

Discussion of “Co-citation and Co-authorship Networks of Statisticians”

Xiaojing Zhu & Eric D. Kolaczyk

To cite this article: Xiaojing Zhu & Eric D. Kolaczyk (2022) Discussion of “Co-citation and Co-authorship Networks of Statisticians”, Journal of Business & Economic Statistics, 40:2, 494-496, DOI: [10.1080/07350015.2022.2044335](https://doi.org/10.1080/07350015.2022.2044335)

To link to this article: <https://doi.org/10.1080/07350015.2022.2044335>



Published online: 21 Apr 2022.



Submit your article to this journal [↗](#)



Article views: 326



View related articles [↗](#)



View Crossmark data [↗](#)



Discussion of “Co-citation and Co-authorship Networks of Statisticians”

Xiaojing Zhu and Eric D. Kolaczyk

Department of Mathematics and Statistics, Boston University, Boston, MA

1. Introduction

We thank the authors for their new contribution to a high quality dataset and interesting findings from the modeling and analysis of the co-citation and co-authorship networks of statisticians. Leveraging this dataset, there are lots of additional questions that might be answered, and analyses done. Network motif analysis is one such, with roots in the triad census of traditional social network analysis (Wasserman and Faust 1994, chap. 14.2.1) and first introduced in its modern form by Milo et al. (2002) in systems biology. It has since been applied to various scientific domains, for example, social science, neuroscience, to study network structures and the underlying complex systems (see Stone, Simberloff, and Artzy-Randrup (2019) for a survey article).

While the notion of network motif was originally defined for static networks as small subgraph patterns occurring frequently in a given network, several ways have been proposed to extend it to dynamic networks consisting of a set of vertices and a collection of timestamped edges. One widely used one is from Paranjape, Benson, and Leskovec (2017), where temporal motifs are defined as an ordered sequence of timestamped edges among a subset of nodes conforming to a specified pattern as well as a specified duration of time δ in which the edges must occur. In contrast to their static counterparts, such temporal motifs take into account not only subgraph isomorphism but also edge ordering and duration, which can be regarded as the simple building blocks for temporal structures of dynamic networks.

There are a few works in the literature on motif analysis for journal citation networks (Wu, Han, and Li 2008; Zeng and Rong 2021) and author collaboration networks (Chakraborty, Ganguly, and Mukherjee 2015), but none of them seem to be from the perspective of temporal motifs. In this discussion, we construct temporal citation networks among statisticians using the publication data provided in the article, and focus on analyzing the frequency and distribution of temporal motifs in such dynamic networks. This analysis provides initial insights into the temporal patterns of citing behaviors among authors of various statistics journals from 1975 to 2015.

2. Definition of Temporal Author Citation Networks

The co-citation and co-authorship networks studied in the article are in the form of matrices or time series of matrices. Here, we consider networks in a different form consisting of a set of vertices and a collection of timestamped directed edges.

The dataset provided in the article contains two data resources: one is the citation records consisting of pairs of citing and cited article; the other is the article related information including authors, year and journal name of publication for each article. Using both sources of information, we construct a temporal author citation network as a series of timestamped directed edges, where nodes represent authors and a directed timestamped edge pointing from a citer to a citee represents that an author cited at least one article from another author within a given year. We exclude all self-loops. The resulting network includes 1,768,050 citing interactions among 43,521 authors from year 1975 to year 2015.

We also construct temporal networks in the same manner for the following three different subcategories of journals: (a) the four flagship journals consisting of AoS, Bka, JASA and JRSSB; (b) the IMS family, consisting of AoAS, AoP, AoS, StSci, EJS and JCGS; and (c) the RSS family, consisting of JRSSA, JRSSB and JRSSC (see Table B.1 in the article for full journal names), where only citing interactions among authors and their articles published within each journal subcategory are considered for each of the three temporal networks constructed. The resulting networks for the three journal subcategories contain 230,821, 93,180 and 25,940 citing interactions among 8998, 7363 and 4001 authors, respectively, from year 1975 to year 2015.

With all these networks, we are interested in understanding the temporal patterns of citing behavior among authors and its change over time.

3. Temporal Motif Analysis Results

For each of the constructed temporal networks, we use the `snap`¹ package to count the number of occurrences of several

¹<https://snap.stanford.edu/temporal-motifs/code.html>.

