

Introduction to Social Influence Models

Auto-Logistic Actor Attribute Models (ALAAMs)

Johan Koskinen

Department of Statistics
Stockholm University
University of Melbourne

February 22, 2024



Preamble

- All material is on the workshop repository
<https://github.com/johankoskinen/CHDH-SNA>
 - ▶ Download the RMarkdown file CHDH-SNA-2.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 MultivarALAAMalt.R from GitHub

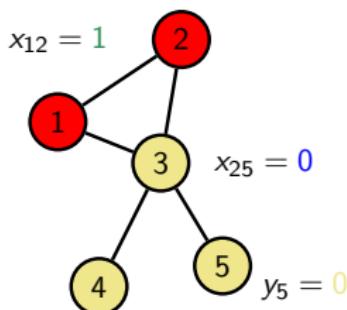


The issue

Tie-variables:

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

$$y_2 = 1$$



Adjacency matrix

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

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

$$\mathbf{y} = [1 \ 1 \ 0 \ 0 \ 0]^\top$$



How model binary outcomes?

i	y_i	x_i
1	0	2
2	1	2
3	1	3
4	1	3
5	0	3
6	1	4
7	1	4
8	1	5
9	1	5

Table: Nine binary observations on a binary dependent variable and covariate

Regressing raw values of y on X ?

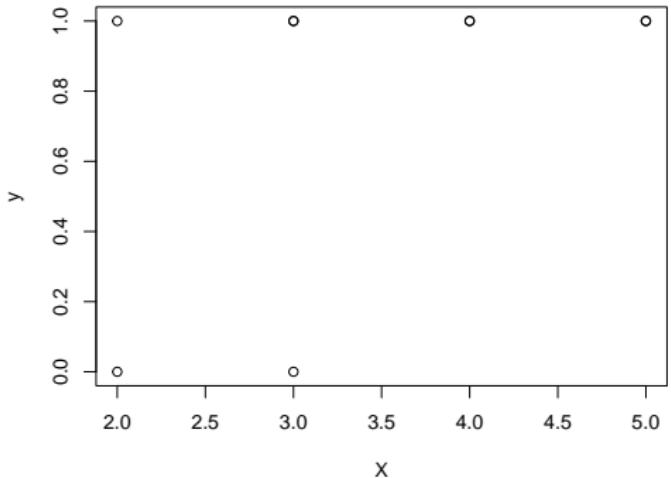


Figure: Scatter plot of data in Table 1



Regressing proportions?

Conditional means in regression:

$$\hat{p}_x = \alpha + \beta x + \epsilon_x$$

	\hat{p}_x	x
1	0.5000000	2
2	0.6666667	3
3	1.0000000	4
4	1.0000000	5

Table: The conditional proportions of the nine binary observations in Table 1 grouped by outcomes on x

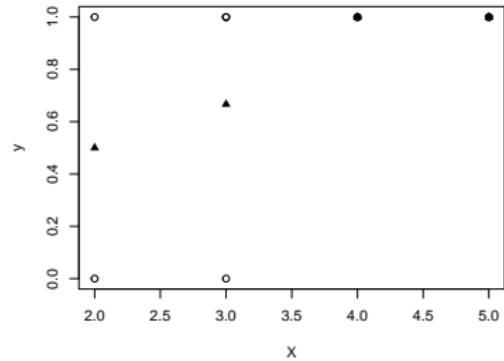


Figure: Scatter plot of data in Table ?? with conditional means (triangles)

Modelling binary dependent variable



Log odds and the logit function

For a probability p , we define the odds as

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

Familiar for gamblers.

