Time Evolving Community Detection in Dynamic Networks

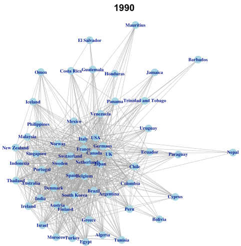
Amal Agarwal
Joint with Kevin Lee and Lingzhou Xue

PENNSTATE
1855

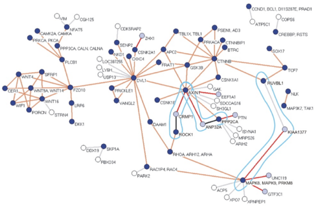
May 17, 2018

Network Science

- “Relational Data" in the form of a graph/adjacency matrix.
  - “Static Network" vs “Dynamic Network"
  - “Undirected Network" vs “Directed Network"
  - “Binary Network" vs “Weighted Network"



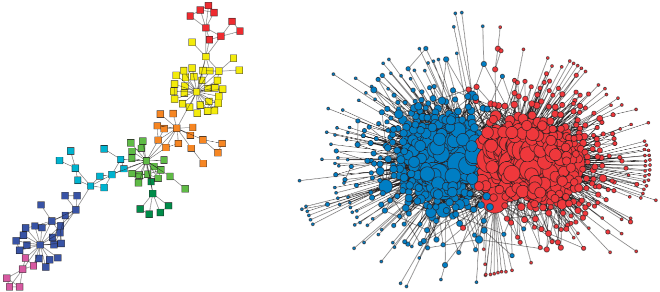
International Trade Network in 1990 *(Ward and Hoff 2007)*



A Human Protein-Protein Interaction Network *(Stelzl, Ulrich, et al. 2005)*

Clustering Network (Community Detection)

- Clustering based on algorithm optimizing community criterion
- Clustering based on probability model
  - Stochastic Block Model (SBM)
  - Exponential-family Random Graph Model (ERGM)



(Left) Network of collaborations among scientists *(Newman 2011)*

(Right) Network of links between US political blogs *(Adamic and Glance 2005)*

Exponential-family Random Graph Model (ERGM)

- **ERGM can be written in the following form:** *(Wasserman and Pattison 1996)*

$$P_\theta(\boldsymbol{Y} = \boldsymbol{y}) = \exp\{\boldsymbol{\theta}' \boldsymbol{g}(\boldsymbol{y}) - \psi(\boldsymbol{\theta})\}$$

- $\mathbf{Y} = (Y_{ij})_{1 \leq i,j \leq n}$ is a random network.
- $\boldsymbol{\theta}$ are network parameters, $\boldsymbol{g}(\mathbf{y})$ is sufficient network statistics and $\psi(\boldsymbol{\theta})$ is the normalization constant.

- **Temporal ERGM can be written as:** *(Hanneke et al. 2010)*

$$P_\theta(\boldsymbol{Y}_t = \boldsymbol{y}_t | \mathbf{y}_{t-1}) = \exp\{\boldsymbol{\theta}' \boldsymbol{g}(\boldsymbol{y}_t, \mathbf{y}_{t-1}) - \psi_t(\boldsymbol{\theta})\}$$

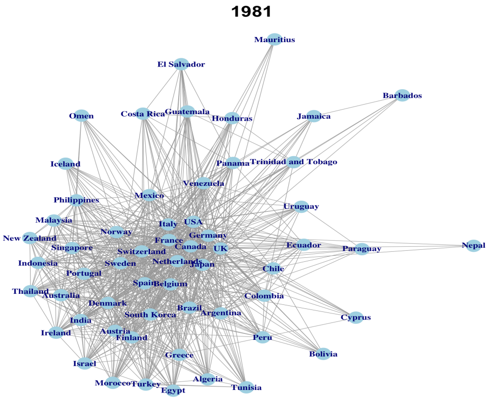
- $\boldsymbol{Y} = (Y_{t,ij})_{1 \leq i,j \leq N, 1 \leq t \leq T}$ is a dynamic network with $\boldsymbol{Y}_t$ as network snapshot at time t.
- $\boldsymbol{g}(\boldsymbol{y}_t, \boldsymbol{y}_{t-1})$ is sufficient dynamic network statistics and $\psi_t(\boldsymbol{\theta})$ is the normalization constant.

