Contribution to the discussion of

"Statistical Modelling of Citation Exchange Among Statistics Journals"

by Christiano Varin, Manuela Cattelan and David Firth

given at the RSS Meeting on May 13, 2015

Authors: Jane Carlen and Mark S. Handcock[1]

Titles and Affiliation: Jane Carlen and Mark S. Handcock (University of California, Los Angeles)

Citation data can be fruitfully thought of from a networks perspective. Visualising the network of journals via graphical layout algorithms offers a summary of relationships, clustering and centrality, but is coupled with artifact.

The Stigler model estimates "export scores", u_i, such that c_{ij} is assumed binomially distributed with $E(c_{ij}) = t_{ij} \exp(\alpha_i + \beta_j)$ and $u_i = \alpha_i - \beta_i$, as in 'quasi-symmetry' formulation (4). We can place these assumptions in the context of an exponential-family random graph model (ERGM) on a valued network, where edge weights are directed citation counts Krivitsky (2012). A direct extension of the Stigler model would retain the assumption of binomially distributed citations. Here we consider a Poisson model with canonical link and mean modeled with sender and receiver effects. The corresponding estimates of export scores (i.e. journal-specific receiver minus sender coefficient) are highly correlated (0.95) with those of the Stigler model reported in Table 5.

A benefit of the network model is extensibility, both theoretically and computationally. To illustrate, consider the two-dimensional latent space model with sender and receiver effects (Krivitsky et al., 2009). This model posits distances between journals as latent variables that effect edge weights (citation counts) (see, Hoff et al., 2002; Krivitsky and Handcock, 2008, 2015). The corresponding estimates of export scores are very highly correlated (0.99) with those of the Stigler model.

Model-based estimates of journal positions are shown in Figure 1. Although there is no clustering term in the latent space model, the clustering presented in the paper (Section 3) is fairly well captured. The plots illustrate how individual journals and clusters fit together. However, we should be careful not to reify these point estimates of positions. The right-hand panel displays the uncertainty in the positions using a sample of draws from the model.

Figure 1 offers a visual aid to the observation (Section 7.2) that many journals are not significantly different in rank, and therefore grouped rankings are often more appropriate than traditional ordering. We see a periphery of low-ranked journals on the left and a small cluster of leading journals around *JRSS B*, but beyond that a widely dispersed middle. Centrality does not equate to rank or prestige, as shown by *Annals of Statistics* and *Bernoulli* in the bottom right.

The models are readily fit with the `latentnet` package (Krivitsky and Handcock, 2015) and the code is available from the first author's website.

[1]Department of Statistics, University of California, Los Angeles, CA 90095-1554 (E-mail: *jcarlen@ucla.edu*).

References

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association 97*(460), 1090–1098.

Krivitsky, P. N. (2012). Exponential-Family Random Graph Models for Valued Networks. *Electronic Journal of Statistics 6*, 1100–1128.

Krivitsky, P. N. and M. S. Handcock (2008). Fitting latent cluster models for social networks with **latentnet**. *Journal of Statistical Software 24*(5).

Krivitsky, P. N. and M. S. Handcock (2015). *latentnet: Latent Position and Cluster Models for Statistical Networks*. The Statnet Project (`http://www.statnet.org`); University of Wollongong. R package version 2.7.0.

Krivitsky, P. N., M. S. Handcock, A. E. Raftery, and P. D. Hoff (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks 31*, 204–213.