For logistic regression it is convenient to work with the logarithm of the odds, the log odds, or *logit*

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

We shall see that $\text{logit}(p)$ takes values in **all of \mathbb{R}**



Logit is a function that maps probabilities \mathbb{R}

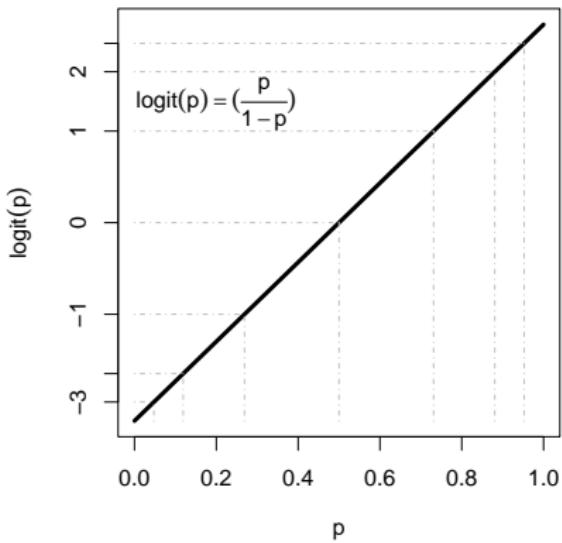
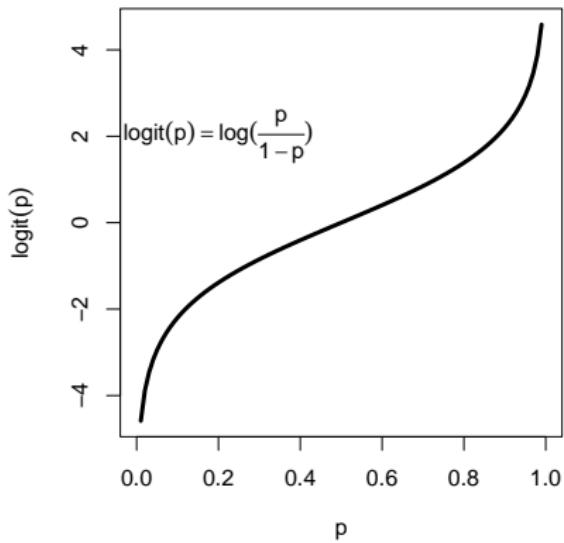


Figure: Probabilities v logit function. Logit scale on vertical axis (right)

Logistic regression

We can regress the logit on any combination of variables

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

We only have a linear predictor for the logit and *NO* random errors.

- increasing function of linear predictor - larger linear comb, large prob
- takes all values on real line



Logistic regression - how probabilities?

For the equation

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

we can easily get the predicted probabilities $\Pr(Y_i = 1 | \mathbf{x}_i)$

$$p_i = \frac{e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}}}$$

to ensure $p+q=1$

A probability in $[0, 1]$

$$p_i + (1 - p_i) = 1$$



Logistic regression

When logit is the **link function**, calling the inverse logit

$$\text{logit}^{-1}(\eta) = G(\eta)$$

G is the **inverse link function**

Logistic regression, is regression of the linear predictor

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

So that

$$E(Y_i | \mathbf{x}) = G(\eta_i)$$



Logistic regression - how probabilities?

Note that we can write

$$\Pr(Y_i = \textcolor{red}{y} | \mathbf{x}_i) = \frac{e^{\textcolor{red}{y}\eta_i}}{1 + e^{\eta_i}}$$

A probability in $[0, 1]$

$$\Pr(Y_i = 0 | \mathbf{x}_i) + \Pr(Y_i = 1 | \mathbf{x}_i) = \frac{e^{0 \times \eta_i}}{1 + e^{\eta_i}} + \frac{e^{1 \times \eta_i}}{1 + e^{\eta_i}} = \frac{1 + e^{\eta_i}}{1 + e^{\eta_i}} = 1$$



Example (cont)

