(Large-)Team Size Reinforces Inequality for Women under Unequal Gender Ratio, Homophily and Triadic Closure

Zeyuan Chen

February 19, 2026

Contents

1 Analysis of a Single Step in Team Assembly

Given a population (with an infinite size) of researchers exhibiting homophily with a specified gender ratio, we ask whether the expected proportion of male members in a team of a certain size differs from the global proportion of males, and how this difference varies with the parameters.

In other words, we aim to express it as a function of team size (T), gender ratio, and homophily. Formally, let the population consist of N_m males and N_f females, with corresponding proportions P_m and P_f. Assume that both males and females have the same tendency for homophily: each individual chooses to collaborate with a same-gender partner with probability s ($s > 0.5$ indicates homophily). The gender ratio is defined as $r = \frac{P_m}{P_f} = \frac{N_m}{N_f}$. We are then interested in computing

$$E\left(\frac{n_m}{T} \,\middle|\, T, r, s\right) - P_m,$$

where n_m denotes the number of male members in a team.

In the seed author model (I believe which is a **stronger** version of Guimera's team-assembly model [1], where the other members are seleted based on a random member among all existing members.), at each step we randomly select a seed author from the population, who is male with probability P_m and female with probability P_f. Next, based on the seed author's gender and the homophily parameter s, we determine the genders of the remaining $T-1$ authors in the team: if a candidate has the same gender as the seed author, they are selected with probability $s \cdot \frac{1}{N}$; otherwise, with probability $(1-s) \cdot \frac{1}{N}$.

Consequently, if the seed author is male, the expected number of males among the remaining $T-1$ members is

$$(T-1) \times \frac{sN_m}{sN_m + (1-s)N_f}.$$

Including the male seed author, the expected proportion of males in a team seeded by a male author is

$$E_{mm} = \frac{1}{T}\left(1 + (T-1) \times \frac{sN_m}{sN_m + (1-s)N_f}\right).$$

Similarly, if the seed author is female, the expected number of males among the remaining $T-1$ members is

$$(T-1) \times \frac{(1-s)N_m}{(1-s)N_m + sN_f}.$$

Including the female seed author, the expected proportion of males in a team seeded by a female author is

$$E_{fm} = \frac{1}{T}\left(0 + (T-1) \times \frac{(1-s)N_m}{(1-s)N_m + sN_f}\right).$$

Therefore, in any given step, when randomly assembling a team of size T, the expected proportion of males is

$$E_m = P_m E_{mm} + P_f E_{fm}.$$

After some calculations, we arrive at the following final formula:

$$E_m - P_m = \frac{T-1}{T}\left(\frac{r\,(1-s)\,(2s-1)\,(r-1)}{(r+1)\,(sr+1-s)\,(r-sr+s)}\right).$$

From this expression, two observations can be immediately made:

1. This function is **always positive**. Under settings with homophily ($0.5 < s < 1$) and unequal gender ratio ($r > 1$), the numerator $r\,(1-s)\,(2s-1)\,(r-1)$ is strictly greater than zero. In other words, regardless of team size T (which is at least 2), as long as there is some degree of gender imbalance and homophily, male members are always more likely than female members to be selected into a team.

2. This function is **monotonically increasing with team size** T. That is, larger teams are more biased toward selecting male members compared to smaller teams.

