Mutual Information Measure for Glass Ceiling Effect in Preferential Attachment Models

Rui Luo, Buddhika Nettasinghe, and Vikram Krishnamurthy, *Fellow, IEEE*

***Abstract*—This article introduces a novel mutual information-based measure to assess the glass ceiling effect in preferential attachment networks, which advances the analysis of inequalities in attributed networks. Using Shannon entropy and generalizing to Rényi entropy, our measure evaluates the conditional probability distributions of node attributes given the node degrees of adjacent nodes, which offers a more nuanced understanding of inequality compared to traditional methods that emphasize node degree distributions and degree assortativity alone. To evaluate the efficacy of the proposed measure, we evaluate it using an analytical structural inequality model as well as historical publication data. Results show that our mutual information measure aligns well with both the theoretical model and empirical data, underscoring its reliability as a robust approach for capturing inequalities in attributed networks. Moreover, we introduce a novel stochastic optimization algorithm that utilizes a parameterized conditional logit model for edge addition. Our algorithm is shown to outperform the baseline uniform distribution based approach in mitigating the glass ceiling effect. By strategically recommending links based on this algorithm, we can effectively hinder the glass ceiling effect within networks.**

***Index Terms*—Citation gap, directed mixed preferential attachment (DMPA) model, glass ceiling effect, mutual information, Rényi entropy.**

I. Introduction

THE glass ceiling effect is a subtle yet profound barrier that keeps certain groups of people from attaining equal advancement or visibility. Social networks are an important area where the glass ceiling effect arises. For instance, faculty hiring networks have been used to examine inequality and hierarchical structure in university prestige and gender [1]. Coauthorship of articles enables us to examine how researchers collaborate and what factors have the strongest impact on that collaboration [2]. Studies of citation networks have uncovered structural inequalities that adversely affect the research influence of scholars from particular groups [3] or based on gender [4]. Notably, female researchers often face underrecognition, with their work frequently appearing in less prestigious journals [5] and receiving fewer citations [6].

In this article, we employ the mutual information measure, derived from Shannon entropy and more generally Rényi entropy, to quantify the glass ceiling effect in an attributed network. Specifically, we demonstrate that the mutual information between the conditional probabilities of two nodes linked by a randomly selected edge provides critical insights into the reduction of uncertainty regarding the attributes of the nodes, conditioned on their degrees. In this context, the conditional probability indicates the likelihood of encountering particular attribute values for a node, given its degree. This measure reflects the complex dependencies in an attributed network and allows us to identify the presence of glass ceiling effect. It helps identify whether a node's demographic traits, such as gender, have an impact on the recognition or attention she receives within the network.

The proposed mutual information measure offers several advantages over existing measures such as power inequality [7], moment glass ceiling [7], tail glass ceiling [7], and gendered citation gap [4], which rely solely on the separate degree distributions of two gender groups.

1) *Capturing General Dependencies:* The mutual information measure takes into account the entire degree distribution and its relationship with node attributes, such as gender. This allows it to capture general dependencies between degrees and attributes, beyond simple linear relationships. In contrast, measures such as assortativity [8] only consider linear correlations, potentially overlooking more complex associations.
2) *Efficient Optimization:* By transforming the discrete optimization problem into a stochastic optimization [9] over the probability space of edge addition, the mutual information measure enables the development of computationally efficient algorithms. This transformation makes it easier to explore and optimize network structures while considering the interplay between node degrees and attributes. Existing measures, which focus on separate degree distributions, may not lend themselves to such optimization techniques.

Manuscript received 27 August 2023; revised 16 April 2024, 21 May 2024, and 9 July 2024; accepted 18 July 2024. Date of publication 8 August 2024; date of current version 3 December 2024. This work was supported in part by the U.S. Army Research Office under Grant W911NF-21-1-0093, in part by the National Science Foundation under Grant CCF-2112457, and in part by the City University of Hong Kong under Grant 9610639. *(Corresponding author: Rui Luo.)*

Rui Luo is with the Department of Systems Engineering, City University of Hong Kong, Kowloon 999077, Hong Kong SAR (e-mail: ruiluo@cityu.edu.hk).

Buddhika Nettasinghe is with Tippie College of Business, University of Iowa, Iowa City, IA 52242 USA (e-mail: buddhika-nettasinghe@uiowa.edu).

Vikram Krishnamurthy is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850 USA (e-mail: vikramk@cornell.edu).

Digital Object Identifier 10.1109/TCSS.2024.3432600

3) *Analytical and Computational Utility:* The proposed method represents a novel application of information theory, specifically mutual information, to network science [10]. By quantifying the reduction in uncertainty about node attributes given node degrees, it provides a principled and interpretable framework for analyzing attributed networks. This approach offers both analytical insights into the relationships between degrees and attributes and computational tools for optimizing network structures based on these relationships.

In summary, the mutual information measure goes beyond simple comparisons of separate degree distributions by capturing general dependencies, enabling efficient optimization, and providing a principled information-theoretic framework for analyzing attributed networks. These advantages make it a valuable addition to the existing set of network inequality measures, offering new perspectives and computational possibilities for understanding and mitigating disparities in network structures.

This article is organized as follows.

Section II provides an overview of related work, including discrete choice models for networks, the glass ceiling effect, and information-theoretic measures. In Section III, we introduce the concept of an attributed network and propose a novel mutual information measure to quantify the glass ceiling effect. We also introduce a data structure, the joint degree and attribute matrix (JDAM), for efficient computation. Section IV presents a stochastic optimization algorithm designed to mitigate the glass ceiling effect by minimizing the proposed mutual information measure. Finally, Section V evaluates the effectiveness of our measure through simulations and real-world citation network analysis.

II. Related Work

A. Discrete Choice in Networks

An array of studies [11], [12], [13] delve into the process of network formation through a discrete choice theory framework. These studies propose that the origins of network inequality could be traced back to factors that influence decision-making processes.

In particular, Overgoor et al. [11] portray network growth as a sequence of discrete decisions made by each node to form connections with others. They utilized the multinomial logit (MNL) model to articulate the probability of selecting each alternative, integrating various network features such as preferential attachment and homophily. Gupta and Porter [13] subsequently introduced the repeated-choice (RC) model as an advanced tool to study network formation. This RC model addresses certain limitations of the MNL model, notably the presumption that each node makes only one decision and shares a common choice set. Furthermore, to comprehend the effects of degree and attribute assortativity on network formation, Sadler [14] formulated a diffusion game. In this game, each node elects whether to adopt a new behavior. To demonstrate the diffusion of this new behavior, Sadler employed a branching process on a randomly generated network, where nodes are chosen based on a configuration model. Concurrently, this study investigates how to extend degree and attribute assortativity to their nonlinear counterparts using mutual information.

In this present work, we similarly employ a conditional logit model to add edges based on an optimized probability distribution, thus extending the application of the proposed mutual information measure.

B. Glass Ceiling Effect in Networks

A variety of discrete-time network models [3], [7], [15] have been proposed to effectively represent the dynamics of the glass ceiling effect.

Avin et al. [7] put forth a preferential attachment model of undirected networks to study the glass ceiling effect. In this model, the presence of preferential attachment, homophily, and minority groups are deemed necessary and sufficient conditions for the emergence of the glass ceiling effect. Subsequently, Nettasinghe et al. [3], [15] developed the directed mixed preferential attachment (DMPA) model. This model weaves directed preferential attachment and homophily into the growth dynamics of a directed graph, containing both a minority and a majority group. They conducted thorough analyses of the in-degree and out-degree distributions of the minority and majority groups within the DMPA model. As such, the DMPA model delves into the glass ceiling effect by examining the interaction between the structure and dynamics of directed networks. Teich et al. [4] recently investigated the under-citation of women's contributions in contemporary physics and explores ways to mitigate this gendered citation gap. The authors analyzed over 1 million articles published in 35 physics journals between 1995 and 2020, using a name-based gender categorization. The authors define the over/under-citation of an author gender category as the percent difference between the actual number of citations received and the expected number of citations based on a gender-blind model. The article also discussed strategies to mitigate the gender citation gap at individual, journal, and community levels, such as including citation diversity statements and reconsidering reference list length limits.