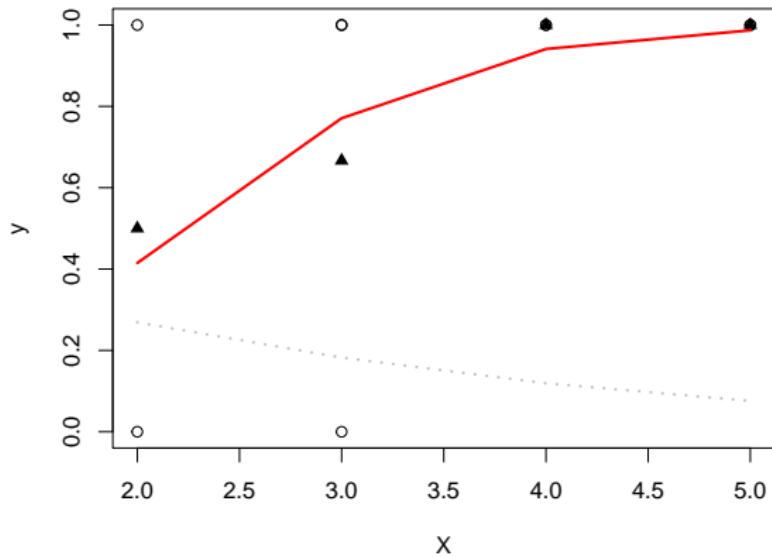


Figure: Original data points of Table 1, with grouped proportions (triangles), the fitted probabilities using β_{ML} (red, solid)

Using network statistics - The Markov model

Auto-Logistic Actor Attribute Model (ALAAM)

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

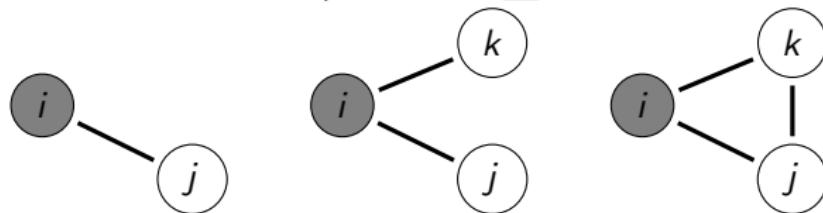
For example

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

which gives us a model

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

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



$$z_2 = \sum_i y_i x_{i+}$$

$$z_3 = \sum_i y_i \sum_{j,k} x_{ij} x_{ik}$$

$$z_4 = \sum_i y_i \sum_{j,k} x_{ij} x_{ik} x_{jk}$$



Auto-Logistic Actor Attribute Model (ALAAM)

If $\beta_{\text{deg}} > 0$ 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 a model for interdependent network *ties* from a Markov dependence assumption

Markov dependence assumption (Robins et al., 2001)

Considering the collection of variables $\mathbf{M} = (\mathbf{y}, \mathbf{X})$ 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 $\{i\} \cap \{i,j\} = \{i\}$

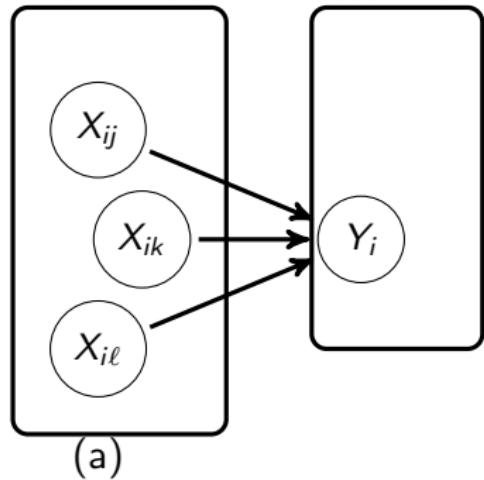
Example (Conditionally independent variables)

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

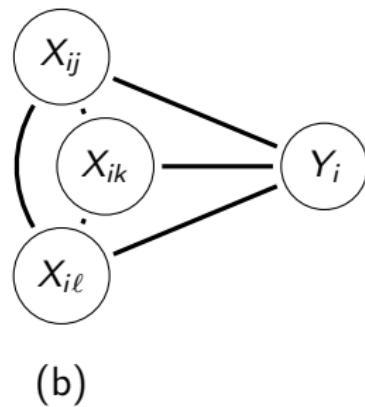


Deriving model from dependence (as in ERGM)

Network Block Attribute Block



(a)



(b)

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}

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



The network activity ALAAM - logistic regression

The network activity ALAAM is equivalent to logistic regression with

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

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 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 modern contraceptives
mcUseConflict	0.68	The use of modern contraceptives 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)

