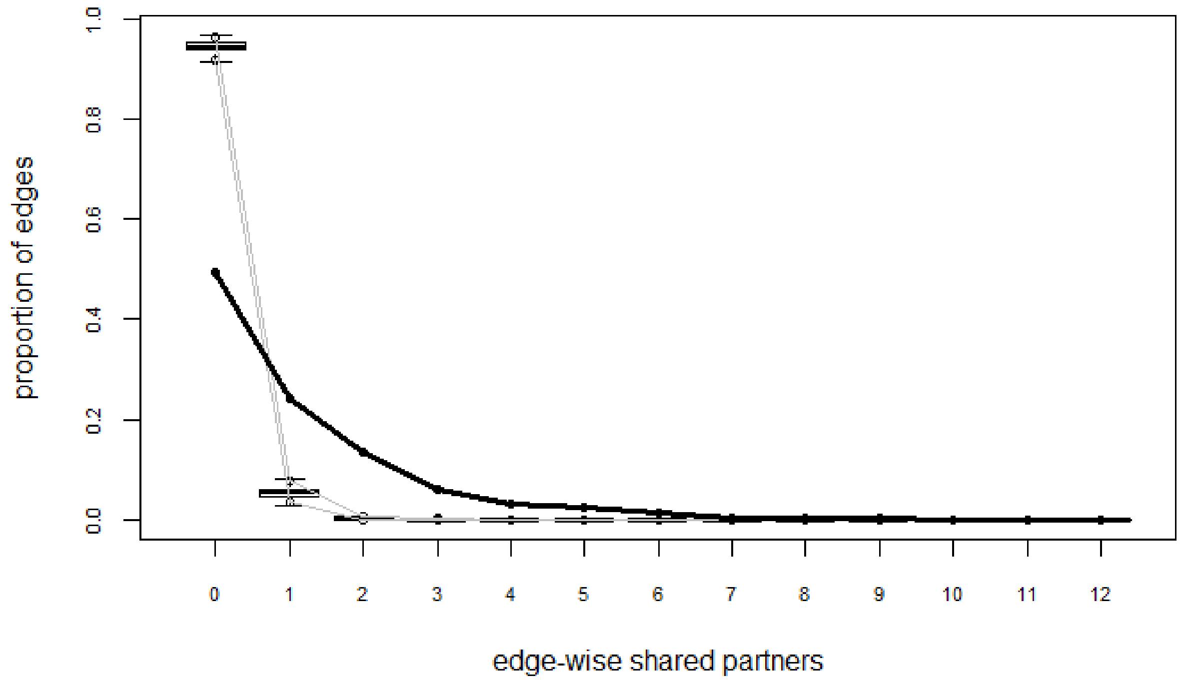
DS8006: Lab 7 "Hypotheses testing with Network Data in R""
Student's name: NAJLIS, BERNARDO (#500744793)

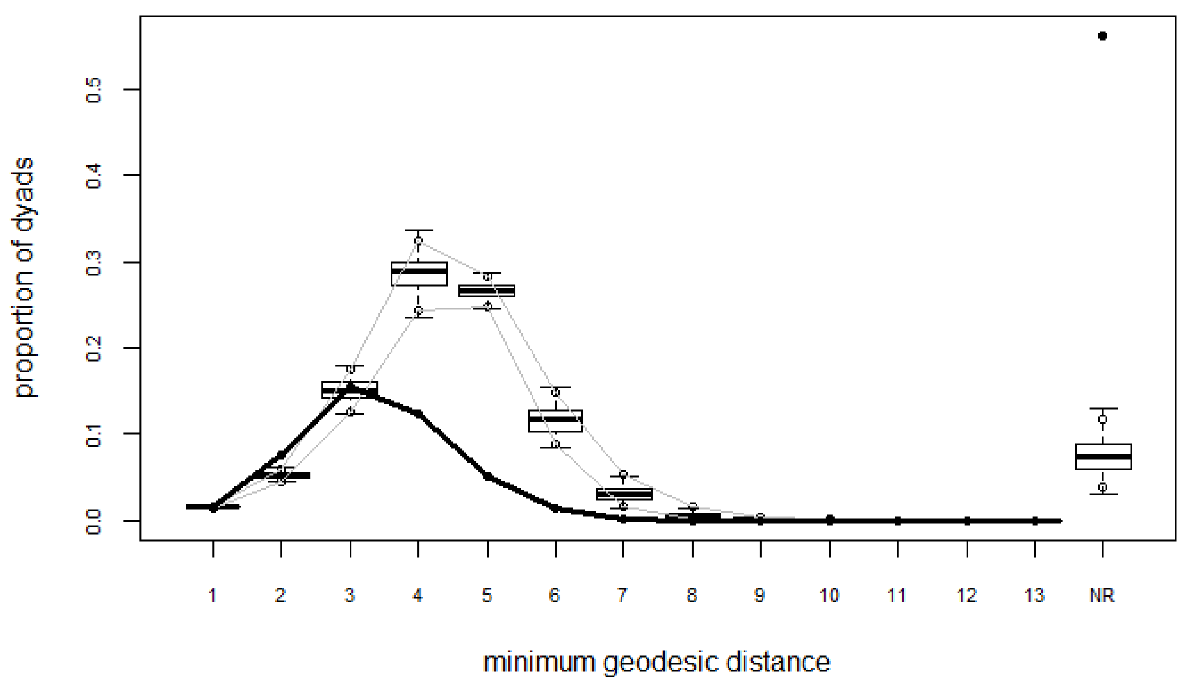
1.

