Mathematically Modeling to Reduce Bias in the Workplace

Jackson Carlberg, Meghan Flack, Jasmine Samadi, Samara Silverman

ORIE 4580/5580: Simulation Modeling and Analysis

Ziv Scully

FA 2024

Executive Summary

As students in ORIE 4580/5580, we are seeking to explore biases present in hiring processes that lead to gender disparity in the workplace, specifically the disproportionate favor given to male candidates in applicant pools of varying qualification. Through this modeling exercise, we are seeking to explore hiring processes that can reduce gender disparity in the workplace via altered methods of evaluating candidates for both new hires and promotion. We have elected to explore gender, seniority, homophily bias, and education level as evaluation parameters for hiring and promotion, and have designed multiple models that take into account each of these factors independently, as well as in combination with each other.

This exercise, like all simulation models, presents multiple challenges and limitations that need to be taken into account during analysis of the results produced by each model. Starting with the obvious, our workplace model lacks all of the social nuance of an actual workplace, and looks at employees strictly as numbers/objects in a data set. This type of modeling removes all the nuances of workplace culture and performance in an employee's current position within the company. Our model looks strictly at random departure of employees, so we are not able to analyze how workplace culture impacts the ratio of men and women who choose to work for and leave the company. Because we have modeled based on random departure and immediate promotion or hiring to fill empty positions, employee performance cannot be evaluated as a factor. This also removes charisma/personality and professionalism in interviews as an aspect of the hiring process, which is commonly a strong contributor to the decision to hire an individual. These are just some of the many factors that limit our model, and have a degree of randomness that cannot be easily quantified, and therefore are not easily modeled in a context such as this without some kind of external assessment to quantify these "human elements" that cannot be separated from the hiring and promotion processes. It also significantly simplifies the inputs, having a strict gender binary (male versus female) and a strictly linear promotion process.

For our simplified hiring and promotion process, we explored the impact of the initial number men and women at each level, seniority of each employee based on how long they had been working at their current level, homophily bias (e.g. men are more likely to hire/promote other men, and women other women), and education level, which was strictly random. When looking at an employee, each one is assigned a gender, a hire date, and one level of education

from high school diploma, undergraduate degree, master's degree, and PhD. We explored each of these factors independently from each other, including models that promoted strictly based on seniority, a combination of seniority and gender of both the interviewer and interviewee (homophily model), and education level with preference being given to individuals with higher levels of education regardless of gender. These models all follow the same general structure when modeled as a Markov Chain, with different hiring processes being represented by changing the probabilities of transitioning from one state to another.

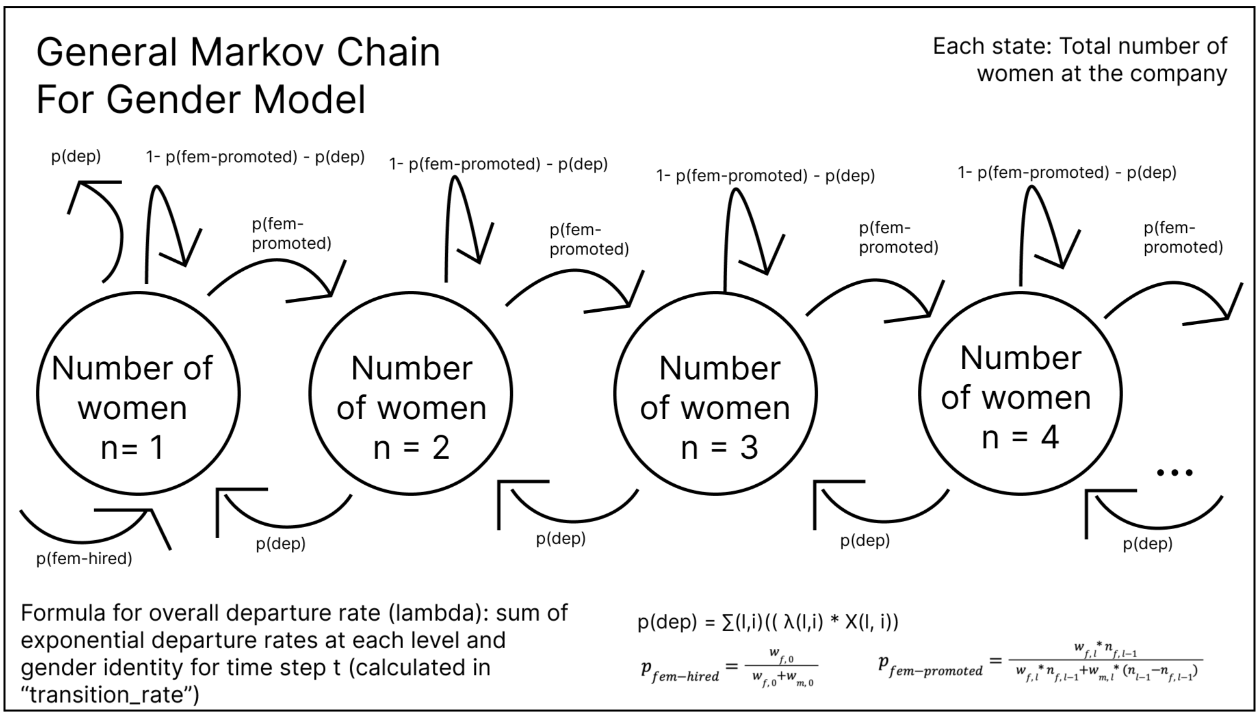


Figure 1: Continuous Time Markov Chain Model used to analyze the number of women at each corporate level. Probabilities vary depending on what factors are considered in the hiring and promoting processes.

