The (basic) Exponential Random Graph Model

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

What does it mean to model a network?

Construct a probability distribution that accurately approximates the network

Why build models?

- Test hypotheses
  Example: Does the Krackhardt 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?

Major Advantage of the ERGM

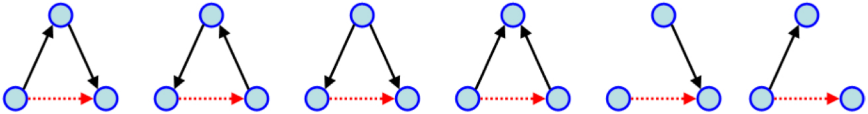
Can model how ties depend upon each other

Modeling Interdependence

Two Classes of Questions:

1. Exogenous (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. Endogenous (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



The Exponential Random Graph Model (ERGM)

Notation:

- Network N is composed of n nodes
- Assume only binary edges between nodes i and j, $N_{ij} = \{0, 1\}$ (relaxable, details in later lectures)
- $\mathcal{N}$ possible permutations of N
- Modeling objective: the probability of observing the network we *did observe*, $\mathcal{P}(N)$, over the networks we *could have* observed
- If there are $\mathcal{N}$ possible networks with the same number of nodes, then it must be $N \in \mathcal{N}$ and we want to know $p(N)$

The Exponential Random Graph Model (ERGM)

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

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

Decomposition:

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

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

Normalizing constant can make estimation difficult

The Exponential Random Graph Model (ERGM)

Conceptual trick:

- ERGM treats the data differently: treat a network as a *single* multivariate observation, rather than a large number of relational observations.
- If the whole network is only one observation, no need for i.i.d. assumptions. The model can now accommodate (almost) arbitrarily complex relationships.

ERGM Assumptions

Given a set of network statistics:

1. There is equal probability of observing any two networks with the same values of those statistics

   Assumes the model is completely and correctly specified

ERGM Assumptions

Given a set of network statistics:

2. The observed network exhibits the average value of those statistics over the networks that could have been observed

   Identifies parameters

   No different from the assumption that the average relationships in a dataset are representative of the population (implicit in OLS and MLE)

   After all, what better indication of the average do we have than the one value we observe?

Limitations of (Basic) ERGM

- Challenges in estimation
  $\rightarrow$ A major theme in our lecture on estimation and degeneracy
- The ERGM cannot be used to model longitudinally observed networks
  $\rightarrow$ We'll show how to extend this with the Temporal ERGM
- The ERGM cannot accommodate non-binary networks
  $\rightarrow$ We'll show how to extend this with the Generalized ERGM

Unpacking **h**: Exogenous Covariates

- A general statistic that captures the relationship between X and N is

$$h_D(N, X) = \sum_{ij} N_{ij} X_{ij}$$

- Naturally accommodates dyad-level covariates

- Node level covariates must be "scaled up" :
$X_{ij} = Z_i$ for node-level covariate Z

- Interpretation: If X has a positive effect on edge formation, then $h_X(N)$ will be higher than if the covariate values were randomly assigned to edges, controlling for the effect of the other $h(N)$s on the network configuration

Endogenous Effects in **h**

How would we measure **reciprocity**?

A statistic we would expect to be high if ties were reciprocated a lot and low if they were not reciprocated.

Unpacking **h**: Endogenous Effects

- Note: Endogenous specifications vary by the type of dependency being captured

- Reciprocity

$$h_R(N) = \sum_{i<j} N_{ij} N_{ji}$$

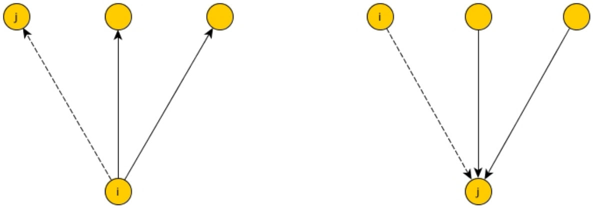
Unpacking $\mathbf{h}$

- Popularity

$$h_P(N) = \sum_{i,j,k} N_{ji} N_{ki} + N_{kj} N_{ij} + N_{ik} N_{jk}$$

- Sociality

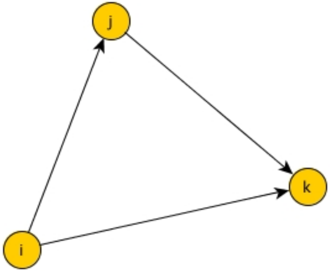
$$h_S(N) = \sum_{i,j,k} N_{ij} N_{ik} + N_{jk} N_{ji} + N_{ki} N_{kj}$$



