An Intro. to Latent-Variable Network Models

Santiago Olivella

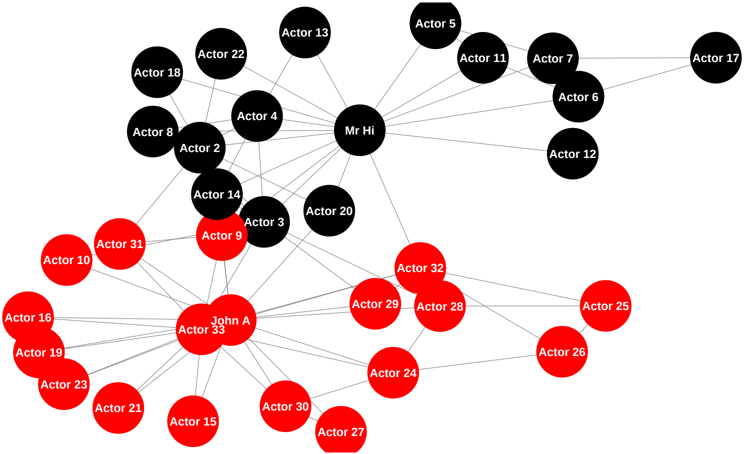
UNC-CH Political Science

June 18, 2020

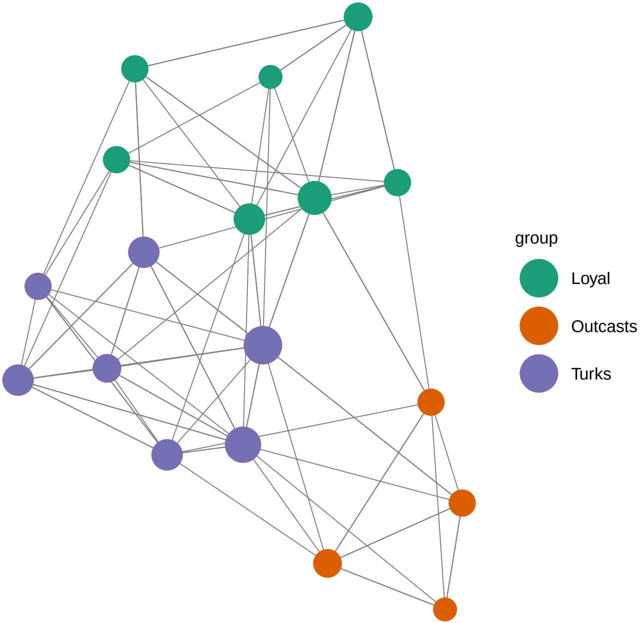
What is a network? What is a graph?

- A **network** is a collection of elements *and* their inter-relations.
  - Sometimes called *relational data*, or *dyadic data*:
    - Friendships
    - Citations
    - Online following/reactions/engagement
    - Co-membership/affiliation
    - Aggressions
    - Spatial proximity
    - ...
- Networks are everywhere!
  - Fun to visualize
  - Hard to describe
  - Harder to model!

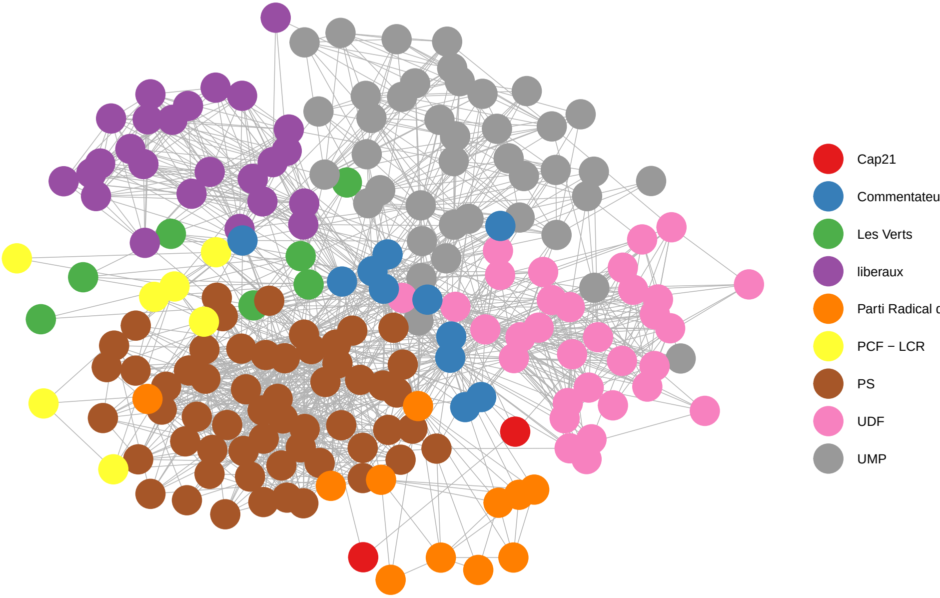
Examples of networks: Karate!