The main questions that we sought to explore with these models were how the proportion of women at the hiring level impacts the amount of women able to enter the company over time, and whether weighting education as a heavy priority would have a significant impact in combatting homophily bias. We hypothesized that there would be a significant decrease in the number of women both hired and promoted if the starting state at each level was biased significantly toward men, and that it would create a lack of accessibility for women seeking to enter the workforce. We also hypothesized that there would be a significant change in the disparity once we introduced education as a factor, and that randomizing education would give the most realistic view of what the hiring process looks like.

With the initial models that we developed, both of these hypotheses seem to be supported by the results. For a next step, we would recommend adjusting the departure rate of men and women to depend on the number of men and women currently employed at the company, as this could be used to represent workplace culture (e.g. women are less likely to want to work in a male dominated office). We would also want to look further into how identities such as race and sexual identity could further impact the homophily bias, as well as increasing the complexity of existing models such that gender is not considered as a strict binary and there are opportunities for lateral movement within a company rather than a solely linear promotion structure.

Main Body

This report outlines the simulation experiments conducted to study gender dynamics in corporate hierarchies. It aims to address key questions regarding how factors such as initial representation, departure rates, and promotion biases impact the representation of women at various levels within a company. The report provides a summary of findings for high-level decision-makers while explaining the technical details behind the simulations for specialists.

Our model takes in a number of different factors to determine theoretical dynamics in hierarchical corporate structures. The main parameters of focus are seniority, gender, and education level. Each of these are expected to have a positive or negative bias associated with them for promotion. The model employs a continuous-time Markov chain framework, where states are the gender distribution at different levels and transitions between states are determined probabilistically based on the above parameters.

This model, like all models, is a simplification of a complicated system. In order to sample meaningfully, simplifications have been made. The crucial assumption of this model is that gender distribution at a given level influences the pool of candidates for promotion to the next. This is meant to sample from possible interviewers and simulate the inherent homophily bias in human nature to hire people more like themselves. In different iterations of the model, this homophily is stronger or weaker, but it is always present. Furthermore, the model assumes that employee departure rates are constant parameters, varying by position and gender, and filling of newly empty positions is exclusively from the pool of candidates in the lower level.

The structure of the model follows four different echelons representing different corporate levels. Level 0 represents the entry-level position with the largest employee base, and level 3 represents the level with the highest positions and smallest number of employees within this company. The distribution is as follows: Junior employees ($\ell = 0$) with 100 employees, Management ($\ell = 1$) with 50 employees, Senior management ($\ell = 2$) with 30 employees, and Executives ($\ell = 3$) with 10 employees. Employees within this system all have a gender, an education level, and a unique employee ID. Promotions depend on qualifications, gender biases, and availability of positions at higher levels. In practice, this means that promotions are available only when employees in higher levels leave the company. These departure rates are parameterized by gender and level, which means they can be manipulated for investigations.

This model was internally validated through a number of measures. Primarily, the simulation measures are reproducible. Though re-running the model naturally changed based on statistical noise, the trends remain the same. Beyond pure reproducibility, there is overall internal consistency in results that match expectations. For simple examples, such as equal starting proportions and promotion probabilities, theoretical results are simple to calculate, and the model follows expected results.

The focus of the research, and thus the model, was the end behavior of female versus male employees under different conditions. These conditions were decided based on a number of research questions, most pressingly homophily bias. In practice, this is analyzing how the amount of women at a level responsible for promotion affects the amount of women able to enter the level. Other factors considered include biasing for qualification, considering the departure rate of women across levels, and considering starting proportions. The simplest analysis considered was how the starting proportion of women at different levels influence their representation over time, and the output of this initial model is shown below in Figure 2.

