The Temporal Exponential Random Graph Model

Skyler J. Cranmer
Carter Phillips and Sue Henry Professor
The Ohio State University

The Temporal ERGM (TERGM)

- Extension of ERGM to accommodate inter-temporal dependence
- Accomplished by adding functions involving past realizations of N to $h(\cdot)$
- Assume the network is observed in T discrete time periods
- N^t is the network at some discrete period of observation
- Time dependencies can be built into the model by conditioning on previous realizations of the network
- Instead of $\mathcal{P}(N, \boldsymbol{\theta})$ as the probability of interest, estimate a model with K-order dependencies: $\mathcal{P}(N^t | N^{t-K}, \ldots, N^{t-1})$
- The realization of N at time t is conditioned on the previous K realizations
- Assumes that if the time-dependent model is well specified, the N^t will be independent of networks more than K periods removed

The TERGM

- The probability of observing N^t in the TERGM of order K is written as

$$\mathcal{P}(N^t|N^{t-K},\dots,N^{t-1},\boldsymbol{\theta}) = \frac{1}{c(\boldsymbol{\theta},N^{t-K},\dots,N^{t-1})}\exp\{\boldsymbol{\theta}'\mathbf{h}(N^t,N^{t-1},\dots,N^{t-K})\}$$

- The joint probability of observing the networks between times $T-K$ and T is then established by taking the product of the probabilities of the individual networks given the others:

$$\mathcal{P}(N^{K+1},N^{K+2},\dots,N^T|N^1,\dots,N^K,\boldsymbol{\theta}) = \prod_{t=K+1}^{T}\mathcal{P}(N^t|N^{t-K},\dots,N^{t-1},\boldsymbol{\theta})$$

- Thus, we can account for arbitrary order time dependence in a single network or over an arbitrary number of networks

Example: First-Order Autocorrelation with TERGM

- The best predictor of a given network may be its last realization
- Simplest special case of the general model
- N^t is dependent only on N^{t-1}
- The general formula for a first-order model would then be written

$$\mathcal{P}(N^t|N^{t-1}, \boldsymbol{\theta}) = \frac{1}{c(\boldsymbol{\theta}, N^{t-1})} \exp\{\boldsymbol{\theta}'\mathbf{h}(N^t, N^{t-1})\}.$$

Example: First-Order Autocorrelation with TERGM

- To see the full specification, suppose we included a measure of the change in edge values from one time to the next – accounting for memory in the edges

$$\mathbf{h}_a(N^t, N^{t-1}) = \sum_{i \neq j} N_{ij}^t N_{ij}^{t-1},$$

- And a structural parameter for single-period-delayed reciprocation of ties

$$\mathbf{h}_r(N^t, N^{t-1}) = \sum_{ji} N_{ij}^t N_{ji}^{t-1}.$$

- Then the first order model would be

$$\mathcal{P}(N^t | N^{t-1}, \boldsymbol{\theta}) = \frac{1}{c(\boldsymbol{\theta}, N^{t-1})} \exp\left\{ \sum_{i \in \{a,r\}} \boldsymbol{\theta}_i' \mathbf{h}_i(N^t, N^{t-1}) \right\}.$$

"Memory" Terms

- Covariates are included in the ERGM equation as

$$h_x = \sum N_{ij}^t X_{ij}^t, \tag{1}$$

- Can equivalently include them as change statistics in

$$P(N_{ij} = 1 | N_{-ij}, \boldsymbol{\theta}) = \text{logit}^{-1}\left(\sum_{r=1}^{k} \theta_r \delta_r^{(ij)}(N)\right) \tag{2}$$

  - N_{-ij} indicates the network excluding N_{ij}
  - $\delta_r^{(ij)}(N)$ is equal to the change in h_r when N_{ij} is changed from zero to one
  - logit^{-1} is the inverse logistic function such that $\text{logit}^{-1}(x) = 1/(1 + \exp(-x))$
- $\delta_r^{(ij)}(N)$ from a thought experiment of $h_x(N_{ij}^-) - h_x(N_{ij}^+)$

Positive autoregression (AKA a lagged outcome network)

- The term is specified as a lagged network included as an edge covariate

$$h_a = \sum N_{ij}^t N_{ij}^{t-1} \tag{3}$$

- $\delta^{(ij)} = 1$ when $N_{ij}^{t-1} = 1$ and $\delta^{(ij)} = 0$ otherwise
- In other words, imagine $h_x(N_{ij}^-) - h_x(N_{ij}^+)$: the change will be 0 when $N_{ij}^{t-1} = 0$ and 1 when $N_{ij}^{t-1} = 1$.
- So, we can also see that directly coding the change statistic as X captures the process of interest.
- Substantive interpretation: it counts the number of edges that persist from $t-1$ to t.

Stability

- The term is specified as
$$h_s = \sum_{ij} (N_{ij}^{t-1} == N_{ij}^{t}), \tag{4}$$
which should be 1 when true. So, better to write
$$h_s = \sum_{ij} N_{ij}^{t} N_{ij}^{t-1} + (1 - N_{ij}^{t})(1 - N_{ij}^{t-1}), \tag{5}$$
- Reflect on $h_x(N_{ij}^{-}) - h_x(N_{ij}^{+})$
  - The statistic will lose 1 whenever $N_{ij}^{t-1} = 0$
  - It will gain 1 whenever $N_{ij}^{t-1} = 1$
  - So, can code $X_{ij}^{t} = -1$ when $N_{ij}^{t-1} = 0$ and $X_{ij}^{t} = 1$ when $N_{ij}^{t-1} = 1$
- Substantive interpretation: count the stable dyads
- Note: stability should not be included along with positive autocorrelation
- Note: modeling stability in this way is the closest one can get with TERGM to SAOM

Edge Innovation

- Create a covariate X^t coded 1 if $N_{ij}^{t-1} = 0$ and $N_{ij}^{t} = 1$
- Change statistic: $\delta^{(ij)} = 1$ if $N_{ij}^{t-1} = 0$ and 0 otherwise
- Substantive interpretation: count the edges that were created between $t-1$ and t
- Note: one cannot do the inverse of this: modeling edge loss directly
- To see this, consider a covariate X^t coded 1 if $N_{ij}^{t-1} = 1$ and $N_{ij}^{t} = 0$. Including this X^t in equation (1) assures that the only time $X_{ij}^{t} = 1$ is when $N_{ij}^{t} = 0$. So, no value is ever added to the statistic.
- However, it is the case that edge loss is the omitted category given the intercept and edge persistence parameters.

