

# Assignment 2: Exponential Random Graph Model

2025-01-13

By Marco Di Stasio, Alessandra Gandini, Gaudenzia Genoni, Yishak Tadele Nigatu

## Introduction

This study applies ERGM analysis to a directed, valued friendship network of 32 network scientists, derived from the Electronic Information Exchange System (EIES) dataset. After dichotomizing the network, three models are estimated:

1. a purely structural model, capturing key network-level features such as reciprocity, degree distribution, and triadic structures;
2. a dyadic-independent model incorporating gender attributes to assess individual-level effects;
3. a combined model integrating both structural and attribute-based effects.

The coefficients of the models are interpreted with a substantial focus on Model 3, followed by goodness-of-fit diagnostics to evaluate how well each model replicates the observed network structure and relational patterns. Model 3 provides the most comprehensive representation of the network. However, there is still room for refinement, as certain effects, such as rare triadic configurations or extreme degree values, may not yet be fully controlled for.

## Part One: the Network and the three Models

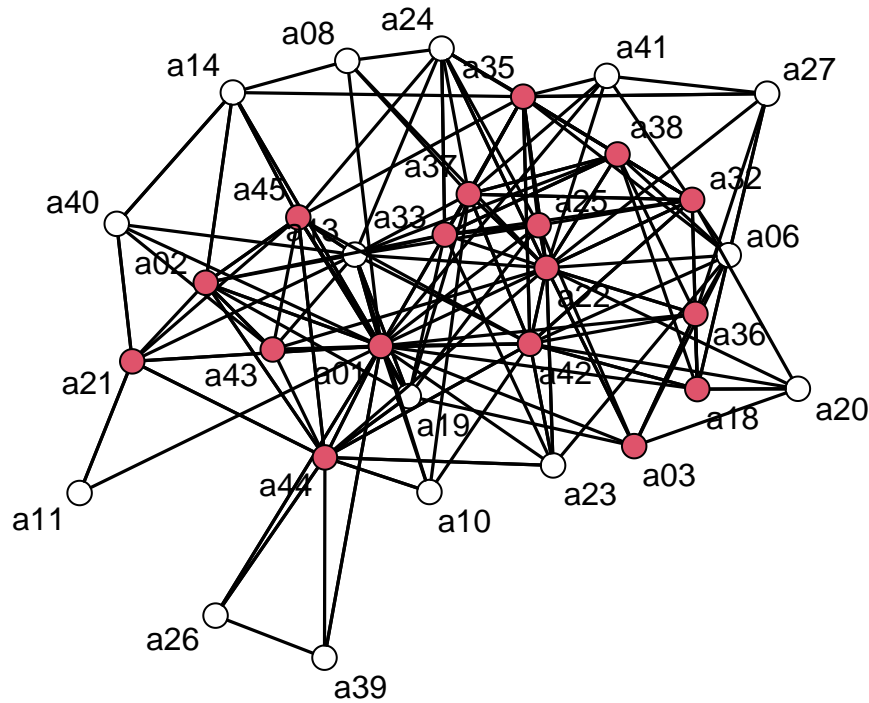
The study examines acquaintance relationships among 32 network scientists using the EIES dataset at Time 2. To focus on strong ties, the directed, valued network is dichotomized: edges with values of 3 (friend) or 4 (close personal friend) are coded as 1, and all others (0, 1, 2) are coded as 0. Gender attributes (from an external dataset) are then assigned to each node, creating an enriched network object for analysis.

```
library(sna)
library(ergm)
library(readxl)
library(knitr)
library(broom)
```

```
EIES_T2F_n <- read.csv("data/EIES_T2.csv", row.names=1)
EIES_T2F_n <- EIES_T2F_n > 2 #consider only friends (3) and good friends (4)

EIES_T2F_n<-as.network(EIES_T2F_n, directed=T) #transform into a network object
```

```
Attributes <- read_excel("data/Data_Gender_3.xlsx")
EIES_T2F_n %v% "Gender"<- Attributes$Gender
```



Three exponential random graph models (ERGMs) are estimated. **Model 1** captures purely structural effects, including edges, mutual ties, geometrically weighted in-degree and out-degree distributions, and two-path dependencies (outgoing and incoming). **Model 2** examines dyadic independent effects by incorporating individual and dyadic attributes, specifically gender as a nodal covariate for both in-degree and out-degree, as well as gender homophily through nodal matching. **Model 3** combines structural and attribute-based effects, integrating the terms from both Models 1 and 2 to provide a comprehensive analysis of the network.

### Model 1

```
## Call:
## ergm(formula = EIES_T2F_n ~ edges + mutual + gwidegree(decay = 0.3,
##   fixed = TRUE) + gwodegree(decay = 0.3, fixed = TRUE) + dgwesp(type = "OTP",
##   decay = 0.5, fixed = TRUE) + dgwesp(type = "ITP", decay = 0.5,
##   fixed = TRUE), control = control.ergm(seed = 102, MCMC.runtime.traceplot = TRUE),
##   verbose = TRUE)
##
## Monte Carlo Maximum Likelihood Results:
##
##               Estimate Std. Error MCMC % z value Pr(>|z|)
## edges             -4.8246    0.3630    0 -13.293  <1e-04 ***
## mutual              2.6026    0.3437    0  7.571   <1e-04 ***
## gwideg.fixed.0.3    2.1954    1.1351    0  1.934   0.0531 .
## gwodeg.fixed.0.3    4.3716    2.7916    0  1.566   0.1174
## gwesp.OTP.fixed.0.5  1.9668    0.2176    0  9.040   <1e-04 ***
## gwesp.ITP.fixed.0.5 -0.6267    0.1103    0 -5.682   <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Null Deviance: 1375.2 on 992 degrees of freedom
## Residual Deviance: 792.3 on 986 degrees of freedom
##
## AIC: 804.3 BIC: 833.7 (Smaller is better. MC Std. Err. = 0.7868)
```