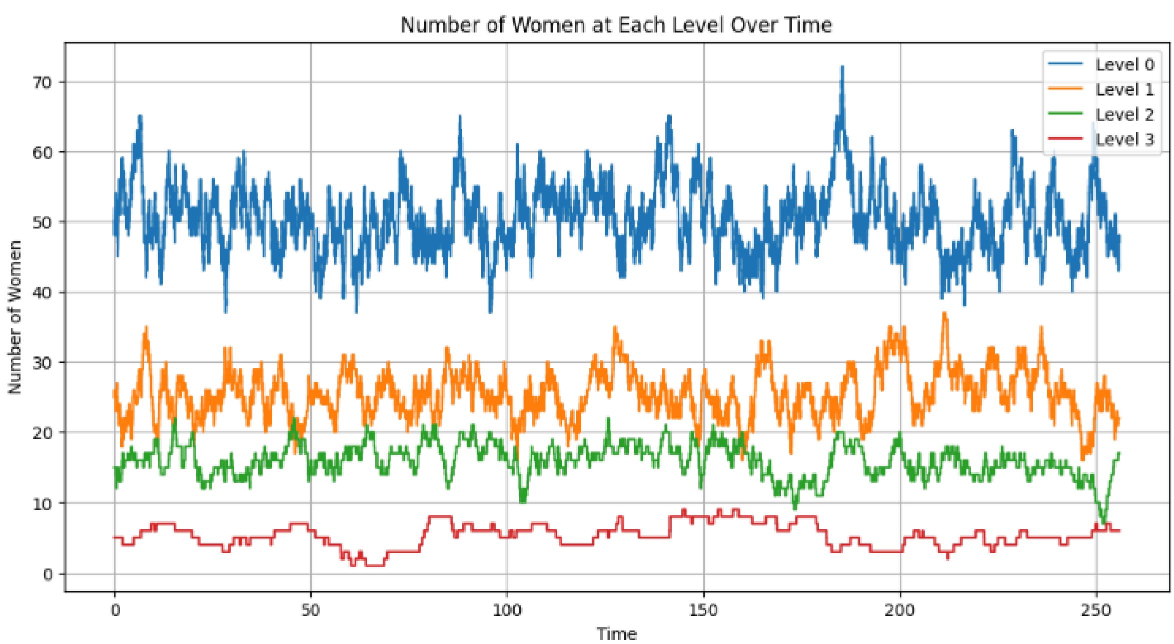


Figure 2: Number of Women at each Corporate Level over Time when the starting ratios and promotion rates of each gender are equal

Figure 2 demonstrates that starting with about equal men and women at each level and about equal chances of promotion across all levels, that the ending proportions at all levels are about the same as the beginning proportions. While a simple example, it is important to note the noise involved even here. Within a margin of error, the findings are what you would assume, that the number of women stays about the same.

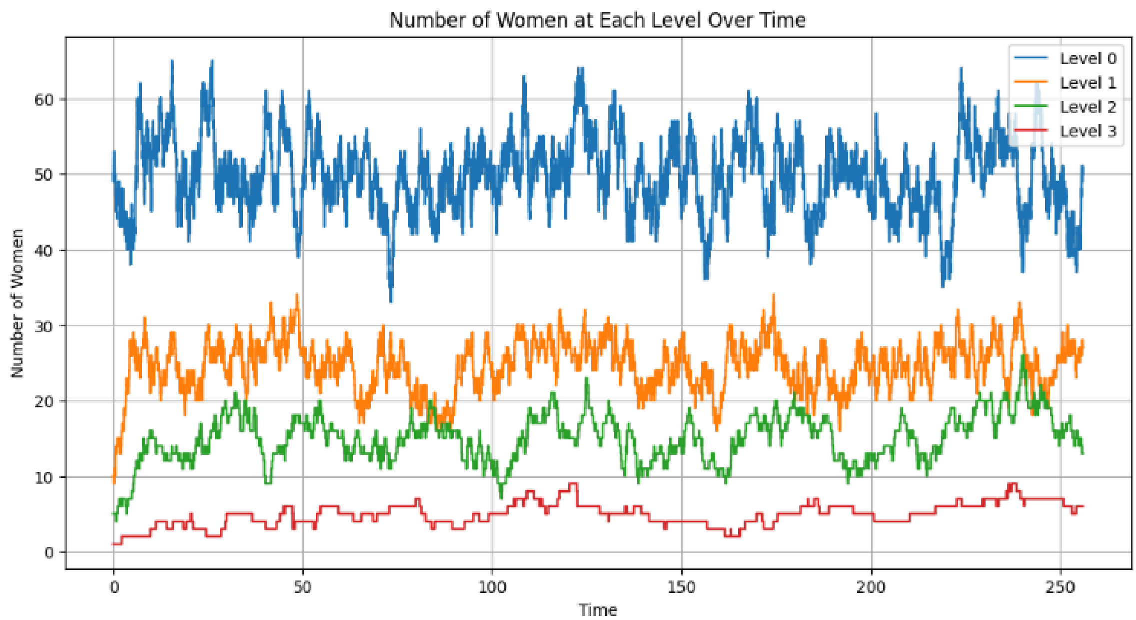


Figure 3: Number of Women at each Corporate Level over Time when the starting ratios of only level one are equal, but the promotion rates of each gender are equal

A much more interesting conclusion can be drawn from Figure 3. Though the starting ratios of women to men were abysmally low at levels 1, 2, and 3, because the hiring pool for promotion candidates was being drawn from level 0 and the promotion was unbiased by gender, the number of employees quickly evened out. This finding is what promoted further research into promotion practices and how those must be affecting overall workplace equity.

Changing from the simple, unbiased model to a model with homophily biased interviewers, the model in theory is most likely to continue the status quo, maintaining the ratio of women to men at a given level. However, with a relatively equal pool of candidates in level 0 that is not weighted by homophily, at lower levels the model trends towards balance, from either side. This effect lessens the more echelons in the system as can be seen in Figure 4, below. Level 1 began with significantly more female employees and Levels 2 and 3 began with significantly fewer female employees. Level 3, being the most removed from the control group of Level 0, is relatively unchanged, but levels 1 and 2 converged closer to equal numbers over time since there was a waterfall effect of available employees for promotion.