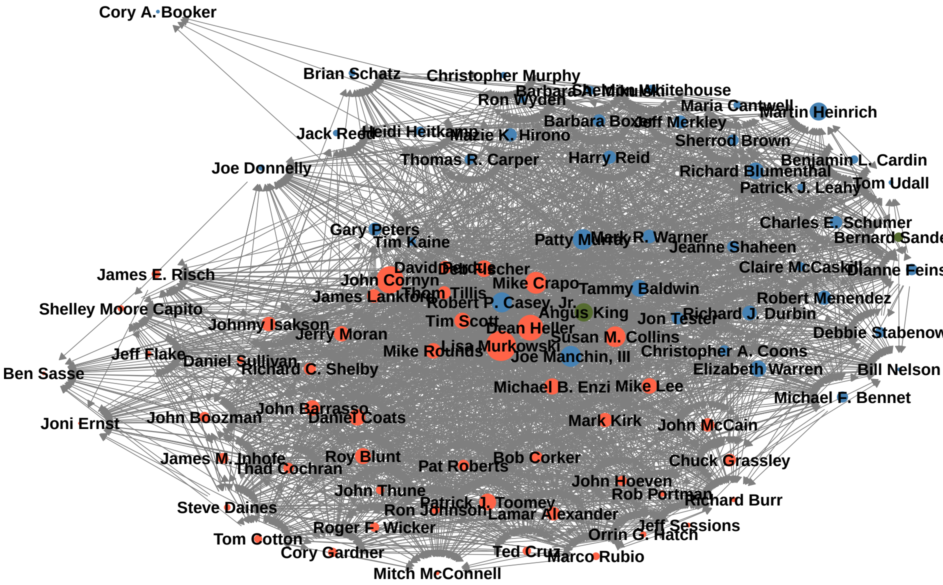
Examples of networks: Monks!



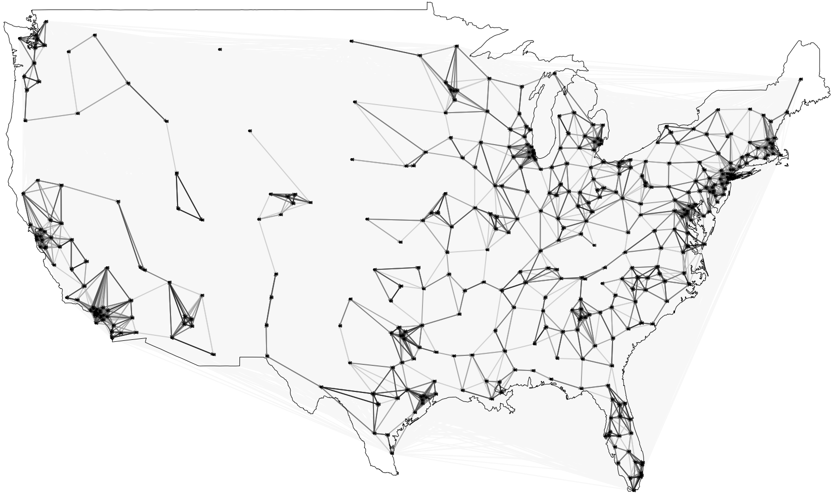
Examples of networks: Blogs!



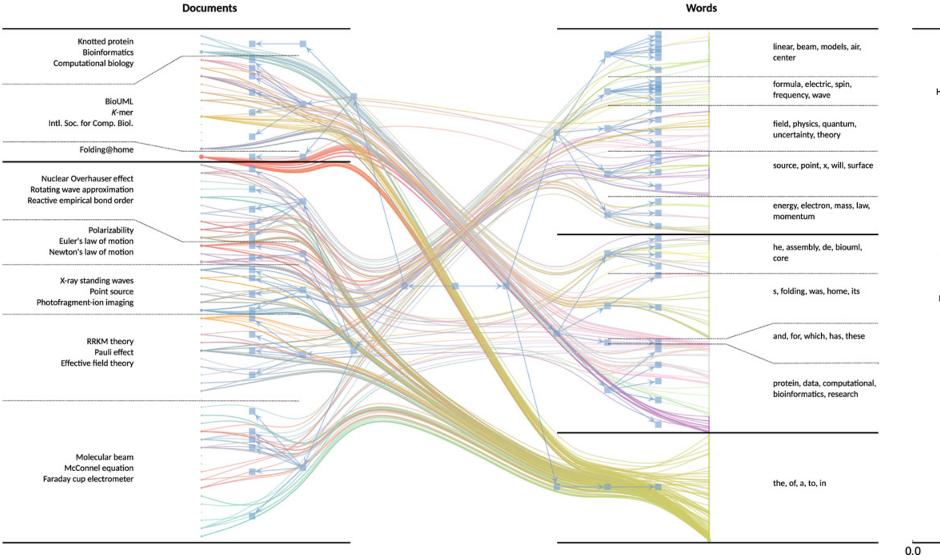
Examples of networks: Tweets!



