Analysing populations of networks with mixtures of generalized linear mixed models

Mirko Signorelli[1]

: mirkosignorelli.github.io
: m.signorelli@math.leidenuniv.nl
: @signormirko

Joint work with Ernst Wit[2]

[1]Mathematical Institute, Leiden University (NL)
[2]Institute of Computational Science, University of Lugano (SW)

COMPSTAT2022
August 26, 2022

1987: populations of networks are born

Social Networks 9 (1987) 109–134
North-Holland

109

COGNITIVE SOCIAL STRUCTURES

David KRACKHARDT *

Cornell University

There are problems within the area of network analysis that can be fruitfully explored with cognitive social structures (CSS). Such structures can be modeled as three-dimensional ($N \times N \times N$) network structures. A definition of such structures is presented, along with a review of some of the problems CSS might address. Three types of aggregations of CSS – Slices, Locally Aggregated Structures (LAS), and Consensus Structures (CS) – are proposed to reduce CSS to a tractable two dimensions for analysis. As an illustration, the CSS of a management team of a small manufacturing firm is analyzed comparing all three types of aggregations.

- In modern terms: CSS = population of networks (PoN)

From networks to populations of networks (PoN)

- Network science = a framework to study relational data
- Network represented with a graph $\mathcal{G} = (V, E)$
  1. V set of vertices / nodes (subjects / objects)
  2. E set of edges (relationships)

From networks to populations of networks (PoN)

- Network science = a framework to study relational data
- Network represented with a graph $\mathcal{G} = (V, E)$
  1. V set of vertices / nodes (subjects / objects)
  2. E set of edges (relationships)
- Often, multiple instances $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \ldots$ of the "same" network are observed:
  1. consecutive snapshots of same network $\Rightarrow$ dynamic network
  2. different types of relationships between same actors $\Rightarrow$ multilayer network
  3. independent realizations of a network $\Rightarrow$ population of networks (PoN)

Motivation

- PoN increasingly common (Krackhardt (1987); Sweet et al. (2014); Taya et al. (2016); Reyes and Rodriguez (2016),...)
- Statistical modelling of networks traditionally focused on models for a single graph
- PoN with many graphs → analysing each network separately cumbersome / unfeasible

Our research question:

how to **model a population of networks in a thrifty way**?

Our proposal: model-based clustering for PoN

Statistical Modelling 2020; **20(1)**: 9–29

Model-based clustering for populations of networks

Mirko Signorelli[1] and Ernst C. Wit[2,3]
[1]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.
[2]Institute of Computational Science, Università della Svizzera italiana, Lugano, Switzerland.
[3]Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands.

Abstract: Until recently obtaining data on populations of networks was typically rare. However, with the advancement of automatic monitoring devices and the growing social and scientific interest in networks, such data has become more widely available. From sociological experiments involving cognitive social structures to fMRI scans revealing large-scale brain networks of groups of patients, there is a growing awareness that we urgently need tools to analyse populations of networks and particularly to model the variation between networks due to covariates. We propose a model-based clustering method based on mixtures of generalized linear (mixed) models that can be employed to describe the joint distribution of a populations of networks in a parsimonious manner and to identify subpopulations of networks that share certain topological properties of interest (degree distribution, community structure, effect of covariates on the presence of an edge, etc.). Maximum likelihood estimation for the proposed model can be efficiently carried out with an implementation of the EM algorithm. We assess the performance of this method on simulated data and conclude with an example application on advice networks in a small business.

Key words: cognitive social structure, EM algorithm, graph, mixture of generalized linear models, model-based clustering, network modelling, population of networks

Model-based clustering for PoN (1)

Let $\mathcal{S} = \{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_K\}$ be a PoN and Y_k the adjacency matrix of $\mathcal{G}_k$

Naive approach

Assume that each $\mathcal{G}_k$ has its own generative model:

1. model each $Y_k \sim f(Y_k|\theta_k)$ separately
2. interpret and compare K different models

Limitations:

- cumbersome if $K > 5$
- unfeasible if $K > 20$

Model-based clustering for PoN (2)

Model-based clustering approach (Signorelli and Wit, 2020)

We model jointly all graphs in $\mathcal{S} = \{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_K\}$:

1. assume there are $M < K$ subpopulations $\mathcal{S}_1, ..., \mathcal{S}_M$ of networks, each with generative component $f(Y|\theta_m)$
2. model each $f(Y|\theta_m)$ with a GLM / GLMM
3. unknown subpopulation membership ($\mathcal{G}_k \in \mathcal{S}_m$)

