Autoregressive Networks with Stylized Features

Qiwei Yao

Department of Statistics

London School of Economics and Political Science

q.yao@lse.ac.uk

Outline

1. Autoregressive dynamic network models: a basic framework
   - Basic properties: stationarity, α-mixing
   - Maximum likelihood estimation
   - Model diagnostics: A permutation test
   - Illutrations: stochastic block models, a change-point
2. Two-way heterogeneity dynamic network models
   - AR(1) β-models: Node-degree heterogeneity
   - Local convexity of log-likkelihood
   - A new MME
3. AR networks with dependent edges
   - Models for transitivity, density-dependence & persistence
   - Relationship to temporal ERGM
   - Stationarity, technical challenges
   - Email interaction: a real data example

Background

Dynamic networks: a large body literature

Evolution analysis of network snapshots: Aggarwal and Subbian (2014), Donnat and Holmes (2018)

Networks at different times are assumed to be conditionally independent (on some latent processes), or independent (Pensky 2019)

Exponential family conditional distributions (Krivitsky and Handcock 2014)

Inference relying on Bayesian/computational methods (Durante et al 2016, Matias & Miele 2017)

Asymptotic theory for independent network data (Bhattacharjee et al 2020, Enikeeva & Klopp 2021)

Our goal: Model dynamic changes, with stylized features, explicitly in a simple manner

Autoregressive Network Models: a basic framework

Joint work with

Binyan Jiang, HKPU

Jialiang Li, NUS

Jiang, B., Li, J. and Yao, Q. (2023). Autoregressive networks. *Journal of Machine Learning Research*, **24** (227), 1-69.

Let $\mathbf{X}_t \equiv (X_{i,j}^t)$ be $p \times p$ adjacency matrix of a network on p nodes $\{1, \cdots, p\}$ at time t, $X_{i,j}^t = 0$ or 1 only.

Assumption: p nodes unchanged over time, edges are indep with each other)

Definition. For $t \geq 1$,

$$X_{i,j}^t = X_{i,j}^{t-1} I(\varepsilon_{i,j}^t = 0) + I(\varepsilon_{i,j}^t = 1), \quad (i,j) \in \mathcal{J},$$

where innovations $\varepsilon_{i,j}^t$, $(i,j) \in \mathcal{J}$ and $t \geq 1$, are independent, and

$$P(\varepsilon_{i,j}^t = 1) = \alpha_{i,j}^t, \;\; P(\varepsilon_{i,j}^t = -1) = \beta_{i,j}^t, \;\; P(\varepsilon_{i,j}^t = 0) = 1 - \alpha_{i,j}^t - \beta_{i,j}^t.$$

For undirected networks w/o selfloops,

$$\mathcal{J} = \{(i,j) : 1 \leq i < j \leq p\}, \quad X_{j,i}^t \equiv X_{i,j}^t, \quad X_{i,i}^t \equiv 0$$

AR(1) networks: a Markov chain

Autoregressive version of 'regression model' of Chang, Kolaczyk & Yao (2022)

$$P(X_{i,j}^t = 1 | X_{i,j}^{t-1} = 0) = \alpha_{i,j}^t, \quad P(X_{i,j}^t = 0 | X_{i,j}^{t-1} = 1) = \beta_{i,j}^t,$$

$\mathbf{X}_t$ is a Markov chain:

$$P(\mathbf{X}_t | \mathbf{X}_{t-1}, \cdots, \mathbf{X}_0) = P(\mathbf{X}_t | \mathbf{X}_{t-1}) = \prod_{(i,j) \in \mathcal{J}} P(X_{i,j}^t | X_{i,j}^{t-1})$$

$$= \prod_{(i,j) \in \mathcal{J}} (\alpha_{i,j}^t)^{X_{i,j}^t (1 - X_{i,j}^{t-1})} (1 - \alpha_{i,j}^t)^{(1 - X_{i,j}^t)(1 - X_{i,j}^{t-1})} (\beta_{i,j}^t)^{(1 - X_{i,j}^t) X_{i,j}^{t-1}} (1 - \beta_{i,j}^t)^{X_{i,j}^t X_{i,j}^{t-1}}.$$

AR(1) networks: stationarity

$\{\mathbf{X}_t, t = 0, 1, \cdots\}$ is strictly stationary if

1. $\alpha_{i,j}^t \equiv \alpha_{i,j}$ and $\beta_{i,j}^t \equiv \beta_{i,j}$ *for all* $(i,j) \in \mathcal{J}$, *and*
2. $P(X_{i,j}^0 = 1) = \pi_{i,j} = 1 - P(X_{i,j}^0 = 0)$, *and* $\pi_{i,j} = \frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}$.

Then

$$E(X_{i,j}^t) = \frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}, \qquad \mathrm{Var}(X_{i,j}^t) = \frac{\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2},$$

$$\rho_{i,j}(|t-s|) = \mathrm{Corr}(X_{i,j}^t, X_{i,j}^s) = (1 - \alpha_{i,j} - \beta_{i,j})^{|t-s|}.$$

Yule-Walker equation: $\rho_{i,j}(k) = (1 - \alpha_{i,j} - \beta_{i,j})\rho_{i,j}(k-1)$.