Examples of networks: Space!



Examples of networks: Text-as-data!



Main concepts I

- A **graph** is the mathematical construct most commonly used to formally represent networks.
  - A graph $G = (V, E)$ is defined by
    - ⋆ A set of **vertices**, or nodes, V
    - ⋆ A set of **edges**, or links, E, s.t. its elements are pairs (i, j) of distinct vertices $i, j \in V$.
- Two vertices $i, j \in V$ are said to be **adjacent** if joined by an edge in E.
  - Also called *neighbors*.
  - Elements in the set of all possible adjacent vertices are called **dyads**.
- Two common representations of a graph:
  - A square **adjacency matrix** (or sociomatrix): if $|V| = N$, $N \times N$ matrix of ones and zeros
    - ⋆ Typically memory-inefficient: most elements are zero (i.e. **sparse**)
  - A two-column **edge-list**: one row per element in E.
- When order of pairs in E matters (i.e. $(i, j) \neq (j, i)$), then graph is **directed** (also called a *digraph*).
  - Typically represented graphically using arrows (rather than line segments)
  - Directed graph $\rightsquigarrow$ asymmetric adjacency matrix

Main concepts II

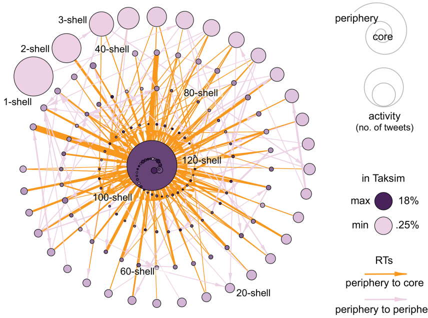
- A graph with multiple kinds of edges (e.g. friendship, romantic, co-authorship) is called a **multigraph** (a.k.a. multiplex/multilevel/multilayer network).
- When edges are not binary, they result in **weighted networks**.
  - E.g. trade volume between countries, in millions of US dollars
  - Sometimes, multigraphs can be turned into weighted networks, but beware of measurement error bias.
- Two edges are said to be adjacent if joined by a common vertex.
  - A vertex i is **incident** on an edge e if i is an endpoint of e, and e is said to be incident on its endpoints.
- The **degree** of a node is the number of edges incident on it (i.e. number of direct connections)
  - For directed graphs, we can distinguish between a node's **in-degree** (total incoming connections, also known as *popularity*) and **out-degree** (total outgoing connections)

Main concepts III

- A **path** is a sequence of unique edges and unique vertices connecting two vertices in V.
  - The length of a path is the number of edges in the sequence.
  - In directed graphs, typically restricted to edges with the same direction.
- A **geodesic** is the shortest path between two nodes
  - The geodesic distance is the length of the shortest path between two nodes
- A **cycle** is a path that begins and ends in the same vertex
  - An **acyclic graph** is one in which there are no cycles.
- A graph is **connected** if there is a path between any two vertices in V
  - A **component** is a maximally connected subgraph of G.

What this talk is *not* about

- Networks are a fascinating and rich topic. But we won't get to some common issues around them:
    - Network visualization
    - Network descriptives (e.g. Centrality)
    - Clustering on networks
    - SUTVA violations
    - Optimization on networks



Source: Barberá et al. 2015

Description vs. inference on graphs

- Until recently, most network analyses where descriptive in nature: characteristics of an observed network where computed and reported
  - A network's 2D representation, produced following algorithms that enable discovery of network features (informal community detection)
  - Describe network characteristics
    - ⋆ Centrality
    - ⋆ Connectivity
    - ⋆ Assortativity
    - ⋆ ...
- Though description can be illuminating, researchers usually care about *inference*:
  - Are there *systematic* associations between node and dyad characteristics and edge formation probabilities?
  - Are these associations the result of mere chance or a properties of a population of networks?