Outline

Outline

Methodology Challenge

No framework for time evolving community detection in ERGMs!



Animation showing Dynamic Trade Network during 1981-2000

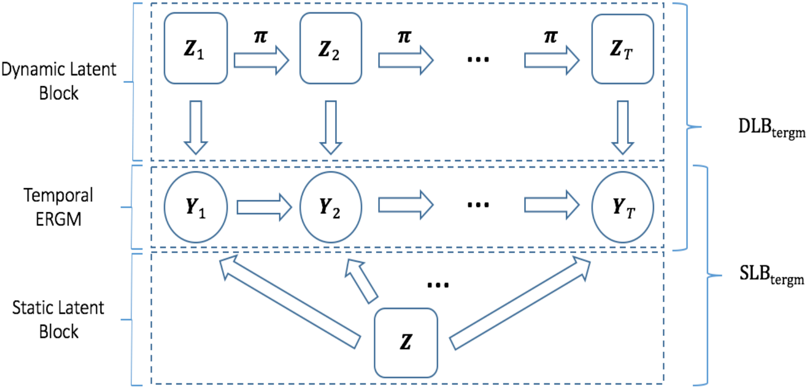
(Ward and Hoff 2007)

Outline

Notations and Model Setup

- $\mathbf{Y}_t = (Y_{t,ij})_{1\leq i,j\leq N}$: adjacency matrix for network snapshot at time t.
- $\boldsymbol{Z} = (Z_{ti,1\leq t\leq T,1\leq i\leq N})$: time evolving community memberships.
- $g(\boldsymbol{y}_t)$: Static Network Statistic
- $g(\boldsymbol{y}_t, \boldsymbol{y}_{t-1})$: Dynamic Network Statistic
- π: Mixture proportions/ Transition probabilities
- θ: Network Parameters

Time Evolving Community Detection



Dynamic Latent Block Model based on Hidden Markov Structure

Time Evolving Community Detection

$$
\begin{aligned}
P_{\pi,\theta}(\boldsymbol{Y}) = &\sum_{z \in \mathcal{Z}} P_\theta(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{Z} = \boldsymbol{z}) P_\gamma(\boldsymbol{Z} = \boldsymbol{z}) \\
= &P(\boldsymbol{Y}_1) \sum_{\boldsymbol{z} \in \mathcal{Z}} \Bigg[\prod_{t=2}^{T} P_\theta(\boldsymbol{Y}_t = \boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{Z}_t = \boldsymbol{z}_t) \times \\
&{\color{red} P_\pi(\boldsymbol{Z}_t = \boldsymbol{z}_t \mid \boldsymbol{z}_{t-1})} \Bigg]
\end{aligned}
$$

- **Advantages:**
  - Incorporated dynamic network features
  - No label switching
  - Incorporated time evolving communities

Dynamic Network Statistics

- **Density statistic**

$$g^{\mathrm{d}}(\mathbf{y}_t, \mathbf{y}_{t-1}) = \sum_{i,j} y_{t,ij}$$

- **Stability statistic**

$$g^{\mathrm{s}}(\mathbf{y}_t, \mathbf{y}_{t-1}) = \sum_{i,j} \left[y_{t,ij} y_{t-1,ij} + (1 - y_{t,ij})(1 - y_{t-1,ij}) \right]$$

- **Transitivity statistic**

$$g^{\mathrm{tr}}(\mathbf{y}_t, \mathbf{y}_{t-1}) = \sum_{i,m,j} y_{t,ij} y_{t-1,im} y_{t-1,mj}$$