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

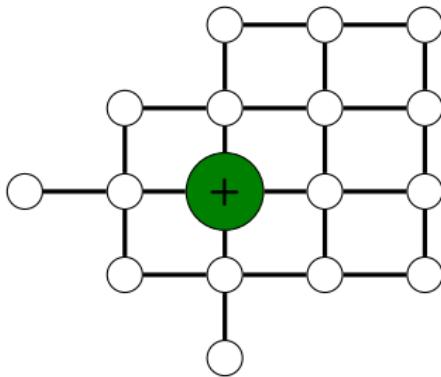
How much is the increase in the probability of mcUse if you acquire another friend?

A network contagion ALAAM



Ising model (Besag, 1972)

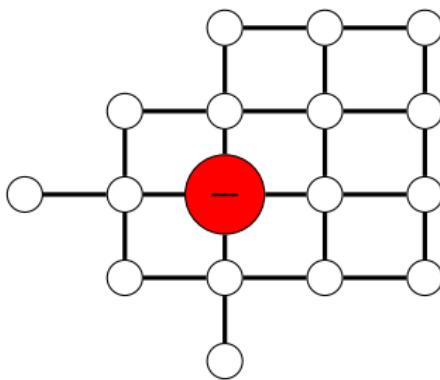
Probability spin + $\approx \#$ neighbours $j \in N(i)$ with spin +



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

Ising model (Besag, 1972)

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

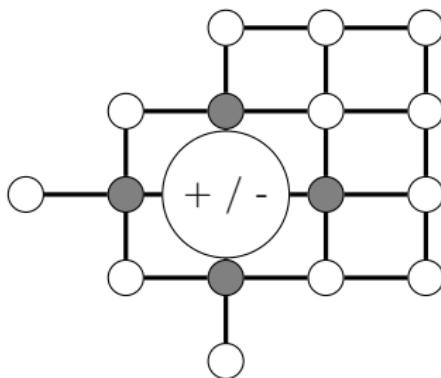


$$\Pr(Y_i = 1 | Y_{N(i)} = y_{N(i)}) \propto \exp\{\theta_1 + \theta_2 \sum_{j \in N(i)} y_j\}$$



Ising model (Besag, 1972)

Probability spin + \approx # 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

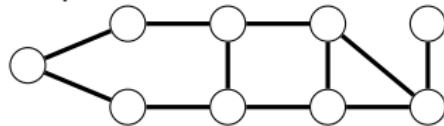
In Graphical models

Conditional independence graph: $i \sim j$ unless

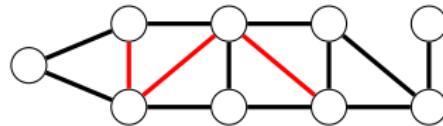
$$X_i \perp X_j | X_{V \setminus \{i,j\}}$$

each node represents one variable (with many observations)

some dependence structures are easier than others



not decomposable

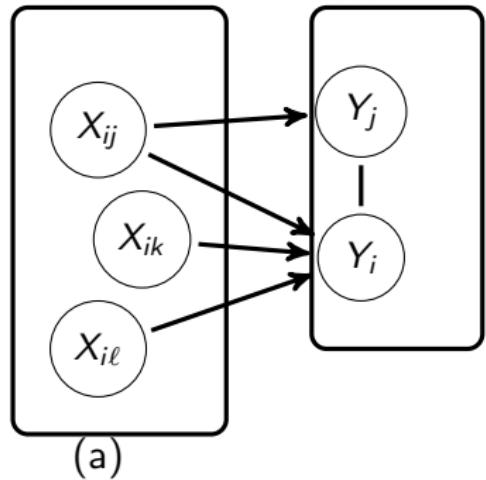


decomposable

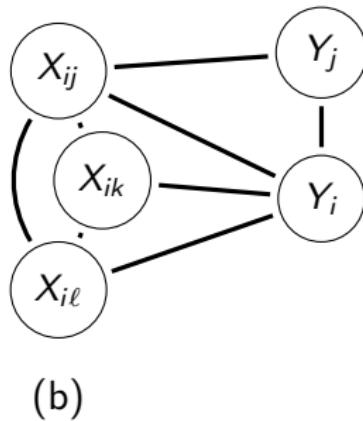


Adding dependence between outcomes

Network Block Attribute Block



(a)



(b)

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

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

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.