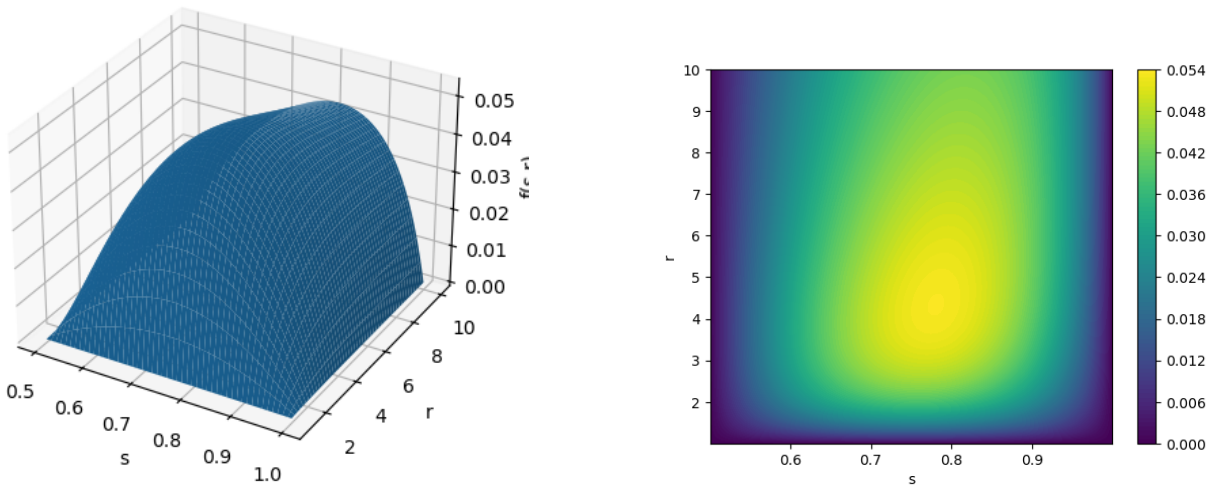
As T only functions as a scaling factor, we could get rid of it and focus on the expression ($f(s, f)$) inside. As Figure 2 shows, gender disparity regarding be assembled into a team is most severely amplified in environments with moderate homophily.

Next, considering studies showing that homophily strength differs by gender, i.e., women researchers exhibit lower homophily than men [2], I allow for different homophily parameters, s_1 for men and s_2 for women. In this case, we have

$$E_m - P_m = \frac{T-1}{T}\left(\frac{r(1-s_2)}{(r+1)(r-rs_2)+s_2)} + \frac{r(s_1-1)}{(r+1)(rs_1-s_1+1)}\right)$$

In fact, the partial derivatives with respect to s_1 and s_2 are consistently positive and negative, respectively. This implies that when men exhibit stronger homophily and women weaker homophily, teams (especially larger ones) become increasingly likely to recruit male researchers. This is intuitive, since homophily only affects how the seed author selects the remaining members; given a certain gender, the optimal strategy under homophily is always to choose collaborators of the same gender.

We are however more concerned with the conditions under which gender disparity can disappear. The contour plot of $f(s_1, s_2, r)$ shows that (when $r = 2/4$), achieving balance requires

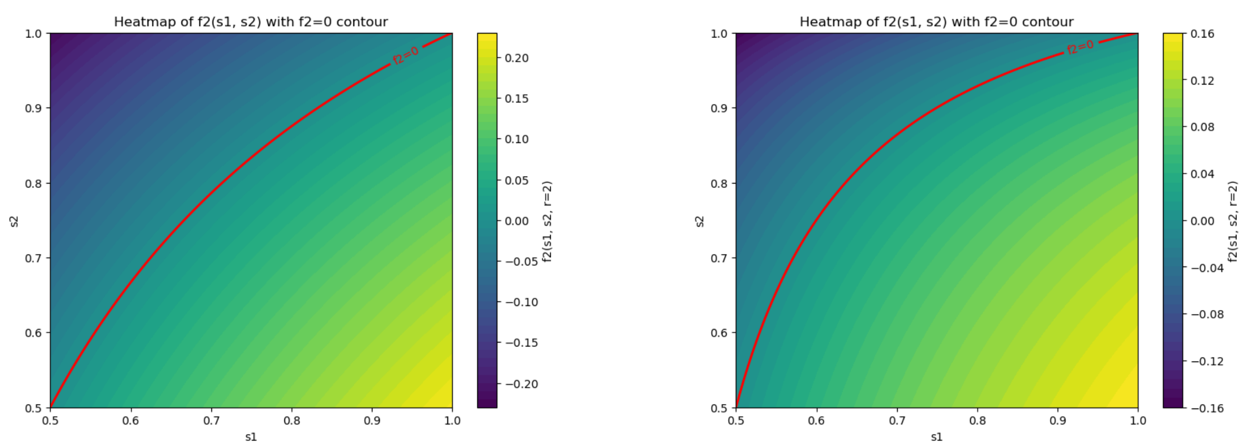


(a) Surface plot of $f(s, r)$ (b) Contour plot of $f(s, r)$

Figure 1: Visualization of the function $f(s, r)$ showing (a) the 3D surface and (b) the contour representation.

women to display slightly stronger homophily than men, in other words, they would need to be more inclined to seek out female collaborators. However, as Torre et al. document, women are generally more willing than men to engage in cross-gender collaboration. Our results thus suggest that this tendency may unintentionally marginalize women's positions in collaboration networks.

