9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Gender Disparities in Brokerage of Scientific Collaboration

Keywords: gender inequalities, brokerage, triadic closure, temporal motifs, scientific collaboration

Extended Abstract

In academia, structural inequalities between women and men still persist today. For instance, women have a higher dropout rate in certain fields and they participate in less successful, but more homophilic collaborative patterns [1]. Modelling scientific collaboration as link formation between co-authors provides a formal framework to study social mechanisms and their effects. In particular, an abundance of same-gender links and triangles have been identified in scientific collaborations [1]. These characteristic structures are often explained by social mechanisms, such as homophily and triadic closure. Homophily describes the tendency of individuals to interact with others similar to them, while triadic closure refers to people's tendency to connect to existing contacts of their direct peers, i.e., friends-of-friends. An important aspect to which gender disparities in these two mechanisms could propagate is that of brokerage. In social networks, brokers are considered to take the role of forming bridges across structural holes [2], that is, between otherwise distant actors or communities. As soon as the distant individuals are introduced to one another through the broker, they may interact and hence form a triangle in the network. In the context of scientific collaboration, novel connections might foster new ideas, spread knowledge and provide access to funding opportunities. Albeit the importance of brokers in mediating social capital, gender disparities in brokerage have mainly been studied ignoring the temporal order of links or at small scales. Applying the lens of temporal networks allows us to understand if men and women execute this role differently and how it affects the subsequent careers of the broker and the newly connected pairs of scientists. Here, we want to close this gap by studying gender-disparities in brokerage through the mechanisms of triadic closure and homophily in a temporally growing scientific collaboration network.
Methods. To explore this relationship, we study the temporal formation of triangles in a large-scale growing co-authorship network from over 600 000 articles published by the American Physical Society (APS). In this dataset, 34 596 (15%) women and 195 490 (85%) men co-authoring a paper between 1893 to 2020 are connected pairwise at the time of publication. In this setting, we capture the formation of triangles via 16 temporal network motifs. Temporal motifs are equivalence classes of small subgraphs, defined by the number of participating nodes and the temporal order of their links [3]. If two authors a and c who share a mutual former colleague b collaborate, they close a triangle in the network (see Figure 1.A). We focus on motifs involving three authors and their publication activities up to the point where all three of them have collaborated. Moreover, we enforce a temporal separation between each of the three publications ($t_1 < t_2 < t_3$) to identify the broker b among the three participating authors. Differentiating whether all three authors jointly collaborated on the final closure paper, yields two motif types (see Figure 1.A). Adding the authors' gender as binary node labels and propagating it over the three nodes produces a total of 16 motifs (see Figure 1.B).

To understand if the proposed classification is capable of capturing social mechanisms, we examine the formation time and expected occurrence of different motif types. Following the motif specification, we define the opening time of a triangle as the time between the first two collaborations $\Delta t_{open} = t_2 - t_1$ and its closure time as the time delta between the last two links $\Delta t_{close} = t_3 - t_2$. Motif counts and related metrics are typically compared to distributions inferred from null models which maintain some characteristics of the observed system while randomizing everything else. Gauvin *et al.* [4] recently proposed a taxonomy for randomizing temporal networks based on the fixed features. We plan to utilize this taxonomy to systematically study the potential contribution of varying characteristics of the whole system to the observed motif counts. For a start, we create 500 permutations by shuffling the gender attribute across all authors. As this preserves everything besides the individual gender assignment, we use it to test whether the observed motif counts depend solely on the disparity in the overall gender group sizes (see [5] for a similar application). From the permutations, we infer distributions of motif counts (or other metrics) with regards to the specified null model.

Results In Figure 2, we report the ratio difference of the two time intervals Δt_{open} and Δt_{close}. The value is zero if both intervals are identical, it tends towards $+1$ if the triangle closes faster than it opens, and to -1 otherwise. Reporting the distribution over all motifs, dissected by motif type and the gender of the broker, we find no strong disparity between men and women. More notably, however, there is a clear distinction between the two motif types. When the broker participates in the final closure event ($M_{(a-b-c)}$), the closure time is generally faster. This hints towards the presence of a social mechanism captured by the motif specification.

Motivated by this finding, we test whether the identified motif counts could be explained by the disparity in gender-group sizes. We compute z-scores to compare the observed motif counts to the distribution retrieved by the gender-label permutations and find that most scores strongly differ from zero ($|z| > 0$, see Figure 2). Thus, we conclude that motifs do not only depend on the group size disparity. This observation is by no means a surprising finding, given the likely presence of other dynamics such as changes of the group sizes and their activity in time. However, the proposed motifs allow us to systematically test these mechanisms by designing more elaborate null models which motivates the next steps. Untangling these mechanisms will reveal how men and women differ in mediating their social capital in collaboration and ultimately contribute to understanding why women are often still disadvantaged in the academic system.