In our study, we employ the DMPA model to examine the efficacy of the proposed mutual information measure. In particular, we compute mutual information from each network snapshot generated from the DMPA model simulation. This approach enables the depiction of the evolution of the glass ceiling effect over time, making it highly compatible with existing models.

Some recent works provide critical insights that challenge conventional models such as the Barabási–Albert (BA) model, and offer new perspectives that could enrich our understanding of glass ceiling effect in networks. The embedding model discussed by Papadopoulos et al. [16] introduces a new perspective on how node popularity and similarity contribute to network growth, diverging from the traditional view that primarily emphasizes node degree. Similarly, the works by Benson et al. [17] and Shang et al. [18] provide further evidence of the evolving nature of network theories. Benson et al. challenge the simplistic view of network organization by introducing the concept of higher order structures. On the other hand, Shang et al. offer a model that incorporates community structures into network growth, which is crucial for understanding assortativity and heterogeneity within networks.

C. Information Theoretic Measures for Networks

Another line of research relevant to this study is information theory for networks. Wiedermann et al. [19] adapted Shannon entropy to network science. They introduced a statistical complexity measure that utilizes the divergence between the actual network and an appropriate reference model, such as a random network, to classify complex networks. Radicchi et al. [20] utilized maximum entropy principle which quantifies the tradeoff between the entropy of the network ensemble and the entropy of its compressed representation to explain the heterogeneity in a spatial network.

In addition to network entropy, [21] proposed a network fisher information measure to assess the degree of disorder within a network and evaluate regime transitions of random networks in the Shannon–Fisher plane. Furthermore, [22] suggested an information content measure that captures the distinction between a network's degree heterogeneity and the assortative noise present within it.

Inspired by these noteworthy network information-theoretic measures, our study introduces a novel Rényi mutual information measure. In addition to considering the degree and joint degree distribution, we also incorporate attribute and joint attribute distributions to quantify attribute inequality within the network.

III. Proposed Methodology

This section begins by introducing the key concepts of attributed networks, remaining degree distribution, and the joint remaining degree distribution of adjacent nodes, which are essential for defining the information content of a network. Then, we propose the mutual information, that incorporates node attributes to measure the information content of an attributed network. Then, we introduce an extension of mutual information, mutual information, and demonstrate its superior correlation coefficient with node attribute assortativity. This analysis offers critical insights into the role of node attributes in network edge formation and emphasizes the importance of incorporating these attributes into network inequality analysis. Our proposed mutual information measure provides a valuable tool for assessing the impact of node attributes on network structure and can aid in mitigating structural inequalities in networks. Notations used in this article are summarized in Table I.

A. Definition of an Attributed Network

We describe an attributed network as a directed multigraph with node labels, $G=(V,E,w,f)$, where V is the set of nodes; $E\subseteq V\times V$ is the set of directed edges between nodes; $w: E\to\mathbb{Z}^+$ denotes the edge multiplicities; $f: V\to\{\pm1\}$ assign a binary label to each node based on a demographic attribute such as gender.[1] The adjacency matrix A can be defined as

TABLE I
SYMBOLS USED IN THIS ARTICLE

Symbols	Description
G	The attributed network
V	The fixed node set of the network
E	The edge set of the network
w	The edge multiplicities function
f	The node attribute function
$p_{\text{out}}(k), p_{\text{in}}(k)$	The node degree distribution
$q_{\text{out}(k)}, q_{\text{in}}(k)$	The remaining node degree distribution
$e(k,k')$	The joint remaining degree distribution
H	The Shannon entropy
H_α	The Rényi entropy
I_α	The mutual information measure

$$A_{ij}=\begin{cases} w(i,j), & \text{if } (i,j)\in E\\ 0, & \text{otherwise.}\end{cases} \tag{1}$$

Drawing a parallel to a citation network for illustration, an edge (i,j) indicates that the target node j cites the source node i, meaning the edge's orientation represents the flow of information or influence from node i to node j. Consequently, the out-degree of node i, denoted as $d_{\text{out}}(i)$, quantifies the number of times others cite node i, whereas the in-degree, denoted as $d_{\text{in}}(i)$, quantifies the number of times node i cites other nodes. We have that

$$d_{\text{out}}(i)=\sum_{(i,j)\in E} w(i,j), \quad d_{\text{in}}(i)=\sum_{(j,i)\in E} w(j,i). \tag{2}$$

The out-degree distribution is defined as $p_{\text{out}}(k)$, $k=0,1,\ldots,k_{\text{out}}^{\max}$, where $p_{\text{out}}(k)$ denotes the probability that a randomly chosen node in the graph will have an out-degree of k. Similarly, the in-degree distribution is defined as $p_{\text{in}}(k)$, $k=0,1,\ldots,k_{\text{in}}^{\max}$, where $p_{\text{in}}(k)$ denotes the probability that a randomly chosen node in the graph will have an in-degree of k. Here, $k_{\text{out}}^{\max}$ and $k_{\text{in}}^{\max}$ represent the maximum out-degree and in-degree observed in the graph, respectively.

We are interested here in the remaining out-degree distribution [8] defined as

$$q_{\text{out}}(k)=\frac{(k+1)p_{\text{out}}(k+1)}{\sum_{j\in\{1,\ldots,k_{\text{out}}^{\max}\}} j p_{\text{out}}(j)}, k\in\{0,\ldots,k_{\text{out}}^{\max}-1\} \tag{3}$$

which quantifies the distribution of the number of edges leaving a node, other than the edge that we arrived along. Similarly, the remaining in-degree distribution is defined as

$$q_{\text{in}}(k)=\frac{(k+1)p_{\text{in}}(k+1)}{\sum_{j\in\{1,\ldots,k_{\text{in}}^{\max}\}} j p_{\text{in}}(j)}, k\in\{0,\ldots,k_{\text{in}}^{\max}-1\} \tag{4}$$

which quantifies the distribution of the number of edges entering a node, other than the edge that we arrived along. The remaining degree distributions are important because they capture

[1]We can do the partition based on other attributes as well in citation networks, such as the prestige of authors' institutional affiliation [3] and the majority/minority identity.

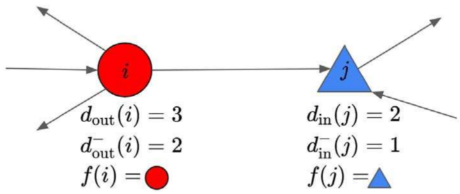

Fig. 1. We display the out-degree $d_{\text{out}}(i)$, the remaining out-degree $d^-_{\text{out}}(i)$, and the node attribute $f(i)$ of the source node i and the in-degree $d_{\text{in}}(j)$, the remaining in-degree $d^-_{\text{in}}(j)$, and the node attribute $f(j)$ of the target node j in an attributed network.

the probability of reaching a node with a certain degree when following a randomly chosen edge in the graph.

Now, we define $e(k,k')$ to be the joint probability distribution of the remaining out-degree k of the source node and the remaining in-degree k' of the target node at the ends of a randomly chosen edge. This quantity is not necessarily symmetric, meaning that $e(k,k')$ may not be equal to $e(k',k)$. It obeys the sum rules $\sum_{k,k'} e(k,k') = 1$, $\sum_{k'} e(k,k') = q_{\text{out}}(k)$, and $\sum_{k'} e(k',k) = q_{\text{in}}(k)$. Fig. 1 shows the statistics of a pair of nodes connected by a directed edge, including the remaining out-degree d^-_{out} of the source node, the remaining in-degree d^-_{in} of the target node, and their respective attributes.

