Social Contagion in Science

Keywords: science of science, social contagion, collaboration networks, information diffusion, computational social science

Extended Abstract

Modern science has become increasingly collaborative over the past decades. A powerful representation of the collaborative nature of science is given by a collaboration network, in which nodes are authors, and two nodes are connected if they have coauthored at least one paper (see Fig. 1B). With the growing availability of bibliometric data, collaboration networks have been extensively studied, and their structural properties are now well known [1, 2]. In particular, the interactions spurred by collaboration favor the circulation of knowledge and ideas between scientists. Coauthors often expose us to new tools, methods, and theories, even when the latter is not being used for the team's specific project. Once scholars are made aware of new research topics, they may decide to work on them in the future. While switching topics is a scientist's free choice, we argue that other scientists may effectively induce it by influencing their coauthors. In this respect, topic switches can result from a social contagion process, where scholar A, who spreads the new topic, influences scholar B to adopt it. The process is similar to the spreading of an epidemic, where *infected* individual A exposes the *susceptible* individual B to a disease, and B has a certain probability of being infected. In the case of an idea or a topic, the infection spreads if B adopts the new idea or works on the new topic.

Model setup and data

We consider a topic t, a scholar I who is an expert in t (the *influencer*), and a scholar S who has never worked on t before a certain time T_0 (the *influencee*). We want to quantitatively assess whether I can induce a coauthor S to start working on t after time T_0. We identify two windows before and after a reference year T_0. The duration of each time window is $T = 5$ years. The earlier range $[T_0 - T, T_0)$ is the *exposure window* (EW), where we keep track of the interactions among the authors in terms of papers written together. The later range $[T_0, T_0 + T)$ is the *observation window* (OW), where we look at the effect of these interactions. For the given topic t, we identify all authors who have produced papers tagged with this topic in the EW. These are the *active authors*. For *e.g.*, in Fig. 1B, authors a_0, a_1, a_4, and a_5 are active. Likewise, *inactive authors* are those who did not publish any paper on topic t before time T_0. We look at *productivity* as a proxy measure to quantify an active author's potential to influence others. It is defined as the count of papers on topic t produced by an author in the EW.

We use the OpenAlex dataset and report the results for 6 topics, *viz.*, String Theory, General Relativity, Cryptography, Support Vector Machine, Microbiome, and Genome across three fields: Physics, Computer Science, and Biology & Medicine.

Results and Conclusions

In our analysis, we want to calculate the probability that an inactive author S becomes active in the OW if they had a certain number of contacts with active authors in the EW. In Fig. 1D, we plot the inverse cumulative probability $C(k)$ that an inactive author becomes active in the OW as a function of the number k of active contacts in the EW [3]. We first focus on the first part of the plot. As expected, we see an increasing trend. In particular, the jump from $k = 0$ to $k = 1$ is remarkable, showing that the probability of *spontaneous* activation, in the absence of previous contacts with active authors ($k = 0$), is much lower than that of induced activation via contagion ($k \geq 1$). The higher the number of contacts, the larger the activation probability. Most growth occurs for low values of k, after which the curve flattens.

Next, we assess whether the cumulative influence of active authors can be considered the result of a process in which they act independently of each other (*simple contagion*) or not (*complex contagion*). We derive the baseline curve corresponding to independent actions (teal dashed line). If the number of contacts is low, the process is consistent with simple contagion. As the number of contacts grows, the empirical curve differs from the baseline, supporting the complex contagion hypothesis.

Now that we find evidence of contagion, we check if it depends on the features of the influencers. Specifically, we want to see whether it matters that the inactive author had contacts with prominent authors, which we define based on productivity on topic t. We create two pools of active authors in the EW: *prominent* authors, the top 10% most productive authors, and *baseline*, a random sample of 10% of the remaining active authors. To mitigate confounding effects, we only consider those inactive authors who had contact with prominent authors or the baseline during the EW, but not both. We now revisit Fig. 1D, focusing on the bottom half. The red circles and blue squares indicate the activation probabilities when the contacts are prominent authors and the baselines, respectively. We find a clear separation between the red and blue curves in each case, indicating the higher effectiveness of prominent authors in influencing their inactive coauthors. Like the plots above, the red curves increase monotonically with an increasing number of contacts. The baseline curves, however, display interesting behavior for String Theory, where the probabilities decrease as k increases. Upon closer inspection, it becomes clear that it results from incomplete summary statistics, as String Theory is the smallest of all the topics.

In conclusion, we shed light on complex information diffusion dynamics in collaboration networks by studying how peer pressure renders future topic switches. Furthermore, we examined how the prestige of influencers positively affects their ability to influence others. This study also paves the way for a more systematic analysis covering topics from various disciplines, incorporating other scientometric indicators of scholarly output, like average citations and the h-index.

References

[1] M. E. Newman, “The structure of scientific collaboration networks,” *Proceedings of the national academy of sciences*, vol. 98, no. 2, pp. 404–409, 2001.