## Model 2

```
## Call:
## ergm(formula = EIES_T2F_n ~ edges + mutual + nodeicov("Gender") +
##       nodeocov("Gender") + nodematch("Gender"), control = control.ergm(seed = 102,
##       MCMC.runtime.traceplot = TRUE), verbose = TRUE)
##
## Monte Carlo Maximum Likelihood Results:
##
##              Estimate Std. Error MCMC % z value Pr(>|z|)
## edges          -2.94132    0.18397      0 -15.988  <1e-04 ***
## mutual           2.48594    0.29055      0   8.556  <1e-04 ***
## nodeicov.Gender   0.05310    0.21679      0   0.245    0.807
## nodeocov.Gender   1.27429    0.20970      0   6.077  <1e-04 ***
## nodematch.Gender -0.02734    0.16159      0  -0.169    0.866
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 1375  on 992  degrees of freedom
## Residual Deviance:  853  on 987  degrees of freedom
##
## AIC: 863  BIC: 887.5  (Smaller is better. MC Std. Err. = 0.7932)
```

## Model 3

```
## Call:
## ergm(formula = EIES_T2F_n ~ edges + mutual + gwidegree(decay = 0.3,
## fixed = TRUE) + gwodegree(decay = 0.3, fixed = TRUE) + dgwesp(type = "OTP",
## decay = 0.5, fixed = TRUE) + dgwesp(type = "ITP", decay = 0.5,
## fixed = TRUE) + nodeicov("Gender") + nodeocov("Gender") +
## nodematch("Gender"), control = control.ergm(seed = 102, MCMC.runtime.traceplot = TRUE),
## verbose = TRUE)
##
## Monte Carlo Maximum Likelihood Results:
##
##              Estimate Std. Error MCMC % z value Pr(>|z|)
## edges          -5.562211    0.446245      0 -12.464  <1e-04 ***
## mutual           2.678476    0.359588      0   7.449  <1e-04 ***
## gwideg.fixed.0.3   1.640805    1.164139      0   1.409    0.1587
## gwodeg.fixed.0.3  16.286160    7.005113      0   2.325    0.0201 *
## gwesp.OTP.fixed.0.5 1.852469    0.223386      0   8.293  <1e-04 ***
## gwesp.ITP.fixed.0.5 -0.632633    0.117689      0  -5.375  <1e-04 ***
## nodeicov.Gender     0.005062    0.188986      0   0.027    0.9786
## nodeocov.Gender     1.235591    0.308959      0   3.999  <1e-04 ***
## nodematch.Gender   -0.071385    0.161228      0  -0.443    0.6579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 1375  on 992  degrees of freedom
## Residual Deviance:  755  on 983  degrees of freedom
##
## AIC: 773  BIC: 817.1  (Smaller is better. MC Std. Err. = 0.6992)
```

The results are presented in the following table:

Effect	Model 1 Estimate (SE)	Model 1 Sign	Model 2 Estimate (SE)	Model 2 Sign	Model 3 Estimate (SE)	Model 3 Sign
edges	-4.825 (0.363)	***	-2.941 (0.185)	***	-5.562 (0.446)	***
mutual	2.603 (0.344)	***	2.486 (0.291)	***	2.678 (0.360)	***
gwidegree	2.195 (1.135)	.	NA	NA	1.641 (1.164)	
gwodegree	4.377 (2.792)		NA	NA	16.286 (7.005)	*
dgwesp.OTP	1.967 (0.218)	***	NA	NA	1.853 (0.223)	***
dgwesp.ITY	-0.627 (0.110)	***	NA	NA	-0.637 (0.118)	***
nodeicov.Gender	NA	NA	0.053 (0.217)		0.005 (0.189)	
nodeocov.Gender	NA	NA	1.274 (0.210)	***	1.236 (0.309)	***
nodematch.Gender	NA	NA	-0.027 (0.162)		-0.071 (0.161)	

## Part Two: Discussion of ERGM Results

In Model 3, six effects are significant (edges, mutual, gwodeg.fixed, gwesp.OTP.fixed, gwesp.ITY.fixed, and nodeocov.Gender), while the remaining three (gwideg.fixed, nodeicov.Gender, and nodematch.Gender) are not.

Below, **Model 3** is discussed in detail and notable differences from **Models 1** (structural only) and **2** (attributes only) are highlighted.