Unpacking **h**: Endogenous Effects

- Example 2: Transitivity

$$h_T(N) = \sum_i \sum_{i \neq j,k} N_{ij} N_{ik} N_{jk}$$



Interpretation of ERGM

ERGM offers an incredibly flexible model – it can be used to investigate individual, dyad, node and network level effects.

Two most common method of interpretation

1. (**Network Level**) The relative likelihood of observing N^{j+} to observing N^{j-} is $\exp(\theta_j)$, where
   - θ_j is the estimate of the parameter that corresponds to statistic j.
   - N^{j+} is one unit greater than N^{j-} on statistic j (e.g., N^{j+} has one more closed triangle, one more edge, etc.), holding all other statistics constant.
2. (**Edge**) $P(N_{ij} = 1 | N_{-ij}, \boldsymbol{\theta}) = \text{logit}^{-1}\left(\sum_{r=1}^{k} \theta_r \delta_r^{(ij)}(N)\right)$
   - N_{-ij} indicates the network excluding N_{ij}
   - $\delta_r^{(ij)}(N)$ is equal to the change in Γ_r when N_{ij} is changed from zero to one
   - logit^{-1} is the inverse logistic function such that $\text{logit}^{-1}(x) = 1/(1 + \exp(-x))$

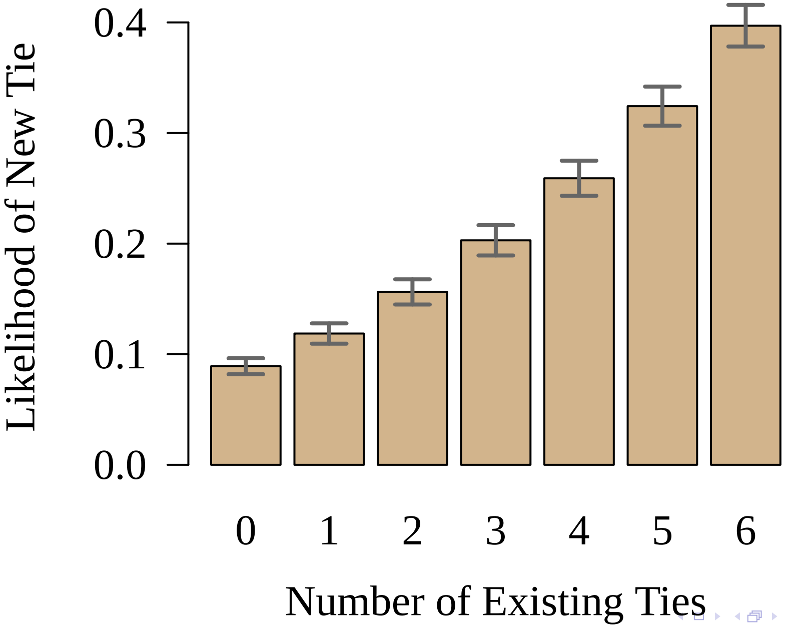
Interpretation by Block-wise Conditional Distributions

The primary inferential benefit of using ERGM is that it can represent how ties in the network depend upon each other.