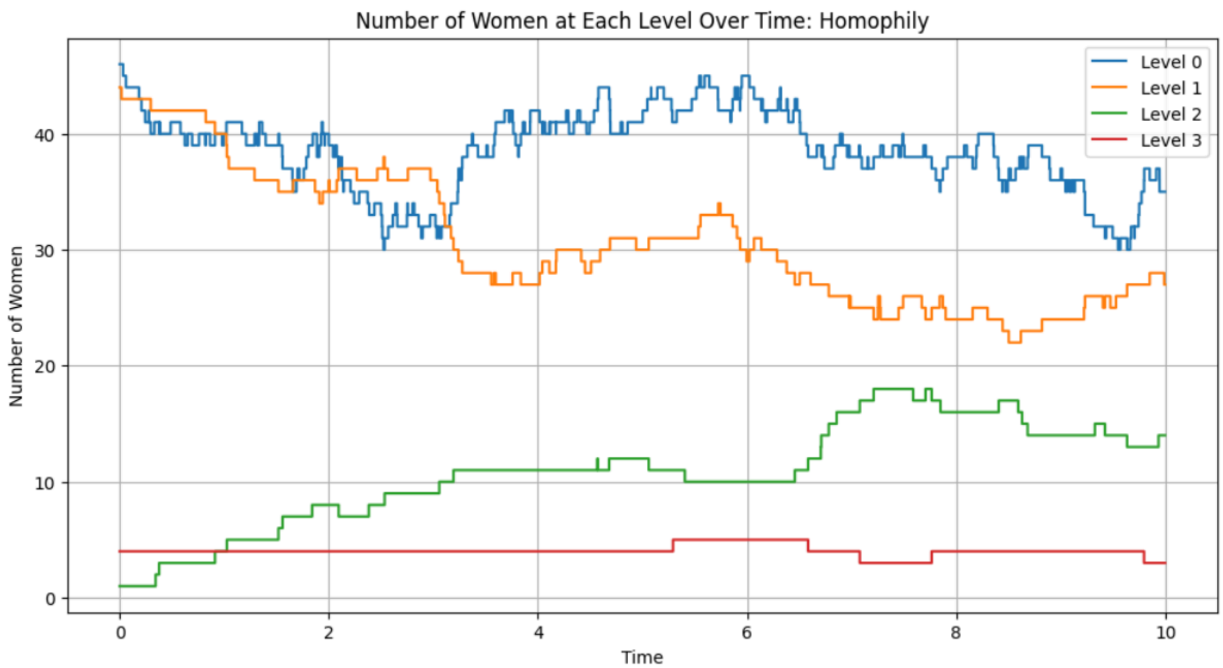


Figure 4: Women at Each Corporate Level over time in a Homophily-biased model with starting Women-to-Men ratios of Level 0: 0.5, Level 1: 0.9, Level 2: 0.1, Level 3: 0.1

When considering representation in a long term view, it is important to consider how the departure rates for women in senior positions affects the hiring and promotion decisions at lower levels, and by extension female representation in the company as a whole. According to a report from McKinsey and LeanIn, departure rates for women in industry is substantially higher than that of men, even more so when looking at more senior positions.[1] This impact can be felt at all levels of employment, especially when homophily bias is considered in the hiring process; as more women leave the company, especially in senior positions, more men are being hired and promoted, and in turn are themselves hiring and promoting more men, contributing to a self-perpetuating cycle that will trend toward maintaining strongly male-dominated workplaces as time goes on. This can be attributed to a number of factors, and the McKinsey report specifically calls out the increase in microaggressions directed at women in industry, including be mistaken for a junior employee, not receiving credit for their own ideas, and long standing

1 https://www.cnbc.com/2022/10/18/women-leaders-are-leaving-companies-at-highest-rate-ever-leanin-mckinsey-co-report.html

and continued bias against hiring and promoting women to senior position, especially c-suite positions, where only one in every four positions are held by women.[1] We have also considered that poor work/life balance, especially for family-oriented individuals, is a factor at play here. We believe that it is likely the case that companies would refrain from promoting women to senior positions on account of the possibility of having to offer maternity leave to individuals that elect to have children. In a similar vein, it is also possible that women would neglect to pursue certain positions due to insufficient or complete lack of maternity leave being offered, which is just one of many factors that contribute to a poor work/life balance.

When considering the impact on the hiring and promotion processes when there is a lack of women in more senior positions, it is easy to see the long term detrimental effects on the number of women at the company as a whole, as well as at each level. Figure 5 shows a case that starts with no women at the seniormost level and only one at the level below, in which there is a sharp decrease in women at the two junior levels as time goes on, with both numbers being roughly halved, and level zero experiencing a decrease of roughly 70% of their female employees between Time = 0 and Time = 5. This creates a bottleneck at the junior levels that prevents women from having any upward mobility, as shown by the consistent figure of zero women at the seniormost level for the entire iteration shown. In this case, a roadblock is created, as a lack of representation leads to a lack of opportunity for observed potential, and a continuation of bias toward male employees regardless of qualifications.