### Interpretation of Model 3 Estimates

1. **Edges** (Estimate: **-5.562**,  $p < 0.001$ ) The strong negative coefficient indicates a low baseline propensity for any given pair to be “friends”. In a friendship context, this is typical because only a minority of all possible ties are strong friendships.
2. **Mutual** (Estimate: **2.678**,  $p < 0.001$ ) This significant positive effect suggests a strong tendency for reciprocity. In other words, if Person A calls Person B a friend, Person B is much more likely to call Person A a friend in return. This effect remains comparably large and significant across all models (it is also highly significant in Models 1 and 2).
3. **gwidegree** (Estimate: **1.641**, **not significant**) The a positive coefficient would suggest that popular individuals (i.e., those who receive many friendship ties) may have a slight tendency to attract additional ties. The effect, however, is not strong enough to reach statistical significance.
4. **gwodegree** (Estimate: **16.286**,  $p < 0.05$ ) This effect captures “activity” (i.e., actively initiating multiple ties, connected to a high out-degree). In Model 3 it is both larger and significant, implying that once we control for gender and other structural factors, some individuals indeed nominate friends at a higher-than-expected rate and create opportunities for others to become integrated into the friendship network.
5. **dgwesp.OTP** (Estimate: **1.853**,  $p < 0.001$ ) A positive and significant OTP parameter indicates that triadic closure is especially likely when an individual sends ties to two different people (i.e., “friends of my out-friends tend to be my friends too”). This finding aligns with typical friendship processes where one’s out-contacts may be more likely to know each other, and it is very similar in magnitude to Model 1, implying that adding gender attributes does not substantially alter this triadic closure dynamic.
6. **dgwesp.ITY** (Estimate: **-0.637**,  $p < 0.001$ ) The negative and significant ITY parameter implies that if two individuals both send ties to the same person, they are less likely than random to form a tie with each other. In friendship networks, this suggests that sharing a common recipient does not necessarily foster direct relationships between the two individuals. This pattern remains consistent from Model 1 through Model 3.

7. **nodeicov.Gender - male** (Estimate: **0.005, not significant**) This coefficient tests whether being male (coded as 0) makes one more likely to receive friend nominations. It is essentially zero and non-significant, indicating that once we control for other network tendencies, male actors do not appear to receive significantly more (or fewer) friendship ties than female actors.
8. **nodeocov.Gender - female** (Estimate: **1.236,  $p < 0.001$** ) This significant positive coefficient indicates that females (coded as 1) are more likely to send friendship ties than males. Notably, it is also significant in Model 2 at a similar magnitude (1.274). Thus, even after accounting for structural processes like reciprocity and triadic closure, women in this network remain more active in naming friends than men.
9. **nodematch.Gender** (Estimate: **-0.071, not significant\*\***) This nonsignificant coefficient suggests that there is no strong homophily or heterophily by gender for close friendships—i.e., men do not show a strong tendency to befriend men, nor women to befriend women, relative to cross-gender friendships, once other factors are considered. This is consistent across Models 2 and 3.

Generally, **Model 3** suggests that both structural forces (reciprocity, triadic closure, activity patterns) and gender attributes shape who becomes friends in this network, though some anticipated effects (like gender-based popularity or homophily) do not appear to be present.

### Notable Differences from Models 1 and 2 to Model 3

As can be seen from the table, the only effects that vary in significance from Model 1 to Model 3 are **gwidegree** (which decreases slightly and loses its marginal significance in Model 3) and **gwodegree** (which is not significant in Model 1 but becomes statistically significant in Model 3). Their interpretation is detailed below. With respect to significance, no effect changes from Model 2 to Model 3.

- For **gwidegree**, the marginally significant positive coefficient in Model 1 ( $=2.195, p=0.053$ ) suggests that individuals who receive many ties (are “popular”) have a slightly increased likelihood of forming additional friendships. When gender attributes are introduced in Model 3, the previously observed in-degree popularity effect is partly explained by other structural or gender-related factors. This suggests that “popularity” in terms of receiving ties is less crucial in explaining tie formation when accounting for sender gender and other structural dynamics.
- For *gwodeg.fixed*, the non-significant coefficient in Model 1 ( $=4.372, p=0.117$ ) indicates that individuals who send many ties (“outgoing activity”) do not significantly influence the overall network structure at this stage. However, when gender attributes are introduced in Model 3, the role of individuals who are more proactive in initiating friendships becomes more evident in shaping the network, especially when considering the sender’s gender.

## Part 3: Goodness of Fit

When performing an ERGM analysis, it is of crucial importance to examine the goodness of fit, as it evaluates how well the model captures the observed network’s structure and relational patterns. Here, goodness-of-fit (GOF) diagnostics are performed by setting a random seed to ensure reproducibility across runs. Subsequently, 2000 networks are simulated for comparison with the observed network, using the same seed for consistency. The GOF is calculated for the three models, comparing observed network statistics with those simulated from the model. The diagnostics include: - *model*: it evaluates how well the model itself fits the data. - *idegree and odegree*: it assesses the in-degree and out-degree distributions. - *distance*: it examines the geodesic distance distribution (shortest path lengths between nodes). - *triadcensus*: it evaluates the distribution of triadic configurations in the network. These additional metrics (degree distributions, distance, and triadic configurations) are used to complement the evaluation of the model fit. While the model statistic assesses overall fit, the additional metrics help pinpoint specific network features the model may under- or overestimate.

```
set.seed(565)
gof.choices<-control.gof.ergm(nsim=2000, seed = 565)
```

```
#Model 1
IES_Model1sim2000<-gof(IES_Model1,
                        GOF=~model+idegree+odegree+distance+triadcensus, control=gof.choices)
IES_Model1sim2000$summary.model
```

##	obs	min	mean	max	MC	p-value
## edges	204.00000	144.00000	204.59850	269.00000		1.000
## mutual	60.00000	36.00000	60.09950	88.00000		1.000
## gwideg.fixed.0.3	41.74622	35.93676	41.75617	43.13333		0.856
## gwodeg.fixed.0.3	42.57393	40.25029	42.58771	43.16086		0.806
## gwesp.OTP.fixed.0.5	271.15878	156.71861	272.05117	388.84664		0.988
## gwesp.ITP.fixed.0.5	217.35306	95.38425	217.62944	359.60525		0.965