Also, we notice that when r increases, women need to display a higher level of homophily to achieve gender balance.



(a) $r = 2$ (b) $r = 4$

Figure 2: Contour plots of the function $f(s_1, s_2)$ when $r = 2$ (a) or 4 (b)

2 Simulation

Following existing studies [3, 4], I computed the following network metrics for 1,000 assembled teams in a network of 2,000 nodes (with females accounting for 40%):

- Segregation of men / women: Measures the level of segregation, defined as

$$\frac{(1-p_a)\,T_{aa}}{p_a(1-T_{aa})+(1-p_a)T_{aa}},\ \text{for } a \in \{m, f\}$$

  The closer to 1, the stronger the segregation. 0.5 indicates no segregation (random).

- Gini index: Measures the inequality of the degree distribution within groups. A higher value indicates a more unequal degree distribution within the group.

- MW-inequality: Measures the inequality between men and women, based on the Mann-Whitney rank test. A value of 0.5 indicates equality; a value below 0.5 indicates that women have, on average, lower degrees than men.

- Avg.Deg-inequality: Also measures the inequality between men and women, based on the ratio of average degree of men and women. A value of 1 indicates equality; a value above 1 indicates that women have, on average, lower degrees than men. Compared with MW-inequality, this metric considers the magnitude of degree distribution.

I tested the following four scenarios:

1. Homophily (Scenario 1): There is no triadic closure (TC) mechanism in the network. Each time members are recruited, they are selected from all remaining nodes in the network. The choice homophily strength is $h = 0.75$.

2. Homophily with Unbiased TC (Scenario 2): Each time a member is recruited, there is a probability of $\tau = 0.25$ that the selection is made from the 2-hop neighborhood, and in this case there is no choice homophily ($h_\tau = 0.5$).

3. Homophily with Biased TC (Scenario 3): Each time members are recruited, there is a probability of $\tau = 0.25$ that the selection is made from the 2-hop neighborhood, and in this case homophily equal to the global homophily is present ($h_\tau = 0.75$).

4. Homophily with Unbiased TC and Preferential Attachment (Scenario 4): Scenario 2 plus preferential attachment model, meaning that the probability of a member being recruited is weighted by its degree at current step.

Table 1: Simulation Results (100 times, avg.T = 3)

Scenario	Seg.Ma	Seg.Fe	Gini.Ma	Gini.Fe	MW-inequity	Deg-inequity
Scenario1	0.6556	0.6905	0.4212	0.4546	0.4384	1.1927
Scenario2	0.6434	0.6754	0.4533	0.4857	0.4459	1.1809
Scenario3	0.6571	0.6930	0.4535	0.4846	0.4437	1.1925
Scenario4	0.6663	0.6535	0.7455	0.7393	0.4715	1.3451

Table 2: Simulation Results (100 times, avg.T = 7)

Scenario	Seg.Ma	Seg.Fe	Gini.Ma	Gini.Fe	MW-inequity	Deg-inequity
Scenario1	0.6444	0.6668	0.3108	0.2822	0.4013	1.2064
Scenario2	0.6258	0.6383	0.3223	0.3519	0.4168	1.1965
Scenario3	0.6507	0.6745	0.3224	0.3495	0.4112	1.2141
Scenario4	0.6617	0.6067	0.7632	0.7536	0.4702	1.4398

Each scenario was tested 100 times, all with an average team size of 3, and the average value of the metrics was obtained. The results are shown in Table 1.

According to our previous findings, the large size of the team could reinforce gender inequality. Therefore, each scenarios were tested 100 times again with an average team size of 7. The results are shown in Table 2.

We have the following observations:

- When team size increased from 3 to 7, segregation between genders decreased slightly; the Gini coefficient showed a significant decline, indicating a flatter overall distribution of degrees. In other words, larger teams help narrow the gap in collaborative networks between non-academic stars and academic stars. However, inequality between genders surged dramatically: in the first three scenarios, MW-inequity decreased by 0.03. After introducing the preferential attachment mechanism, although MW-inequity did not show a significant decline, the ratio of average degrees increased markedly (because the preferential attachment mechanism elevated the rankings of some advantaged females, but widened the gap with advantaged males).
- Unbiased triadic closure functions as a moderating mechanism that can slightly alleviate gender segregation and inequality. However, when triadic closure itself is influenced by choice homophily, this moderating effect is weakened or even eliminated. It is important to note that triadic closure increases the Gini coefficient, as it reduces segregation

and gender inequality by increasing connections between advantaged females and the male group.