Application: International Conflict

- Replicate and extend Maoz et al. (2006)
- Article is appealing because the authors explicitly argue that international conflict should be treated as a network
- Original application considers all dyads in the state system from 1870 – 1996
- DV: MID =1, No MID = 0
- Maoz et al. (2006) argue that structural equivalence, which measures the similarity of ties held by nodes in a number of important international networks, is a measure of international affinity
- We use Maoz et al.'s replication data and specify the same theoretical model they did:

$$\begin{aligned}\text{MIDs} = &\ \text{dyad's weak-link regime score}\\ &+ \text{military-industrial capabilities ratio}\\ &+ \text{distance between capitols}\\ &+ \text{integrated structural equivalence score}\end{aligned}$$

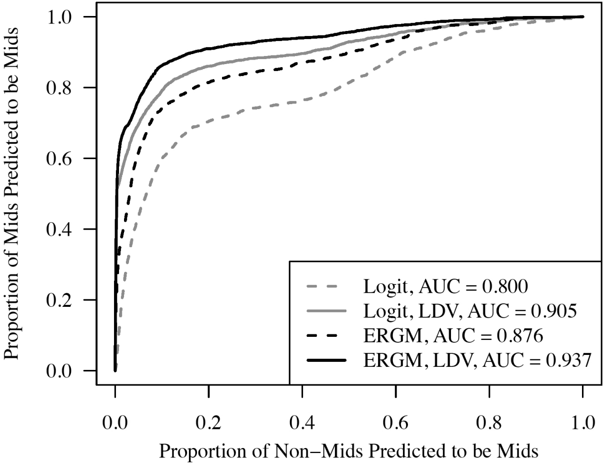
Application: International Conflict

- We expand the Maoz et al. (2006) model to include two likely structural characteristics of the conflict network
- An effect akin to the "popularity" of a state in the conflict network
- We capture this with a two-star statistic
  (the number of times in the network where two states are at war with the same state)
- Also contend that triangles are especially unlikely to be present in the conflict network
- Would not make sense from a strategic perspective

Application: International Conflict

	Logit	Logit, LDV	ERGM	ERGM, LDV
Edges	-3.14	-4.42	-3.61	-4.6
	[-3.43, -2.87]	[-4.68, -4.17]	[-3.79, -3.44]	[-4.79, -4.41]
Min Reg Score	-0.003	-0.002	-0.001	-0.001
	[-0.006, -0.00043]	[-0.0054, -0.00047]	[-0.0035, 0.00010]	[-0.0035, 0.00031]
Capability Ratio	0.00029	0.00027	0.00011	0.00021
	[0.0001, 0.0004]	[0.0001, 0.00039]	[-0.00010, 0.00028]	[0.00031, 0.00034]
Distance	-0.0005	-0.0003	-0.0005	-0.0003
	[-0.0006, -0.0004]	[-0.0004, -0.0002]	[-0.0006, -0.0005]	[-0.0004, -0.0003]
Integrated SEq	-0.867	-0.605	-0.511	-0.344
	[-1.09, -0.645]	[-0.822, -0.39]	[-0.682, -0.352]	[-0.544, -0.171]
Lagged MID	–	5.13	–	4.67
	–	[4.88, 5.35]	–	[4.48, 4.86]
Two Stars	–	–	0.335	0.272
	–	–	[0.302, 0.363]	[0.237, 0.308]
Triads	–	–	-0.583	-0.482
	–	–	[-0.743, -0.426]	[-0.707, -0.279]

Application: International Conflict



Easy Software Implementation

The `btergm` package on CRAN.

Small Example: Golden Age of Hollywood

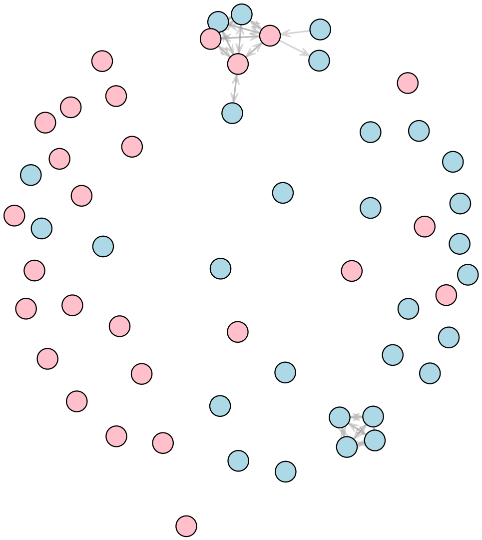
Application from INA book.
Data referece: "Eigenvector-Based Centrality Measures for Temporal Networks" Dane Taylor, Sean A. Myers, Aaron Clauset, Mason A. Porter, Peter J. Mucha. Preprint, arXiv:1507.01266 (2015)

```
attributes <- read.table("HollywoodGoldenAge_names.txt", header=TRUE)

t1 <- as.matrix(read.table("HollywoodGoldenAge_matrix_1920-1929_s2.txt", header=FALSE)) # read in the adjacency matrix
t1[t1 > 0] <- 1 # replace edge weights with binary values
n1 <- network(t1) # create object of class "network"
network.vertex.names(n1) <- as.character(attributes$name) # add vertex names
set.vertex.attribute(n1, "female", attributes$female) # add sex attribute of vertices

t2 <- as.matrix(read.table("HollywoodGoldenAge_matrix_1930-1939_s2.txt", header=FALSE))
t2[t2 > 0] <- 1
n2 <- network(t2)
network.vertex.names(n2) <- as.character(attributes$name)
set.vertex.attribute(n2, "female", attributes$female)

t3 <- as.matrix(read.table("HollywoodGoldenAge_matrix_1940-1949_s2.txt", header=FALSE))
t3[t3 > 0] <- 1
```

