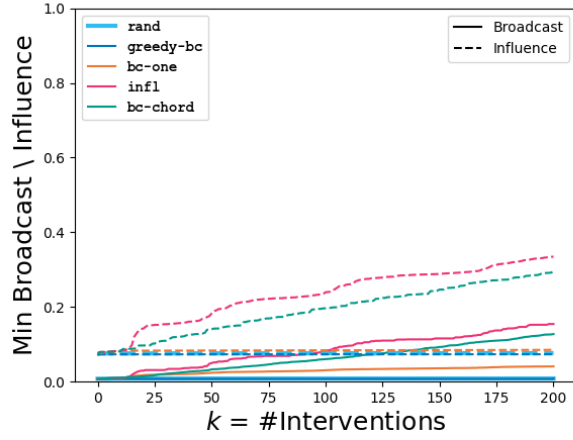
(a) **Email-EU (Left:** $\alpha = 0.1$, **Right:** $\alpha = 0.3$)
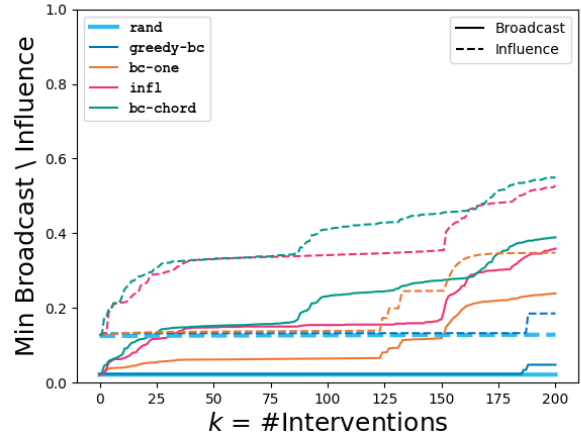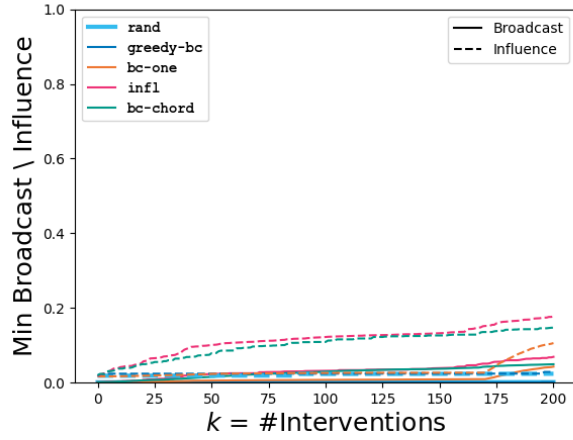
(b) **Email-Arenas (Left:** $\alpha = 0.2$, **Right:** $\alpha = 0.4$)
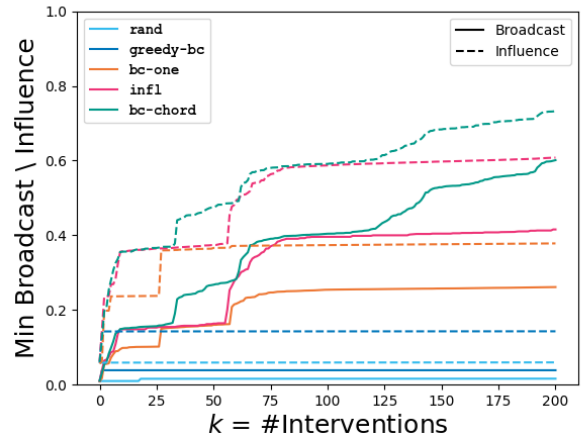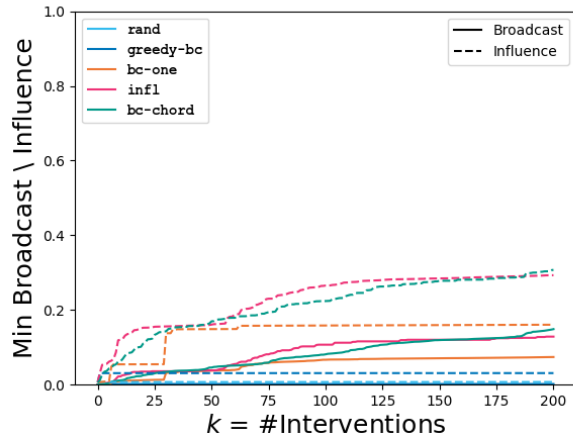
(c) **Irvine (Left:** $\alpha = 0.2$, **Right:** $\alpha = 0.4$)

