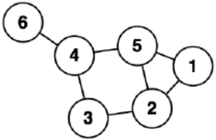
STA6800 - Statistical Analysis of Network

ERGM for Dynamic Networks

Ick Hoon Jin

Yonsei University, Department of Statistics and Data Science

Introduction



An example of cross-sectional structure

- However, a need for statistical models representing the evolving phenomena $\Rightarrow$ ”Dynamic Models” with a temporal structure

Temporal ERGM

- ERGM $\rightarrow$ TERGM $\rightarrow$ STERGM
- One-step transition probability (t-1) $\rightarrow$ (t) (Markov Assumption)

$$\boxed{Pr_{\eta,g}(Y^t = y^t | Y^{t-1} = y^{t-1}; \theta) = \frac{exp(\eta(\theta) * g(y^t, y^{t-1}))}{c_{\eta,g}(\theta, y^{t-1})}}$$

- TERGM: Temporal ERGM $\Rightarrow$ The network at time t is a single draw from an ERGM conditional on the network at time $t-1$ (and possibly time $t-2$)

Temporal ERGM

- Simplify a statistical model for evolving social networks is to make a Markov assumption on the network from one time step to the next.
- If A^t is the weight matrix representation of a single-relation social network at time t, then we might make the assumption that A^t is independent of $A^1, \cdots, A^{t-2}$ given A^{t-1}.
- Temporal ERGM

$$P\left(A^2, A^3, \cdots, A^t \mid A^1\right) = P\left(A^t \mid A^{t-1}\right) P\left(A^{t-1} \mid A^{t-2}\right) \cdots P\left(A^2 \mid A^1\right).$$

- Given our Markov assumption, one natural way to generalize ERGMs for evolving networks is to assume $A^t \mid A^{t-1}$ admits an ERGM representation.

Temporal ERGM

- Specify a function $\Psi : \mathbb{R}_{n\times n} \times \mathbb{R}_{n\times n} \to \mathbb{R}^k$, which can be understood as a temporal potential over cliques across two time-adjacent networks, and parameter vector $\boldsymbol{\theta} \in \mathbb{R}^k$, such that the conditional PDF has the following form:

$$P\left(A^t \mid A^{t-1}, \boldsymbol{\theta}\right) = \frac{1}{\kappa(\boldsymbol{\theta}, A^{t-1})} \exp\left\{\boldsymbol{\theta}^T \Psi\left(A^t, A^{t-1}\right)\right\}.$$

- In particular, we will be especially interested in the special case of these models in which

$$\Psi\left(A^t, A^{t-1}\right) = \sum_{ij} \Psi_{ij}\left(A^t_{ij}, A^{t-1}\right).$$

  This form of the temporal potential function represents situations where the conditional distribution of $A^t \mid A^{t-1}$ factors over the entries A^t_{ij} of A^t.

Network Statistics for Temporal ERGM

- Density: The number of ties in the network as a whole.

$$\Psi_D\left(A^t, A^{t-1}\right) = \frac{1}{n-1}\sum_{ij} A^t_{ij}.$$

- Stability: The tendency of a link that does exist at time $t-1$ to continue existing at time t.

$$\Psi_S\left(A^t, A^{t-1}\right) = \frac{1}{n-1}\sum_{ij}\left\{A^t_{ij}A^{t-1}_{ij} + \left(1 - A^t_{ij}\right)\left(1 - A^{t-1}_{ij}\right)\right\}.$$

Network Statistics for Temporal ERGM

- Reciprocity: The tendency of a link from i to j to result in a link from j to i at the next time step.

$$\Psi_R\left(A^t, A^{t-1}\right) = n\left(\sum_{ij} A^t_{ji} A^{t-1}_{ij}\right) / \left(\sum_{ij} A^{t-1}_{ij}\right).$$

- Transitivity: The tendency of a tie from i to j and from j to k to result in a tie from i to k at the next time step.

$$\Psi_T\left(A^t, A^{t-1}\right) = n\left(\sum_{ijk} A^t_{ik} A^{t-1}_{ij} A^{t-1}_{jk}\right) / \left(\sum_{ijk} A^{t-1}_{ij} A^{t-1}_{jk}\right).$$

Notation

- Use the sequence of observed networks, $N^1, N^2, \cdots, N^T$, to find an estimator $\hat{\theta}$ that is close to the actual parameter value $\boldsymbol{\theta}$.
- The normalizing constant is computationally intractable, often making explicit solutions of MLE difficult.
- Use MCMC stochastic approximation to estimate parameters.