Identifiability

- **Generic Identifiability:** All non-identifiable parameters lie on a subset of the full parameter space whose Lebesgue measure is zero.
- **Theorem:** Let N be the number of nodes in temporal networks. The conditional probability $P_{\theta_{kl}}(Y_{t,ij} = 1 | \boldsymbol{y}_{t-1}, \boldsymbol{z}_t)$ and the transition matrix π, are **generically identifiable** up to label switching, if

$$\begin{cases} N^{1/2} \geq K - 1 + (K+2)^2/4, & \text{for } K \text{ even;} \\ N^{1/2} \geq K - 1 + (K+1)(K+3)/4, & \text{for } K \text{ odd.} \end{cases}$$

Challenges for Time Evolving Community Detection

1. ERGMs are **not scalable** for modeling large networks.

   (Snijders 2002; Hunter and Handcock 2006)

2. Computationally **intractable log-likelihood** function.

Conditional Dyadic Independence Assumption

- **Dyadic Independence (DI)** *(Handcock 2003)*

$$P_\theta(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{y}_{t-1}) = \prod_{i<j} P_\theta(Y_{t,ij} = y_{t,ij} | \mathbf{y}_{t-1}).$$

  - Advantages: Easy to simulate and estimate; solves degeneracy
  - Disadvantage: Too restrictive assumption

- **Conditional Dyadic Independence (CDI)** *(Vu, Hunter and Schweinberger, 2013; Lee, Xue and Hunter 2016)*

$$P_\theta(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{z}) = \prod_{i<j} P_{\theta_{z_i z_j}}(Y_{t,ij} = y_{t,ij} | \mathbf{y}_{t-1}, \boldsymbol{z}).$$

  - Advantage: Much less restrictive

- **We extend CDI to time evolving communities** i.e. $\forall t = 2, \ldots T$,

$$P_\theta(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{z}_t) = \prod_{i<j} P_{\theta_{z_{ti} z_{tj}}}(Y_{t,ij} = y_{t,ij} | \mathbf{y}_{t-1}, \boldsymbol{z}_t)$$

EM Algorithm?

- **MLE:** $(\hat{\pi}, \hat{\theta}) = \arg\max_{\boldsymbol{\pi},\boldsymbol{\theta}} \log \mathbb{P}_{\boldsymbol{\pi},\boldsymbol{\theta}}(Y)$
- **Is EM algorithm the most obvious choice?**
  - E-step: Calculate $E_{\pi_{t0},\theta_0}(l_{\text{comp}(\boldsymbol{\pi}_t,\boldsymbol{\theta})}(\boldsymbol{\pi},\boldsymbol{\theta})|\boldsymbol{Y}_{\text{obs}})$ given initial π_{t0} and θ_0.
  - M step: Maximize the calculated expectation w.r.t. $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$.
- **A critical** *computing challenge* **for large networks:**
  - E-step requires computation of $\mathbb{P}_\theta(\boldsymbol{Z}|\boldsymbol{Y})$.

$$\mathbb{P}_\theta(\boldsymbol{Z}|\boldsymbol{Y}) = \mathbb{P}_\theta(Z_1|Y_1)\prod_{t=2}^{T}\mathbb{P}_\theta(Z_t|Z_{t-1}, Y_t)$$

  - $\mathbb{P}_\theta(Z_t|Z_{t-1}, Y_t)$ can't be factored further over nodes.

