Informal collaboration in financial economics

Final report

Patrick Altmeyer

27 March, 2021

Rose and Georg (2018) set out to analyse how informal collaboration among scholars affects research outcomes. To do so they use a dataset comprising meta data of more than 5000 articles from the academic literature on Financial Economics published between 1997 and 2011. Although they are hesitant to draw and causal connections, the authors find significant links between informal networking - conferences, seminars and paper reviews - and the success of papers and ultimately researchers. The estimated relationships are stronger than for formal co-authorship, which had been the primary focus of previous research in the line of literature. Perhaps not surprisingly then, more than half of the papers analysed in Rose and Georg (2018) involve all three kinds of informal collaboration. The findings set out in the paper are relevant not just to the hiring committees of academic institutions, but also young researchers at the beginning of their academic careers, who Rose and Georg (2018) find to more actively seek informal collaboration than their predecessors. Perhaps their methodological and data work is an even more valuable contribution. Their network of informal collaborations is publicly available and can be used to understand the drivers of scientific productivity. More importantly, their methodology is reproducible and in fact has already been applied to other fields of science (Dong et al. (2021) and Rose, Georg, and Opolot (2020)). In what follows I will first briefly summarize the methodology and findings in Rose and Georg (2018). I will then critically review of their work in context of the broader literature.

Rose and Georg (2018)

Data and methodology

The network of informal collaboration in Financial Networks is a novel data set derived from the acknowledgement sections of over 5000 published research papers from three top journals – The Journal of Finance, The Review of Financial Studies, the Journal of Financial Economics – and three second-tier journals – the Journal of Financial Intermediation, the Journal of Money, Credit & Banking, and the Journal of Banking and Finance. The raw data has been made freely available along with a short demonstration of how to link authors and commenters in a graph.

Rose and Georg (2018) use the data to construct two graphs: one describing the network informal collaboration and one describing the network of co-authorship. In the former, edges are directed and weighted. Arrows point from author i to her commenter j who she acknowledges. Weights simply correspond to the inverse number of authors of any paper k. In the case of co-authorship network, links are still weighted, but undirected with elements h_{ij} of the adjacency matrix corresponding to the number of publications any two co-authors produced. Both networks are built from the same time series of research publications introduced above. Figure 1 shows the evolution over time of the number of links in the two networks. It illustrates two interesting points: firstly, both networks have seen massive growth over the sample period, which the authors attribute partially to more active engagement in networking by younger researchers; and secondly, the number of informal ties between researchers far exceeds the number of co-authorships.

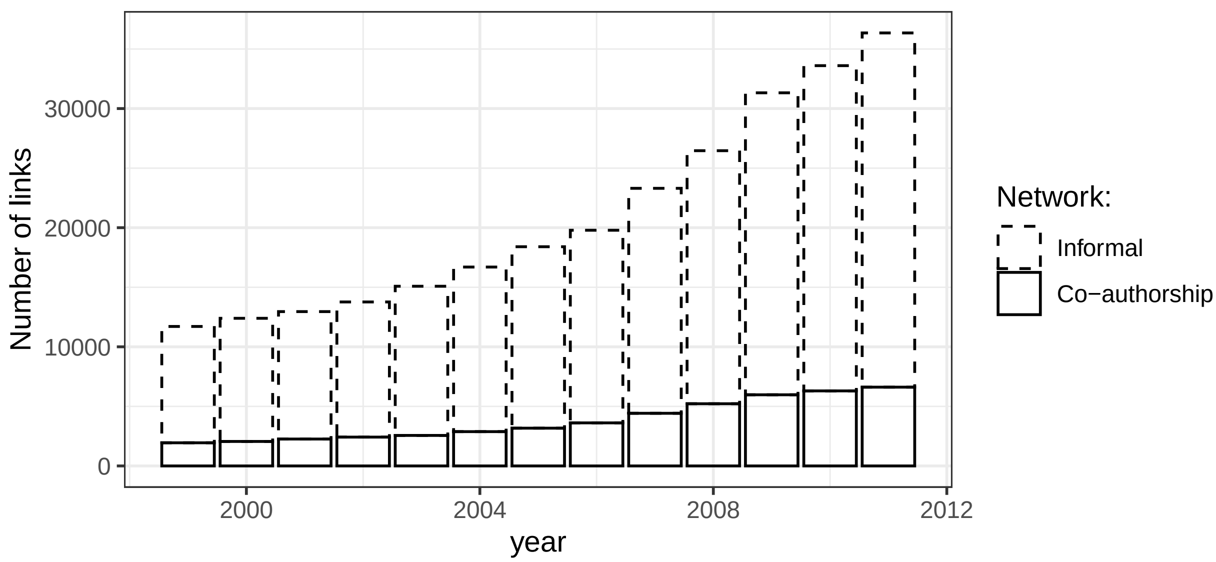

Figure 1: The evolution of the number of links in the two types of networks. At each time t the networks include relationships formed in $t-2$, $t-1$ and t as in Rose and Georg (2018).