Advantages:

- parsimony: $M << K$ models
- quantify similarity between graphs $[P(\mathcal{G}_k \in \mathcal{S}_m)]$
- find clusters of similar graphs

A mixture of network models

1. Each graph is drawn from a mixture model

$$Y_k \sim \sum_{m=1}^{M} \pi_m f(Y|\theta_m)$$

A mixture of network models

1. Each graph is drawn from a mixture model

$$Y_k \sim \sum_{m=1}^{M} \pi_m f(Y|\theta_m)$$

2. f? → any network model that can be specified as GLM/GLMM
   - Examples: p_1, p_2, SBM / dcSBM a priori, saturated network model, node/edge-specific covariates, ...

A mixture of network models

1. Each graph is drawn from a mixture model

$$Y_k \sim \sum_{m=1}^{M} \pi_m f(Y|\theta_m)$$

2. f? → any network model that can be specified as GLM/GLMM
   - Examples: p_1, p_2, SBM / dcSBM a priori, saturated network model, node/edge-specific covariates, ...

3. Latent $Z_k = m$ if $\mathcal{G}_k \in \mathcal{S}_m$, $m \in \{1, \dots, M\}$

4. Likelihood depends on observed $\mathbf{Y}$ and unobserved Z:

$$L(\mathbf{Y}, Z|\Theta) = \prod_{k=1}^{K} \pi_{Z_k} f\left(Y_k|\theta_{Z_k}\right)$$

$\Rightarrow$ EM algorithm

Model estimation

EM algorithm implementation

1. Choose $M \in \{1, 2, 3, \dots\}$
2. Initialize $p^0_{km} = P(Z_k = m)$; gather probs in matrix P^0
3. Estimate $f(Y|\theta_m)$ through a mixture of GLMM with weights $\left(p^0_{1m}, \dots, p^0_{Km}\right)$ for component $f(Y|\theta_m) \Rightarrow$ obtain $\hat{\Theta}^0 = \left(\hat{\theta}^0_1, \dots, \hat{\theta}^0_M\right)$
4. For $t = 1, 2, 3, \dots$ (until convergence reached):
   - **E step**: given $\hat{\Theta}^{t-1}$, update P^t as

$$p^t_{km} = \frac{P(\mathcal{G}_k|\hat{\theta}^{t-1}_m)}{\sum_{j=1}^M P(\mathcal{G}_k|\hat{\theta}^{t-1}_j)}$$

   - **M step**: estimate a mixture of GLMMs with weights $\left(p^t_{1m}, \dots, p^t_{Km}\right)$ for component $f(Y|\theta_m)$ and derive $\hat{\Theta}^t$

1. Motivation

2. Methods

3. Simulations

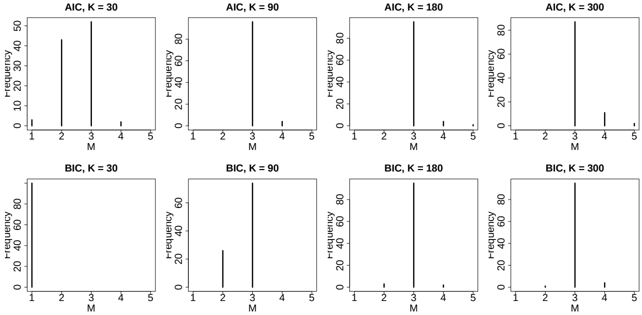
4. Application

5. Conclusion

Model selection

With real data, M typically unknown

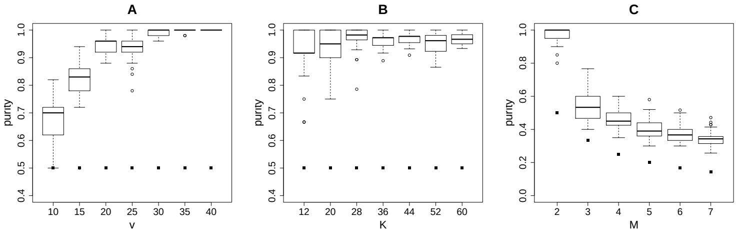
- choice of optimal M through model selection criteria
- comparison between AIC and BIC in MC simulations (true $M = 3$):



- Conclusion: AIC outperforms BIC when K small

Accuracy

- Simulations A-C with p_1 model:



- Purity of estimated cluster membership
  1. increases rapidly with v
  2. doesn't depend much on K
  3. decreases with M
- More simulations (D-I) in Signorelli and Wit (2020)

1. Motivation
2. Methods
3. Simulations
4. Application
5. Conclusion

Krackhardt's data

- Back to Krackhardt (1987):

Social Networks 9 (1987) 109–134
North-Holland 109

COGNITIVE SOCIAL STRUCTURES

David KRACKHARDT *

Cornell University

There are problems within the area of network analysis that can be fruitfully explored with cognitive social structures (CSS). Such structures can be modeled as three-dimensional ($N \times N \times N$) network structures. A definition of such structures is presented, along with a review of some of the problems CSS might address. Three types of aggregations of CSS – Slices, Locally Aggregated Structures (LAS), and Consensus Structures (CS) – are proposed to reduce CSS to a tractable two dimensions for analysis. As an illustration, the CSS of a management team of a small manufacturing firm is analyzed comparing all three types of aggregations.

Krackhardt's data

Krackhardt (1987) collected data about advice relationships between 21 managers of a US high-tech company:

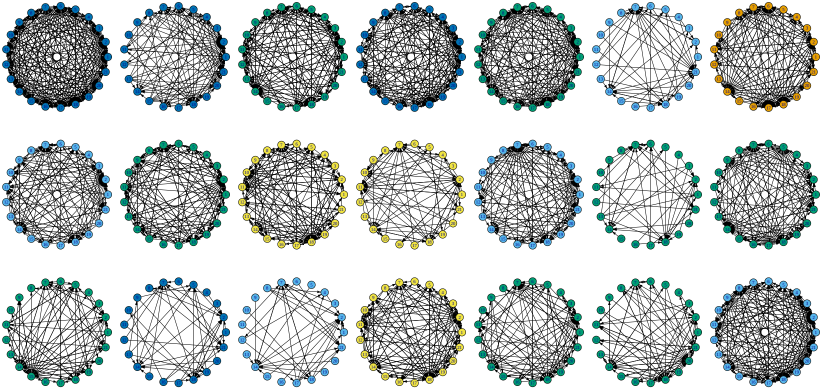
- $K = 21$ directed networks (no self-loops)

$$y_{ij}^{k} = 1 \text{ if } k \text{ thinks that } i \text{ goes to } j \text{ for advice}$$

$$i = \text{sender},\ j = \text{receiver},\ k = \text{perceiver}$$

- covariates:
  1. age and length of service (tenure) of each employee
  2. role in the firm (CEO + 2 directors + 18 managers)
  3. department (1,2,3,4) where the employee works

The (21) networks, actually:



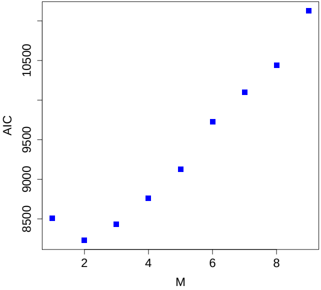
- Colors denote different departments. Orange = CEO
- Density $\in [0.1, 0.66]$, mean $= 0.32$

Selection of optimal M

Starting point:

- a mixture of "saturated" network models where $y_{ij}^k | z_k \sim Bern(\pi_{ij}^{z_k})$

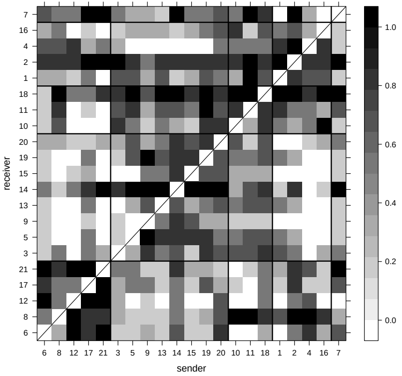
Model selection based on AIC minimization:



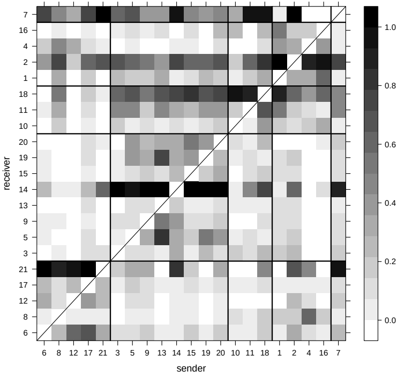
$$\Rightarrow \hat{M} = 2$$

Saturated model: parameter estimates

$\hat{\mathcal{S}}_1 = \{\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_4, \mathcal{G}_5, \mathcal{G}_{10}, \mathcal{G}_{21}\}$