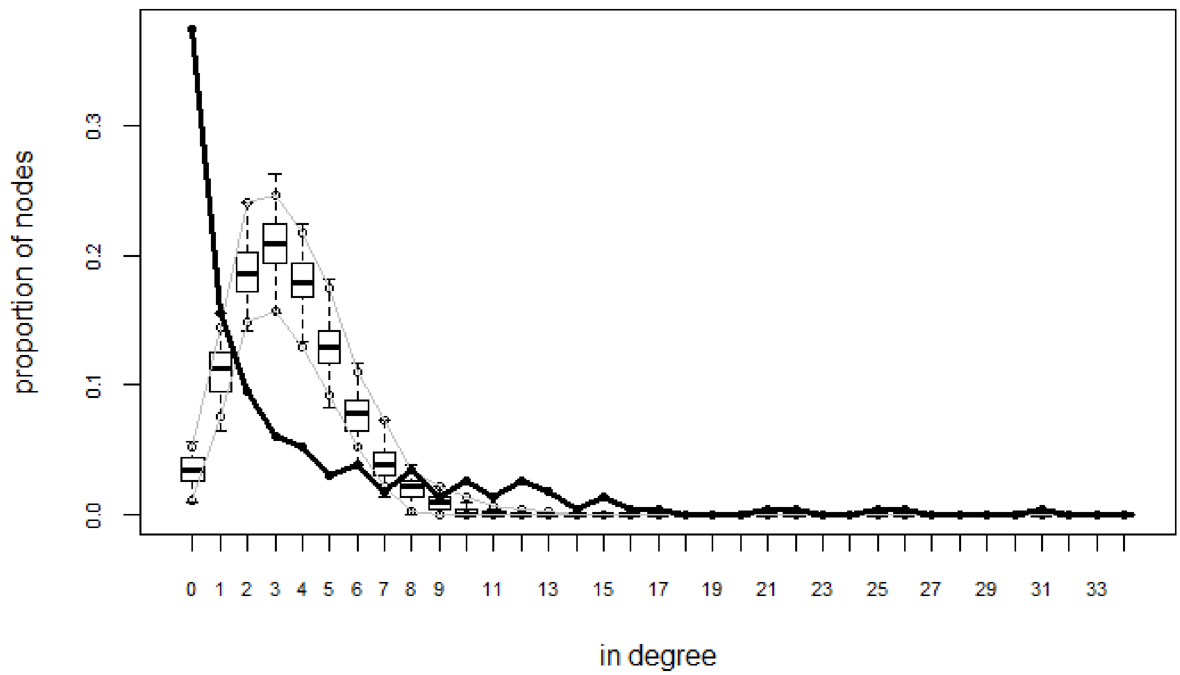
a. Include the ERGM output for the "Role" model.