Leading questions

Before turning to what we commonly understand as network analysis, Rose and Georg (2018) demonstrate the usefulness of their data through a number of simple regression models. The purpose of these empirical exercises is to shed light on a few leading questions: whether informal collaboration actually improves research, what type individuals engage (or are engaged) in the informal network and why researchers choose to form informal ties.

To be able to answer these questions Rose and Georg (2018) use the data underlying the network to derive a number of researcher characteristics including an individual's prolificness, experience and gender. The former is measured through a simple proxy measure based on their citations and publications. Experience is defined as the number of years between the first publication and the year of publication of the paper. Gender is estimated from authors names. Rose and Georg (2018) limit their attention to academic authors, omitting from the analysis any research assistants, non-academic advisers and editorial board members. The latter are removed from the analysis since their names can be expected to show up regularly in the acknowledgements section, simply due to their role as editors. Aside from researcher characteristics, the data comes equipped with a set of variables relating to the intensity of informal collaboration associated with individual papers. Those include the number of seminars attended, the number of conferences attended as well as the number of commenters.

With that data at hand the authors proceed to answer the leading questions empirically. With respect to the first question they estimate several modifications of the following model

$$\begin{aligned} \text{Success}_p = \alpha_1 \text{Paper Characteristics}_{p,t-1} + \beta_1 \text{No. of seminars}_p + \beta_2 \text{No. of conferences}_p \\ + \beta_3 \text{No. of commenters}_p + \beta_4 \text{Commenter quality}_p + \gamma \mathbf{D}_{\text{Journal}_p} + \mathbf{D}_t + \varepsilon_p \end{aligned} \tag{1}$$

where success of a paper p is measured in terms of its citation and acceptance by a top journal. Their estimates provide statistically significant evidence that with the exception of *conferences* all forms of informal collaboration contribute to the success of a paper. Descriptive statistics seem to demonstrate that it is mainly younger authors who increasingly rely on informal networking.

Turning to the second question about who engages in informal collaboration Rose and Georg (2018) estimate variants of the following model

$$
\begin{aligned}
\text{Collaboration}_{i,t} = \beta_0 + \beta_1 \text{female}_i + \beta_2 \text{Euclid}_{i,t} + \beta_3 \text{Publication Stock}_{i,t} \\
+ \beta_4 \text{Citation Stock}_{i,t} + \beta_5 \text{Experience}_{i,t} + \gamma \mathbf{D}_{Female_i} + \varepsilon_{i,t}
\end{aligned} \tag{2}
$$

where the dependent variable is a measure of how many times a reasearcher was included in the acknowledgments section of a paper. Only a minority of researchers acts as both authors and commenters at the same time, so this separation into producers and helpers of research as the authors describe it seems reasonable. The empirical estimates suggest that researcher charactersitics have significant impact on likelihood of being acknowledged. Most strikingly perhaps, gender appears to play an important role with female researchers up to 20% less likely to find their names in the acknowledgement sections.

Finally, with respect to the third leading question about researcher's incentives to seek informal ties, the authors simply look at distributional properties of the underlying sample. They find that there is a tendency to help researchers in return for past or future help: more than two thirds of all papers in the sample involve commenters commenting on the work of co-authors, their own commenters or department colleagues. Another possible driver of informal collaboration put forward by the authors revovlves around inter-generational transfers: older, more experienced researchers may choose to comment in order to support younger colleagues. The sample period is too small to establish conclusive evidence, but sample distributions nonetheless reveal that commenters ten to have between 7 and 20 years of experience.

Network analysis

In the context of network science, the part of Rose and Georg (2018) we turn to next is really the core of the paper. The authors first briefly present summary statistics for two different networks: firstly, the network of informal collaboration already introduced above and secondly a network of co-authorship. They then move on to perform a forecasting exercise in the spirit of Ductor et al. (2014) in order to gauge how well information derived from the two networks can forecast future productivity of researchers. Next, Rose and Georg (2018) also analyse the impact of network variables on the success of individual papers. Finally, they also investigate what drives centrality of researchers. We shall now briefly review each of these exercises.

Macroproperties of both networks The network of informal collaboration is much more connected, in the sense that one giant component captures more than 95% of all researchers. But the average clustering coefficient, which can be interpreted as the network's local density (Barabási 2013), is higher for the co-authorship network. Intuitively, these observations indicate that the social network of informal collaboration is more inclusive overall. The co-authorship network is composed of many components, but ties within these tend to be stronger on average than in the more loosely connected giant component comprising most of the informal network.