B. Mutual Information Measure of an Attributed Network

In this section, we define the mutual information measure of an attributed network, where network degree information is readily accessible, such as the number of citations an author has received. However, node attributes, such as gender, research area, or academic rank, may not be entirely determined by node degrees. To capture this, we propose the mutual information framework, which assesses the decrease in uncertainty about node attributes when considering node degrees. This framework quantifies the extent to which node degrees provide information about node attributes, even when the attributes are not entirely dependent on the degrees, which proves to be a useful measure of glass ceiling effect.

1) Shannon Mutual Information: The Shannon entropy[2] of a network with respect to the remaining degree distribution (3) and (4) are defined as follows:

$$\begin{aligned} H(q_{\text{out}}) &= -\sum_{k\in\{0,\ldots,k^{\max}_{\text{out}}-1\}} q_{\text{out}}(k)\log(q_{\text{out}}(k)) \\ H(q_{\text{in}}) &= -\sum_{k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}} q_{\text{in}}(k')\log(q_{\text{in}}(k')) \end{aligned} \tag{5}$$

which quantifies the heterogeneity of the network's remaining out- and in-degree distributions.

Furthermore, the network's information content [22] (also referred to as information transfer in [10]) can be determined by considering the joint distribution of the remaining out-degree j of the source node and the remaining in-degree k of the target node at the ends of a randomly chosen directed edge. Specifically, the information content is defined as

$$I(q_{\text{out}}; q_{\text{in}}) = H(q_{\text{out}}) - H(q_{\text{out}}|q_{\text{in}}) \tag{6}$$

[2]We assume logarithm to base 2 throughout the article.

where

$$\begin{aligned} H(q_{\text{out}}|q_{\text{in}}) &= \sum_{k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}} q_{\text{in}}(k') \sum_{k\in\{0,\ldots,k^{\max}_{\text{out}}-1\}} \pi_{\text{out}}(k|k')\log\frac{1}{\pi_{\text{out}}(k|k')} \\ &= \sum_{\substack{k\in\{0,\ldots,k^{\max}_{\text{out}}-1\}\\ k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}}} e(k,k')\log\frac{q_{\text{in}}(k')}{e(k,k')} \end{aligned} \tag{7}$$

is the conditional entropy defined via conditional probabilities $\pi_{\text{out}}(k|k') = (e(k,k')/q_{\text{in}}(k'))$ of observing a node with k edges leaving it provided that the node at the other end of the chosen edge has k' incoming edges. In information-theoretic terms, $H(q_{\text{out}}|q_{\text{in}})$ is equivocation (nonassortativity) between q_{out} and q_{in}, i.e., the assortative noise within the network's information channel [23].

Given the entropy and the conditional entropy, the information content (6) can be expressed as

$$I(q_{\text{out}}; q_{\text{in}}) = \sum_{\substack{k\in\{0,\ldots,k^{\max}_{\text{out}}-1\}\\ k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}}} e(k,k')\log\frac{e(k,k')}{q_{\text{out}}(k)q_{\text{in}}(k')} \tag{8}$$

which measures the amount of network degree correlations, i.e., the information of an author's degree conveyed by degrees of authors she cites.

2) Rényi Mutual Information: The Rényi entropy [24], [25] is a generalization of Shannon entropy and has many applications in fields such as unsupervised learning, source adaptation, and image registration. The Rényi entropy of order α is defined as follows for the node remaining degree distribution q:

$$\begin{aligned} H_\alpha(q_{\text{out}}) &= \frac{1}{1-\alpha}\log\left(\sum_{k\in\{0,\ldots,k^{\max}_{\text{out}}-1\}} q_{\text{out}}(k)^\alpha\right) \\ H_\alpha(q_{\text{in}}) &= \frac{1}{1-\alpha}\log\left(\sum_{k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}} q_{\text{in}}(k')^\alpha\right) \end{aligned} \tag{9}$$

where $\alpha \in (0,1)\cup(1,+\infty)$. The limiting value of H_α as $\alpha \to 1$ is the Shannon entropy (5). The conditional Rényi entropy (7) is defined following $H_\alpha(q|q_{\text{in}}) = H_\alpha(q_{\text{out}}, q_{\text{in}}) - H_\alpha(q_{\text{in}})$:

$$H_\alpha(q_{\text{out}}|q_{\text{in}}) = \frac{1}{1-\alpha}\log\left(\frac{\sum_{\substack{k\in\{0,\ldots,k^{\max}_{\text{out}}-1\}\\ k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}}} e(k,k')^\alpha}{\sum_{k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}} q_{\text{in}}(k')^\alpha}\right). \tag{10}$$

Replacing the Shannon entropy in the definition of mutual information (6) by the Rényi entropy, one obtains the Rényi mutual information

$$\begin{aligned} I_\alpha(q_{\text{out}}; q_{\text{in}}) &= H_\alpha(q_{\text{out}}) - H_\alpha(q_{\text{out}}|q_{\text{in}}) \\ &= \frac{1}{1-\alpha}\log\left(\frac{\sum_{\substack{k\in\{0,\ldots,k^{\max}_{\text{out}}-1\}\\ k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}}} q_{\text{out}}(k)^\alpha q_{\text{in}}(k')^\alpha}{\sum_{\substack{k\in\{0,\ldots,k^{\max}_{\text{out}}-1\}\\ k'\in\{0,\ldots,k^{\max}_{\text{in}}-1\}}} e(k,k')^\alpha}\right). \end{aligned} \tag{11}$$

Now taking the node attributes into account, we define the joint mutual information as

$$I_\alpha(q_{\text{out}}, m_{\text{out}}; q_{\text{in}}, m_{\text{in}}) = H_\alpha(q_{\text{out}}, m_{\text{out}}) - H_\alpha(q_{\text{out}}, m_{\text{out}} | q_{\text{in}}, m_{\text{in}}) \tag{12}$$

which is the mutual information between the two bivariate variables $(q_{\text{out}}, m_{\text{out}})$ and $(q_{\text{in}}, m_{\text{in}})$ that stand for the degrees and the attributes of a pair of two adjacent nodes connected by a directed edge. It can be expanded as

$$\begin{aligned} & I_\alpha(q_{\text{out}}, m_{\text{out}}; q_{\text{in}}, m_{\text{in}}) \\ & = \frac{1}{1-\alpha} \log \left(\frac{\sum\limits_{\substack{k \in \{0,\ldots,k_{\text{out}}^{\max}-1\} \\ k' \in \{0,\ldots,k_{\text{in}}^{\max}-1\}}} \sum\limits_{\substack{c \in \{\pm 1\} \\ c' \in \{\pm 1\}}} p(k,c)^\alpha p(k',c')^\alpha}{\sum\limits_{\substack{k \in \{0,\ldots,k_{\text{out}}^{\max}-1\} \\ k' \in \{0,\ldots,k_{\text{in}}^{\max}-1\}}} \sum\limits_{\substack{c \in \{\pm 1\} \\ c' \in \{\pm 1\}}} p(k,c,k',c')^\alpha} \right) \end{aligned} \tag{13}$$

where $p(k, c)$ and $p(k', c')$ represent the joint distribution of degree and attribute for the source node and the target node, respectively, whereas $p(k, c, k', c')$ represents the joint distribution of degree and attribute of a pair of nodes connected by a directed edge.

The proposed mutual information measure is defined as the conditional mutual information of the attributes of neighboring nodes given the degrees of the nodes

$$\begin{aligned} I_\alpha &= I_\alpha(m_{\text{out}} | q_{\text{out}}; m_{\text{in}} | q_{\text{in}}) \\ &= I_\alpha(q_{\text{out}}, m_{\text{out}}; q_{\text{in}}, m_{\text{in}}) - I_\alpha(q_{\text{out}}; q_{\text{in}}) \end{aligned} \tag{14}$$

where the mutual information is specifically defined between the conditional probability distributions of node attributes given node degrees of adjacent nodes.