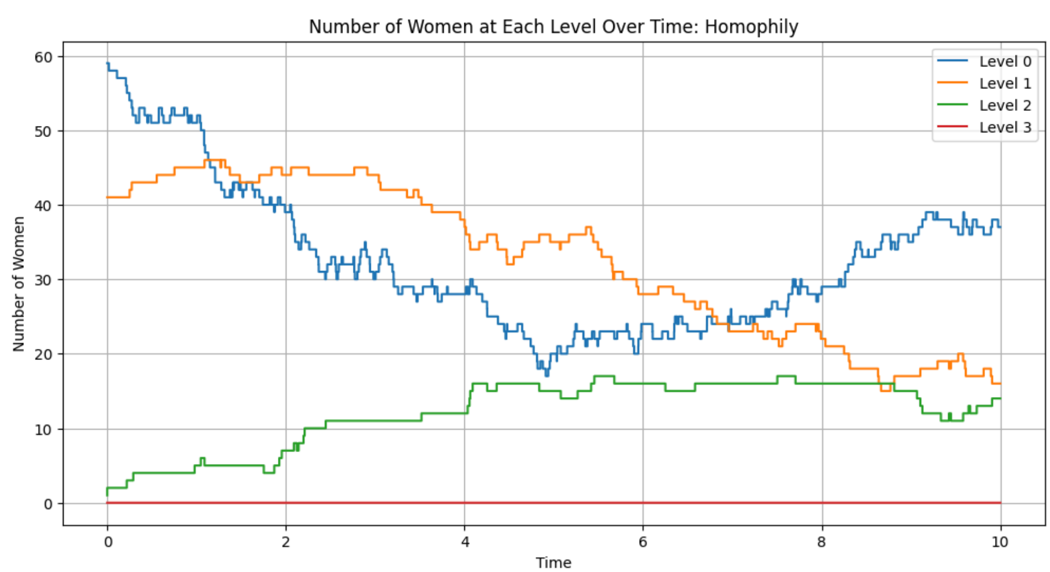


Figure 5: Women at Each Corporate Level over time in a Homophily-biased model with starting Women-to-Men ratios of Level 0: 0.5, Level 1: 0.9, Level 2: 0.1, Level 3: 0.1, extreme case

Figure 5 is a perfect demonstration of the problem at hand, and especially the complete lack of female representation at level 3, gives us an extreme case to compare to. As a method of mitigating this strong homophily bias, we implemented the random assignment of education level to each employee as they were generated at the hiring stage. Any employee, regardless of gender, is equally likely to have a high school diploma, undergraduate degree, master's degree, or PhD as their highest level of completed education. After adding this parameter to the employee object, we implemented a weighted preference in the hiring process that favored employees with more advanced degrees for promotions. As expected, employees with PhDs tended to get promoted faster, and the education level tended to roughly correlate with relative position within the company. By comparing Figures 6 to Figure 5, we can first see just how detrimental the introduction of homophily is with existing gender disparity in the workplace, and also how providing a lot of value to education can help to further reduce the disparity, even in the presence of homophily bias.

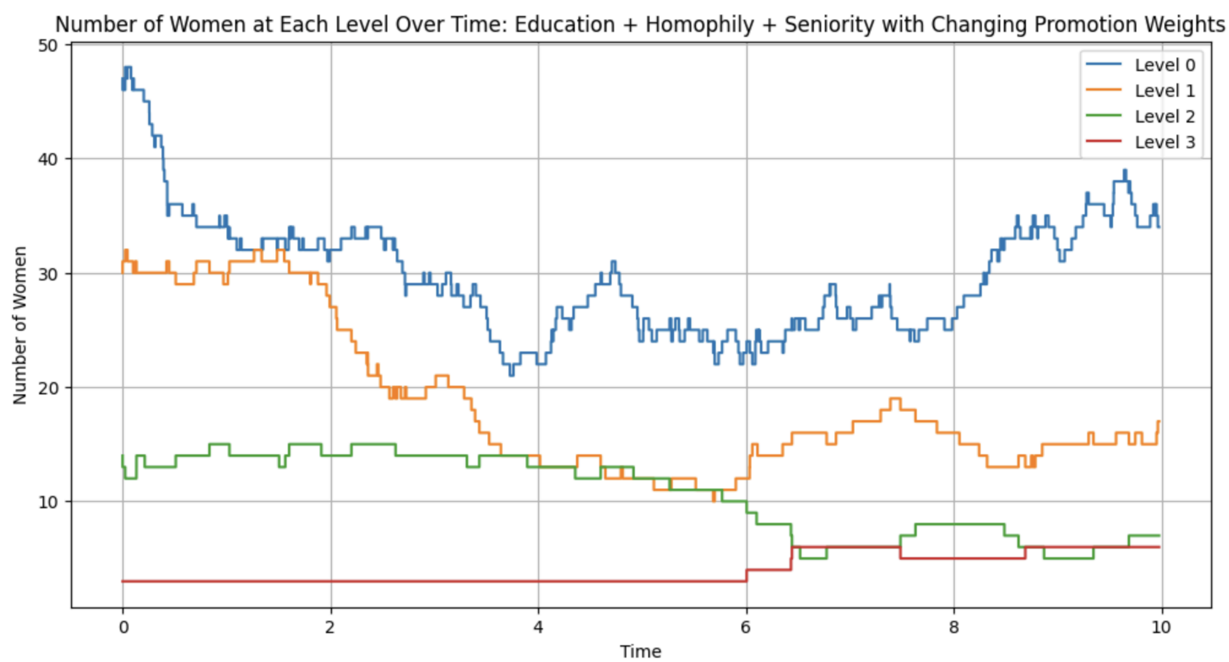


Figure 6: Women at Each Corporate Level over time with promotion weights: level 0 female: 0.5, level 0 male: 1.0, level 1 female: 0.3, level 1 male: 2, level 2 female: 0.2, level 2 male: 5, level 3 female: 0.1, level 3 male: 12

In Figure 6, one can see that with drastic differences in promotion weight for different genders across levels, the number of women at higher levels does not begin to increase for a large amount of time. It is worth noting that this model has a high homophily bias, which tends to drive level three toward and even split of men and women once the first woman is promoted to level 3.