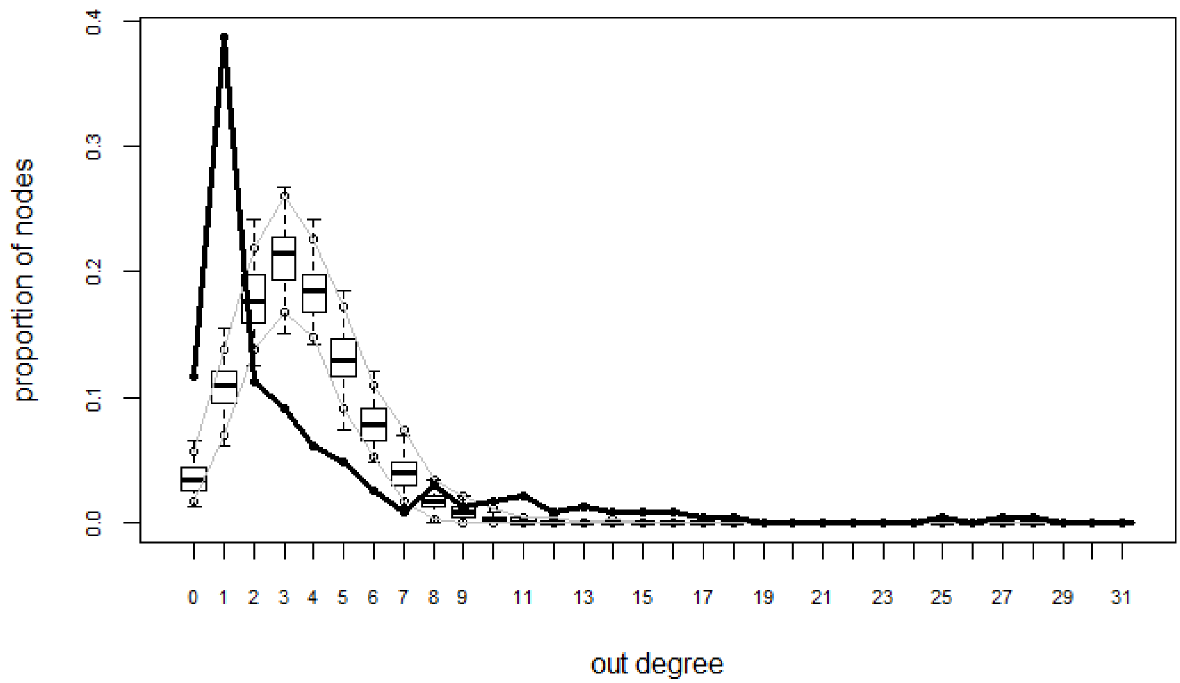
```
> summary(model6)

==========================
Summary of model fit
==========================

Formula:   net ~ edges + nodematch("gender", diff = TRUE, keep = c(1)) +
    nodecov("experience") + nodematch("region", diff = TRUE,
    keep = c(1, 4)) + mutual + nodematch("role1", diff = TRUE,
    keep = c(5, 7, 8, 11))

Iterations:  2 out of 20

Monte Carlo MLE Results:
                               Estimate Std. Error MCMC %  p-value
edges                          -4.98902    0.13589      0  < 1e-04 ***
nodematch.gender.female         0.22169    0.06888      0 0.001289 **
nodecov.experience              0.10292    0.02887      0 0.000364 ***
nodematch.region.International  1.16753    0.25323      0  < 1e-04 ***
nodematch.region.South          0.15636    0.08225      0 0.057307 .
mutual                          2.98152    0.14351      0  < 1e-04 ***
nodematch.role1.libmedia        0.79526    0.34954      0 0.022901 *
nodematch.role1.other           2.09756    0.80493      1 0.009166 **
nodematch.role1.otheredprof     0.58004    0.26967      0 0.031489 *
nodematch.role1.specialed       2.09647    0.35486      1  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 74294  on 53592  degrees of freedom
 Residual Deviance:  8044  on 53582  degrees of freedom

AIC: 8064    BIC: 8153    (Smaller is better.)
```

b. What is the goodness of fit for this model?









c. Explain the result and general implications of the "Role" model (~100 words).

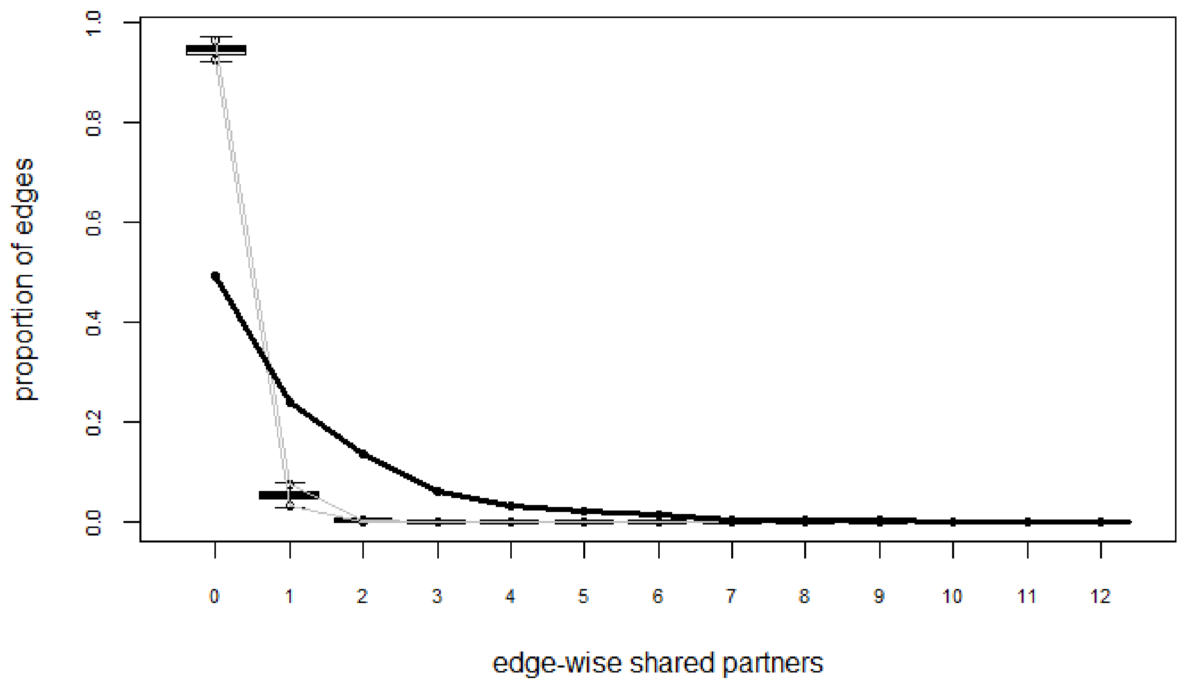
The first attempt to add “role” to the existing model (‘model4’) generated -Inf values for the coefficient estimate with attributes “Operations”, “profdev” and “techinfrastructure” making it impossible to continue the evaluation the model. This required to remove these two attribute values as the first step before even evaluating all other ones for statistical significance.
After removing them, only a couple of roles improve the model (are statistical significant): ‘libmedia’, ‘other’, ‘otherprof’ and ‘specialed’, making AIC somewhat lower but BIC higher.
By looking at the charts generated, goodness-of-fit is still not good when comparing the original network with the simulated network.

