SUPPLEMENT TO LATENT SPACE MODELS FOR MULTIVIEW NETWORK DATA

By Michael Salter-Townshend and Tyler H. McCormick

University of Oxford and University of Washington

1. Introduction. We present results of applying the methods in our paper "Latent Space Models for Multiview Network Data" to the well known Sampson's monks dataset (Sampson, 1969). This dataset comprises 8 relations (4 negative and 4 positive) on 18 actors (monks in a cloister). The relations are antagonism, disesteem, negative influence, blame, like, esteem, influence and praise. Previous analyses of this dataset has often focussed on the like network and includes clustering (Hoff *et al.*, 2002; Handcock *et al.*, 2007; Salter-Townshend and Murphy, 2013).

We then include latent space representations from a subset of villages. We also provide some MCMC diagnostics for the chains used for inference of the Karnataka results in the main body of the paper.

Fienberg *et al.* (1985), whose methods are perhaps closest to our approach, aggregate the networks into two views; positive and negative. They fit various models that are extensions of the p_1 model of Holland and Leinhardt (1981) to directed multiview networks and find the best fit model to be one that uses group specific exchange parameters. This model requires known groupings of the actors (although in this case the groups were found via a clustering based on the patterns of linkage in each view) and the exchange parameters are denoted as $\rho_{12}^{[d_r]}$ and $\rho_{12}^{[.e]}$, where d and e index over the known groups.

 $\rho_{12}^{[d,r]}$ refers to the propensity of group d to exchange relations 1 and 2 asymmetrically. i.e. the probability of a link from a monk in group d depends on the group specific tendency to have a have a view 1 link to someone with a view 2 link to him. Similarly, the probability of a link to a monk in group e in view 1 depends on the group specific tendency to send a view 1 link to someone with a view 2 link from him $(\rho_{1}^{[l,e]})$. This model is

^{*}This work is partially supported by the U.S. Army Research Office and a Google Faculty Research Award and by Science Foundation Ireland under Grant No. 08/SRC/I1407: Clique: Graph & Network Analysis Cluster.

Keywords and phrases: Latent space model, Multiview relational data, Social network

an extension to the p_1 model and is only defined for asymmetric networks; it thus cannot be applied to the Karnataka dataset. Interpretation of the exchange parameters is somewhat similar to our interaction parameters; if $\rho_{12}^{[d,.]}$ is positive it would suggest that a link from j to i in view 2 indicates a high(er) probability of a link from i to j in view 1. Our interaction parameter is non-directional due to the symmetry of our motivating example but interpretation is similar. Note that in the exchange model of Fienberg *et al.* (1985) the groupings are playing a similar role to a (reduced) form of the latent space positions to capture the network topology.

However, whereas there are 2279 degrees of freedom in their model our more parsimonious approach requires just $18 \times 2 + 1$ free variables in a fully Bayesian approach. The authors of Fienberg *et al.* (1985) do not provide the estimates for our comparison, however we will report our results for this dataset here.

2. Results. The raw graph covariance for the aggregated positive network with the aggregated negative network was -0.05 and the raw correlation was -0.22. Our corresponding mean posterior values were -0.02 and -0.09 respectively. These values are lower in magnitude due to the inclusion of the latent space component of our model but are in the same (negative) direction indicating that even when accounting for position in both latent spaces (the negative and positive relations networks) there is a negative correlation between links in these views. The first result means that there is a higher likelihood of a positive link between two monks if they do **not** have a negative link and vice versa (raw correlation). The second result means that there is a higher likelihood of a positive link between two monks of known separation in both negative and positive social space if they do **not** have a negative link and vice versa. The first result is a marginal estimate and the second is a conditional statement based on the posterior.

No positive values for the interaction term were sampled and a t-test for the null hypothesis that the correlation ρ is zero returns a p-value less than 2.2×10^{-16} . Figure 1 shows the trace plot and histogram for the per-sample correlation parameter ρ .

We also fit a model with the interaction term set to zero (i.e. $\phi = \sigma = \rho = 0$). This is equivalent to modelling the two network views independently. We then calculated the Area Under the Curve for Receiver Operating Characteristic curves for the two models and found values of 0.825 for

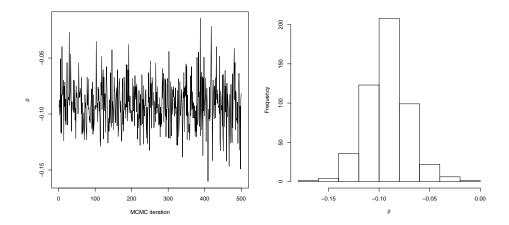


Fig 1. Correlation parameter ρ trace plot (left) and posterior histogram (right).

the independent model and 0.8265 for the joint model. This indicates that both models fit the data well but that the model with our interaction term performs slightly better and is consistent with the result that the posterior for ρ does not contain 0.