We can rewrite (14) as

$$I_\alpha = \frac{1}{1-\alpha} \cdot \log \left(\frac{\sum\limits_{\substack{k \in \{0,\ldots,k_{\text{out}}^{\max}-1\} \\ k' \in \{0,\ldots,k_{\text{in}}^{\max}-1\}}} \sum\limits_{\substack{c \in \{\pm 1\} \\ c' \in \{\pm 1\}}} e(k,k')^\alpha p(k,c)^\alpha p(k',c')^\alpha}{\sum\limits_{\substack{k \in \{0,\ldots,k_{\text{out}}^{\max}-1\} \\ k' \in \{0,\ldots,k_{\text{in}}^{\max}-1\}}} \sum\limits_{\substack{c \in \{\pm 1\} \\ c' \in \{\pm 1\}}} q_{\text{out}}(k)^\alpha q_{\text{in}}(k')^\alpha p(k,c,k',c')^\alpha} \right). \tag{15}$$

The Shannon mutual information is the special case of (15) when $\alpha \to 1$

$$I = \sum_{\substack{k \in \{0,\ldots,k_{\text{out}}^{\max}-1\} \\ k' \in \{0,\ldots,k_{\text{in}}^{\max}-1\}}} \sum_{\substack{c \in \{\pm 1\} \\ c' \in \{\pm 1\}}} \left(p(k,k',c,c') \log \frac{q_{\text{out}}(k) q_{\text{in}}(k') p(k,k',c,c')}{e(k,k') p(k,c) p(k',c')} \right). \tag{16}$$

I_α quantifies the increase in information content of an attributed network due to the inclusion of node attributes. In other words, I_α provides insight into the additional information that node attributes can offer beyond what can be explained by degree assortativity alone, thereby enabling a more accurate assessment of the underlying network inequality and glass ceiling effect.

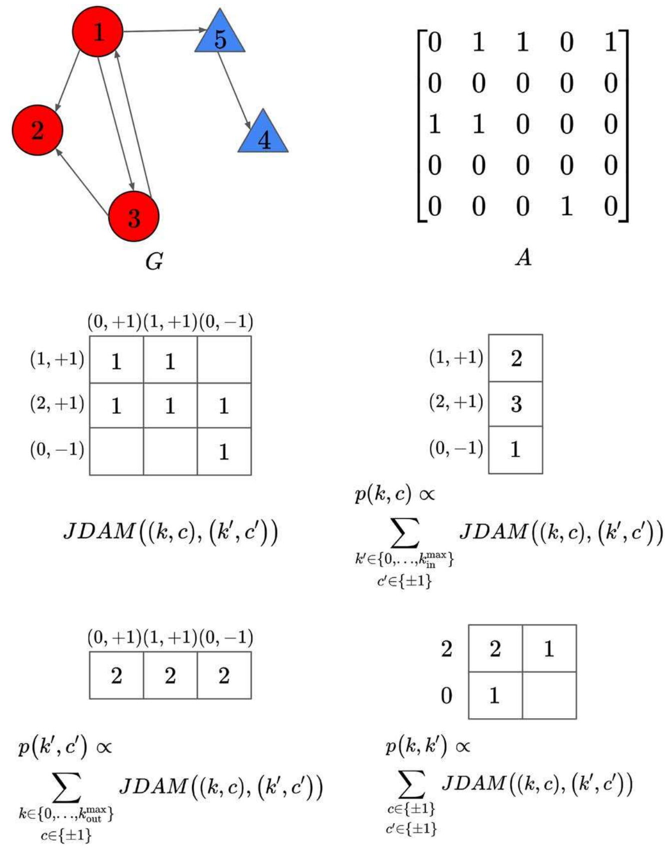

Fig. 2. Representation of the JDAM (17) for an attributed network. Upper section displays an attributed network G alongside its corresponding adjacency matrix A; centrally located on the left is the JDAM. The other three visuals illustrate aggregating the JDAM across rows, columns, and attributes to derive the joint distributions (18). These visuals demonstrate the usefulness of the JDAM in computing the information content (6) and the mutual information (15). We disregard any (k, c) where $p(k, c) = 0$ and (k', c'), where $p(k', c') = 0$ to represent JDAM more concisely.

To visualize I_α for an attributed network, we first define the (out-degree, attribute) group $B_{\text{out}}(k, c) = \{i | q_{\text{out}}(i) = k, f(i) = c\}$ as the set of nodes that have remaining out-degree k and attribute c. The number of (out-degree, attribute) groups is capped at $2k_{\text{out}}^{\max}$, each representing a unique combination of out-degree and attribute values. Similarly, we define the (in-degree, attribute) group $B_{\text{in}}(k, c) = \{i | q_{\text{in}}(i) = k, f(i) = c\}$ as the set of nodes that have remaining in-degree k and attribute c. The JDAM [26], is defined by the count of edges that link nodes from the (out-degree, attribute) group (k, c) to the (in-degree, attribute) group (k', c')

$$\text{JDAM}((k,c),(k',c')) = \sum_{\substack{i: q_{\text{out}}(i)=k, \\ f(i)=c}} \sum_{\substack{j: q_{\text{in}}(j)=k', \\ f(j)=c'}} w(i,j) \tag{17}$$

which capture the co-occurrence of node degrees and attributes for nodes that are connected by a directed edge within the network. Fig. 2 shows an example citation network and its corresponding JDAM.

Driven by the observation that the JDAM corresponds to the joint degree and attribute distribution of connected node pairs

$p(k, k', c, c')$, we can group its rows or columns by node degrees or attributes, leading to the following formulations:

$$\begin{aligned} e(k,k') &\propto \sum_{\substack{c\in\{\pm1\}\\ c'\in\{\pm1\}}} \mathrm{JDAM}((k,c),(k',c')) \\ p(k,c) &\propto \sum_{\substack{k'\in\{0,\dots,k_{\text{in}}^{\max}-1\}\\ c'\in\{\pm1\}}} \mathrm{JDAM}((k,c),(k',c')) \\ p(k',c') &\propto \sum_{\substack{k\in\{0,\dots,k_{\text{out}}^{\max}-1\}\\ c\in\{\pm1\}}} \mathrm{JDAM}((k,c),(k',c')) \\ q_{\text{out}}(k) &\propto \sum_{\substack{k'\in\{0,\dots,k_{\text{in}}^{\max}-1\}\\ c'\in\{\pm1\},c\in\{\pm1\}}} \mathrm{JDAM}((k,c),(k',c')) \\ q_{\text{in}}(k') &\propto \sum_{\substack{k\in\{0,\dots,k_{\text{out}}^{\max}-1\}\\ c\in\{\pm1\},c'\in\{\pm1\}}} \mathrm{JDAM}((k,c),(k',c')). \end{aligned} \tag{18}$$

Given the above reformulation, we are now able to calculate I_α (15) derived entirely from the JDAM.

C. Existing Glass Ceiling Measures

In this part, we examine alternative measures for assessing glass ceiling effect. For illustrative purposes, we use gender categories as node labels within a citation network context.

1) Power Inequality [7]: Power inequality is defined as the ratio of the average degree of male authors to that of female authors

$$\text{power} = \frac{\frac{\sum_{i:f(i)=+1} d(i)}{|i:f(i)=+1|}}{\frac{\sum_{i:f(i)=-1} d(i)}{|i:f(i)=-1|}} \tag{19}$$

where label $+1$ represents male and -1 represents female.