Challenges of network inference

- Answers to inferential questions on graphs present important challenges
- Relational data tend to display outcomes that are the result of endogenous network structure rather than of exogenous characteristics of nodes and dyads
  - E.g. *Popularity*: some nodes tend to have larger in- or out-degrees
  - E.g. *Reciprocity*: if a node dislikes another, the feeling tends to be mutual.
  - E.g. *Homophily/Heterophily*: nodes with similar attributes tend to be connected/disconnected.
  - E.g. *Stochastic equivalence*: nodes with similar attributes tend to have similar connectivity patterns.
  - E.g. *Triadic closure*: two nodes that are friends of a third tend to be friends themselves (consider also Ramsey's theorem)
  - ...
- If these structural outcomes are not accounted for
  - their effect is incorrectly ascribed to exogenous characteristics
  - our estimates of uncertainty will be anti-conservative

Modeling networks: smart first year approach

- Standard statistical tools for inference assume conditional independence across observations (think of binomial regression models)
  - Conditional on predictors, outcomes are assumed to be independent of one another.
  - This allows us to create a simple expression for the *joint* distribution from which data are believed to have been drawn.
- This is clearly false with data that arise due to interdependence.
  - This is not unrelated to the point above about the effects of network topology.

ERGMs: Let's get this out of the way...

- One approach to modeling network data side-steps the issue of independence by defining a distribution over *all* dyads (rather than assuming they are conditionally independent): graph as a random variable G, with observed instance g.
  - This is the approach of **exponential random graph models** (also known as p^* models)
- First, define a set of *sufficient statistics* $\mathbf{S}(g)$:
  - Sum of monadic or dyadic exogenous characteristics
  - Endogenous *motifs* defined in terms of a graph G
    - ⋆ Structural characteristics of a network expressible as sums of products of $y_{i,j}$'s (e.g. reciprocity, transitivity, etc.)

ERGMs: Let's get this out of the way...

- The Hammersley-Clifford Theorem guarantees that any probability distribution over networks $\Pr(G = g)$ can be expressed as

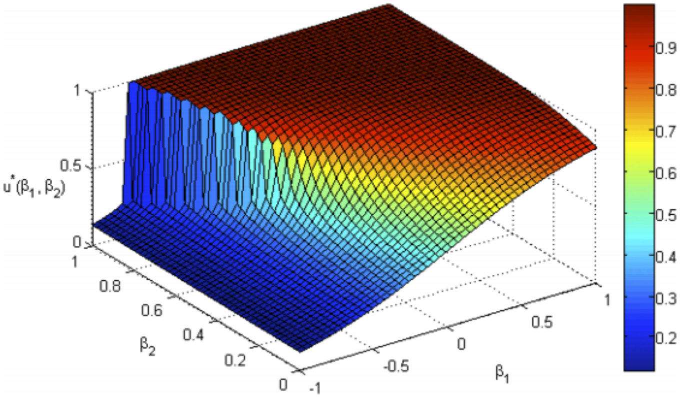
$$\begin{aligned} \Pr(g) &= \exp[\boldsymbol{\theta}^\top S(g) - \psi(g)] \\ &= \frac{\exp(\boldsymbol{\theta}^\top \mathbf{S}(g))}{\sum_{g' \in \mathcal{G}} \exp[\boldsymbol{\theta}^\top \mathbf{S}(g')]} \end{aligned}$$

  where $\mathcal{G}$ is the set of all graphs that can be defined on a node-set of the observed cardinality
    - Thus, the normalizing denominator sums over all possible graphs.
- Pretty powerful stuff!
    - If we get invoke it...

Why not ERGMs?

- The Hammersley-Clifford Theorem only bites if two conditions are met:
  - A correct model specification: $\mathbf{S}(G)$ is the *exhaustive* set of statistics that determine the probability with which graphs are observed.
  - The observed network is "representative": the observed test statistics are *equal* to their expectations
- Both are very tall orders. . .
  - But, okay, *most* models tend to assume both, too.
- A much more serious issue: **even if correctly specified, implied likelihoods may not be well behaved**

Why not ERGMs?



Source: Chattarjee and Diaconis (2013)

Why not ERGMs?

