

## DISCUSSION OF “COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS”

BY PEDRO REGUEIRO, ABEL RODRÍGUEZ AND JUAN SOSA

*University of California*

**1. General comments.** We would like to start by congratulating the authors for taking the initiative in gathering this very interesting and novel data set. Given the substantial amount of work involved in scrapping the data, their focus on four periodicals (Journal of the American Statistical Association, Journal of the Royal Statistical Society Series B, *Biometrika* and *Annals of Statistics*) is understandable. However, this relatively narrow choice raises some concerns. The most obvious one relates to the robustness of the results to the choice of periodicals, particularly for authors/papers concentrating on areas for which specialized high-quality alternative publications exist. Two examples are biomedical applications and Bayesian methods. Furthermore, although the four journals selected are mainly methodological, the inclusion of the Applications and Case Studies section of JASA was unfortunate. Manuscripts published there can be expected to have more in common with papers published in the *Annals of Applied Statistics* or the Journal of the Royal Statistical Society, Series C than with manuscripts in the Theory and Methods section of JASA itself.

The analysis in the paper feels a little bit like a “fishing expedition.” The paper lacks a clear question that motivates and shapes the data collection. The use of multiple alternative methods (both for constructing the networks and for analyzing them) yielding different results also detracts from a sense of purpose. This is a pity because there are a number of interesting questions that could be explored if the data collection exercise had been slightly expanded with a clear objective in mind. Some examples include the following:

1. What are the main drivers of collaboration in statistics?
2. How have the collaboration networks evolved over time?
3. How likely are researchers to publish with their Ph.D. mentors as time goes by?
4. Are there regional biases in citation and/or publication patterns?
5. How prevalent is “self-referencing” (both at the author and journal level)?

The feeling of lack of focus is reinforced by the fact that the clusters generated by the community identification methods in the paper are puzzling. For example, the fact that only three clusters are identified in the connected component of the author citation network is quite surprising. This small number could be driven by