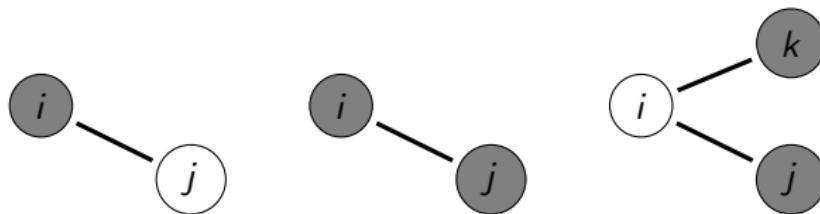
Auto-Logistic Actor Attribute Model (ALAAM)

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

ALAAM pmf

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

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

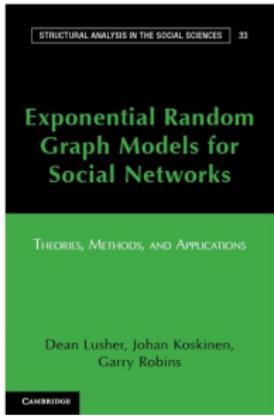


$\sum_i y_i x_{i+}$ (degree) $\sum_{i,j} y_i y_j x_{ij}$ (contagion) $\sum_{i,j} y_i y_k x_{ij} x_{jk}$ (indirect contagion)



... not new

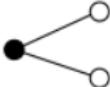
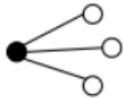
In book



And Robins, Elliot, Pattison (2001); Daraganova (2009)

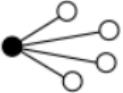
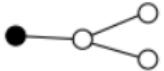
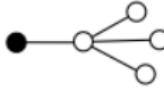
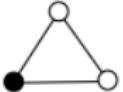


CUP ERGM book - available in MPNet (1)

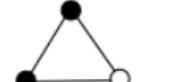
Configuration	Statistic	Parameter
	$\sum_i y_i$	attribute-density
	$\sum_i y_i \sum_j x_{ij}$	actor activity
	$\sum_i y_i \sum_{j,k} x_{ij} x_{ik}$	actor 2-star
	$\sum_i y_i \sum_{j,k,l} x_{ij} x_{ik} x_{il}$	actor 3-star



CUP ERGM book - available in MPNet (2)

	\dots	$\sum y_i \binom{x_{i+}}{k}$	actor k -star
		$\sum_i y_i \sum_{j,k} x_{ij} x_{jk}$	partner activity actor 2-path
		$\sum_i y_i \sum_j x_{ij} \binom{x_{j+}}{2}$	partner 2-star
	\dots	$\sum y_i \sum_j x_{ij} \binom{x_{j+}}{m}$	partner m -star
		$\sum_i y_i \sum_{j,k} x_{ij} x_{ik} x_{jk}$	actor triangle

CUP ERGM book - available in MPNet (3)

Configuration	Statistic	Parameter
	$\sum_{i < j} y_i y_j x_{ij}$	partner attribute
	$\sum_{i < k} y_i y_k \sum_{j \neq i, k} x_{ij} x_{jk}$	indirect partner attribute
	$\sum_{i,j} y_i y_j x_{ij} x_{j+}$	partner attribute activity
	$\sum_{i,j,k} y_i y_j y_k x_{ij} x_{jk}$	partner-partner attribute
	$\sum_{i,j,k} y_i y_j x_{ij} x_{ik} x_{jk}$	partner attribute triangle
	$\sum_{i,j,k} y_i y_j y_k x_{ij} x_{ik} x_{jk}$	partner-partner attribute triangle