Note. Recall model $X_{i,j}^t = I(\varepsilon_{i,j}^t = 0)X_{i,j}^{t-1} + I(\varepsilon_{i,j}^t = 1)$, the Y-W equation is

$$\rho_{i,j}(t) = E\{I(\varepsilon_{i,j}^t = 0)\}\rho_{i,j}(t-1).$$

AR(1) networks: α-mixing

Let $\mathcal{F}_a^b = \sigma(X_{i,j}^k, a \le k \le b)$, and

$$\alpha^{i,j}(\tau) = \sup_{k \in \mathbb{N}} \sup_{A \in \mathcal{F}_0^k, B \in \mathcal{F}_{k+\tau}^\infty,} |P(A \cap B) - P(A)P(B)|.$$

Then under the stationarity,

$$\alpha^{i,j}(\tau) = \rho_{i,j}(\tau) = (1 - \alpha_{i,j} - \beta_{i,j})^\tau, \quad \tau \ge 1.$$

Hamming distance: $D_H(\mathbf{A}, \mathbf{B}) = \sum_{i,j} I(A_{i,j} \neq B_{i,j})$ for any two matrices $\mathbf{A} = (A_{i,j})$, $\mathbf{B} = (B_{i,j})$ of the same size.
Let $d_H(|t-s|) = E\{D_H(\mathbf{X}_t, \mathbf{X}_s)\}$, then

$$
\begin{aligned}
d_H(k) &= d_H(k-1) + \sum_{(i,j)\in\mathcal{J}} \frac{2\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} (1-\alpha_{i,j}-\beta_{i,j})^{k-1} \\
&= \sum_{(i,j)\in\mathcal{J}} \frac{2\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2} \{1-(1-\alpha_{i,j}-\beta_{i,j})^k\}.
\end{aligned}
$$

Thus $d_H(d)$ increases strictly, as k increases, initially from $d_H(0) = 0$ towards the limit $d_H(\infty) = \sum \frac{2\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2}$ which is the expected Hamming distance of the two independent networks sharing the same marginal distribution of $\mathbf{X}_t$.

Processes with alternating ACF

Define

$$X_{i,j}^t = (1 - X_{i,j}^{t-1}) I(\varepsilon_{i,j}^t = 0) + I(\varepsilon_{i,j}^t = 1).$$

Then for $k = 0, 1, 2, \cdots,$

$$\mathrm{Corr}(X_{i,j}^t, X_{i,j}^{t+k}) = (-1)^k (1 - \alpha_{i,j} - \beta_{i,j})^k,$$

and

$$E\{D_H(\mathbf{X}_t, \mathbf{X}_{t+k})\} = \sum_{(i,j)\in\mathcal{J}} \frac{2(1-\alpha_{i,j})(1-\beta_{i,j})}{(2-\alpha_{i,j}-\beta_{i,j})^2} \{1 - (-1)^k (1-\alpha_{i,j}-\beta_{i,j})^k\}.$$

Maximum likelihood estimation

Since $X_{i,j}^t$, for different $(i,j) \in \mathcal{J}$, are independent, $(\alpha_{i,j}, \beta_{i,j})$, for different (i,j), can be estimated separately.

Observations: $\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_n$.

Log-likelihood (conditional on $\mathbf{X}_0$):

$$l(\alpha_{i,j}, \beta_{i,j}) = \log(\alpha_{i,j}) \sum_{t=1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1}) + \log(1 - \alpha_{i,j}) \sum_{t=1}^{n} (1 - X_{i,j}^t)(1 - X_{i,j}^{t-1})$$

$$+ \log(\beta_{i,j}) \sum_{t=1}^{n} (1 - X_{i,j}^t) X_{i,j}^{t-1} + \log(1 - \beta_{i,j}) \sum_{t=1}^{n} X_{i,j}^t X_{i,j}^{t-1}.$$

MLEs:

$$\widehat{\alpha}_{i,j} = \frac{\sum_{t=1}^n X_{i,j}^t (1 - X_{i,j}^{t-1})}{\sum_{t=1}^n (1 - X_{i,j}^{t-1})}, \qquad \widehat{\beta}_{i,j} = \frac{\sum_{t=1}^n (1 - X_{i,j}^t) X_{i,j}^{t-1}}{\sum_{t=1}^n X_{i,j}^{t-1}}.$$

Asymptotic properties of MLE

Under conditions C1 and C2,

$$\max_{(i,j)\in\mathcal{J}} |\widehat{\alpha}_{i,j} - \alpha_{i,j}| = O_p\left(\sqrt{\frac{\log p}{n}}\right), \quad \max_{(i,j)\in\mathcal{J}} |\widehat{\beta}_{i,j} - \beta_{i,j}| = O_p\left(\sqrt{\frac{\log p}{n}}\right),$$

$$\sqrt{n}\{\begin{pmatrix} \widehat{\alpha}_{i,j} \\ \widehat{\beta}_{i,j} \end{pmatrix} - \begin{pmatrix} \alpha_{i,j} \\ \beta_{i,j} \end{pmatrix}\} \xrightarrow{D} N\big(0, \mathrm{diag}(\sigma_{i,j}, \sigma^\star_{i,j})\big),$$

where

$$\sigma_{i,j} = \frac{\alpha_{i,j}(1-\alpha_{i,j})(\alpha_{i,j}+\beta_{i,j})}{\beta_{i,j}}, \qquad \sigma^\star_{i,j} = \frac{\beta_{i,j}(1-\beta_{i,j})(\alpha_{i,j}+\beta_{i,j})}{\alpha_{i,j}}.$$

C1. There exists a constant l such that $0 < l \le \alpha_{i,j}, \beta_{i,j}, \alpha_{i,j} + \beta_{i,j} \le 1$ holds for all $(i,j) \in \mathcal{J}$.

C2. $n, p \to \infty$, and $(\log n)(\log\log n)\sqrt{\frac{\log p}{n}} \to 0$.