Overall, the goodness-of-fit diagnostics for Model 1 indicate that the observed network statistics align closely with the simulated values across all metrics. This suggests that Model 1 provides an adequate fit for capturing the structural properties of the observed network.

```
#Model 2
IES_Model2sim2000<-gof(IES_Model2,
                        GOF=~model+idegree+odegree+distance+triadcensus, control=gof.choices)
IES_Model2sim2000$summary.model
```

##	obs	min	mean	max	MC	p-value
## edges	204	154	203.3955	245		1.000
## mutual	60	40	59.9185	80		1.000
## nodeicov.Gender	130	98	129.2015	164		0.962
## nodeocov.Gender	155	114	154.4205	194		1.000
## nodematch.Gender	111	78	111.3525	148		1.000

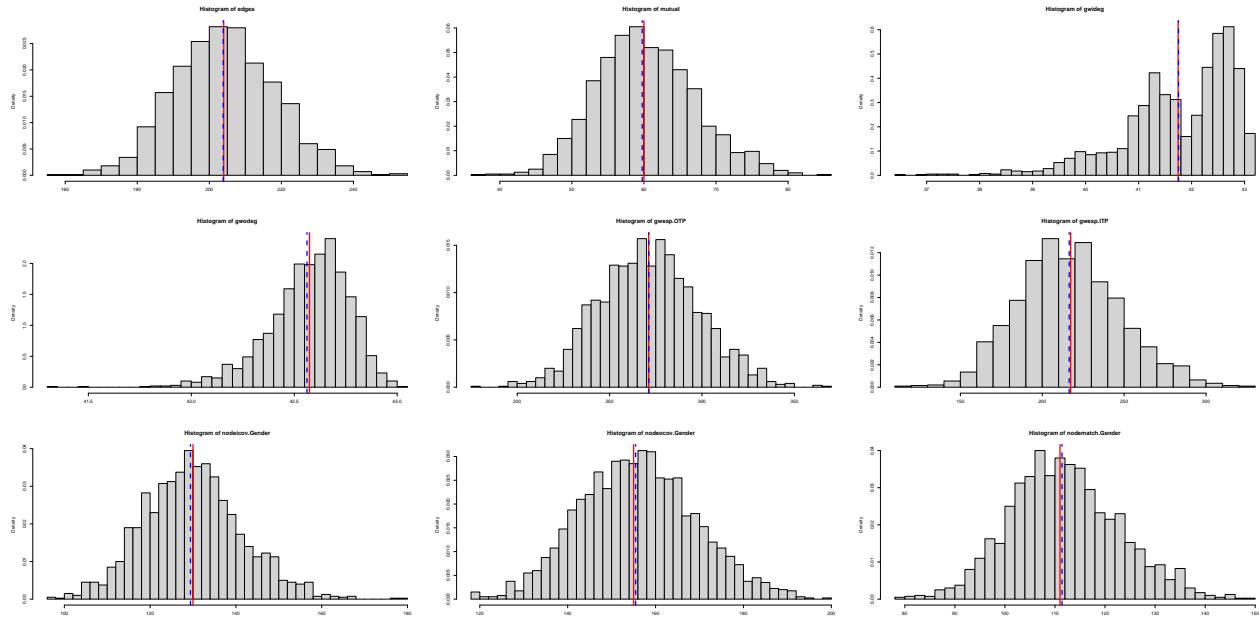
The goodness-of-fit diagnostics for Model 2 demonstrate that the observed network statistics are well-replicated by the simulated networks, too.

```
#Model 3
IES_Model3sim2000<-gof(IES_Model3,
                        GOF=~model+idegree+odegree+distance+triadcensus, control=gof.choices)
IES_Model3sim2000$summary.model
```

##	obs	min	mean	max	MC	p-value
## edges	204.00000	159.00000	203.84650	250.00000		1.000
## mutual	60.00000	37.00000	59.75300	85.00000		0.979
## gwideg.fixed.0.3	41.74622	36.50725	41.75348	43.14454		0.919
## gwodeg.fixed.0.3	42.57393	41.29307	42.56191	43.01802		0.957
## gwesp.OTP.fixed.0.5	271.15878	176.65631	271.25023	369.00044		0.987
## gwesp.ITP.fixed.0.5	217.35306	110.67811	216.44038	327.69236		0.932
## nodeicov.Gender	130.00000	96.00000	129.40250	179.00000		0.964
## nodeocov.Gender	155.00000	118.00000	155.45850	198.00000		1.000
## nodematch.Gender	111.00000	79.00000	111.40350	148.00000		1.000

