Team formation: expected or unexpected

Seokkyun Woo1, Zejian Lyu2, Oh-Hyun Kwon3, Noly Higashide4
1Northwestern University, 2University of Chicago, 3Pohang University of Science and Technology, 4 The University of Tokyo

Keywords: Team science, Co-author network

Extended Abstract

At the “record-breaking” ICSSI 2023, Aaron Clauset stated the importance to study the interaction of scientists to understand more about epistemic inefficiency or inequality in science. Understanding the impact of co-authorship interactions would provide valuable insights into the scientific discovery processes as well as the science ecosystem. It could also lead to fresh ideas and approaches for creating a more diverse and innovative scientific community. Looking at interactions from an individual perspective, co-authorship with top researchers has more impact on the number of citations than individual ability [1]. Superstrong ties, co-authorship over a substantially long period, also have a substantial and positive effect on productivity and citations [2]. From the team perspective, in contrast, teamwork has more impact than solo work and the impact of a team is determined more by the less-cited members rather than the highly-cited members [3].

We hypothesized that team compositions are typically characterized by expected relationships, such as collaborations between individuals in the same field or country, with top researchers, or as mentor-mentee pairs. However, what about relationships that are unexpected, namely outside of these typical patterns, such as collaboration with non-top researchers, or strangers? Are teams in unexpected formations more successful compared to teams in expected formations? Here, we specifically focused on determining the likelihood of a team combination being formed by examining the structure of co-authors' networks.

We used SciSciNet data to investigate how the likelihood of team formation affects the impact, disruption, and novelty of the papers in the field of biology, chemistry, psychology, and sociology. In this project, we took two approaches to measuring the proximity score from 2001 to 2010, to connect it with the collaborations in 2011 and 2012. First, we calculated the shortest path distances between authors within collaboration teams. Second, we get author proximity from all pair of authors for each paper. If the paper is written by authors who newly joined, the average author proximity is lower value which capture the team formation is unexpected (Figure 1). In contrast, if authors have collaborated before, in this case time window of 2001-2010, the score of the paper is high.