Notation

Let

$$
\begin{aligned}
L\left(\theta : N^1, N^2, \cdots, N^T\right) &= \log P\left(N^2, N^3, \cdots, N^t \mid N^1, \theta\right), \\
M\left(t, \theta\right) &= E_\theta\left(\Psi\left(\underline{N}^t, N^{t-1}\right) \mid N^{t-1}\right), \\
C\left(t, \theta\right) &= E_\theta\left(\Psi\left(\underline{N}^t, N^{t-1}\right)\Psi\left(\underline{N}^t, N^{t-1}\right)^T \mid N^{t-1}\right),
\end{aligned}
$$

where expectations are taken over the random variable $\underline{N}^t$, the network at time t.

Notation

Note that

$$\Delta L\Big(\boldsymbol{\theta} : N^1, N^2, \cdots, N^T\Big) = \sum_{t=2}^{T} \Big(\Psi(N^t, N^{t-1}) - M(t, \boldsymbol{\theta})\Big)$$

$$\Delta^2 L\Big(\boldsymbol{\theta} : N^1, N^2, \cdots, N^T\Big) = \sum_{t=2}^{T} \Big(M(t, \boldsymbol{\theta}) M(t, \boldsymbol{\theta})' - C(t, \boldsymbol{\theta})\Big).$$

- The expectations can be approximated by Gibbs sampling from the conditional distributions.
- Perform an unconstrained optimization procedure akin to Newton's method: Approximate the expectations, update parameter values in the direction that increases the likelihood, repeat until convergence.

Algorithm

Randomly initialize $\boldsymbol{\theta}^{(1)}$.
for $i = 1$ up until convergence **do**
    for $t = 2, \cdots, T$ **do**
        Sample $\hat{N}^{t,1}_{(i)}, \cdots, \hat{N}^{t,B}_{(i)} \sim P\left(\underline{N}^t \mid N^{t-1}, \boldsymbol{\theta}^{(i)}\right)$
        $\hat{\mu}^t_{(i)} = \frac{1}{B}\sum_{b=1}^{B} \Psi\left(\hat{N}^{t,b}_{(i)}, N^{t-1}\right)$
        $\hat{C}^t_{(i)} = \frac{1}{B}\sum_{b=1}^{B} \Psi\left(\hat{N}^{t,b}_{(i)}, N^{t-1}\right)\Psi\left(\hat{N}^{t,b}_{(i)}, N^{t-1}\right)'$
    end for
    $\hat{H}_{(i)} = \sum_{t=2}^{T}\left(\hat{\mu}^t_{(i)}\hat{\mu}^t_{(i)} - \hat{C}^t_{(i)}\right)$
    $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \hat{H}^{-1}_{(i)}\sum_{t=2}^{T}\left(\Psi\left(\hat{N}^{t,b}_{(i)}, N^{t-1}\right) - \hat{\mu}^t_{(i)}\right)$
end for

- The choice of B can affect the convergence of this algorithm. Generally, larger B values will give more accurate updates, and thus fewer iterations needed until convergence.