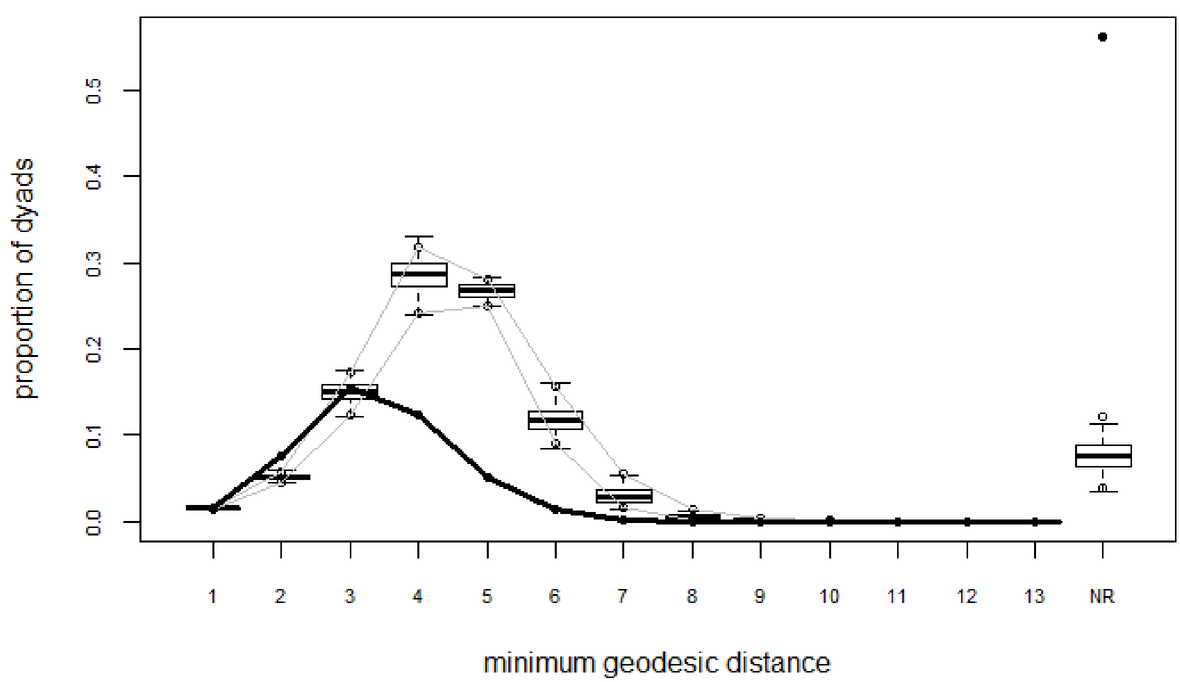
2.

