Auto-Logistic Actor Attribute Models (ALAAMs)

INSNA Sunbelt XLV Paris 2025

Johan Koskinen

Department of Statistics
Stockholm University
University of Melbourne

June 22, 2025

Preamble

- All material is on the workshop repository https://github.com/johankoskinen/ALAAM
  - Download the RMarkdown file `ALAAM-INSNA-XVL.Rmd`
  - Download the (proto) manual https://github.com/johankoskinen/ALAAM/blob/main/alaam_effects.pdf
  - After the introductory `Rmd` we can do selected parts of `Advanced-ALAAM-INSNA-XVL.Rmd`
- In order to run the Markdown you need
  - The R-package
  - The RStudio interface
- We will predominantly use the packages
  - `sna`
  - `network`
- as well as `balaam.R` from GitHub

What is new

If you have used BayesALAAM before

- Entirely new
  - MultivarALAAM.R ⇒ balaam.R
  - Documentation: alaam_effects.pdf
- Define and estimate the model using
  - Standard formula agree ∼ odegree + mood + sex + simple
  - Main function estimate.alaam returns estimate.alaam.obj
- The object prevBayes
  - Continue previous estimation estimate.alaam.obj
  - recalibrate the proposal variance-covariance matrix
- Model selection
  - Obtain posterior deviance from post.deviance.alaam applied on estimate.alaam.obj
  - Calculate DIC using alaam.dic directly on object returned by post.deviance.alaam
- ... and a lot of other tweaks that may or may not have broken the functionality

Agenda

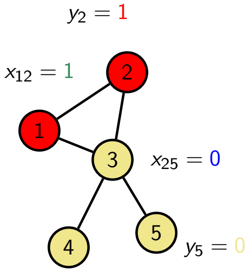
Data structure and homophily

Data structure and homophily

Data - binary outcomes

Tie-variables:

$$X_{ij} = \begin{cases} 1, & \text{if tie from } i \text{ to } j \\ 0, & \text{else} \end{cases}$$



Nodes: $V = \{1, 2, \ldots, n\}$

Adjacency matrix

$$\mathbf{X} = (X_{ij})_{ij \in V \times V} = \begin{bmatrix} \cdot & 1 & 1 & 0 & 0 \\ 1 & \cdot & 1 & 0 & 0 \\ 1 & 1 & \cdot & 1 & 1 \\ 0 & 0 & 1 & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot \end{bmatrix}$$

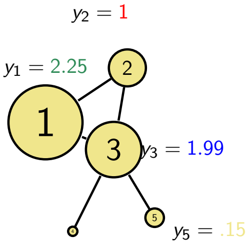
Attribute vector

$$\boldsymbol{y} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \end{bmatrix}^\top$$

Data - Continuous outcomes

Tie-variables:

$$X_{ij} = \begin{cases} 1, & \text{if tie from } i \text{ to } j \\ 0, & \text{else} \end{cases}$$



Nodes: $V = \{1, 2, \dots, n\}$

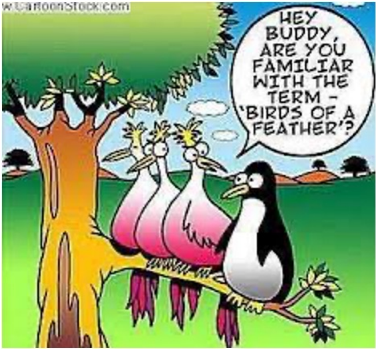
Adjacency matrix

$$\mathbf{X} = = \begin{bmatrix} \cdot & 1 & 1 & 0 & 0 \\ 1 & \cdot & 1 & 0 & 0 \\ 1 & 1 & \cdot & 1 & 1 \\ 0 & 0 & 1 & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot \end{bmatrix}$$

Attribute vector

$$\boldsymbol{y} = \begin{bmatrix} 2.25 & 1 & 1.99 & .05 & .15 \end{bmatrix}^{\top}$$

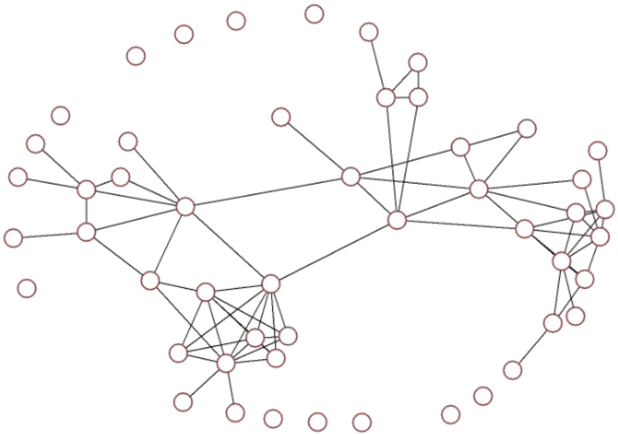
Homophily



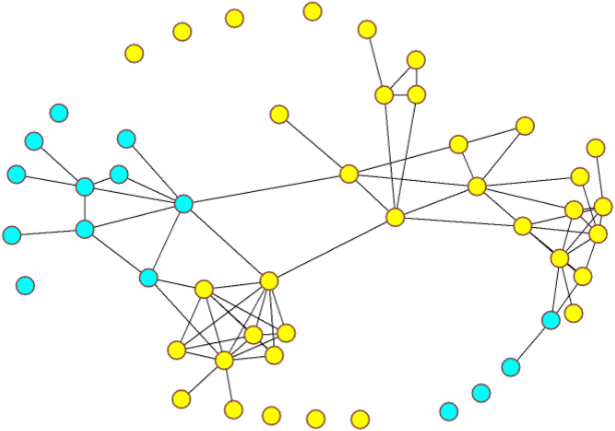
Homophily - empirical evidence

We repeatedly **observe** that people that are *similar* hang together

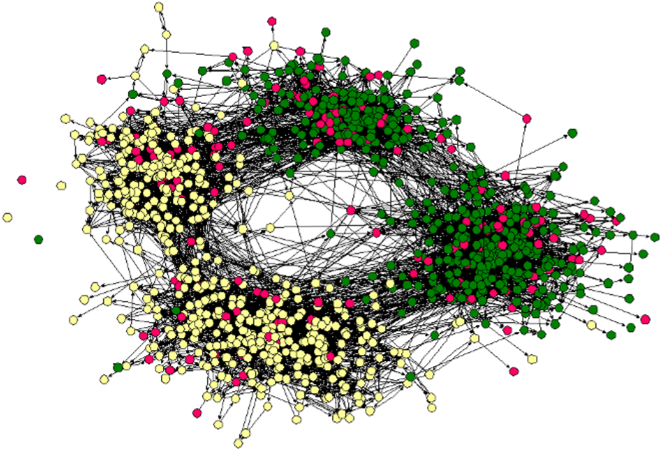
Yuval Kalish - school kids Israel (1)