$\hat{\mathcal{S}}_2 = \mathcal{S} \setminus \hat{\mathcal{S}}_1$



Saturated model: pros and cons

Mixture of saturated network models

- useful for:
  1. selection of M through a very flexible network model
  2. quick visual comparison of the subpopulations
- limitations:
  1. $21 \cdot 20 \cdot 2 = 840$ parameters to look at!
  2. no information on effect of covariates on π_{ij}^k

Which factors influence advice relationships?

Effect of covariates on π_{ij}^k:

1. What affects the choice to ask advice from a colleague?
   - Age and seniority?
   - Role within company?
   - Department affiliation?
2. Do perceivers report more advice relationships involving themselves?

Refining the model

$$y_{ij}^{k} \,|\, \left(z_k, u_i^{z_k}, v_i^{z_k}\right) \sim Bern(\pi_{ij}^{z_k})$$

$$\begin{aligned}\text{logit}(\pi_{ij}^{z_k}) = \beta_0^{z_k} &+ u_i^{z_k} + v_j^{z_k} + \beta_1^{z_k} A_i + \beta_2^{z_k} T_i + \beta_3^{z_k} L_i + \beta_4^{z_k} I(i = k) \\ &+ \beta_5^{z_k} A_j + \beta_6^{z_k} T_j + \beta_7^{z_k} L_j + \beta_8^{z_k} I(j = k) \\ &+ \sum_{r=1}^{4} \gamma_r^{z_k} I[D_i = r] + \sum_{s=1}^{4} \delta_s^{z_k} I[D_j = s] + \sum_{r=1}^{4}\sum_{s=1}^{4} \xi_{rs}^{z_k} I[D_i = r] I[D_j = s],\end{aligned}$$

where

- $u_i^{z_k} \sim N\left[0, (\sigma^{z_k})^2\right]$ and $v_j^{z_k} \sim N\left[0, (\tau^{z_k})^2\right]$ model the in- and out-degree distributions
- A_i = age, T_i = tenure, L_i = leading role of employee asking (i) / receiving (j) advice
- γ_r^m, δ_s^m and ξ_{rs}^m are blockmodel main effects and interactions[1] for the departments (D_i)

[1]constraints: $\sum_{r=1}^{4} \gamma_r^m = 0$, $\sum_{s=1}^{4} \delta_s^m = 0$ and $\sum_{r=1}^{4}\sum_{s=1}^{4} \xi_{rs}^m = 0$ for every $m \in \{1, 2\}$

Effect of department affiliation

Subpopulation 1

Dept. sender	Dept. receiver 1	2	3	4
1	$\oplus$	$\ominus$	$\ominus$	$-$
2	$\ominus$	$\oplus$	$+$	$\ominus$
3	$\ominus$	$\oplus$	$\oplus$	$+$
4	$-$	$\ominus$	$-$	$\oplus$

Subpopulation 2

Dept. sender	Dept. receiver 1	2	3	4
1	$\oplus$	$\ominus$	$\ominus$	$-$
2	$\ominus$	$\oplus$	$+$	$\ominus$
3	$\ominus$	$+$	$\oplus$	$\ominus$
4	$-$	$\ominus$	$-$	$\oplus$

$\oplus$, $\ominus$: significant effects ($p < 0.05$)

- Strong community structure induced by department in both $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$

Node-specific covariates

Parameter	$\hat{\theta}^1$	$\hat{\theta}^2$	$SE(\hat{\theta}^1)$	$SE(\hat{\theta}^2)$	p-value ($\theta^1 = \theta^2$)
β_0	0.809	-1.997*	0.671	0.439	0.000
β_1 (age sender)	-0.014	-0.006	0.012	0.010	0.972
β_2 (tenure sender)	-0.035*	-0.016	0.017	0.009	0.930
β_3 (sender in lead pos.)	0.014	0.008	0.016	0.013	0.977
β_4 (perceiver = sender)	1.128*	1.020*	0.231	0.146	0.675
β_5 (age receiver)	0.034	0.044*	0.022	0.012	0.964
β_6 (tenure receiver)	0.543*	0.582*	0.214	0.170	0.876
β_7 (receiver in lead pos.)	1.407*	2.058*	0.287	0.150	0.017
β_8 (perceiver = receiver)	1.353*	1.354*	0.231	0.149	0.998