2) Tail Glass Ceiling [7]: Tail glass ceiling interprets the most powerful positions as those held by the highest degree nodes, and is defined as the ratio of the number of male authors with degree at least k to that of female authors with degree at least k

$$\text{tail} = \frac{|i : f(i) = +1, d(i) \ge k|}{|i : f(i) = -1, d(i) \ge k|}. \tag{20}$$

3) Moment Glass Ceiling [7]: Moment glass ceiling extends from (19) by considering the second moment of the degree sequences of male authors and female authors

$$\text{moment} = \frac{\frac{\sum_{i:f(i)=+1} d^2(i)}{|i:f(i)=+1|}}{\frac{\sum_{i:f(i)=-1} d^2(i)}{|i:f(i)=-1|}}. \tag{21}$$

Nevertheless, neither moment glass ceiling nor the previously mentioned measures are inherently tailored for directed graphs.

4) Gendered Citation Gap [4]: This is an additional metric for comparison which specifically examines the phenomenon where publications by women are cited less frequently, while those by men receive disproportionately more citations. Define MW as the total count of instances where a female author cites a male author, and WM as the total count of instances where a male author cites a female author. Let n_1, n_2, n_3, n_4 be the number of citations received by male authors, received by female authors, given by male authors, and given by female authors. It is evident that the sum of citations received by males and females equals the sum of citations given by males and females, i.e., $n_1 + n_2 = n_3 + n_4$. The over/under-citation of an author gender category is defined as the percent difference between the actual number of citations received and the expected number of citations based on a gender-blind null model

$$\text{gap} = \frac{\frac{\text{MW}}{n_1 n_4/(n_1+n_2)}}{\frac{\text{WM}}{n_2 n_3/(n_1+n_2)}} = \frac{\text{MW} n_2 n_3}{\text{WM} n_1 n_4} \tag{22}$$

where $n_1 n_4/(n_1 + n_2)$ denotes the expected number of citations received by male authors given by female authors.

IV. Stochastic Optimization for Reducing the Network Inequality

As discussed in Section III, the proposed mutual information measure quantifies the glass ceiling effect in an attributed network. In this section, our goal is to mitigate the glass ceiling effect by optimizing the mutual information measure. Our approach involves using a parameterized distribution to sample the edges to be added to the network. By updating the parameterized sampling distribution through stochastic optimization, we aim to maximize the expected value of the mutual information measure. This approach is superior to deterministic edge addition or removal as it guarantees a local optimum, and it is effective in reducing inequalities such as the glass ceiling effect. It achieves this objective through random link recommendation, which preserves privacy [27].

A. Stochastic Optimization Problem Formulation

In order to optimize the proposed mutual information measure of a network, we consider adding an edge between node in (out-degree, attribute) group (k, c) and node in (in-degree, attribute) group (k', c'). By doing so, we increase the entry $[(k+1, c), (k'+1, c')]$ in JDAM by one while reducing the entry $[(k, c), (k', c')]$ by one. The set of possible (out-degree, attribute) groups and (in-degree, attribute) groups before and after the edge addition are $\mathcal{V}_{\text{out}} = \{0, \dots, k_{\text{out}}^{\max}\} \times \{\pm1\}$ and $\mathcal{V}_{\text{in}} = \{0, \dots, k_{\text{in}}^{\max}\} \times \{\pm1\}$, respectively. The edge space is thus $\mathcal{E} = \mathcal{V}_{\text{out}} \times \mathcal{V}_{\text{in}}$, which is $4(k_{\text{out}}^{\max} + 1)(k_{\text{in}}^{\max} + 1)$ dimensional.

To formulate the optimization problem, we define the objective function as the change in mutual information resulting from adding an edge to the network, i.e., $Q(x) = I_\alpha(E \cup \{x\}) - I_\alpha(E)$, where $x \in \mathcal{E}$ and $E \cup \{x\}$ represents the edge set after the edge addition. However, we demonstrate with a counterexample in Fig. 3 that the discrete optimization problem

$$x^* \in \operatorname*{argmin}_{x\in\mathcal{E}} Q(x) \tag{23}$$

is not submodular. Specifically, we find that $I_\alpha(E_a \cup \{x\}) - I_\alpha(E_a) = -0.218 < I_\alpha(E_c \cup \{x\}) - I_\alpha(E_c) = 0.551,$ even though $E_a \subset E_c$.

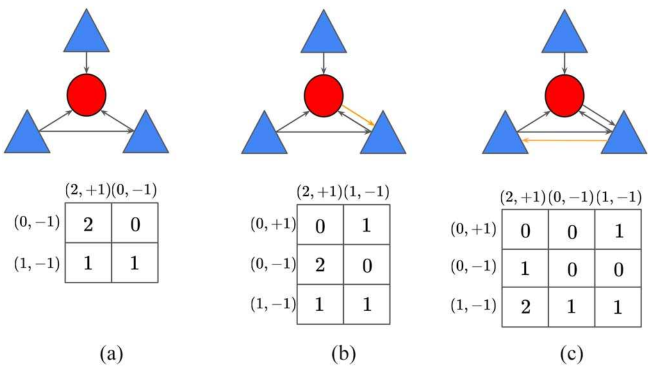

Fig. 3. Example illustrates a scenario where the discrete optimization problem (23) is shown to be not submodular. We present the attributed network at the top and the corresponding JDAM at the bottom. (a) Original network. (b) Network with one additional edge. (c) Network with another additional edge. Sequentially, we add one edge from (a) to (b) and then from (b) to (c). As a result, the I_α increases from 0 to 0.551 but subsequently decreases to 0.333. Therefore, the performance of greedy algorithm cannot be guaranteed. To address this, we relax the original problem into an optimization over the probability distribution space of edge additions. We describe the application of the stochastic optimization algorithm in Algorithm 1, demonstrating how it navigates the solution space to optimize the objective function in a nonsubmodular landscape.

To overcome this challenge, we relax the problem into a stochastic optimization across probability distributions of edges, with these distributions defined by a $|\mathcal{E}|$-dimensional parameter θ. The stochastic optimization algorithm allows us to find locally optimal solutions in a randomized manner. Specifically, we employ a distribution to select which edges to add to the network and update the distribution via stochastic optimization to maximize the expected value of I_α. This approach effectively resolves the edge addition problem while reducing the glass ceiling effect.

Consider a family of probability mass functions $f(x;\theta)$ on $\mathcal{E}$, parameterized by the conditional logit model with separate fixed effect θ_i for each edge i

$$f_i(\mathcal{E};\theta) = \frac{\exp(\theta_i)}{\sum_{i=1}^{|\mathcal{E}|}\exp(\theta_j)}. \tag{24}$$

Then, we formulate the optimization problem as

$$\begin{aligned}\theta^* &= \operatorname*{argmin}_{\theta\in\mathbb{R}^{|\mathcal{E}|}} C(\theta) \\ &= \operatorname*{argmin}_{\theta\in\mathbb{R}^{|\mathcal{E}|}} \mathbb{E}_{x\sim f(\mathcal{E};\theta)}\{Q(x)\}.\end{aligned} \tag{25}$$

Here, we optimize the probability distribution of the edges in (25), where the objective function is the expected increase in mutual information with respect to the probability distribution.

In the following section, we explain how we use the simultaneous perturbation stochastic approximation (SPSA) algorithm [28] to solve this stochastic optimization problem effectively.

B. Stochastic Optimization Algorithm

We use the SPSA algorithm [28], [29] to estimate the gradient of the new objective function with respect to each component of θ. One of the key advantages of SPSA is that it requires only two simulations of the objective function, regardless of the dimension $|\mathcal{E}|$ of the optimization problem.

Algorithm 1 SPSA based algorithm to estimate θ^*

Input: Initial parameterization $\theta^{(0)}$; JDAM of the current network.

