(Leone Sciabolazza et al. 2017)

Leone Sciabolazza, Valerio, Raffaele Vacca, Therese Kennelly Okraku, and Christopher McCarty. 2017. “Detecting and Analyzing Research Communities in Longitudinal Scientific Networks.” *PLoS ONE* 12 (8). https://doi.org/10.1371/journal.pone.0182516.

[Summary of the paper’s research]

“this paper studies the formation of communities and interdisciplinary collaborations by analyzing longitudinal collaboration network extracted from peer-reviewed publications and awarded grants at the University of Florida” (p 1-2). Studies three years (2013-2015). “We propose a method that uses community-detection algorithms to identify longitudinal research communities, i.e., sets of investigators who have been part of the same collaborative subgroup consistently over a certain number of years. We analyze the drivers of interdisciplinary collaborations that cut across different research communities using Exponential Random Graph Models (ERGM)” (p 2). “We analyze the evolution of this community structure by developing a two-step strategy to identify persistent research communities. First, we detect clusters that signal the presence of collaborative subgroups in yearly cross-sectional networks (Gu,2013, Gu,2014, Gu,2015). Second, we identify sets of researchers who have consistently been part of the same collaborative subgroups, filtering out temporary collaborations that were only in place for a limited time” (p. 6) They use the Louvain community detection algorithm. “the Louvain algorithm returns a partition of investigators into collaborative subgroups within the network of each year t” (p.6-7).

“Broad institutional classifications such as institutes, departments, and colleges are unlikely to describe these groups, particularly when they include investigators from different disciplinary backgrounds and affiliations” (p. 2). [This points out that the node attributes are not sufficient to identify disciplines.]

“By contrast, network methods for community detection can help to identify collaborative subgroups as they emerge from actual interactions between scientists” (p.2). [This claims that community detection methods are a way to identify disciplines. I’d use it to explain a way to continue the SNAP research.]

“Spatial proximity, homophily, transitivity, past collaboration experiences, shared funding sources, disciplinary background, and department and college affiliation all play a role in shaping the structure of scientific networks” (p.3). [This points out ways of describing the network’s structure. I’d use it in the introduction of CUPID methods]

“Network formation models, including ERGMs, view a social network as the realization of a probabilistic social process in which the creation of a tie between researchers may result from different factors, including node-level attributes (e.g., the collaborators’ disciplinary background) and network topology (e.g. transitivity effects)” (p.3).

“While in some cases groups of collaborators work on both publications and grants at approximately the same time, it is often the case that researchers collaborate on publications first, to then be able to demonstrate research results and apply for grants together; or conversely, frequently scientists obtain a grant first, which then enables them to conduct research resulting in publications. Thus, at a specific point in time, for certain pairs and groups of scientists, collaboration might only occur through publications and be detectable in publication data; for others, collaboration might only take place through grants and be visible in grant data” (p. 3-4).

The co-authorship ties are weighted by the total publications the dyad shared in a one-year period. The awarded grants are weighted by the total awarded grants the dyad shared in a one-year period. “We define as union collaboration the union of a publication and a grant collaboration” (p.4). A dyad has a union tie in a year if articles are published by the dyad and if the dyad shares an active grant in that same year.

“It should be noted that even the most recent publication and awarded grant data have an inherent temporal lag, in that published articles and awarded grants normally follow in time, by at least several months, the actual start of a collaborative research project” (p. 4).

“The rich set of interactions captured in our data entails two major challenges. First, the data include a large amount of interactions over time. Analyses of such large network data normally require extremely computer-intensive calculations. Second, there is a constant variation of these interactions over time (see Appendix A in S1 File), which is a possible source of confounding factors: sporadic and temporary instances of collaboration may generate a noise that can bias our comprehension of systematic collaboration patterns.” (p.5) [My data also has these two major challenges. I have found that it is computer-intensive, where my ERGMs take 12 hours to run. It is time-consuming to run a model for each year while also incrementally increasing the \alpha parameter. This is why I focused on the 5-year network. Future SNAP research can work to model each individual year. Second, the data has this same variation where the nodes enter and leave the network in progressing years.]   
  
“We address these issues by developing a method to aggregate and summarize data from multiple years, reducing the computational burden of network analyses, while also accounting for a significant portion of the variation in interactions between years” (p. 5). [I didn’t have a fancy method. I just set to time frame as a 5 year time frame instead of a 1 year time frame.]

“Our aim is to identify persistent patterns of collaboration over time.” (p.5) [This study has the same goal as my study]

“The result is the emergence of a community network structure [46–47], in which subgroups of nodes are densely connected with each other, and have a high number of adjacent nodes in common (high internal clustering), while being more sparsely connected with the rest of the network.” [This is a description of my networks, too. It must be the structure of research networks in general. I could use this to say that when describing my network structures.]