$^* = p < 0.05$ ($H_0 : \beta_j^m = 0$)

Similarities between $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$:

- perceivers report more in- and out- advice involving themselves ($\hat{\beta}_4 > 0$ and $\hat{\beta}_8 > 0$)
- advice sought from employees with longer tenure ($\hat{\beta}_6 > 0$) and from CEO and directors ($\hat{\beta}_7 > 0$)

Node-specific covariates

Parameter	$\hat{\theta}^1$	$\hat{\theta}^2$	$SE(\hat{\theta}^1)$	$SE(\hat{\theta}^2)$	p-value ($\theta^1 = \theta^2$)
β_0	0.809	-1.997*	0.671	0.439	0.000
β_1 (age sender)	-0.014	-0.006	0.012	0.010	0.972
β_2 (tenure sender)	-0.035*	-0.016	0.017	0.009	0.930
β_3 (sender in lead pos.)	0.014	0.008	0.016	0.013	0.977
β_4 (perceiver = sender)	1.128*	1.020*	0.231	0.146	0.675
β_5 (age receiver)	0.034	0.044*	0.022	0.012	0.964
β_6 (tenure receiver)	0.543*	0.582*	0.214	0.170	0.876
β_7 (receiver in lead pos.)	1.407*	2.058*	0.287	0.150	0.017
β_8 (perceiver = receiver)	1.353*	1.354*	0.231	0.149	0.998

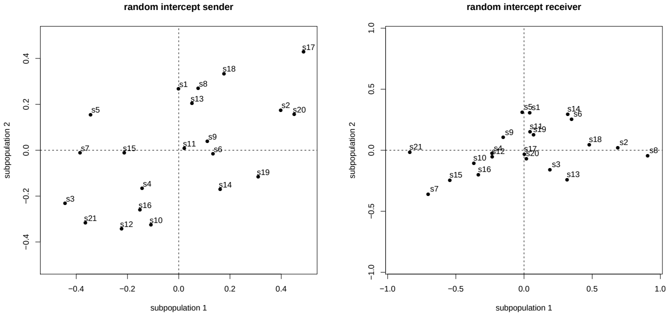
$^* = p < 0.05$ ($H_0 : \beta_j^m = 0$)

Differences between $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$:

- graphs denser in $\hat{\mathcal{S}}_1$
- tendency to seek advice from CEO and directors stronger in $\hat{\mathcal{S}}_2$

Predicted random effects

$\hat{u}_i$ and $\hat{v}_j$ in $\hat{\mathcal{S}}_1$ (x axis) and $\hat{\mathcal{S}}_1$ (y):



- Perceivers in $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$ have similar ideas about
  - who asks advice from more colleagues
  - who is more frequently consulted for advice
- $\hat{\sigma}^1 \approx \hat{\sigma}^2$, but $\hat{\tau}^1 >> \hat{\tau}^2$ (out-degrees more heterogeneous in $\hat{\mathcal{S}}_1$)

Take-home message

- Proposal: model-based clustering of PoN (Signorelli and Wit, 2020)
- Advantages:
  1. parsimonious: a single mixture model for the whole PoN
  2. efficient ML estimation through EM algorithm
  3. flexible: both binary and edge-valued graphs + easy to combine multiple types of network models ($p_1/p2$, SBM, node-specific covariates, ...)
  4. by-product: clusters of "similar" graphs
- Limitation: cannot deal with ERGMs (Frank and Strauss, 1986)

References

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Krackhardt, D. (1987). Cognitive social structures. *Social networks*, 9(2):109–134.

Reyes, P. and Rodriguez, A. (2016). Stochastic blockmodels for exchangeable collections of networks. *arXiv preprint arXiv:1606.05277*.

Signorelli, M. and Wit, E. C. (2020). Model-based clustering for populations of networks. *Statistical Modelling*, 20(1):9–29.

Sweet, T. M., Thomas, A. C., and Junker, B. W. (2014). Hierarchical mixed membership stochastic blockmodels for multiple networks and experimental interventions. *Handbook on mixed membership models and their applications*, pages 463–488.

Taya, F., de Souza, J., Thakor, N. V., and Bezerianos, A. (2016). Comparison method for community detection on brain networks from neuroimaging data. *Applied Network Science*, 1(1):8.