- The dependencies of interest vary by application.
- **Conditional Distributions** offer a unifying concept for analyzing interdependencies. For instance
  - How expected values depend upon another
  - How variance in ties propagates through the network.
  - Which nodes are relatively 'independent'
- Desmarais and Cranmer (2013) show

$$P(\mathbf{N}_b|\mathbf{N}_{-b}) = \frac{\exp(\sum_{j=1}^{k} \theta_j h_j(\mathbf{N}_b \cup \mathbf{N}_{-b}))}{\sum_{\mathrm{N}_b^* \in \mathcal{N}_b} \exp(\sum_{j=1}^{k} \theta_j h_j(\mathbf{N}_b^* \cup \mathbf{N}_{-b}))}$$

- The notation $\mathbf{N}_b^* \cup \mathbf{N}_{-b}$ stands for the complete network created by holding $\mathbf{N}_{-b}$ constant and inserting $\mathbf{N}_b^*$ into the b^{th} block of N



Likelihood of New Tie
0.0
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
Number of Existing Ties

Example ERGM

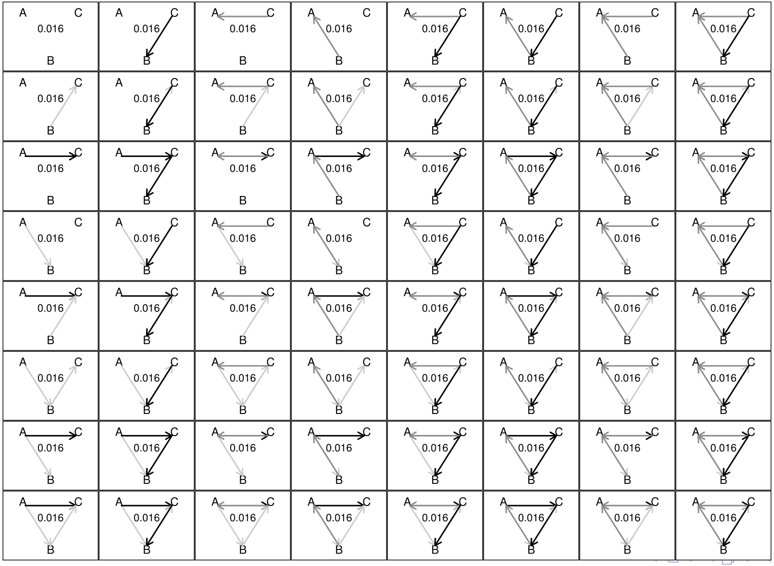
A small scale example

1. Directed network defined on three nodes $\{A, B, C\}$

2. An exogenous edge attribute X with three values: low, moderate and high

3. 64 possible realizations of this network

Naive uniform assumption $\frac{1}{64}$

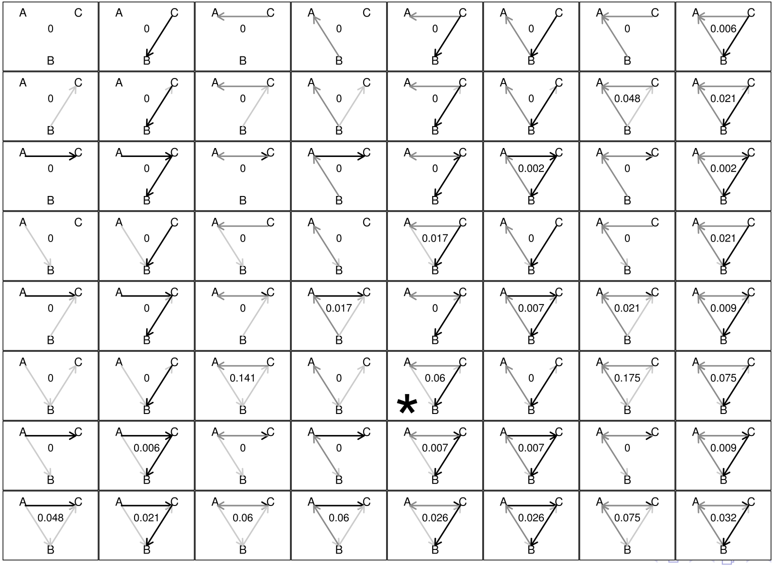


Example ERGM

A small scale example

1. Suppose a network N is observed
2. This network contains four out of six possible edges, one reciprocal edge, two edges corresponding to low covariate values, one edge corresponding to a moderate covariate value, and one edge corresponding to a high covariate value.
3. Specify an ERGM including: the number of edges (E), the number of reciprocal edges (R), and the sum of the covariate values corresponding to the edges (S_X)
4. The maximum likelihood estimates of the effects of these statistics, estimated from the starred network, are $\theta_E = 16.96$, $\theta_R = -14.61$, and $\theta_{S_x} = -1.07$
5. We can see the distribution over three-node directed networks implied by these ERGM parameters and statistics.
6. $P(N)$ is 0.06, almost four times as high as the uniform probability of 0.016.