[2] R. Guimera, B. Uzzi, J. Spiro, and L. A. N. Amaral, “Team assembly mechanisms determine collaboration network structure and team performance,” *Science*, vol. 308, no. 5722, pp. 697–702, 2005.

[3] G. Kossinets and D. J. Watts, “Empirical analysis of an evolving social network,” *science*, vol. 311, no. 5757, pp. 88–90, 2006.

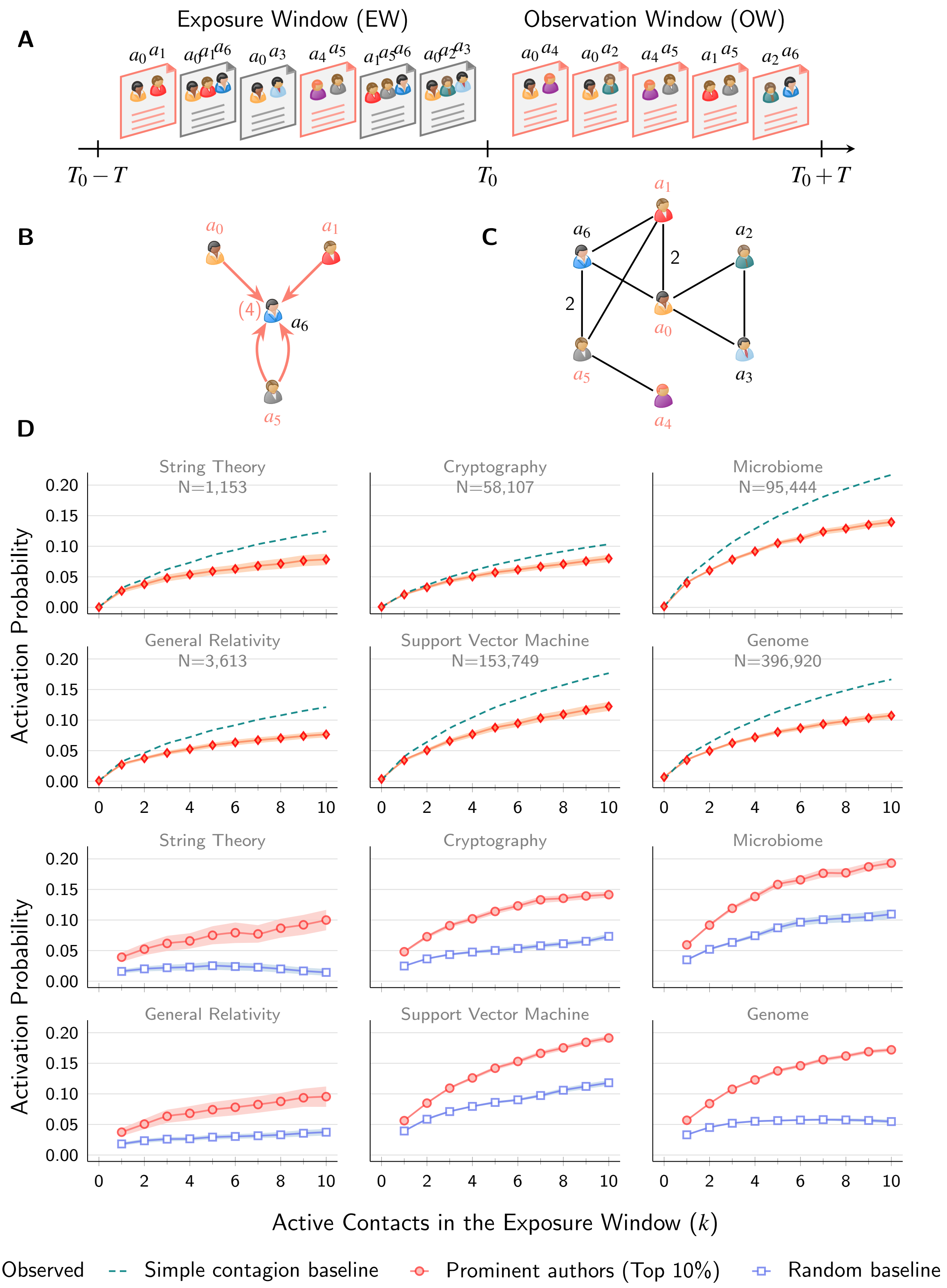

Figure 1: (A) Stream of papers across exposure (EW) and observation (OW) windows. Papers on topic t are in red. (B) Author collaboration graph at the end of EW. (C) Active contact sources (a_0, a_1, and a_5) of an inactive author (a_6) derived from the collaboration graph, resulting in 4 contacts. (D) Above: the activation probability of inactive authors in the OW with shaded standard errors; below: a comparison of the activation probability of prominent authors (red) and random samples (blue). At each k, the y-value indicates the proportion of inactive authors who became active in the OW having had at least k active contacts in the EW. N is the average number of active authors in the OW.