Hierarchical Latent Space Models for Multiplex Social Network Analysis

Alex Loewi

November 30, 2017

Motivation

- I have multiplex data – five edge types ("layers") over 80 nodes
  1. Discuss Academics
  2. Discuss Behavior
  3. Talk Socially
  4. Talk Frequently
  5. Helpful
- I want to understand this data in as many ways as possible BUT
- Models are rare
- The data is complicated
- Visualization is hard

Visualizing Multigraphs

Visualizing Multigraphs

LATENT SPACE MODELS FOR MULTIVIEW NETWORK DATA 1223

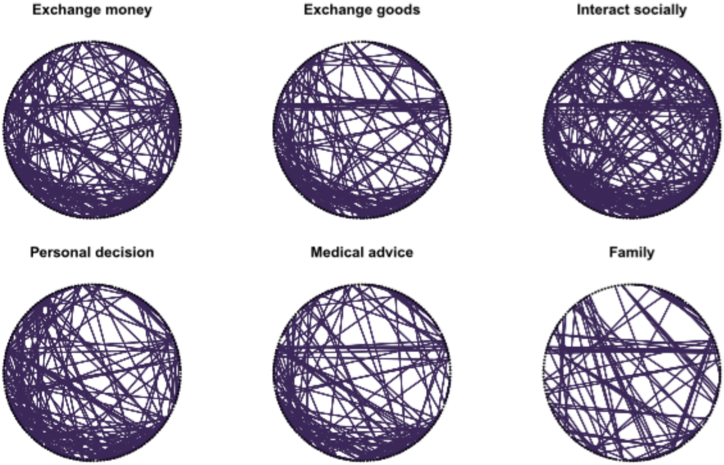
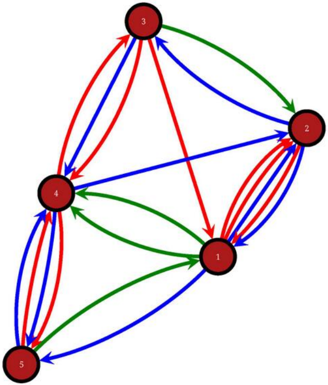


FIG. 1. *Social networks within a village. Actors are arranged in identical order around the outside of the graphs with lines representing edges on a given relationship.*

Visualizing Multigraphs



Visualizing Multigraphs

- Graph **layouts** require node positions to be **flexible**
- Graph **comparisons** require node positions to be **stable**
- Is there a way to manually tune between the extremes?

A Proposal

Hierarchical models are tunable in this way

1. Have a single set of "base" positions for the nodes
2. Positions for a given layer are perturbations from that base position

$$\prod_k \prod_{ij} \sigma_{ijk}^{y_{ijk}} (1 - \sigma_{ijk})^{1 - y_{ijk}}$$

$$\sigma_{ijk} = \text{logit}^{-1}(\alpha_k - d(z_{ik}, z_{jk}))$$

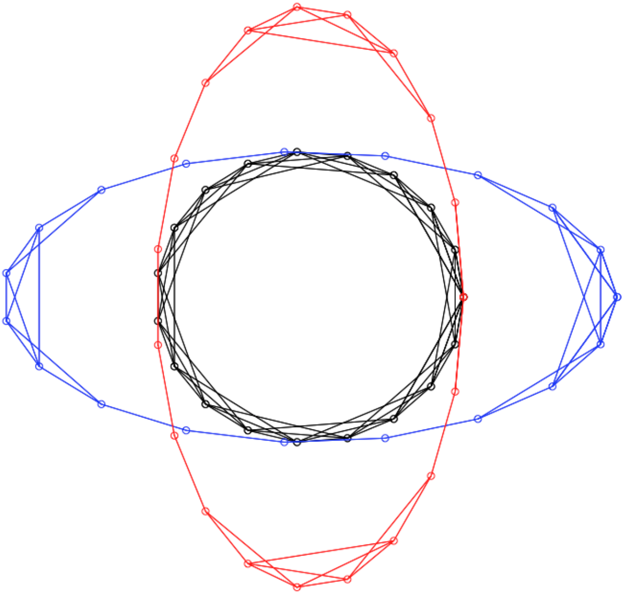
$$z_{ik} = b_i + \epsilon_{ik}$$

3. Regularizing these perturbations allows for tuning
4. (and a lasso penalty grouped layer-wise could reduce the dimensionality)