Output: Estimate of the (locally) optimal solution θ^* of the conditional logit model for edge addition (25).

for $k = 0, 1, \cdots$ **do**
  Simulate the $|\mathcal{E}|$-dimensional vector d_k with random elements

$$d_k(i) = \begin{cases} +1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5 \end{cases} \tag{26}$$

  Sample an edge x from $\mathcal{E}$ using $f(\mathcal{E};\theta^{(k)} + \Delta d_k)$ (24), where $\Delta > 0$.
  Compute the change in the mutual information, $C(\theta^{(k)} + \Delta d_k)$.
  Sample an edge x from $\mathcal{E}$ using $f(\mathcal{E};\theta^{(k)} - \Delta d_k)$.
  Compute $C(\theta^{(k)} - \Delta d_k)$.
  Obtain the gradient estimate using (28).
  Update $\theta^{(k)}$ via stochastic gradient descent

$$\theta^{(k+1)} = \theta^{(k)} + \epsilon\hat{\nabla}C(\theta^{(k)}) \tag{27}$$

end for

Algorithm 1 outlines the SPSA implementation. In the kth iteration, all elements of θ undergo a random perturbation. A vector d_k made up of $\{\pm 1\}$ with a Bernoulli (0.5) distribution is scaled by a factor[3] $\Delta > 0$ and added and subtracted from θ. Then, by sampling the edge addition using the distribution represented by the two perturbed θ, the objective function $C(\theta)$ (25) is evaluated twice. The derivative of $C(\theta)$ with respect to θ is approximated by

$$\hat{\nabla}C(\theta^{(k)}) = \frac{C(\theta^{(k)} + \Delta d_k) - C(\theta^{(k)} - \Delta d_k)}{2\Delta} d_k. \tag{28}$$

Finally, the stochastic gradient algorithm can be used to update θ. As the SPSA algorithm employs stochastic gradient descent (28), and the objective function may not be convex, the algorithm can only converge to a local stationary point. However, in practice, local optima can still provide useful solutions. Specifically, SPSA converges to a sampling distribution that is a local maximum of the objective function.

In Section V-C, we demonstrate the practical effectiveness of our approach through numerical results. This application holds significant implications for fostering a more inclusive and equitable scientific community.

V. Numerical Evaluation

In this section, we demonstrate the ability of the proposed mutual information measure to quantify the glass ceiling effect

[3] Δ affects the bias-variance tradeoff as the bias in the derivative estimate (28) is proportional to Δ^2 whereas the variance is proportional to $1/\Delta^2$ [29]. In practice, we carry out a grid search over the parameters $\{\Delta, \epsilon\}$ to identify an optimal parameter set for the SGD algorithm (27).

in citation networks, using both analytical models and real-world datasets. Additionally, we apply Algorithm 1 (Section IV) to a real-world citation network to optimize the distribution over edge addition, with the goal of reducing glass ceiling effect. Our results demonstrate that the optimized conditional logit model outperforms uniformly random edge addition, with a significant improvement in reducing the growth of the proposed measure. These findings highlight the importance of considering the distribution of edge additions in reducing the glass ceiling effect in citation networks. The code and datasets used in the experiments are publicly available at GitHub.

A. DMPA Model

We presented the DMPA model [3] in this section. We also employed a condensed version of it to validate the mutual information proposed in Section III.

The DMPA model examines glass ceiling effect by modeling connection patterns in a growing citation network based on preferential attachment, minority group and homophily, as well as the rate at which new authors join. The two node types, m and f where f represents the minority group, correspond to the binary labels $c \in \{\pm 1\}$ defined in Section III-A. The network starts with a single node of type m that has bidirectional directed edges with another node of type f. At each time step, one of the three edge addition events occurs:

1) With probability p, a new node appears and an existing node cites it. The new node is assigned type f with probability $p(f) \leq (1/2)$ and m with probability $1 - p(f)$. The potential citing node is chosen from the existing nodes with probability proportional to their in-degrees plus a constant δ. A new citation edge is then created based on the probability matrix

$$P_{\text{att}} = \begin{bmatrix} \rho_{\text{att}} & 1 - \rho_{\text{att}} \\ 1 - \rho_{\text{att}} & \rho_{\text{att}} \end{bmatrix}. \quad (29)$$

Specifically, an edge is added with probability $\rho_{\text{att}} \in (0, 1)$ if both nodes have the same type. If not, an edge is added with probability $1 - \rho_{\text{att}}$. A value of $\rho_{\text{att}} > 0.5$ suggests the presence of homophily, indicative of a glass ceiling effect. We simplify the model by assuming nodes with different types share the ρ_{att}, making P_{att} symmetric.

2) With probability q, a new node appears and cites an existing node. The new node's type is assigned in the same manner as in case (1). The potential cited node is chosen from the existing nodes with probability proportional to their out-degrees plus a constant δ. Then, a new citation edge is created based on P_{att} (29).

3) With probability $1 - p - q$, a new edge is created between two existing nodes. Both the citing and cited nodes are chosen independently based on their in- and out-degree, and they are connected based on P_{att} (29).

In this DMPA model, $p(f)$ controls the imbalance between type m and f, while ρ_{att} determines whether the nodes are homophilic or heterophilic, which is integral to understanding the glass ceiling effect. At $\rho_{\text{att}} = 0.5$, nodes display no preference, being equally likely to connect with similar or different types.

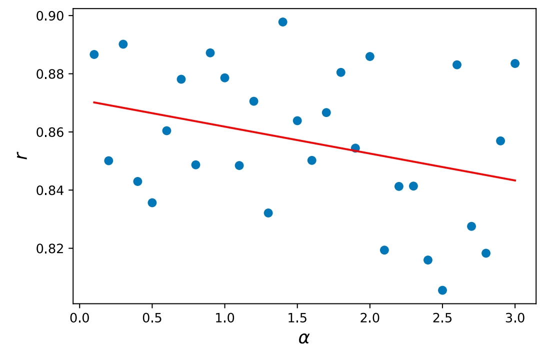

Fig. 4. We plotted the correlation coefficient r between the proposed mutual information measure I_α and the analytical model's parameter ρ_{att} (Section V-A1), across various values of α ranging from 0.1 to 3. Correlation coefficient r consistently exceeded 0.8. Additionally, we fitted a linear regression curve based on r and α, but the results were not significant, suggesting that the choice of α is not crucial for analyzing the glass ceiling effect.

In the simulation, we set $\rho_{\text{att}} \geq 0.5$, representing scenarios that include the glass ceiling effect. Then, we vary its value from 0.5 to 0.95 and calculate the correlation between ρ_{att} and the proposed mutual information measure I_α.

1) Effects of Rényi Mutual Information Order: To identify the optimal order α for the mutual information measure, we carried out a series of numerical experiments by adjusting α values from 0.1 to 3. For each α, we calculated the correlation coefficient r between I_α and ρ_{att}. This correlation was used to determine the most effective α value.

As illustrated in Fig. 4, regardless of the chosen order α, the correlation coefficient r between I_α and ρ_{att} consistently exceeds 0.8. Additionally, a linear regression curve was fitted to demonstrate the weak relationship between α and r, supporting our assertion that the order α is not crucial in analyzing the glass ceiling effect.

Choosing an order $\alpha = 1$ which is equivalent to Shanon's entropy and standard mutual information captures information on all the elements on the distributions, while still emphasizing the elements with high probabilities. In the subsequent experiments, we will use I_α with $\alpha = 1$.

2) Evaluation of the Rényi Mutual Information With the DMPA Model: In the simulation,[4] we set $p(f) \in [0.1, 0.5]$ so that type f corresponds to the minority. We vary ρ_{att} from 0.5 to 0.95 which covers a spectrum of homophily. We set $p = q = 0.15$ and $\delta = 10$. For each combination of the parameters $\{p(f), \rho_{\text{att}}\}$, we generate a network with 5000 edges.

Fig. 5 displays I_α of networks simulated using various combination of $\{p(f), \rho_{\text{att}}\}$. The scatter plot reveals a positive correlation between the mutual information measure and ρ_{att}. Additionally, we display a plot of I_α versus ρ_{att} for different $p(f)$ in Fig. 6. This further supports the positive correlation between I_α and ρ_{att}, affirming the effectiveness of the proposed mutual information measure in capturing the glass ceiling effect. On the other hand, Fig. 7 illustrates

[4]We used the code provided by authors of [3]: https://github.com/ninoch/DMPA.