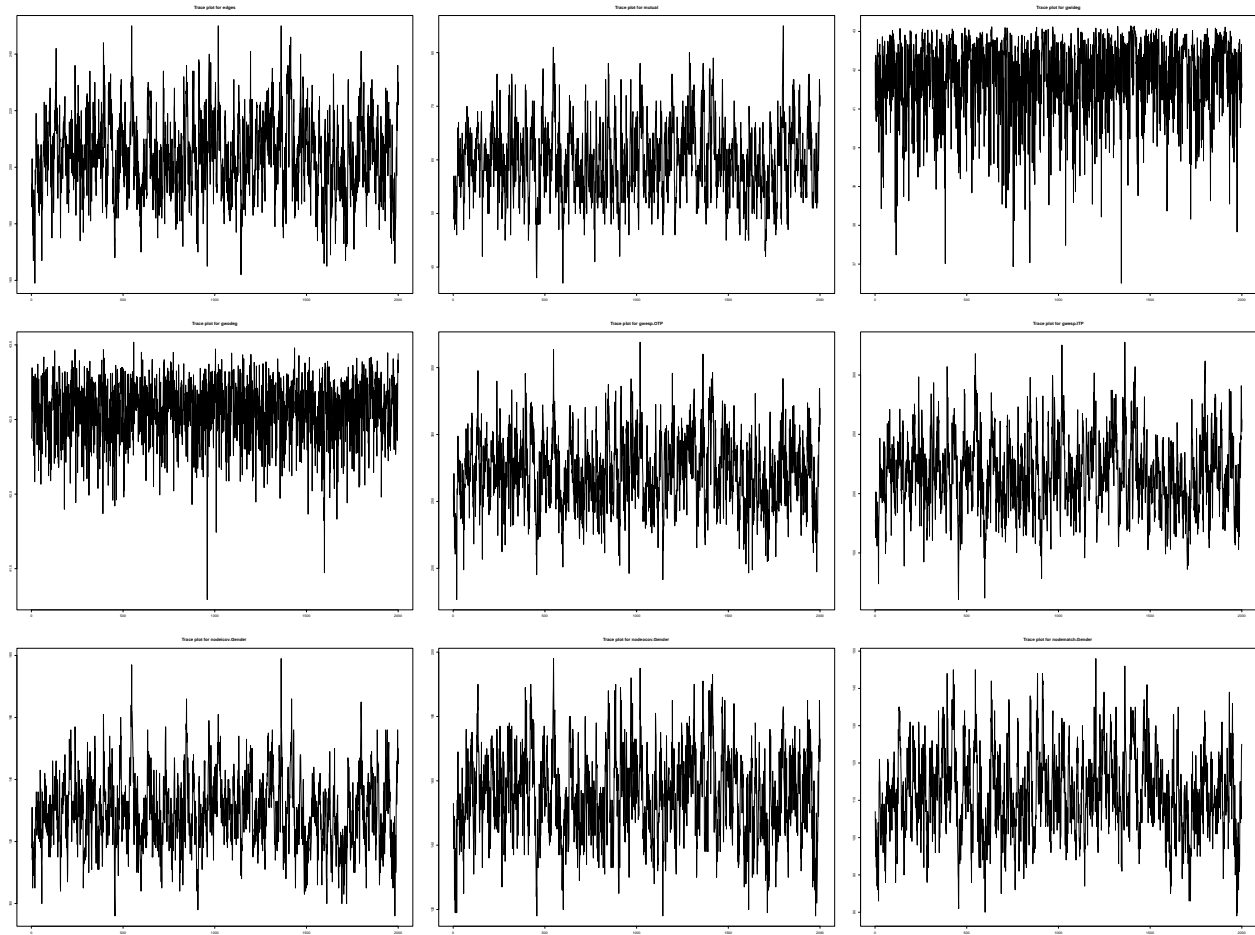
The goodness-of-fit diagnostics for Model 3 indicate a strong alignment between the observed network statistics and the simulated values, as reflected by high MC p-values across all metrics. Structural effects, such as edges, mutual ties, and measures of triadic closure (gwesp.OTP.fixed.0.5 and gwesp.ITP.fixed.0.5), are well-captured by the model, showing minimal discrepancies between observed and simulated values. Additionally, gender-related covariates, including nodeicov.Gender, nodeocov.Gender, and nodematch.Gender, are also accurately represented, suggesting that the model effectively incorporates both structural and attribute-based dynamics. This demonstrates that Model 3 provides a robust and comprehensive fit to the observed friendship network.

These findings are illustrated in the histograms below, where the blue dotted line (representing the mean of the predictions) and the red line (representing the value of the observations) are almost overlapping.



### Trace plots for Model 3

The *trace plots* showed below are visual diagnostics used to assess the convergence of the Markov Chain Monte Carlo (MCMC) algorithm in Model 3. In a well-behaved chain, we expect the sampled values to fluctuate around some stable mean (the chain's stationary distribution) without systematic trends upward or downward. The presence of major trends, drifts, or extremely prolonged excursions could indicate that the sampler has not “mixed” well or that convergence to a stable distribution has not yet occurred. By contrast, if the samples appear roughly like a tight, random cloud of points—oscillating within a limited range but neither wandering away nor getting stuck at extremes—this generally suggests good mixing and supports the notion that the MCMC process has reached stationarity. Here, each statistic (e.g., the number of edges or mutual ties in a simulated network) appears to remain within a relatively consistent band of values across the 2,000 iterations. For instance, the trace for **edges** hovers between roughly 160 and 240, showing no noticeable increasing or decreasing trend over time. Similarly, **mutual** shifts between about 40 and 80, and the plots for **gwideg** and **gwodeg** fluctuate around stable mid-40s and mid-42s, respectively, without drifting away. The **gwesp** parameters (OTP and ITP) also move within their respective ranges, and the gender covariates (both **nodecov**, **nodecov**, and **nodematch** for Gender) do not show signs of a systematic trend or long-term drift. Overall, the absence of any major trends, plateaus, or extremely abrupt jumps indicates that the MCMC procedure is performing reliably for each term in the model. This reliability underpins the validity of the fitted ERGM and supports the interpretation that Model 3 has converged, providing stable parameter estimates that capture the observed network structure and attribute effects.



### Goodness-of-Fit Diagnostics for Model 3

Finally, the evaluation of Model 3's fit is extended to include the other additional metrics (complementing the previously discussed GOF statistics): these are node-level degree distributions (in-degree and out-degree), geodesic distances, and triadic configurations. The values are reported below, followed by their graphical representation and a brief description for each.