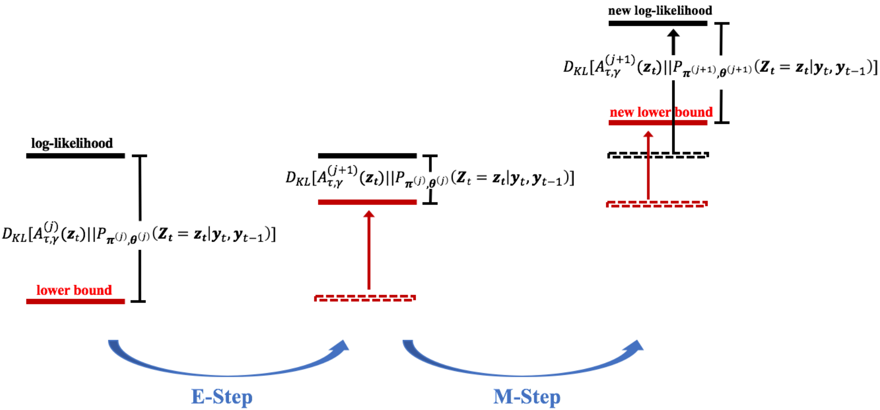
Variational EM Algorithm

- **Why Variational EM (VEM) instead of EM?**

1. VEM uses $A(\boldsymbol{z})$ to approximate $\mathbb{P}_\theta(\boldsymbol{Z}|\boldsymbol{Y})$ (e.g. mean field approximation)
2. Construct a tractable lower bound of the intractable log-likelihood (by using Jensen's inequality)
3. Maximize the lower bound, yielding approximate ML estimates. *(Wainwright and Jordan 2008, Hunter and Lange, 2004)*

- Variational inference: A review for statisticians *(Blei et al. 2017)*

Variational EM Algorithm: Ascent Property



Approximate MLE and ascent property

Constructing tractable lower bound

- **Using Jensen's inequality:**

$$
\begin{aligned}
&\sum_{t=2}^{T} \log P_\theta(\mathbf{Y}_t = \mathbf{y}_t|\mathbf{y}_{t-1}) \\
&\quad = \sum_{t=2}^{T} \log \left[\sum_{\mathbf{z}_t \in \mathcal{Z}} \frac{P_{\pi_t,\theta}(\mathbf{Y}_t = \mathbf{y}_t, \mathbf{Z}_t = \mathbf{z}_t|\mathbf{y}_{t-1})}{A_{\tau,\gamma}(\mathbf{z}_t)} A_{\tau,\gamma}(\mathbf{z}_t) \right] \\
&\quad \geq \sum_{t=2}^{T} \sum_{\mathbf{z}_t \in \mathcal{Z}} \left[\log \frac{P_{\pi_t,\theta}(\mathbf{Y}_t = \mathbf{y}_t, \mathbf{Z}_t = \mathbf{z}_t|\mathbf{y}_{t-1})}{A_{\tau,\gamma}(\mathbf{z}_t)} \right] A_{\tau,\gamma}(\mathbf{z}_t). \\
&\quad = \mathrm{ELBO}(\pi, \theta, \alpha; \tau, \gamma)
\end{aligned}
$$

Outline

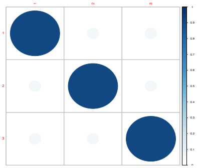
Models

- **Simulation settings:**
  - $T = 10$, $N = 500$ and $K = 3$.
  - $\theta = (-0.3, 0, 0.3)$
  - $\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
- **4 different models:**
  - Model 1: *Undirected*; Network Statistic: *Density*
  - Model 2: *Undirected*; Network Statistic: *Stability*
  - Model 3: *Directed*; Network Statistic: *Density*
  - Model 4: *Directed*; Network Statistic: *Transitivity*
- Each model has 4 different transition probabilities.

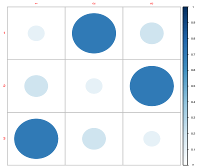
Homogeneous Settings for π

$$\pi_t^{(S_1)} = \begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{pmatrix}$$

$$\pi_t^{(S_2)} = \begin{pmatrix} 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.7 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}$$



$\pi_{1 \leq t \leq 9}^{(S_1)}$

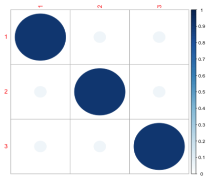


$\pi_{1 \leq t \leq 9}^{(S_2)}$

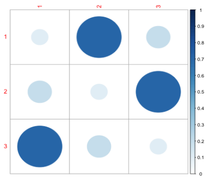
Inhomogeneous Settings for π

$$\pi_t^{(S_3)} = \begin{bmatrix} 0.9-0.1t & 0.05+0.05t & 0.05+0.05t \\ 0.05+0.05t & 0.9-0.1t & 0.05+0.05t \\ 0.05+0.05t & 0.05+0.05t & 0.9-0.1t \end{bmatrix}$$

$$\pi_t^{(S_4)} = \begin{bmatrix} 0.1 & 0.7-0.05t & 0.2+0.05t \\ 0.2+0.05t & 0.1 & 0.7-0.05t \\ 0.7-0.05t & 0.2+0.05t & 0.1 \end{bmatrix}$$