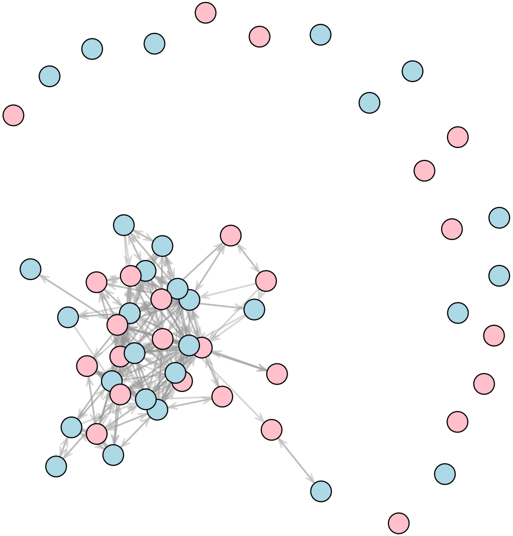
```
t4 <- as.matrix(read.table("HollywoodGoldenAge_matrix_1950-1959_s2
    .txt", header=FALSE))
t4[t4 > 0] <- 1
n4 <- network(t4)
network.vertex.names(n4) <- as.character(attributes$name)
set.vertex.attribute(n4, "female", attributes$female)
t5 <- as.matrix(read.table("HollywoodGoldenAge_matrix_1960-1969_s2
    .txt", header=FALSE))
t5[t5 > 0] <- 1
n5 <- network(t5)
network.vertex.names(n5) <- as.character(attributes$name)
set.vertex.attribute(n5, "female", attributes$female)
t6 <- as.matrix(read.table("HollywoodGoldenAge_matrix_1970-1979_s2
    .txt", header=FALSE))
t6[t6 > 0] <- 1
n6 <- network(t6)
network.vertex.names(n6) <- as.character(attributes$name)
set.vertex.attribute(n6, "female", attributes$female)
hga <- list(n1, n2, n3, n4, n5, n6) # create list from oldest to
    newest. This will be the outcome object for TERGM analysis
```

```
## Plot the network
set.seed(5)
par(mfrow = c(3,2))
hgat <- c("20's", "30's", "40's", "50's", "60's", "70's")
for (i in 1:length(hga)){
  plot(hga[[i]],displaylabels=F,label=network.vertex.names(n1),
      vertex.cex=2,label.cex=1,edge.col=rgb(150,150,150,100,
      maxColorValue=255),label.pos=5,vertex.col=c("lightblue", "pink")[get.vertex.attribute(n1,"female")+1], main=hgat[[i]]
      )
}
```

