Social Network Analysis on Food Web and Dispute Data

F. Swen Kuh*, Anton H. Westveld† and Grace S. Chiu†

*Department of Statistics, The University of Auckland

†Research School of Finance, Actuarial Studies & Statistics, The Australian National University

1 What is it?

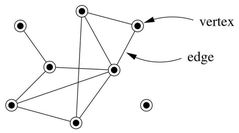

Figure 1: A basic network with 8 vertices and 10 edges[6]

A NETWORK is any collection of objects in which some pairs of these objects are connected by links. Network analysis can be traced back to at least three different disciplines: anthropology, psychology and sociology, but it is now widely applicable to many other disciplines too in the physical and biological sciences. We are particularly interested in **Social Network Analysis**, where it uses network theory to analyse social networks – a network that often involves individual social actors (people) and relations between them. Social science data are different to physical data of natural sciences as it is organised around meanings, motives, and definitions. This feature of the data that involves complex relationships and interactions between the individuals meant that the social actors and their actions (links) should be viewed as interdependent, as opposed to independent, autonomous units. Network analysis treats the relational data by using *nodes* or *vertices* to represent the individuals and *edges* to illustrate presence of a specified relation between actors.[3]

Model specification

Our inferences and analyses are done by using the package **latentnet**[5] in the statistical software **R**. We consider the following model fitted by the package which includes $\theta_{ij} = (\beta_0, Z_i, Z_j, s_i, r_j)$ for $i, j = 1, \ldots, n$ and $i \neq j$:

$$\Pr(Y = y \mid \theta, x) = \prod_{\forall\,(i,j),\ i \neq j} \Pr(Y_{i,j} = y_{i,j} \mid \theta_{i,j}, x_{.,i,j}),$$

$$\Pr(Y_{i,j} = y_{i,j} \mid \theta_{i,j}, x_{.,i,j}) = f(y_{i,j} \mid \mu_{i,j}),$$
$$\mu_{i,j} = g^{-1}(\eta_{i,j}),$$

$$\eta_{i,j} = \beta_0 + s_i + r_j + Z_i' Z_j \qquad (1)$$

where β_0 - fixed mean; s_i - sender effect due to i; r_j - receiver effect due to j; $Z_i' Z_j$ - inner product of latent positions of respective *actors* in an unobserved social space.

We consider the following hierarchical model: distributions on the latent space positions, clustering, sender and receiver components, and dimension $d = 2$ was prespecified for easy visualisation. We estimate the parameters through a Bayesian context with default priors defined as discussed in Krivitsky's paper[5] and used in the **statnet** and **latentnet** packages:

$$\beta_0 \sim \mathrm{N}(\xi_0, \psi_0^2)$$

$$Z_i \overset{i.i.d.}{\sim} \sum_{g=1}^{G} \lambda_g \mathrm{MVN}_d(\mu_g, \sigma_g^2 \mathrm{I}_d) \qquad i = 1, \ldots, n$$

$$s_i \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma_s^2) \qquad i = 1, \ldots, n$$

$$r_i \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma_r^2) \qquad i = 1, \ldots, n$$

$$\mu_g \overset{i.i.d.}{\sim} \mathrm{MVN}_d(0, \omega^2 \mathrm{I}_d) \qquad g = 1, \ldots, G$$

$$\sigma_g^2 \overset{i.i.d.}{\sim} \sigma_0^2 \mathrm{Inv}\chi_\alpha^2 \qquad g = 1, \ldots, G$$

$$(\lambda_1, \ldots, \lambda_G) \sim \mathrm{Dirichlet}(v_1, \ldots, v_G)$$

2 Examples

In order to evaluate the statistical latent position and cluster models for networks, we will use a social network analysis (SNA) technique known as *latent space modelling*. The *space* corresponds to a space of unobserved latent characteristics that represent potential transitive tendencies in network relations. It is assumed each actor i has an unknown position Z_i in this space.[3] The sender effect s_i generally reflects the propensity for node i to be influenced by others, while the receiver effect r_j reflects the propensity for node j to influence others.[7] We apply the **latentnet** package to two different datasets: **food web data** and **militarized interstate dispute data**.