Model diagnostics – A permutation test

'Residual' $\widehat{\varepsilon}_{i,j}^t$ is defined as the estimated value of $E(\varepsilon_{i,j}^t | X_{i,j}^t, X_{i,j}^{t-1})$:

$$\widehat{\varepsilon}_{i,j}^t = \frac{\widehat{\alpha}_{i,j}}{1-\widehat{\beta}_{i,j}} I(X_{i,j}^t = 1, X_{i,j}^{t-1} = 1) - \frac{\widehat{\beta}_{i,j}}{1-\widehat{\alpha}_{i,j}} I(X_{i,j}^t = 0, X_{i,j}^{t-1} = 0)$$
$$+ I(X_{i,j}^t = 1, X_{i,j}^{t-1} = 0) - I(X_{i,j}^t = 0, X_{i,j}^{t-1} = 1)$$

for $(i,j) \in \mathcal{J},\ t = 1, \cdots, n$.
To check the adequacy of the model: to test for the independence of $\widehat{\boldsymbol{\mathcal{E}}}_t \equiv (\widehat{\varepsilon}_{i,j}^t)$ for $t = 1, \cdots, n$
Since $\widehat{\varepsilon}_{i,j}^t$, $t = 1, \cdots, n$, only take 4 different values for each $(i,j) \in \mathcal{J}$, we adopt two-way, or three-way contingency table to test the independence of $\widehat{\boldsymbol{\mathcal{E}}}_t$ and $\widehat{\boldsymbol{\mathcal{E}}}_{t-1}$, or $\widehat{\boldsymbol{\mathcal{E}}}_t, \widehat{\boldsymbol{\mathcal{E}}}_{t-1}$ and $\widehat{\boldsymbol{\mathcal{E}}}_{t-2}$.

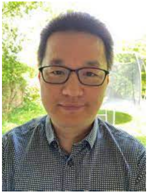
Outlook

- AR(p), or even ARMA?
- Incorporating heterogeneity, sparsity, transitivity, homophily and other stylized features, and dealing with networks with dependent edges.
- Networks with weighted edges: matrix time series models (Tensor decomposition): Wang, Liu and Chen (2019), Chang, He, Yang and QY (2023).

Two-way heterogeneity dynamic network models (TWHM)

Joint work with

Binyan Jiang
HKPU



Chenlei Leng
Warwick

Ting Yan
CCNU

Xinyang Yu
HKPU

Jiang, Leng, Yan, Yao and Yu (2023). A two-way heterogeneity model for dynamic networks. `arXiv:2305.12643`.

TWHM: a parsimonious AR(1) & a dynamic β-model

Let $\quad X_{i,j}^t = I(\varepsilon_{i,j}^t = 0) + X_{i,j}^{t-1} I(\varepsilon_{i,j}^t = 1)$, where $\varepsilon_{i,j}^t$, $(i,j) \in \mathcal{J}$ and $t \geq 1$, are independent, and

$$P(\varepsilon_{i,j}^t = 1) = \frac{e^{\beta_{i,1}+\beta_{j,1}}}{1 + \sum_{k=0}^{1} e^{\beta_{i,k}+\beta_{j,k}}}, \quad P(\varepsilon_{i,j}^t = 0) = \frac{e^{\beta_{i,0}+\beta_{j,0}}}{1 + \sum_{k=0}^{1} e^{\beta_{i,k}+\beta_{j,k}}},$$

$$P(\varepsilon_{i,j}^t = -1) = \frac{1}{1 + \sum_{k=0}^{1} e^{\beta_{i,k}+\beta_{j,k}}}.$$

There are only $2p$ (instead of $2p^2$) parameters:

$$\boldsymbol{\beta}_0 = (\beta_{1,0}, \cdots, \beta_{p,0})^\top, \qquad \boldsymbol{\beta}_1 = (\beta_{1,1}, \cdots, \beta_{p,1})^\top.$$

Then

$$P(X_{i,j}^t = 1) = \frac{e^{\beta_{i,0}+\beta_{j,0}}}{1 + e^{\beta_{i,0}+\beta_{j,0}}} = 1 - P(X_{i,j}^t = 1)$$

i.e. the (static) β-model: the larger $\beta_{i,0}$ is, the larger node i's degree

We call $\boldsymbol{\beta}_0 = (\beta_{1,0}, \cdots, \beta_{p,0})^\top$ static heterogeneity parameters.

Furthermore,

$$P(X_{i,j}^t = 1 | X_{i,j}^{t-1} = 0) = \frac{e^{\beta_{i,0}+\beta_{j,0}}}{1 + \sum_{k=0}^{1} e^{\beta_{i,k}+\beta_{j,k}}},$$

$$P(X_{i,j}^t = 0 | X_{i,j}^{t-1} = 1) = \frac{1}{1 + \sum_{k=0}^{1} e^{\beta_{i,k}+\beta_{j,k}}},$$

i.e. the larger $\beta_{i,1}$ is, the more likely $X_{i,j}^t$ will retain the value of $X_{i,j}^{t-1}$ for all j.

We call $\boldsymbol{\beta}_1 = (\beta_{1,1}, \cdots, \beta_{p,1})^\top$ dynamic heterogeneity parameters.

AR(1) β-models for dynamic networks

$$E(X_{i,j}^t) = \frac{e^{\beta_{i,0}+\beta_{j,0}}}{1+e^{\beta_{i,0}+\beta_{j,0}}}, \qquad \mathrm{Var}(X_{i,j}^t) = \frac{e^{\beta_{i,0}+\beta_{j,0}}}{(1+e^{\beta_{i,0}+\beta_{j,0}})^2},$$

$$\rho_{i,j}(|t-s|) \equiv \mathrm{Corr}(X_{i,j}^t, X_{i,j}^s) = \left(\frac{e^{\beta_{i,1}+\beta_{j,1}}}{1+\sum_{r=0}^{1} e^{\beta_{i,r}+\beta_{j,r}}} \right)^{|t-s|}.$$

We allow $\beta_{i,0} \to -\infty$ as $p \to \infty$, to allow edge-sparsity.

Condition (A1) below ensures that $\rho_{i,j}(|t-s|)$ are bounded away from 1, hence observations contain the info of dynamic changes.

Then $\{X_{i,j}^t, t \geq 1\}$ is α-mixing with uniformly exponential decaying rates.