Rose and Georg (2018) look at three different centrality measures: degree, eigenvector and betweenness. Each time attention is limited to the largest component only, since centrality measures are not comparable across components. As the authors point out, including three different measures of centrality is not simply due diligence. They explain that while degree centrality provides a good measure of how central researchers are within their communities, eigenvector centrality roughly corresponds to the effort brought forward in equilibrium - a nuanced, but important difference. Finally, betweenness centrality can help us understand which researchers help information to flow between communities of researchers. An interesting finding here is that correlation between eigenvector centrality and betweenness centrality is not high: it appears that researchers with high influence are not necessarily the ones that connect their colleagues with each other.

Forecasting researcher's productivity Ductor et al. (2014) use a social network of co-authors to study if information about individual researchers' relations with others in the network can be used to forecast their future productivity. Comparing forecasts from two benchmark models that merely include information about an individual's past output to a model that incorporates network variables, they find that the latter leads to slightly more accurate forecasts. Rose and Georg (2018) apply that methodology to both the co-author network and the network of informal collaboration, in each case benchmarking forecasts to a model that only uses information about individuals' past output. In particular, the additional network information included involves the various centrality measures introduced above. The reduction in the root mean forecasting error (RMSE) is significant in both cases, though higher with respect to the network of informal collaboration. The highest forecasting accuracy is attained when using information from both networks. The authors' main conclusion from this exercise is that incorporating network information can support the process of hiring researchers.

Success of papers To understand further the effect of researcher characteristics on the success of a paper (first leading question above), the authors repeat the estimation from above (Equation (1)), this time including centrality measures instead. Perhaps somewhat unsurprisingly, they find that centrality of researchers does contain significant information about how well a paper is received. The most striking result here is that information from the informal network once again appears to yield more significant estimates. Also consistent with observations already made above, the findings here suggest that high betweenness centrality negatively affects the success of a paper, in sharp contrast to other measures of centrality.

Centrality of researchers Broadly in line with the last point, the authors also find that while experience and prolificness of researchers tend to be associated with higher betweenness centrality, prolificness has no significant effect on eigenvector centrality of authors, that is their overall influence within the network. There is also some evidence that female researchers are less likely to have high betweenness centrality.

Critical review

Threats to validity

There are a few caveats, some of which the author's point out themselves. With respect to incentives for including commenters in the acknowledgements section, Rose and Georg (2018) point out that acknowledging may be strategic: name-tagging a prolific, senior researcher may be beneficial to a young researcher for reasons not related to their actual comments. While this is perhaps the main concern regarding the validity of the data, other small issues are worth pointing out. Firstly, it is potentially problematic to focus on the largest component only, especially since the network of informal collaboration is so much more connected. Secondly, the fact that gender had to be estimated from authors names adds some additional uncertainty regarding any conclusions with respect to gender. Finally, the rationale for excluding industry professionals and central bankers from the analysis is not entirely clear. After all, these professionals play an important role in academic research around Financial Economics: they attend and organise conferences, can often provide a unique first-hand perspective on financial markets and regularly produce research published in academic journals themselves.

Concluding thoughts

In terms of the author's main conclusions, I would have expected a slightly different tone. While they do not fail to mention potential pitfalls associated with networks characterized by a few central agents with high influence (Azoulay, Fons-Rosen, and Graff Zivin 2019), I would have expected more of the discussion to revolve around this, rather than the benefits of network information for hiring committees. In particular, more investigation could have gone into the growth dynamics of the networks: do the networks exhibit success-breeds-success dynamics? Figure (2) shows the distribution of out-degrees (logs) of the two types of networks. Evidently, in particular for the informal network this distribution is heavily right-skewed, indicative of a few highly influential researchers. There is some indication that this skew has mildly increased over the sample period. Interestingly, a recent publication that cites Rose and Georg (2018) investigates just that at the institutional level and finds that success-breeds-success dynamics appear to lead to disparity of institutional output (Dong et al. 2021). With all that said, Rose and Georg (2018) is a comprehensive paper as it is and admittedly a thorough investigation of network dynamics would have exceeded its scope.

