Homophily and the Glass Ceiling Effect in Social Networks

Chen Avin* †[1], Barbara Keller‡[2], Zvi Lotker[1], Claire Mathieu[3], David Peleg[4] and Yvonne-Anne Pignolet[5]

[1]Ben Gurion University, Israel
[2]ETH Zürich, Switzerland
[3]École Normale Supérieure, France
[4]The Weizmann Institute, Israel
[5]ABB Corporate Research, Switzerland

February 3, 2015

Abstract

The glass ceiling effect has been defined in a recent US Federal Commission report as "the unseen, yet unbreakable barrier that keeps minorities and women from rising to the upper rungs of the corporate ladder, regardless of their qualifications or achievements". It is well documented that many societies and organizations exhibit a glass ceiling. In this paper we formally define and study the glass ceiling effect in social networks and propose a natural mathematical model, called the *biased preferential attachment* model, that partially explains the causes of the glass ceiling effect. This model consists of a network composed of two types of vertices, representing two sub-populations, and accommodates three well known social phenomena: (i) the "rich get richer" mechanism, (ii) a minority-majority partition, and (iii) homophily. We prove that our model exhibits a strong moment glass ceiling effect and that all three conditions are necessary, i.e., removing any one of them will prevent the appearance of a glass ceiling effect. Additionally, we present empirical evidence from a mentor-student network of researchers derived from the DBLP database that exhibits both a glass ceiling effect and the above three phenomena.

1 Introduction

Attaining *equality of opportunity* is a fundamental value in democratic societies, therefore existing inequalities present us with a major concern. A particularly sore example is that many highly-qualified women and members of minority groups are unable to realize their full potential in society (and specifically in the workforce) due to a phenomenon commonly referred to as the *glass ceiling*, a powerful visual image for an invisible barrier blocking women and minorities from advancing past middle management levels [20]. This concern was raised in a recent US Federal commission report [18]:

> *The "glass ceiling"... is the unseen, yet unbreakable barrier that keeps minorities and women from rising to the upper rungs of the corporate ladder, regardless of their qualifications or achievements.*

The existence of the glass ceiling effect is well documented [8, 16, 31]. In academia, for example, gender disparities have been observed in the number of professors [36], earnings [42, 13, 36]

*Supported in part by the Israel Science Foundation (grant 1549/13).
†Part of this work was done while the author was a long term visitor at ICERM, Brown University
‡Part of this work was done while the author was a visiting student at the Weizmann Institute.

funding [30] and patents [10]. A recent study [26] analyzed gender differences in research output, research impact and collaborations based on Thomson Reuters Web of Science databases. When prominent author positions were analyzed by sole authorship, first-authorship and last-authorship, it was discovered that papers with women in those leading roles were less frequently cited. The question we focus on in this article concerns the causes of this phenomenon. What are the invisible mechanisms that combine to create the glass ceiling effect, and in particular, what is the role of the social network in creating this effect? Many papers discuss possible causes of the glass ceiling effect and potential solutions to it, e.g., [9, 15, 24, 28], but to the best of our knowledge, the present work is the first attempt to define the glass ceiling mathematically, study it in the context of the social network structure, and to propose a *mathematical model* capturing this phenomenon.

In order to talk about the glass ceiling we have to agree on a measure of success in a social network. Following the traditional approach that sees network edges as the "social capital" of the network, we define successful members of a social network to be high degree vertices, namely, the vertices that maintain a large number of connections, corresponding to high influence. Based on this we propose formal definitions for glass ceiling effects as a first contribution. Note that it is not clear how to capture the nature of a delicate dynamic mechanism like the glass ceiling in a concise yet precise way. To represent the dynamic nature we examine sequences of networks and their behavior when the number of vertices grows.

Consider the following three well-accepted observations on human behavior related to forming networks, namely (i) the "rich get richer" mechanism, (ii) minority-majority partition (slower growth rate of the red group in the network), and (iii) homophily (affinity towards those similar to oneself). The main result of the paper is that under these three simple and standard assumptions the glass ceiling effect naturally arises in social networks. Let us first briefly describe these three social phenomena.

The "rich get richer" mechanism. This mechanism describes and explains the process of wealth concentration. It follows the basic idea that newly created wealth is distributed among members of society in proportion to the amount they have already amassed. In our setting, where the degree of the vertex captures its level of social wealth, this mechanism predicts that people may try to connect more often to people who already have many connections. In order to profit from their social wealth or because they are more visible in the network.

Minority-majority partition. Many social groups exhibit unequal proportions of men and women. Certain occupations, such as construction, law enforcement, politics and computer science, tend to have a higher proportion of men. For example, the ratio of women taking up studies in the computing discipline varies per year and region between 10% and 35% [3, 21, 40, 46]. Other professions, such as elementary school teaching, nursing, and office administration, are occupied by a higher proportion of women. In fact, it is difficult to find an occupation with a balanced ratio of genders (this also holds for many other social partitions, e.g., ones based on ethnicity or family background). This imbalance is the second phenomenon underlying our model.

Homophily. It is a well established social phenomenon that people tend to associate with others who are similar to themselves. Characteristics such as gender, ethnicity, age, class background and education influence the relationships among human beings [27] and similarities make communication and relationship formation easier.

Based on these phenomena we propose a model obtained by applying the classical preferential attachment model (see Barabasi and Albert [2]) to a bi-populated minority-majority network augmented with homophily. The resulting model is hereafter referred to as the *Biased Preferential Attachment* Model. We proof that networks generated by this model exhibit a glass ceiling structure.

As a running example serving to illustrate the issue, let us consider the social network of mentor-student relationships in academia. With time, new (male and female) PhD students

arrive and join the network. Upon arrival, each student needs to select a mentor. Over time, graduated students may become mentors themselves and some mentors become more successful than others (e.g., in terms of the number of students they advise). How can one determine that there is a glass ceiling effect in this network? And if such an effect exists, what are the roots of its emergence? Is it merely a result of the females being a minority in the network, or is it some sort of discriminatory process? To complement our theoretical analysis with data from real networks we study these questions on data from publications in computer science. Using the definitions and insights from the model we observe homophily and glass ceiling effects in a mentor-student network derived from this data.

Overview of Contributions

The paper's main contributions are the following.

(1) Formal definitions for the glass ceiling effect *in social networks* using graph sequences of growing size. The definitions capture the dynamic nature of the glass ceiling effect by measuring the decreasing fraction of minorties among higher degree nodes and the ratio of the second moment of the degree sequences of the minority and majority.

(2) A model for bi-populated social networks extending the classical preferential attachment model [2], and augment it by including two additional basic phenomena, namely, a minority-majority partition, and homophily.

(3) A rigorous analysis of this extended model to study its suitability as a possible mechanism for the emergence of a glass ceiling effect. The main tool used is the study of the degree distribution with Doob martingales. We also show that omitting any one of the three ingredients of our model prevents the occurrence of a glass ceiling effect.

(4) Empirical evidence for a mentor-student network exhibiting preferential attachment, minority-majority partition, homophily, and a glass ceiling effect.

Roadmap

The rest of the paper is organized as follows. In the next section we review related work, then in Section 3 we introduce the model and the formal definitions of the involved properties: glass ceiling, influence inequality and homophily tests. In Section 4 we state our two main theorems, and in Section 5 we provide empirical evidence for the existence of all our necessary ingredients and for the glass ceiling effect in a student-mentor network of researchers in computer science. We conclude with a discussion.

2 Related Work

Homophily in social networks. Different characteristics such as gender, ethnicities, age, class background and education influence the relationships human beings form with each other [27]. McPherson et al. [34] survey a variety of properties and how they lead to particular patterns in bonding. Gender-based homophily can already be observed in play patterns among children at school [32, 43]. Eder and Hallinan [12] discovered that young girls are more likely to resolve intransitivity by deleting friendship choices, while young boys are more likely to add them. Overall, children are significantly more likely to resolve intransitivity by deleting a cross-sex friendship than by adding another cross-sex friendship [47]. These results show that gender influences forming cliques and larger evolving network structures. These trends displaying homophily and gender differences in resolving problems in the structure of relationships mean that boys and girls gravitate towards different social circles. As adults, homophilic behavior persists, and men still tend to have networks that are more homophilic than women do. This

behavior is even more pronounced in areas where they form the majority and in relationships exchanging advice and based on respect, e.g., mentoring [22, 23, 5, 41]. A homophilic network evolution model was studied in [4]. In this model new nodes connect to other nodes in two phases. First they choose their neighbors with a bias towards their own type (the model allows a positive as well as a negative bias). In a second phase they choose their neighbors unbiased from the neighbors of their biased neighbors. The authors show, that the second phase overcomes the bias in the first phase and if the second phase is unbiased, the network ends up in an integrated state. They illustrate their model with data on citations in physics journals.

Gender disparity in science and technology. Gender disparities have been observed in the number of professors [13, 36], earnings [42], funding [30] and patenting [10]. A related aspect is the "productivity puzzle": men are more successful when it comes to number of publications and name position in the author list [48], for reasons yet unclear. Some conjectures raised involve (unknown) biased perceptions related to pregnancy/child care [6]. E.g., it was observed in [36] that science faculty members of both sexes exhibit unconscious biases against women. Simulations showed that even small male-female differences in work performance ratings can lead to substantially lower promotion rates for women, resulting in proportionately fewer women than men at the top levels of the organization [33]. Gender differences in research output, research impact and collaborations was analyzed in a study based on Thomson Reuters Web of Science databases [26]. It was not only revealed that papers with women in prominent author positions (sole authorship, first-authorship and last-authorship) were cited less frequently but the authors also found that age plays an important role in collaborations, authorship position and citations. Thus many of the trends observed therein might be explained by the under-representation of women among the elders of science. In other words, fixing the "leaky pipeline" [45] is key for a more equal gender distribution in science.

Minority of women in Computer Science. In the computing discipline, the ratio of women taking up studies varies by year and region between 10% and 35% [3, 21, 40, 46] (except in Malaysia, where women form a narrow majority [37]). This under-representation has been investigated [44, 19, 49] and remedial strategies have been propoesd [39, 17]. There is a positive feedback loop [25]: the lack of women leads to a strong male stereotype which drives away even more women. Thus the increase of the relative number of women in computer science is argued to be the best of the investigated strategies, up to a "critical mass" of women. However, as pointed out by Etzkowitz [14], even achieving a critical mass of 15% women might not guarantee that the effects of a critical mass come into play.

3 Model and Definitions

3.1 Biased preferential attachment model

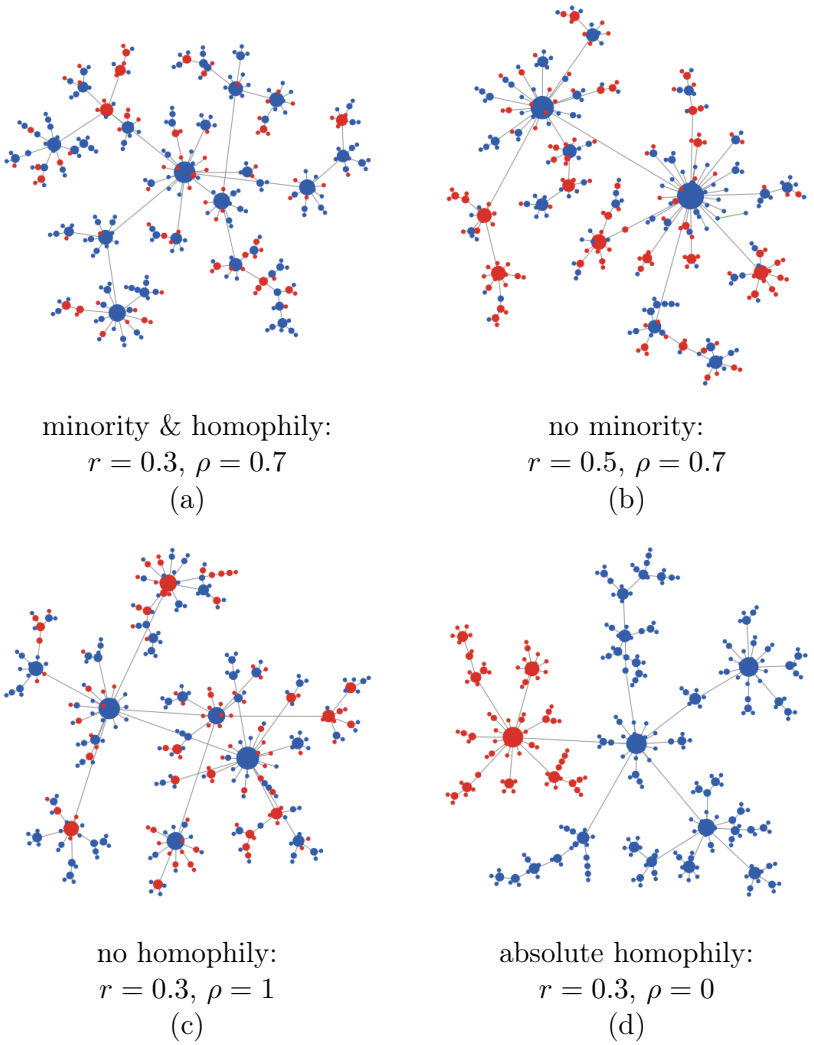
Our first contribution is in proposing a simple bi-populated preferential attachment model. In a gist, our model is obtained by applying the classical preferential attachment model (see Barabasi and Albert [2]) to a bi-populated minority-majority network augmented with homophily. The resulting model is hereafter referred to as the *Biased Preferential Attachment* Model. Formally, for $r \leq 1/2$ and $0 \leq \rho \leq 1$ let $G(n, r, \rho)$ be a variant of the preferential attachment model in which r represents the relative arrival rate of the red vertices (and hence the expected fraction of red vertices in the network converges to r as well, as the relative size of the initial population becomes smaller over time), and ρ represents the level to which homophily (incorporated by using rejection sampling) is expressed in the system: for $\rho = 1$ the system is uniform and exhibits no homophily, whereas for $\rho = 0$ the system is fully segregated, and all added edges connect vertex pairs of the same color.