(A1) The true parameters satisfy $\beta_{i,1}^* - \max(\beta_{i,0}^*, 0) < K$ for any $i = 1, 2, \cdots, p$, where $K > 0$ is a constant.

Likelihood

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_0^\top, \boldsymbol{\beta}_1^\top)^\top$, and

$L(\boldsymbol{\theta}; \mathbf{X}^n, \mathbf{X}^{n-1}, \cdots, \mathbf{X}^1 | \mathbf{X}^0) = \prod_{t=1}^n L(\boldsymbol{\theta}; \mathbf{X}^t | \mathbf{X}^{t-1})$

Then

$$
\begin{aligned}
l(\boldsymbol{\theta}) &= -\frac{1}{np} \log L(\boldsymbol{\theta}; \mathbf{X}^n, \mathbf{X}^{n-1}, \cdots, \mathbf{X}^1 | \mathbf{X}^0) \\
&= -\frac{1}{p} \sum_{1 \le i < j \le p} \log \left(1 + e^{\beta_{i,0} + \beta_{j,0}} + e^{\beta_{i,1} + \beta_{j,1}}\right) + \frac{1}{np} \sum_{1 \le i < j \le p} \Bigg\{ (\beta_{i,0} + \beta_{j,0}) \sum_{t=1}^n X_{i,j}^t + \\
&\log \left(1 + e^{\beta_{i,1} + \beta_{j,1}}\right) \sum_{t=1}^n \left(1 - X_{i,j}^t\right) \left(1 - X_{i,j}^{t-1}\right) + \log \left(1 + e^{\beta_{i,1} + \beta_{j,1} - \beta_{i,0} - \beta_{j,0}}\right) \sum_{t=1}^n X_{i,j}^t X_{i,j}^{t-1} \Bigg\}.
\end{aligned}
$$

Unfortunately, MLE is no longer explicitly available, and $l(\boldsymbol{\theta})$ is not convex.

A roadmap to derive MLE for $\boldsymbol{\theta}$

1. Show $l(\boldsymbol{\theta})$ is (locally) convex in a neighbourhood of the true value $\boldsymbol{\theta}^*$
2. Let $\widehat{\boldsymbol{\theta}}$ be the local MLE in the neighbourhood specified above. Derive the bounds for $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$ in both ℓ_2 and ℓ_∞ norms
3. Construct a new MME which is asymptotically in the neighbourhood specified above
4. Setting this MME as the initial value, the local MLE is obtained via a gradient decent algorithm

A new MME for $\boldsymbol{\beta}_0$

Note $E(X_{i,j}^t) = \frac{e^{\beta_{i,0}+\beta_{j,0}}}{1+e^{\beta_{i,0}+\beta_{j,0}}}$, hence $\frac{1}{n}\sum_{t=1}^n X_{i,j}^t = \frac{e^{\beta_{i,0}+\beta_{j,0}}}{1+e^{\beta_{i,0}+\beta_{j,0}}}$.

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1,\, j\neq i}^{p} X_{i,j}^t - \sum_{j=1,\, j\neq i}^{p} \frac{e^{\beta_{i,0}+\beta_{j,0}}}{1+e^{\beta_{i,0}+\beta_{j,0}}} = 0, \quad i = 1, \cdots, p.$$

Those p equations are $\frac{\partial}{\partial \beta_{i,0}} f(\boldsymbol{\beta}_0) = 0$, $i = 1, \cdots, p$, where

$$f(\boldsymbol{\beta}_0) = \sum_{1\le i,j\le p} \log\{1 + e^{\beta_{i,0}+\beta_{j,0}}\} - n^{-1}\sum_{i=1}^{p}\{\beta_{i,0}\sum_{t=1}^{n}\sum_{j=1,\, j\neq i}^{p} X_{i,j}^t\}$$

Hence MME $\widetilde{\boldsymbol{\beta}}_0$ is the unique minimizer of convex function $f(\boldsymbol{\beta}_0)$

AR networks with dependent edges

Joint work with

			
Jinyuan Chang SWUFE, Chengdu	Qin Fang Sydney	Eric Kolaczyk McGill	Peter MacDonald McGill

Chang, Fang, Kolaczyk, MacDonald and Yao (2024). Autoregressive networks with dependent edges. arXiv:2404.15654

AR(m) networks with dependent edges

Consider undirected networks without selfloops ($X_{i,j}^t \equiv X_{j,i}^t$, $X_{i,i}^t \equiv 0$)

Conditionally on $\{\mathbf{X}_s\}_{s \le t-1}$, the edges $\{X_{i,j}^t\}_{1 \le i < j \le p}$ are mutually independent with

$$\begin{aligned}
\alpha_{i,j}^{t-1} &\equiv P\big(X_{i,j}^t = 1 \,|\, X_{i,j}^{t-1} = 0, \mathbf{X}_{t-1} \setminus X_{i,j}^{t-1}, \mathbf{X}_{t-k} \text{ for } k \ge 2\big) \\
&= P\big(X_{i,j}^t = 1 \,|\, X_{i,j}^{t-1} = 0, \mathbf{X}_{t-1} \setminus X_{i,j}^{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-m}\big)\,, \\
\beta_{i,j}^{t-1} &\equiv P\big(X_{i,j}^t = 0 \,|\, X_{i,j}^{t-1} = 1, \mathbf{X}_{t-1} \setminus X_{i,j}^{t-1}, \mathbf{X}_{t-k} \text{ for } k \ge 2\big) \\
&= P\big(X_{i,j}^t = 0 \,|\, X_{i,j}^{t-1} = 1, \mathbf{X}_{t-1} \setminus X_{i,j}^{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-m}\big)\,,
\end{aligned}$$

