Temporal Exponential Random Graph Models: Modeling Network Panel Data

Instructor: Bruce A. Desmarais
University of Massachusetts Amherst

George Mason University

January 15, 2013

Step 1: Playing Along

1. Go to my website

   `people.umass.edu/bruced`

2. Click or scroll down to "Teaching"
3. Download "GMU TERGM Materials"
4. Unzip and open 'TERGM.pdf'

Workshop Outline

1. (T)ERGM Derivation and Definition
2. Model Specification
3. Parameter Estimation
4. Example Application
5. `R` Tutorial

What does it mean to model the network?

Construct a probability distribution that accurately approximates the network

Why build models?

- Test hypotheses
  Example: Does the cosponsorship network exhibit reciprocity?
- Simulation for theoretical exploration
  Example: How should seats be assigned in a classroom to encourage cross-racial friendships?
- Tie prediction
  Example: Will Canada attack next year?

Advantage of ERGM

Can model how ties depend upon each other

Modeling Interdependence

Two Classes of Questions: Covariate and Interdependence

1. Covariate
   - Do legislators in the same political party collaborate more frequently than those in opposite parties?
   - Do states with democratic governments have more alliances than those with autocratic regimes?
2. Interdependence
   - Are two states at war with the same third state less likely to be at war with each other?
   - Are there popularity effects in the choice of co-authors?

ERGM: integrate effects for any forms of (1) and (2) into a unified model of a network.

The Exponential Random Graph Model (ERGM)

The probability (likelihood function) of observing network N is:

$$\mathcal{P}(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}'\boldsymbol{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}'\boldsymbol{h}(N^*)\}}$$

Decomposition:

$$\underbrace{\boldsymbol{h}(N)}_{Net\ Stats} \qquad \underbrace{\boldsymbol{\theta}}_{Effects} \qquad \underbrace{\exp\{\boldsymbol{\theta}'\boldsymbol{h}(N)\}}_{+\ Weight} \qquad \underbrace{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}'\boldsymbol{h}(N^*)\}}_{Normalizer}$$

Flexible: $\boldsymbol{h}$ can capture virtually any form of interdependence among the edges + covariates

Normalizing constant can make estimation difficult

Concise ERGM Definition

Motivations

1. Statistical/probabilistic model for the entire network
2. Probability of a network depends upon its topology
3. Differentiate the effects of different topological features
4. Use model to simulate, forecast and test competing theories

Mathematical Derivation

1. Statistics (topoligical features) of the network – $\boldsymbol{h}(N)$
2. Parameters that give the effects of the statistics – $\boldsymbol{\theta}$
3. Positively weight each configuration – $\omega(N) = \exp(\boldsymbol{\theta}'\boldsymbol{h}(N))$
4. Normalize into a probability distribution –

$$p(N, \boldsymbol{\theta}) = \frac{\omega(N)}{\sum_{i=1}^{m} \omega(N^i)}$$

TERGM via ERGM

Two differences account for the "T"

1. Inter-temporal network statistics
2. Block-diagonal Adjacency Matrix

Networks Over (discrete) Time

Two types of dynamic network data

1. **Network Snapshot**: Typically by surveys administered at different times
   - Who are your current friends?

2. **Interval Census**: Ties are aggregated over time intervals
   - What wars were fought in the past year?

TERGM is most appropriate for *interval census*.

TERGM

1. Statistics computed on the time series of networks– $\boldsymbol{h}(\boldsymbol{N})$
2. N^t dependent upon k previous networks (k-order Markov)
3. $p(N^t|\boldsymbol{\theta}, \{N^t, N^{t-1}, \ldots, N^{t-k}\}) = \frac{\omega(\{N^t, N^{t-1}, \ldots, N^{t-k}\})}{\sum_{i=1}^{m} \omega(\{N^t, N^{t-1}, \ldots, N^{t-k}\})}$
4. $p(\{N^T, N^{T-1}, \ldots, N^{k+1}\}|\boldsymbol{\theta}, \{N^k, N^{k-1}, \ldots, N^1\}) =$

$$\prod_{t=k+1}^{T} p(N^t|\boldsymbol{\theta}, \{N^t, N^{t-1}, \ldots, N^{t-k}\})$$