Let us describe the model in more detail. Denote the social network at time t by $G_t = (V_t, E_t)$, where V_t and E_t, respectively, are the sets of vertices and edges in the network at time t, and let $\delta_t(v)$ denote the degree of vertex v at time t. The process starts with an arbitrary initial (connected) network G_0 in which each vertex has an arbitrary color, red or blue. (For simplicity we require that a minimal initial network consists of one blue and one red vertex connected by an edge, but this requirement can be removed if $\rho > 0$). This initial network evolves in time as follows. In every time step t a new vertex v enters the network. This vertex is red with probability r and blue with probability $1-r$. On arrival, the vertex v chooses an existing vertex $w \in V_t$ to attach to with probability p proportional to w's degree at time t, i.e., $\mathbb{P}[w \text{ is chosen}] = \delta_t(w) / \sum_{u \in V_t} \delta_t(u)$. Next, if w's color is the same as v's color, then an edge is inserted between v and w; if the colors differ, then the edge is inserted with probability ρ, and with probability $1-\rho$ the selection is *rejected*, and the process of choosing a neighbor for v is restarted. This process is repeated until some edge $\{v, w\}$ has been inserted. Thus in each time step, one new vertex and one new edge are added to the existing graph.

Figure 1 presents four examples of parameters for our model in the case of a 300-vertex bi-populated social network. First, Figure 1(a) provides an example for the *minority & homophily* case with $r = 0.3$ and $\rho = 0.7$ so the red vertices are a strict minority in the network and there is some homophily in the edge selection. The next three sub-figures present special cases. Figure 1(b) illustrates the *no minority* case (equal-size populations, i.e., $r = 0.5$) with homophily ($\rho = 0.7$). Figure 1(c) considers the *no homophily* case ($\rho = 1$) with minority ($r = 0.3$). The last extreme case, shown in Figure 1(d), is *absolute homophily*, where $\rho = 0$, but the red vertices are still in the minority ($r = 0.3$). This case results in *fully segregated* societies, namely, societies where members connect *only* to members of their own color. In this extreme case, the society in effect splits into two separate networks, one for each of the two populations (except for the single edge connecting the initial red and blue vertices).

Consider as an example for our model the social network of mentor-student relationships in academia. With time, new PhD students arrive, but for some fields female students arrive at a lower rate than male students. Upon arrival, each student needs to select exactly one mentor, where the selection process is governed by the mechanisms of preferential attachment and homophily. Namely, initially the student selects the mentor according to the rules of preferential attachment and then homophily takes its role, rejecting the selection with some probability if their gender is different enforcing a re-selection. Over time, graduated students may become mentors and some mentors become more successful than others (in terms of the number of students they advise). A glass ceiling effect can be observed in this network if, after a long enough time interval, the fraction of females among the most successful mentors tends to zero.

We would like to emphasize that the homophily effect that we look at is quite minor and "seemingly harmless", in two ways. First, it is "symmetric", i.e., it applies both to male students with respect to female mentors and to female students with respect to male mentors. Second, it does not adversely affect the student, in the sense that the student always gets admitted in our model. The only tiny (but ominous) sign for the potential dangers of this homophilic effect is that it does affect the professor: a male professor who rejects (or is rejected by) some fraction of the female candidates risks little, whereas a female professor who rejects (or is rejected by) some fraction of the male candidates will eventually have fewer students overall, since most of the applicants are male. In fact, as we show later on, this homophily-based consequence will only impact her if her future potential students use preferential attachment to select their mentors.



minority & homophily:
$r = 0.3$, $\rho = 0.7$
(a)

no minority:
$r = 0.5$, $\rho = 0.7$
(b)

no homophily:
$r = 0.3$, $\rho = 1$
(c)

absolute homophily:
$r = 0.3$, $\rho = 0$
(d)

Figure 1: Examples of the Biased Preferential Attachment (BPA) model with various parameters. All examples depict a 300-vertex bi-populated network generated by our BPA model starting from a single edge connecting a blue and a red vertex (with vertex size proportional to its degree). (a) Minority & homophily: $r = 0.3$ (resulting in about 30% red vertices) and $\rho = 0.7$ (meaning that a new edge that connects red-blue vertices (i.e., a "mixed" edge) is accepted with probability 0.7 and otherwise rejected and sampled again, according to Preferential Attachment). (b) No minority & homophily: $r = 0.5$ and $\rho = 0.7$. (c) Minority & no-homophily: $r = 0.3$ and $\rho = 1$. (d) Minority & absolute homophily: $r = 0.3$ and $\rho = 0$ (indicating complete homophily in edge selection which results in two separate networks, one for the red vertices and the second for the blue vertices, plus a single initial connecting edge).

3.2 Influence inequality and glass ceiling

Our second contribution is to propose formal definitions of the glass ceiling effect in social networks. Consider a *bi-populated* network $G(n)$ consisting of m edges and n nodes of two types, the group $\mathtt{R}$ and the group $\mathtt{B}$. We assume that the network size n tends to infinity with time. Let $n(\mathtt{R})$ and $n(\mathtt{B})$, respectively, denote the number of red and blue nodes, where $n(\mathtt{R}) + n(\mathtt{B}) = n$. The red nodes are assumed to be a minority in the social network, i.e., denoting the percentage of red nodes in the network by r, we assume $0 \leq r < \frac{1}{2}$. For a node v in G, let $\delta(v)$ denote its degree. Let $d(\mathtt{R})$ and $d(\mathtt{B})$ denote the sum of degrees of the red and blue nodes, respectively, where $d(\mathtt{R}) + d(\mathtt{B}) = 2m$. As the degree of a node corresponds to its

power, the sum of the degrees of a certain kind of nodes represents the power of this group in the network. Let $\text{top}_k(\mathtt{R})$ (respectively, $\text{top}_k(\mathtt{B})$) denote the number of red (resp., blue) nodes that have degree *at least* k in G. When $G(n)$ is a random graph, we replace variables by their expectations in the definitions below, e.g., we use $\mathbb{E}[n(\mathtt{R})]$, $\mathbb{E}[d(\mathtt{R})]$, and $\mathbb{E}[\text{top}_k(\mathtt{R})]$. Next we provide formal definitions for the social phenomena discussed in the introduction. *Influence inequality* for the minority is defined in the following way.

Definition 1 (Influence inequality)**.** *A graph sequence $G(n)$ exhibits a* influence inequality *effect for the red nodes if the average power of a red node is lower than that of a blue (or a random) node, i.e., there exists a constant $c < 1$ such that*

$$\lim_{n\to\infty} \frac{\frac{1}{n(\mathtt{R})}\sum_{v\in\mathtt{R}}\delta(v)}{\frac{1}{n(\mathtt{B})}\sum_{v\in\mathtt{B}}\delta(v)} = \frac{d(\mathtt{R})/n(\mathtt{R})}{d(\mathtt{B})/n(\mathtt{B})} \leq c\ . \tag{1}$$

The definition of the glass ceiling effect is more complex. We interpret the most powerful positions as those held by the highest degree nodes, and offer two alternative definitions. The first tries to capture the informal, "dictionary" definition, which describes a decreasing fraction of women among higher degree nodes, i.e., in the *tail* of the graph degree sequence. Formally:

Definition 2 (Tail glass ceiling)**.** *A graph sequence $G(n)$ exhibits a* tail glass ceiling *effect for the red nodes if there exists an increasing function $k(n)$ (for short k) such that $\lim_{n\to\infty}\text{top}_k(\mathtt{B}) = \infty$ and*

$$\lim_{n\to\infty}\frac{\text{top}_k(\mathtt{R})}{\text{top}_k(\mathtt{B})} \ = \ 0\ .$$

The second definition considers a more traditional, distribution-oriented measure, the second moment of the two degree sequences. Formally:

Definition 3 (Moment glass ceiling)**.** *A graph sequence $G(n)$ exhibits a* moment glass ceiling *g for the red nodes where*

$$g \ = \ \lim_{n\to\infty}\frac{\frac{1}{n(\mathtt{R})}\sum_{v\in\mathtt{R}}\delta(v)^2}{\frac{1}{n(\mathtt{B})}\sum_{v\in\mathtt{B}}\delta(v)^2}\ .$$

When $g = 0$, we say that $G(n)$ has a *strong* glass ceiling effect. The intuition behind this definition is that a larger second moment (and assuming a similar average degree, i.e., no influence inequality) will result in a larger variance and therefore a significantly larger number of high degree nodes. As we show later, the above two definitions for the glass ceiling are independent, in the sense that neither of the effects implies the other.

Note that these definitions are very general and do not rely on any assumptions of the degree distribution. In particular it is not necessary for networks that exhibit a glass ceiling effect to follow a power law degree distribution.

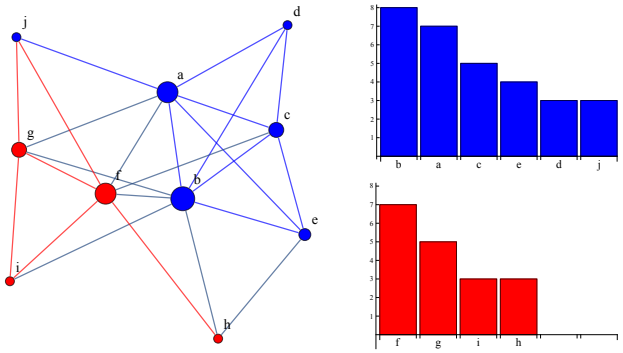
Testing for *homophily* in a bi-populated network is based on checking whether the number of *mixed* (i.e., red-blue) edges is significantly lower than to be expected if neighbors were to be picked randomly and independently of their color. Formally:

Definition 4 (Homophily Test)**.** *[11] A bi-populated social network exhibits homophily if the fraction of mixed edges is significantly less than $2r(1-r)$.*

The above definition implicitly assumes that there is power equality between the colors and therefore is not always accurate. A more careful test should take the average degree of each gender into account.

Definition 5 (Normalized Homophily Test)**.** *A bi-populated social network exhibits homophily if the fraction of mixed edges is significantly less than $2\frac{d(\mathtt{R})}{2m}\left(1-\frac{d(\mathtt{R})}{2m}\right)$.*

An illustration of these definitions can be found in Figure 2.



(a) (b)

Figure 2: (a) An example bi-populated social network with blue and red populations of 6 and 4 vertices respectively. (b) The degree sequences of both populations (i.e., the sequence specifying for each vertex its degree in the network). Considering the *tail glass ceiling* definition, there are four blue vertices of degree greater or equal to 4, but only two such red vertices so $\text{top}_4(\texttt{R})/\text{top}_4(\texttt{B}) = 1/2$. For the *moment glass ceiling* definition, the second moment for the blue vertices is $\frac{1}{6}(8^2 + 7^2 + 5^2 + 4^2 + 3^2 + 3^2) = 28.6$, while for the red vertices it is $\frac{1}{4}(7^2 + 5^2 + 3^2 + 3^2) = 23$ and the ratio is $23/28.6$. To exhibit a glass ceiling, these ratios should converge to zero as the network size increases. The average degree of the blue vertices in the network is 5 while the average for the red vertices is 4.5. It is possible that this numbers remain (almost) the same while the network size increases and the network exhibits a glass ceiling. Regarding homophily, in a random network with the same population, i.e., 60% blue vertices and 40% red vertices, one expects to find 36% blue-blue edges, 16% red-red edges and 48% mixed edge. If we take the degree sequences into account we would expect to see 46.8% mixed edge. In the above example network we observe only about 33% mixed edges, which indicates the effect of homophily.

4 Theoretical Results

4.1 Influence inequality and glass ceiling

Our main theoretical result (Thm. 4.1) is that in the biased preferential attachment model, $G(n, r, \rho)$, the glass ceiling effect emerges naturally. Additionally, this process generates a *influence inequality*, an independent property that is weaker than the glass ceiling effect. Influence inequality describes the situation where the average degree of the minority is lower than that of the majority (although their members possess the same qualifications). Moreover, we also show (Thm. 4.2) that all three ingredients (unequal entry rate, homophily, preferential attachment) are necessary to generate what we call a *strong* glass ceiling effect, i.e., removing any one of them will prevent the appearance of a glass ceiling effect. One may suspect that the glass ceiling effect is in fact a byproduct of influence inequality or unequal qualifications; we show that this is not the case. Minorities can have a smaller average degree without suffering from a glass ceiling effect. We also note that our results are independent of the starting condition. Even if the network initially consisted entirely of vertices of one color, if a majority of the vertices being added are of the opposite color, then eventually the vertices that rise to the highest positions will be of the new color.

Theorem 4.1. *Let $0 < r < \frac{1}{2}$ and $0 < \rho < 1$. For $G(n, r, \rho)$ produced by the Biased Preferential Attachment Model the following holds:*

1. *$G(n, r, \rho)$ exhibits influence inequality, and*
2. *$G(n, r, \rho)$ exhibits both a tail and a strong glass ceiling effects.*

Moreover, all three ingredients are necessary to generate a *strong* glass ceiling effect.