Convergence

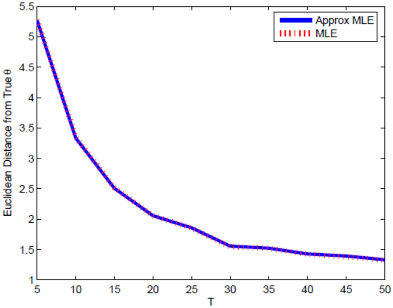


FIG 1. *Convergence of estimation algorithm on simulated data, measured in Euclidean distance of the estimated values from the true parameter values. The approximate MLE from the sampling-based algorithm is almost identical to the MLE obtained by direct optimization.*

Degeneracy of Temporal ERGMs

- In the simple case where the transition distribution factors over the edges, it turns out these models avoid such problems entirely.
- The intuitive reason for this is that, since the edges of A^t are conditionally independent given A^{t-1}, as long as the individual conditional distributions for the A^t_{ij} given A^{t-1} are not too extreme, the conditional entropy of A^t given A^{t-1} should be large, and thus the entropy of A^t itself must be large.
- Of course, this argument only works if the dependence of A^t on A^{t-1} is not too strong, and the strength of this dependence can be controlled by the magnitudes of the parameters.

Degeneracy of Temporal ERGMs

- Calculate it for equivalence classes of graphs which can be analytically shown to have identical probability values, and weight each class according to its size in the entropy calculation.
- For the first plot, since the conditional probability of A^2 given A^1 is only a function of how many edges are present in A^2 and how many ij values have $A^2_{ij} = A^1_{ij}$, and since the edges of A^1 are exchangeable, we can write the marginal distribution of A^2 purely in terms of the number of edges.
- Thus we need only calculate $n(n-1)$ probability values, and the entropy is a weighted sum, where the weights are combinatorial quantities reflecting the number of graphs with that many edges.

Degeneracy of Temporal ERGMs

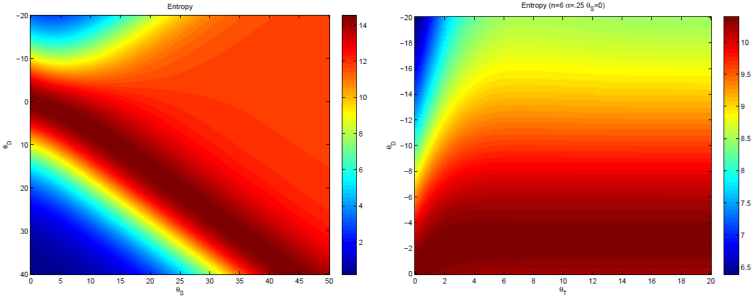


FIG 2. *Entropy plots for the example model. In both plots, small magnitudes of the parameters give distributions with high entropy, as predicted.*

108th U.S. Senate Network Example

- Model the network transitions of the 108th U.S. Senate network.
- The dynamic network has 100 nodes and 12 time points
- We perform two types of experiments here:
    - The first is simply to assess which statistics are important for modeling the network transitions, by observing the magnitudes of the estimated parameters.
    - The second assesses the quality of fit of a model with a cross-validation experiment.
- Network Statistics Used
    1. Density, Stability, and Reciprocity
    2. 1 + Transitivity, Reverse-Transitivity, Co-Supported, and Co-Supporting
    3. 2 + Popularity and Generosity

Three Parameter Model

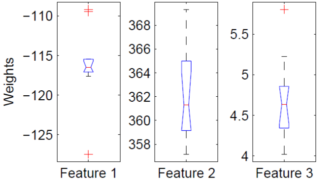


FIG 3. *Estimated parameter values (weights) for a TERGM with 3 statistics (features). Feature 1 is* Density*; Feature 2 is* Stability*; Feature 3 is* Reciprocity.

Description of Network Statistics

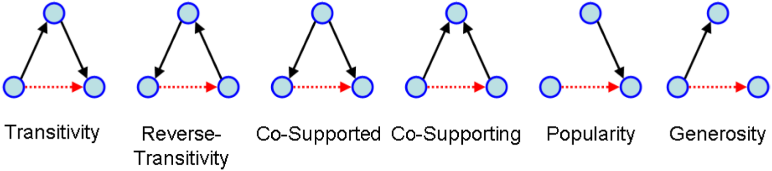


FIG 4. *Graph illustrations of six 3-node statistics corresponding to Features 4—9 in Figure 5 and 6. Blue circles are nodes; black solid arrows represent links (or a supporting relationship) at time $(t-1)$; red dotted arrows represent an edge at time t.*