Figure 2: For smaller networks, we plot the improvement in min. broadcast and influence for low-moderate- and well-spreading α**.**
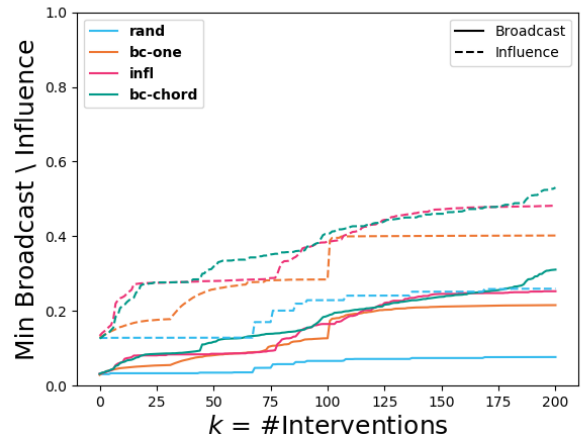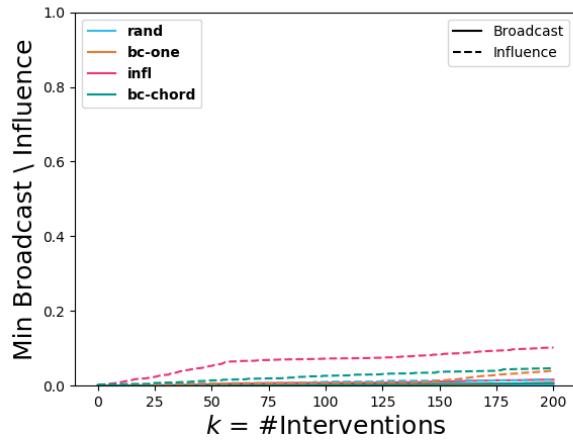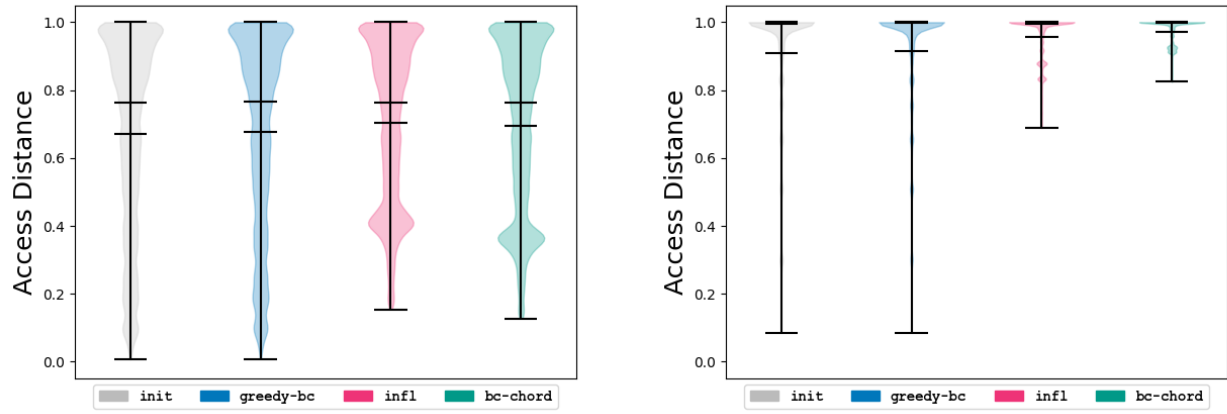
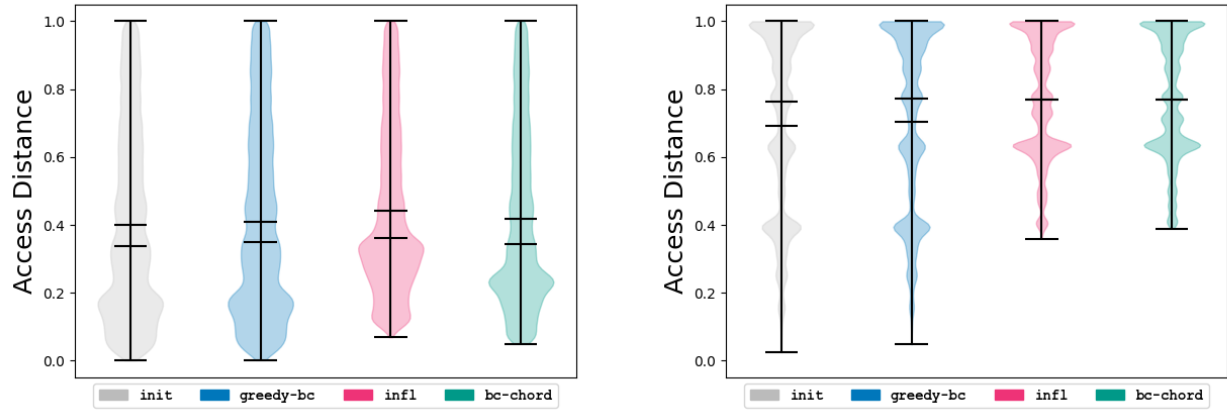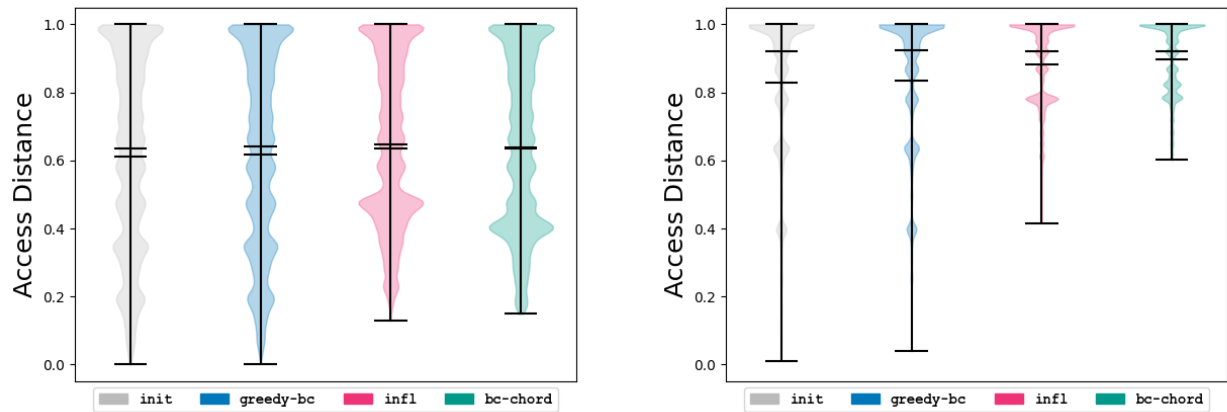