Implied distribution



A more realistic example

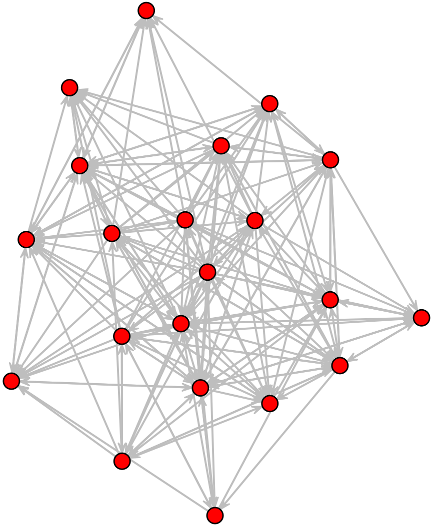
Consider a network. . .

Krackhardt's manager advice network

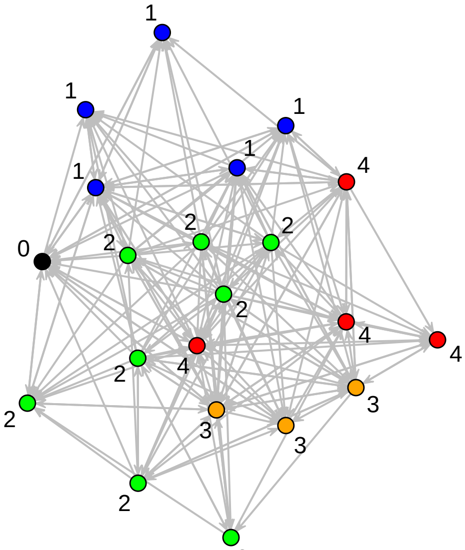
- Nodes: 21
- Edges: Directed, binary
- Density: 0.4524
- Covariates: Age, Department, Level, Tenure, Reports to

Exploratory analysis of the Krackhardt network.

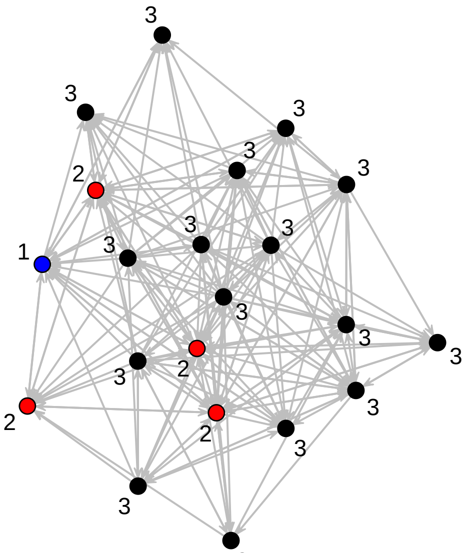
Krackhardt Manager Advice Network



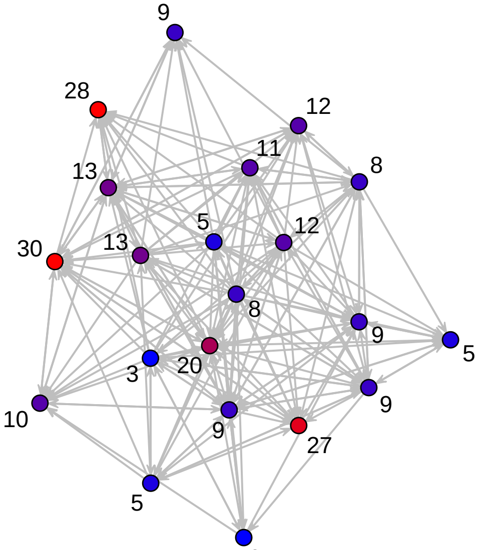
Krackhardt Manager Advice Network: Department