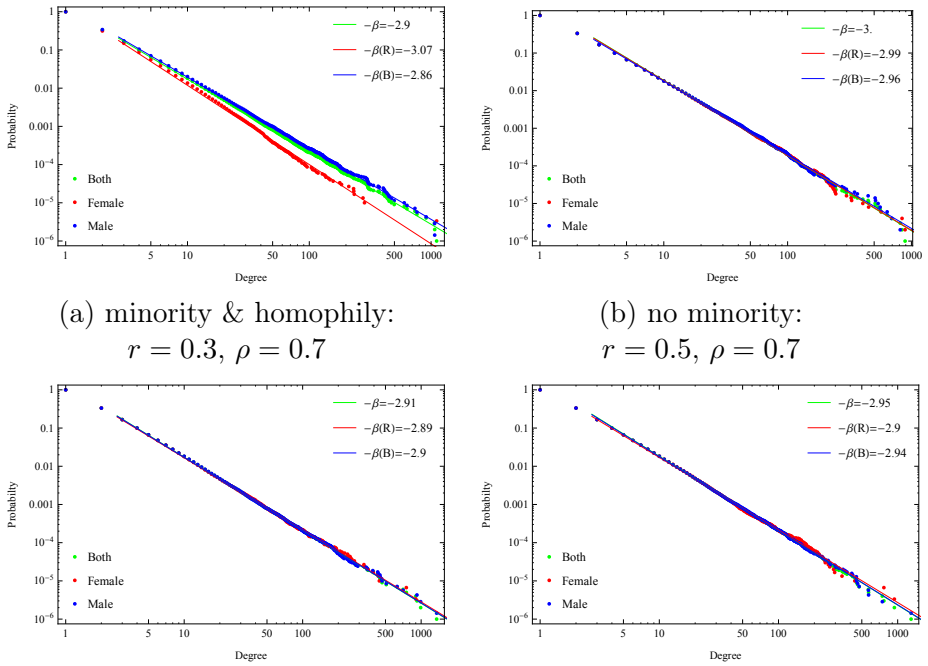
Theorem 4.2. *1. $G(n, r, \rho)$ will not exhibit a glass ceiling effect in the following cases:*

(a) If the rate $r = \frac{1}{2}$ (no minority).

(b) If $\rho = 1$ (no homophily)

(c) If $\rho = 0$ (no heterophily).

2. $G(n, r, \rho)$ will not exhibit a strong *glass ceiling effect if the attachment process is uniform rather than preferential, i.e., a new vertex at time t selects an existing vertex to attach to uniformly at random from all vertices present at time $t - 1$ (and for any value of r and ρ).*



(a) minority & homophily: $r = 0.3$, $\rho = 0.7$

(b) no minority: $r = 0.5$, $\rho = 0.7$

(c) no homophily: $r = 0.3$, $\rho = 1$

(d) absolute homophily: $r = 0.3$, $\rho = 0$

Figure 3: Graphical illustrations of our formal claims concerning the glass ceiling effect in the Biased Preferential Attachment model. Each figure presents the degree distribution (on a log-log scale) of the red and blue populations from a 1,000,000-vertex network generated by the BPA model with the same parameters as the corresponding figure in Figure 1. (a) Minority & homophily: $r = 0.3$ and $\rho = 0.7$. Both populations exhibit a power-law degree distribution but with different exponents. Since $\beta(\mathtt{R}) > \beta(\mathtt{B})$, there is a glass ceiling effect for the red vertices. (The "noise" on the right-hand side of the graph stems from the fact that there are much fewer samples at the high-end of the range.) (b) No minority & homophily: $r = 0.5$ and $\rho = 0.7$. Both populations exhibit a power-law degree distribution with $\beta = 3$, which indicates no glass ceiling effect. (c) Minority & no-homophily: $r = 0.3$ and $\rho = 1$. Again, the distributions indicate no glass ceiling effect. (d) Minority & absolute homophily: $r = 0.3$ and $\rho = 0$. Again, the distributions indicate no glass ceiling effect.

Let us graphically illustrate the above results. Figure 3 presents the degree distributions of both the red and blue populations (as well as of the entire population) for four 1,000,000-vertex networks with parameters identical to the examples in Figure 1. The plots clearly show (and we prove this formally) that in all cases the degree distribution of both populations

follows a power-law. (A subset W of vertices in a given network obeys a power-law degree distribution if the fraction $P(k)$ of vertices of degree k in W behaves for large values of k as $P(k) \sim k^{-\beta}$ for parameter β.) All figures present (in log-log scale) the cumulative degree distributions, so a power-law corresponds to a straight line (we present the samples together with the best-fit line). Theorem 4.1 corresponds to Figure 3(a) with the *minority & homophily* settings of $0 < r < \frac{1}{2}$ and $0 < \rho < 1$. In this case (and only in this case), the power-law exponents of the red and blue populations, $\beta(\mathtt{R})$ and $\beta(\mathtt{B})$ respectively, are *different*, where $\beta(\mathtt{R}) > \beta(\mathtt{B})$; we prove that this will eventually lead to both tail and strong glass ceiling effect for the red vertices. Theorem 4.2 corresponds to Figures 3(b) and 3(c). The figures show that in the case of *no minority* (i.e., $r = 0.5$) or *no homophily* (i.e., $\rho = 1$), both $\beta(\mathtt{R})$ and $\beta(\mathtt{B})$ are the same (in particular they are equal to 3 as in the classical Preferential Attachment model), and therefore there will be no glass ceiling effect. Figure 3(d) considers the last extreme case of *absolute homophily*. Perhaps surprisingly, in this case a glass ceiling effect also does not occur, as each sub-population forms an absolute majority in its own network (see again Figure 1(d)). The case of no preferential attachment (which does not lead to a glass ceiling) is more delicate and presented in Section 4.5.

Proof Overview of Theorem 4.1. The basic idea behind the proof of Theorem 4.1 is to show that both populations in $G(n, r, \rho)$ have a power law degree distribution but with different exponents. Once this is established, it is simple to derive the glass ceiling effect for the population with a higher exponent in the degree distribution. To study the degree distribution of the red (and similarly the blue) population, we first define α_t to be the random variable that is equal to the ratio of the total degree of the red nodes (i.e., the sum of degrees of all red nodes) divided by the total degree (i.e., twice the number of edges). We show that the expected value of α_t converges to a fixed ratio independently of how the network started. The proof of this part is based on tools from dynamic systems. Basically, we show that there is only one fixed point for our system. However, determining the expectation of α_t is not sufficient for analyzing the degree distribution, and it is also necessary to bound the rate of convergence and the concentration of α_t around its expectation. We used Doob martingales for this part. Using the high concentration of the total degree, we were able to adapt standard techniques to prove the power law degree distribution. Next we give an overview of the proofs and the helping lemmas.

4.2 Proof of Theorem 4.1 Part 1

An urn process. The biased preferential attachment model $G(n, r, \rho)$ process can also be interpreted as a Polya's urn process, where each edge in the graph corresponds to two balls, one for each endpoint, and the balls are colored by the color of the corresponding vertices. When a new (red or blue) ball y arrives, we choose an existing ball c from the urn uniformly at random; if c is of the same color as y, then we add to the urn both y and another ball of the same color as c; otherwise (i.e., if c is of a different color), with probability ρ we still add to the urn both y and another ball of the same color as c, and with probability $1-\rho$ we reject the choice of c and repeat choosing an existing ball c' from the urn uniformly at random. To analyze influence inequality, there is no need to keep track of the degrees of individual vertices; the sum of the degrees of all vertices of $\mathtt{R}$ is exactly the number of red balls in the urn.

Denote by $u_t(\mathtt{R})$ (respectively, $u_t(\mathtt{B})$) the number of red (resp., blue) balls present in the urn at time $t \geq 0$. Altogether, the number of balls at time t is $u_t = u_t(\mathtt{R}) + u_t(\mathtt{B})$. Initially, the system contains u_0 balls. Noting that exactly two balls join the system in each time step, we have $u_t = u_0 + 2t$. Note that while $u_t(\mathtt{R})$ and $u_t(\mathtt{B})$ are random variables, u_t is not. Denote by α_t the random variable equal to $u_t(\mathtt{R})/u_t$, the fraction of red balls in the system at time t.

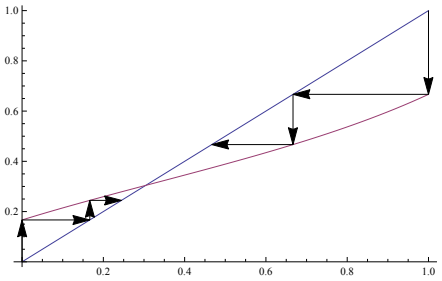


Figure 4: Function $F(x)$ with the parameters $r = 1/3$ and $\rho = 1/2$, with its only fix point in the interval $[0, 1]$ and $y = x$. The arrows represent the iterations of applying $F(x)$ repetitively onto itself, converging to its fix point.

Convergence of expectations. We first claim that the process of biased preferential attachment converges to a ratio of α red balls in the system. More formally, we claim that regardless of the starting condition, there exists a limit

$$\alpha \;=\; \lim_{n\to\infty} \mathbb{E}[\alpha_t] \;.$$

We prove our claim step by step and start with presenting a function $F(x)$ describing the expected percentage growth of red balls from time step t to time step $t+1$. $F(\alpha_t)$ is the expected number of red balls at time $t+1$ given the ratio of red balls at time t.

Lemma 4.3. $\mathbb{E}[\alpha_{t+1}|\alpha_t] \;=\; \alpha_t + \dfrac{F(\alpha_t) - \alpha_t}{t+1}$, *where*

$$F(x) \;=\; \left(1 - (1-r)\frac{(1-x)}{1-x(1-\rho)} + r\frac{x}{1-(1-x)(1-\rho)}\right)/2.$$

Proof. We start from an arbitrary ratio $\alpha_0 = u_0(\mathtt{R})/u_0$. Observe that given that the new vertex is blue, the probability p that it attaches to a blue vertex satisfies $p = (1-\alpha_t) + \alpha_t(1-\rho)p$, hence $p = (1-\alpha_t)/(1-\alpha_t(1-\rho))$. Given that the new vertex is red, the probability p' that it attaches to a red vertex satisfies $p' = \alpha_t + (1-\alpha_t)(1-\rho)p'$, hence $p' = \alpha_t/(1-(1-\alpha_t)(1-\rho))$. We know that in each step the sum of the degrees increases by 2 in total so $u_{t+1} = u_t + 2$ and if X_t is the random variable that denotes the number of new red balls at time t, we obtain:

$$X_{t+1} = \begin{cases} 0 & \text{with probability } (1-r)\frac{(1-\alpha_t)}{1-\alpha_t(1-\rho)}, \text{(a blue ball entered and chose a blue ball)} \\ 2 & \text{with probability } r\frac{\alpha_t}{1-(1-\alpha_t)(1-\rho)}, \text{(a red ball entered and chose a red ball)} \\ 1 & \text{with the remaining probability, (a blue ball chose a red ball or vice versa)} \end{cases}$$

and we have $u_t(\mathtt{R}) = \sum_0^t X_i$. We now define:

$$\mathbb{E}[u_{t+1}(\mathtt{R}) - u_t(\mathtt{R})|\alpha_t] = \left(1 - (1-r)\frac{(1-\alpha_t)}{1-\alpha_t(1-\rho)} + r\frac{\alpha_t}{1-(1-\alpha_t)(1-\rho)}\right) = 2F(\alpha_t).$$

Substituting $u_{t+1}(\mathtt{R}) = 2(t+1)\alpha_{t+1}$ and $u_t(\mathtt{R}) = 2t\alpha_t$ and rewriting yields the Lemma. □

We now have a function for the expected value of α_{t+1} given α_t. To prove that a_t actually converges to α we have to analyze the function $F(x)$ in more detail. We prove the following properties of this function:

Lemma 4.4. *1. $F(x)$ is monotonically increasing.*

2. $F(x)$ has exactly one fixed point, denoted α^*, in $[0,1]$.

3. The image of the unit interval by $F(x)$ is contained in the unit interval: $F([0,1]) = \left[\frac{r}{2}, \frac{1+r}{2}\right] \subset [0,1]$

4. If $x < \alpha^*$ then $x < F(x) < \alpha^*$ and if $x > \alpha^*$ then $x > F(x) > \alpha^*$.

5. $\alpha^* < r$.

Proof. With a little bit of algebra, we can, and for some reason we prefer, to rewrite $F(x)$ as

$$F(x) \;=\; \frac{1}{2}\left(r + \frac{rx}{x+(1-x)\rho} + (1-r)\frac{x\rho}{x\rho+(1-x)}\right).$$

For the first property, using simple algebra we compute

$$\frac{\partial F(x)}{\partial x} = \frac{1}{2}\left(\frac{\rho-\rho r}{(1+(\rho-1)x)^2} + \frac{\rho r}{(\rho+x-\rho x)^2}\right) > 0$$

for each $x, r, \rho \in [0,1]$.

For the second property, we define the function $G(x) = F(x) - x$. The roots of $G(x)$ correspond to the fixpoints of $F(x)$ so it is enough to show that $G(x)$ has exactly one real root in the interval $[0,1]$. Using simple algebra it follows that $\frac{\partial G(x)}{\partial x} > 0$ for each $x, r, \rho \in [0,1]$. Setting $G(x) = 0$ we get the following equation:

$$(2-4\rho+2\rho^2)x^3 + (5\rho-2-3\rho^2-2r+2r\rho)x^2 + (2r-2\rho-2r\rho+\rho^2)x + r\rho \;=\; 0. \quad (2)$$

We observe that $G(x) = -\infty$ when $x \to -\infty$ for each $\rho \in [0,1)$ and that $G(0) = r\rho \geq 0$. Which induces that there are 1 or 3 roots in the interval $(-\infty, 0)$. Observing that $G(x) = \infty$ when $x \to \infty$ and evaluating $G(1) = rp - p \leq 0$ for each $\rho, r \in [0,1]$ we see that there are 1 or 3 roots in the interval $(1, \infty)$. Knowing that $G(x)$ has exactly 3 roots concludes the claim that G has exactly one root in $[0,1]$ which leads to the conclusion that the function $F(x)$ has exactly one fixed point in $[0,1]$.