Hence

$$P\big(X_{i,j}^t = 1 \,|\, \mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-m}\big) = \alpha_{i,j}^{t-1} + X_{i,j}^{t-1}(1 - \alpha_{i,j}^{t-1} - \beta_{i,j}^{t-1}) \equiv \gamma_{i,j}^{t-1}\,,$$

i.e. $X_{i,j}^t \,|\, \mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-m} \sim \text{Bernoulli}(\gamma_{i,j}^{t-1})\,, \qquad 1 \le i < j \le p\,.$

Transitivity models

Let

$$\alpha_{i,j}^t = \xi_i \xi_j \frac{e^{aU_{i,j}^{t-1}}}{1 + e^{aU_{i,j}^{t-1}} + e^{bV_{i,j}^{t-1}}}, \quad \beta_{i,j}^t = \eta_i \eta_j \frac{e^{bV_{i,j}^{t-1}}}{1 + e^{aU_{i,j}^{t-1}} + e^{bV_{i,j}^{t-1}}},$$

where $U_{i,j}^{t-1} = \sum_k X_{i,k}^{t-1} X_{j,k}^{t-1}$ is no. of common friends of nodes i and j at time $t-1$ — used by Facebook and LinkedIn,

$V_{i,j}^{t-1} = \sum_k \{X_{i,k}^{t-1}(1 - X_{j,k}^{t-1}) + (1 - X_{i,k}^{t-1}) X_{j,k}^{t-1}\}/2$ is a distance measure bwt nodes i and j,

ξ_i, η_i, a and b are non-negative parameters.

Density-dependent models

$$\alpha_{i,j}^{t-1} = \frac{\xi_i \xi_j \vartheta_{i,j}^{t-1}}{1 + \vartheta_{i,j}^{t-1} + \varpi_{i,j}^{t-1}}, \qquad \beta_{i,j}^{t-1} = \frac{\eta_i \eta_j \varpi_{i,j}^{t-1}}{1 + \vartheta_{i,j}^{t-1} + \varpi_{i,j}^{t-1}}$$

where $\vartheta_{i,j}^{t-1} = \exp\{a_0 D_{-i,-j}^{t-1} + a_1(D_i^{t-1} + D_j^{t-1})\}$,

$\varpi_{i,j}^{t-1} = \exp\{b_0(1 - D_{-i,-j}^{t-1}) + b_1(2 - D_i^{t-1} - D_j^{t-1})\}$, and

$$D_{-i,-j}^{t-1} = \frac{1}{(p-2)(p-3)} \sum_{k,\ell:\, k,\ell \neq i,j,\, k \neq \ell} X_{k,\ell}^{t-1}, \quad D_i^{t-1} = \frac{1}{p-1} \sum_{\ell:\, \ell \neq i} X_{i,\ell}^{t-1}.$$

Assume $a_i, b_i > 0$. The propensity to form new edge between nodes i and j at time t is positively impacted by densities $D_{-i,-j}^{t-1}$, D_i^{t-1} and D_j^{t-1},

The propensity to dissolve an existing edge between nodes i and j at time t is negatively impacted by those three densities.

Persistence models

$$\alpha_{i,j}^{t-1} = \theta_i \theta_j \exp[-1 - a\{(1 - X_{i,j}^{t-2}) + (1 - X_{i,j}^{t-2})(1 - X_{i,j}^{t-3})\}],$$
$$\beta_{i,j}^{t-1} = \eta_i \eta_j \exp\{-1 - b(X_{i,j}^{t-2} + X_{i,j}^{t-2} X_{i,j}^{t-3})\}.$$

This is AR(3) model with parameters $a, b, \theta_i, \eta_i > 0$.

Hence if an edge status between two nodes are unchanged for a time period of 2 or 3, the probability for it remaining unchanged next time is larger than that otherwise.

The persistent connectivity or non-connectivity is widely observed in, for example, brain networks, gene connections and social networks.

Relationship to temporal exponential random graph models

- A TERGM with conditional independent edges (given the past networks) (Hanneck et al 2010) is the AR(m) network model with

$$\alpha_{i,j}^{t-1} = \frac{e^{\phi(\boldsymbol{\theta})^\top f_{ij}(\mathbf{X}_{t-1} \backslash X_{ij}^{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-m})}}{1 + e^{\phi(\boldsymbol{\theta})^\top f_{ij}(\mathbf{X}_{t-1} \backslash X_{ij}^{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-m})}},$$

$$\beta_{i,j}^{t-1} = \frac{e^{\psi(\boldsymbol{\theta})^\top g_{ij}(\mathbf{X}_{t-1} \backslash X_{ij}^{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-m})}}{1 + e^{\psi(\boldsymbol{\theta})^\top g_{ij}(\mathbf{X}_{t-1} \backslash X_{ij}^{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-m})}}.$$

- A separable TERGM with conditional independent edges (given the past networks) (Krivitsky & Handcock 2014) is the AR(m) network model for which $\alpha_{i,j}^{t-1}$ and $\beta_{i,j}^{t-1}$ are defined as above but with $\phi(\boldsymbol{\theta})$ and $\psi(\boldsymbol{\theta})$ replaced, respectively, by $\phi(\boldsymbol{\theta}_\alpha)$ and $\psi(\boldsymbol{\theta}_\beta)$.

Stationarity

A general setting:

$$\alpha_{i,j}^{t-1} = f_{i,j}(\mathbf{X}_{t-1} \setminus X_{i,j}^{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-m}; \boldsymbol{\theta}_0)\,,$$
$$\beta_{i,j}^{t-1} = g_{i,j}(\mathbf{X}_{t-1} \setminus X_{i,j}^{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-m}; \boldsymbol{\theta}_0)\,,$$

where $f_{i,j}$'s and $g_{i,j}$'s are known functions.

When $m = 1$ and $f_{i,j},\, g_{i,j} \in (0, 1)$ (non-sparse networks), $\{\mathbf{X}_t\}_{t\geq 1}$ is an irreducible homogeneous Markov chain with $2^{p(p-1)/2}$ states. Hence for any fixed p, $\{\mathbf{X}_t\}_{t\geq 1}$ is strictly stationary and ergodic.