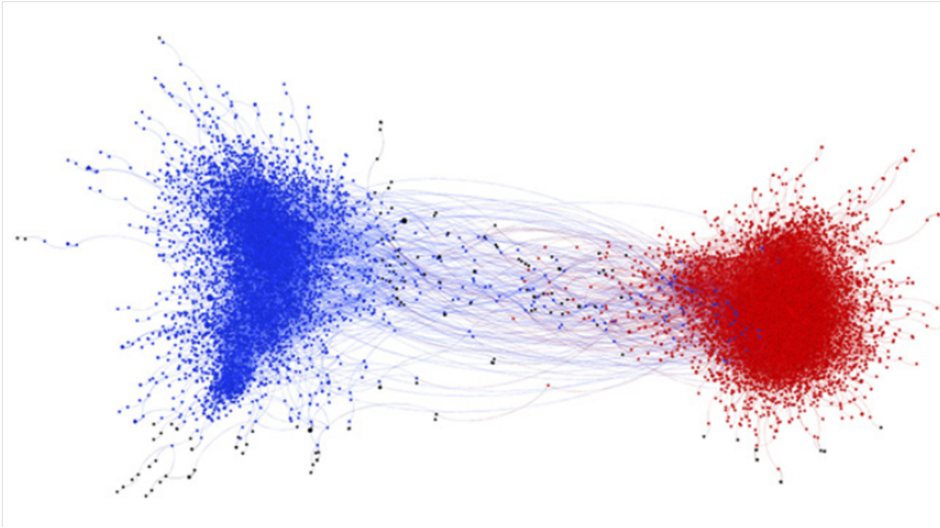
Yuval Kalish - school kids Israel (2)



Ad Health - school kids of different races (Moody et al)



Democrats and Republicans on twitter (Brady et al., 2017)

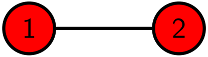


Homophily

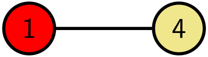
Hanging together, being similar

Binary Y_i:

$\mathbb{1}_{\{y_i=y_j\}}x_{ij}$



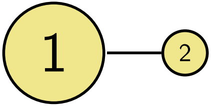
$y_1 = 1$ $y_2 = 1$



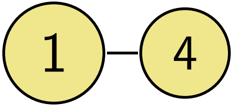
$y_1 = 1$ $y_2 = 0$

Continuous Y_i:

$|y_i - y_j|x_{ij}$

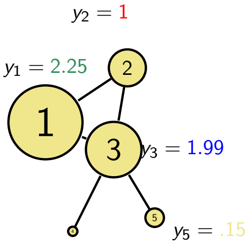


$y_1 = 2.25$ $y_2 = 1$



$y_1 = 2.25$ $y_2 = 1.99$

Correlations of continuous outcomes (1)



Nodes: $V = \{1, 2, \ldots, n\}$

Correlation for variables U and V:

$$corr(U, V) = \frac{\sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^{n}(u_i - \bar{u})^2 \sum_{i=1}^{n}(v_i - \bar{v})^2}}$$

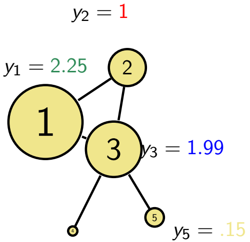
For **network correlation** we only need associations

$$(y_i - \bar{y})(y_j - \bar{y})$$

Between i and j that are connected, i.e.

$$x_{ij} = 1$$

Correlations of continuous outcomes (2)



Nodes: $V = \{1, 2, \dots, n\}$

Moran's I:

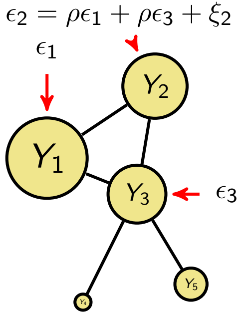
$$I_k = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} (y_i - \bar{y})(y_j - \bar{y}) x_{ij}^{(k)}}{\sum_{i,j} x_{ij}^{(k)} \sum_{j=1} y_j^2}$$

Where

$$x_{ij}^{(1)} = x_{ij}$$

and $x_{ij}^{(2)}$ if i and j are at a distance of 2, etc
If $I_1, I_2, \dots$ are large, neighbouring nodes have similar values

Network autocorrelation model



Nodes: $V = \{1, 2, \ldots, n\}$

The **Network autocorrelation model**:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$$

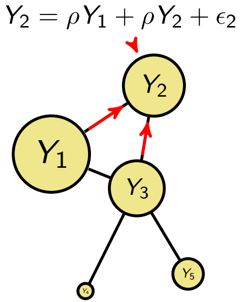
assumes a linear model where there is autocorrelation between the errors

$$\epsilon_i = \rho \sum_{j=1}^{n} w_{ij}\epsilon_j + \xi_i,\ \xi_i \overset{iid}{\sim} N(0, \sigma^2)$$

where w_{ij} is either x_{ij} or a scaled version of it

Note: if $\rho = 0$, we have regular regression with independent observations

Network effects model



Nodes: $V = \{1, 2, \ldots, n\}$

The **Network effects model**:

$$Y_i = \rho \sum_{j=1}^{n} w_{ij} Y_j + \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$$

where w_{ij} is either x_{ij} or a scaled version of it, assumes a non-linear model where there is autocorrelation between the outcomes. Errors are assumed independent

$$\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

Note: if $\rho = 0$, we have regular regression with independent observations

ALAAM - The basic model

The Model

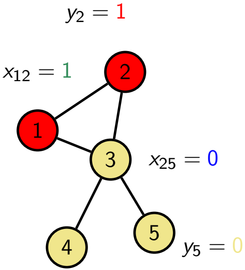
Data - binary outcomes

We want to model binary **Attribute vector**

$$\boldsymbol{y} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \end{bmatrix}^{\top}$$

conditional on **Adjacency matrix**

$$\mathbf{X} = \begin{bmatrix} \cdot & 1 & 1 & 0 & 0 \\ 1 & \cdot & 1 & 0 & 0 \\ 1 & 1 & \cdot & 1 & 1 \\ 0 & 0 & 1 & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot \end{bmatrix}$$



Marginally, we can think of modelling the probabilities

$$p_i = \mathbb{E}(Y_i \mid \mathbf{X}) = \Pr(Y_i = 1 \mid \mathbf{X})$$

Logit: Log-odds

Odds

For a probability p, we define the odds as

$$\frac{p}{1-p}.$$

which is always positive, and increases with p (e.g. $\frac{0.9}{0.1} > \frac{0.5}{0.5} > \frac{0.1}{0.9} > 0$)