Gender diversity in the workplace is a multifaceted issue that requires strategic intervention to overcome. There are a number of specific challenges highlighted that must be overcome in order to make strides towards workplace equality. The findings of this report indicate that initial representation, particularly in mid- and high-level positions, is a key point for long-term gender diversity. Furthermore, keeping women in these roles by reducing female departure rates, especially at senior levels, is critical. This ensures gender-balanced promotion pools and increases the likelihood of overall female retention. Also demonstrated was how objective merit-focused evaluations, such as education levels, significantly improve the promotion of women in the workplace.

The courses of action to be taken based on these findings are clear. Entry level “quotas” are not sustainable, so if that is a policy that the company wishes to uphold, they should be at mid-level management positions rather than for the Junior Employees. Also recommended is to emphasize merit based promotion decisions. These objective measures, while they do not need to be the only metric evaluated, reduce bias in decision making, and benefit historically marginalized groups and individuals in the promotion process. By implementing these data-driven strategies, companies can make meaningful progress toward workplace equity and inclusion. Perhaps even more important than those, though, is to create a company culture that women want to stay in. Surveys have found that women increasingly value companies that prioritize career advancement, flexibility, employee well-being, and diversity, equity, and inclusion—and are departing from their workplaces in record numbers when these priorities are not fulfilled to find work elsewhere. With this in mind, companies that want to ensure they have equitable workplaces should ensure that they not only have equitable hiring processes, but also equitable working conditions.

Technical Appendix

Research Questions:

- How does the amount of women at a level responsible for promotion affect the amount of women able to enter the level.
- How does biasing for qualification stronger than biasing against gender affect the amount of women at each level?
- How does the departure rate of female employees at each level affect the amount of women at the company over time?
- How does the starting amount of women at different levels affect the behavior over time of women at that company?

There are two classes with models built on top of them in this paper. Both the simple workplace model and complex workplace model take in the same basic workplace parameters. This includes: number of employees at each level, the departure rate of each level for both identities (lambda), and the weight of promotion/hiring for each level and identity. The first is the "SimpleWorkplaceModel". The simple workplace state includes integers representing the number of female employees at each level.

The complex workplace model tracks the amount of time each employee has been at the company. To keep track of this, there is a class "Employee", which includes a unique identifier, the gender identity, as well as a tuple containing the times that employee was hired and when they got promoted to other levels. The "Employee" class contains methods that create new instances of employees when randomly sampling a new hire, or when an employee gets promoted and their promotion times have changed. The complex workplace state is composed of tuples with instances of the "Employee" class for each hierarchy level. To simplify this input, there is a method that converts the complex workplace state into a simple workplace state. This way, it outputs the number of female employees at each level. The method iterates through every employee level, and sums the number of employees that have a female identity. It then passes this count into the "SimpleWorkplaceState" constructor to output a simple workplace state with a number of females at each level, while encapsulating the extra complex attributes in the "Employee" class.

Transition Rates and Sampling the Next Employees:

All of our models have the same general structure. They have a transition_rate function and a sample_next function. The transition_rate function calculates the total departure rate by summing up the departure rates for each employee (determined by lambdas that are inputted into the model). The sample_next function changes the state of the model. It simulates having one person departing their role and promotes one person at each level below to fill in the empty spots. Some of our more complex models have helper functions to determine who will get promoted.

There is a specific departure rate for the employees with distinct gender identities at each level ($\lambda(l, i)$). The overall departure rate for a given level and identity at a certain time is the specific departure rate mentioned above ($\lambda(l, i)$) multiplied by current state: $X(l, i)$ (number of employees with a specific gender and level at time t). In order to compute the transition rates, both the simplex and complex workplace models aggregate the rate, starting from 0, and at each level and identity add (number of employees $X(l, i)$ * specified departure rate $\lambda(l, i)$). In a continuous-time Markov chain, the total departure rate is the sum of each individual departure rate at each level. Therefore, this function outputs the sum of the individual departure rates for each level and identity: $\sum_{l,i}((\lambda\,(l, i) \ \star\ X(l,\ i))$.

The sample_next function has similar logic for both the simple and complex workplace models. A probability list was created to determine the probability of an employee with a certain identity and level leaving the company. This way an event of a gender identity from a certain level that leaves is randomly chosen based on the departure probability list. That is how the model knows when to fill a vacancy. For each level and gender, the probability of departure is equal to the departure rate for that specific group divided by the total departure rate calculated in the transition_rate method.

$$P[Employee\ from\ Level\ l\ and\ Identity\ i\ leaves] \ = \frac{(\lambda\,(l,i) \star X(l,\,i))}{\sum_{l,i}((\lambda\,(l,i) \star X(l,\,i))}$$

$$= \frac{departure\ rate\ for\ level\ l\ and\ identity\ i \times number\ of\ employees\ at\ level\ l\ with\ identity\ i}{total\ departure\ rate}$$

This probability list is populated to start with females at level 0, then males at level 0, and so on until the last two entries which are females at level 3 and males at level 3. This means that probabilities[0] is the probability of a female at level 0 leaving, and probabilities[1] is the probability of a male at level 0 leaving. There is a variable "event" defined by randomly picking a number in the length of the probability list based on the calculated probability of departures. This tells the simulation which identity is departing from which level.