- Other reasons to not use ERGMS:
  - Not robust to missing (even MAR!) edges
  - Normalization constant of likelihood is, in most cases, impossible to compute exactly
    - ★ Standard approximation procedures will take exponential time to explore space of graphs, unless the links in the network are approximately independent
    - ★ Approximation methods also vulnerable to pathologies in likelihood/posterior
  - Hard to specify!
    - ★ What is a distribution over "dyadwise shared partners", and (more importantly) what does it mean substantively?!

Why not ERGMs?

Dyadwise shared partners: The d argument is a vector of distinct integers. This term adds one network statistic to the model for each element in d; the ith such statistic equals the number of dyads in the network with exactly d[i] shared partners. This term can be used with directed and undirected networks. For directed networks the count is over homogeneous shared partners only (i.e., only partners on a directed two-path connecting the nodes in the dyad).

Imported from Launchpad using lp2gh.

- date created: 2010-06-28T17:42:50Z
- owner: gabor.csardi
- the launchpad url was https://bugs.launchpad.net/bugs/599471

digitaldust commented on Mar 27, 2014

hello, is there any implementation of this measure on igraph so far?

gaborcsardi commented on Mar 27, 2014 Contributor Author

To be honest, I am not even sure what this is....

Viable alternative: Latent Network Models

- A popular alternative: a **latent variable model**
  - If structural and exogenous characteristics that determine interdependence can be controlled for, conditional independence can still be justified
- It is possible to account for structural characteristics
  - First order

The social relations model

- Define

$$\Pr(y_{ij} = 1) = \Pr(\mu + \eta_i^s + \eta_j^r + \epsilon_{ij} > 0)$$
$$[\eta_i^s \ \eta_j^r]^\top \sim N_2(\mathbf{0}, \mathbf{\Sigma}_\eta)$$
$$\mathbf{\Sigma}_\eta = \begin{bmatrix} \sigma_{\eta^s} & \sigma_{\eta^s,\eta^r} \\ \sigma_{\eta^s,\eta^r} & \sigma_{\eta^r} \end{bmatrix}$$

  where ϵ_{ij} is Normally distributed, η_i^s is a *sender* random intercept (sociability) and η_j^r is a *receiver* random intercept (popularity): **first order dependence**.
- We can capture **second order dependence** by imposing structure on ϵ_{ij}:

$$[\epsilon_{ij} \ \epsilon_{ji}]^\top \sim N_2(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$$

  such that

$$\mathbf{\Sigma}_\epsilon = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

The social relations regression model

- Can be expanded to incorporate node-level and dyad-level exogenous covariates:

$$\Pr(y_{ij} = 1) = \Pr(\mu_{ij} + \eta_i^s + \eta_j^r + \epsilon_{ij} > 0)$$
$$\mu_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}_d + \mathbf{x}_i\boldsymbol{\beta}_s + \mathbf{x}_j\boldsymbol{\beta}_r$$

- Allow random intercepts and correlation of errors to capture endogenous relationships, estimate effects net of those relationships.
  - Interpret coefficients exactly as you would in other mixed GLM settings.
- Easily extensible to other types of edge types.

The social relations regression model

- Bayesian inference for the SRRM can rely on tried-and-true Gibbs sampling schemes
  - Simply define Normal priors for $\boldsymbol{\beta}_d, \boldsymbol{\beta}_s, \boldsymbol{\beta}_r \ldots$
  - $\ldots$ and an inverse-Wishart for $\boldsymbol{\Sigma}_\eta \ldots$
  - $\ldots$ and an inverse-gamma for $sigma_\epsilon^2$ (and a change of variables for ρ)
- For Normal outcomes (i.e. a weighted network, for instance), the set of full conditional posterior distributions is no different from a random-effects model with a slightly more structured data-level covariance
- For non-Normal outcomes, data augumentation can come to our rescue
  - Augment with Normal RV for probit
  - Augment with Pólya-Gamma RV for logit
- For other models (e.g. Poisson edges), MH can always be implemented.

The AME model

- SRRM great, but still missing ability to capture higher order relationships
  - Homophily
  - Transitivity
  - Stochastic equivalence
  - ...
- For *latent* vectors $\mathbf{u}_{\cdot}$, the latent variable model is given by

$$\Pr(y_{ij} = 1) = \Pr(\mu_{ij} + \eta_i^s + \eta_j^r + \mathbf{u}_i^\top \mathbf{u}_j + \epsilon_{ij} > 0)$$

- Simple inclusion of a **multiplicative term** involving latent vectors $\mathbf{u}_i$ and $\mathbf{u}_j$
  - For unit-length vectors, is simply the cosine of the angle between them: how "close" are they?
  - Also known as *bilinear* model
- Letting the $\mathbf{u}_{\cdot}$ have independent MVN priors, the Gibbs sampler for the SRRM is easy to update
  - Note that the model depends on $\mathbf{U}$ only though inner products of the form $\mathbf{u}_i^\top \mathbf{u}_j$