Logit

The logarithm of the odds, the log odds, is called *logit*

$$\mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

which is increases with p but $\mathrm{logit}(p)$ takes all values in $\mathbb{R}$

The logit maps probabilities to all values in $\mathbb{R}$

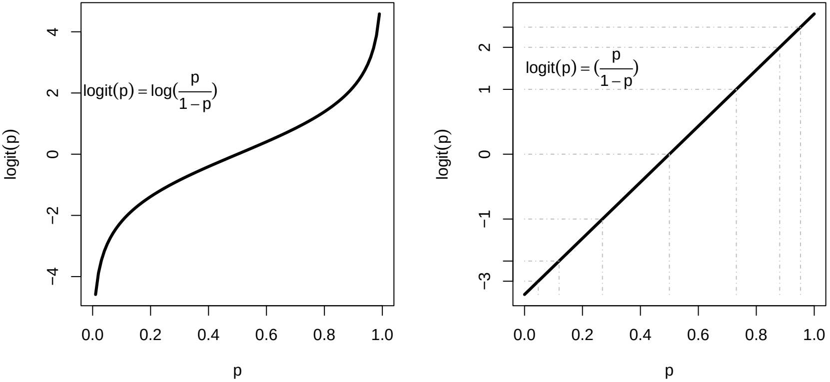


Figure: Probabilities against logit function. Logit scale on vertical axis (right)

Logit link function and linear predictors

While

$$0 \leq p \leq 1$$

as $\mathrm{logit}(p) \in \mathbb{R}$, we write

$$\mathrm{logit}(p_i) = \eta_i$$

where η_i is the **linear predictor**

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

which is a linear regression

The logit-link function

For any $x_{i1}, \ldots, x_{ik}$, and parameters $\beta_0, \beta_1, \ldots, \beta_k$, we can calculate

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

and use the *inverse logit*

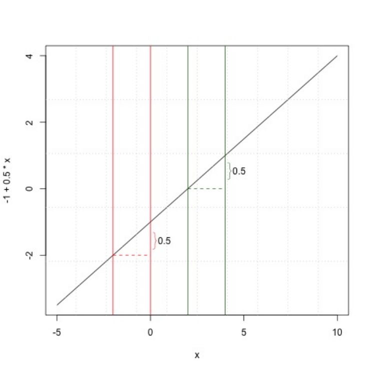
$$\eta_i = \underbrace{\log\left(\frac{p_i}{1-p_i}\right)}_{\textit{logit}} \quad \underbrace{\Rightarrow}_{\text{solve for } p_i} \quad p_i = \underbrace{\frac{e^{\eta_i}}{1+e^{\eta_i}}}_{\text{inverse logit}}$$

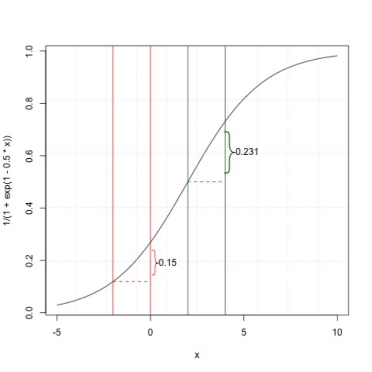
to calculate the probability p_i

Probabilities and the linear predictor: non-linearity

Example (Hypothetical example)

Assume $\text{logit}\,[E(Y \mid x)] = \eta$, where $\eta = -1 + 0.5x$





$$\eta_1 - \eta_0 = \beta_1(x_1 - x_0), \text{ but } \text{logit}^{-1}(\eta_1) - \text{logit}^{-1}(\eta_0)$$

Binary outcomes

We want to model

$$p_i = \Pr(Y_i = 1 \mid \mathbf{X})$$

and if the Y_i are independent

$$p(\boldsymbol{y} \mid \mathbf{X}) = \prod_{i=1}^{n} \Pr(Y_i = y_i \mid \mathbf{X})$$

where for each $i = 1, \dots, n$

$$\Pr(Y_i = 1 \mid \mathbf{X}) = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$

Auto-Logistic Actor Attribute Model (ALAAM)

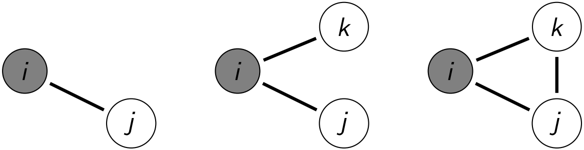
What if we let $\Pr(Y_i = 1|\mathbf{X})$ depend on i's position in the network? For example

$$\eta_i = \beta_0 + \beta_{\text{deg}} \sum_j x_{ij} + \beta_{\text{var}} \sum_{j,k} x_{ij} x_{ik} + \beta_{\text{tri}} \sum_{j,k} x_{ij} x_{ik} x_{jk}$$

which gives us a model

$$p(\boldsymbol{y} \mid \mathbf{X}) = \exp\left\{\boldsymbol{\beta}^\top z(\boldsymbol{y}, \mathbf{X}) - \psi(\boldsymbol{\beta})\right\}$$

where $z(\boldsymbol{y}, \mathbf{X}) = (z_1, \ldots, z_p)^\top$, $z_1 = \sum y_i$, and



$z_2 = \sum_i y_i x_{i+}$ $\quad$ $z_3 = \sum_i y_i \sum_{j,k} x_{ij} x_{ik}$ $\quad$ $z_4 = \sum_i y_i \sum_{j,k} x_{ij} x_{ik} x_{jk}$

Auto-Logistic Actor Attribute Model (ALAAM)

If $\beta_{\mathrm{deg}} > 0$ then nodes with high degree centrality are more likely to have $y_i = 1$ than nodes with low degree

The network activity ALAAM

Frank and Strauss (1986) derived an ERGM for interdependent network *ties* from a Markov dependence assumption. For attributes:

Markov dependence assumption (Robins et al., 2001)

Considering the collection of variables $\mathbf{M} = (\boldsymbol{y}, \mathbf{X})$
Let variables M_u and M_v be conditionally independent if $u \cap v = \emptyset$

Example (Conditionally dependent variables)

The outcomes Y_i and X_{ij} are conditionally dependent as $\{i\} \cap \{i, j\} = \{i\}$

Example (Conditionally independent variables)

The outcomes Y_i and X_{kj} are conditionally independent as $\{i\} \cap \{i, j\} = \emptyset$

Deriving model from dependence (as in ERGM)