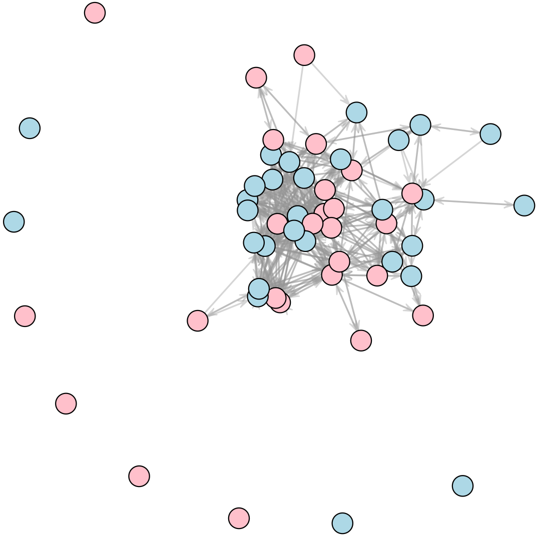
20's



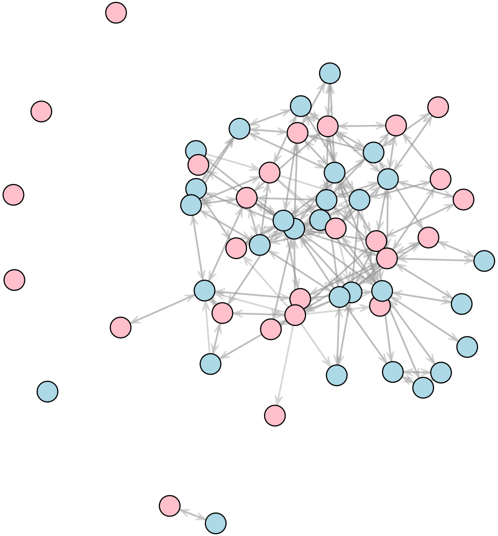
30's



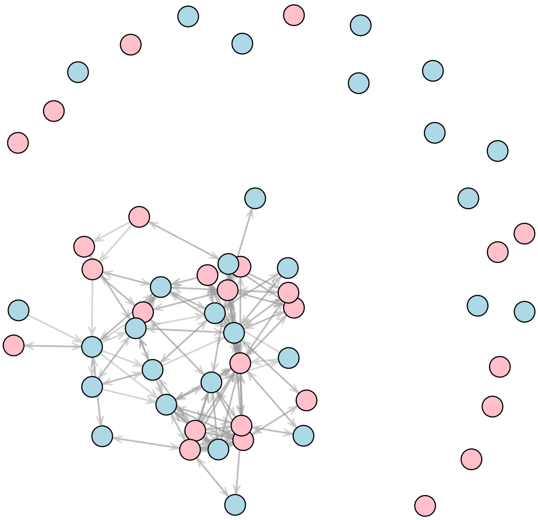
40's



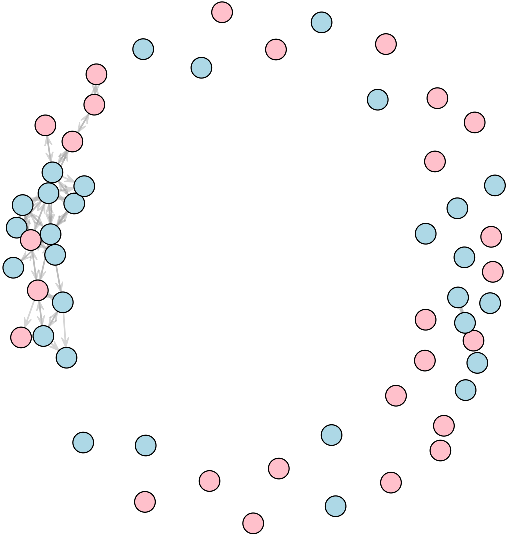
50's



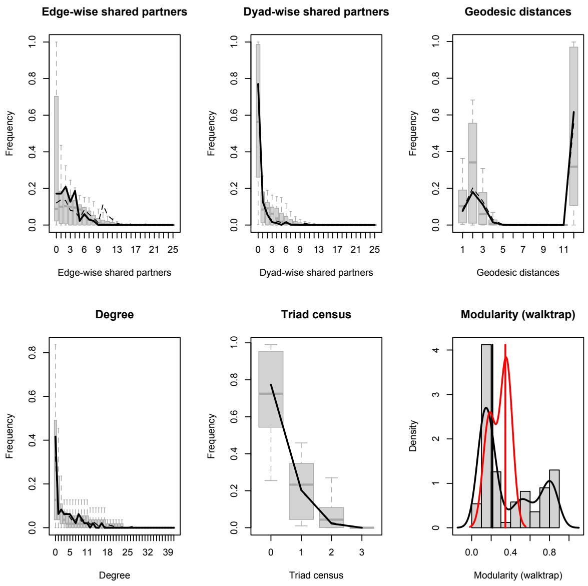
60's



70's



```r
library(bergm)

# Basic model
set.seed(12345)
m1 <- btergm(hga ~ edges
             + mutual
             + gwesp(0.5, fixed=TRUE)
             + idegreepopularity
             + absdiff("female")
             + nodefactor("female")
             + delrecip
             + memory(type="stability")
             , R=100)
summary(m1)
gof1 <- gof(m1, statistics = c(esp, dsp, geodesic,deg, triad.
    undirected, walktrap.modularity))
plot(gof1)
```

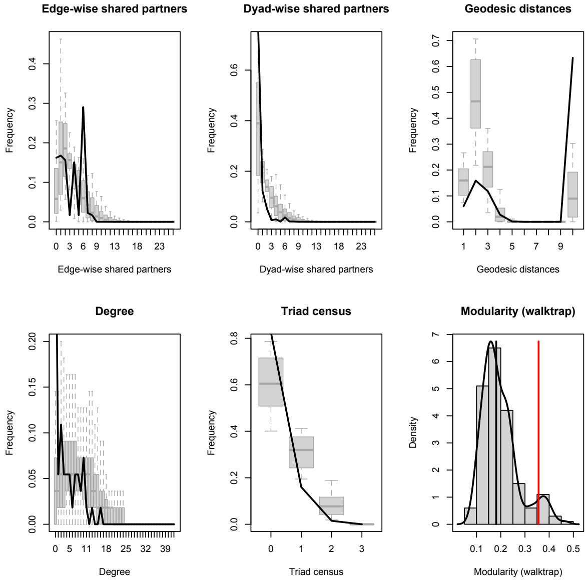

Edge-wise shared partners
Frequency
Edge-wise shared partners
Dyad-wise shared partners
Frequency
Dyad-wise shared partners
Geodesic distances
Frequency
Geodesic distances
Degree
Frequency
Degree
Triad census
Frequency
Triad census
Modularity (walktrap)
Density
Modularity (walktrap)

```
Estimates and 95% confidence intervals:
                      Estimate Boot mean    2.5%   97.5%
edges                -6.423693 -6.429535 -6.8580 -6.0120
mutual                6.007808  6.015379  5.1297  6.8191
gwesp.fixed.0.5       0.617819  0.668626  0.4235  1.0193
idegreepopularity     0.299247  0.284951  0.1581  0.3601
absdiff.female        0.120250  0.102582 -0.0944  0.2266
nodefactor.female.1  -0.056032 -0.056389 -0.1313  0.0348
edgecov.delrecip[[i]] 0.671244  0.657701  0.0194  1.3773
edgecov.memory[[i]]  -0.201031 -0.197819 -0.5495  0.0942
```

```
# Same model estimated by MCMC
set.seed(5)
m1.5 <- m1 <- btergm(hga ~ edges + mutual + gwesp(0.5, fixed=TRUE)
     + idegreepopularity + absdiff("female") + nodefactor("female")
     + delrecip + memory(type="stability"))
summary(m1.5)
```

```
Estimates and 95% confidence intervals:
                        Estimate Boot mean     2.5%   97.5%
edges                  -6.423693 -6.429535 -6.8580 -6.0120
mutual                  6.007808  6.015379  5.1297  6.8191
gwesp.fixed.0.5         0.617819  0.668626  0.4235  1.0193
idegreepopularity       0.299247  0.284951  0.1581  0.3601
absdiff.female          0.120250  0.102582 -0.0944  0.2266
nodefactor.female.1    -0.056032 -0.056389 -0.1313  0.0348
edgecov.delrecip[[i]]   0.671244  0.657701  0.0194  1.3773
edgecov.memory[[i]]    -0.201031 -0.197819 -0.5495  0.0942
```

```r
# Out of sample prediction
set.seed(12345)
m1oos <- btergm(hga[2:4] ~ edges + mutual + gwesp(0.5, fixed=TRUE)
    + idegreepopularity + absdiff("female") + nodefactor("female")
    + delrecip + memory(type="stability"), R=100)
gof1oos <- gof(m1oos, target=hga[[5]], statistics = c(esp, dsp,
   geodesic,deg, triad.undirected, walktrap.modularity))
plot(gof1oos)
```