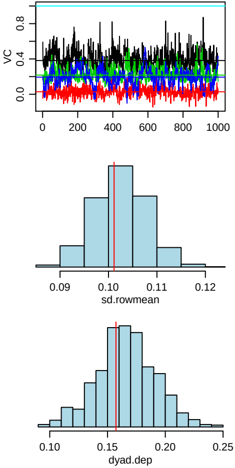
The AME model: Example

```
fit_AME <- ame(Y, Xd=Xd, Xr=Xno, Xc=Xno,
               R=3,
               model="bin",
               seed = 83123,
               nscan = 10000,
               burn = 1000,
               odens = 10,
               plot=FALSE)
```

The AME model: Example

```
##
## Regression coefficients:
##                 pmean   psd z-stat p-val
## intercept       0.092 1.067  0.087 0.931
## status.row     -0.071 0.301 -0.237 0.813
## female.row     -0.133 0.225 -0.590 0.555
## seniority.row -0.006 0.019 -0.341 0.733
## age.row        -0.025 0.015 -1.680 0.093
## practice.row   -0.043 0.196 -0.222 0.825
## status.col     -0.467 0.268 -1.744 0.081
## female.col     -0.014 0.199 -0.071 0.943
## seniority.col -0.004 0.016 -0.284 0.776
## age.col        -0.004 0.012 -0.303 0.762
## practice.col    0.013 0.171  0.078 0.938
## advice.dyad    -0.129 0.113 -1.146 0.252
## cowork.dyad     1.580 0.101 15.693 0.000
##
## Variance parameters:
##     pmean   psd
## va  0.395 0.100
## cab 0.029 0.053
## vb  0.225 0.070
## rho 0.191 0.096
## ve  1.000 0.000
```

AME: Measures of model fit



AME in Political Science I

170 Peter D. Hoff and Michael D. Ward

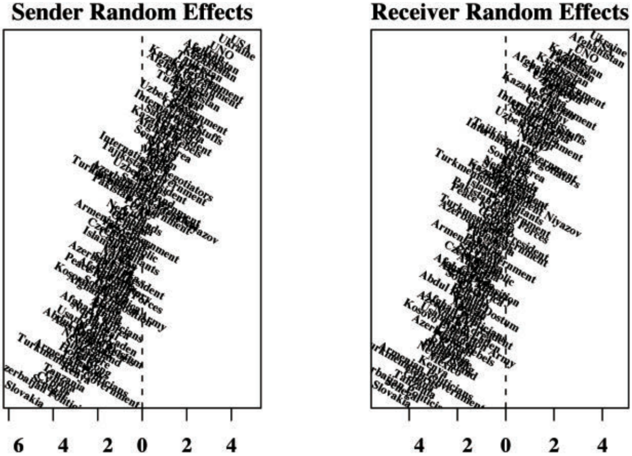


Fig. 2 Orderings of sender-specific and receiver-specific random effects are similar for the conflict data. The United States, the United Nations, Ukraine, Afghanistan, Iran, Kazakhstan, and Pakistan are actors with large, positive random effects sending and receiving both conflict and cooperation

AME in Political Science II

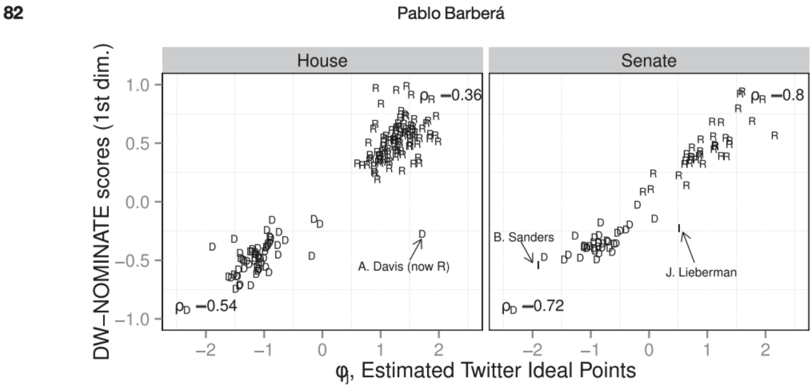


Fig. 1 Ideal point estimates for members of US Congress.