Part I: Food web data

The first set of data we will consider is on trophic food webs. Trophic food webs have nodes that are intertwined by directed links that point from prey to predator. The data was observed from Goose Creek Bay in the St. Marks National Wildlife Refuge in the southeastern United States. As discussed in Chiu and Westveld's study, the use of SNA to comprehend network relational structure allows us to gain deeper understanding of food web structure and its contributing factors.[1]

Model specification

The food web data is presence-absence binary data, where $y_{ij} = 1$ if the link i → j is observed, and $y_{ij} = 0$ otherwise. We specify a logistic regression model to (1) using bilinear distances similar to Chiu & Westveld[1]:

$$\log \mathrm{odds}(y_{ij} = 1 \mid \beta_0, s_i, r_j, Z_i, Z_j) = \beta_0 + s_i + r_j + Z_i' Z_j$$

where $\log \mathrm{odds}(p) = \log \frac{p}{1-p}$.

Results

The 2-cluster model was found to be the best model based on its BIC value[9]. However, the two clusters are only graphically noticeable when we use the 3-cluster model. Based on Figs. 2 & 3, the two main clusters appear to have been divided based on the general notion of "prey" and "predators". A similar plot but for s vs. r (not shown) suggests a negative correlation. It implies that a node with a high sending effect tends to have a low receiver effect, and vice versa. The formal clustering and associated plots have provided further insight into the food web structure in Goose Creek Bay. Node labels used here are available in Chiu & Westveld[1].

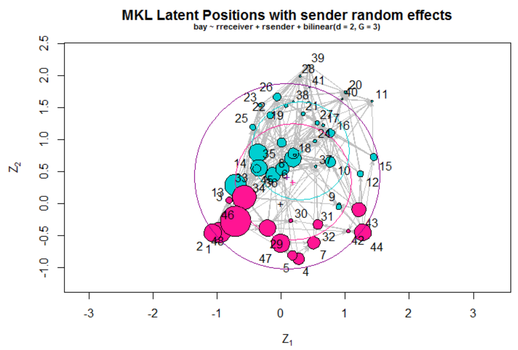

Figure 2: Food web graph after fitting latent cluster model with 3 clusters. The size of the nodes is proportional to the posterior mean of its sender effects, s.

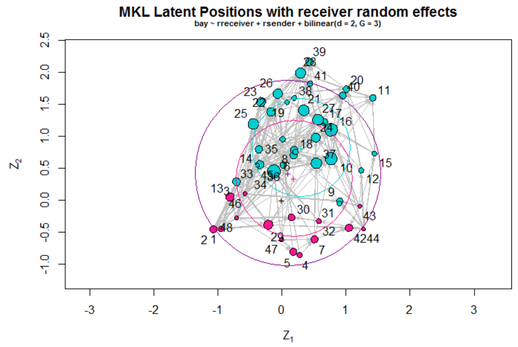

Figure 3: As before, the food web graph after fitting latent cluster model with 3 clusters but the node sizes are proportional to the receiver effects r instead.

Part II: Militarized interstate dispute data

A militarized interstate dispute (MID) is described as an event "in which the threat, display or use of military force ... by one member state is explicitly directed toward the government, official representatives, official forces, property, or territory of another state."[4] The data was obtained from the Correlates of War project[2] and contains records of conflicts between 196 countries from the year 1816 to 2010. Due to limited time frame for computations, our analysis was done on only 20 countries including the countries in the Middle East, USA, Malaysia and New Zealand from 2001 to 2010.

Model specification

The MID data used in our analysis consists of a ten-year period of whether or not one nation initiated a dispute against another nation. For that reason, we have binomial data with a total of 10 trials. We specify a logistic regression model for (1) using Euclidean distances:

$$\log \mathrm{odds}(y_{ij} = 1 \mid \beta_0, s_i, r_j, Z_i, Z_j) = \beta_0 + s_i + r_j - \mid Z_i - Z_j \mid$$

Results

Similarly in this example, the 2-cluster model was also found to be the best model based on its BIC value. From Figs. 4 & 5, we can see that one cluster contains only 3 countries (Iran, Iraq and United Arab Emirates), and the other cluster contains the rest of the countries. A similar plot but for s vs. r (not shown) indicates that s and r tend to be highly positively correlated. This is unlike the negative correlation for the food web data.