EIES\_Model3sim2000

```
##
## Goodness-of-fit for model statistics
##
##               obs      min      mean      max MC p-value
## edges          204.00000 159.00000 203.84650 250.00000      1.000
## mutual          60.00000  37.00000  59.75300  85.00000      0.979
## gwidth.fixed.0.3  41.74622  36.50725  41.75348  43.14454      0.919
## gwodeg.fixed.0.3  42.57393  41.29307  42.56191  43.01802      0.957
## gwesp.OTP.fixed.0.5 271.15878 176.65631 271.25023 369.00044      0.987
## gwesp.ITP.fixed.0.5 217.35306 110.67811 216.44038 327.69236      0.932
## nodecov.Gender    130.00000  96.00000 129.40250 179.00000      0.964
## nodecov.Gender    155.00000 118.00000 155.45850 198.00000      1.000
## nodematch.Gender   111.00000  79.00000 111.40350 148.00000      1.000
##
## Goodness-of-fit for in-degree
##
```



##		obs	min	mean	max	MC	p-value
##	idegree0	0	0	0.5665	4		1.000
##	idegree1	3	0	1.1410	6		0.207
##	idegree2	3	0	1.9955	8		0.644
##	idegree3	4	0	2.8830	11		0.678
##	idegree4	5	0	3.4905	9		0.531
##	idegree5	1	0	3.8815	12		0.169
##	idegree6	3	0	3.7950	10		0.926
##	idegree7	3	0	3.4810	11		1.000
##	idegree8	2	0	2.9165	10		0.865
##	idegree9	1	0	2.2875	8		0.650
##	idegree10	2	0	1.7760	6		1.000
##	idegree11	1	0	1.3005	6		1.000
##	idegree12	2	0	0.9150	5		0.462
##	idegree13	0	0	0.6145	4		1.000
##	idegree14	0	0	0.3940	3		1.000
##	idegree15	1	0	0.2250	3		0.414
##	idegree16	0	0	0.1730	2		1.000
##	idegree17	0	0	0.0790	2		1.000
##	idegree18	0	0	0.0380	1		1.000
##	idegree19	0	0	0.0295	1		1.000
##	idegree20	0	0	0.0120	1		1.000
##	idegree21	0	0	0.0020	1		1.000
##	idegree22	0	0	0.0035	1		1.000
##	idegree24	1	0	0.0000	0		0.000

##  
## Goodness-of-fit for out-degree  
##

##		obs	min	mean	max	MC	p-value
##	odegree1	0	0	0.1605	4		1.000
##	odegree2	6	0	4.5150	11		0.588
##	odegree3	2	0	5.9685	13		0.068
##	odegree4	4	0	3.7400	10		1.000
##	odegree5	3	0	2.4550	8		0.876
##	odegree6	2	0	1.9800	8		1.000
##	odegree7	5	0	1.9615	7		0.079
##	odegree8	0	0	1.9275	7		0.254
##	odegree9	5	0	1.9180	8		0.069
##	odegree10	2	0	1.7415	7		1.000
##	odegree11	1	0	1.4930	7		1.000
##	odegree12	1	0	1.2210	5		1.000
##	odegree13	0	0	0.9885	6		0.704
##	odegree14	0	0	0.6880	4		0.963
##	odegree15	0	0	0.4710	4		1.000
##	odegree16	0	0	0.3410	3		1.000
##	odegree17	0	0	0.1985	3		1.000
##	odegree18	0	0	0.1225	2		1.000
##	odegree19	0	0	0.0610	2		1.000
##	odegree20	1	0	0.0280	1		0.056
##	odegree21	0	0	0.0130	1		1.000
##	odegree22	0	0	0.0035	1		1.000
##	odegree23	0	0	0.0030	1		1.000
##	odegree24	0	0	0.0005	1		1.000