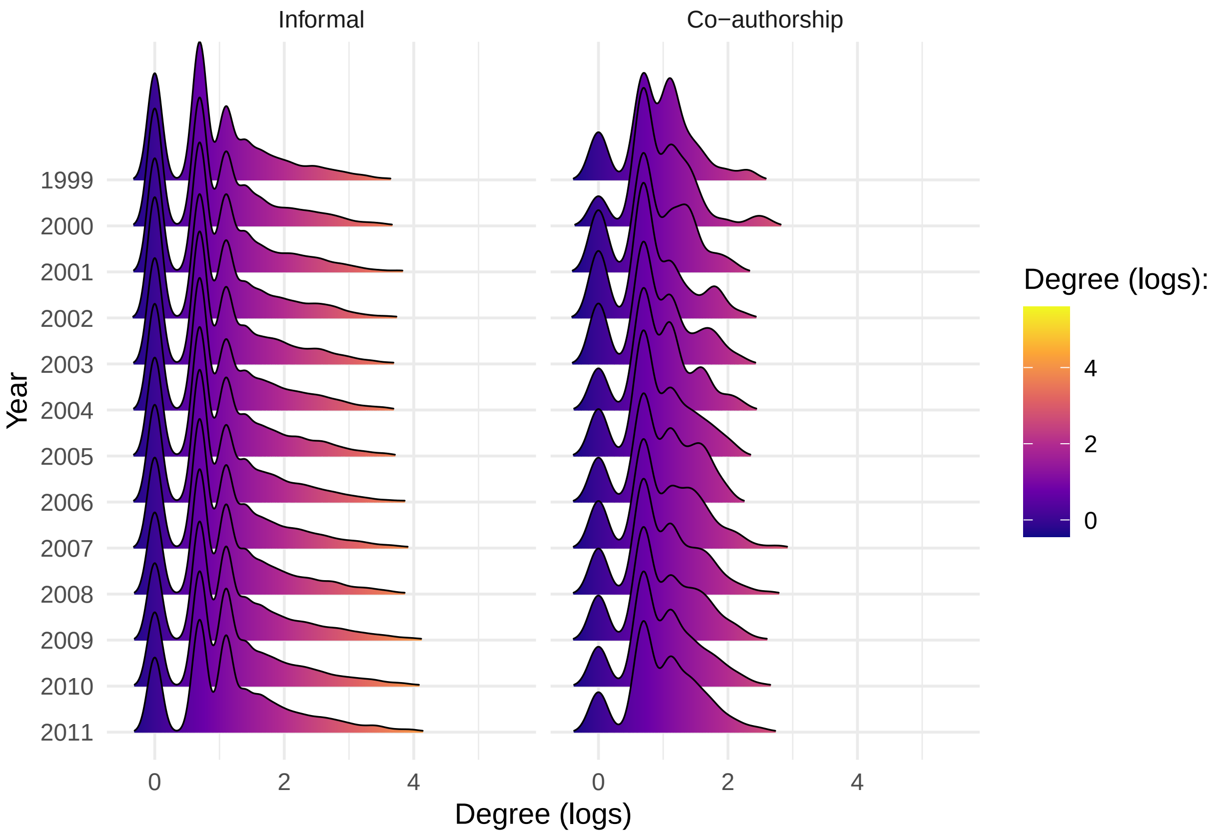

Figure 2: Distribution of out-degrees of the two types of networks. At each time t the networks include relationships formed in $t-2$, $t-1$ and t as in Rose and Georg (2018).

Future research

Overall Rose and Georg (2018) provides an interesting, reproducible exhibit of how network science can be applied to social science research. In a recent follow-up paper Rose, Georg, and Opolot (2020) demonstrate that their methodology can be readily applied to other, related questions: the find that a paper's citation count increases in the discussant's prolificness. Using social network analysis they rule out a diffusion of information about the paper within the discussant's social network as a factor. Finally, let me propose another interesting avenue for future research: the role of researchers in cross-disciplinary research. Rose and Georg (2018) and related papers have tended to limit their attention to one field of research. An interesting idea could be to use betweenness centrality to identify researchers that contribute to bridging the gaps between different lines of thought and even separate disciplines. This information could then be used to analyse the effectiveness of cross-disciplinary research towards increased overall research productivity.

References

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin. 2019. "Does Science Advance One Funeral at a Time?" *American Economic Review* 109 (8): 2889–2920.

Barabási, Albert-László. 2013. "Network Science." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (1987): 20120375.

Dong, Hui, Dan Luo, Xudong Zeng, and Zhentao Zou. 2021. "Cross-Affiliation Collaboration and Power Laws for Research Output of Institutions: Evidence from Top Journals in Financial Economics."

Ductor, Lorenzo, Marcel Fafchamps, Sanjeev Goyal, and Marco J Van der Leij. 2014. "Social Networks and Research Output." *Review of Economics and Statistics* 96 (5): 936–48.

Rose, Michael E, and Co-Pierre Georg. 2018. "What 5,000 Acknowledgements Tell Us about Informal Collaboration in Financial Economics." *Max Planck Institute for Innovation & Competition Discussion Paper*, no. 11.

Rose, Michael E, Co-Pierre Georg, and Daniel C Opolot. 2020. "Discussants." *Max Planck Institute for Innovation & Competition Research Paper*, no. 20-19.