Differences Between Models:

There are 4 models with different factors that impact promotion. The first model promotes employees based on gender and uses the simple workplace model structure. Once the departed employee is determined, the employee promoted to level ℓ is based on gender using the formula: $p_{female} = \frac{w_{f,\ell} * n_{f,\ell-1}}{w_{f,\ell} * n_{f,\ell-1} + w_{m,\ell} * (n_{l-\ell} - n_{f,\ell-1})}$ where $w_{f,\ell}$= promotion weight for females at level ℓ , $w_{m,\ell}$= promotion weight for males at level ℓ , $n_{f,\ell-1}$= number of females at level ℓ -1, $n_{\ell-1}$= total number of employees at level ℓ -1. This is done for the level that the employee is departing from and every level below to keep the number of employees the same at each level. Someone is also promoted at level 0 based on the weights entered into the model: $p_{female} = \frac{w_{f,0}}{w_{f,0} + w_{m,0}}$.

The second model promotes employees based on seniority and is based on the complex workplace model structure. It also takes gender into account when promoting at level 0. The model has helper functions to promote employees at each level. Employees are promoted by sorting them by the time that they entered their current levels. This value is the last element of the hiring_and_promotion_times attribute. The one with the minimum time (entered the current level the earliest) is chosen to be promoted to the next one. To promote to level 0, we use the same formula as the previous model.

The third model promotes employees based on seniority, homophily, and gender. This model is based on the complex workplace model structure. In order to determine who is promoted, an interviewer is picked randomly from the next level. A list of the top 3 employees from the current level based on seniority is created by sorting the employees in ascending order based on the time they each joined in the current level. The promoted employee is chosen from this list of top candidates with probabilities determined by their genders and the gender of the interviewer. If the interviewer is female, $p_{female} = w_{f,\ell} + (1 + h)$, $p_{male} = w_{m,\ell}$ and if the interviewer is male $p_{female} = w_{f,\ell}$, $p_{male} = w_{m,\ell} + (1 + h)$ where h represents the homophily bias. Each of the top candidates are assigned to p_{female} or p_{male} depending on their gender. These probabilities are then normalized since they may not sum to 1 and used to choose the promoted employee. Employees are promoted to level 0 in a similar way. The only difference is that a top candidates list based on seniority is not made. The probabilities p_{female} and p_{male} are calculated the same way as the other levels and the entering employee's gender is chosen using them.

The last model picks employees based on education, seniority, homophily, and gender. This model is based on the complex workplace model structure. For this model, the employee class was changed to add an extra attribute that keeps track of the highest education level (1 = high school, 2 = undergraduate degree, 3 = masters degree, 4 = PhD). The sample_new_hire function was also changed to randomly assign a new employee an education level. Therefore, when promoting to level 0, education is randomly assigned and not used as a picking criteria. When promoting to levels above 0, the method of picking an employee is almost the same as the previous model. The only difference is that employees are sorted first by education and then by seniority (using the current level time) when picking the top 3 candidates. The probabilities p_{female} and p_{male} are calculated and assigned to the top candidates the same way as the previous model. Then, the probabilities are normalized and a candidate is picked from the top candidates list to be promoted.

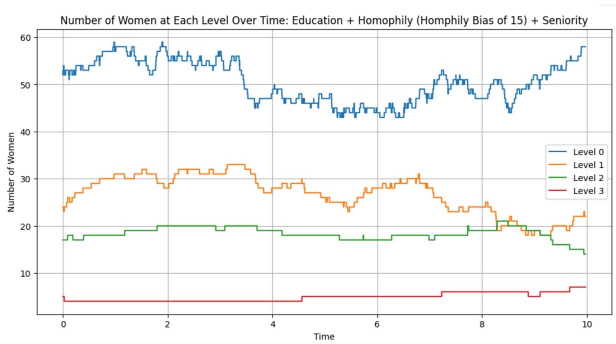
Additional Justification: Changing the Homophily Bias

In order to see the effect of the homophily bias parameter, the model including education, seniority, gender, and homophily were tested with a homophily bias of 15 instead of the previous 5. For chart 7A, the initial number of women at each was randomly sampled with a probability of

0.5 to be female. There was an equal weight of 1 for males and females to be promoted at each level. In chart B, the initial number of females at levels 2 and 3 were 0, with equal weights for both genders to be promoted at every level.

Based on Chart 7A, one can see that the large homophily bias did not have an effect on the number of women at each level over time. This is because with a 0.5 probability of the interviewer being male or female, even if the interviewer is heavily biased towards their own gender there is a relatively similar amount of people from both genders that are heavily biased.

One can see from Chart 7B that a large homophily bias has a greater emphasis when there are 0 level 2 and 3 female employees at the starting point. Even though levels 0 and 1 begin with half of the population being female employees, as time goes on this number decreases because there are less female level 2 interviewers interviewing level 1candidates, and less female level 1 interviewers interviewing level 0 candidates, and less level 0 female interviewers interviewing potential new hires. Therefore, the homophily bias will only be a significant factor of interest when the initial ratios of male:female employees at the company have a large disparity.