References

1. Jadidi, M., Karimi, F., Lietz, H. & Wagner, C. Gender Disparities in Science? Dropout, Productivity, Collaborations and Success of Male and Female Computer Scientists. *Advances in Complex Systems* **21,** 1750011 (Nov. 1, 2017).
2. Burt, R. S. *Structural Holes: The Social Structure of Competition* 324 pp. ISBN: 978-0-674-84371-4 (Harvard University Press, Cambridge, MA, Nov. 8, 1995).
3. Kovanen, L., Karsai, M., Kaski, K., Kertész, J. & Saramäki, J. Temporal Motifs in Time-Dependent Networks. *Journal of Statistical Mechanics: Theory and Experiment* **2011,** P11005 (Nov. 2011).
4. Gauvin, L. *et al.* Randomized Reference Models for Temporal Networks. *SIAM Review* **64,** 763–830 (Nov. 2022).
5. Kovanen, L., Kaski, K., Kertész, J. & Saramäki, J. Temporal Motifs Reveal Homophily, Gender-Specific Patterns, and Group Talk in Call Sequences. *Proceedings of the National Academy of Sciences* **110,** 18070–18075 (Nov. 5, 2013).

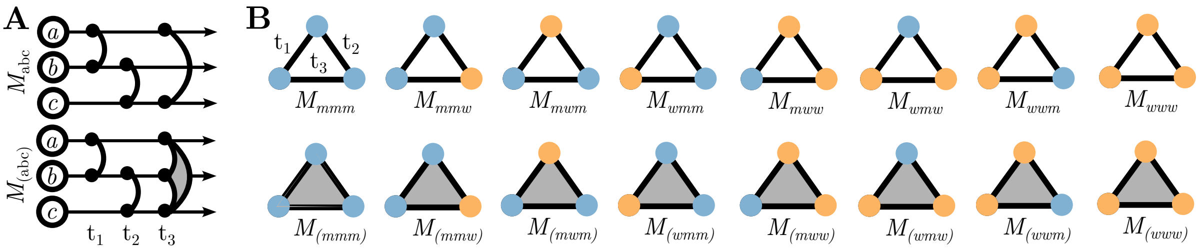

Figure 1: (**A**) Temporal triadic closure between nodes a, b and c over three distinct points in time $t_1 < t_2 < t_3$. Here, node b is the broker, and we differentiate the closure depending on whether b participates in the last collaboration (bottom) or not (top). To capture the initiation of each contact, we count only the first occurrence of each link. (**B**) Including all possible gender-compositions, we distinguish a total of 16 motifs. Note that cases of semi-distinct temporality ($t_i \leq t_j$) are left out for simplicity here, but will be considered in the final manuscript.

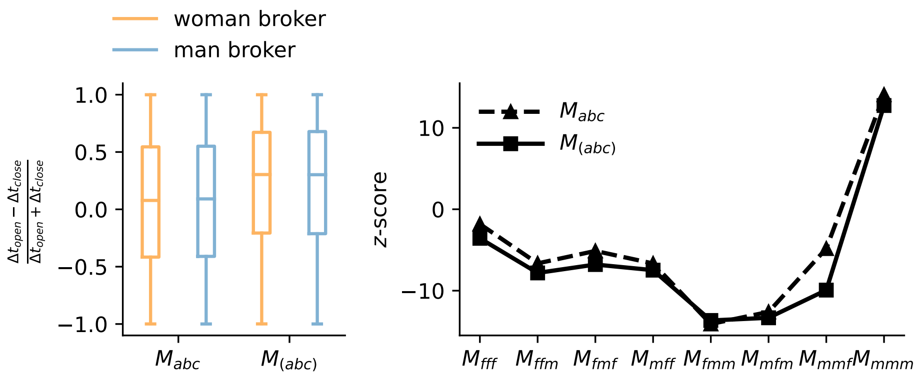

Figure 2: (Left) Comparison between triangle opening and closure time. Values of zero indicate equal intervals. Values towards $+1$ (-1) indicate that the closing occurs faster (slower) than the opening. The x-axis differentiates whether the broker b participates in the final closure event. Colors indicate the gender of the broker b, where orange (blue) indicates a woman (man) broker. Irrespective of the broker's gender, triangles close faster when b participates in the closure. (Right) The 16 z-scores (connecting lines for visual aid). A value of zero indicates that the null hypothesis could not get rejected. As most values are well above or below zero, we conclude that motif counts do not solely depend on the difference in gender-group sizes.