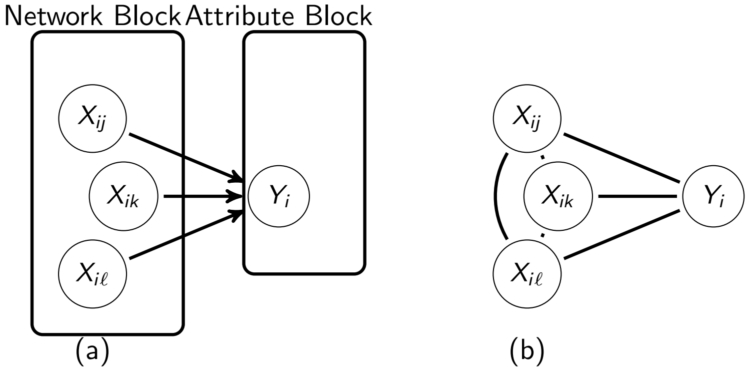


Figure: Dependence graph (a) and Moral graph (b) of network activity dependence model (Robins et al., 2001)

The network activity ALAAM

The statistics z_r correspond to cliques in the Moral graph, and includes

- intercept: $\sum y_i$
- degree: $\sum y_i \sum_j x_{ij}$
- stars: $\sum y_i \sum x_{ij_1} \cdots x_{ij_k}$

But crucially, **no** statistics of the type

$$y_i y_j x_{ij}$$

and thus Y_i and Y_j are independent given $\mathbf{X}$

$$\Pr(Y_i = y_i, Y_j = y_j \mid \mathbf{X}, \boldsymbol{y}_{-i,j}) = \Pr(Y_i = y_i \mid \mathbf{X}, \boldsymbol{y}_{-i,j}) \Pr(Y_j = y_j \mid \mathbf{X}, \boldsymbol{y}_{-i,j})$$

The network activity ALAAM - logistic regression

The network activity ALAAM is equivalent to logistic regression with

$$\mathrm{logit}(p_i) = \beta_0 + \beta_1 z_{i1} + \cdots + \beta_1 z_{ip}$$

where the statistics z_{ih} are summaries of i's network position

The network activity ALAAM - logistic regression

Example (Modern contraceptive use in rural Kenya)

	Mean	Description
mcUse	0.35	Do you use modern contraceptive (MC) techniques?
Age	34.41	Age (sd:16.04)
Female	0.60	Female (1) or Male (0)
HasChildren	0.68	Have one child or more
relevanOthersApprove	0.45	Other people's approval is important
relevanOthersUse	0.67	I care if other people use MC
mcUseConflict	0.68	The use of MC is contentious and causes conflict
numFriends	0.88	Tallied: the number of names of people they spend their free time with

Table: Variables in Kenya study on Modern contraception usage (Not exact question wordings)(NSF-CMMI-2005661). Modi, Koskinen, DeChurch, Contractor, 2025, SocNet

The network activity ALAAM - logistic regression

Example (Modern contraceptive use in rural Kenya (cont.) $n = 1303$)

Estimated logistic regression

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.6340	0.2601	-2.44	0.0148
Age	-0.0554	0.0067	-8.24	0.0000
Female	-1.0232	0.1538	-6.65	0.0000
HasChildren	1.9622	0.2068	9.49	0.0000
relevanOthersApprove	1.4696	0.1514	9.70	0.0000
relevanOthersUse	0.3415	0.1720	1.99	0.0471
mcUseConflict	-0.3835	0.1474	-2.60	0.0093
numFriends	0.3349	0.0828	4.04	0.0001

How much is the increase in the probability of `mcUse` if you acquire another friend?

How account for dependencies through the network

Intuitively[1], we would want the response of i and j not to be independent

$$\Pr(Y_i = 1, Y_j = 1 \mid Y_{-ij}) \neq \Pr(Y_i = 1 \mid Y_{-ij}) \Pr(Y_j = 1 \mid Y_{-ij})$$

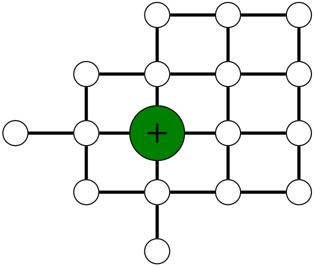
If there is a tie from i to j, $x_{ij} = 1$.
Suggesting a statistic

$$\sum_{i=1}^{n} \underbrace{y_i}_{\text{your succes}} \underbrace{\sum_{j \neq i} y_j x_{ij}}_{\sharp\text{successful friends}}$$

[1] And this is what Robins et al., 2001, did

Ising model (Besag, 1972)

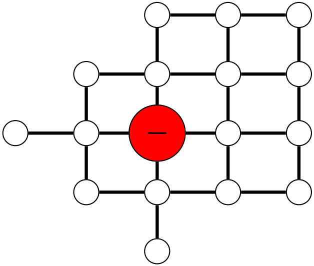
Probability spin $+ \approx \sharp$ neighbours $j \in N(i)$ with spin $+$



$$\Pr(Y_i = 1 | Y_{N(i)} = y_{N(i)}) \propto \exp\{\theta_1 + \theta_2 \sum_{j \in N(i)} y_j\}$$

Ising model (Besag, 1972)

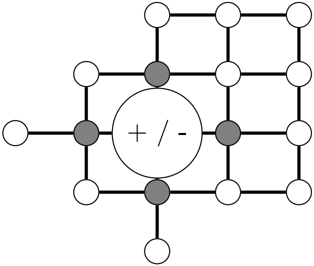
Probability spin $+ \approx \sharp$ neighbours $j \in N(i)$ with spin $+$



$$\Pr(Y_i = 1 | Y_{N(i)} = y_{N(i)}) \propto \exp\{\theta_1 + \theta_2 \sum_{j \in N(i)} y_j\}$$

Ising model (Besag, 1972)

Probability spin $+ \approx \sharp$ neighbours $j \in N(i)$ with spin $+$



$$\Pr(Y_i = 1 | Y_{N(i)} = y_{N(i)}) \propto \exp\{\theta_1 + \theta_2 \sum_{j \in N(i)} y_j\}$$

Markov random fields for Social Networks

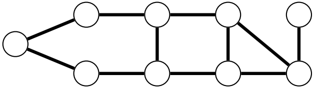
- ppls' networks are not regular lattices
- ppls' attitudes/behaviours also depend on SES, SEX, Education, etc

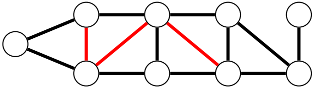
Social dependence is messy

In Graphical models
Conditional independence graph: $i \sim j$ unless

$$X_i \perp X_j | X_{V \setminus \{i,j\}}$$

each node represents one variable (with many observations)
some dependence structures are easier than others





not decomposable

decomposable

Adding dependence between outcomes

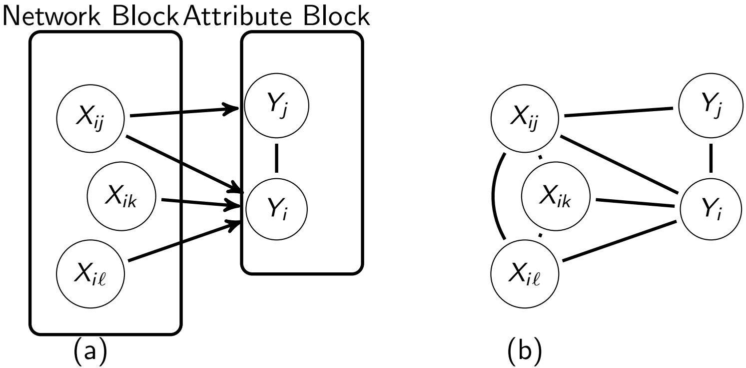


Figure: Dependence graph (a) and Moral graph (b) of model with dependence between attributes that share tie-variables

Deriving contagion statistics is non-trivial

To derive a non-trivial set of statistics use *realization-dependence* (Baddeley & Möller, 1989).