The third property follows from the fact that the function $F(x)$ is strictly monotonically increasing and by evaluating the function $F(x)$ for the two extreme values $x = 0$, and $x = 1$.

The fourth property follows from the fact that the function is strictly monotonically increasing, that there is only one fix point and that $F(x)$ maps $[0,1]$ inside $[0,1]$.

Finally, to show that $\alpha^* < r$, since we know that $F(x) - x$ is positive for $x < \alpha^*$ and negative for $x > \alpha^*$, it suffices to show that $F(r) < r$. This is equivalent to

$$r + \frac{r^2}{r+(1-r)\rho} + (1-r)\frac{r\rho}{r\rho+(1-r)} < 2r,$$

which is true for all $r < 1/2$. □

Now assume $\alpha_t < \alpha^*$. By Lemma 4.4, $\alpha_t < F(\alpha_t) < \alpha^*$, so by Lemma 4.3 we obtain $\alpha_t < \mathbb{E}[\alpha_{t+1}|\alpha_t] < \alpha^*$.

With that we have shown that the expected value of α_t does converge to the fix point α^* of $F(x)$. Figure 4 shows an instance of $F(x)$ with the parameters $r = 1/3$ and $\rho = 1/2$. You can see its only fix point in the interval $[0,1]$ on the intersection of $F(x)$ with the line $y = x$. If the function $F(x)$ is applied repetitively it will converge to its fix point, no matter if the initial value was larger or smaller than the fix point. We still need to bound the rate of convergence of $F(x)$ and show:

Lemma 4.5. $|\alpha^* - \mathbb{E}[\alpha_t]| = O(1/\sqrt[3]{t})$.

Proof. Assume that $\alpha_t < \alpha^*$ (the other case is similar).

$$\mathbb{E}[\alpha_{t+1}] = \mathbb{E}[\alpha_t] + \frac{\mathbb{E}[F(\alpha_t) - \alpha_t]}{t+1} \tag{3}$$

Let $\Delta = \frac{\alpha^* - r/3}{\alpha^*}$ and the line $L(x) = \Delta \cdot x + r/3$. Note that L shares the point $(\alpha^*, F(\alpha^*))$ with $F(x)$ but is strictly below $F(x)$ in the range $[0, \alpha^*]$ it (See Lemma 4.20). Thus

$$\mathbb{E}[F(\alpha_t] \geq \mathbb{E}[\alpha^* - \epsilon_t \Delta],$$

where $\epsilon_t = \alpha^* - \alpha_t$.

Substituting into Equation (3), we get

$$\begin{aligned}\mathbb{E}[\alpha_{t+1}] &\geq \mathbb{E}[\alpha_t] + \frac{\mathbb{E}[\alpha^* - \epsilon_t\Delta - \alpha_t]}{t+1} \\ &= \mathbb{E}[\alpha_t] + \frac{\mathbb{E}[\epsilon_t](1-\Delta)}{t+1} = \alpha^* - \mathbb{E}[\epsilon_t] + \frac{\mathbb{E}[\epsilon_t](1-\Delta)}{t+1},\end{aligned}$$

so the expected error at time $t+1$ is

$$\mathbb{E}[\epsilon_{t+1}] \leq \mathbb{E}[\epsilon_t](1 - \frac{1-\Delta}{t+1}).$$

Solving for $\mathbb{E}[\epsilon_t]$ we have

$$\mathbb{E}[\epsilon_t] = \epsilon_0(1 - \frac{1-\Delta}{2})(1 - \frac{1-\Delta}{3}) \cdots (1 - \frac{1-\Delta}{t}) = \epsilon_0 \exp(-\sum_{i=1}^{t} \frac{1-\Delta}{i}) = O(\frac{\epsilon_0}{t^{1-\Delta}}),$$

Note that since $r/2 \leq \alpha^* < r$ we have $\Delta < 2/3$, and so $\mathbb{E}[\epsilon_t] = O(1/\sqrt[3]{t})$.

□

With the previous steps we have proven both, the convergence of α to α^* as well as we have bound the convergence rate. We now investigate the fix point α^*.

Theorem 4.6. *For any initial configuration, as t goes to infinity, the expected fraction of red balls in the urn, $\mathbb{E}(\alpha_t)$, converges to the unique α^* in $[0, 1]$ satisfying the equation*

$$2\alpha^* \;=\; 1 - (1-r)\frac{(1-\alpha^*)}{1-\alpha^*(1-\rho)} + r\frac{\alpha^*}{1-(1-\alpha^*)(1-\rho)}. \tag{4}$$

Hence the limit α^* is the solution of the cubic equation Eq. (5).

$$(4\rho - 2\rho^2 - 2)\alpha^3 + (2 + 3\rho^2 - 5\rho + 2r - 2r\rho)\alpha^2 + (2\rho - 2r + 2r\rho - \rho^2)\alpha - r\rho = 0$$

Note that this limit is independent of the initial values u_0 and α_0 of the system. Having shown the independence of the fix point α^* of the initial configuration of the urn and the convergence to it, we can now claim:

Corollary 4.7. *Let $0 < \rho < 1$, $0 < r < 1/2$. Then $G(n, r, \rho)$ has a influence inequality effect.*

Proof. The expected degree of a red node tends to $2\alpha^*/r$, which is strictly less than 2, the expected degree of a random node, because of Lemma 4.4 Part 5. □

Thus we have proven the influence inequality part of Theorem 4.1.

4.3 Proof of Theorem 4.1 Part 2

Concentration. To prove the glass ceiling effect we first bound the degree distribution. To do this we need to bound the rate by which $u_t(\mathtt{R})$ converge to $\alpha \cdot t$. Let $X_i \in \{0, 1, 2\}$ be the number of new red balls in the system at time i. Note that $u_t(\mathtt{R}) = \sum_0^t X_i$. Let

$$\bar{X}_i = (X_1, X_2, \ldots, X_i)$$

be a tuple that captures all random variables $X_1, X_2, \ldots, X_i$ and let

$$\Psi_i = \mathbb{E}_{X_{i+1}, X_{i+2}, \ldots, X_t}\Big[\sum_{j=0}^{t} X_j | \bar{X}_i\Big].$$

Observe that $(\Psi_i)_i$ is a Doob Martingale [35], and note that $\Psi_0 = \mathbb{E}\Big[\sum_{i=0}^t X_i\Big] = \mathbb{E}\Big[u_t(\mathtt{R})\Big]$.

Theorem 4.8 (Azuma's inequality [1])**.** *Let Ψ_t be a martingale such that for all i, almost surely $|\Psi_i - \Psi_{i-1}| < c_i$. Then for all positive t and all positive reals x,*

$$\Pr(\Psi_t - \Psi_0 \geq x) \leq exp\left(\frac{-x^2}{2\sum_i c_i^2}\right).$$

Lemma 4.9. *Let $C_i = |\Psi_i - \Psi_{i-1}|$. Then $C_i = O(\sqrt{t/i})$.*

Proof. Observe that for $c = 0, 1, 2$,

$$\Psi_i - \Psi_{i-1} = \mathbb{E}_{X_{i+1}, \ldots, X_t}\left[\left(\mathbb{E}(\sum_{j=i}^{t} X_j | X_i = c) - \mathbb{E}_{X_i}(\sum_{j=i}^{t} X_j)\right) | \bar{X}_i\right].$$

To bound $C_i = |\Psi_i - \Psi_{i-1}|$, since each additional ball creates an *independent* effect on its descendents (the red balls that are connected to it), we have: $\mathbb{E}(C_i) \leq 2z_i$, where

$$z_i = \mathbb{E}\Big[a_{i,t} \mid b_i\Big]$$

where $a_{i,t}$ = number of additional red balls at times $[i, t]$, and b_i = one additional red ball at time i. We have the recurrence: $z_t = 1$, and for $i < t$

$$z_i = 1 + \frac{\gamma}{2(i+1)} z_{i+1} + \frac{\gamma}{2(i+2)} z_{i+2} + \cdots + \frac{\gamma}{2t} z_t$$

where $\gamma/(2i)$ is the probability of selecting a particular marked red ball at time i: we always have $\gamma \leq 1$, and γ depends on the homophily parameter ρ.

We apply some algebraic changes and let $y_i = z_i/2i$. It is easy to see that the recurrence becomes $y_t = 1/(2t)$, and $y_i = (1 + (2+\gamma)/(2i))y_{i+1}$. Solving the recurrence yields

$$y_i = \left(1 + \frac{2+\gamma}{2i}\right)\left(1 + \frac{2+\gamma}{2(i+1)}\right)\cdots\frac{1}{2t} = O\left(\frac{1}{2t}\exp\left(\sum_{j=i}^{t}\frac{2+\gamma}{2j}\right)\right) = O\left(\frac{1}{2t}\left(\frac{t}{i}\right)^{\frac{2+\gamma}{2}}\right). \tag{5}$$

Since $\gamma \leq 1$, we obtain $y_i = O((1/i)\sqrt{t/i})$ and $|C_i| = O(z_i) = O(\sqrt{t/i})$. □

Recall that $\alpha_t = u_t(\mathtt{R})/(2t)$. By Theorem 4.8 and Lemma 4.9 then

Lemma 4.10. $\Pr\Big[|u_t(\mathtt{R}) - 2t\mathbb{E}(\alpha_t)| > O(2\sqrt{t}\log t)\Big] \leq \frac{1}{t^4}$.

Proof. By Lemma 4.9

$$\sum_{i=1}^{t} C_i^2 \leq O(\sum_{i=1}^{t} \frac{t}{i}) \leq ct \log t.$$

We now use Theorem 4.8 for $x = \Theta(2\sqrt{ct \log t})$, note that $\Psi_t = u_t(\mathtt{R})$ and $\Psi_0 = \mathbb{E}[u_t(\mathtt{R})]$, and obtain:

$$\Pr\left[|u_t(\mathtt{R}) - \mathbb{E}[u_t(\mathtt{R})]| > O(\sqrt{4t \log t})\right] \leq 2\exp\left(\frac{-4t\log^2 t}{t \log t}\right) = O\left(\frac{1}{t^4}\right).$$

□

Combining Lemmas 4.5 and 4.10 yields:

Corollary 4.11.

$$\Pr\left[|\alpha_t - \alpha^*| > \max\left\{\frac{2\log t}{\sqrt{t}}, \frac{1}{\sqrt[3]{t}}\right\}\right] < \frac{1}{t^4}.$$

4.4 Degree distribution

We investigate the degree distribution of the red and blue vertices in a graph generated by the above described process, following the analysis outline of [7] for the basic preferential attachment model.

Let $m_{k,t}(\mathtt{B})$ (resp., $m_{k,t}(\mathtt{R})$) denote the number of blue (resp., red) vertices of degree k at time t. For $\mathbf{x} \in \{\mathtt{R}, \mathtt{B}\}$, define

$$M_k(\mathbf{x}) = \lim_{t\to\infty} \frac{\mathbb{E}(m_{k,t}(\mathbf{x}))}{t} \ . \tag{6}$$

Theorem 4.12. *The expected degree distributions of the blue and red vertices follow a power law, namely, $M_k(\mathtt{B}) \propto k^{-\beta(\mathtt{B})}$ and $M_k(\mathtt{R}) \propto k^{-\beta(\mathtt{R})}$. If $0 < r < 1/2$ and $0 < \rho < 1$ then $\beta(\mathtt{R}) > 3 > \beta(\mathtt{B})$.*

Equipped with Theorem 4.12, Part 2 of Theorem 4.1 follows easily. Indeed, for the *tail glass ceiling* effect, let $k(n) = n^{\frac{1}{\beta(\mathtt{R})}}$. Then

$$\mathbb{E}[\mathrm{top}_k(\mathtt{R})] = n(\mathtt{R}) \sum_{k' \geq k} M_{k'}(\mathtt{R}) \ ,$$

$$\mathbb{E}[\mathrm{top}_k(\mathtt{B})] = n(\mathtt{B}) \sum_{k' \geq k} M_{k'}(\mathtt{B}).$$

For $k' = n^{\frac{1}{\beta(\mathtt{R})}}$ we have $nM_{k'}(\mathtt{R}) = O(n \cdot n^{-\frac{\beta(\mathtt{R})}{\beta(\mathtt{R})}}) = O(1)$ while $nM_{k'}(\mathtt{B}) = \Omega\left(n \cdot n^{-\frac{\beta(\mathtt{B})}{\beta(\mathtt{R})}}\right) = \Omega(n^{1-\frac{\beta(\mathtt{B})}{\beta(\mathtt{R})}}) = \Omega(n^{\epsilon})$ for $\epsilon > 0$. The result then follows since $n(\mathtt{R}) < n(\mathtt{B})$ and $M_{k'}(\mathtt{R}) < M_{k'}(\mathtt{B})$ for $k' > k$.

For the *moment glass ceiling* effect we can show similarly:

$$g = \lim_{n\to\infty} \frac{\sum k^2 M_k(\mathtt{R})}{\sum k^2 M_k(\mathtt{B})} \ = \ \lim_{n\to\infty} \frac{O(n^{3-\beta(\mathtt{R})})}{\Omega(n^{3-\beta(\mathtt{B})})} = \lim_{n\to\infty} O\left(\frac{1}{n^{\epsilon'}}\right) \ = \ 0$$

for some $\epsilon' > 0$.