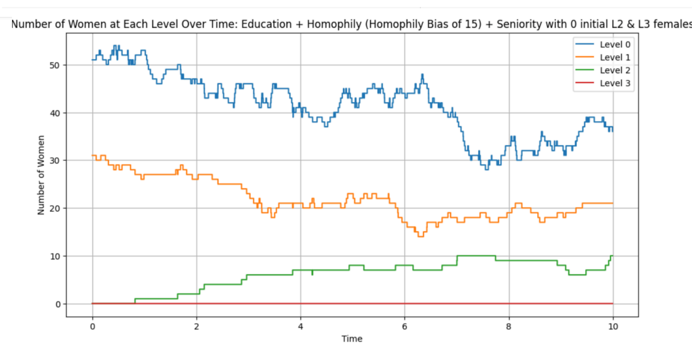


Figure 7A: Homophily bias of 15 with equal starting number of men and women at each level

Figure 7B: Homophily bias of 15 with no women starting at levels 2 and 3

<u>Education Model:</u>

In order to see the impact of education, the last model chose people to promote based on their education. It was run with equal weights for males and females being chosen. Over time, the higher levels quickly became more populated with employees having higher education and less populated with employees having lower education. On the other hand, the amount of employees with education level 1 and 2 in level 0 increased, which makes sense as they were more likely to get chosen to move up to higher positions.

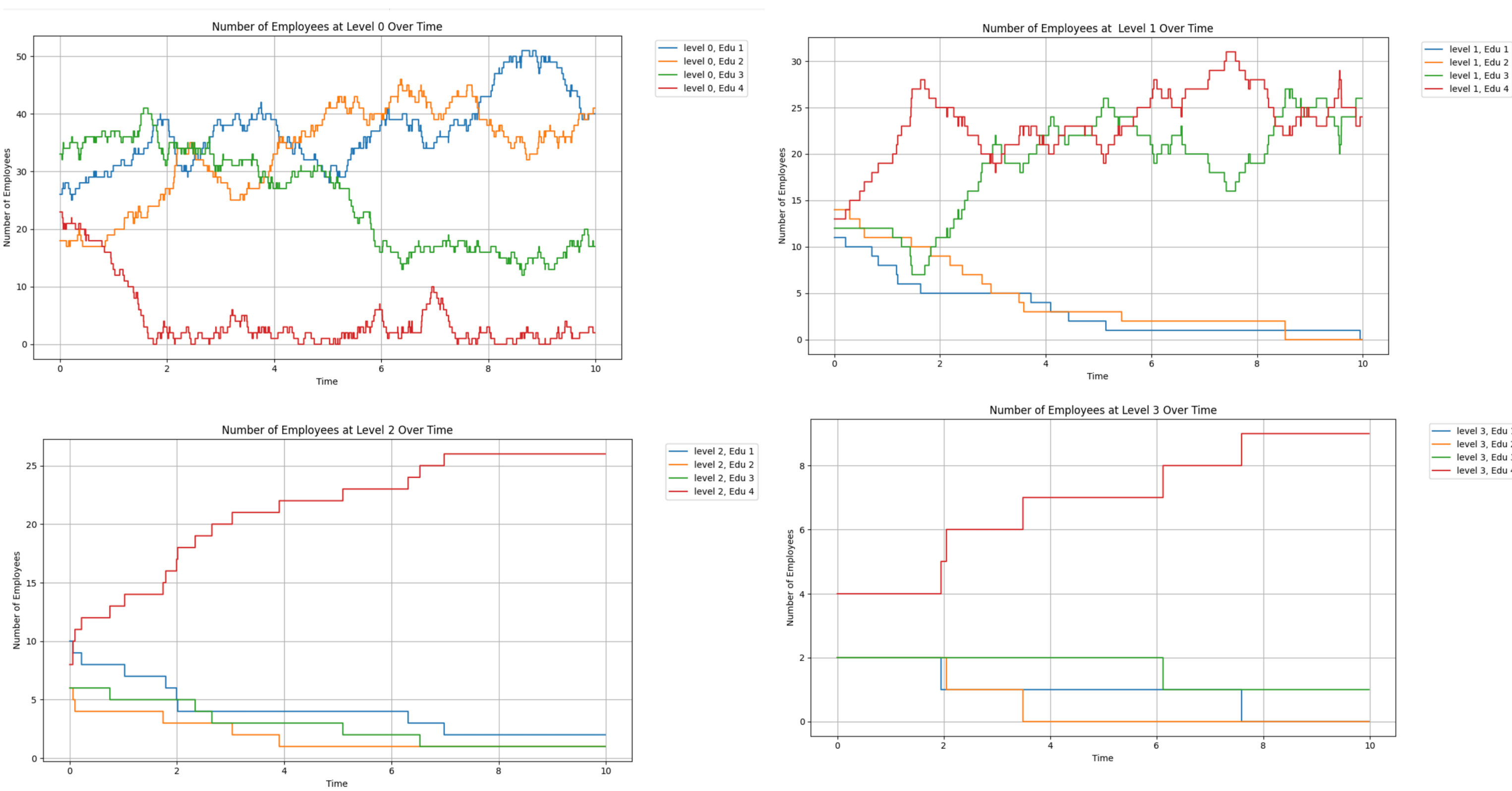


Figure 8: Employees with each Education Level at each Corporate Level and over time in an education-biased model