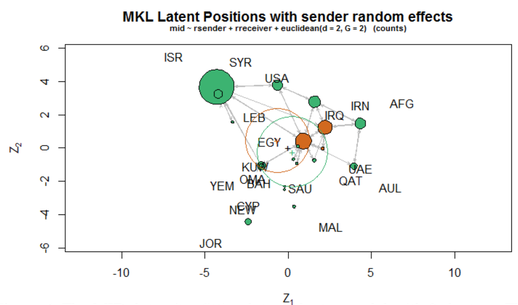

Figure 4: The MID data after fitting latent cluster model with 2 clusters. The size of the nodes is also proportional to the posterior mean of its sender effect.

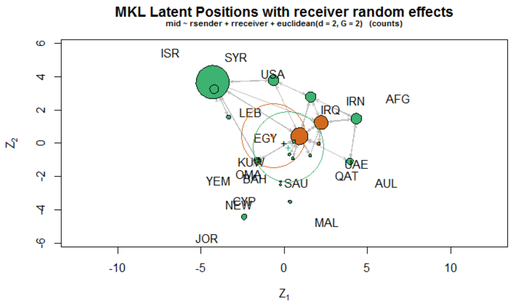

Figure 5: Latent cluster model with 2 clusters on MID data with the node sizes proportional to the receiver effects instead.

3 Discussion & Future Work

- We additionally fitted models to include covariates but excluding clustering. The **ergmm** function from the **latentnet** package in R by Krivitsky[5] was used. This inference/computation was compared to the **gbme** function used by Hoff[3] (which does not allow for clustering). To compare the two codes, we consider the covariate from Sampson's[8] Monk data ($x_{i,j} = 1$ if i and j are in the same group as identified by Sampson, $x_{i,j} = 0$ otherwise). We note a difference in the means of the posterior distributions:
  - Krivitsky's : 3.760 with a 95% credible interval of [2.841, 4.769]
  - Hoff's : 5.297 with a 95% credible interval of [3.404, 7.938]

  This may be due to the model by Hoff correlating sender and receiver random effects and also accounting for pairwise reciprocity. With the developments discussed below, we intend to include covariates.
- Directed network data are more naturally modelled with a sending preference space $\boldsymbol{U}$ that is different from a receiving preference space $\boldsymbol{V}$.[1] We would like to cluster on each of the two spaces, thus we would extend our model from (1) to:

$$\eta_{i,j} = \beta_0 + s_i + r_j + U_i' V_j$$

- Additionally, we would also like to cluster on the joint sending and receiving space (s_i, r_i), as well as allowing the clusters to evolve over time. (See Chiu & Westveld[1] and Westveld & Hoff[10])

4 References

[1] Chiu, G. S., & Westveld, A. H. (2011). A unifying approach for food webs, phylogeny, social networks, and statistics. *Proceedings of the National Academy of Sciences*, *108*(38), 15881-15886.

[2] Diehl, P., Geller, D., Gibler, D., Gochman, C., Hensel, P., Moaz, Z., Palmer, G., Pollins, B., Ray, J. L., Regan, P., & Stoll, R.. Correlates of War Project: Militarized Interstate Dispute (MID) Data, 1816-2001. ICPSR24386-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-03-05.

[3] Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, *97*(460), 1090-1098.

[4] Jones, D. M., Bremer, S. A., & Singer, J. D. (1996). Militarized interstate disputes, 18161992: Rationale, coding rules, and empirical patterns. *Conflict Management and Peace Science*, *15*(2), 163-213.

[5] Krivitsky, P. N., & Handcock, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, *24*(2).

[6] Newman, M. (2010). *Networks: an introduction*. Oxford University Press.

[7] OMalley, A. J., & Marsden, P. V. (2008). The Analysis of Social Networks. *Health Services & Outcomes Research Methodology*, 8(4), 222269.

[8] Sampson, S. F. (1969). Crisis in a cloister (Doctoral dissertation, Ph. D. Thesis. Cornell University, Ithaca).

[9] Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

[10] Westveld, A. H., & Hoff, P. D. (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics*, *5*(2A), 843-872.