(Large-)team exclusion of women researchers under unequal gender ratio and homophily

Zeyuan Chen

January 26, 2026

Question: 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, 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

$$E_m=P_mE_{mm}+P_fE_{fm}.$$

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

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

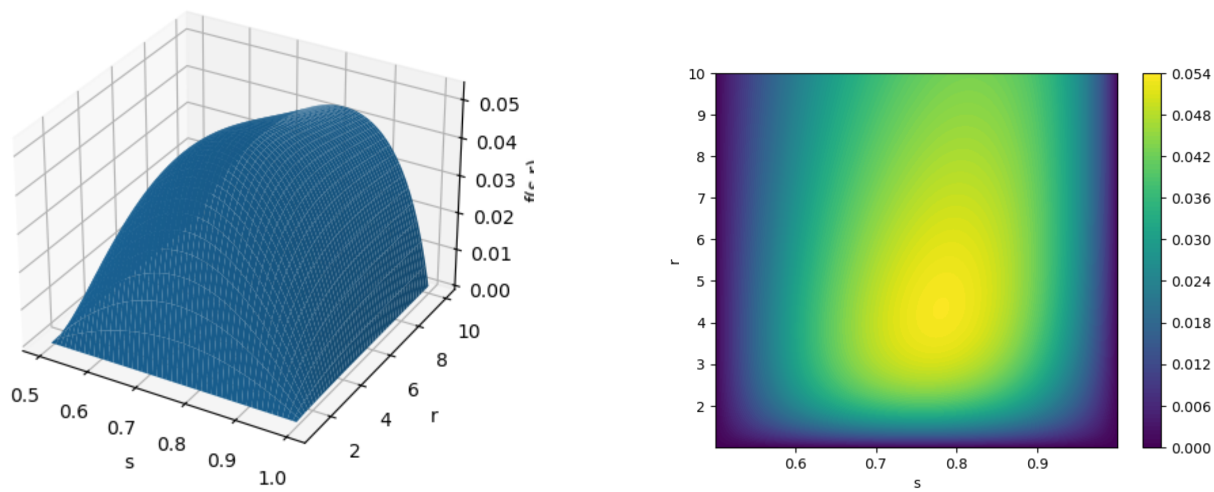
From this expression, two observations can be immediately made:

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 1 shows, gender disparity regarding be assembled into a team is most severely amplified in environments with moderate homophily.



(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.