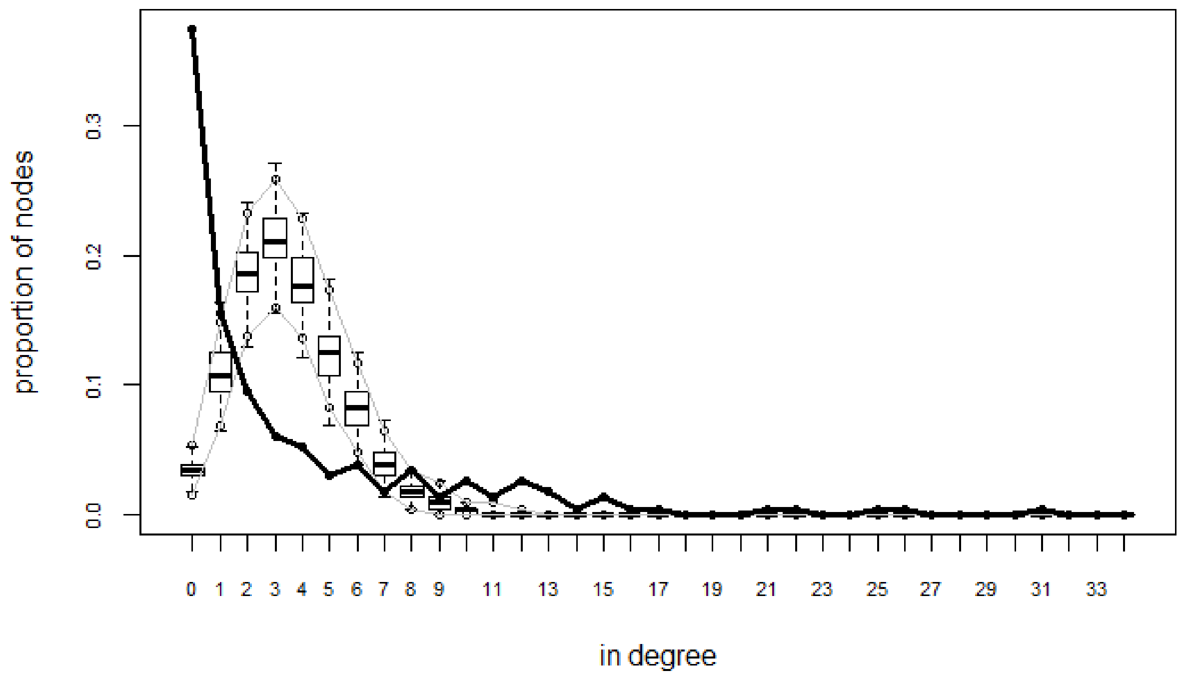
a. Include the ERGM output for the “<u>Grades</u>” model.

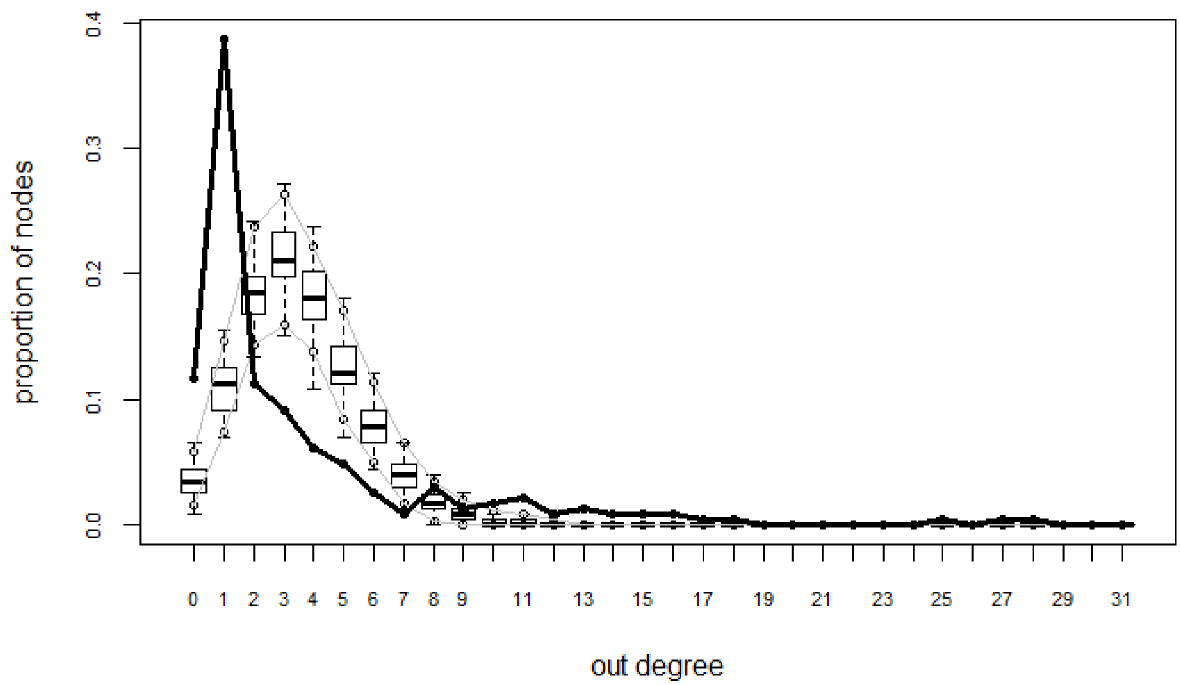
```
> summary(model7)

==========================
Summary of model fit
==========================

Formula:   net ~ edges + nodematch("gender", diff = TRUE, keep = c(1)) + 
    nodecov("experience") + nodematch("region", diff = TRUE, 
    keep = c(1, 4)) + mutual + nodematch("role1", diff = TRUE, 
    keep = c(5, 7, 8, 11)) + nodematch("grades", diff = TRUE, 
    keep = c(1, 2, 4))

Iterations:  8 out of 20 

Monte Carlo MLE Results:
                               Estimate Std. Error MCMC %  p-value    
edges                          -4.94324    0.13321      0  < 1e-04 ***
nodematch.gender.female         0.20505    0.06958      0 0.003211 ** 
nodecov.experience              0.09921    0.02888      0 0.000592 ***
nodematch.region.International  1.07671    0.24328      0  < 1e-04 ***
nodematch.region.South          0.14094    0.08085      0 0.081287 .  
mutual                          2.96859    0.12925      0  < 1e-04 ***
nodematch.role1.libmedia        0.70729    0.30563      0 0.020661 *  
nodematch.role1.other           1.90779    0.84997      1 0.024802 *  
nodematch.role1.otheredprof     0.61389    0.25572      0 0.016371 *  
nodematch.role1.specialed       2.00974    0.59086      0 0.000671 ***
nodematch.grades.college        1.36103    0.44053      1 0.002006 ** 
nodematch.grades.generalist    -0.19573    0.09386      0 0.037038 *  
nodematch.grades.primary        0.35778    0.15104      0 0.017848 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 74294  on 53592  degrees of freedom
 Residual Deviance:  8031  on 53579  degrees of freedom
 
AIC: 8057    BIC: 8172    (Smaller is better.)
```