$\pi_{1\leq t\leq 9}^{(S_3)}$



$\pi_{1\leq t\leq 9}^{(S_4)}$

Performance Metrics

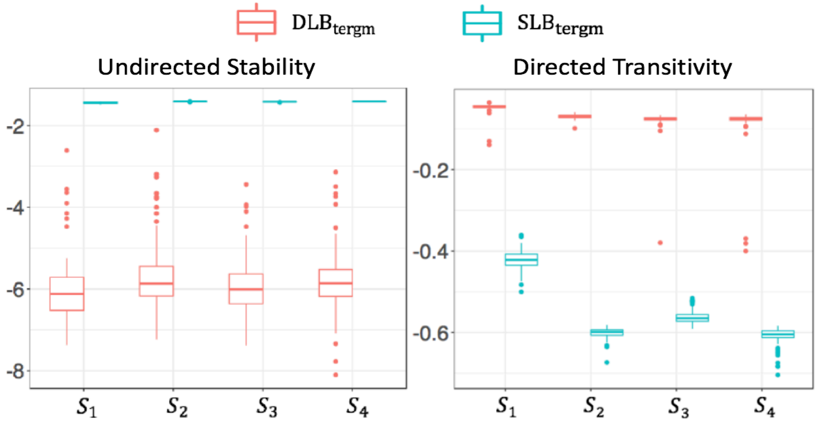
- **To check estimation performance:**
  - Root Average Squared Error for network parameters θ,

$$\mathrm{RASE}_{\theta} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\hat{\theta}_k - \theta_k\right)^2}$$

- **To check clustering performance:**
  - Average Rand Index (ARI) at time point t, *(Rand, 1971)*

$$\mathrm{ARI}(\boldsymbol{z}, \hat{\boldsymbol{z}}) = \frac{1}{T\binom{n}{2}}\sum_{t=1}^{T}\sum_{i<j}\big(I\{z_{t,i} = z_{t,j}\}I\{\hat{z}_{t,i} = \hat{z}_{t,j}\} + I\{z_{t,i} \neq z_{t,j}\} I\{\hat{z}_{t,i} \neq \hat{z}_{t,j}\}\big)$$

Estimation and Clustering Results

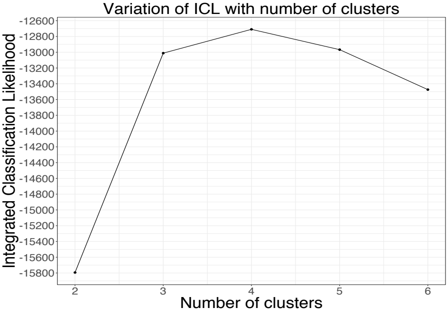


(Left) θ Estimation Performance measured using Log of RASE and (Right) Clustering Performance measured using Log of Average Rand Index, over 100 repetitions

Outline

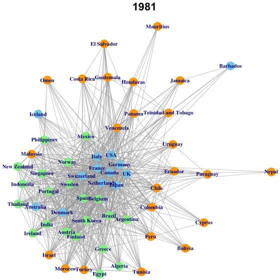
International Trade Network

- Trade Network for 58 countries *(Ward and Hoff 2007)* .
- 20 time points, 1981 – 2000 .



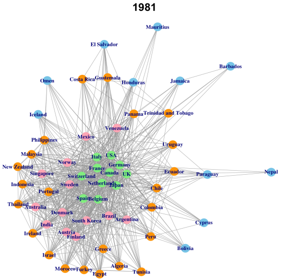
Model Selection: Choosing **4 communities** with density statistic using max ICL criterion