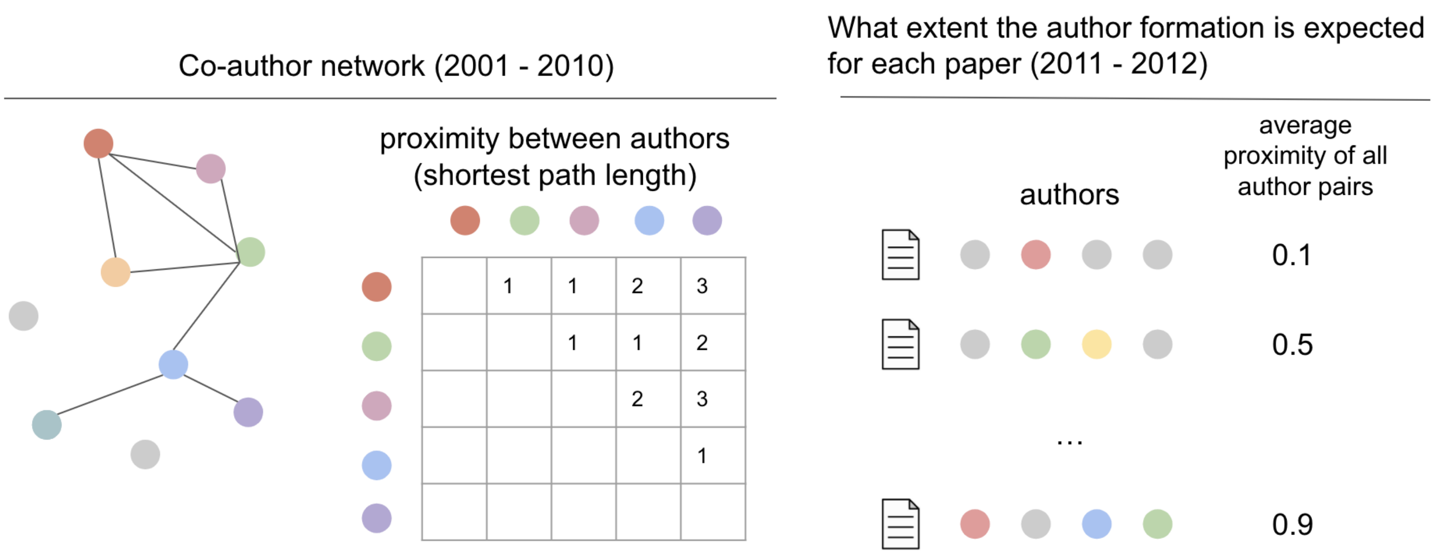

Figure 1 Method to calculate author proximity

Figure 2 shows the relation between the author proximity and the number of citation. The patterns differ depends on discipline. For chemistry and biology, the pattern shows the inverse U-shape. The papers with lower or higher author proximity have a relatively lower impact while the papers with moderate proximity have a higher impact. For psychology, impact increase as proximity increase and higher impact remains over 0.7 of proximity. These indicate that unexpected collaboration with newly joined authors is less likely to correlate positively with citation success. Note that sociology shows no clear dependency pattern.