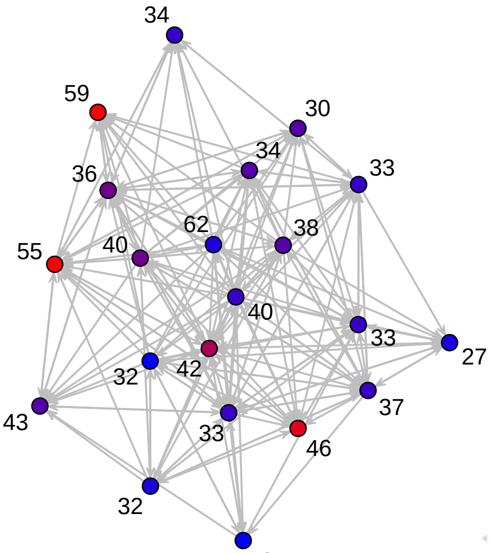
Krackhardt Manager Advice Network: Level



Krackhardt Manager Advice Network: Tenure

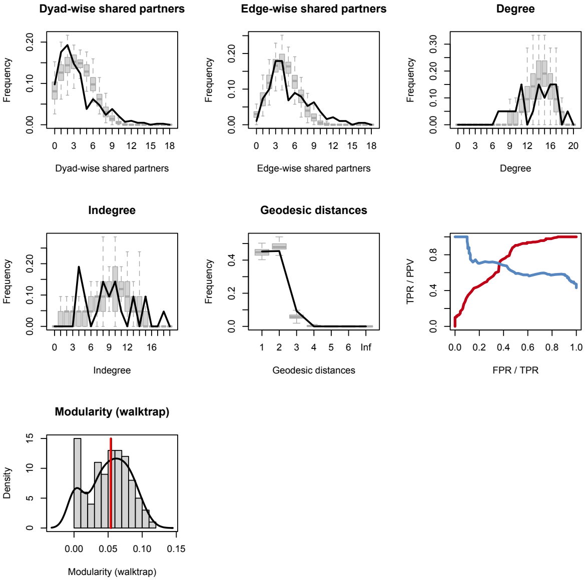


Krackhardt Manager Advice Network: Age

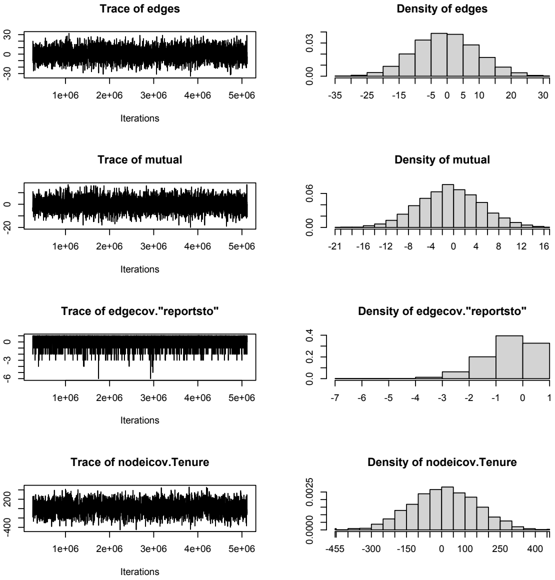


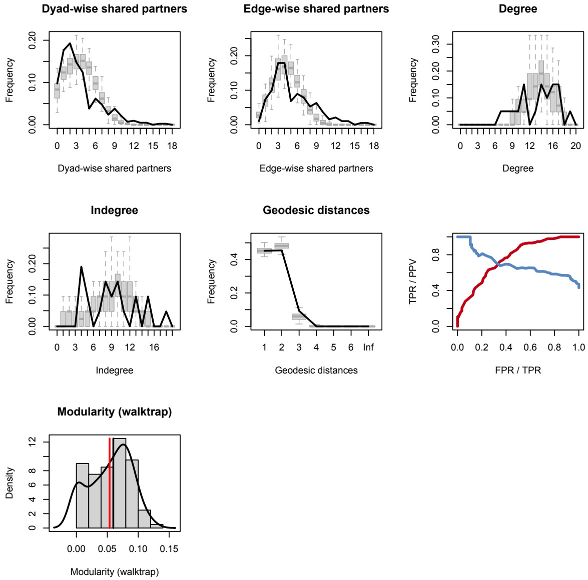
```r
## Covariate model without network structural parameters.
mod0 <- ergm(Krack ~ edges
          + edgecov("reportsto")
          + nodeicov("Tenure") + nodeocov("Tenure") + absdiff("Tenure")
          + nodeicov("Age") + nodeocov("Age") + absdiff("Age"),
          verbose=TRUE)
gof0 <- gof(mod0, verbose=TRUE)
par(mfrow=c(2,2)); plot(gof0)
summary(mod0)
```

```
                     Estimate Std. Error MCMC % z value Pr(>|z|)
edges                -0.22089    0.83101      0  -0.266 0.790389
edgecov."reportsto"   3.17309    1.06300      0   2.985 0.002835 *
nodeicov.Tenure       0.10798    0.02056      0   5.253  < 1e-04 *
nodeocov.Tenure      -0.05448    0.01765      0  -3.086 0.002029 *
absdiff.Tenure       -0.07551    0.02212      0  -3.413 0.000643 *
nodeicov.Age         -0.02184    0.01781      0  -1.226 0.220146
nodeocov.Age          0.02489    0.01502      0   1.657 0.097576
absdiff.Age          -0.02618    0.01792      0  -1.461 0.143914
```


Dyad-wise shared partners
Frequency
Dyad-wise shared partners
Edge-wise shared partners
Frequency
Edge-wise shared partners
Degree
Frequency
Degree
Indegree
Frequency
Indegree