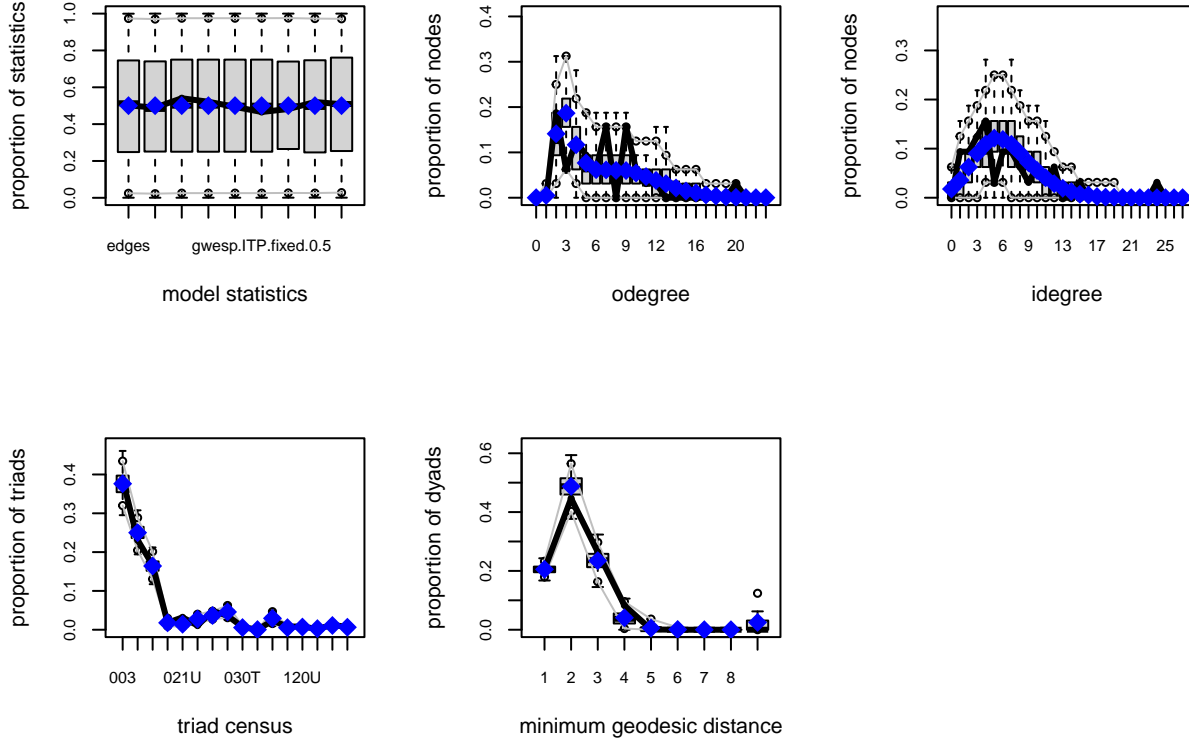
##

```

## Goodness-of-fit for minimum geodesic distance
##
##      obs min      mean max MC p-value
## 1   204 159 203.8465 250      1.000
## 2   443 355 483.5300 621      0.310
## 3   262 109 232.9405 339      0.417
## 4    82   0  40.2935 147      0.126
## 5     1   0   5.7425  96      0.989
## 6     0   0   0.8045  60      1.000
## 7     0   0   0.1155  59      1.000
## 8     0   0   0.0070   9      1.000
## Inf   0   0  24.7200 267      1.000
##
## Goodness-of-fit for triad census
##
##      obs min      mean max MC p-value
## triadcensus.003 1926 1413 1865.9060 2353      0.693
## triadcensus.012 1146  798 1240.3890 1621      0.384
## triadcensus.102  826  491  814.4885 1149      0.874
## triadcensus.021D   90   26   88.6065  201      0.892
## triadcensus.021U  158   26   70.1980  140      0.000
## triadcensus.021C   53   53  136.0185  238      0.001
## triadcensus.111D  241  106  182.0200  273      0.040
## triadcensus.111U  180  113  227.1695  373      0.218
## triadcensus.030T   35    4   27.5155   71      0.464
## triadcensus.030C    0    0    1.7240    9      0.472
## triadcensus.201  116   47  145.7095  291      0.460
## triadcensus.120D   42    8   26.6230   56      0.086
## triadcensus.120U   40    5   33.9055   74      0.541
## triadcensus.120C   10    1   13.9735   45      0.604
## triadcensus.210   62   19   54.2680  113      0.549
## triadcensus.300   35   10   31.4850   64      0.649

```

## Goodness-of-fit diagnostics



**Boxplots of Key Statistics (Edges, Mutual, gwodeg, gwesp, nodecov, Gender, etc.)** As already mentioned, the fitted model reproduces the fundamental network features very well, evidenced by high MC p-values for edges ( $p=1.000$ ), mutual ties ( $p=0.979$ ), and other key statistics (gwdeg, gwodeg, gwesp, and nodematch by Gender). For instance, the observed number of edges (204) sits comfortably within the simulated range (159 to 250), and the same holds for mutual ties (observed=60, mean=59.75). These results indicate that the model captures both basic connectivity (edges) and more nuanced structural tendencies (reciprocity, geometrically weighted terms, gender effects).

**Out-Degree and In-Degree Distributions** The second and third plots show the *distribution* of out-degree (number of ties sent) and in-degree (number of ties received) across all nodes. The outdegree plot focuses on how many friends each person nominates, whereas the indegree plot shows how many friend nominations each person receives.

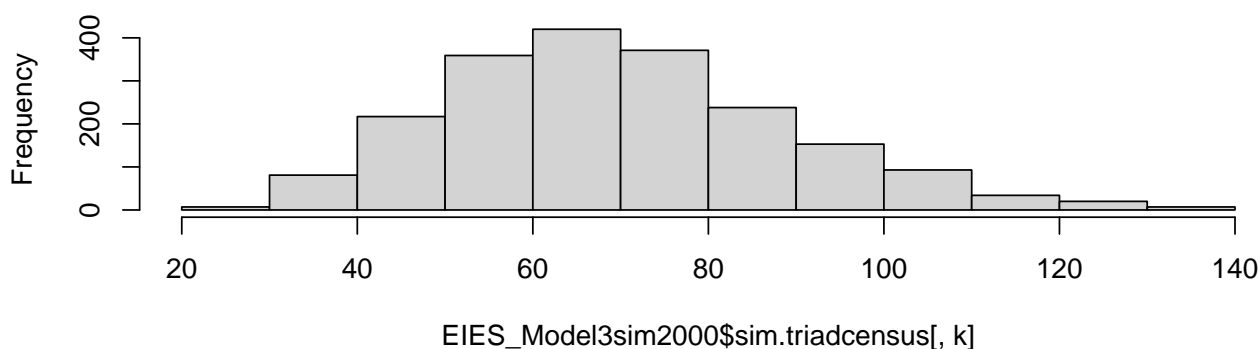
Overall, the model does a good job fitting most in-degree categories, as shown by high p-values for degrees such as 6 ( $p=0.926$ ), 7 ( $p=1.000$ ), and 10 ( $p=1.000$ ). However, certain counts depart from the observed data. For example, the frequency of nodes with 1 incoming tie (observed=3) is underestimated (mean=1.1410;  $p=0.207$ ), whereas the number of nodes with 5 incoming ties (observed=1) is overestimated (mean=3.8815;  $p=0.169$ ). The biggest mismatch occurs at the extreme end (idegree24): the model never predicts this very high in-degree (mean=0.0000;  $p=0.000$ ), suggesting it struggles with nodes that receive a very large number of ties.