Block-Diagonal Network Representation

	a_1	b_1	c_1	d_1	a_2	b_2	c_2	d_2	a_3	b_3	c_3	d_3
a_1	0	1	0	0	×	×	×	×	×	×	×	×
b_1	0	0	1	0	×	×	×	×	×	×	×	×
c_1	0	1	0	0	×	×	×	×	×	×	×	×
d_1	0	0	0	0	×	×	×	×	×	×	×	×
a_2	×	×	×	×	0	1	0	0	×	×	×	×
b_2	×	×	×	×	1	0	0	0	×	×	×	×
c_2	×	×	×	×	1	1	0	1	×	×	×	×
d_2	×	×	×	×	1	0	0	0	×	×	×	×
a_3	×	×	×	×	×	×	×	×	0	0	1	0
b_3	×	×	×	×	×	×	×	×	1	0	0	0
c_3	×	×	×	×	×	×	×	×	0	1	0	1
d_3	×	×	×	×	×	×	×	×	1	1	0	0

TERGM Background

Methodological Literature

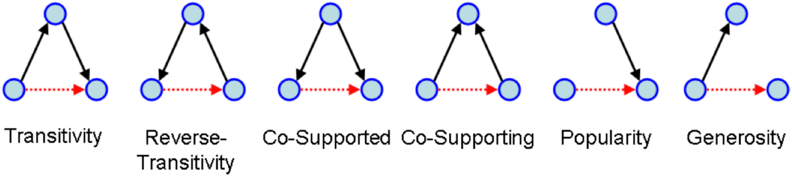
- Robins, Garry and Philippa Pattison. 2001. "Random Graph Models for Temporal Processes in Social Networks." *Journal of Mathematical Sociology*, 25(1):5 – 41.
- Hanneke, Steve, Wenjie Fu and Eric P. Xing. 2010. "Discrete Temporal Models of Social Networks." *The Electronic Journal of Statistics*, 4:585–605.
- Desmarais, Bruce A. and Skyler J. Cranmer. 2012. "Statistical Mechanics of Networks: Estimation and Uncertainty." *Physica A*, 391(4):1865–1876.
- Krivitsky, Pavel N. and Mark S. Handcock. Working Paper. "A Separable Model for Dynamic Networks." *arXiv:1011.1937v1.*

Separability and Approximation

- Hanneke, Fu and Xing (2010) analyze separable TERGMs
- $p(N^t|\boldsymbol{\theta}, \{N^t, N^{t-1}, \ldots, N^{t-k}\}) =$

$$\prod_{i=1}^{N}\prod_{j=1}^{N} p(N_{ij}^t|\boldsymbol{\theta}, \{N^t, N^{t-1}, \ldots, N^{t-k}\})$$

- Contemporaneous ties are independent given the past.
- Likelihood can be maximized using logit software
- Proven to avoid degeneracy
- Formulated as a subgraph completion process

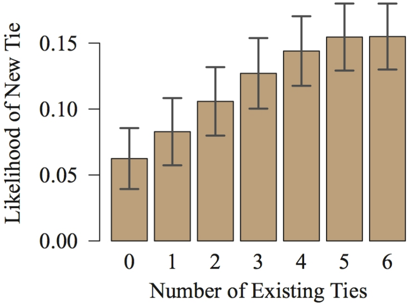


Core Considerations in Specification

- Common functional form questions
- What effects function at each lag?
- What is k?
- Is there temporal heterogeneity in $\boldsymbol{\theta}$?
- Are edge creation and loss separate processes?

Functional Form Example: Popularity

Domain: Information-seeking among local estuary management organizations in the US



Relevant Specification/Estimates:

$$0.588 \times (In - Two - Stars) - 0.097 \times (In - Three - Stars)$$

Source: Bruce A. Desmarais and Skyler J. Cranmer. "Micro-level interpretation of exponential random graph models with application to estuary networks." *Policy Studies Journal*, 40(3):402-434, 2012.