Edge-wise shared partners
Frequency
0.0 0.1 0.2 0.3 0.4
0 3 6 9 13 18 23
Edge-wise shared partners
Dyad-wise shared partners
Frequency
0.0 0.2 0.4 0.6
0 3 6 9 13 18 23
Dyad-wise shared partners
Geodesic distances
Frequency
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
1 3 5 7 9
Geodesic distances
Degree
Frequency
0.00 0.05 0.10 0.15 0.20
0 5 11 18 25 32 39
Degree
Triad census
Frequency
0.0 0.2 0.4 0.6 0.8
0 1 2 3
Triad census
Modularity (walktrap)
Density
0 1 2 3 4 5 6 7
0.1 0.2 0.3 0.4 0.5
Modularity (walktrap)

```
> interpret(m1, type="dyad", i=1, j=2, t=3)
$`t = 3`
            j->i = 0    j->i = 1
i->j = 0 0.90922441 0.009997232
i->j = 1 0.01476589 0.066012474



> dyads <- edgeprob(m1) # also errors
> checkdegeneracy(m1, nsim=1000) # produces error "object 'mat'
    not found"
```

BONUS Forecasting with ERGMs

B.A. Desmarais and S.J. Cranmer. 2011. "Forecasting the Locational Dynamics of Transnational Terrorism: A Network Analytic Approach" *Proceedings of the European Intelligence and Security Informatics Conference, IEEE Computer Society.*

S.J. Cranmer and B.A. Desmarais. "What Can We Learn from Predictive Modeling?" *Political Analysis* 25.2 (2017), 145–166.

The Place of Prediction

Explanation and prediction are distinct but complementary

- Prediction often seen as a tool of "applied science"
- The failure to predict is not a policy problem:
  It is a symptom of a failure to adequately explain
- e.g. models of international conflict predict notoriously poorly

Our aim: develop the best predicting model(s) we can and use it to improve our explanatory models

The Difference Between Explanatory and Predictive Modeling

Explanatory modeling:

- Statistical model is careful operationalization of theory
- Coefficients are the objects of interest
- Therefore, the statistical model itself is the object of interest

Predictive modeling:

- Objects of interest are the variables
- Statistical model chosen to produce best predictions of y
- Model does not necessarily correspond to any theory

Further Differences

Prediction does not aim to capture causality

Predict new observations rather than those already observed

Predictive models minimize bias *and* estimation variance
(the result of using a sample)

The Explanatory Utility of Predictive Models

Prediction can help refine/improve explanation:

- Uncover previously unknown patterns / causal mechanisms
- Refine operationalizations of theoretical concepts
- Compare competing theories of the same outcome
- Examine how well our explanatory theories capture the phenomena of interest
- Set benchmarks for improvements of existing theory

A Predictive “Gold Standard”

“Good” predictive benchmarks must meet the following criteria:

- Use only the outcome variable
- Out-of-sample prediction only
- Criteria for judging predictive accuracy must be appropriate for the rarity of the event being predicted

Application 1: predicting transnational terrorism

The Problem

- Forecasting transnational terrorist attacks important for security policy
- Problem: most forecasting techniques need substantial data before they can work
  - We often/usually do not have lots of data
  - If we have lots of data, we already know where to send resources
- Related Problem: need to be able to identify *new* sources of threat before they materialize
  - The "who will attack us who has not done so before" problem

Background 1: Predictors of Terrorism

Much of the empirical research on terrorism focuses on identifying covariates that *explain* the within-country frequency of terrorism.

Terrorism is more likely in

- Countries with **democratic governments** (Li 2005)
- Countries that **grant concessions to terrorists** (Kydd and Walter 2006)
- Countries with **further economic reach** (Li and Schaub 2004)

Problems

- Research is not explicitly focused on forecasting
- Covariates can be costly and time consuming to collect
- No information on the source of terrorism

Background 2: Forecasting Frequency

The line of research that explicitly addresses forecasting focuses on predicting the number of events in a country given the recent history of events.

- Frequency of terrorist attacks exhibit cycling (Enders, Parise and Sandler 1992)
- Extreme increases in attack frequency are unsustainable (Enders and Sandler 2000)
- Series can be accurately forecast using Poisson Changepoint (Brandt and Sandler 2010)

Problems

- No information on the source of terrorism
- Without recent series of terrorism, forecasts fail

Our Research Objectives

1. Develop an approach to forecasting the **network** of transnational terrorist attacks in order to **forecast source information**.
2. To overcome the **sparsity** in the network and **lack of innovation** in country-specific time series, we seek to leverage indirect ties to improve prediction.
3. Methodologically, we **improve upon extant proximity-based prediction algorithms**

Our Approach

- A network analytic perspective to foreasting
- The structure of the network can predict the its evolution
- We integrate
  - A deterministic, similarity-based, edge prediction framework (Liben-Nowell and Kleinberg, 2003)
  - A model-based probabilistic approach (Hanneke, Fu, and Xing, 2010)
- Result: a likelihood-based forecasting model that can predict new edge formation
- Predict by substituting the network structure into the model that best predicted the network up to time $t - 1$
- True out-of-sample prediction: the predictive models for t are not based on the data from t

Proximity-Based Prediction

- Foundational work: Liben-Nowell and Kleinberg (2003)
- Insight: vertices similar to one another are likely to link
- Each measure of proximity results in a score $\delta(i, j)$ for each dyad of vertices ij
- Scores computed on a training (past) network
- Subset to potential edges that meet a certain degree threshold
- Dyads ranked w.r.t. proximity scores
- Dyads with high δ are predicted to be edges in the next interval

Liben-Nowell and Kleinberg (2003)

Strengths:

- Prediction is based solely on graph topology – no covariates
- Can experiment with t_0 to optimize prediction

Weaknesses:

- Does not permit precise probability statements about future edges
- Cannot integrate/combine proximity measures

The TERGM

- Developed by Hanneke, Fu, and Xing (2010)
- Estimation routines by Cranmer and Desmarais (2011) and Desmarais and Cranmer (2011)
- Let N^t be the observed network at time t
- Can condition N^t on K previous realizations of the network to account for temporal dependencies

$$\mathcal{P}(N^t|K,\boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}'\,\mathbf{h}(N^t, N^{t-1}, \dots, N^{t-K})\}}{C(\boldsymbol{\theta}, N^{t-K}, \dots, N^{t-1})}.$$