b. What is the goodness of fit for this model?









c. Explain the result and general implications of the "<u>Grades</u>" model (~100 words).
For this model, we first added all possible values for "grades" and then only kept the ones that are statistically significant: 'college', 'generalist' and 'primary'. This made AIC

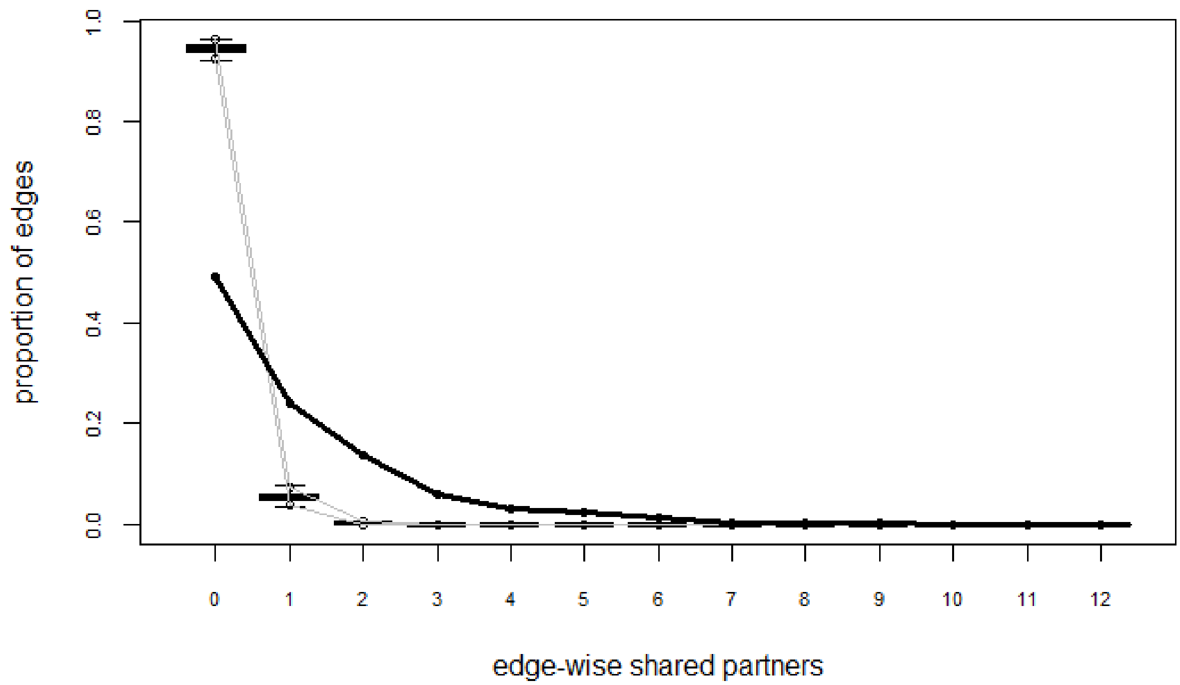
stay the same as the previous model with BIC dropping from 8191 to 8172 (still not back at the value of 8129 we had at 'model4').
As for goodness-of-fit, there is no significant variance between the charts generated with the previous model and this one, and the difference with the original network remains visually the same.