- Partial dependence graph $\mathcal{Q}_{\mathcal{B}}$, is a graph on $\mathcal{V}_{-\mathcal{B}}$
- where $\{i, j\} \in \mathcal{Q}_{\mathcal{B}}$ if
  - ✓ variables i and j are not conditionally independent conditional on variables $\mathcal{V}_{-\mathcal{B},i,j}$,
  - ✓ and all variables corresponding to the index set $\mathcal{B}$ are zero.

In the model, the parameter for the statistic $A \subset \mathcal{V}$ is non-zero only if A is a clique of $\mathcal{M}$ and A is a clique of $\mathcal{Q}_{\mathcal{B}}$ for all $\mathcal{B}$.
Daraganova (2009) - derived statistics

Standard ALAAM

From this, and

- Making some Homogeneity assumptions and
- setting some higher-order statistics to zero,

we arrive at the following contagion model

$$p_{\theta}(\boldsymbol{y}|\mathbf{X}) = \exp\left\{\theta_0 \sum_{i=1}^{n} y_i + \theta_{out} \sum_{i=1}^{n} y_i \sum_{j \neq i} x_{ij} + \theta_{in} \sum_{i=1}^{n} y_i \sum_{j \neq i} x_{ji} + \theta_{con} \sum_{i,j:i \neq yj} y_i y_j (x_{ij} + x_{ji}) - \psi(\theta)\right\}$$

This includes an interaction term similar to that of Besag's (1972) classic auto-logistic model but it is subtly different in the definition of the neighbourhood.

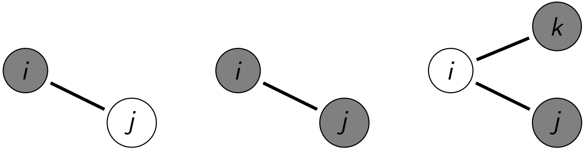
Auto-Logistic Actor Attribute Model (ALAAM)

ALAAM defines a distribution on attributes $\boldsymbol{y} \in \mathcal{Y} = \{0,1\}^V$

ALAAM pmf

$$p_\theta(\boldsymbol{y}|\mathbf{X}) = \exp\{\theta^\top z(\boldsymbol{y};\mathbf{X}) - \psi(\theta)\}$$

ERGM-like model for cross-sectional contagion, e.g.



$\sum_i y_i x_{i+}$ (degree) $\sum_{i,j} y_i y_j x_{ij}$ (contagion) $\sum_{i,j} y_i y_k x_{ij} x_{jk}$ (indirect contagion)

The network activity ALAAM - social influence

Example (Modern contraceptive use in rural Kenya (cont.))

Estimated ALAAM

	Posterior		95% CI	
	Estimate	sd	0.025	0.975
intercept	−0.762	0.291	−1.273	−0.188
contagion	0.457	0.076	0.303	0.592
Age	−0.049	0.007	−0.063	−0.035
Female	−1.091	0.178	−1.461	−0.747
HasChildren	1.710	0.233	1.240	2.154
relevanOthersApprove	1.473	0.165	1.140	1.802
relevanOthersUse	0.353	0.179	−0.005	0.697
mcUseConflict	−0.359	0.164	−0.678	−0.026

How much is the increase in the probability of `mcUse` if your friend uses?

The network activity ALAAM - social influence

A closer look at the pmf

$$p(\boldsymbol{y} \mid \mathbf{X}) = \exp\{\theta^\top z(\boldsymbol{y}; \mathbf{X}) - \underbrace{\psi(\theta)}_{\text{norm. const.}}\} = \frac{e^{\theta^\top z(\boldsymbol{y},\mathbf{X})}}{\underbrace{\sum_{\boldsymbol{y}\in\mathcal{X}} e^{\theta^\top z(\boldsymbol{y},\mathbf{X})}}_{2^n \text{ terms}}}$$

We can **only** evaluate *conditional* probabilities

$$\Pr(Y_i = 1 \mid \mathbf{X}, \boldsymbol{y}_{-i}) = \frac{e^{\theta^\top z(\boldsymbol{y}^{i+},\mathbf{X})}}{e^{\theta^\top z(\boldsymbol{y}^{i+},\mathbf{X})} + e^{\theta^\top z(\boldsymbol{y}^{i-},\mathbf{X})}}$$

where $\boldsymbol{y}^{i+}$ is $\boldsymbol{y}$ with $y_i = 1$, and $\boldsymbol{y}^{i-}$ is $\boldsymbol{y}$ with $y_i = 0$

Estimation

Markov chain Monte Carlo

Simulating from likelihood

We cannot evaluate likelihood for any θ, but for any θ we can simulate Y_i given $y_1, \ldots, y_{i-1}, y_{i+1}, \cdots, y_n$ using probabilities

$$\text{logit}\left\{ \Pr_\theta(Y_i = 1 | \boldsymbol{y}_{-i}, \mathbf{X}) \right\} = \theta^\top \{ z(\boldsymbol{y}^{i+}, \mathbf{X}) - z(\boldsymbol{y}^{i-}, \mathbf{X}) \}$$

giving us samples from

$$\boldsymbol{y} \mid \mathbf{X}, \theta$$

We will use this for

- estimation, and
- goodness-of-fit (GOF)

`MPNet` uses samples in stochastic approximation for MLE

Simulating from likelihood: Metropolis algorithm

Initialising in vector $\boldsymbol{y} := \boldsymbol{y}_0$, in each iteration t

1. Pick $i \in V$ at random
2. Propose to set $y_i := 1 - y_i$
3. Accept and set $\boldsymbol{y}_t := \Delta_i \boldsymbol{y}$, with probability

$$\min \left\{ 1, \exp\{\theta^\top [z(\Delta_i \boldsymbol{y}, \mathbf{X}) - z(\boldsymbol{y}, \mathbf{X})]\} \right\}$$

4. Otherwise set $\boldsymbol{y}_t := \boldsymbol{y}_{t-1}$

This gives us a sequence

$$\underbrace{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k}_{\text{first } k \text{ will remember } \boldsymbol{y}_0}, \overbrace{\boldsymbol{y}_{k+1}, \ldots, \boldsymbol{y}_{T+1}, \boldsymbol{y}_T}^{\text{a } \textit{dependent} \text{ sample}}$$

For sufficiently large burnin k, $\boldsymbol{y}_{k+1}$ a draw from model.