The sample means of some summary statistics of $\mathbf{X}_t$ may converge faster that the sample mean of $\mathbf{X}_t$ for any fixed p.

When p diverges together with n, the sample means may no longer converge even when $\mathbf{X}_t$ is stationary.

Stationarity is not an asymptotic property but ergodicity is.

Technical challenges

Edges are no longer independent

Number of (local) parameters may be of the size of network p

A silver lining: Based on conditional independence, a martingale difference structure can be constructed, which paves the way for the asymptotic analysis of the statistical inference.

The limiting distributions of MLEs are not normal in general, and they reduce to normal when the underlying process satisfies some mixing conditions.

Local and global parameters

Put $\gamma_{i,j}^{t-1}(\boldsymbol{\theta}) = \alpha_{i,j}^{t-1}(\boldsymbol{\theta}) + X_{i,j}^{t-1}\{1 - \alpha_{i,j}^{t-1}(\boldsymbol{\theta}) - \beta_{i,j}^{t-1}(\boldsymbol{\theta})\}$,

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)$. Write $\gamma_{i,j}^{t-1} = \gamma_{i,j}^{t-1}(\boldsymbol{\theta}_0)$, and $\boldsymbol{\theta}_0$ is the true value.

Let

$$\mathcal{G} = \left\{l \in [q] : \gamma_{i,j}^{t-1}(\boldsymbol{\theta}) \text{ involves } \theta_l \text{ for all } 1 \le i < j \le p \text{ and } t \in [n] \setminus [m]\right\}$$

$\boldsymbol{\theta}_{\mathcal{G}}$: global parameters

$\boldsymbol{\theta}_{\mathcal{G}^c}$: local parameters

In all the 3 models, ξ_i, η_i are local parameters, all other parameters are global.

Note. $\boldsymbol{\theta}_{\mathcal{G}}$ and $\boldsymbol{\theta}_{\mathcal{G}^c}$ need to be treated differently, their estimators may entertain different convergence rates.

Assume $|\mathcal{G}|$ is finite and fixed (when $n, p \to \infty$).

Log likelihood, and martingale differences

Let $\mathcal{F}_t = \sigma(\mathbf{X}_1, \cdots, \mathbf{X}_t)$. For any $l \in [q]$, put

$$\mathcal{S}_l = \left\{(i,j) : 1 \le i < j \le p \text{ and } \gamma_{i,j}^{t-1}(\boldsymbol{\theta}) \text{ involves } \theta_l \text{ for any } t \in [n] \setminus [m]\right\},$$

$$\widehat{\ell}_{n,p}^{(l)}(\boldsymbol{\theta}) = \frac{1}{(n-m)|\mathcal{S}_l|} \sum_{t=m+1}^{n} \sum_{(i,j)\in\mathcal{S}_l} \left[\log\{1-\gamma_{i,j}^{t-1}(\boldsymbol{\theta})\} + X_{i,j}^t \log\left\{\frac{\gamma_{i,j}^{t-1}(\boldsymbol{\theta})}{1-\gamma_{i,j}^{t-1}(\boldsymbol{\theta})}\right\}\right],$$

$$\ell_{n,p}^{(l)}(\boldsymbol{\theta}) = \frac{1}{(n-m)|\mathcal{S}_l|} \sum_{t=m+1}^{n} \sum_{(i,j)\in\mathcal{S}_l} \mathbb{E}_{\mathcal{F}_{t-1}}\left[\log\{1-\gamma_{i,j}^{t-1}(\boldsymbol{\theta})\} + X_{i,j}^t \log\left\{\frac{\gamma_{i,j}^{t-1}(\boldsymbol{\theta})}{1-\gamma_{i,j}^{t-1}(\boldsymbol{\theta})}\right\}\right].$$

Then $\widehat{\ell}_{n,p}^{(l)}(\boldsymbol{\theta})$ is a part of log-likelihood involving θ_l, and

$$\widehat{\ell}_{n,p}^{(l)}(\boldsymbol{\theta}) - \ell_{n,p}^{(l)}(\boldsymbol{\theta}) \equiv \frac{1}{n-m} \sum_{t=m+1}^{n} M_t$$

where $\{M_t\}$ is a sequence of martingale differences.

Real data analysis: Email interactions

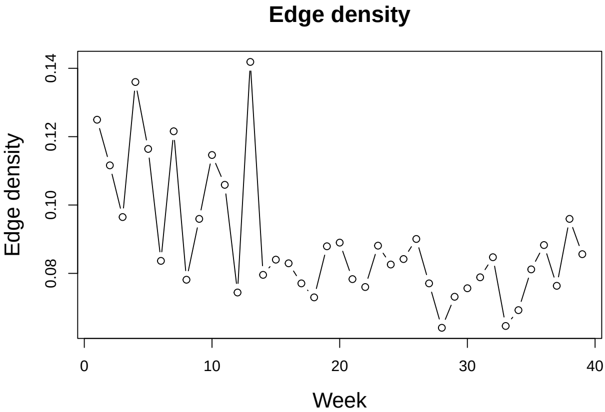
The email interactions in a medium-sized Polish manufacturing company in January – September 2010 (Michalski et al., 2014)

Consider $p = 106$ of the most active participants out of an original 167 employees

$n = 39$ represents 39 weeks, and $X_{i,j}^t = 1$ if participants i and j exchanged at least one email during Week t.

To gain some insight, we first present some preliminary summaries of the data.