3. MCMC Diagnostics for the Karnataka Dataset. We wish to check the mixing and convergence of our MCMC chains. Fortunately, STAN provides a convergence diagnostic. This is a scalar associated with each sampled latent variable in the MCMC chain called \hat{R} that is 1 at convergence. For the Karnataka data, the values of \hat{R} all α and ϕ mix well in relatively few iterations and satisfy the Gelman-Rubin convergence criteria $(\hat{R}_{max} < 1.1)$. Identifiability issues make it more difficult to access convergence in the latent positions, Z. As with previous latent space models, the likelihood we propose depends on the latent positions only through pairwise distances. We could use a common rotation to improve the identifiability of the latent positions. Since the remaining parameters mix well and since we use the ϕ parameters most extensively in the results, we chose not to restrict rotations. Across all villages the latent positions have a have a mean \hat{R} of 1.54. We also experimented with running longer MCMC chains and found similar parameters values and convergence diagnostics. Along with the Gelman-Rubin diagnostic, we also visually inspected the chains for each village. An example of typical output for village 72 is given in Figure 2.

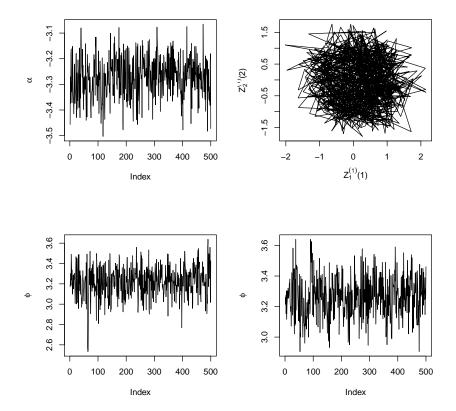


Fig 2. MCMC trace plot for $\alpha^{(1)}$, $z_1^{(1)}$, $\phi_{1,2}$ and ϕ_{35} from village 72. The latent positions Z were rotated and translated but not scaled using Procrustes method to the initial configuration. This was done as both the likelihood and the prior are invariant to translations and rotations of Z in each latent space.

4. Stan Codes.

```
# likelihoods and priors
data {
  int<lower=0> N;
  int<lower=0> R;
  int<lower=0> R2;
  int<lower=0> R3;
  int x[N,N];
  real<lower=0> sigma_alpha;
```

```
real<lower=0> sigma_rho;
  real<lower=0> sigma_z;
  int b_y[R,R2];
  int b_rho[R2,R3];
    }
parameters {
 real alpha[R];
  real rho[R3];
 real z[R,N,2];
          }
model {
 vector[R2] p;
 real LP[R2];
 real up[R2];
 real sp;
  real f[R];
  # first the likelihood using MVB theory
  for (i in 1:N) {
  for (j in 1:N) {
  for (r in 1:R) {
      f[r] \leftarrow alpha[r] - sqrt(pow(z[r,i,1]-z[r,j,1],2) +
    pow(z[r,i,2]-z[r,j,2],2));
      f[r] < -alpha[r] - (pow(z[r,i,1]-z[r,j,1],2) +
          pow(z[r,i,2]-z[r,j,2],2));
      }
  for (r in 1:R2) {
     LP[r] \leftarrow 0;
     for (k in 1:R) {
       LP[r] \leftarrow LP[r] + b_y[k,r]*f[k];
      for (k in 1:R3) {
        LP[r] \leftarrow LP[r] + rho[k]*b_rho[r,k];
   }
  for (r in 1:R2) {
     up[r] \leftarrow exp(LP[r]);
     }
  sp <- 0;
  for (r in 1:R2) {
    sp \leftarrow sp + up[r];
  for (r in 1:R2) {
      p[r] <- up[r]/sp;</pre>
```

5. Goodness-of-fit. An open problem with latent space approaches involves choosing the dimension of the latent space. Using a higher dimensional latent space provides more flexibility to capture more nuanced interaction patters, but also involves estimating a greater number of parameters. To evaluate goodness-of-fit in our model, we fit the model across multiple villages using latent spaces of dimension ranging from one to five. Figure 3 shows boxplots of BIC across villages. For visualization, we plot the log of the absolute value of BIC. This presentation means that higher values are preferred. We see that the two and three dimensional latent spaces perform better than either lower or higher dimensional options. The performance between two and three dimensions is slightly superior for three dimensional latent spaces but overall similar.