Network Statistics for Seven Parameter Model

- Recerse-Transitivity:

$$\Psi_{RT}\left(A^t, A^{t-1}\right) = n\Big(\sum_{ijk} A^t_{ij} A^{t-1}_{jk} A^{t-1}_{ki}\Big) \Big/ \Big(\sum_{ijk} A^{t-1}_{jk} A^{t-1}_{ki}\Big).$$

- Co-Supported:

$$\Psi_{CSd}\left(A^t, A^{t-1}\right) = n\Big(\sum_{ijk} A^t_{ij} A^{t-1}_{ki} A^{t-1}_{kj}\Big) \Big/ \Big(\sum_{ijk} A^{t-1}_{ki} A^{t-1}_{kj}\Big).$$

- Co-Supporting:

$$\Psi_{CSd}\left(A^t, A^{t-1}\right) = n\Big(\sum_{ijk} A^t_{ij} A^{t-1}_{ik} A^{t-1}_{jk}\Big) \Big/ \Big(\sum_{ijk} A^{t-1}_{ik} A^{t-1}_{jk}\Big).$$

Network Statistics for Nine Parameter Model

- Popularity

$$\Psi_P\left(A^t, A^{t-1}\right) = n\Big(\sum_{ijk} A^t_{kj} A^{t-1}_{ij}\Big) / \Big(\sum_{ij} A^{t-1}_{kj}\Big).$$

- Generosity

$$\Psi_G\left(A^t, A^{t-1}\right) = n\Big(\sum_{ijk} A^t_{ik} A^{t-1}_{ij}\Big) / \Big(\sum_{ij} A^{t-1}_{ik}\Big).$$

Seven Parameter Model

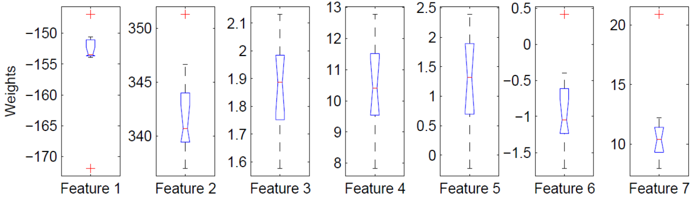


FIG 5. *Estimated parameter values (weights) for a TERGM with 7 statistics (features).*

Nine Parameter Model

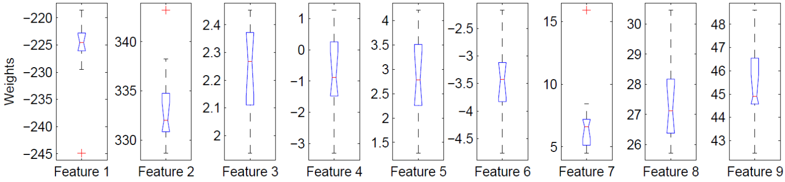


FIG 6. *Estimated parameter values (weights) for a TERGM with 9 statistics (features).*

Goodness-of-Fit

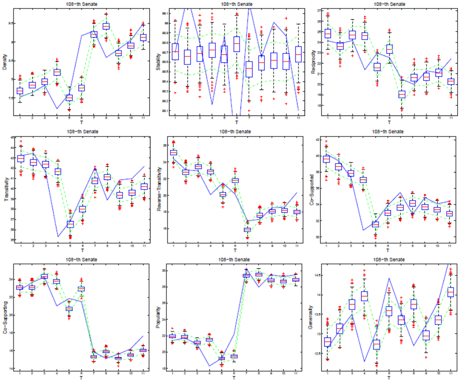


FIG 7. *Statistic values of real networks and sampled networks based on a TERGM with 9 statistics. The comparisons are grouped by statistic. Blue solid lines indicate the observed (true) network statistics. Box plots are for the sampled networks (in the described cross-validation experiments) and green dotted lines indicate 5- and 95-percentiles.*