- The preferential attachment mechanism causes a significant increase in the Gini coefficient, which is consistent with expectations. Simultaneously, it widens the average degree disparity between genders, potentially leading to a ceiling effect. Finally, prior to introducing the preferential attachment mechanism, females exhibit a higher degree of segregation than males. However, after its introduction, males exhibit stronger segregation than females, despite both possessing identical choice homophily. This may explain why females exhibit lower homophily in real-world scenarios.

3 Mean-Field Modelling of the Network Evolution

3.1 Deriving

To get some clues about the mathematics underpinning the simulation results above, in this section, we derive the mean-field model for a team-network with choice homophily (same for male and female) and unbiased triadic closure. We consider a network in which the evolution depends on the probabilities of edges connecting nodes of different genders, denoted as

$$\{p_{mm}, p_{mf}, p_{ff}\},$$

with

$$p_{mm} + p_{mf} + p_{ff} = 1,$$

representing the probability that a randomly selected edge in the network connects two nodes of a particular gender combination.

From this set of probabilities, we can define a gender transition matrix TM, where TM_{mf} represents the probability that an edge from a male node connects to a female node. Our goal is to track the evolution of p_{mm} and p_{ff} over time to determine whether they reach a steady state.

To simplify the analysis, we assume the total number of edges L in the network is constant. This implies

$$p_{mm} = \frac{N_{mm}}{L} \propto N_{mm},$$

so that tracking

$$\frac{dp_{mm}}{dt} \propto \frac{dN_{mm}}{dt}.$$

This assumption requires that when assembling a team of size T, existing edges are simultaneously removed. Specifically, we assume that the T members of the team randomly delete $T-1$ existing edges, which can be interpreted as previous collaborations that have broken.

Let the fractions of males and females in the population be p_m and p_f, respectively. Denote the expected number of male members in a team seeded by a male as k_{mm}, and in a team seeded by a female as k_{fm}. Given the transition matrix TM and team size T, we have

$$\begin{aligned}\frac{dp_{mm}}{dt} \propto p_m\left(\frac{1}{2}k_{mm}(k_{mm}-1) - \frac{1}{2}k_{mm}\,TM_{mm}(T-1)\right)\\ + p_f\left(\frac{1}{2}k_{fm}(k_{fm}-1) - \frac{1}{2}k_{fm}\,TM_{fm}(T-1)\right)\end{aligned} \tag{1}$$

where

$$k_{mm} = E_{mm}(T-1)+1, \tag{2}$$

$$k_{fm} = E_{fm}(T-1). \tag{3}$$

Let $h \in [0.5, 1)$ denote the homophily strength, and $\tau \in [0,1]$ the probability of forming edges via (unbiased) triadic closure. Then, the expected fraction of males in a male-seeded team is

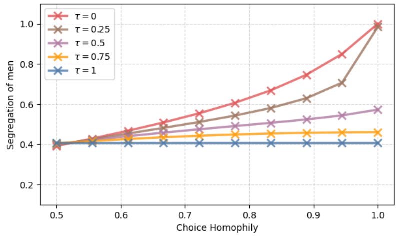
$$E_{mm} = \tau(TM^2)_{mm} + (1-\tau)\frac{hp_m}{hp_m + (1-h)p_f}, \tag{4}$$

and in a female-seeded team is

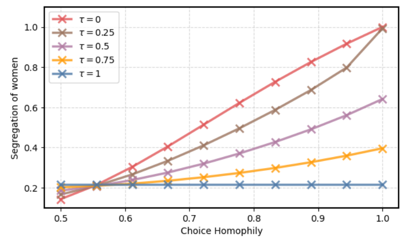
$$E_{fm} = \tau(TM^2)_{fm} + (1-\tau)\frac{(1-h)p_m}{(1-h)p_m + hp_f}. \tag{5}$$