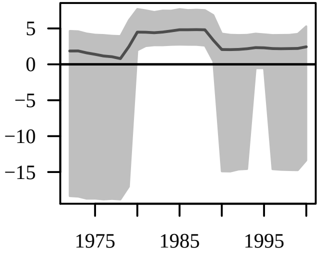
Common Process, Varying Lag Effects

Multiple-lags can illuminate complex dynamic effects

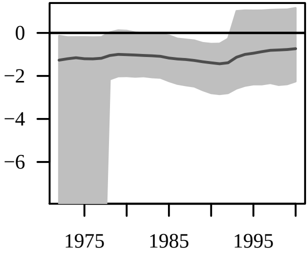
- Edges form in response to developments in the network
- Multiple lags can represent different stages in response

International Sanctions and Retaliation

Reciprocity at t



Reciprocity at $t-1$



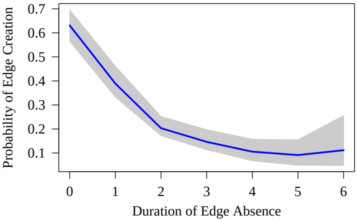
Source: Skyler J. Cranmer, Tobias Heinrich, and Bruce A. Desmarais. Reciprocity and the Structural Determinants of the International Sanctions Network. *Social Networks*, 36(January):5-22, 2014.

Different Processes of Edge Creation and Loss

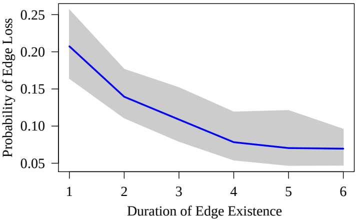
In many networks, the likelihood of a tie between i and j at time t depends upon the history of the relationship between i and j in a *long* fashion.

Example: Legislative Collaboration in the US Senate

Edge Creation



Edge Loss



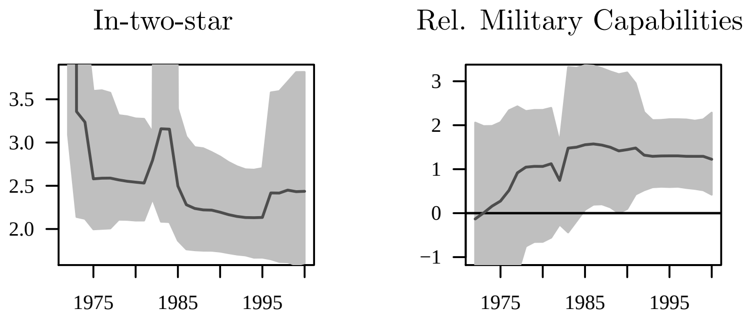
Source: Bruce A. Desmarais and Skyler J. Cranmer. "Statistical Mechanics of Networks: Estimation and Uncertainty." *Physica A*, 391(4):18651876, 2012.

Temporal Homogeneity: Do Effects Change Over Time?

Possibilities Include

1. Trends
2. Cycling
3. White Noise in Effects
4. Distinct structural shifts (change points)

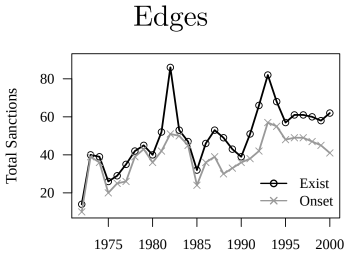
International Sanctions: Parameter Heterogeneity

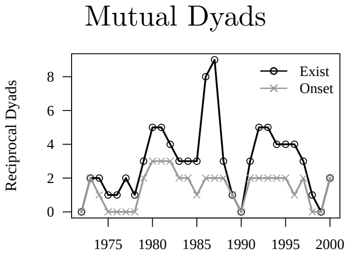


Exploring Temporal Heterogeneity: Network Statistics

- ERGM has a *Canonical Exponential Family Form*
- Network statistics (i.e., $\boldsymbol{h}(N)$) are *sufficient statistics*
- This means $\boldsymbol{h}(N)$ tells us all we can learn about $\boldsymbol{\theta}$
- Substantial change in the statistics implies change in the ERG, change in $\boldsymbol{\theta}$.

International Sanctions: Sufficient Statistics Over Time





Estimation: Numerical Challenges

The big normalizing constant creates a big headache...