The stochastic block model

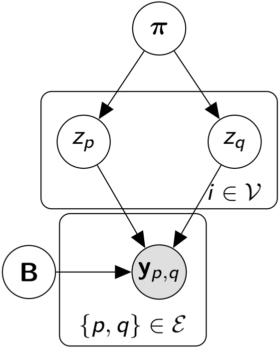
- Now, consider a categorical (i.e. indicator) $\mathbf{u}_i$, so that $u_{i,k} = 1$ iff $i \in k$, and $u_{i,k} = 0$ otherwise. We can define a multiplicative term

$$\mathbf{u}_i^\top \mathbf{B} \mathbf{u}_j$$

  where $\mathbf{B}$, called the **blockmodel**, is the matrix of group-to-group connection probabilities
- This is a kind of model-based **community detection** algorithm
  - Nodes in the same community are said to be members of a *stochastic equivalence* class.

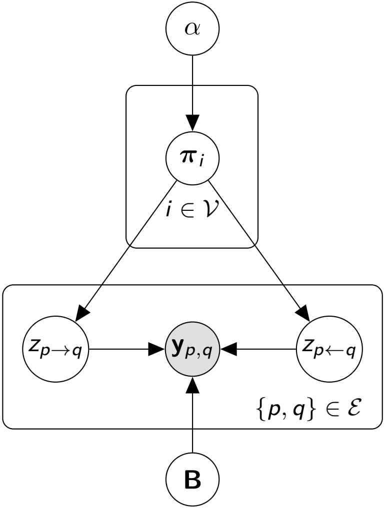
The stochastic Blockmodel

- Let
  - $\mathbf{B}$ be a $K \times K$ matrix of block-dependent edge probabilities for K latent blocks
  - $\boldsymbol{\pi}$ be a K-dimensional vector of block membership probabilities.
  - For each node, sample membership indicator $z_i \sim \text{Multinom}(\boldsymbol{\pi})$.
  - For each dyad, sample edge $y_{p,q} \sim \text{Bern}(z_p^\top \mathbf{B} z_q)$.
- Effectively a mixture model for relational data.



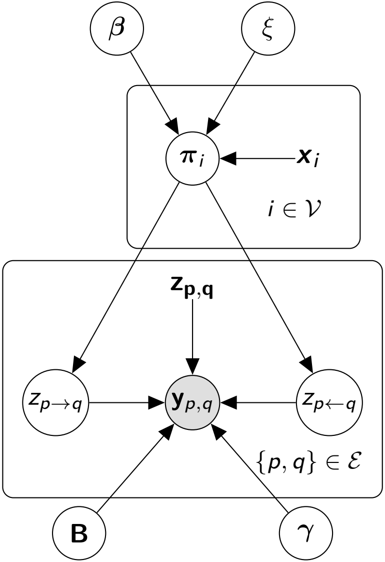
The mixed-membership SBM

1. *For each node*, mixed-membership vector $\boldsymbol{\pi}_i \sim \text{Dirichlet}(\boldsymbol{\alpha})$
2. For each dyad
   - Draw a membership indicator for sender $z_{p \to q} \sim \text{Multinom}(\boldsymbol{\pi}_p)$
   - Draw a membership indicator for receiver $z_{p \leftarrow q} \sim \text{Multinom}(\boldsymbol{\pi}_q)$
   - Draw the edge $y_{pq} \sim \text{Bern}(z_{p \to q}^{\top} \mathbf{B} z_{p \leftarrow q})$
3. Essentially LDA for networks.