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

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

$\underbrace{2^n \text{ terms}}$

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

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

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



How did we get estimates - Estimation

Simulating from likelihood

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

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

giving us samples from

$$\mathbf{y} | \mathbf{X}, \theta$$

We will use this for

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

MPNet uses samples in stochastic approximation for MLE



MCMC for un-normalized distributions

MCMC: Sample $\theta^{(0)}, \theta^{(1)}, \dots$ 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 distribution $\pi(\theta|z)$

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

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

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

so that we use Monte Carlo estimators

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

as well as approximate probabilities $\Pr(\theta \in C)$



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

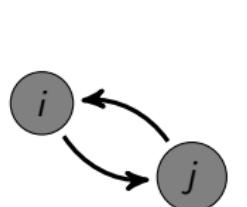
In our implementation the proposal distribution in each iteration

$$\theta^* | \theta^{(t)} \sim N(\theta^{(t)}, \Sigma_p)$$

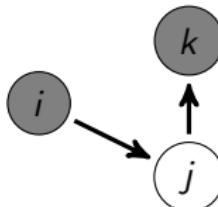
SACF can be lowered and mixing improved through improved Σ_p .

More elaborate effects

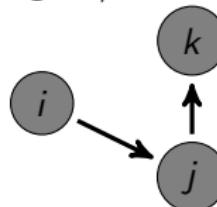
A number of more elaborate forms of contagion/influence are admissible



reciprocal



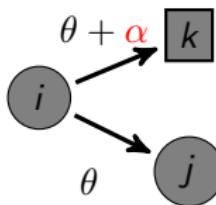
indirect



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



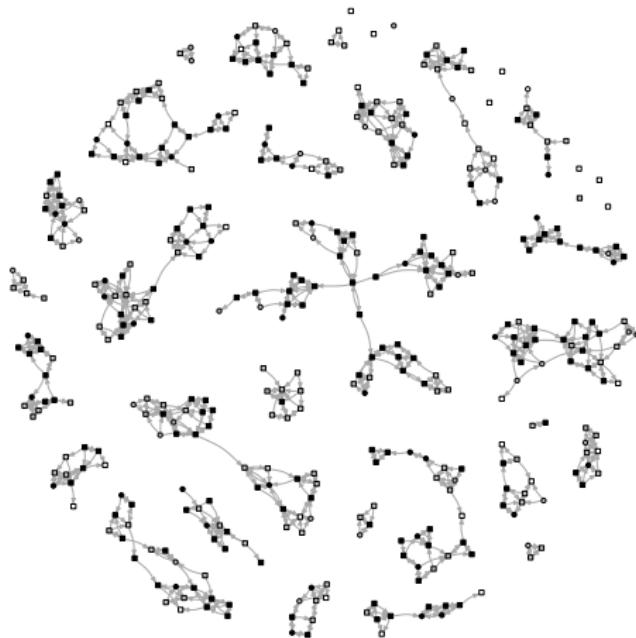


Figure: Bffs in 4 schools. Squares (girl) and circles (boys), and outcome black (y_i), grey ($y_i = 0$), and white for missing.

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**



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

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



Other issues

Model selection: technical issue

BMI: basic marginal likelihood identity

Evidence (= probability of data marginalising with respect to parameters)

$$m(\mathbf{y}) = \frac{p_{\theta}(\mathbf{y}|\mathbf{X})\pi(\theta)}{\pi(\theta|\mathbf{y})}$$

requires that we evaluate p and π in a point $\theta^* \in \Theta$.

(technically possible but computationally costly)



Model selection: philosophical issue

BMI: basic marginal likelihood identity

Evidence (= probability of data marginalising with respect to parameters)

$$m(\mathbf{y}) = \frac{p_{\theta}(\mathbf{y}|\mathbf{X})\pi(\theta)}{\pi(\theta|\mathbf{y})}$$

requires that we have a **subjective** (and interpretable) **prior** $\pi(\theta)$.
(... if your mother told you this was easy she lied to you)



Goodness-of-fit (GOF)

Once we have a draw

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

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

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

each from

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

GOF evaluation

If

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

are 'similar' to y , then model has good fit