3.2 Computing

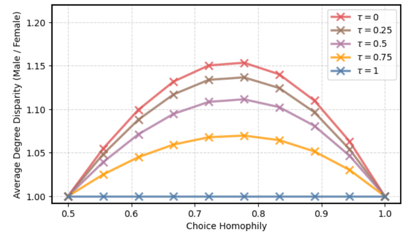
Figure 3 presents the numerical results of the mean-field model. First, we consider a range of parameter combinations in which the strength of choice homophily varies from 0.5 to 1, and the probability of activating triadic closure, τ, varies from 0 to 1. In each scenario ($p_m = 0.7, p_f = 0.3$), with initial values set to $p_{mm}, p_{ff} = 0.49, 0.09$, the equations converge to a fixed point. Second, to examine the effect of team size on network evolution, we compute



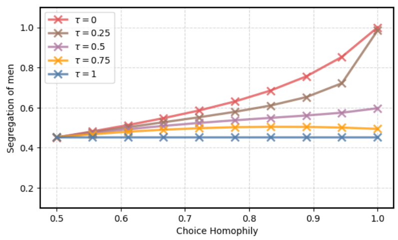
(a) Ma. Segregation (T=3)



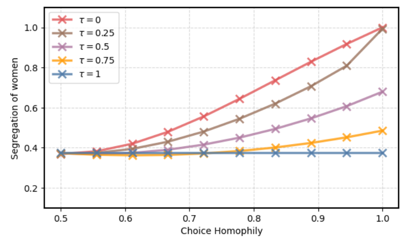
(b) Fe. Segregation (T=3)



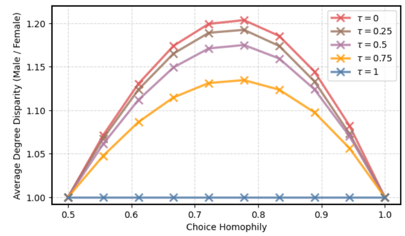
(c) Avg. Degree Diff (T=3)



(d) Ma. Segregation (T=7)



(e) Fe. Segregation (T=7)



(f) Avg. Degree Diff (T=7)

Figure 3: Average Field Calculation Results of Gender Disparity in Average Degree, and Degree of Segregation of Both Genders (Team Size = 3 / 7)

the network metrics separately for $T = 3$ and $T = 7$, representing small-team and large-team collaboration cultures, respectively.

Finally, we calculate three indicators: the normalized segregation level of men, the normalized segregation level of women, and the ratio of average degrees between the two genders (men relative to women). Specifically, in the mean-field framework, these three indicators are defined as follows:

$$\text{Segregation}_a = \frac{(1-p_a)\,T_{aa}}{p_a(1-T_{aa}) + (1-p_a)T_{aa}}, \quad \text{for } a \in \{m, f\}$$

$$\text{AvgDegRatio} = \frac{(2P_{mm} + P_{mf})(1-p_m)}{(2P_{ff} + P_{mf})p_m}$$

From the numerical analysis, we obtain results consistent with previous simulations and existing studies:

- First, it is evident that, regardless of team size, when $\tau < 1$ (when $\tau = 0$, only triadic closure is at work), choice homophily is positively correlated with the level of segregation. This positive relationship weakens as τ increases; in other words, triadic closure mitigates the segregation caused by choice homophily.

- Second, considering the ratio of average degrees, gender inequality in average degree is most pronounced at a moderate level of choice homophily, which is consistent with previous findings from the single-step team assemblage analysis. On the other hand, an increase in τ reduces gender inequality, further supporting the conclusion that triadic closure functions as a regulatory mechanism.
- Finally, when comparing the cases of $T = 3$ and $T = 7$, we observe that increasing team size both raises the levels of inequality, and weakens the moderating effect of the triadic closure mechanism. For example, in the figure of average degree differences, when $T = 3$, the gap between the curves for $\tau = 0.75$ and $\tau = 0$ is wider than that for $T = 7$. This finding is also consistent with previous simulation experiments.

References

[1] Roger Guimera et al. "Team assembly mechanisms determine collaboration network structure and team performance". In: *Science* 308.5722 (2005), pp. 697–702.

[2] Margarita Torre, Jesús A Prieto-Alonso, and Iñaki Ucar. "The uneven effects of gender parity: Trends in gender homophily in scientific publications, 1980–2019". In: *Social Science Research* 132 (2025), p. 103228.

[3] Aili Asikainen et al. "Cumulative effects of triadic closure and homophily in social networks". In: *Science Advances* 6.19 (2020), eaax7310.

[4] Jan Bachmann et al. "Network Inequality through Preferential Attachment, Triadic Closure, and Homophily". In: *arXiv preprint arXiv:2509.23205* (2025).