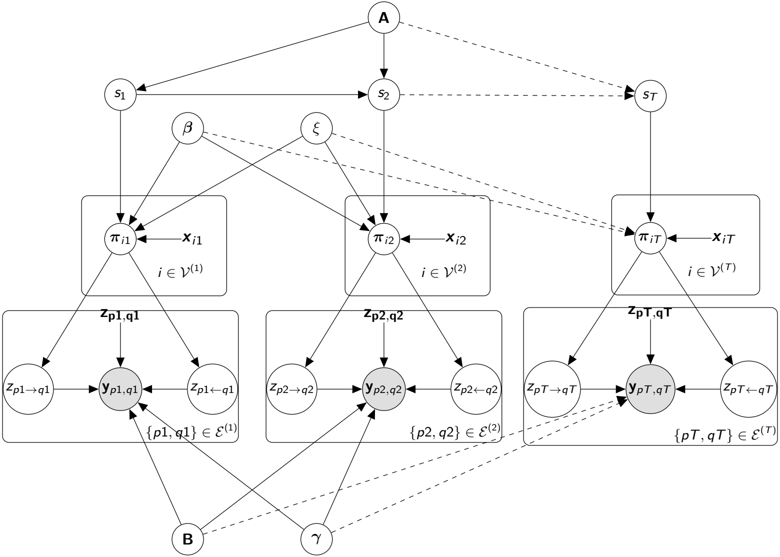


A dynamic, regression MM-SBM

- Regression structure on mixed-membership vectors:
  - For each node i, sample $\boldsymbol{\pi}_i \sim \text{Dirichlet}(g^{-1}(\mathbf{x}_i^\top \beta_{s_t}), \xi)$
  - HMM for mixed-membership regression $s_t \sim \text{Multinom}(A_{s_{t-1}})$.
- For each dyad
  - Sample group membership vectors as before.
  - Draw $y_{pt,qt} \sim \text{Bern}(g^{-1}(z_{p\rightarrow q}^\top \mathbf{B} z_{p\leftarrow q} + \mathbf{z}_{p,q}\gamma))$
- Fast, parallelized collapsed variational EM implemented in C++ and R.



The full model

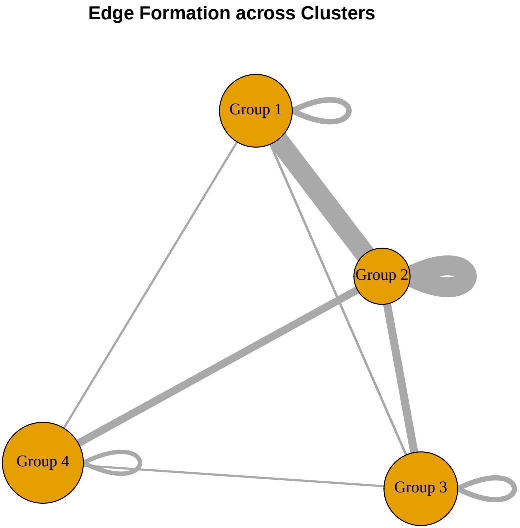


Empirical illustration: MIDS over time

- Undirected MIDs between 1816-2010
- 164 nodes, 195 years, $625,134$ dyads
- Monadic covariates: Polity and CINC scores
- Dyadic covariates: incld. alliance, distance, peace spell spline

Empirical illustration: MIDS over time

- Undirected MIDs between 1816-2010
- 164 nodes, 195 years, $625,134$ dyads
- Monadic covariates: Polity and CINC scores
- Dyadic covariates: incld. alliance, distance, peace spell spline
- "Meta" network of blocks

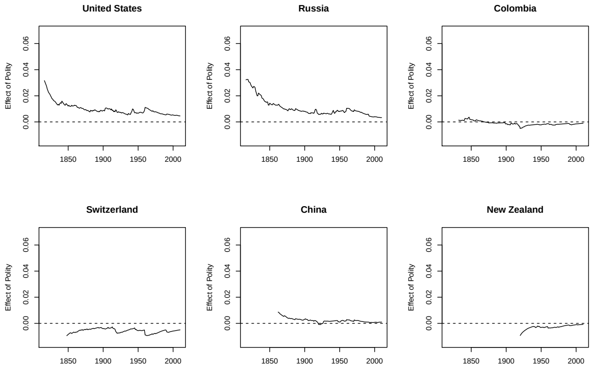


Empirical illustration: MIDS over time

- Undirected MIDs between 1816-2010
- 164 nodes, 195 years, 625, 134 dyads
- Monadic covariates: Polity and CINC scores
- Dyadic covariates: incld. alliance, distance, peace spell spline
- “Meta” network of blocks

Empirical illustration: MIDS over time

- Undirected MIDs between 1816-2010
- 164 nodes, 195 years, 625, 134 dyads
- Monadic covariates: Polity and CINC scores
- Dyadic covariates: incld. alliance, distance, peace spell spline
- "Meta" network of blocks
- Marginal effects of node predictors



Wrapping up

- Networks, and networked data, is everywhere
  - Seriously big data
- All models are wrong, but some are more wrong than others...
  - Stay away from *ERGMs!
  - Maybe design a LASSO ERGM? haven't seen one around, but might work.
- Lots of flavors of latent variable network models
  - Latent distance, bilinear, SBMs...
- Latent variable models offer an easily extensible framework!
  - Hierarchical
  - Dynamic
  - Bipartite? (coming soon!)
- Computational challenges
  - Collapsed Gibbs sampler
  - Approximate methods
    - ★ Variational EM
    - ★ MLE-2 (a.k.a. Empirical Bayes)