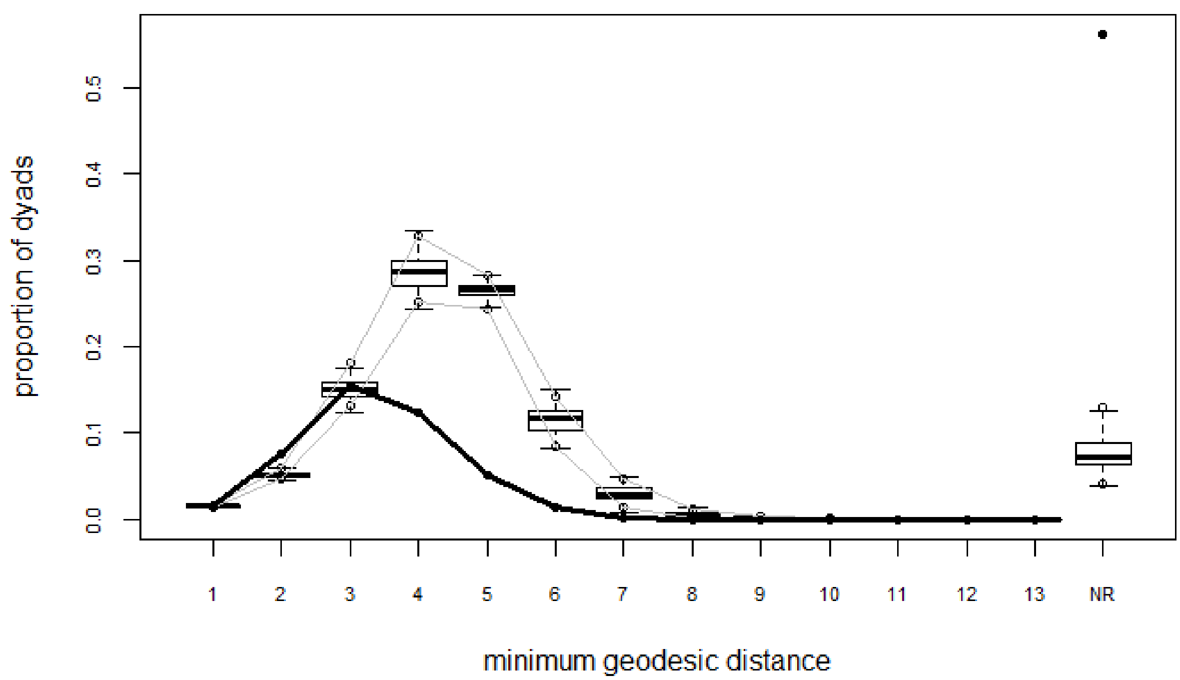
3.

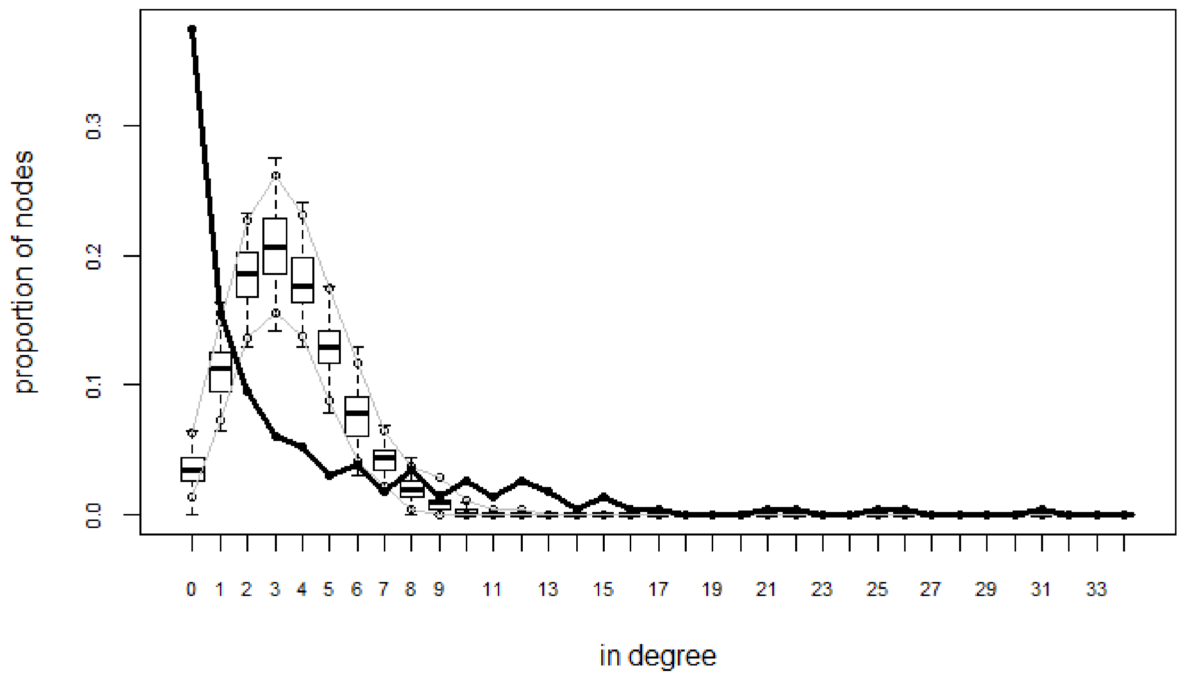
a. Include the ERGM output for the "<u>Expert</u>" model.