Nodes	m (i.e., unique undirected networks)
5	1,024
10	35,184,370,000,000
15	40,564,820,000,000,000,000,000,000,000,000

- The log-likelihood contribution of a single time point is

$$ll(\boldsymbol{\theta}) = \ln\left[\frac{\omega(N)}{\sum_{i=1}^{m}\omega(N^i)}\right]$$

- We don't have time to compute $\sum_{i=1}^{m}\omega(N^i)$

Estimation: Conditional MLE

Simulation-based approximation

- $\sum_{i=1}^{m} \omega(N^i)$ is a sum over a population of m networks.
- Use a random sample from the population to approximate.
- **Markov Chain Monte Carlo Maximum Likelihood**
  - In MLE, replace the exact $ll(\boldsymbol{\theta})$ with an approximation computed on a *sufficiently large* sample of networks.
  - Use importance sampling (i.e., the MCMC) to focus on the most likely networks.

Conditional MCMC-MLE

- Each time point conditioned on the previous k
- Time points enter *separably* into the log-likelihood

$$\hat{ll}(\boldsymbol{\theta}) = \sum_{t=k+1}^{T} \hat{ll}^{t}(\boldsymbol{\theta})$$

Degeneracy

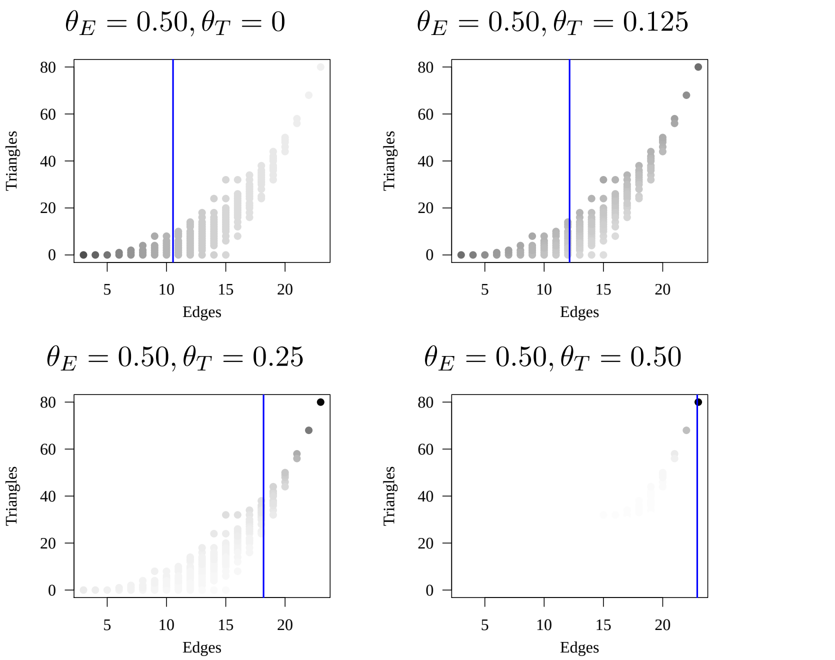
Complexity and Flexibility: Combinatorial properties of high order dependence functions induce unintended consequences.

- Each edge can be involved in $n - 2$ triangles

Degeneracy: Most probability mass concentrated on a few networks, most commonly, the **completely full** or **completely empty** network.

Avoid Degenerate Models!! They constitute completely unrealistic characterizations of the data generating process.

5 node directed net with the number of **Edges** and **Triangles**



Dealing with Degeneracy

Common Solution: use statistics that down-weight repeated structures that involve the same edge.