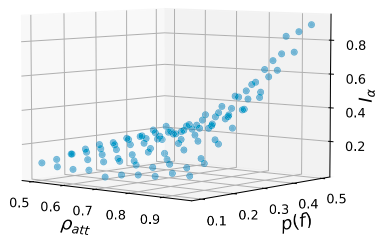

Fig. 5. Mutual information I_α of networks that were simulated from various combinations of $p(f)$ and ρ_{att} under the DMPA model, which indicates the positive correlation between I_α and ρ_{att}. $p(f) \in [0.05, 0.5]$ represents the proportion of the female nodes, while $\rho_{\text{att}} \in [0.50, 0.95]$ represents the probability that two nodes with the same gender connect with each other.

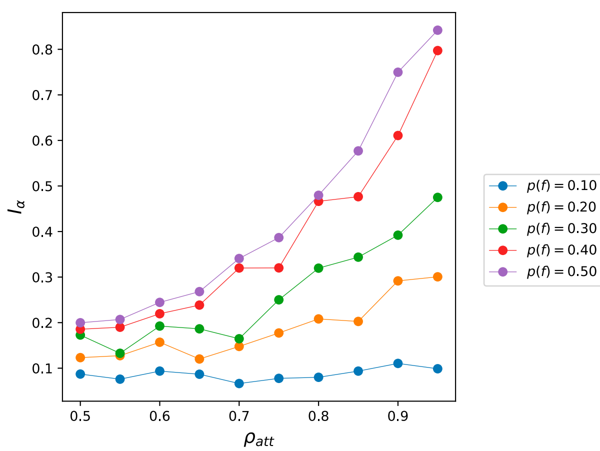

Fig. 6. Relationship between I_α and ρ_{att} of networks simulated from the DMPA model. Different colors represent different proportions of female nodes $p(f) \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. Positive correlation supports our claim that increased homophily leads to a higher mutual information measure, indicating that greater emphasis is placed on node attributes and there is a stronger manifestation of the glass ceiling effect.

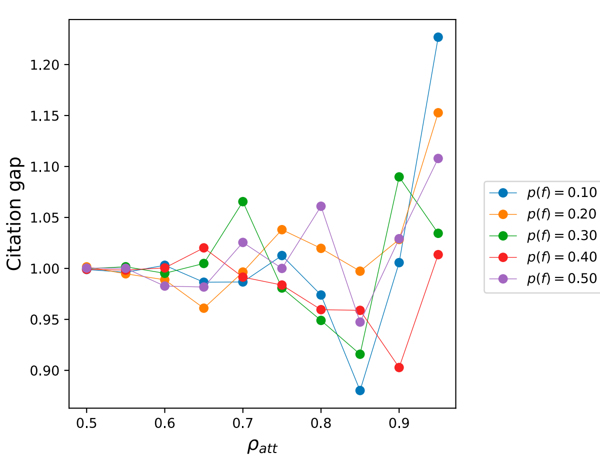

Fig. 7. Relationship between the gendered citation gap (22) and ρ_{att}. Plot demonstrates that the gendered citation gap fails to exhibit a positive correlation with the parameter ρ_{att}, which is associated with the glass ceiling effect.

the relationship between the gendered citation gap (22) and ρ_{att}, which fails to reveal a strong correlation between these two metrics.

Fig. 8. This figure displays the evolution of the mutual information measure across five academic fields, alongside the gendered citation gap for each field. Mutual information measure reveals varying trends: a decrease over time in political science suggests reduced gender influence in citations, while steady increases in physics and psychology indicate a growing glass ceiling effect. Management shows a stable trend with a notable spike in 2018, likely due to data anomalies. Except for physics, all fields exhibit a positive correlation between the mutual information measure and the gendered citation gap, demonstrating the measure's effectiveness in capturing the dynamics of the glass ceiling effect across disciplines.

B. Empirical Analysis of Publication Data

To confirm that mutual information measure is useful in quantifying the glass ceiling effect in real-world data, we use the publishing records from public datasets. We use the bibliographic data of publications in the field of physics, management, psychology, political science, and economics [3]. The authors' genders are extracted using third-party APIs.[5] We construct citation networks as follows. We group the articles in each field based on the publication year to build the network per field per year. Because the dataset for the field of Physics comprises of articles published before 2019, we build citation network for all the five fields from 2003 to 2018.

Fig. 8 illustrates the evolution of the proposed mutual information measure across the five fields examined in this study. The figure also displays the gendered citation gap (22) in relation to the mutual information measure. In four out of the five fields studied, a positive correlation is observed between the proposed measure I_α and the gendered citation gap, indicating that the proposed measure is consistent with existing metrics. Notably, the physics (APS) dataset is an exception to this trend.

Furthermore, the analytical DMPA model (Fig. 7) demonstrates that the gendered citation gap fails to exhibit a positive correlation with the parameter ρ_{att}, which is associated with the

[5]https://namsor.app/; https://www.gender-api.com/.

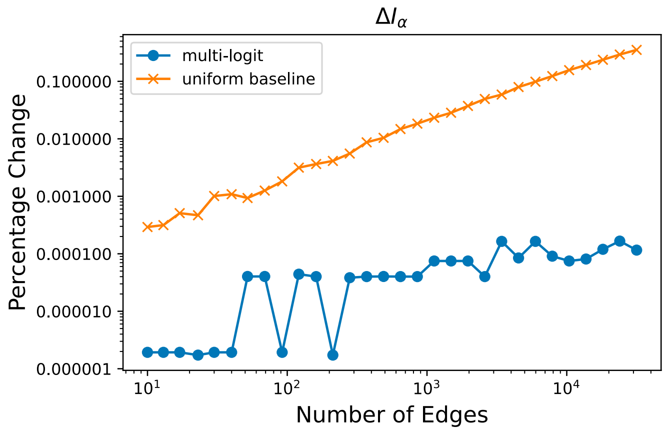

Fig. 9. This figure depicts the impact of edge addition on the mutual information measure within the 2012 physics citation network. Horizontal axis shows the number of edges added, with both axes presented on a logarithmic scale. Contrary to the baseline approach that uses a uniform distribution for edge addition, the results indicate that the optimized conditional logit model (Section IV) results in a less significant increase in the mutual information measure I_α. This suggests that the optimized model is effective in reducing the glass ceiling effect.

glass ceiling effect. In contrast, the proposed measure successfully captures the glass ceiling effect (Fig. 6), suggesting its superiority over existing metrics in quantifying this phenomenon.

Specifically, in political science, I_α shows a gradual decline over time, suggesting a decreasing focus on gender as a factor influencing citation patterns. Conversely, in physics and psychology, I_α consistently rises over the years, indicating an intensifying glass ceiling effect where gender increasingly influences citation dynamics. In management, the glass ceiling effect remains relatively stable, except for a notable increase in 2018, which may be attributed to incomplete citation data. These trends provide critical insights into the changing landscape of gender-based disparities in citation practices within various academic fields.

C. Optimization on Randomized Edge Addition

Previous numerical experiments on real-world citation networks show that mutual information measure I_α depicts the glass ceiling effect of a network over time. We implement Algorithm 1 in Section IV and show that the I_α can be optimized using randomized edge addition.

We use the citation network in physics from year 2012 as an example. As can be seen in Fig. 8, I_α is relatively high. We use a parameterized conditional logit model (24) to select edges to add to the network. $\theta_i, i = 1, \ldots, |\mathcal{E}|$ is initialized using a standard normal distribution. We run the algorithm for 10 000 iterations.

We add various numbers of edges to the original network using the conditional logit model to evaluate the optimization result. The baseline method adds the same amount of edges but uses a uniform distribution. Fig. 9 shows that both the attribute assortativity γ_{att} and the degree assortativity γ_{deg} increase more in the baseline method compared with the optimized conditional logit model. As such, by recommending links in a randomized fashion, optimization on mutual information could reduce inequality in a network while preserving the privacy of the network participants.