The rest of this section sketches a proof of Theorem 4.12. Note that $m_{0,0}(\mathtt{B}) = u_0(\mathtt{B})$. We derive a recurrence for $\mathbb{E}(m_{k,t}(\mathtt{B}))$. A blue vertex of degree k at time t could have arisen from three scenarios: (s1) at time $t-1$ it was already a blue vertex of degree k and no edge

was added to it at time t. (s2) at time $t-1$ it was a blue vertex of degree $k-1$ and an edge was added to it at time t. (s3) in the special case where $k=1$, at time $t-1$ it did not exist yet and it has arrived as a new blue vertex at time t. Thus letting $\mathcal{F}_t$ be the history of the process up to time t, for any $k>1$, the expectation of $m_{k,t+1}(\mathtt{B})$ conditioned on $\mathcal{F}_t$ satisfies

$$\begin{aligned}\mathbb{E}(m_{k,t+1}(\mathtt{B})|\mathcal{F}_t) = \quad & m_{k,t}(\mathtt{B})\left(1-\frac{ru_t(\mathtt{B})\rho\frac{k}{u_t(\mathtt{B})}}{u_t(\mathtt{R})+u_t(\mathtt{B})\rho}-\frac{(1-r)u_t(\mathtt{B})\frac{k}{u_t(\mathtt{B})}}{u_t(\mathtt{R})\rho+u_t(\mathtt{B})}\right)\\ & + m_{k-1,t}(\mathtt{B})\left(\frac{ru_t(\mathtt{B})\rho\frac{k-1}{u_t(\mathtt{B})}}{u_t(\mathtt{R})+u_t(\mathtt{B})\rho}+\frac{(1-r)u_t(\mathtt{B})\frac{k-1}{u_t(\mathtt{B})}}{u_t(\mathtt{R})\rho+u_t(\mathtt{B})}\right).\end{aligned}$$

For $k=1$ we similarly have

$$\mathbb{E}(m_{1,t+1}(\mathtt{B})|\mathcal{F}_t) = m_{1,t}(\mathtt{B})\left(1-\frac{\rho r}{u_t(\mathtt{R})+u_t(\mathtt{B})\rho}-\frac{1-r}{u_t(\mathtt{R})\rho+u_t(\mathtt{B})}\right)+(1-r)\ . \tag{7}$$

Recalling again that $\alpha_t = u_t(\mathtt{R})/(2t)$, the above can be rewritten as

$$\begin{aligned}\mathbb{E}(m_{k,t+1}(\mathtt{B})|\mathcal{F}_t) = \quad & m_{k,t}(\mathtt{B})\left(1-\frac{r\rho k}{2t(\alpha_t+(1-\alpha_t)\rho)}-\frac{(1-r)k}{2t(\alpha_t\rho+(1-\alpha_t))}\right)\\ & + m_{k-1,t}(\mathtt{B})\left(\frac{r\rho(k-1)}{2t(\alpha_t+(1-\alpha_t)\rho)}+\frac{(1-r)(k-1)}{2t(\alpha_t\rho+(1-\alpha_t))}\right)\end{aligned}$$

and for $k=1$,

$$\mathbb{E}(m_{1,t+1}(\mathtt{B})|\mathcal{F}_t) \ = m_{1,t}(\mathtt{B})\left(1-\frac{\rho r}{2t(\alpha_t+(1-\alpha_t)\rho)}-\frac{1-r}{2t(\alpha_t\rho+(1-\alpha_t))}\right)+(1-r). \tag{8}$$

This can be expressed as

$$\begin{aligned}\mathbb{E}(m_{k,t+1}(\mathtt{B})|\mathcal{F}_t) &= m_{k,t}(\mathtt{B})\left(1-A_t\frac{k}{t}\right)+m_{k-1,t}(\mathtt{B})A_t\frac{k-1}{t}\ ,\\ \mathbb{E}(m_{1,t+1}(\mathtt{B})|\mathcal{F}_t) &= m_{1,t}(\mathtt{B})\left(1-\frac{A_t}{t}\right)+(1-r),\end{aligned} \tag{9}$$

using the notation

$$A_t = \frac{r\rho}{2\alpha_t+2(1-\alpha_t)\rho}+\frac{(1-r)}{2\alpha_t\rho+2(1-\alpha_t)}\ .$$

Note that A_t is a random variable so we next bound its divergence. Let

$$\begin{aligned}C_{\mathtt{B}} &= \frac{r\rho}{2\alpha+2(1-\alpha)\rho}+\frac{(1-r)}{2\alpha\rho+2(1-\alpha)}\\ C_{\mathtt{R}} &= \frac{(1-r)\rho}{2(\alpha\rho+1-\alpha)}+\frac{r}{2(\alpha+(1-\alpha)\rho))}.\end{aligned}$$

We have

Lemma 4.13. $\Pr\left[|A_t-C_{\mathtt{B}}|>\max\left\{\frac{2\log t}{\sqrt{t}},\frac{1}{\sqrt[3]{t}}\right\}\right] \ < \ \frac{1}{t^4}.$

We use the following lemma.

Lemma 4.14. *[7] Let $(a_t),(b_t),(c_t)$ be three sequences such that $a_{t+1}=(1-\frac{b_t}{t})a_t+c_t$, $\lim_{t\to\infty}b_t=b>0$ and $\lim_{t\to\infty}c_t=c$. Then $\lim_{t\to\infty}a_t/t$ exists and its value is*

$$\lim_{t\to\infty}\frac{a_t}{t} \ = \ \frac{c}{1+b}\ . \tag{10}$$

Lemma 4.15.

- $M_1(\mathtt{B})$ *exists and equals* $(1-r)/(1+C_{\mathtt{B}})$,
- *For* $k \geq 2$, $M_k(\mathtt{B})$ *exists and equals* $M_{k-1}(\mathtt{B}) \cdot (k-1)C_{\mathtt{B}}/(1+kC_{\mathtt{B}})$,
- $M_1(\mathtt{R})$ *exists and equals* $r/(1+C_{\mathtt{R}})$, *and*
- *For* $k \geq 2$, $M_k(\mathtt{R})$ *exists and equals* $M_{k-1}(\mathtt{R}) \cdot (k-1)C_{\mathtt{R}}/(1+kC_{\mathtt{R}})$,

It is possible to show the following about $C_{\mathtt{B}}$ and $C_{\mathtt{R}}$:

Lemma 4.16.

- *If* $0 < r < 1/2$ *and* $0 < \rho < 1$ *then* $C_{\mathtt{R}} < \frac{1}{2} < C_{\mathtt{B}}$
- *If* $r = 1/2$ *then* $C_{\mathtt{R}} = C_{\mathtt{B}} = 1/2$.
- *If* $\rho = 0$ *or* $\rho = 1$ *then* $C_{\mathtt{R}} = C_{\mathtt{B}} = 1/2$.

To show that the degree distributions of both the red and the blue vertices follow power laws we recall that a power law distribution has the following property: $M_k \propto k^{-\beta}$ for large k, where β is independent of k. If $M_k \propto k^{-\beta}$, then

$$\frac{M_k}{M_{k-1}} \;=\; \frac{k^{-\beta}}{(k-1)^{-\beta}} \;=\; \left(1-\frac{1}{k}\right)^{\beta} \;=\; 1-\frac{\beta}{k}+O\left(\frac{1}{k^2}\right) .$$

Solving for the blue vertices, $M_k(\mathtt{B})$ and the blue exponent $\beta(\mathtt{B})$, and using Lemma 4.15, we get:

$$\frac{M_k(\mathtt{B})}{M_{k-1}(\mathtt{B})} = \frac{(k-1)\cdot C_{\mathtt{B}}}{1+k\cdot C_{\mathtt{B}}} \;=\; 1-\frac{C_{\mathtt{B}}+1}{k\cdot C_{\mathtt{B}}+1} = 1-\frac{1+\frac{1}{C_{\mathtt{B}}}}{k}+O\left(\frac{1}{k^2}\right)$$

hence $\beta(\mathtt{B}) = 1 + 1/C_{\mathtt{B}}$. Similarly, for red vertices of degree k, $M_k(\mathtt{R})$ decays according to a power law with exponent $\beta(\mathtt{R}) = 1 + 1/C_{\mathtt{R}}$. Note that when $C_{\mathtt{R}} < \frac{1}{2} < C_{\mathtt{B}}$ we have $\beta(\mathtt{R}) > 3 > \beta(\mathtt{B})$ thus proving Theorem 4.12.

4.5 Testing Competing Explanations: Proof of Theorem 4.2

No Minority.

If $r = 1/2$ then by Lemma 4.16 $C_{\mathtt{R}} = C_{\mathtt{B}}$, so the degree distribution of the two sub-populations is the same and a glass ceiling effect does not emerge.

No Homophily.

If there is no homophily, namely, $\rho = 1$ and $r < 1/2$, then by Lemma 4.16 $C_{\mathtt{R}} = C_{\mathtt{B}}$, so the degree distribution of the two sub-populations is the same and again no glass ceiling effect emerges.

No Preferential Attachment.

Let $U(n,r,\rho)$ be a random graph model similar to $G(n,r,\rho)$ except that when a new vertex z arrives at time t it chooses its neighbor $v \in V_t$ uniformly at random (with probability $1/t$). For simplicity we start at time $t=0$ with a single vertex (with arbitrary color). Let $\tilde{M}_k(\mathtt{R})$ and $\tilde{M}_k(\mathtt{B})$ denote, respectively, the expected number of red and blue vertices of degree k in $U(n,r,\rho)$. First we prove the following Lemma.

Lemma 4.17. *For any $0 < r < 1/2$ and any $0 \le \rho \le 1$, the expected number of red and blue vertices of degree k in a random graph $U(n,r,\rho)$ follows a geometric distribution. In particular:*

$$\tilde{M}_k(\mathtt{R}) = r \cdot p_r \cdot (1-p_r)^{k-1}$$

and

$$\tilde{M}_k(\mathtt{B}) = (1-r) \cdot p_b \cdot (1-p_b)^{k-1},$$

where $p_r = \frac{1}{1+p_r^*}$, $p_b = \frac{1}{1+p_b^*}$, $p_r^* = \frac{r}{1-(1-r)(1-q)} + \frac{q(1-r)}{1-r(1-q)}$ and $p_b^* = \frac{rq}{1-(1-r)(1-q)} + \frac{1-r}{1-r(1-q)}$.

Proof. As before, let $\tilde{m}_{k,t}(\mathtt{B})$ (resp., $\tilde{m}_{k,t}(\mathtt{R})$) denote the number of blue (resp., red) vertices of degree k at time t. Again, we will derive a recurrence for $\mathbb{E}(\tilde{m}_{k,t}(\mathtt{B}))$. Consider a red vertex v and let p_{rr} be the probability that v is selected given that the new arrived vertex is a red vertex. Then $p_{rr} = \frac{1}{t} + \frac{t-1}{t}(1-r)(1-q)p_{rr}$, so $p_{rr} = \frac{1}{1-(1-r)(1-q)+(1-r)(1-q)/t}\frac{1}{t}$. If the new arrived vertex is blue then the probability that v will be selected is $p_{br} = \frac{q}{t} + \frac{t-1}{t}r(1-q)p_{br}$, so $p_{br} = \frac{q}{1-r(1-q)+r(1-q)/t}\frac{1}{t}$. Similarly, if v is blue then the probability it is selected if the new vertex is red is $p_{rb} = \frac{q}{1-(1-r)(1-q)+(1-r)(1-q)/t}\frac{1}{t}$ and $p_{bb} = \frac{1}{1-r(1-q)+r(1-q)/t}\frac{1}{t}$ is the new vertex is blue.

Let $\mathcal{F}_t$ be the history of the process up to time t. Thus for any $k > 1$,

$$\mathbb{E}(\tilde{m}_{k,t+1}(\mathtt{B})|\mathcal{F}_t) = \tilde{m}_{k,t}(\mathtt{B})(1 - rp_{rb} - (1-r)p_{bb}) + \ \tilde{m}_{k-1,t}(\mathtt{B})(rp_{rb} + (1-r)p_{bb}).$$

For $k = 1$ we have

$$\mathbb{E}(\tilde{m}_{1,t+1}(\mathtt{B})|\mathcal{F}_t) = \tilde{m}_{1,t}(1 - rp_{rb} - (1-r)p_{bb}) + (1-r).$$

Taking the expectation on both sides we have

$$\mathbb{E}(\tilde{m}_{k,t+1}(\mathtt{B})) = \mathbb{E}(\tilde{m}_{k,t})(\mathtt{B})(1 - rp_{rb} - (1-r)p_{bb}) + \ \mathbb{E}(\tilde{m}_{k-1,t})(\mathtt{B})(rp_{rb} + (1-r)p_{bb}).$$

$$\mathbb{E}(\tilde{m}_{1,t+1}(\mathtt{B})) = \mathbb{E}(\tilde{m}_{1,t})(1 - rp_{rb} - (1-r)p_{bb}) + (1-r).$$

Let

$$\tilde{M}_k(\mathtt{B}) = \lim_{t\to\infty} \frac{\mathbb{E}(\tilde{m}_{k,t}(\mathtt{B}))}{t}. \tag{11}$$

and

$$p_b^* = \lim_{t\to\infty}(rp_{rb} - (1-r)p_{bb})t = \frac{rq}{1-(1-r)(1-q)} + \frac{1-r}{1-r(1-q)}\ .$$

Then using Lemma 4.14 and setting a_t, b_t and c_t accordantly we have

- $\tilde{M}_1(\mathtt{B}) = \frac{1-r}{1+p_b^*}$
- $\tilde{M}_k(\mathtt{B}) = \tilde{M}_{k-1}\frac{p_b^*}{1+p_b^*} = \tilde{M}_{k-1}(1 - \frac{1}{1+p_b^*}) = \tilde{M}_1(1 - \frac{1}{1+p_b^*})^{k-1}$.