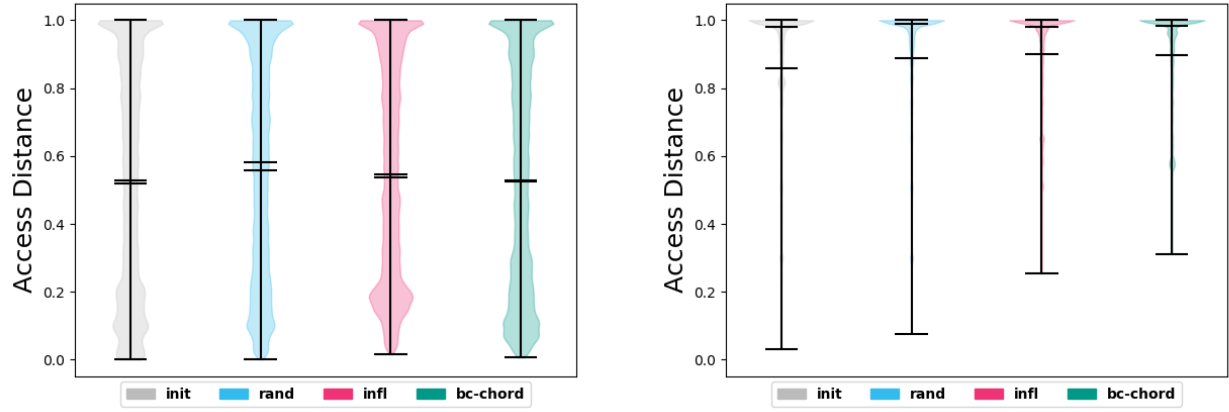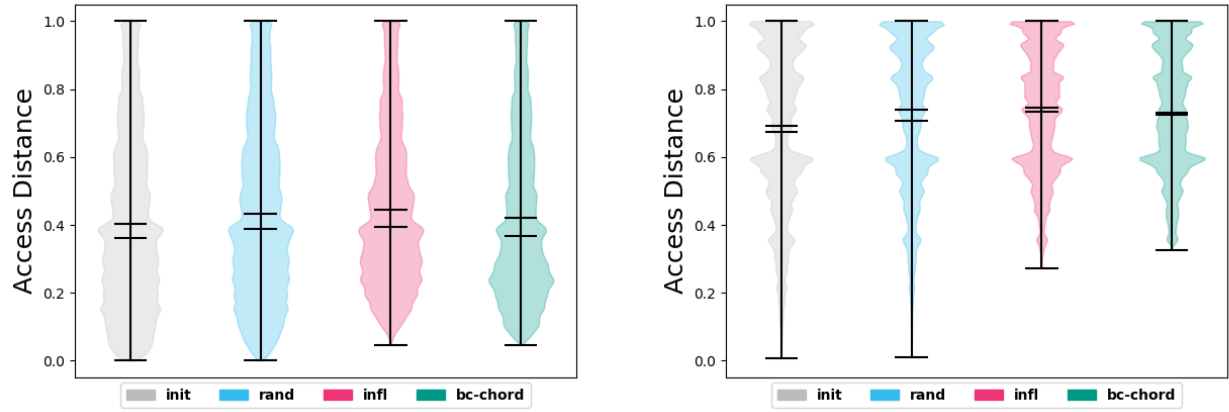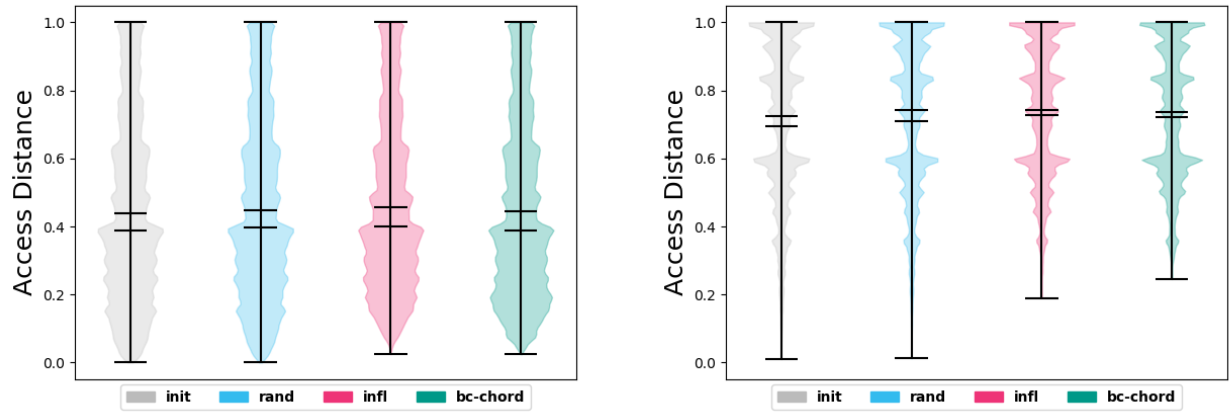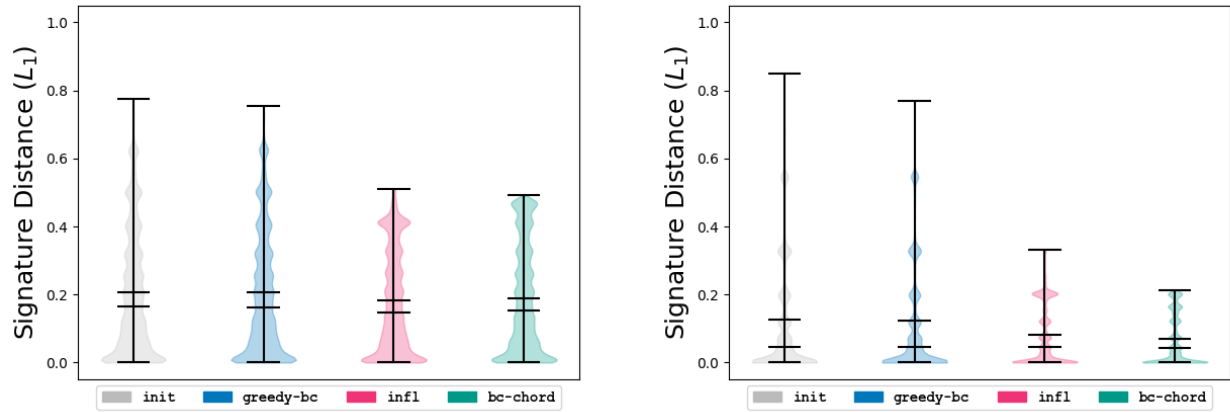(a) Facebook (Left: $\alpha = 0.1$, Right: $\alpha = 0.3$)
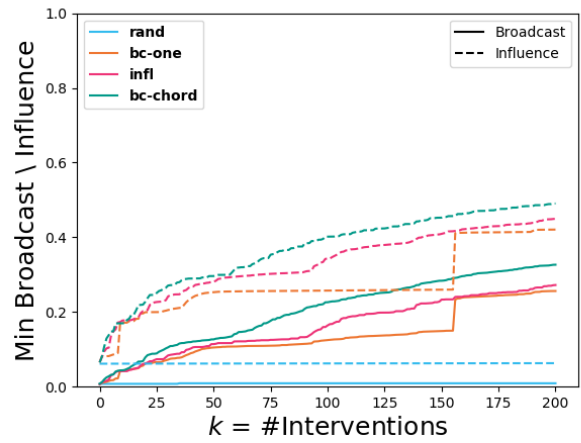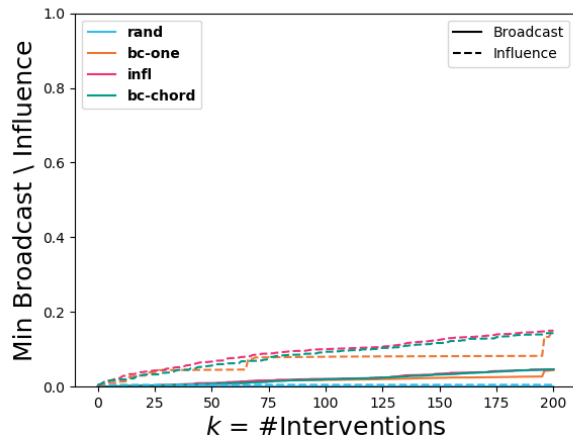
(b) ca-GrQc (Left: $\alpha = 0.4$, Right: $\alpha = 0.6$)