Network Visualization



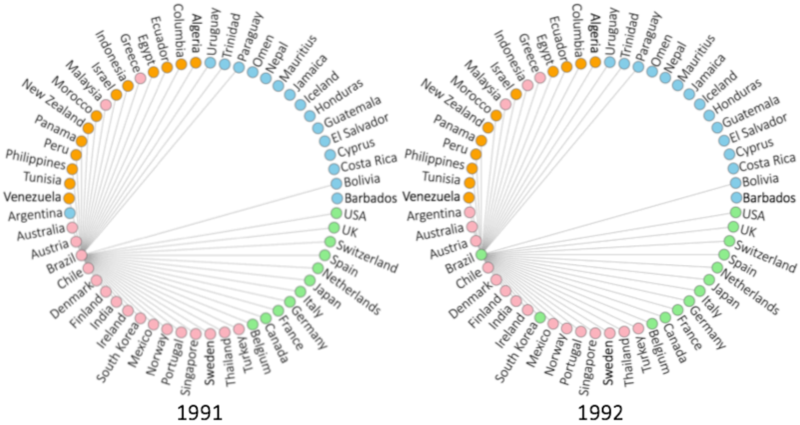
Fixed Community Detection through TERGM

(Lee, Xue and Hunter 2017)



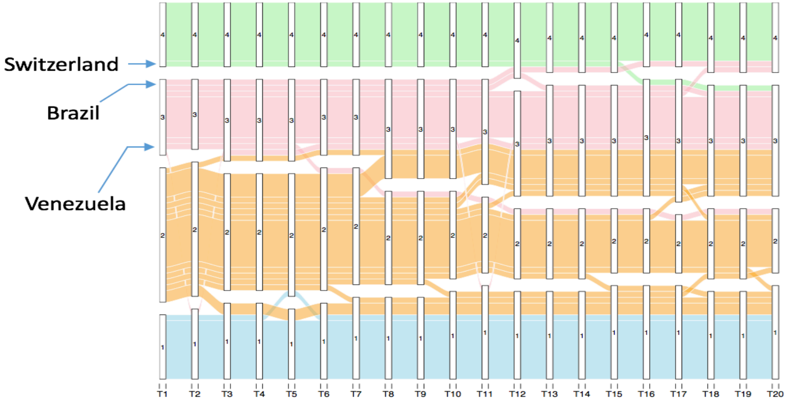
Time Evolving Community Detection through dynamic ERGM

Case Study: Brazil's Trade Network



Brazil's Trade Network with Time Evolving Communities

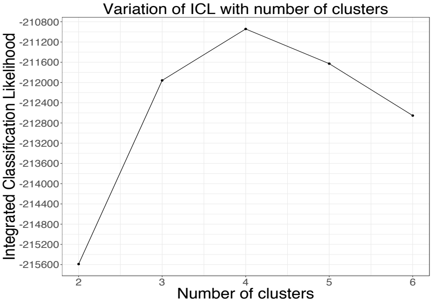
Time Evolving Community Detection Visualization



Alluvial plot for visualizing community membership transitions; T1-T20 represents 1981-2000 in the trade network

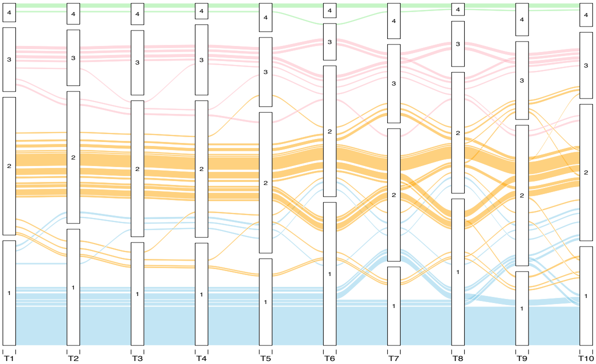
Email Network

- Email dataset *(Yin et al. 2017)* from a large European research institution.
- N=559, 10 time points: .



Model Selection: Choosing **4 communities** with density statistic using max ICL criterion

Time Evolving Community Detection Visualization



Alluvial plot for visualizing cluster membership transitions; T1-T10 represents 10 time segments in the email network

Outline

Conclusion

Time evolving community detection:

- Dynamic network statistics could give significant insights.
- ERGM scalability and log likelihood intractability issues solved.
- Parsimonious model, easy interpretations, network parameters identifiable.