Introduction

STERGM = A Separable Model for Dynamic Network

- Dynamic : social networks that evolve over time
- Time(discrete) : $\cdots$ (t-2) $\rightarrow$ (t-1) $\rightarrow$ (t) $\rightarrow \cdots$
- Shows longitudinal properties based on the ERGM
- Separable
  - formation : new ties
  - duration : lasting ties

Temporal ERGM Interpretation

- However, caution must be used in interpreting their parameters.
- Property1: incidence of ties (the rate at which new ties are formed)
- Property2: duration of ties (how long they tend to last once they do)
- Network statstic
  - ex) edge count $\boxed{g(y^t, y^{t-1}) = |y^t|}$
  - $\uparrow$ coefficient on g $\rightarrow \uparrow$ possibility of a network with many ties
  - But, this term simultaneously increases the weight of preservation of extant ties (fewer dissolved) $\Rightarrow$ Both incidence and duration $\uparrow$

Temporal ERGM Interpretation

- The two-sided nature of these effects tends to muddle parameter interpretation. $\Rightarrow$ STERGM which separates the incidence and duration of ties and allows for the separate interpretation.
- + : incidence/tie formation $\quad y^{+} = y^{t-1} \cup y^{t}$
- – : duration/tie dissolution $\quad y^{-} = y^{t-1} \cap y^{t} \Rightarrow y^{t} = y^{-} \cup (y^{+} \setminus y^{t-1})$

Separable Temporal ERGM Specification

- If Y^+ is conditionally independent of Y^- given Y^{t-1} then,

$$Pr(Y^+ = y^+ | Y^{t-1} = y^{t-1}; \theta^+) = \frac{exp(\eta^+(\theta^+) * g^+(y^+, y^{t-1}))}{c_{\eta^+, g^+}(\theta^+, y^{t-1})}$$

$$Pr(Y^- = y^- | Y^{t-1} = y^{t-1}; \theta^-) = \frac{exp(\eta^-(\theta^-) * g^-(y^-, y^{t-1}))}{c_{\eta^-, g^-}(\theta^-, y^{t-1})}$$

$$\begin{aligned} Pr(Y^t = y^t | Y^{t-1} = y^{t-1}; \theta) &\times \text{incidence} \times \text{duration} \\ &= Pr(Y^+ = y^+ | Y^{t-1} = y^{t-1}; \theta^+) \\ &\times Pr(Y^- = y^- | Y^{t-1} = y^{t-1}; \theta^-) \end{aligned}$$

Separable Temporal ERGM Specification

Definition

Definition 1 We say that a dynamic model is separable if Y^+ is conditionally independent of Y^- given Y^{t-1} and the parameter space of θ is the product of the individual parameter spaces of θ^+ and θ^-.

- Assumption: During a given discrete time step, the process by which the ties form does not interact with the process by which they dissolve.

Separable Temporal ERGM Specification

- Lost: In the parameterization in terms of formation and dissolution, some flexibility(formation and dissolution processes interact within a given time step) is lost
- Gain: Ease of specification, tractability of the model and substantial improvement in interpretability of TERGM

Interpretability of STERGM

- Now the parameter and its interpretation have an implicit direction. (formation or duration)
- Formation network
  - $\Pr(\mathbf{Y}^{+} = \boldsymbol{y}^{+} | \mathbf{Y}^{\boldsymbol{t-1}} = y^{t-1}; \theta^{+}) = \frac{\exp\{(\theta^{+})^{\cdot T} g^{+}(y^{+}, y^{t-1})}{c_{g^{+}}(\theta^{+}, y^{t-1})}$
  - θ^{+} is related to formation network **only**
- Duration network (or Dissolution network)
  - $\Pr(\mathbf{Y}^{-} = \boldsymbol{y}^{-} | \mathbf{Y}^{\boldsymbol{t-1}} = y^{t-1}; \theta^{-}) = \frac{\exp\{(\theta^{-}) \cdot g^{-}(y^{-}, y^{t-1})}{c_{g^{-}}(\theta^{-}, y^{t-1})}$
  - $\theta^{-}$is related to duration network **only**