(1)
$$BIC = 2\log\text{-likelihood} - \left(N(D+1) + \frac{R(R-1)}{2}\right)\log(N),$$

where D is the dimension of the latent space. The effective number of parameters is the sum of the number of latent position parameters (ND),

the number of intercept parameters (*D*), and the number of association parameters $\frac{R(R-1)}{2}$.

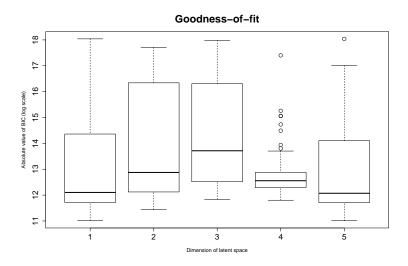


Fig 3. Goodness-of-fit. Plot of the log of absolute value of BIC. Note that after taking the log larger values are preferred.

6. Regression results with 3 dimensional latent spaces. Given the results from the goodness of fit evaluation above, we also fit the model using 3 dimensional latent spaces. The figures below show the results of our hierarchical regression model comparing posterior credible intervals for two dimensional and three dimensional latent spaces.

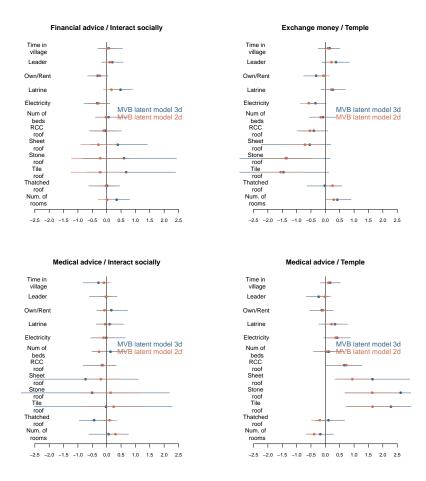


FIG 4. Association results using Karnataka data. Each plot represents a regression model for a particular view pair. Solid dots represent coefficient estimates in a model where covariates are village-level measures of socioeconomic and demographic characteristics for villages. The outcome is either the graph correlation (blue) or the association parameter from the MVB latent model (orange). The results indicate that accounting for network structure through the MVB cluster model can have a substantial impact. All variables are standardized for comparison across models. Additional results are presented in Appendix ??.

7. Additional replication codes.

7.1. *Accessing the data.* Data used for our paper are available through the Harvard IQSS Dataverse, with citation and link below:

Banerjee, Abhijit; Chandrasekhar, Arun G.; Duflo, Esther; Jackson, Matthew

- O., 2013, "The Diffusion of Microfinance", http://hdl.handle.net/1902.1/21538, Harvard Dataverse, V9
- 7.2. *Replication codes.* These codes provide the necessary formatting and functions to move the data from the formate provided in the link above into Stan for analysis with the codes provided above.

```
# move from natural parameter space to probabilities space
f_to_p<-function(f,m.rho,R)</pre>
  {
  b.y<-matrix(FALSE,R,2^R)
  for (i in 1:(2^R))
   b.y[,i]<-as.logical(binary(i-1,R-1))</pre>
  n<-nrow(f)
  # or use these two lines
  eS<-matrix(0,n,2^R)
  for (i in 2:(2^R))
    eS[,i]<-apply(matrix(f[,b.y[,i]],nrow=n),1,sum)
    +rep(0.5*sum(m.rho[b.y[,i],b.y[,i]]),n)
  eS<-exp(eS)
  eb <- apply(eS, 1, sum)
  eS / eb
  }
require(boot)
# vague normal priors on everything
prior.z<-function(z)</pre>
  sum(dnorm(c(z),0,sigma.z,log=T))
prior.alpha<-function(alpha)</pre>
  sum(dnorm(alpha,0,sigma.alpha,log=T))
prior.rho<-function(m.rho)</pre>
  sum(dnorm(m.rho[upper.tri(m.rho)],0,sigma.rho,log=T))
like<-function(decimal.Y,z,alpha,m.rho)</pre>
  # use MVB theory here
  f<-NULL
  for (r in 1:R)
    #f<-cbind(f,c(alpha[r]-as.matrix(dist(z[r,,]))))</pre>
   f<-cbind(f,c(alpha[r]-as.matrix(dist(z[r,,])^2))) # for
        compatibility with rstan
  p<-f_to_p(f,m.rho,R)
  out<-0
  for (i in 1:ncol(p))
```

```
out<-out+sum(log(p[,i])*((decimal.Y+1)==i))</pre>
 out
 }
require(ergm)
require(sna)
require(rstan)
#set mcmc parameters if no defaults
if (!exists("BURN")) BURN=4000
if (!exists("STEPS")) STEPS=8000
if (!exists("THIN")) THIN=5e0
if (!exists("nchains")) nchains=4
#for different datasets, Karnataka replicates the paper
if (!exists("START")) START<-"Karnataka"</pre>
if (START=="SIM") source("sim_setup.R")
if (START=="s50") source("s50.R")
if (START=="Karnataka") source("karnataka.R")
if (START=="LAW") source("lawyers.R")
if (START=="FLO") source("florentine_setup.R")
#generate network matrices
binary <- function(x,n)</pre>
 out<-rev(as.integer(intToBits(as.integer(x))))</pre>
 out[(length(out)-n):length(out)]
b.y<-matrix(FALSE,R,2^R)</pre>
for (i in 1:(2^R))
 b.y[,i]<-as.logical(binary(i-1,R-1))</pre>
# prior standard deviations for latent variables
if (!exists("sigma.z")) sigma.z=10
if (!exists("sigma.alpha")) sigma.alpha=10
if (!exists("sigma.rho")) sigma.rho=10
t.rho<-matrix(0,R,R)
1<-0
for (i in 1:(R-1))
 for (j in (i+1):R)
   {
   1<-1+1
```

```
t.rho[i,j]<-l
b.rho<-matrix(FALSE, 2^R, R*(R-1)/2)
for (r in 1:2^R)
 for (i in 1:(R-1))
   for (j in (i+1):R)
     \verb|b.rho[r,t.rho[i,j]|<-b.y[i,r]&b.y[j,r]|
#reformat so that it is an array for stan
x<-array(0,c(N,N)) # necessary to recast the array Y for
    compatibility with jags / stan
n < -ncol(x)
decimal<-function(y) sum(2^(which(as.logical(rev(y)))-1))</pre>
decimal.Y<-c(apply(Y,c(2,3),decimal)) # takes ages!</pre>
for (i in 1:N)
 for (j in 1:N)
   x[i,j] < -decimal.Y[(j-1)*N+i]
#ready for stan
dat<-list(N=N,R=R,R2=2^R,R3=R*(R-1)/2,x=x,sigma_alpha=sigma.alpha,
sigma_rho=sigma.rho,sigma_z=sigma.z,b_y=b.y*1,b_rho=b.rho*1)
         #simrho=sim.m.rho[upper.tri(sim.m.rho)])
```

References.

- S. E. Fienberg, M. M. Meyer, and S. S. Wasserman. Statistical Analysis of Multiple Sociometric Relations. *Journal of the American Statistical Association*, 80(389):51–67, 1985.
- M. Handcock, A. E. Raftery, and J. M. Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society Series A*, 170:301–354, 2007.
- P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97:1090–1098, 2002.
- P. W. Holland and S. Leinhardt. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373):33–50, 1981.
- M. Salter-Townshend and T. B. Murphy. Variational Bayesian inference for the latent position cluster model for network data. *Computational Statistics & Data Analysis*, 57(1):661 671, 2013.
- S. F. Sampson. Crisis in a cloister. Unpublished doctoral dissertation, Cornell University, 1969.

E-маіl: salter@stats.ox.ac.uk

DEPARTMENT OF STATISTICS UNIVERSITY OF OXFORD

E-маіl: tylermc@uw.edu

DEPARTMENT OF STATISTICS
DEPARTMENT OF SOCIOLOGY
UNIVERSITY OF WASHINGTON

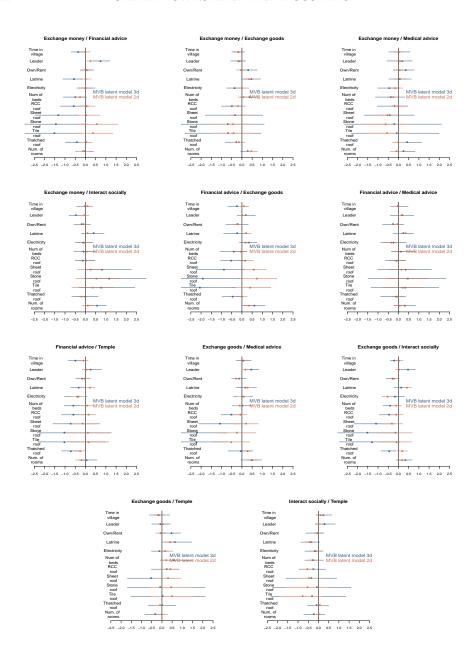


FIG 5. Results for additional view pairs. These plots present coefficients and error bars for the remaining relations not presented in Figure 4. Each plot represents a single view pair. Dots represent point estimates in a regression model where the outcome is either graph correlation or the association parameter in our MVB latent model for a particular village and village level covariates. Bars represent 80% uncertainty intervals. We standardized all variables for comparison across outcomes.