Similarly, letting

$$p_r^* = \lim_{t\to\infty}(rp_{rr} - (1-r)p_{br})t = \frac{r}{1-(1-r)(1-q)} + \frac{q(1-r)}{1-r(1-q)}$$

we get

- $\tilde{M}_1(\mathtt{R}) = \frac{r}{1+p_r^*}$
- $\tilde{M}_k(\mathtt{R}) = \tilde{M}_{k-1}\frac{p_r^*}{1+p_r^*} = \tilde{M}_{k-1}(1-\frac{1}{1+p_r^*}) = \tilde{M}_1(1-\frac{1}{1+p_r^*})^{k-1}$

Setting $p_b = \frac{1}{1+p_b^*}$ and $p_r = \frac{1}{1+p_r^*}$ the result follows. □

Theorem 4.18. *For any $0 < r < 1/2$ and any $0 \le \rho \le 1$, the random graph $U(n, r, \rho)$ does not exhibit a* strong *glass ceiling effect.*

Proof. Since both degree distributions follow a geometric distribution we have:

$$g = \lim_{n\to\infty} \frac{\sum k^2 \tilde{M}_k(\mathtt{R})}{\sum k^2 \tilde{M}_k(\mathtt{B})} = \frac{r(1-2p_r)/p_r^2}{(1-r)(1-2p_b)/p_b^2} = \Omega(1).$$

□

4.6 Testing Competing Explanations: Unequal Rates and Unequal Qualification

Another possible competing explanation to the glass ceiling effect may be based on the conjecture that the effect occurs in areas where women have lower qualifications and skills than men. As we interpret the degree of a vertex as representing its power, unequal qualifications can be modeled by assuming that when a red (minority) vertex joins the network, it does so with a lower degree (fewer new edges) than does a blue vertex. This provides a "trivial" explanation to influence inequality, but will it cause a glass ceiling effect? We show that this is not the case: assuming no homophily, even if the minority has lower average degree, no glass ceiling effect emerges.

To formally model unequal qualifications, consider a random model similar to the *unbiased* preferential attachment model $G(n, r, \rho = 1)$ that we denote by $G_\Delta(n, r)$, operating as follows. At each time t a new vertex w joins the graph. Its color is red with probability r and blue with probability $(1-r)$. If w is red, then it generates one new edge according to preferential attachment as before. If w is blue then it generates Δ new edges, one at a time, according to preferential attachment.

We prove the following.

Theorem 4.19. *Let $0 < r < \frac{1}{2}$ and Δ a constant integer. Then $G_\Delta(n, r)$ does not exhibit a tail* glass ceiling *effect. Formally, for every k s.t. $\lim_{n\to\infty} \mathrm{top}_k(\mathtt{B}) = \infty$,*

$$\lim_{n\to\infty} \frac{\mathrm{top}_k(\mathtt{R})}{\mathrm{top}_k(\mathtt{B})} > c,$$

where $c > 0$ is a constant that depends only on r and Δ.

Proof. Instead of $G_\Delta(n, r)$, let us consider an equivalent process $G_1^\Delta(n, r)$ defined as follows. First generate $G_1(\Delta n, r)$ without coloring the vertices, according to the preferential attachment model. Then, consider the vertices of $G_1(\Delta n, r)$ in order of arrival, $v^1, v^2, \ldots, v^{\Delta n}$. Generate $G_1^\Delta(n, r)$ and its vertices $v_1, v_2, \ldots, v_n$ as follows. Initially $j = i = 0$. Assume by induction that vertices $v_1, v_2, \ldots, v_i$ were already generated by processing vertices $v^1, v^2, \ldots, v^j$. With probability r, v_{i+1} is red, in which case set the neighbors of v_{i+1} to be the vertices of G_1^Δ corresponding to the neighbors of v^{j+1} and increment j by 1. With probability $1-r$ vertex v_{i+1} is blue, in which case set v_{i+1} to be the "merging" of $v^{j+1}, v^{j+2}, \ldots, , v^{j+\Delta}$ into a single blue vertex. That is, the set of neighbors of v_{i+1} is taken to be the union the sets of vertices of G_1^Δ corresponding to the neighbors of $v^{j+1}, v^{j+2}, \ldots, , v^{j+\Delta}$; then increment j by Δ. Once n vertices are generated in this way, we ignore all remaining vertices $v^{j'}$ that were not used, as well as their edges. This defines $G_1^\Delta(n, r)$.

For the analysis, let n^* denote the number of vertices used from $G_1(\Delta n, r)$ and note that $n \leq n^* \leq \Delta n$. Let $G_1(\Delta n, r)[n^*]$ denote the induced subgrapgh of $G_1(\Delta n, r)$ from vertices $v^1, v^2, \ldots, v^{n^*}$. We prove that $G_1^{\Delta}(n, r)$ does not exhibit a glass ceiling effect and therefore $G_{\Delta}(n, r)$ doesn't either.

Let $\text{top}_k(\texttt{R}, G)$ denote the expected number of red vertices of degree at least k in the graph G. Consider a red vertex of degree k in $G_1^{\Delta}(n, r)$. The expected number of red vertices of degree at least k in $G_1^{\Delta}(n, r)$ is the same as in $G_1(\Delta n, r)[n^*]$.

Let $\hat{M}_k(\texttt{R})$ and $\hat{M}_k(\texttt{B})$ denote, respectively, the expected number of red and blue vertices of degree k in $G_1(\Delta n, r)[n^*]$. Note that $G_1(\Delta n, r)[n^*]$ follows a power law for both the red and blue vertices, and with the same β, namely, $\hat{M}_k(\texttt{R}) \propto k^{-\beta}$ and $\hat{M}_k(\texttt{B}) \propto k^{-\beta}$ for large k. Hence for constants c' and c'' we have:

$$\begin{aligned}
\text{top}_k(\texttt{R}, G_1^{\Delta}(n, r)) &= \text{top}_k(\texttt{R}, G_1(\Delta n, r)[n^*]) = \frac{r}{r + \Delta(1-r)} \text{top}_k(G_1(\Delta n, r)[n^*]) \\
&\geq \frac{r}{r + \Delta(1-r)} \text{top}_k(G_1(\Delta n, r)[n]) = \frac{r}{r + \Delta(1-r)} \int_k^n n \cdot c' \cdot i^{-\beta} di \\
&= \frac{r}{r + \Delta(1-r)} \frac{c'' \cdot n}{\beta - 1} k^{-(\beta-1)}.
\end{aligned}$$

Similarly, the number of blue vertices of degree at least k in $G_1^{\Delta}(n, r)$ can be upper bounded by the number of blue vertices of degree at least k/Δ in $G_1(\Delta n, r)[n^*]$, since merging Δ blue vertices of degree less than k/Δ cannot yield a blue vertex of higher degree than k. For constants c' and c'' we have:

$$\begin{aligned}
\text{top}_k(\texttt{B}, G_1^{\Delta}(n, r)) &\leq \text{top}_{\frac{k}{\Delta}+}(\texttt{B}, G_1(\Delta n, r)) = \frac{\Delta(1-r)}{r + \Delta(1-r)} \text{top}_{\frac{k}{\Delta}+}(G_1(\Delta n, r)) \\
&= \frac{\Delta(1-r)}{r + \Delta(1-r)} \int_{\frac{k}{\Delta}}^{\Delta n} \Delta n \cdot c' \cdot i^{-\beta} di \\
&= \frac{\Delta(1-r)}{r + \Delta(1-r)} \frac{c'' \cdot \Delta n}{\beta - 1} k^{-(\beta-1)} \Delta^{\beta-1}.
\end{aligned}$$

Putting it all together, we have that in $G_1^{\Delta}(n, r)$ there is no *tail* glass ceiling effect:

$$\frac{\text{top}(\texttt{R})_k}{\text{top}(\texttt{B})_k} \geq \frac{r}{(1-r)\Delta^{\beta+1}} .$$

A similar argument shows that there is also no *strong* glass ceiling effect. □

4.7 Auxiliary Lemma

The next lemma states that for any $r/2 \le \alpha \le r$, the straight line joining the points $(0, r/3)$ and $(\alpha, F(\alpha))$ is below the function $F(x)$ for all $x \in [0, \alpha]$.

Lemma 4.20. *For all $r \in (0, 1/2)$, $\rho \in [0, 1]$, $r/2 \le \alpha \le r$, and $x \in [0, \alpha]$*

$$F(x) \ge \frac{r}{3} + x\frac{F(\alpha) - r/3}{\alpha}$$

Proof. We need to show that $\psi(x) \equiv F(x) - \frac{F(\alpha)-r/3}{\alpha} \cdot x - \frac{r}{3} \ge 0$. Note that $\psi(x)$ can be rewritten as

$$\psi(x) = \frac{1}{6}\left(r\Big(x\Big(3w - \frac{1}{\alpha}\Big) + 1\Big) + 3\rho x\left(\frac{1}{\alpha(-\rho) + \alpha - 1} + \frac{1}{(\rho-1)x + 1}\right)\right) \tag{12}$$

$$\text{where } w = -\frac{1}{\alpha(-\rho) + \alpha + \rho} + \rho\Big(\frac{1}{\alpha(\rho-1)+1} + \frac{1}{-\rho x + x - 1}\Big) + \frac{1}{\rho - \rho x + x}.$$

We arrange $\psi(x)$ as $\frac{N(x)}{D(x)}$, then the denominator can be written as:

$$D(x) = 6\alpha(\alpha\rho - \alpha + 1)(\alpha\rho - \alpha - \rho)(\rho x - x + 1)(-\rho + \rho x - x)$$

which is positive, so the sign of $\psi(x)$ is determined by the numerator $N(x)$. The advantage of working with the numerator is that it is a polynomial. Some calculations yield that

$$\begin{aligned} N(x) = {} & (x-\alpha)r\rho(1 + \alpha(\rho-1))(-\rho + \alpha(\rho-1)) \\ & + (x-\alpha)r(\rho-1)^2x^2\Big(2(\alpha-1)\alpha(\rho^2+\rho-2) + \rho\Big) \\ & - (x-\alpha)r(\rho-1)x\Big(2\alpha^2(\rho^3 - 3\rho + 2) - \alpha(\rho(\rho(2\rho+3)-3)+4) + (\rho-1)\rho\Big) \\ & - (x-\alpha)3\alpha(\rho-1)\rho x(\alpha(\rho-1)-\rho)((\rho-1)x - \rho)\ . \end{aligned}$$

Clearly $N(x)$ has a root at α, i.e., $N(\alpha) = 0$. Since the degree of $N(x)$ is 3 it follows that it has at most three real roots $\lambda_1 \le \lambda_2 \le \lambda_3$. Assume that $\lambda_2 = \alpha$. A simple calculation shows that the leading coefficient of the polynomial $N(x)$ is $\alpha\rho(-r)(\alpha(\rho-1)+1)(\alpha(\rho-1)-\rho)$ and therefore it is positive. This implies that

$$\lim_{x \to -\infty} N(x) \ = \ -\infty$$

and that

$$\lim_{x \to \infty} N(x) \ = \ \infty.$$

Next we claim that $N[0] > 0$. Setting $x = 0$ in $N(x)$ we get that $N(0) = \alpha\rho(-r)(\alpha(\rho-1)+1)(\alpha(\rho-1)-\rho)$, which is positive in the range of the variables r, α, ρ. This shows that some root of N is less than 0, i.e., $\lambda_1 < 0$.

Next we show that there is one root of N that is greater than $1/2$, i.e., $\lambda_3 > 1/2$. To do this, we consider $N(1/2)$ and show that it is negative, i.e., $N(1/2) < 0$. A simple calculation shows that

$$\begin{aligned} N(1/2) = {} & (\frac{1}{2} - \alpha)r\frac{1}{2}(1-\rho)\Big(2\alpha^2(\rho^3 - 3\rho + 2) - \alpha(\rho(\rho(2\rho+3)-3)+4) + (\rho-1)\rho\Big) \\ & + (\frac{1}{2} - \alpha)r\Big(\frac{1}{4}(\rho-1)^2(2(\alpha-1)\alpha(\rho^2+\rho-2)+\rho) + \rho(\alpha(\rho-1)+1)(\alpha(\rho-1)-\rho)\Big) \\ & - (\frac{1}{2} - \alpha)\frac{3}{2}\alpha(\frac{\rho-1}{2} - \rho)(\rho-1)\rho(\alpha(\rho-1)-\rho). \end{aligned}$$

Viewing $N(0)$ as a function of r, we notice that $N(0)$ is a monotonically linear descending function of r and therefore one can assume that $r = \alpha$. In this case, i.e., $r = \alpha$, we get that

$$\begin{aligned} N(1/2) < (\frac{1}{2} - \alpha)\alpha\frac{1}{2}(1-\rho)\Big(2\alpha^2(\rho^3 - 3\rho + 2) \ - \alpha(\rho(\rho(2\rho+3)-3)+4) + (\rho-1)\rho\Big) \\ + (\frac{1}{2} - \alpha)\alpha\Big(\frac{1}{4}(\rho-1)^2(2(\alpha-1)\alpha(\rho^2+\rho-2)+\rho) + \rho(\alpha(\rho-1)+1)(\alpha(\rho-1)-\rho)\Big) \\ - (\frac{1}{2} - \alpha)\frac{3}{2}\alpha(\frac{\rho-1}{2} - \rho)(\rho-1)\rho(\alpha(\rho-1)-\rho). \end{aligned}$$

Now a simple calculation show that this is a monotonically descending function of ρ. Setting $\rho = 0$, we get

$$N(1/2) \ < \ \left(\frac{1}{2} - \alpha\right)\alpha\left(\frac{1}{2}\left(4\alpha^2 - 4\alpha\right) + (1-\alpha)\alpha\right),$$

which is less than 0 for all $0 < \alpha < 1/2$ so $\lambda_2 > 1/2$. This in turn imply that $\psi(x) \geq 0$ when $x \in [0, \alpha]$. □

4.8 Independence of tail and moment glass ceiling effects

To establish the independence of the two definitions of the glass ceiling effect, consider two sets of degree sequences, denoted A and B, where each set contains two degree sequences, for the red and blue vertices respectively. For simplicity, each degree sequence is of size n and the combined graph has $2n$ vertices. For each such set it is easy to construct a graph with the given degree sequences, for example by the random configuration model (which generates a random graph for every given degree sequence). Set A exhibits a tail glass ceiling effect but not a strong moment glass ceiling effect, whereas set B exhibits a strong glass ceiling effect but not a tail glass ceiling effect.