Geodesic distances
Frequency
Geodesic distances
TPR / PPV
FPR / TPR
Modularity (walktrap)
Density
Modularity (walktrap)

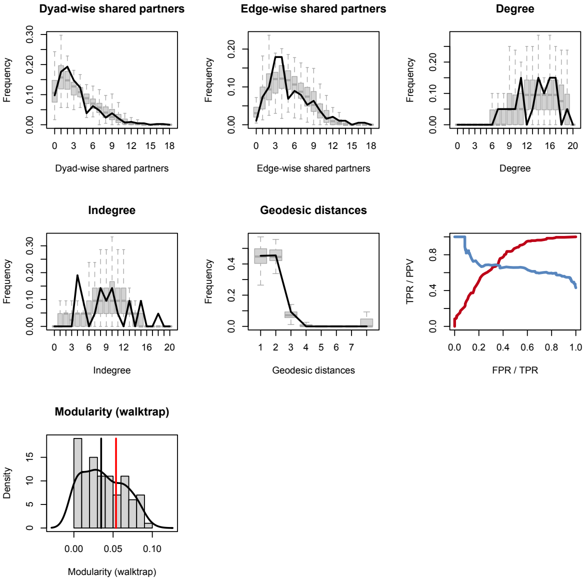
```r
## Model with reciprocity and covariates. NOTE: Burnin and sample sizes are too
## low to get a good est. Increase by a factor of 10 to get reliable estimates.
set.seed(5)
mod1 <- ergm(Krack ~ edges + mutual
             + edgecov("reportsto")
             + nodeicov("Tenure") + nodeocov("Tenure") + absdiff("Tenure")
             + nodeicov("Age") + nodeocov("Age") + absdiff("Age"),
             control=control.ergm(
                 MCMC.samplesize=5000,
                 MCMC.burnin=10000,
                 MCMLE.maxit=10),
             verbose=TRUE)
mcmc.diagnostics(mod1, vars.per.page=5)
plot(mod1$sample)                          # using coda instead
gof1 <- gof(mod1, verbose=TRUE)
par(mfrow=c(2,2)); plot(gof1)
summary(mod1)
```


Trace of edges
Density of edges
30
0
-30
1e+06
2e+06
3e+06
4e+06
5e+06
Iterations
0.00
0.03
-35
-25
-15
-5
0
5
10
20
30
Trace of mutual
Density of mutual
0
-20
1e+06
2e+06
3e+06
4e+06
5e+06
Iterations
0.00
0.06
-21
-16
-12
-8
-4
0
4
8
12
16
Trace of edgecov."reportsto"
Density of edgecov."reportsto"
0
-3
-6
1e+06
2e+06
3e+06
4e+06
5e+06
Iterations
0.0
0.2
0.4
-7
-6
-5
-4
-3
-2
-1
0
1
Trace of nodeicov.Tenure
Density of nodeicov.Tenure
200
-400
1e+06
2e+06
3e+06
4e+06
5e+06
Iterations
0.0000
0.0025
-455
-300
-150
0
100
250
400