MCMC for un-normalized distributions

MCMC: Sample $\theta^{(0)}, \theta^{(1)}, \ldots$ from $\pi(\theta)$ by

- propose update $\theta^{(t)}$ to θ^{*} $q(\theta^{*}|\theta^{(t)})$
- set $\theta^{(t+1)} := \theta^{*}$ w.p. $\min\{1, H\}$

$$H = \frac{\pi(\theta^{*})}{\pi(\theta^{(t)})} \frac{q(\theta^{(t)}|\theta^{*})}{q(\theta^{*}|\theta^{(t)})}$$

(Works when $\pi(\theta) = f(\theta)/c(\theta)$ and $c(\theta)$ intractable)

Inference: ALAAM

For our target distribtuion $\pi(\theta|z)$

$$H = \frac{\exp\{\theta^{*\top} z(\boldsymbol{y};\mathbf{X}) - \psi(\theta^*)\}\pi(\theta^*)}{\exp\{\theta^{(t)\top} z(\boldsymbol{y};\mathbf{X}) - \psi(\theta^{(t)})\}\pi(\theta^{(t)})} \frac{q(\theta^{(t)}|\theta^*)}{q(\theta^*|\theta^{(t)})}$$

normalising constant $\psi(\cdot)$ of *likelihood* cannot be evaluated (*model is doubly intractable*)

Solution to double intractability

Approximate $\hat{\lambda}(\theta, \theta^*) \approx \exp\{\psi(\theta) - \psi(\theta^*)\}$

- off-line importance sample (Koskinen, 2004)
- 'exact' auxiliary variable-based online importance sample with sample size of 1 - (Møller et al., 2006)
- 'exact' online (linked) path sampler auxiliary variable (Koskinen, 2008; Koskinen, 2009)
- online self-tuning auxiliary variable (Murray et al., 2006) [Approximate Exchange Algorithm]

ERGO: we can obtain posterior for θ when y is observed

Monitoring performance of MCMC

Ideally, in our MCMC sample

$$\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(M)}$$

the samples points are independent draws

$$\theta^{(m)} \overset{iid}{\sim} \pi(\theta | \boldsymbol{y}, \mathbf{X})$$

so that we use Monte Carlo estimators

$$\hat{E}(\theta | \boldsymbol{y}, \mathbf{X}) = \bar{\theta} = \frac{1}{M} \sum_{m=1}^{M} \theta^{(m)} \text{ , and } \widehat{Cov}(\theta | \boldsymbol{y}, \mathbf{X}) = \frac{1}{M} \sum_{m=1}^{M} (\theta^{(m)} - \bar{\theta})(\theta^{(m)} - \bar{\theta})^{\top}$$

as well as approximate probabilities $\Pr(\theta \in C)$, for any $C \subset \Theta$

Monitoring performance of MCMC - trace plots

In plots, *trace plots*, of

$$\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(M)}$$

we should **not** see any

- trend/drift (independence of starting point)
  - select the number of initial iterations to discard - burnin
- serial correlation (good mixing)
  - space out sample points $\theta^{(k)}, \theta^{(2k)}, \theta^{(3k)}, \dots$ - thinning of sample

Monitoring performance of MCMC - SACF & ESS

The *sample autocorrelation function* (**SACF**) measures serial correlation between sample points

$$\theta^{(m-k)}, \theta^{(m)}$$

at different lags k

If SACF at lag k is low, say 30 (SIC?), then taking every k'th sample point will yield an approximately independent sample

The *effective sample size* (**ESS**) tells us roughly how many independent sample points we have

Improving mixing

In our implementation the proposal distribution in each iteration

$$\theta^* \mid \theta^{(t)} \sim N(\theta^{(t)}, \boldsymbol{\Sigma}_p)$$

SACF can be lowered and mixing improved through improved $\boldsymbol{\Sigma}_p$.

Goodness-of-fit

Goodness-of-fit (GOF)

Once we have a draw

$$\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(M)}$$

from $\pi(\theta|\boldsymbol{y})$, we can generate draws

$$\boldsymbol{y}^{(0)}, \boldsymbol{y}^{(1)}, \dots, \boldsymbol{y}^{(M)}$$

each from

$$p_{\theta^{(m)}}(\boldsymbol{y}^{(m)}|\mathbf{X})$$

GOF evaluation

If

$$\boldsymbol{y}^{(0)}, \boldsymbol{y}^{(1)}, \dots, \boldsymbol{y}^{(M)}$$

are 'similar' to $\boldsymbol{y}$, then model has good fit

Model selection

Picking the ‘best’ model

Posterior deviance

The deviance is defined as minus twice the log likelihood

$$D(\boldsymbol{\theta}) = -2\log[p_{\boldsymbol{\theta}}(\boldsymbol{y}|\mathbf{X})].$$

Aitkin et al. (2017) graphical comparison of models can be done through comparing the posterior distribution of the deviance

Assume a sample

$$\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T$$

Calculate the deviance $D(\boldsymbol{\theta}_t)$ for the parameters in your posterior.

Posterior deviance: important

We *cannot* evaluate log likelihood

$$p_{\boldsymbol{\theta}}(\boldsymbol{y}|\mathbf{X}),$$

because of $\psi(\theta)$.
But for pairs $\tilde{\theta}$ and θ, we can approximate $\hat{\lambda}(\theta, \tilde{\theta}) \approx \exp\{\psi(\theta) - \psi(\tilde{\theta})\}$.
Intuition: for bridges $\tilde{\theta} = \theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(M)} = \theta$, we draw

$$\boldsymbol{y}_0^{(j)}, \boldsymbol{y}_{2k}^{(j)}, \ldots, \boldsymbol{y}_{3k}^{(j)}, \boldsymbol{y}_{4k}^{(j)}, \ldots, \boldsymbol{y}_{Tk}^{(j)} \sim p_{\theta^{(j)}}(\boldsymbol{y} \mid \mathbf{X})$$

and use[2] $\bar{\boldsymbol{z}}^{(j)} = \frac{1}{T} \sum \boldsymbol{z}(\boldsymbol{y}_t^{(j)}, \mathbf{X})$ to get estimate
$\hat{\lambda}(\theta, \tilde{\theta}) \approx \exp\{\psi(\theta) - \psi(\tilde{\theta})\}$
NB: Sensitive to T and thinning k - samples $\{\boldsymbol{y}_t^{(j)}\}$ have to be good

[2] Requires a bit more thought ...