- We estimate a new $\boldsymbol{\theta}$ for each t to account for temporal heterogeneity

Our approach: integrating proximity-based prediction with TERGMs

- We integrate proximity measures into the **h** of the TERGM
- Individual proximity measures are combined into a single model using the estimated weights ($\boldsymbol{\theta}$)
- We can include many proximity measures
- TERGM estimates permit us to forecast the probability of edges in the future
- Each proximity measure δ is integrated into the TERGM by adding

$$h(N^t, N^{t_0,t-1}) = \sum_{ij} N_{ij}^t \delta(i,j)^{t_0,t-1}.$$

- $\delta(i,j)^{t_0,t-1} = 0$ if i or j was not in the system during training interval
- We perform our analyses setting t_0 at $t-1$ and $t-5$

Forecasting

- The forecast model for t is selected as the best performing model up to $t-1$
- Performance is judged based on the predictive log score (i.e., the forecast log-likelihood)
- This gives us, in expectation, the model with the minimum Kullback-Leibler divergence
- We use $\boldsymbol{\theta}^{t-1}$ to perform the forecast of N^t, which was estimated to fit N^{t-1} based on $N^{t_0-1,t-2}$

In other words. . .

The forecasting algorithm we employ is summarized as:

1. Estimate each of the 6,561 forecasting models for each time point from 1980 up to the previous year. Denote the structural measures in model M as $\mathbf{h}_M$. Let $\boldsymbol{\theta}_M^t$ be the parameters estimated on N^t using $\mathbf{h}_M$.
2. Select as the forecasting model (M^*) for time t to be that model maximizing

$$\sum_{i=1981}^{t-1} \ln \left[\mathcal{P}(N^i | \boldsymbol{\theta}_M^{i-1}, \mathbf{h}_M)\right]$$

3. Forecast the next network from the distribution

$$\frac{\exp\{\left[\boldsymbol{\theta}_{M^*}^{t-1}\right]' \mathbf{h}_{M^*}(N, N^{t_0, t-1})\}}{C\left(\boldsymbol{\theta}_{M^*}^{t-1}\right)}$$

4. Draw many forecast networks from the distribution in item 3 and compute the mean edge value in order to estimate the probability of any particular edge

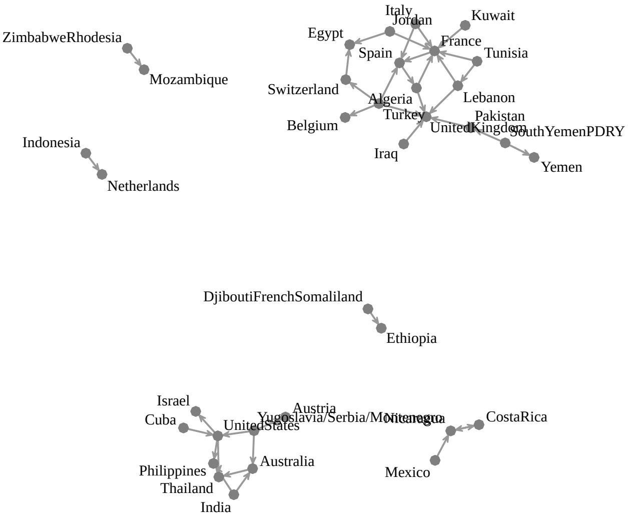
We apply this algorithm to all 23 years of the network under consideration

Data

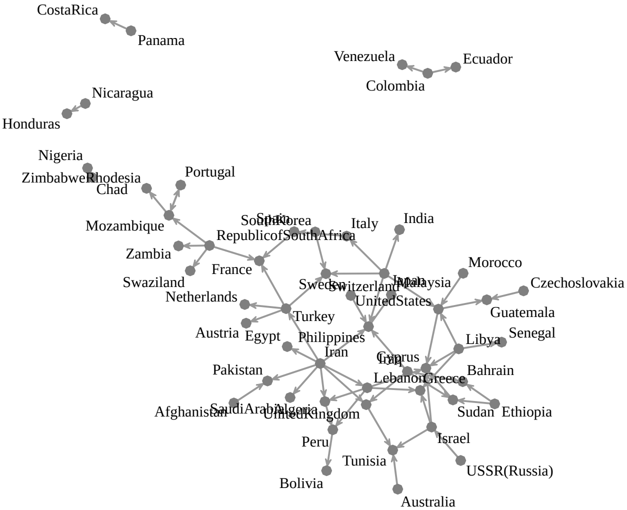
We construct the transnational terror network from events data

- Data from the ITERATE dataset (Mickolus, 2008)
- More than 12,000 transnational terror attacks between 1968 and 2002
- Data codes nationalities of terrorists and the location of their attacks
- We consider more than just states, a few IGOs and contested territories (e.g. the UN and Palestine)
- Edges exist from a terror producer to a terror target
- Networks have a median of 175 vertices over our timespan
- Some self-ties are present
- Some data limitations in terms of where people are from vs where they live vs where they trained