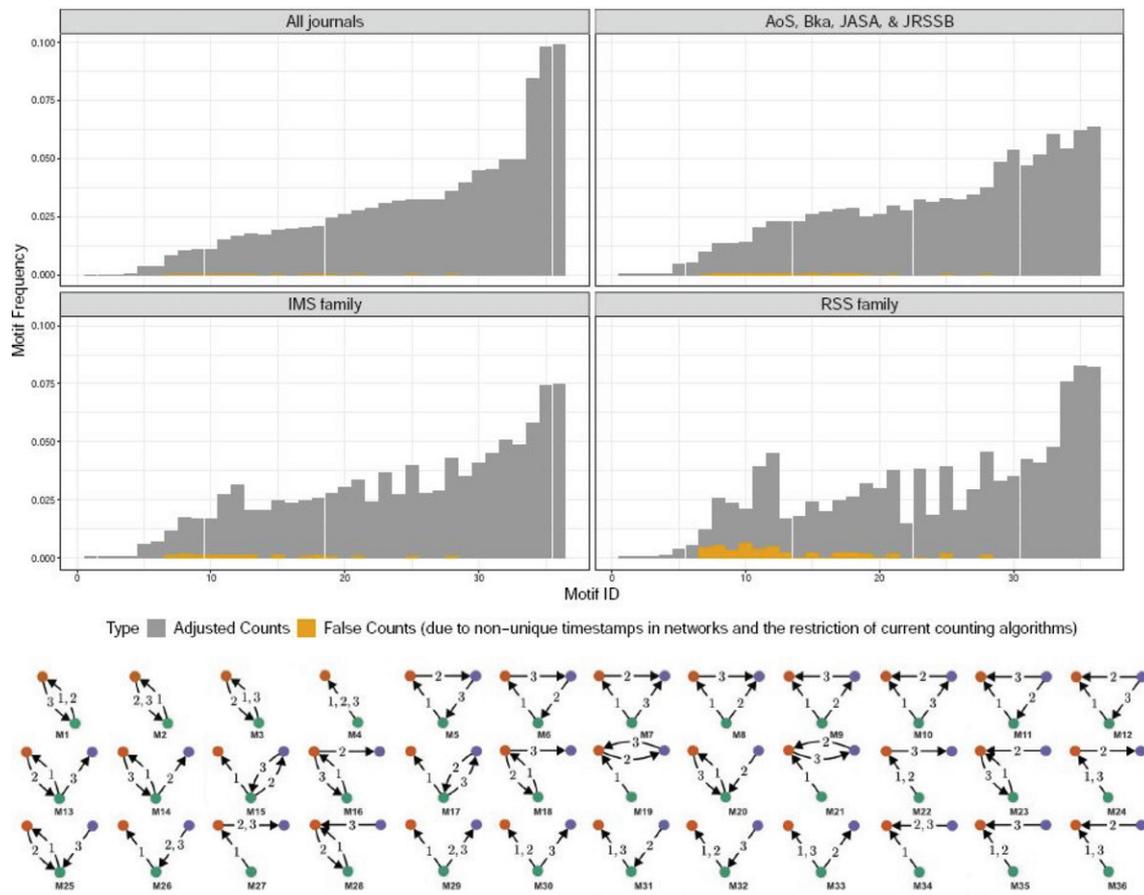


Figure 1. Top: frequency distributions of the 36 two/three-node, three-edge temporal motifs in four temporal citation networks constructed within different categories of journals. Bottom: the 36 temporal motifs with IDs corresponding to the x-axis of the bar plots, ordered by relative frequency for all journals.

temporal motifs in the network within a sliding observation window of $\delta = 5$ years, with results² shown in Figure 1. However, this is not an exact but an overestimated motif count in our networks since all temporal motif counting methods are currently built upon the assumption that each event in the given network has a unique timestamp, and it is not the case for our networks with edge resolution of one year. Consequently, some citing interactions with nonunique timestamps are mistakenly counted as motif instances by the algorithm. We adjust for one source of such error where all three citing interactions in the motif instance are of the same timestamps by subtracting such false counts (as illustrated in yellow in Figure 1) from the total counts, while the other source of false motif counts where only two citing interactions are of the same timestamps cannot be easily adjusted for given the restrictions in current motif counting methods. To fully address this issue, either finer time resolution of timestamped edges or a motif counting method adapted for nonunique timestamps is needed. For now, we use the partially adjusted counts in the following temporal motif analysis, which are the best estimates we can come up with so far for the true motif counts.