Set A. Let the degree sequence of the red vertices consist of $n - \sqrt{n}$ vertices of degree 1 and $\sqrt{n}$ vertices of degree $h = \lfloor \log n \rfloor$. The degree sequence of the blue vertices consists of $n - \sqrt{n}$ vertices of degree 1 and $\sqrt{n}$ vertices of degree $3h$.

Taking $k = 2h$, we get $\lim_{n\to\infty} \mathrm{top}_k(\mathtt{B}) = \infty$ and

$$\lim_{n\to\infty} \frac{\mathrm{top}_k(\mathtt{R})}{\mathrm{top}_k(\mathtt{B})} = \frac{0}{\sqrt{n}} = 0.$$

However, the network does not exhibit a strong moment glass ceiling effect, as

$$\lim_{n\to\infty} \frac{\frac{1}{n(\mathtt{R})}\sum_{v\in\mathtt{R}} \delta(v)^2}{\frac{1}{n(\mathtt{B})}\sum_{v\in\mathtt{B}} \delta(v)^2} = \lim_{n\to\infty} \frac{n - \sqrt{n} + \sqrt{n}h}{n - \sqrt{n} + 9\sqrt{n}h} = 1 - o(1) \geq \frac{1}{2}\ .$$

Set B. Let the degree sequence of the red vertices consist of n vertices of degree 2 (e.g., a ring). The degree sequence of the blue vertices consists of $n-1$ vertices of degree 1 and one vertex of degree $n-1$ (e.g., a star graph).

Taking $k > 1$, we get $\lim_{n\to\infty} \mathrm{top}_k(\mathtt{B}) = 1$ and the condition does not hold. If we take $k = 1$ then

$$\lim_{n\to\infty} \frac{\mathrm{top}_k(\mathtt{R})}{\mathrm{top}_k(\mathtt{B})} = \frac{n}{n} = 1.$$

Hence there is no tail glass ceiling effect. However, the network does exhibit a strong moment glass ceiling effect, as

$$\lim_{n\to\infty} \frac{\frac{1}{n(\mathtt{R})}\sum_{v\in\mathtt{R}} \delta(v)^2}{\frac{1}{n(\mathtt{B})}\sum_{v\in\mathtt{B}} \delta(v)^2} = \lim_{n\to\infty} \frac{\sum_1^n 2^2}{n - 1 + (n-1)^2} = \lim_{n\to\infty} \frac{4n}{n(n-1)} = 0.$$

5 Empirical Observations

To provide empirical evidence illustrating the results of our analysis in real-life, we studied a *mentor-student network* of researchers in computer science, extracted from DBLP [29], a dataset recording most of the publications in computer science. A filtering process creates a list of edges connecting students to mentors. For each edge we determined the gender of the student and the mentor and the year in which the connection was established. The resulting network spans over 30 years and has 434232 authors and 389296 edges. In the remainder of this section we describe the data collection and the assignment of gender to names followed by a temporal analysis of the minority-majority partition, influence inequality and glass ceiling effects.

5.1 Data Collection and Gender Assignment

Assigning Gender to Names

Unfortunately, the DBLP dataset (as well as the genealogy dataset) does not include direct information about the gender of the authors. In order to determine the gender of the authors, we made use of the fact that in most languages the first name also encodes the gender. Difficulties arise with unisex names, names that lose their gender information while being translated to the Latin alphabet (such as Chinese names), as well as single letter abbreviations (such as "A. Smith"). In order to match first names with their corresponding gender we built a dictionary including first names, their corresponding gender and a number between $[0, 1]$ describing the probability of the person with this first name being assigned the correct gender.

In order to build our name / gender dictionary we used four different datasets.

Dataset 1: US Birth Names (85547 names). This dataset uses all names from Social Security card applications for births that occurred in the United States after 1879 until 2012[1]. For each year of birth there is a list of names and their number of occurence for each gender. We summed up the counts over all years, i.e., we produced an entry for the total count of each name for both genders. From this list a name was assigned to be female or male respectively if its gender's count was more than 90% of the total count of occurences for both genders. Otherwise, the name was assumed to fit both genders. The female score of a name would thus be $count_f/(count_f + count_m)$.

Dataset 2: US Census 1990 Data (5163 names). This dataset has been composed by the Census Bureau in the US[2] and contains the names and the frequency of names for the sample male and female population respective according to the 1990 census. As an example: If, in the population of the 1990 census, 2000 people were named "Patricia" and 1900 were female, we assign it the gender "female" with probability 0.95 (score).

Dataset 3: Popular US Baby Names (4411 names). This dataset is from from the US Social Security Administration's statistics[3] for popular baby names and contains for every year between 1960 and 2010 the 100 most popular baby names. For each year and name the average probability of usage between 1960 and 2010 was calculated. In order to assign a name its gender, we compared the male probabiltiy to the female probability and assigned the conditional probability and the gender with the higher probability.

Dataset 4: Baby Name Lists for Parents (19833 names). The last dataset used is based on the gender information on a homepage collecting information on names made to help parents to choose a name for their baby[4]. This site notes for each name whether is used as a female,

[1] http://www.ssa.gov/OACT/babynames/limits.html
[2] http://www.census.gov/genealogy/www/data/1990surnames/names_files.html
[3] http://www.ssa.gov/cgi-bin/popularnames.cgi
[4] http://www.behindthename.com

Total Number of Authors in DBLP	1359616	100%
Female	172532	12.69%
Male	618830	45.52%
Names fits both	53330	3.92%
Excluded from top 1000	452	0.03%
Not found in name list	514472	37.84%

Table 1: The numbers of authors identified as female or male, or as having a name that fits both genders in the dictionary version v1. Also listed is the number of authors that have been excluded from the top 1000 authors. Note that this leads to an gender ratio of 21.08% females for the names with a gender assigned to them.

male or unisex name. However, there is no information on the frequency of its use for each gender. The score for a female only or male only name is hence set to 1.

The datasets mentioned above were unified via the following protocol. First, generate a list L based on Dataset 1. Each entry of the list consists of the tuple $(name, gender, score)$. Second, process Dataset 2: for each name of Dataset 1 that is already in L, check if the gender is the same, in which case the list is not modified. If the gender differs, then the name is declared to be unisex. If the name is not in L, add it together with the score of Dataset 2. Repeat the same process with Datasets 3 and 4.

In total, these sources led to a collection of 96,314 distinct names, including 36,316 names with a score of more than 0.9 for males and 58,827 names with a score of more than 0.9 for females. To assign a gender to the authors in DBLP we looked up their first name in our dictionary. If the probability of the name being female or male was over 90%, then the corresponding gender was assigned to the author. We refer to the resulting dictionary as *version v0*.

Cross Checking and Validation of Influential Authors. To make sure that the very active authors are identified correctly and to prevent a case where a name is representing several authors with the same name (a known problem in DBLP), we carried out a number of heuristic cross checks and validations. In particular, for the top 1000 authors in the dataset, we ran a script that filtered out potentially problematic nodes. Over all this resulted in excluding 452 (0.03%) nodes from the dataset. We refer to the resulting dictionary *version v1*. We believe that our version v1 dictionary is a cleaner and more accurate one, and therefore we present in the main paper results that are based on it. But in fact, the overall results of versions v0 and v1 are very similar. Table 1 summarizes the numbers of the gender assigning process.

Construction of Mentorship Graph

The empirical part of this article focuses on mentor relationships, as they are significantly influenced by homophily, as described in [22]. According to the Oxford Dictionary, a mentor is "an experienced person in a company or educational institution who trains and counsels new employees or students". In a scientific context, this typically includes guidance on writing research articles, especially for the first publications. Thus, mentors are often co-authors of young researchers in the first few years of their career. Clearly, a mentor has to have some experience, i.e., the mentor has started publishing a few years earlier than the mentee. Apart from websites such as the *Mathematics Genealogy Project* ("http://genealogy.math.ndsu.nodak.edu/"), which collects data on PhD students and their advisors in the field of mathematics, we are unaware of databases on scientific mentoring[5]. Therefore, we base our empirical findings on a mentoring graph constructed from the publication data base DBLP, an on-line reference for bibliographic information on major computer

[5]we empirically studied the Mathematics Genealogy Project, with similar results, but do not report it here since the network size is much smaller.

Parameter	Value	%
Number of Females	90035	20.73
Number of Males	344197	79.27
Sum of Females and Males	434232	100.00
Mixed Edges	101607	26.09
Female-Female edges	16074	4.12
Male-Male Edges	271615	69.77
Total number of edges	389296	100.00
Sum of Female degrees	133755	17.18
Sum of Male degrees	644837	82.82
Sum of edges	778592	100.00
Number of Female Mentors	10819	14.34
Number of Male Mentors	64638	85.66
Sum of Female and Male Mentors	75457	100.00
Females Avg. Degree	1.48	
Males Avg. Degree	1.87	
Avg Degree	1.79	
Female Mentors Avg. Degree	4.60	
Male Mentors Avg. Degree	5.25	
Mentors Avg. Degree	5.16	

Table 2: DBLP mentor graph statistics

science publications. It has evolved from an early small experimental web server to a popular open-data service for the computer science community. The entire DBLP dataset is freely available as a large XML file containing all bibliographic records. For each publication, this database provides the authors, the year, and the journal or conference, among other data.

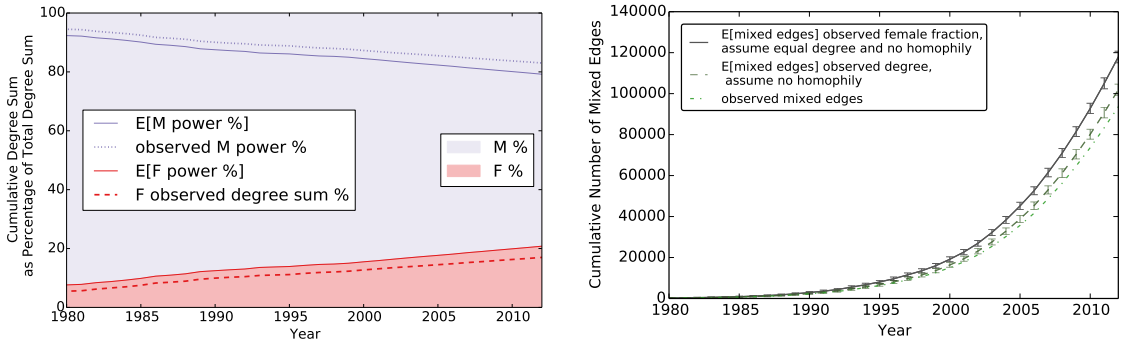
For each author in DBLP, we looked at the set of co-authors to find potential mentors. More precisely, we considered all people that co-authored an article in the first four years of a young researcher (we only considered papers with up to 20 authors) . In addition we relied on the assumption that the mentor has significantly more experience than the mentee. We looked at the years when the researchers wrote their first article. A person was only considered as a potential mentor of a mentee if the difference in the number of years between the dates of their first articles exceeded four. Like this we computed a set of eligible mentoring candidates for each author in DBLP. Among these candidates, we selected the one with the highest number of early papers written (if there are several authors competing for this position, we picked one at random).

The DBLP snapshot downloaded on December 23, 2013 contains 8,867,408 articles with two or more authors, written by a total of 1,282,790 people. Among them 871,839 have a set of at least one mentor candidate, i.e., 68.01 % of the authors in DBLP can be assigned a mentor with this method.

When using the same procedure but requiring an experience difference of at least 5 or 6 years, the percentage of authors than can be assigned an author decreases to 65% and 62% respectively. The changes in our observations however is negligible. Table 2 gives general statistics on the mentor graph. Note that any author with degree 2 or above is a mentor.

5.2 Temporal Analysis

As may be expected based on previously reported studies, our mentor-student network exhibits a minority-majority partition (namely, a low proportion of up to 21% females), homophily, power law distribution and a glass ceiling effect.