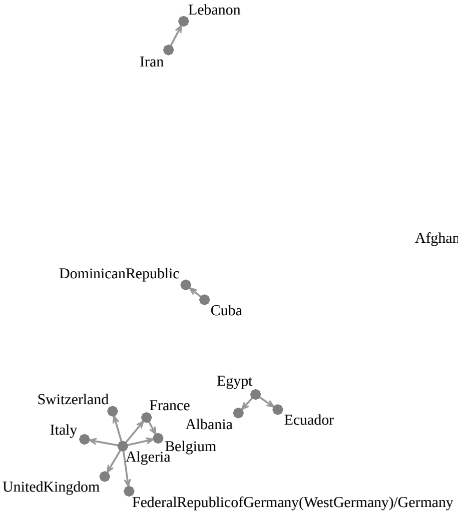
Network: 1978



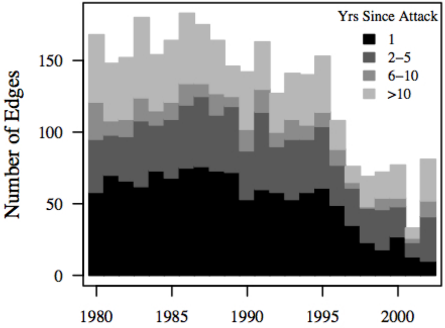
Network: 1988



Network: 1998



Edge Innovation

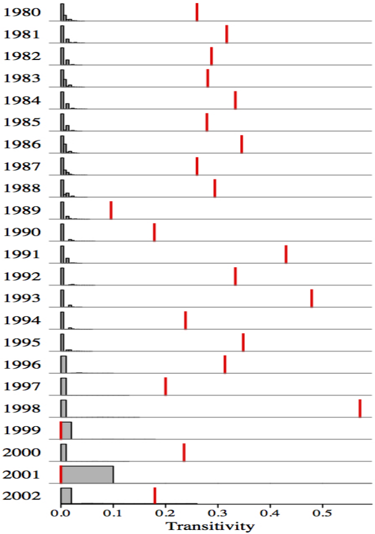


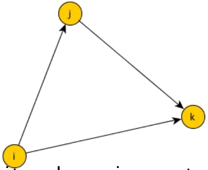
Low degree of edge recurrence

Time series techniques will not work for innovations

Any information that can be leveraged about indirect ties will be valuable

Transitivity





Transitivity: how important is proximity for edge formation

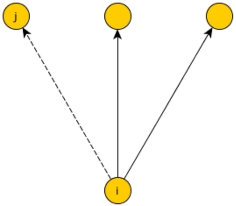
If high transitivity, proximity-based forecasting should be fruitful

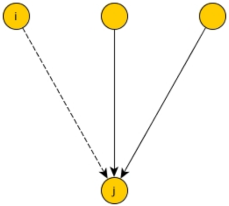
CUG test for transitivity

Transitivity indicates that indirect ties will have predictive power w.r.t. edge formation

Our Proximity Measures: Flow

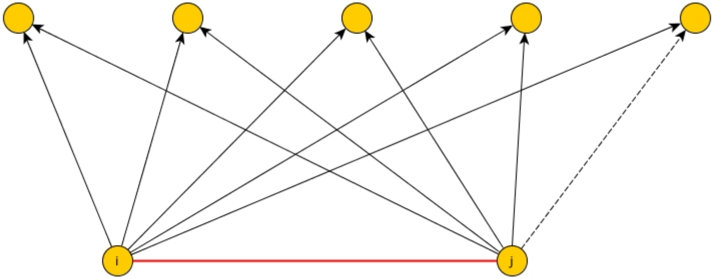
- **Flow**. Generalizes preferential attachment to the directed case
- A process whereby an attack from i to j is likely if i sends many attacks and/or j receives many attacks
- Mathematically: $\delta(i,j) = k_i^o k_j^i$, where k^o and k^i are the out and in-degrees respectively





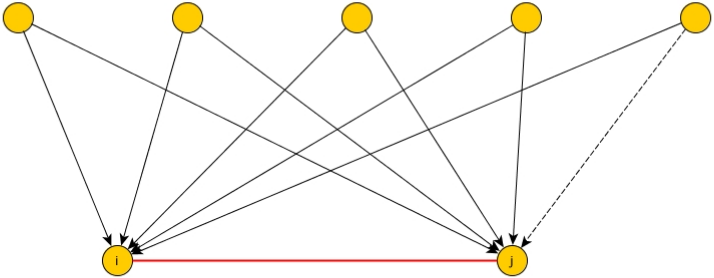
Our Proximity Measures: Common Targets

- **CTarget**. The number of common targets for a dyad
- Mathematically: $\delta(i, j) = \sum_h N_{ih} N_{jh}$.



Our Proximity Measures: Common Attackers

- **CAttacker**. The number of common attackers for a dyad
- Mathematically: $\delta(i,j) = \sum_h N_{hi} N_{hj}$



Our Proximity Measures: Two Similarity Measures

We consider 2 similarity measures

- **JacSim**. The Jaccard similarity between two countries normalizes the measure of common neighbors by the total number of neighbors of the vertices in the dyad

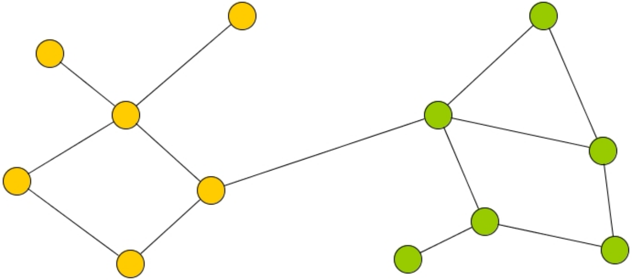
$$\delta(i,j) = [\textbf{CTarget} + \textbf{CAttacker}]/[k_i + k_j]$$

- **AASim**. Adamic/Adar similarity adjusts the measure of common neighbors for the rarity of the neighbors to which the two countries tie

$$\delta(i,j) = \sum_h [\ln(k_h)]^{-1} (N_{ih}N_{jh} + N_{hi}N_{ji})$$

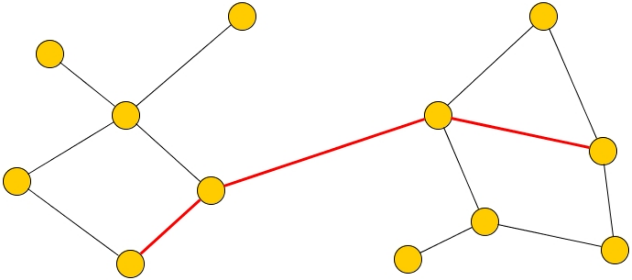
Our Proximity Measures: community membership

- **SameCom**. Common community membership
- We partition the countries into communities using the random walk modularity optimization algorithm "Walktrap" (Pons and Latapy, 2005)
- We then create an indicator, $\delta(i,j) = \mathbf{1}(c_i == c_j)$, of whether i and j are members of the same community.



Our Proximity Measures: Minimum Path Length

- **Distance**. Minimum path length between i and j
- We set $\delta(i, j)$ equal to the number of countries in the network plus one if there is no path from i to j.