Deviance information criterion

Using the posterior distribution of the deviance, we can calculate

$$DIC = E[D(\boldsymbol{\theta})] + V(D(\boldsymbol{\theta}))/2$$

Models with smaller DIC prefered to models with LARGER DIC

Missing data

Missing outcomes

Missing data (cp Bayesian data augmentation for ERGM)

Under assumption of Missing at Random (MAR)
Define the missing data mechanism $f(I|y,\phi)$, where

$$I_i = \begin{cases} 1, & \text{if response } y_i \text{ is unobserved for } i \\ 0, & \text{else} \end{cases}$$

update (impute) missing response by toggling and accepting w.p.

$$\min\left[1, \exp\{\theta^\top(z(\Delta_i y, x) - z(y,x))\}\frac{f(I|\Delta_i y,\phi)}{f(I|y,\phi)}\right]$$

where $\Delta_i y$ is y with element i toggled and set to $1 - y_i$.
Update ϕ, with MH-updating and Hastings ratio

$$\min\left\{1, \frac{f(I|y,\phi^*)\pi(\phi^*)}{f(I|y,\phi)\pi(\phi)}\right\}.$$

Missing data (cp Bayesian data augmentation for ERGM)

In the actual estimation, simply define

$$y_i = \begin{cases} 1, & \text{if response } y_i = 1 \text{ is unobserved for } i \\ 0, & \text{if response } y_i = 0 \text{ is unobserved for } i \\ NA, & \text{if response is missing for } i \end{cases}$$

Sampling will return draws

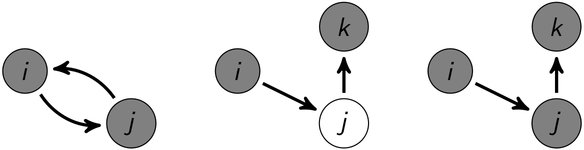
$$(\theta^{(0)}, \boldsymbol{y}_{miss}^{(0)}), (\theta^{(1)}, \boldsymbol{y}_{miss}^{(1)}), \dots, (\theta^{(M)}, \boldsymbol{y}_{miss}^{(M)})$$

Interactions with contagion

More complicated contagion effects

More elaborate effects

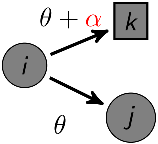
A number of more elaborate forms of contagion/influence are admissible



reciprocal indirect reinforced indir

Interaction:

influence from some nodes can be θ and for others $\theta + \alpha$



SBC (Koskinen and Daraganova, 2022)

Stockholm Birth Cohort (SBC) cohort study, Stockholm Metropolitan area (Stenberg et al., 2006; Stenberg et al. 2007).

- best-friend network with a cap of three nominations (May 1966)
- Let y be indicators $y_i = 1$ of whether pupils i said that they intended to proceed to higher secondary school, and $y_i = 0$ otherwise (see Koskinen and Stenberg, 2012)
- Here: 19 school classes, six of which are from a school in a suburb in the south of Stockholm and the rest are from three inner-city schools
- The proportion of missing entries range from 0 to 0.286

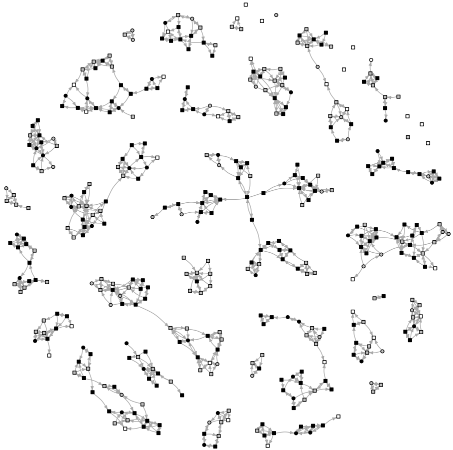


Figure: Bffs in 4 schools. Squares (girl) and circles (boys), and outcome black ($y_i = 1$),

More elaborate effects - interaction example

Example (Simple contagion of intention to go to higher secondary school)

	mean	sd	ESS	SACF 10	SACF 30	2.5 perc	97.5 perc
intercept	-9.67	1.11	178.03	0.68	0.32	-11.83	-7.51
contagion	0.16	0.10	183.10	0.68	0.32	-0.04	0.35
indegree	-0.07	0.11	183.55	0.67	0.32	-0.29	0.13
sex	-0.09	0.29	134.35	0.70	0.39	-0.66	0.47
family attitude	0.48	0.09	164.22	0.70	0.32	0.33	0.65
marks	0.99	0.15	168.66	0.68	0.32	0.69	1.28
social class 1	0.59	0.32	198.40	0.66	0.24	-0.06	1.19

Table: Posterior summaries for model with controls estimated for contagion-model for progression to upper-secondary school in SBC (thinned sample of 10,000 iterations, taking every 20th iteration, with burnin of 1000)

More elaborate effects - interaction example

Example (Contextual contagion of intention to go to higher secondary school)

	mean	sd	ESS	SACF 10	SACF 30	2.5 perc	97.5 perc
intercept	-10.13	1.19	168.32	0.76	0.44	-12.81	-8.04
contagion	0.24	0.12	143.31	0.72	0.39	0.02	0.48
indegree	-0.08	0.12	122.80	0.75	0.41	-0.33	0.13
sex	-0.09	0.28	126.04	0.76	0.45	-0.69	0.47
family attitude	0.48	0.08	140.26	0.72	0.38	0.34	0.65
marks	1.01	0.14	265.08	0.72	0.40	0.76	1.31
composition	0.91	0.55	137.33	0.74	0.39	-0.25	1.97
social class 1	0.57	0.34	143.59	0.73	0.37	-0.07	1.21
contagion int	-0.21	0.16	152.15	0.72	0.37	-0.51	0.11

Table: Posterior summaries for model with controls estimated for contagion-model for progression to upper-secondary school in SBC (thinned sample of 10,000 iterations, taking every 20th iteration, with burnin of 1000) with social class interacted with contagion

Fully Bayesian

Specifying proper priors

Fully Bayesian: what priors?

Assuming

$$\boldsymbol{y} \mid \mathbf{X} \sim ALAAM(\boldsymbol{\theta}, \mathbf{X})$$

The model is

$$P(\boldsymbol{y} \mid \mathbf{X}, \boldsymbol{\theta}) = \exp\{\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{y}, \mathbf{X}) - \psi(\boldsymbol{\theta})\}$$

The aim of the Bayesian inference scheme is to obtain the posterior

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto P(\boldsymbol{y} \mid \mathbf{X}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

where

$$\pi(\boldsymbol{\theta})$$

is the **prior distribution** for the parameters that *quantify our uncertainty about the parameter values prior to observing data*.

Fully Bayesian: what priors?