VI. Conclusion

This article quantifies the glass ceiling effect in networks using the mutual information measure (based on Shannon and more generally, Rényi entropy) between the conditional probability distributions of node attributes given the degree of adjacent nodes. Compared to existing measures, mutual information measure accounts for both demographic information and node degrees, making it a more comprehensive measure of node attribute information in networks. We illustrate the efficacy of the mutual information measure through various examples and simulations using an analytical network growth model (DMPA model). Our results indicate that although this measure aligns with existing metrics, it uniquely identifies the glass ceiling effect in scenarios where existing metrics fall short. Then, we apply the proposed measure to citation networks and demonstrate its ability to identify structural inequality and track the evolution of the glass ceiling effect across different research fields over time. Our results suggest strategies to support minority scientists and demonstrate the potential of mutual information to contribute to a better understanding of glass ceiling effects in networks.

References

[1] A. Clauset, S. Arbesman, and D. B. Larremore, "Systematic inequality and hierarchy in faculty hiring networks," *Sci. Adv.*, vol. 1, no. 1, 2015, Art. no. e1400005.
[2] M. E. Newman, "Coauthorship networks and patterns of scientific collaboration," *Proc. Nat. Acad. Sci.*, vol. 101, no. suppl_1, pp. 5200–5205, 2004.
[3] B. Nettasinghe, N. Alipourfard, V. Krishnamurthy, and K. Lerman, "Emergence of structural inequalities in scientific citation networks," 2021, *arXiv:2103.10944*.
[4] E. G. Teich et al., "Citation inequity and gendered citation practices in contemporary physics," *Nature Phys.*, vol. 18, no. 10, pp. 1161–1170, 2022.
[5] C. Ross, A. Gupta, N. Mehrabi, G. Muric, and K. Lerman, "The leaky pipeline in physics publishing," *Bull. Amer. Phys. Soc.*, vol. 66, no. 5, 2021.
[6] M. L. Dion, J. L. Sumner, and S. M. Mitchell, "Gendered citation patterns across political science and social science methodology fields," *Political Anal.*, vol. 26, no. 3, pp. 312–327, 2018.
[7] C. Avin, B. Keller, Z. Lotker, C. Mathieu, D. Peleg, and Y.-A. Pignolet, "Homophily and the glass ceiling effect in social networks," in *Proc. Conf. Innovations Theor. Comput. Sci.*, 2015, pp. 41–50.
[8] M. E. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, no. 20, 2002, Art. no. 208701.
[9] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Hoboken, NJ, USA: Wiley, 2005.
[10] R. V. Solé and S. Valverde, "Information theory of complex networks: On evolution and architectural constraints," in *Complex Networks*. New York, NY, USA: Springer-Verlag, 2004, pp. 189–207.
[11] J. Overgoor, A. Benson, and J. Ugander, "Choosing to grow a graph: Modeling network formation as discrete choice," in *Proc. World Wide Web Conf.*, 2019, pp. 1409–1420.
[12] J. Overgoor, G. Pakapol Supaniratisai, and J. Ugander, "Scaling choice models of relational social data," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2020, pp. 1990–1998.
[13] H. Gupta and M. A. Porter, "Mixed logit models and network formation," *J. Complex Netw.*, vol. 10, no. 6, 2022, Art. no. cnac045.
[14] E. Sadler, "Diffusion games," *Amer. Econ. Rev.*, vol. 110, no. 1, pp. 225–70, 2020.

[15] B. Nettasinghe, N. Alipourfard, S. Iota, V. Krishnamurthy, and K. Lerman, "Scale-free degree distributions, homophily and the glass ceiling effect in directed networks," *J. Complex Netw.*, vol. 10, no. 2, 2022, Art. no. cnac007.

[16] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguná, and D. Krioukov, "Popularity versus similarity in growing networks," *Nature*, vol. 489, no. 7417, pp. 537–540, 2012.

[17] A. R. Benson, D. F. Gleich, and J. Leskovec, "Higher-order organization of complex networks," *Science*, vol. 353, no. 6295, pp. 163–166, 2016.

[18] K.-k. Shang, B. Yang, J. M. Moore, Q. Ji, and M. Small, "Growing networks with communities: A distributive link model," *Chaos Interdisciplinary J. Nonlinear Sci.*, vol. 30, no. 4, 2020.

[19] M. Wiedermann, J. F. Donges, J. Kurths, and R. V. Donner, "Mapping and discrimination of networks in the complexity-entropy plane," *Phys. Rev. E*, vol. 96, no. 4, 2017, Art. no. 042304.

[20] F. Radicchi, D. Krioukov, H. Hartle, and G. Bianconi, "Classical information theory of networks," *J. Phys.: Complexity*, vol. 1, no. 2, 2020, Art. no. 025001.

[21] C. G. Freitas, A. L. Aquino, H. S. Ramos, A. C. Frery, and O. A. Rosso, "A detailed characterization of complex networks using information theory," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 16689.

[22] M. Piraveenan, M. Prokopenko, and A. Y. Zomaya, "Assortativeness and information in scale-free networks," *Eur. Phys. J. B*, vol. 67, no. 3, pp. 291–300, 2009.

[23] M. Prokopenko, F. Boschetti, and A. J. Ryan, "An information-theoretic primer on complexity, self-organization, and emergence," *Complexity*, vol. 15, no. 1, pp. 11–28, 2009.

[24] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6801–6810, Nov. 2014.

[25] S. Verdú, "α-mutual information," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Piscataway, NJ, USA: IEEE Press, 2015, pp. 1–6.

[26] M. Gjoka, B. Tillman, and A. Markopoulou, "Construction of simple graphs with a target joint degree matrix and beyond," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Piscataway, NJ, USA: IEEE Press, 2015, pp. 1553–1561.

[27] R. Cummings, V. Gupta, D. Kimpara, and J. Morgenstern, "On the compatibility of privacy and fairness," in *Proc. Adjunct Publication 27th Conf. User Model., Adaptation Personalization*, 2019, pp. 309–315.

[28] J. C. Spall, "Implementation of the simultaneous perturbation algorithm for stochastic optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 3, pp. 817–823, Jul. 1998.

[29] V. Krishnamurthy, *Partially Observed Markov Decision Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

Rui Luo received the B.S. degree from Tsinghua University, Beijing, China, and the Ph.D. degree from Cornell University, Ithaca, NY, USA.

He is currently an Assistant Professor with the Department of Systems Engineering, City University of Hong Kong, Kowloon, Hong Kong, SAR. His research interests include graph machine learning, network science, and complex systems.

Buddhika Nettasinghe received the M.A.Sc. degree from the University of British Columbia, Vancouver, BC, Canada, and the Ph.D. degree from Cornell University, Ithaca, NY, USA, both in electrical and computer engineering.

He is currently an Assistant Professor with the Department of Business Analytics, Tippie College of Business, University of Iowa, Iowa City, IA, USA. His research interests include complex systems, network science, and uncertainty quantification.

Vikram Krishnamurthy (Fellow, IEEE) received the Ph.D. degree from the Australian National University, Canberra, Australia, in 1992.

He is currently a Professor with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. From 2002 to 2016, he was a Professor and a Canada Research Chair with the University of British Columbia, Vancouver, BC, Canada. His research interests include statistical signal processing and stochastic control in social networks and adaptive sensing. He is an author of the books titled *Partially Observed Markov Decision Processes* and *Dynamics of Engineered Artificial Membranes and Biosensors* (Cambridge University Press, 2016 and 2018).

Prof. Krishnamurthy served as a Distinguished Lecturer for IEEE Signal Processing Society and the Editor-in-Chief of IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. In 2013, he was awarded an Honorary Doctorate from KTH (Royal Institute of Technology), Sweden.