The out-degree distribution also aligns well with the observed data in most categories, reflected in high p-values for many degrees (e.g., odegree4  $p=1.000$ , odegree6  $p=1.000$ ). Nonetheless, the model underestimates some degrees (e.g., odegree3 observed=2 vs. mean=5.9685,  $p=0.068$ ) and slightly overestimates others (e.g., odegree7  $p=0.079$ , odegree9  $p=0.069$ ). A notable discrepancy occurs at odegree20 (observed=1 vs. mean=0.028,  $p=0.056$ ), indicating the model struggles to capture the presence of a rare “super-connector.”

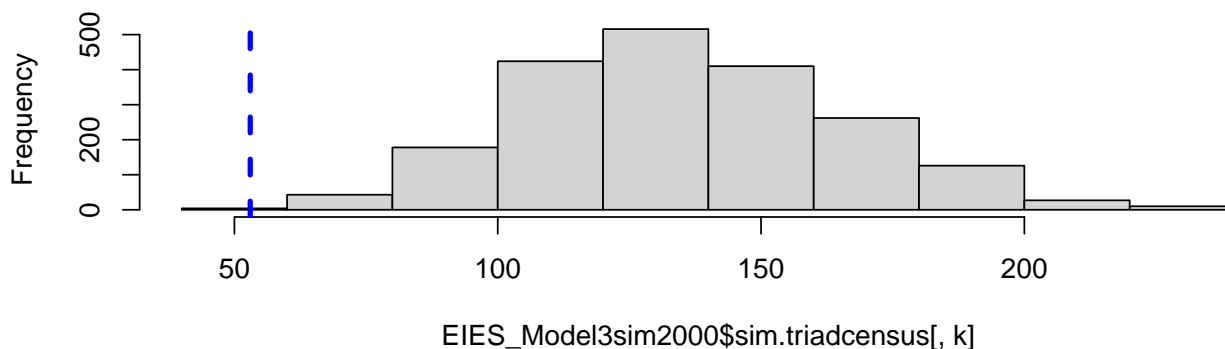
## Triad Census

The triad census (fourth plot) displays the proportion of all triads (three-node subgraphs) of each type—e.g., "003" means no ties among the three nodes, "021D" means a particular pattern of two directed edges, "111U" a fully mutual triad, and so on. In Model 3, most triad types appear in realistic proportions (e.g., 003 has  $p=0.693$ , 102  $p=0.874$ ), which suggests the model handles common triad structures adequately. However, certain directed triads are misestimated. For instance, 021U ( $p=0.000$ ) and 021C ( $p=0.001$ ) deviate markedly from the observed data (as can be seen in detail from the following picture), and fully reciprocal triads (111D, 111U) also show some discrepancies ( $p=0.040$  and  $0.218$ , respectively). Although overall triadic structure is captured, those directed subgraphs remain more challenging for the model. In particular, 021U is the configuration where one node has outgoing ties to two others, but those two nodes are not connected to each other. This type of triad often reflects hierarchical or source-like structures and is harder to capture when the model does not fully account for asymmetrical dependency patterns. Meanwhile, 021C is the configuration where the ties form a directed cycle among the three nodes. This structure often represents feedback or cyclical interactions, which may be underrepresented or poorly modeled when longer dependencies or complex reciprocal relationships are not explicitly incorporated.

### Poor Estimates for 021U



### Poor Estimates for 021C



## Geodesic Distance Distribution

In the final plot, each dyad's minimum geodesic distance (the shortest path length between two people) is tabulated and grouped, then compared across observed vs. simulated networks. Distances of 1 mean direct ties; distance 2 means you can reach one person via a single intermediary, and so forth.

Distances of 1, 5, 6, 7, 8, and etc match closely between observation and simulation, as indicated by p-values near 1.0. Distance 2 (observed=443 vs. mean=483.53,  $p=0.310$ ) and distance 3 (observed=262 vs. mean=232.94,  $p=0.417$ ) also fall within acceptable ranges. The largest gap arises at distance 4 (ob-

served=82, mean=40.29,  $p=0.126$ ), suggesting the model underestimates the number of paths requiring four steps. Even so, the overall shape of the distance distribution is well captured, meaning the model generally reflects how “far apart” nodes are in this network.

## Conclusion

In conclusion, of all the models, Model 3 provides the most comprehensive fit to the data. The analysis of its estimates demonstrates that both structural dynamics, such as reciprocity and triadic closure, and attribute-based effects, such as gender differences in outgoing activity, significantly shape the formation of strong ties in the network. Notably, reciprocity and outgoing activity emerge as key drivers of friendship ties, while cross-gender friendships are as likely as same-gender friendships, suggesting no strong gender homophily.

At the same time, the GOF analysis demonstrates that the model aligns closely with the observed network statistics for edges, mutual ties, degree distributions, and geodesic distances. It also successfully incorporates gender-related covariates, reflecting both structural and relational patterns within the network.

However, certain challenges remain. The model struggles to capture extreme degree values, such as nodes with very high in-degrees or out-degrees, and rare triadic configurations, such as hierarchical (021U) and cyclical (021C) structures. These discrepancies suggest that additional structural or latent effects (e.g., specific terms for directed asymmetry and cycles) may need to be incorporated in future refinements to improve the model’s explanatory power.

## Bibliography

Borgatti, S. P., Everett, M. G., Johnson, J. C., & Agneessens, F. (2022). *Analyzing social networks using R*. SAGE Publications.