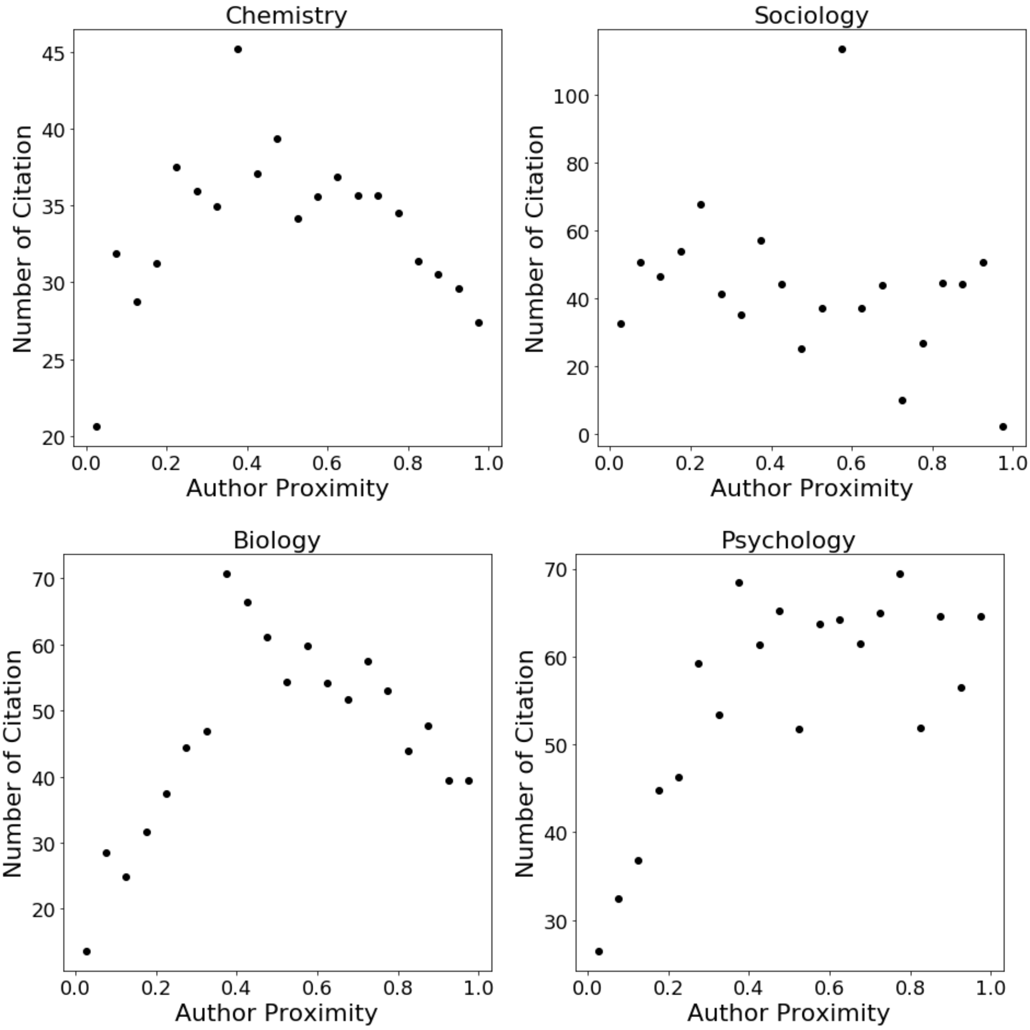

Figure 2 Average citation count distribution binned author proximity

Why does the team with moderate proximity of authors more likely to succeed? Teaming with strangers might take a larger cost to communicate. Maybe the unchanged teams that have been collaborating for a long time are getting harder to produce impactful work.

We uncovered interesting patterns and further investigation is required. Promising questions are the following.

- What is the role of authorship space in facilitating knowledge recombination?
- How does this differ in fields? In particular, with the increasing team science and the division of labor, do repeated ties matter more for lab-based knowledge production?
- How is the authorship network related to actual collaboration? Can they predict future collaboration, or are there factors that encourage unlikely authors to collaborate, or prevent them from working?

References

[1] Li, W. et al. Untangling the network effects of productivity and prominence among scientists. *Nature Communications* 13, 4907 (2022).

[2] Ahmadpoor, M. and Jones, B. F. Decoding team and individual impact in science and invention. *Proceedings of the National Academy of Sciences* 16 (28) 13885-13890 (2019).

[3] Petersen, A. M. Quantifying the impact of weak, strong, and super ties in scientific careers. *Proceedings of the National Academy of Sciences* 112 (34) E4671-E4680 (2015).

Appendix

sociology

	Year	Citation_Count	Disruption
first_TL	-0.017396	0.124557	-0.056096
last_TL	-0.022537	0.123303	-0.053171
avg_TL	-0.019558	0.126771	-0.055171
homo	-0.036026	-0.146225	0.053964

4 rows × 3 columns Open in new tab

	Year	Citation_Count	Disruption
first_TL	0.015498	-0.003064	0.009607
last_TL	0.121347	-0.140418	0.010592
avg_TL	0.067619	-0.052971	0.017983
homo	-0.068538	0.303382	-0.154453

4 rows × 3 columns Open in new tab

chemistry

	Year	Citation_Count	Disruption
first_TL	-0.042658	-0.006147	0.004733
last_TL	0.022763	0.040676	-0.010797
avg_TL	-0.016843	0.009341	0.001065
homo	-0.009841	0.098070	-0.010693

4 rows × 3 columns Open in new tab

	Year	Citation_Count	Disruption
first_TL	0.035459	0.112892	-0.085410
last_TL	-0.095470	-0.223799	0.037114
avg_TL	0.114547	-0.093725	-0.119120
homo	-0.073613	0.108324	0.165796

4 rows × 3 columns Open in new tab

psychology

	Year	Citation_Count	Disruption
first_TL	-0.040848	-0.047309	-0.002401
last_TL	-0.013279	-0.080583	-0.006873
avg_TL	-0.031935	-0.078999	-0.001181
homo	0.005032	0.107838	-0.024830

4 rows × 3 columns Open in new tab

	Year	Citation_Count	Disruption
first_TL	-0.165192	-0.101680	0.002892
last_TL	-0.137656	0.024215	-0.032593
avg_TL	-0.178284	-0.064933	0.004600
homo	-0.039578	-0.045394	-0.050591

4 rows × 3 columns Open in new tab

biology

	Year	Citation_Count	Disruption
first_TL	-0.010775	-0.061455	0.017707
last_TL	0.010844	-0.059049	0.021819
avg_TL	0.004097	-0.092738	0.031014
homo	-0.022817	0.170597	-0.061065

4 rows × 3 columns Open in new tab

	Year	Citation_Count	Disruption
first_TL	0.147230	-0.092892	0.023256
last_TL	0.161163	-0.190246	-0.079472
avg_TL	0.111470	-0.212162	-0.065123
homo	-0.070141	0.047284	0.048770

4 rows × 3 columns Open in new tab

Figure 3 Correlation between paper attribute and team likelihood calculated by different methods