Dyad-wise shared partners
Edge-wise shared partners
Degree
Indegree
Geodesic distances
Modularity (walktrap)
Frequency
Density
TPR / PPV
FPR / TPR
Inf

```
                    Estimate Std. Error MCMC % z value Pr(>|z|)
edges               -0.42817    0.80808      0  -0.530  0.59621
mutual               0.59242    0.33985      0   1.743  0.08130 .
edgecov."reportsto"  3.19355    1.04934      0   3.043  0.00234 *
nodeicov.Tenure      0.11593    0.02137      0   5.425  < 1e-04 *
nodeocov.Tenure     -0.06780    0.01945      0  -3.486  0.00049 *
absdiff.Tenure      -0.06947    0.02155      0  -3.224  0.00126 *
nodeicov.Age        -0.02637    0.01802      0  -1.464  0.14328
nodeocov.Age         0.02689    0.01485      0   1.811  0.07013
absdiff.Age         -0.02225    0.01731      0  -1.286  0.19857
```

```r
## Add out-stars
set.seed(10)
mod2 <- ergm(Krack ~ edges + mutual + ostar(2:3)
             + edgecov("reportsto")
             + nodeicov("Tenure") + nodeocov("Tenure") + absdiff("Tenure")
             + nodeicov("Age") + nodeocov("Age") + absdiff("Age"),
             control=control.ergm(
                 MCMC.samplesize=5000,
                 MCMC.burnin=10000,
                 MCMLE.maxit=10),
             verbose=TRUE)
mcmc.diagnostics(mod2)
plot(mod1$sample)                          # using coda instead
gof2 <- gof(mod2, verbose=TRUE)
par(mfrow=c(2,2)); plot(gof2)
summary(mod2)
```