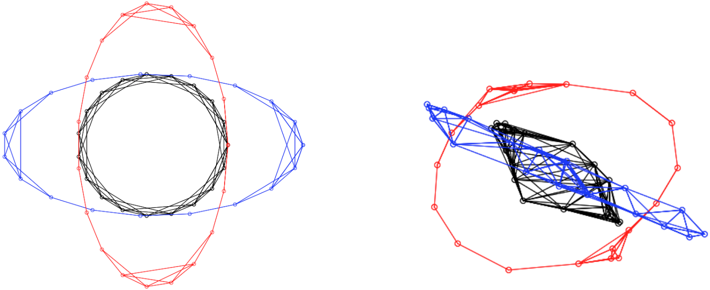
STAN

```
model {
    real d; // just a convenience variable, to split up a long line
    // prior on alpha
    for (k in 1:K) {
        alpha[k] ~ normal(0, sigma_alpha);
    }
    // prior on the base positions
    for(i in 1:N){
        b[i] ~ multi_normal(mu_b, sigma_b); # XXX
    }
    // priors on the layer positions
    for (i in 1:N) {
        for (k in 1:K) {
            z[i,k] ~ multi_normal(b[i], sigma_z[k]);
        }
    }
    # Fitting the parameters to the edges
    for(i in 1:N){
        for(j in 1:N){
            for(k in 1:K){
                d = dot_self(z[i,k] - z[j,k]);
                edges[i,j,k] ~ bernoulli_logit(alpha[k] - d);
            }
```

The Test



Some Results



(a) (b)

Figure: Original, and Estimates

A Recent Model

The Annals of Applied Statistics
2017, Vol. 11, No. 3, 1217–1244
DOI: 10.1214/16-AOAS955

LATENT SPACE MODELS FOR MULTIVIEW NETWORK DATA

BY MICHAEL SALTER-TOWNSHEND[1] AND TYLER H. MCCORMICK[2]

University of Oxford and University of Washington

Social relationships consist of interactions along multiple dimensions. In social networks, this means that individuals form multiple types of relationships with the same person (e.g., an individual will not trust all of his/her acquaintances). Statistical models for these data require understanding two related types of dependence structure: (i) structure within each relationship type, or network view, and (ii) the association between views. In this paper, we propose a statistical framework that parsimoniously represents dependence between relationship types while also maintaining enough flexibility to allow individuals to serve different roles in different relationship types. Our approach builds on work on latent space models for networks [see, e.g., *J. Amer. Statist. Assoc.* **97** (2002) 1090–1098]. These models represent the

A Recent Model

Their model:

1. Separate set of latent variables for each layer
2. Correlations between layers

Intuition: An edge could exist because its nodes are close, OR because its nodes are close in a correlated layer. However ...

“ ... our current model is way too flexible.”

– The Authors

Advantages

- Models are rare
- The data is complicated
- Visualization is hard
- (Model too flexible)

Advantages

- Models are rare
- The data is complicated
- Visualization is hard
- (Model too flexible)

– here's another one!

Advantages

- Models are rare – here's another one!
- The data is complicated – data reduction!
- Visualization is hard
- (Model too flexible)

Advantages

- Models are rare – here's another one!
- The data is complicated – data reduction!
- Visualization is hard – make layers comparable!
- (Model too flexible)

Advantages

- Models are rare – here's another one!
- The data is complicated – data reduction!
- Visualization is hard – make layers comparable!
- (Model too flexible) – this one isn't!

Next Steps

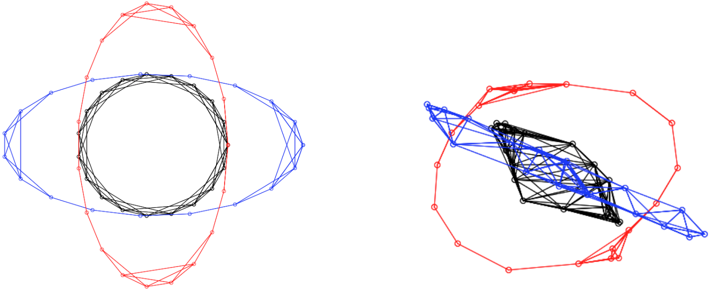
A Full To-Do List

- More test cases
  1. Dimensionality reduction,
  2. Visual comparability,
  3. More complex graphs
- Better theory around fitting
- ... so as to find better *methods* for fitting
- Way to find a good starting point (use normal LSMs?)

Issues already encountered –

- The intercepts are being difficult
- 20k iterations take half an hour, fewer don't burn in
- ...

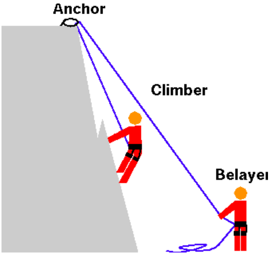
Some Results



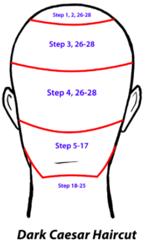
(a) (b)

Figure: Original, and Estimates

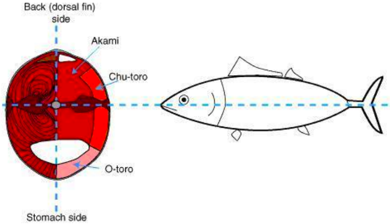
Preparing Your Next Presentation



(a)



(b)



(c)

Figure: Images that Evidently Resemble My Graphics