(c) ca-HepTh (Left: $\alpha = 0.4$, Right: $\alpha = 0.6$)

Figure 3: For larger networks, we plot the improvement in min. broadcast and influence for low-moderate- and well-spreading α.

(a) Email-EU (Left: $\alpha = 0.1$, Right: $\alpha = 0.3$)

(b) Email-Arenas (Left: $\alpha = 0.2$, Right: $\alpha = 0.4$)

(c) Irvine (Left: $\alpha = 0.2$, Right: $\alpha = 0.4$)

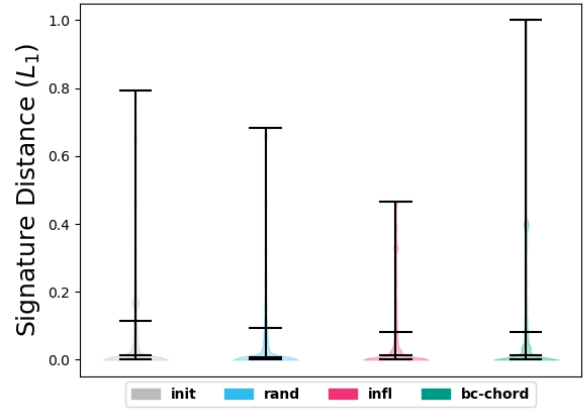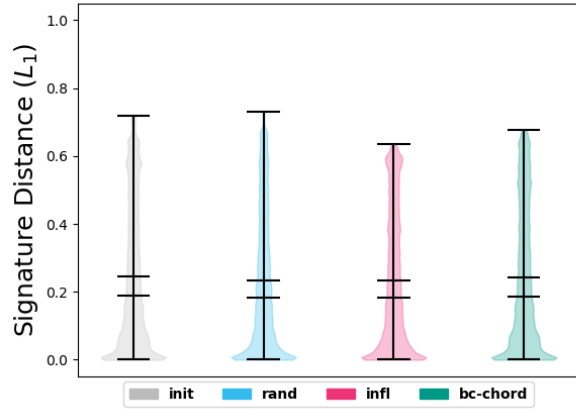Figure 4: For smaller networks, we plot the distribution of pairwise access proximities for low-moderate- and well-spreading α.
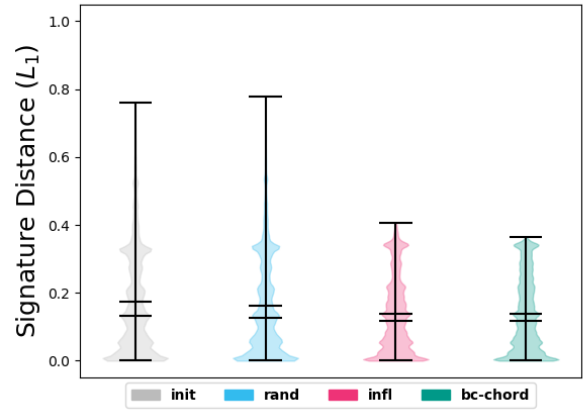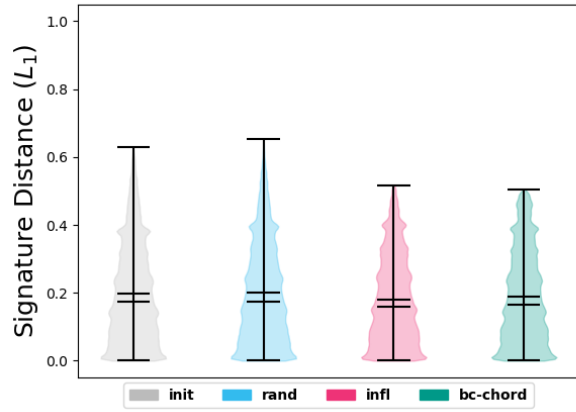
(a) Facebook (Left: $\alpha = 0.1$, Right: $\alpha = 0.3$)

(b) ca-GrQc (Left: $\alpha = 0.4$, Right: $\alpha = 0.6$)

(c) ca-HepTh (Left: $\alpha = 0.4$, Right: $\alpha = 0.6$)