(a) (b)

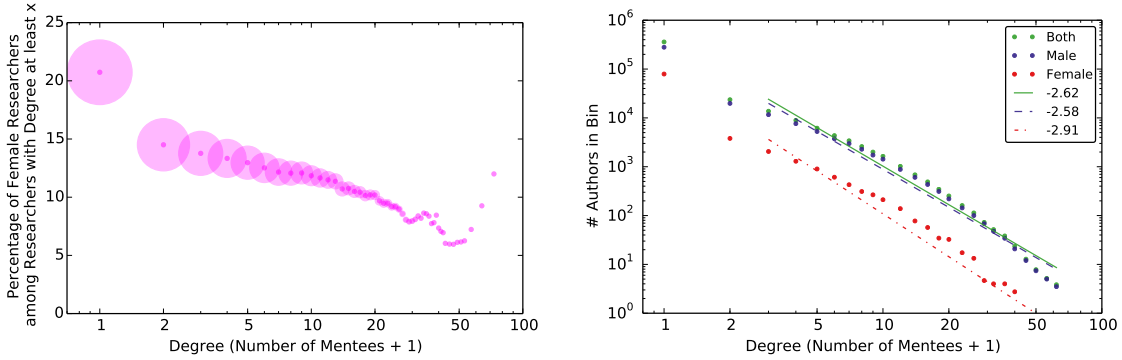
Figure 5: Female rate and *normalized power* in the computer science mentor graph. (a) The rate (i.e., percentage in the population) of females over time, compared with their normalized power, defined as $d(\mathtt{R})/(2m)$. Males have more power than expected by their rate, while females have less power than expected by their rate. (b) Evidence for homophily: a comparison of the observed number of "mixed" edges to the expected value assuming there is no homophily. We consider two cases: (i) the expected number of mixed edges ignoring the difference between the male and female average degree (expected: 127963.09 std: 293.08) and (ii) the expected number of mixed edges while considering the different degree sequences for males and females (expected: 110777.11 std: 281.52). In both cases the observed value (101607 edges) significantly deviates from the expectation (the error bars indicate the expected value $\pm$ 10 times the standard deviation) with extremely low p-values.

Figure 5(a) reveals that over time, the fraction of females in the network ($n(\mathtt{R})/n$, the shaded red area) has increased, but it is still below 21%. Also the average degree for females vertices is lower (1.48 vs 1.87). Figure 5(b) presents an indication for homophily in the mentoring selection process. This is done by the *homophily test* of [11], which compares the expected number of "mixed" (female-male) edges to the observed one (see also Section 3.2).

Figure 6 presents indications for the glass ceiling effect. Figure 6(a) shows that the fraction of females among the vertices of degree k or higher, namely, $\mathrm{top}_k(\mathtt{R})/\mathrm{top}_k(\mathtt{B})$, decreases continuously as k increases. The first major decrease occurs when moving from the group of "students" (i.e., degree 1 vertices) to the group of researchers of degree 2 or higher: the fraction of females drops from $\mathrm{top}_1(\mathtt{R}) \approx 21\%$ to $\mathrm{top}_2(\mathtt{R}) < 15\%$. It is important to note that the data indicates that even at the high end of the graph, a few female researchers with very high degrees are still present; however, our definitions for the glass ceiling ignore this extremal effect, which is caused by a few individuals, and concentrate on the averages over large samples. Indeed, when the sample size is large enough, the fraction of the female researchers decreases. Figure 6(b) shows a strong indication that the degree distribution of the vertices (females, males and combined) follows a power law. This in turn is associated with a preferential attachment mechanism that is known to result in a power law degree distribution. Note that the power-law exponent β for the graph of the female researchers is $\beta = 2.91$ (in the best fit), which is higher than the corresponding exponent in the graph for the male researchers, $\beta = 2.58$. Our analysis (presented in 4.2 and 4.3) establishes that if the degree distribution of both sub-populations follow a power law and the exponent for the minority sub-population is higher than that of the majority sub-population, then a strong moment glass ceiling effect will appear.

5.3 Data and Code Files

All data and code files are available at `http://www.glassceiling.pignolet.ch/`.



(a) (b)

Figure 6: Glass ceiling effect in mentor graph: (a) percentage of females in the mentor population of degree at least k. Female start with 21% in the population and drop to below 15% when considering degree at least 2 (faculty members). It continues to decrease (ignoring small samples at the end, see text). Vertex size and darker color represent larger sample space. (b) The power-law-like degree distribution for both females and males. The exponent β for females is higher than for males, demonstrating the glass ceiling effect.

6 Discussion and Conclusion

One obvious limitation of our model is that it is somewhat simplistic and captures only one possible mechanism for generating a glass ceiling effect. It ignores many important aspects of real life (such as sexual tension, fear, family responsibilities and jealousy, to name a few) and alternative (co-existing) mechanisms that contribute to the effect. For instance, our model cannot be used to explain the occurrence of a glass ceiling effect in contexts where pairwise individual interactions play a less dominant role than in academia. To account for the glass ceiling effect in such contexts as well as others, one may consider alternative explanations. In particular, a common possible explanation is the "leaky pipeline" phenomenon, namely, the phenomenon that women tend to quit or slow down their careers in order to invest more time in their families. This phenomenon can be modeled mathematically in several different ways. One such way is by introducing vertex departures in addition to vertex arrivals, with a bias in the form of increased departure rate of the minority group. But in fact, such a dynamic "leaky pipeline" model allows several reasonable sub-models that will not generate a glass ceiling effect, as well as some other sub-models that do. Moreover, the cause and effect relationships between glass ceiling and leaky pipeline are not necessarily one-directional; while the glass ceiling effect may indeed be the outcome of the "leaky pipeline" phenomenon in certain settings, there are other settings where it may be its (partial) cause. An interesting direction for future work would be to describe a more complete model, most likely combining a number of different mechanisms contributing jointly to the glass ceiling effect. In any case, we find it remarkable that the simple mathematical mechanism presented here (based on homophily) is sufficient to explain (at least parts of) the glass ceiling effect, despite the fact that it does not utilize the "leaky pipeline".

Our findings may suggest ways to deal with the glass ceiling phenomenon. By better understanding the roots of the glass ceiling effect, one can address each of the elements and attempt to mitigate them or deal with those elements that are easier to manage. Our research indicates that for certain mechanisms involved in the formation of a glass ceiling, removing one element may eliminate the glass ceiling effect. Hence, while it might be difficult to modify the human tendencies of homophily and preferential attachment one could attempt to balance the proportions of minorities within the population or impose a proportional representation of successful women at the top level. Both of these options may be classified as variants

of affirmative action, but the latter, even if more common, seems to avoid the roots of the problem. In particular, a more equally represented society could be created by encouraging minorities to enter the system, as our findings indicate that increasing the ratio of minorities at the entry stage may mitigate the glass ceiling effect at least partially. This conclusion is in line with a common view [38, 45], which states that fixing the "leaky pipeline" is key for a more equal gender distribution in science. By determining and examining the causes of the glass ceiling effect, we can work on alleviating the glass ceiling effect, resulting in a richer and more diverse community.

Acknowledgements

The authors thank Eli Upfal for suggesting the use of a Doob Martingale and the anonymous reviewers of this paper.

References

[1] N. Alon and J. H. Spencer. *The probabilistic method*. John Wiley & Sons, 2004.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] S. J. Bock, L. J. Taylor, Z. E. Phillips, and W. Sun. Women and minorities in computer science majors: Results on barriers from interviews and a survey. *WOMEN*, 14(1):143–152, 2013.

[4] Y. Bramoullé, S. Currarini, M. O. Jackson, P. Pin, and B. W. Rogers. Homophily and long-run integration in social networks. *Journal of Economic Theory*, 147(5):1754–1786, 2012.

[5] D. J. Brass. Men's and women's networks: A study of interaction patterns and influence in an organization. *Academy of Management Journal*, 28(2):327–343, 1985.

[6] S. J. Ceci and W. M. Williams. Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8):3157–3162, 2011.

[7] F. R. K. Chung and L. Lu. *Complex graphs and networks*, volume 107 of *CBMS Regional Conference Series in Mathematics*. AMS Bookstore, 2006.

[8] D. A. Cotter, J. M. Hermsen, S. Ovadia, and R. Vanneman. The glass ceiling effect. *Social forces*, 80(2):655–681, 2001.

[9] S. A. Davies-Netzley. Women above the glass ceiling perceptions on corporate mobility and strategies for success. *Gender & Society*, 12(3):339–355, 1998.

[10] W. W. Ding, F. Murray, and T. E. Stuart. Gender differences in patenting in the academic life sciences. *Science*, 313(5787):665–667, 2006.

[11] D. Easley and J. Kleinberg. Networks, crowds, and markets. *Cambridge Univ Press*, 6(1):6–1, 2010.

[12] D. Eder and M. T. Hallinan. Sex differences in children's friendships. *American Sociological Review*, pages 237–250, 1978.

[13] Education at a Glance. Organisation for economic co-operation and development. *OECD*, 2012.

[14] H. Etzkowitz, C. Kemelgor, M. Neuschatz, B. Uzzi, and J. Alonzo. The paradox of critical mass for women in science. *Science*, 266:51–54, 1994.

[15] A. Eyring and B. A. Stead. Shattering the glass ceiling: Some successful corporate practices. *Journal of Business Ethics*, 17(3):245–251, 1998.

[16] E. Falk and E. Grizard. The glass ceiling persists: The 3rd annual appc report on women leaders in communication companies. *The Annenberg Public Policy Center, University of Pennsylvania. Retrieved March*, 4:2005, 2003.

[17] W. Faulkner and M. Lie. Gender in the information society strategies of inclusion. *Gender, Technology and Development*, 11(2):157–177, 2007.

[18] Federal Glass Ceiling Commission. Solid investments: Making full use of the nation's human capital. US Government, Department of Labor. Washington, DC, 1995.

[19] A. Fisher, J. Margolis, and F. Miller. Undergraduate women in computer science: Experience, motivation and culture. In *Proceedings of the Twenty-eighth SIGCSE Technical Symposium on Computer Science Education*, pages 106–110. ACM, 1997.

[20] N. Frenkiel. The up and comers: Bryant takes aim at the settlers-in. *Adweek, March*, 1984.

[21] C. Hill, C. Corbett, and A. St Rose. *Why So Few? Women in Science, Technology, Engineering, and Mathematics.* ERIC, 2010.

[22] H. Ibarra. Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative science quarterly*, pages 422–447, 1992.

[23] H. Ibarra. Paving an alternative route: Gender differences in managerial networks. *Social Psychology Quarterly*, pages 91–102, 1997.

[24] A. P. Kottis. Women in management: The "glass ceiling" and how to break it. *Women in Management Review*, 8(4), 1993.

[25] V. A. Lagesen. The strength of numbers: Strategies to include women into computer science. *Social Studies of Science*, 37(1):67–92, 2007.

[26] V. Lariviere, C. Ni, Y. Gingras, B. Cronin, and C. R. Sugimoto. Bibliometrics: Global gender disparities in science. *Nature*, 504:211–213, 2013.

[27] P. F. Lazarsfeld, R. K. Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.

[28] S.-J. Leslie, A. Cimpian, M. Meyer, and E. Freeland. Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219):262–265, 2015.

[29] M. Ley. Dblp: some lessons learned. *Proceedings of the VLDB Endowment*, 2, 2009.

[30] T. J. Ley and B. H. Hamilton. The gender gap in nih grant applications. *Science*, 322(5907):1472–1474, 2008.

[31] P. Longo and C. J. Straehley. Whack! i've hit the glass ceiling! women's efforts to gain status in surgery. *Gender medicine*, 5(1):88–100, 2008.

[32] E. E. Maccoby. *The two sexes: Growing up apart, coming together.* Harvard University Press, 1998.

[33] R. F. Martell, D. M. Lane, and C. Emrich. Male-female differences: A computer simulation. 1996.

[34] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[35] M. Mitzenmacher and E. Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[36] C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, and J. Handelsman. Science facultyís subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41):16474–16479, 2012.

[37] M. Othman and R. Latih. Women in computer science: No shortage here! *Commun. ACM*, 49(3):111–114, Mar. 2006.

[38] A. N. Pell. Fixing the leaky pipeline: women scientists in academia. *Journal of animal science*, 74(11):2843–2848, 1996.

[39] E. S. Roberts, M. Kassianidou, and L. Irani. Encouraging women in computer science. *SIGCSE Bull.*, pages 84–88, 2002.

[40] S. Sassler, J. Glass, Y. Levitte, and K. Michelmore. The missing women in stem? accounting for gender differences in entrance into stem occupations. In *Annual meeting of the Population Association of America Presentation*, 2011.

[41] J. M. Sheltzer and J. C. Smith. Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences*, 2014.

[42] H. Shen. Mind the gender gap. *Nature*, 2013.

[43] L. Smith-Lovin and J. M. McPherson. You are who you know: A network approach to gender. *Theory on gender/feminism on theory*, pages 223–51, 1993.

[44] E. Spertus. Why are there so few female computer scientists? Technical report, Cambridge, MA, USA, 1991.

[45] P. E. Stephan and S. G. Levin. Leaving careers in IT: gender differences in retention. *The Journal of Technology Transfer*, 30(4):383–396, 2005.

[46] R. Stross. What has driven women out of computer science? *New York Times*, 15, 2008.

[47] N. B. Tuma and M. T. Hallinan. The effects of sex, race, and achievement on schoolchildren's friendships. *Social Forces*, 57(4):1265–1285, 1979.

[48] J. D. West, J. Jacquet, M. M. King, S. J. Correll, and C. T. Bergstrom. The role of gender in scholarly authorship. *PloS one*, 8(7):e66212, 2013.

[49] M. A. Whitecraft and W. M. Williams. *Why Aren't More Women in Computer Science? In Making Software: What Really Works, and Why We Believe*. 2011.