Example of Parameter Interpretation (Edge Count)

- Now the parameter and its interpretation have an implicit direction. (formation or duration)
- Formation network
  - Edge count $g^+(y^+, y^{t-1}) = |y^+|$, $y^+ = y^{t-1} \cup y^t$
  - Recall, y^+ is network about **formation**
  - θ^+ means log-odds of gaining new tie from $y^{t-1} \implies y^t$
- Dissolution network
  - Edge count $g^-(y^-, y^{t-1}) = |y^-|$, $y^- = y^{t-1} \cap y^t$
  - Recall, y^- is network about **duration**
  - θ^- means log-odds of existing tie to *survive* at $y^{t-1} \implies y^t$

Likelihood-based Inference for STERGM

- Fit STERGM by finding **conditional MLE** under an order 1 Markov assumption

$$\hat{\theta} = \arg\max_{\theta} \prod_{t=1}^{T} \Pr(Y^t = y^t | Y^{t-1} = y^{t-1})$$

$$= \prod_{t=1}^{T} \frac{\exp\{(\theta^+)^{\cdot T} g^+(y^+, y^{t-1})\}}{c_{g^+}(\theta^+, y^{t-1})} \cdot \frac{\exp\{(\theta^-) \cdot g^-(y^-, y^{t-1})\}}{c_{g^-}(\theta^-, y^{t-1})}$$

where

$$c_g(\theta, y^{t-1}) = \sum_{y^{,} \in \psi} \exp\{(\theta)^{\cdot T} g(y, y^{t-1})\}$$

- In practical, MLE can be obtained by maximizing the log-likelihood using numerical optimization.

Likelihood-based Inference for STERGM

- The normalizing constant $c_{g^+}(\theta^+, y^{t-1})$ and $c_{g^-}(\theta^-, y^{t-1})$ can not be calculated.
- Each of them can be estimated separately by simulation. with MCMCMLE.
- i.e. maximizing

$$l(\theta) - l(\theta^0) = \{\theta - \theta^0\} \sum_{t=1}^{T} g(y^t, y^{t-1}) - \log\{\prod_{t=1}^{T} \frac{c_g(\theta, y^{t-1})}{c_g(\theta^0, y^{t-1})}\}$$

.

Application to the Dynamics of Friendship

- Friendship relation between 26 students during their first years at secondary school (assessed at 4 time points)
- Covariates: Homophilous on sex, Co-attendance of primary school, Reciprocal (pairwise relationship in directed network), Transitivity & Cyclical (traiangular relationship in directed network), Edge count.

Application to the Dynamics of Friendship

Table 2. MLE parameter estimates for the longitudinal friendship network

Parameter	*Formation*		*Dissolution*	
	Estimate	*Standard error*	*Estimate*	*Standard error*
Edges	−3.336†	0.320	−1.132‡	0.448
Homophily (girls)	0.480	0.269	0.122	0.394
Homophily (boys)	0.973§	0.355	1.168‡	0.523
Female → male heterophily	−0.358	0.330	−0.577	0.609
Primary school	0.650§	0.248	0.451	0.291
Reciprocity	1.384†	0.280	2.682†	0.523
Transitive ties	0.886†	0.247	1.121†	0.264
Cyclical ties	−0.389§	0.133	−1.016†	0.231

†Significance level 0.001.
‡Significance level less than 0.05.
§Significance level less than 0.01.

- Homophily by sex is overally stronger for boys, but the effect is stronger for boys at Dissolution rather than Formation.
- Co-attendance in primary school was significance in Formation but not int Dissolution.
- High degree in reciprocity and transitive ties in both Formation and Dissolution.

Conclusion

- Introduce statistical model for networks that evolve over time.
- Separable parameterization of incidence and duration.
- Greatly improve interpretability of model parameters, with sacrificing a little.
- Identify the structure of incident and durational structure