Figure 5: For larger networks, we plot the distribution of pairwise access proximities for low-moderate- and well-spreading α.
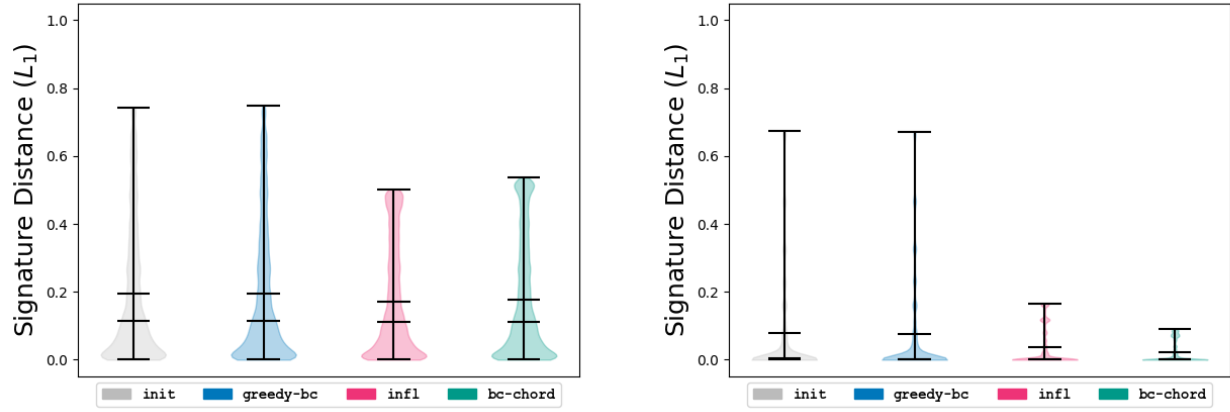
(a) Email-EU (Left: $\alpha = 0.1$, Right: $\alpha = 0.3$)
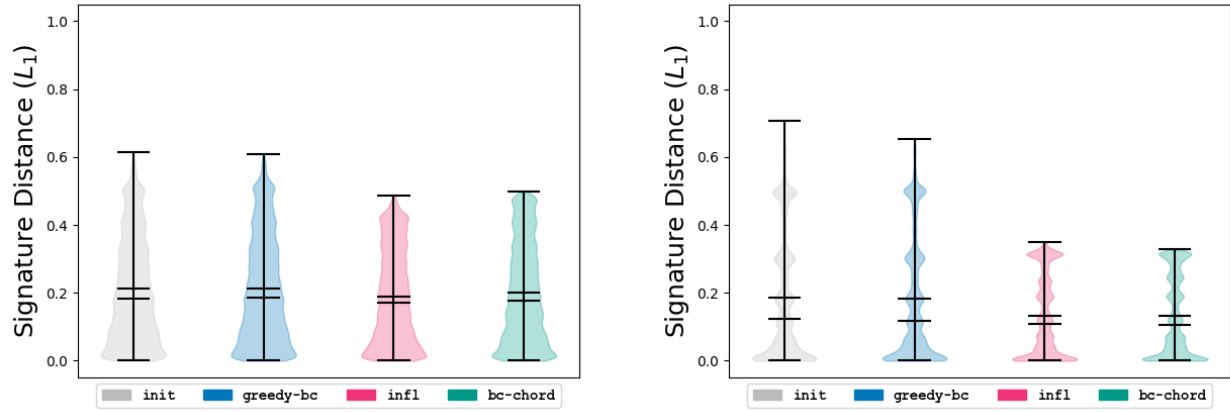
(b) Email-arenas (Left: $\alpha = 0.2$, Right: $\alpha = 0.4$)

(c) Irvine (Left: $\alpha = 0.2$, Right: $\alpha = 0.4$)

Figure 6: For smaller networks, we plot the distribution of pairwise signature manhattan for low-moderate- and well-spreading α.
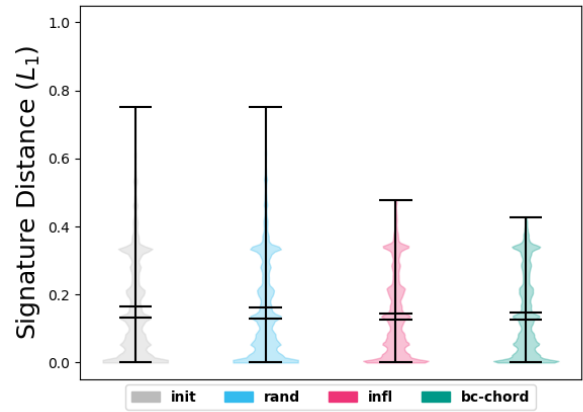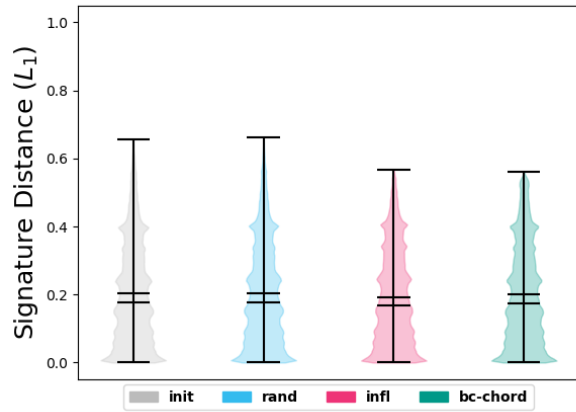
(a) Facebook (Left: $\alpha = 0.1$, Right: $\alpha = 0.3$)

(b) ca-GrQc (Left: $\alpha = 0.4$, Right: $\alpha = 0.6$)

(c) ca-HepTh (Left: $\alpha = 0.4$, Right: $\alpha = 0.6$)

Figure 7: For larger networks, we plot the distribution of pairwise signature manhattan distances for low-moderate- and well-spreading α.