Example: Transitivity

- Classic measure: Number of triangles in the network

$$\sum_{ijk} N_{ij} N_{ik} N_{jk}$$

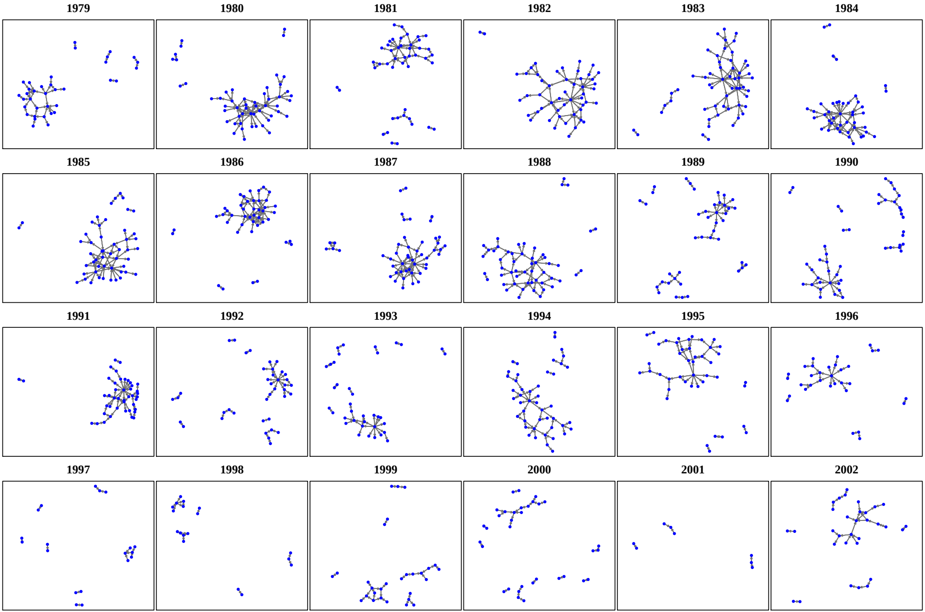
- Prone to degeneracy. There are probably decreasing marginal returns to indirect connections
  - I am probably not twice as likely to befriend the friend of two of my friends as I am the friend of one of my friends.
- Geometrically Weighted Edgewise Shared Partners

$$\sum_{i=1}^{n-2} \left[1 - \left(1 - e^{-\phi}\right)^{i}\right] EP_i(N)$$

Example: Transnational Terrorism Networks 1979–2002

- **Vertices:** states in the international system
- **Edges:** i to j if terrorist from i executed attack in j
- **Time Interval:** One year
- **Data Source:** The *ITERATE* database

Inferential Objective: Develop TERGM to *forecast* the geographic flow of transnational terrorism.


1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002

TErroRGM: Specification

Model based on Purely Structural Features

Pros:

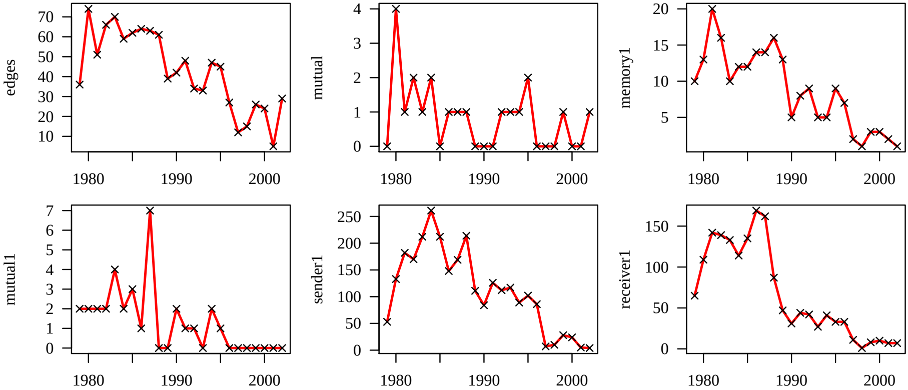
- Minimizes Data collection costs
- Permits forecasting without imputation

Cons:

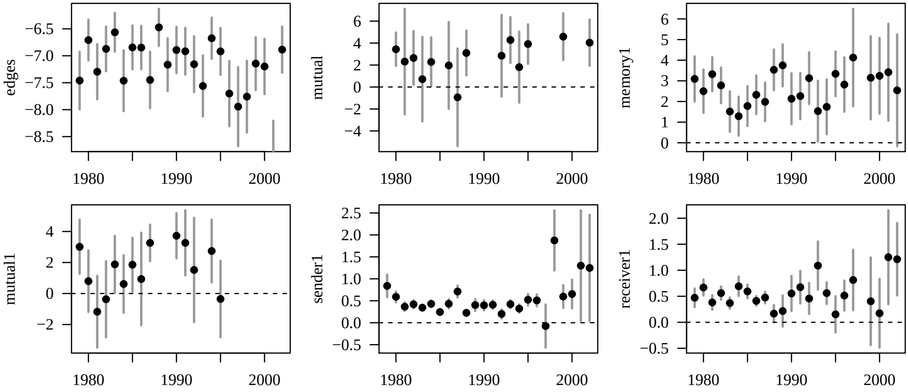
- Probably sacrifices precision and recall
- Parameter estimates subject to omitted variable bias

Concept	**Statistic**
Density	Number of Edges at time t
Memory	Number of Persistent Ties from $t-1$ to t
Reciprocity	Number of mutual dyads at t
Delayed Reciprocity	Num. Reciprocations from $t-1$ to t
Source Activity	Out degree t times out degree $t-1$
Targeting	In degree t times in degree $t-1$

Temporal Homogeneity?