Additional Measures

We also include, in each model, the following

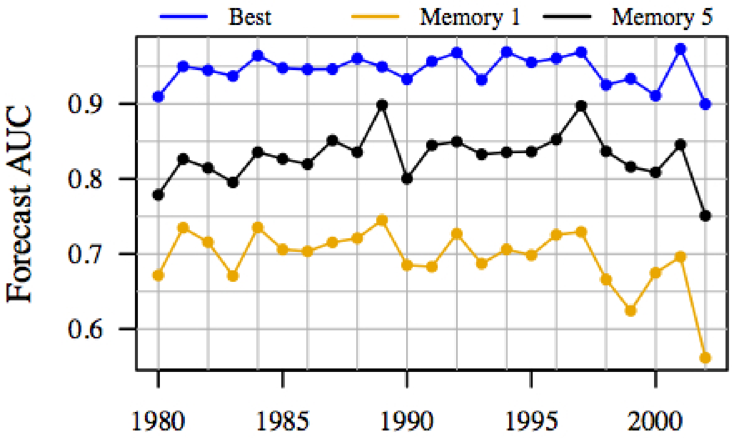
- A count of the number of edges in the network to model the network's density
- A memory term (**PrevAttack**) to capture persistence in the ties between the training network and the current network

Mathematically: memory at time t is $\sum_{ij} N_{ij}^{t} N_{ij}^{t_0,t-1} + (1 - N_{ij}^{t})(1 - N_{ij}^{t_0,t-1})$

Specification

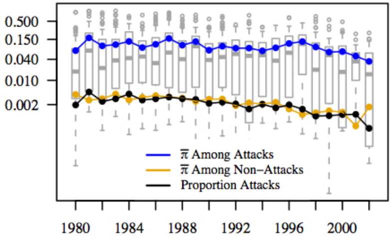
- We try each statistic computed on the networks over the intervals $[t-1, t-1]$ and $[t-5, t-1]$
- Our analysis begins at 1980 and ends at 2002
- The memory term and each of the proximity terms is
  - Included computed on the one year training interval
  - Included computed on the five year training interval
  - excluded from the model.
- This leads to a total of $3^8 = 6,561$ models estimated at each t

Results: Overal Forecasting Performance

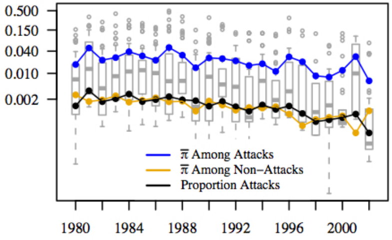


Results: Predicting Edge Innovations

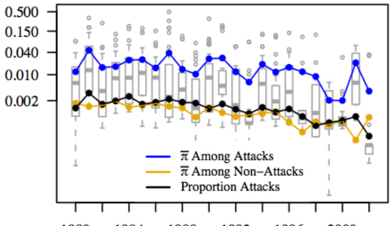
(a) No Attack in the Previous Year



(b) No Attack in the Previous 5 Years

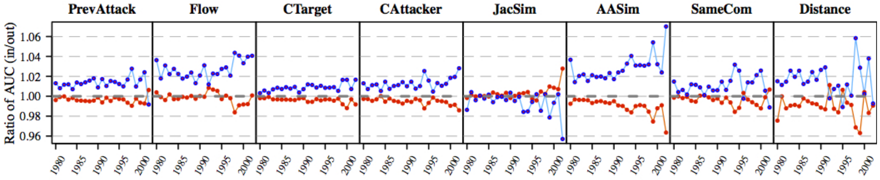


(c) No Attack in the Previous 10 Years



Forecasted probabilities for edges that

Results: Patterns in the Proximity and Memory Features



- Ratio of mean one-year-ahead forecast AUCs with and without the given measure
- A value greater than one indicates that the average forecast AUC is higher when the respective term is included
- The superior performance of the measures computed with five year memories reinforces idea that the transnational terrorism network exhibits long memory

Case Test: The Saudi Link to the U.S. in 2001

Top Ten Forecasted Sources, 2001

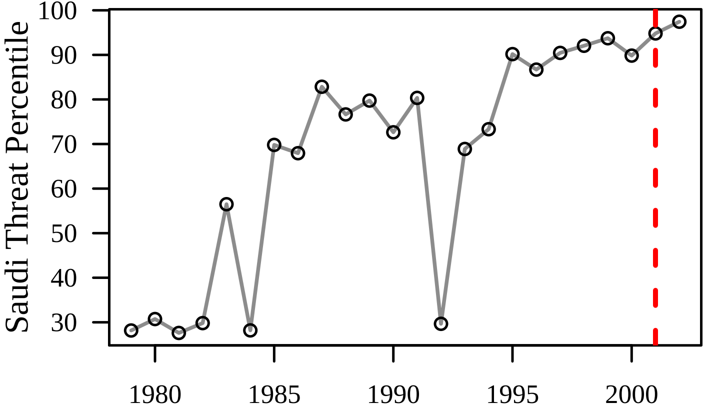
Rank	Country	P(Attack)
1	Algeria	0.126
2	Pakistan	0.055
3	Iraq	0.044
4	Jordan	0.037
5	Cuba	0.037
6	Canada	0.029
7	Romania	0.024
8	**Saudi Arabia**	0.012
9	**Egypt**	0.011
10	Iran	0.011

Overall Probability of Attack on the U.S. 0.41

Predictive Model for 2001

δ	t_0	θ	δ %tile*
PrevAttack	1996	1.64	0
Flow	1996	0.027	99.99
CAttacker	1996	0.24	98.46
AASim	2000	0.5	0
SameComm	2000	0.441	0
Distance	1996	4.07	98.89

*δ %tile is the Percentile rank of $\delta(SA, US)$


Saudi Threat Percentile
100
90
80
70
60
50
40
30
1980
1985
1990
1995
2000

Contributions

- We show that a network-analytic approach can succeed in forecasting transnational terrorism
- We show that indirect ties can be leveraged to forecast innovations in terrorist links
- We advance link forecasting methods by integrating vertex proximity measures into TERGM