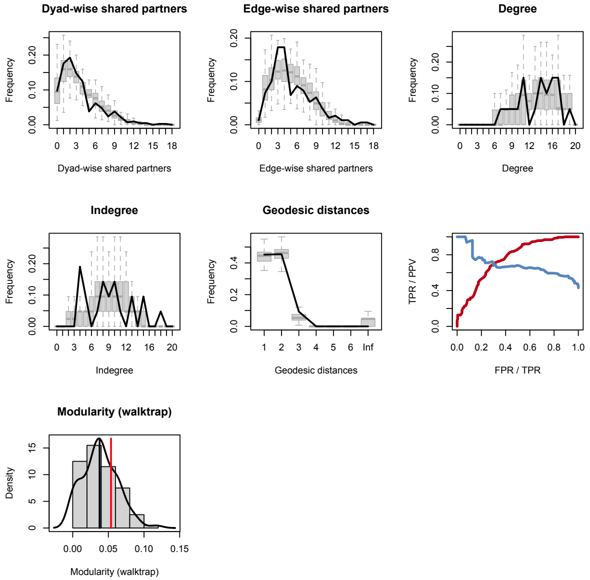

Dyad-wise shared partners
Frequency
0.00 0.10 0.20 0.30
0 3 6 9 12 15 18
Dyad-wise shared partners
Edge-wise shared partners
Frequency
0.00 0.10 0.20
0 3 6 9 12 15 18
Edge-wise shared partners
Degree
Frequency
0.00 0.10 0.20
0 3 6 9 12 16 20
Degree
Indegree
Frequency
0.00 0.10 0.20 0.30
0 3 6 9 12 16 20
Indegree
Geodesic distances
Frequency
0.0 0.2 0.4
1 2 3 4 5 6 7
Geodesic distances
TPR / PPV
0.0 0.4 0.8
0.0 0.2 0.4 0.6 0.8 1.0
FPR / TPR
Modularity (walktrap)
Density
0 5 10 15
0.00 0.05 0.10
Modularity (walktrap)

```
                      Estimate Std. Error MCMC % z value Pr(>|z|)
edges                -2.114480   0.890758      0  -2.374 0.017606 *
mutual                0.576654   0.324189      0   1.779 0.075280 .
ostar2                0.199293   0.100691      0   1.979 0.047788 *
ostar3                0.004144   0.010872      0   0.381 0.703049
edgecov."reportsto"   3.771637   1.088355      0   3.465 0.000529 *
nodeicov.Tenure       0.133947   0.022955      0   5.835  < 1e-04 *
nodeocov.Tenure      -0.019719   0.013548      0  -1.456 0.145524
absdiff.Tenure       -0.048401   0.018613      0  -2.600 0.009312 *
nodeicov.Age         -0.046148   0.019620      0  -2.352 0.018669 *
nodeocov.Age          0.015284   0.009132      0   1.674 0.094187 .
absdiff.Age          -0.017881   0.014945      0  -1.196 0.231527
```

```r
## Add transitivity ; Degenerate
set.seed(15)
mod3 <- ergm(Krack ~ edges + mutual + ostar(2:3) + transitive
             + edgecov("reportsto")
             + nodeicov("Tenure") + nodeocov("Tenure") + absdiff("
                Tenure")
             + nodeicov("Age") + nodeocov("Age") + absdiff("Age"),
             control=control.ergm(
                 MCMC.samplesize=5000,
                 MCMC.burnin=10000,
                 MCMLE.maxit=10),
             verbose=TRUE)
mcmc.diagnostics(mod3)
gof3 <- gof(mod3, verbose=TRUE)
par(mfrow=c(2,2)); plot(gof3)
summary(mod3)
```

Clunk!

```
## Use GWESP
set.seed(20)
mod4 <- ergm(Krack ~ edges + mutual + ostar(2:3) + gwesp(0, fixed=
    TRUE)
             + edgecov("reportsto")
             + nodeicov("Tenure") + nodeocov("Tenure") + absdiff("
                 Tenure")
             + nodeicov("Age") + nodeocov("Age") + absdiff("Age"),
             control=control.ergm(
                 MCMC.samplesize=5000,
                 MCMC.burnin=10000,
                 MCMLE.maxit=10),
             verbose=TRUE)
mcmc.diagnostics(mod4)
gof4 <- gof(mod4, verbose=TRUE)
par(mfrow=c(2,2)); plot(gof4)
summary(mod4)
```


Dyad-wise shared partners
Frequency
0.00 0.10 0.20
0 3 6 9 12 15 18
Dyad-wise shared partners
Edge-wise shared partners
Frequency
0.00 0.10 0.20
0 3 6 9 12 15 18
Edge-wise shared partners
Degree
Frequency
0.00 0.10 0.20
0 3 6 9 12 16 20
Degree
Indegree
Frequency
0.00 0.10 0.20
0 3 6 9 12 16 20
Indegree
Geodesic distances
Frequency
0.0 0.2 0.4
1 2 3 4 5 6 Inf
Geodesic distances
TPR / PPV
0.0 0.4 0.8
0.0 0.2 0.4 0.6 0.8 1.0
FPR / TPR
Modularity (walktrap)
Density
0 5 10 15
0.00 0.05 0.10 0.15
Modularity (walktrap)

```
                      Estimate Std. Error MCMC % z value Pr(>|z|)
edges                -3.710550   1.192170      0  -3.112  0.00186 *
mutual                0.483411   0.340762      0   1.419  0.15601
ostar2                0.228414   0.099066      0   2.306  0.02113 *
ostar3                0.001701   0.010602      0   0.160  0.87251
gwesp.fixed.0         1.390206   0.760309      0   1.828  0.06748 .
edgecov."reportsto"   3.757909   1.148804      0   3.271  0.00107 *
nodeicov.Tenure       0.138634   0.022947      0   6.042  < 1e-04 *
nodeocov.Tenure      -0.022380   0.013812      0  -1.620  0.10516
absdiff.Tenure       -0.049341   0.019107      0  -2.582  0.00981 *
nodeicov.Age         -0.047273   0.019239      0  -2.457  0.01400 *
nodeocov.Age          0.015489   0.009028      0   1.716  0.08623 .
absdiff.Age          -0.017436   0.014925      0  -1.168  0.24273
```