Year-by-Year Estimates



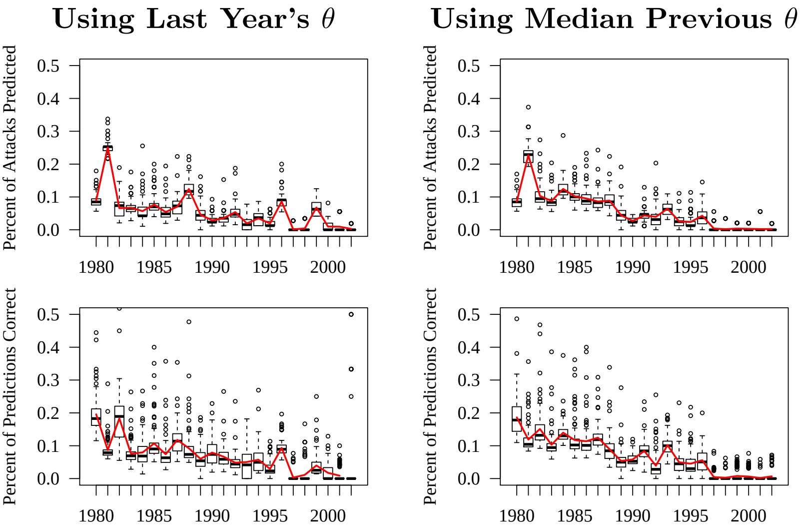
Forecasting with TERGM

Uses/Advantages of Forecasting:

- Out-of-sample validation of model complexity
- Comparison of different models
- Application/societal relevance

Requirements/process

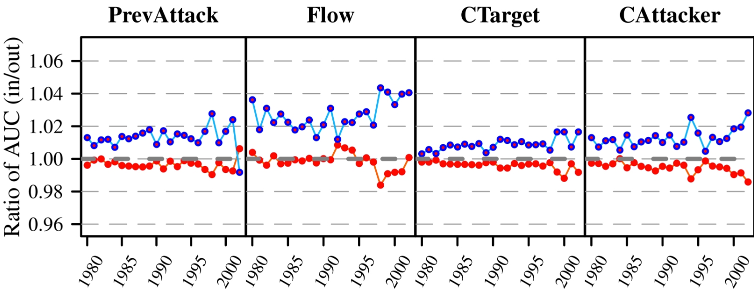
- Use parameters estimated without data from t
- Use information available prior to t
- Design prediction methodology



Using Last Year's θ
Using Median Previous θ
Percent of Attacks Predicted
Percent of Predictions Correct
0.0
0.1
0.2
0.3
0.4
0.5
1980
1985
1990
1995
2000

Example: Predictive Feature Evaluation

Dynamic networks/TERGM offer a simple way to evaluate model features using held-out validation.



- Above 1 supports the feature
- Blue = 5 year interval, Red = 1 year interval

Source: Bruce A. Desmarais and Skyler J. Cranmer. "Forecasting the Locational Dynamics of Transnational Terrorism: A Network Analytic Approach." In *Proceedings of the European Intelligence and Security Informatics Conference (EISIC) 2011*. IEEE Computer Society, 2011.

(T)ERGM Takeaways

1. Offer Complex and Realistic Network Models
   1.1 Extremely Flexible
   1.2 MLE Challenging/Degeneracy Prone
2. TERGMs differ from ERGMs in two ways
   2.1 Inter-temporal network statistics
   2.2 Block-diagonal network structure
3. TERGMs present myriad specification choices
   3.1 Complex inter-temporal dependence
   3.2 Time-varying effects
   3.3 Lag-varying effects
   3.4 Creation vs. Duration
4. TERGMs can be used for forecasting