How can a human *quantify their uncertainty about the parameter values prior to observing data*?
For ALAAM this is really hard!Possible choices

- Default prior: constant $\pi(\boldsymbol{\theta}) \propto 1$
- Convenient: Multivariate normal distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - Diagonal: $\boldsymbol{\Sigma} = \lambda \mathbf{I}$
  - Scaling: $\boldsymbol{\Sigma} = \lambda(\mathbf{X}^\top \mathbf{X})^{-1}$
- Caution: setting $\boldsymbol{\mu} = \mathbf{0}$ pulls posteriors towards 0 - bad for e.g. intercept if $\bar{y}$ small
- Experiemental: use a prior to 'fix' a nuisance parameter

Hierarchical ALAAM

A multilevel version of ALAAM

Hierarchical ALAAM: preamble

The routines for analysing outcomes

$$\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \dots, \boldsymbol{y}^{(G)}$$

for G independently observed networks

$$\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(G)}$$

is implemented in *very old* code in `MultivarALAAMalt.R`

Hierarchical ALAAM: principle

Following Koskinen and Snijders' (2023) work on SAOMs, we assume that independently for each group $g = 1, \ldots, G$,

$$\boldsymbol{y}^{(g)} \mid \mathbf{X}^{(g)}, \boldsymbol{\theta}^{(g)} \sim ALAAM(\boldsymbol{\theta}^{(g)}, \mathbf{X}^{(g)})$$

where parameters are partitioned

$$\boldsymbol{\theta}^{[g]} = \begin{pmatrix} \gamma^{[g]} \\ \boldsymbol{\eta} \end{pmatrix},$$

into a set of group-specific parameters $\boldsymbol{\gamma}^{[g]} \in \boldsymbol{\Gamma} \subset \mathbb{R}^q$, and a common parameter $\boldsymbol{\eta} \in \boldsymbol{H} \subset \mathbb{R}^r$, $q + r = p$.

Hierarchical ALAAM: 'fixed' effects

The common parameter, $\boldsymbol{\eta}$, can be used to parse out **group-level** effects,e.g. Public/private school,gender composition of school class, etc. We may assume that $\boldsymbol{\eta}$, are independent of $\gamma^{[1]}, \gamma^{[2]}, \dots, \gamma^{[G]}$, with prior

$$\pi(\boldsymbol{\eta} \mid \boldsymbol{\mu}_{\boldsymbol{\eta}}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}, \gamma^{[1]}, \gamma^{[2]}, \dots, \gamma^{[G]}) = \pi(\boldsymbol{\eta} \mid \boldsymbol{\mu}_{\boldsymbol{\eta}}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}})$$

which for convenience may be assumed to be $\mathcal{N}_r(\boldsymbol{\mu}_{\boldsymbol{\eta}}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}})$.

Hierarchical ALAAM: ‘random’ effects

The group-level parameters allow, e.g. the intercept and the contagion effect to vary across groups.
We assume

$$\gamma^{[g]} \overset{iid}{\sim} \mathcal{N}_r(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma).$$

Assuming that $\gamma^{[g]}$ follow a multivariate normal, it is common (cp Gelman et al., 1995) to assume a Normal-inverse-Wishard prior for the parameters $\boldsymbol{\mu}_\gamma$ and $\boldsymbol{\Sigma}_\gamma$

$$\boldsymbol{\mu}_\gamma \mid \boldsymbol{\Sigma}_\gamma \sim \mathcal{N}_q(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_\gamma/\kappa_0), \text{ and } \boldsymbol{\Sigma}_\gamma \sim \mathcal{IW}_r(\boldsymbol{\Lambda}_0, \nu_0)$$

Hierarchical ALAAM: the DAG

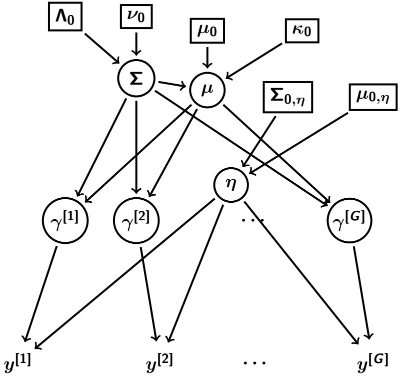
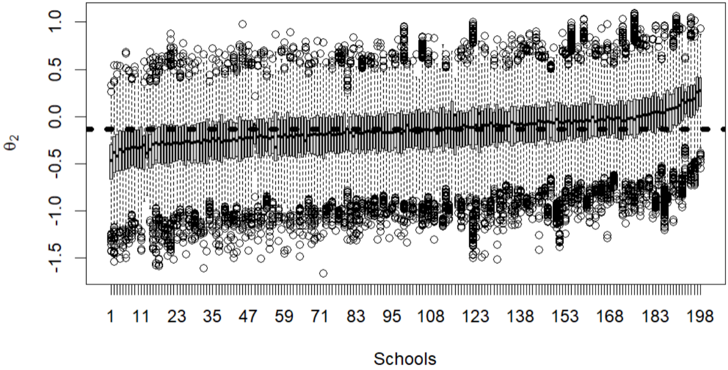


Figure: Dependence structure of hierarchical network model for $\mathbf{Y}^{[g]}$, $g = 1, 2, \dots, G$.

Hierarchical ALAAM: example (1)

Acke Arvidsson (Master's thesis) analysed hundreds of school classes in SBC

Outcome is leader ($y_i = 1$) or not ($y_i = 0$); Network BFF

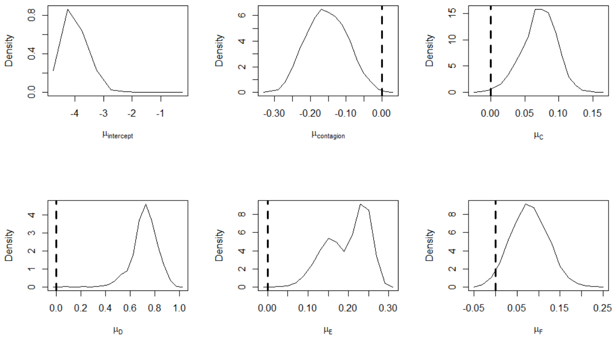


Posteriors for group-level contagion $\gamma_2^{(g)}$

Hierarchical ALAAM: example (2)

Acke Arvidsson (Master's thesis) analysed hundreds of school classes in SBC

Outcome is leader ($y_i = 1$) or not ($y_i = 0$); Network BFF



Posteriors for $\boldsymbol{\mu}$

C: indegree; D: sex; E: average test score; F: communication

Further topics

Further complications

Further topics: topics

- Missing NOT at random (implemented; also, plug for MNAR ERGM, Januar, Gallagher, Koskinen, Friday 27 8.20AM - OS-65)
- Missing network ties
- Marginal effects: Titanic
- Multivariate ALAAM
- Snowball sample or outlier nodes canchange