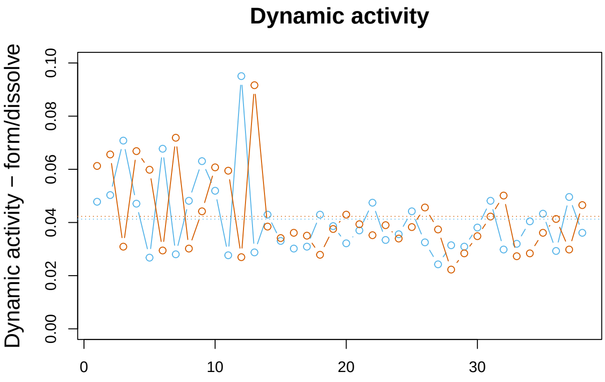
Edge density



Plot of percentage of edges $D_t = \frac{2}{p(p-1)} \sum_{1 \le i < j \le p} X_{i,j}^t$ against t.

A change-point at $t = 14$: Period 1 – first 13 points, Period 2 – last 26 points

Densities of newly formed edges, and newly dissolved edges



Plot of percentage of grown $D_{1,t} = \frac{2}{p(p-1)} \sum_{1 \le i < j \le p} (1 - X_{i,j}^{t-1}) X_{i,j}^t$ and dissolved $D_{0,t} = \frac{2}{p(p-1)} \sum_{1 \le i < j \le p} X_{i,j}^{t-1} (1 - X_{i,j}^t)$ against t.

As $\bar{D}_{1,\cdot} \approx \bar{D}_{2,\cdot} \approx 0.04$, the relative frequency to grow new edge is about 5%, and that to dissolve existing edge is about 45%.

Empirical evidence for transitivity effects

Recall the transitivity model

$$\alpha_{i,j}^t = \xi_i \xi_j \frac{e^{aU_{i,j}^{t-1}}}{1 + e^{aU_{i,j}^{t-1}} + e^{bV_{i,j}^{t-1}}}\,, \quad \beta_{i,j}^t = \eta_i \eta_j \frac{e^{bV_{i,j}^{t-1}}}{1 + e^{aU_{i,j}^{t-1}} + e^{bV_{i,j}^{t-1}}}\,,$$

where $U_{i,j}^t = \sum_{k \neq i,j} X_{i,k}^t X_{j,k}^t,\ V_{i,j}^t = \sum_{k \neq i,j} \{X_{i,k}^t(1 - X_{j,k}^t) + (1 - X_{i,k}^t) X_{j,k}^t\}$.

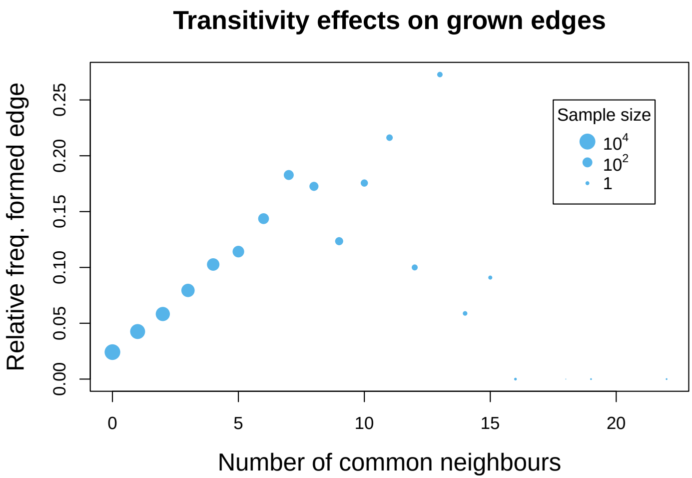
Let

$$\mathcal{U}_\ell = \big\{(i,j,t) : 1 \le i < j \le p\,,\ t \in [n] \setminus \{1\}\,,\ X_{i,j}^{t-1} = 0\,,\ U_{i,j}^{t-1} = \ell\big\}\,,$$

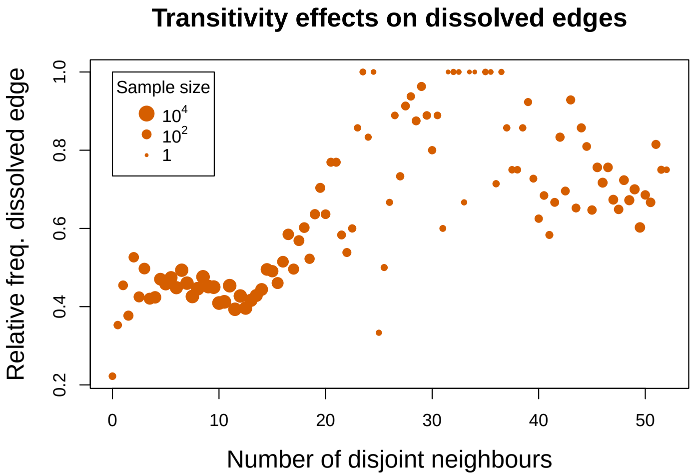
$$\mathcal{V}_\ell = \big\{(i,j,t) : 1 \le i < j \le p\,,\ t \in [n] \setminus \{1\}\,,\ X_{i,j}^{t-1} = 1\,,\ V_{i,j}^{t-1} = \ell\big\}\,,$$

$$\mathcal{U}_\ell^1 = \big\{(i,j,t) \in \mathcal{U}_\ell,\ X_{i,j}^t = 1\big\}, \quad \mathcal{V}_\ell^0 = \big\{(i,j,t) \in \mathcal{V}_\ell,\ X_{i,j}^t = 0\big\}\,.$$

Transitivity: both $|\mathcal{U}_\ell^1|/|\mathcal{U}_\ell|$ and $|\mathcal{V}_\ell^0|/|\mathcal{V}_\ell| \nearrow$, as $\ell \nearrow$.