Figure 1 shows frequency distribution of all the two/three-node, three-edge temporal motifs (36 motifs in total shown on the bottom of Figure 1 with motif ID from 1 to 36), reflecting

the behavior of citing patterns among two or three authors publishing in different categories of journals. The plots are rich and space here precludes a comprehensive analysis, but here are a few examples of findings that can be obtained from the plots. In the full journal analysis, we can see that the most frequently occurring citing patterns (M34, M35, M36) are two authors successively citing another author in 5 years which reflects the broad impact of some seminal works, while the least frequent ones (M1, M2, M3, M4) are two authors citing each other, or one author citing another author multiple times in 5 years, which demonstrates that reciprocal citations across time occur relatively more rarely in the community of statistics. The motif frequency distributions for the three subsets of journals look different from each other and also different from the one for all journals. The most eye-catching differences are: (a) the frequency of motif M34, M35 and M36 decreases in the AoS, Bka, JASA and JRSSB category compared with other journal categories although they still occurred quite frequently; (b) while all the triadic patterns seem to occur less frequently in each journal category, the frequency of two triadic patterns (M11 and M12) seems to be more prominent in the RSS family.

Figure 2 shows the frequency distribution of the 36 temporal motifs across four decades, from which we can see a change in the behavior of citing patterns from 1975 to 2015. One interesting observation is that the two lines for 1996–2005 and

²Analysis code is available at <https://github.com/KolaczykResearch/TempMotifOnStatCitationNets>.

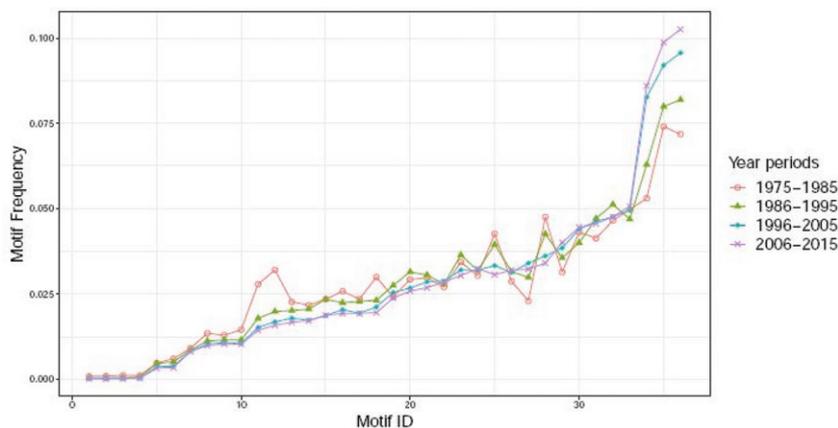


Figure 2. The comparison of motif frequency distributions across four decades: 1975–1985, 1986–1995, 1996–2005, and 2006–2015. Motif IDs in the x-axis correspond to the temporal motifs shown on the bottom of Figure 1.

2006–2015 align quite well with each, indicating that the citation pattern during 2006–2015 didn't change much from that during 1996–2005. However, the citation patterns from 1975 to 2005 seem to evolve every 10 years. For example, the frequency of motif M27, M34, M35 and M36 increases while that of motif M11, M12, M25 and M28 decreases from 1975 to 2005.

Funding

This work was supported by NSF award SES-2120115.

References

- Chakraborty, T., Ganguly, N., and Mukherjee, A. (2015), "An Author is Known by the Context She Keeps: Significance of Network Motifs in Scientific Collaborations," *Social Network Analysis and Mining*, 5, 16. [494]
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002), "Network Motifs: Simple Building Blocks of Complex Networks," *Science*, 298, 824–827. [494]
- Paranjape, A., Benson, A. R., and Leskovec, J. (2017), "Motifs in Temporal Networks," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 601–610. [494]
- Stone, L., Simberloff, D., and Artzy-Randrup, Y. (2019), "Network Motifs and their Origins," *PLoS Computational Biology*, 15, e1006749. [494]
- Wasserman, S., and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press. [494]
- Wu, W., Han, Y., and Li, D. (2008), "The Topology and Motif Analysis of Journal Citation Networks," in *2008 International Conference on Computer Science and Software Engineering* (Vol. 1), pp. 287–293. IEEE. [494]
- Zeng, X., and Rong, Z. (2021), "Evolution of the Physics Citation Network with Motifs," in *2021 40th Chinese Control Conference (CCC)*, pp. 776–780. IEEE. [494]