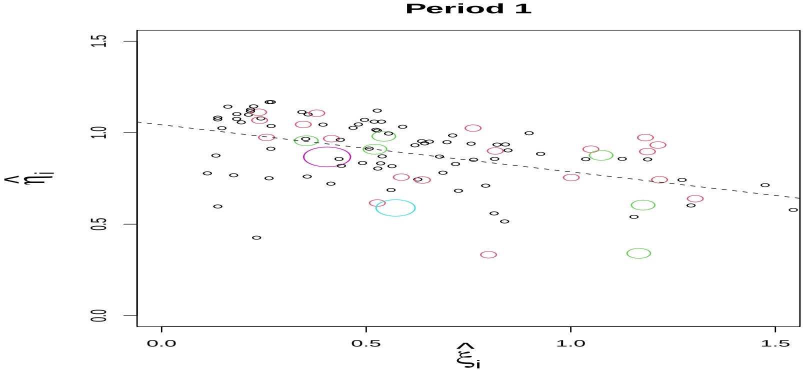
Plot of relative edge frequency $|\mathcal{U}_\ell^1|/|\mathcal{U}_\ell|$ against ℓ for $\ell = 0, 1, \cdots$.



Plot of relative non-edge frequency $|\mathcal{V}_\ell^0|/|\mathcal{V}_\ell|$ against ℓ for $\ell = 0, 1, \cdots$.

Fitting for Period 1

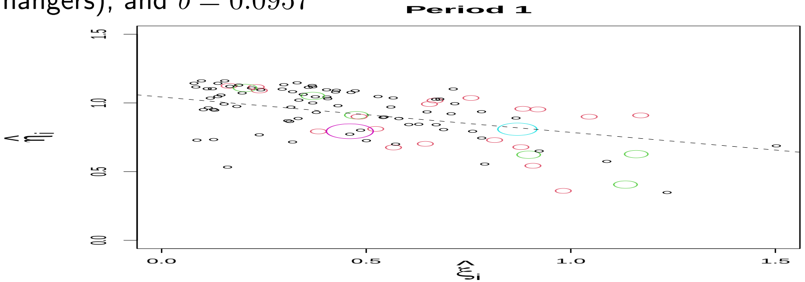
$\widehat{a} = 0.1273$ and $\widehat{b} = 0.0916$



$\widehat{\xi}_i$ and $\widehat{\eta}_i$ are negatively correlated: employees who tend to grow new edges also tend to maintain existing edges.

Fitting for Period 2

$\widehat{a} = 0.2099$ – stronger transitivity effect (more email activities among mangers), and $\widehat{b} = 0.0957$



Circles are sized and coloured according to hierarchical levels in the company: the smallest black circles have no direct reports, while the largest purple circle is CEO.

The means of $\widehat{\xi}_i$ for managers and non-managers are, respectively, 0.68 and 0.42: managers are more likely to grow edges. However, this increasing pattern does not continue at higher levels.

Stronger transitivity and lower edge density: concentration of email activities among a smaller group of employees, many of them managers.

Comparison with other models by AIC & BIC

Global AR model:

$$P(X_{i,j}^t = 1 | X_{i,j}^{t-1} = 0) = \alpha, \quad P(X_{i,j}^t = 0 | X_{i,j}^{t-1} = 1) = \beta$$

Edgewise AR model:

$$P(X_{i,j}^t = 1 | X_{i,j}^{t-1} = 0) = \alpha_{i,j}, \quad P(X_{i,j}^t = 0 | X_{i,j}^{t-1} = 1) = \beta_{i,j}$$

Edgewise mean model: $X_{i,j}^t \stackrel{\text{iid}}{\sim} \text{Bernoulli}(P_{i,j})$

Degree parameter mean model: $X_{i,j}^t \stackrel{\text{iid}}{\sim} \text{Bernoulli}(\nu_i \nu_j)$

No edge dependence in the above 4 models

No dynamic dependence in the last 2 models

No. of parameters is, respectively, 2, $p(p-1)$, $\frac{1}{2}p(p-1)$ and p.

AR transitivity model has $2p+2$ parameters.

	Period 1		**Period 2**	
Model	**AIC**	**BIC**	**AIC**	**BIC**
Transitivity AR model	**33227**	**35176**	52548	**54654**
Global AR model	36309	36327	58267	58287
Edgewise AR model	42717	144102	55840	165394
Edgewise mean model	33248	83941	**47133**	101910
Degree parameter mean model	41730	42695	68969	70013

For Period 1, AR transitivity model achieves the lowest AIC and BIC.

For Period 2, it achieves the lowest BIC, and the 2nd lowest AIC (behind the edgewise mean model).

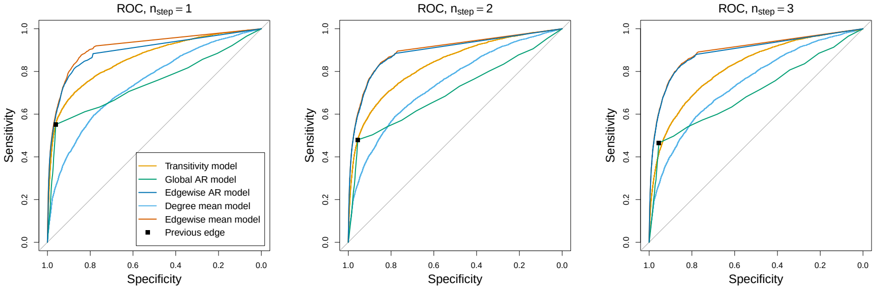
Post-sample edge forecasting

For 26 networks in Period 2, train models based on the first n_{train} data for $n_{\text{train}} = 10, \ldots, 23$.

Based on the fitted model, we make n_{step}-step forward prediction for $\mathbf{X}_{n_{\text{train}}+n_{\text{step}}}$ for $n_{\text{step}} = 1, 2, 3$.

The combined results are presented in ROC curves.

ROC curves: Sensitivity $= \frac{TP}{TP+FN}$, Specificity $= \frac{TN}{TN+FP}$



The two edgewise models (with $O(p^2)$ parameters) perform about the same, are clearly better than all the other models.

The transitivity model (with $O(p)$ parameters) outperform the other three models.