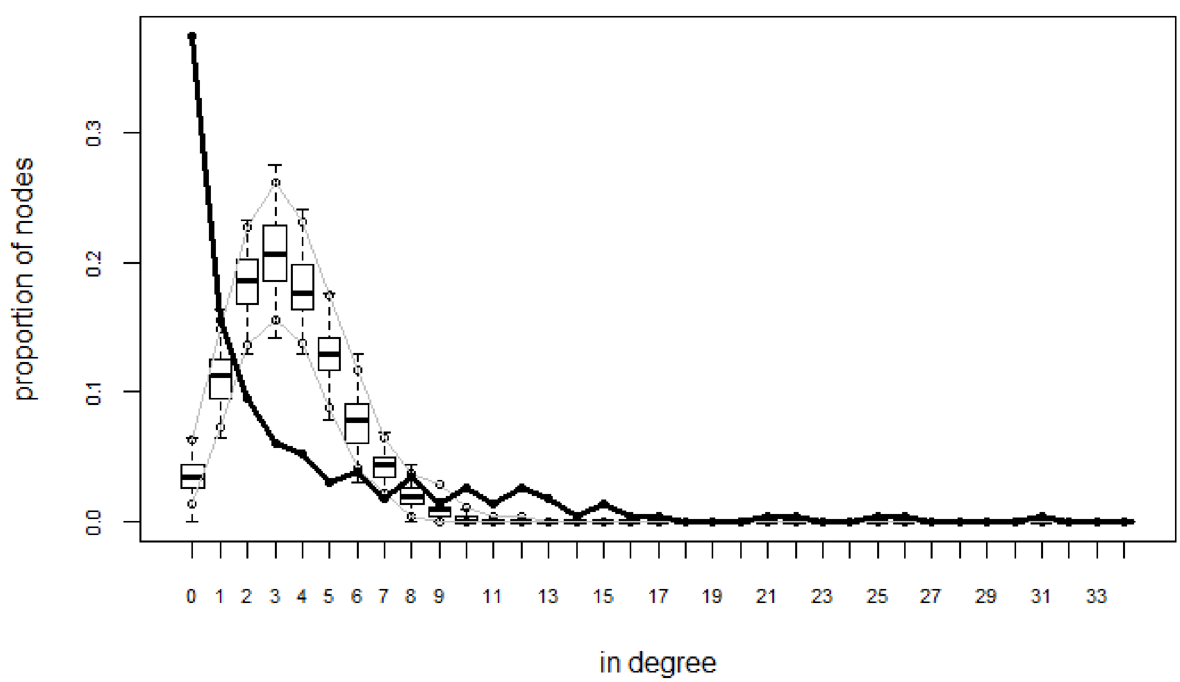
```
> summary(model8)

==========================
Summary of model fit
==========================

Formula:   net ~ edges + nodematch("gender", diff = TRUE, keep = c(1)) +
    nodecov("experience") + nodematch("region", diff = TRUE,
    keep = c(1, 4)) + mutual + nodematch("role1", diff = TRUE,
    keep = c(5, 7, 8, 11)) + nodematch("grades", diff = TRUE,
    keep = c(1, 2, 4)) + nodecov("expert")

Iterations:  2 out of 20

Monte Carlo MLE Results:
                               Estimate Std. Error MCMC %  p-value
edges                          -4.92956    0.13851      0  < 1e-04 ***
nodematch.gender.female         0.21491    0.06798      0 0.001571 **
nodecov.experience              0.09777    0.02999      0 0.001116 **
nodematch.region.International  1.09768    0.29922      0 0.000244 ***
nodematch.region.South          0.12745    0.08106      0 0.115891
mutual                          2.96748    0.14488      0  < 1e-04 ***
nodematch.role1.libmedia        0.75348    0.30926      0 0.014837 *
nodematch.role1.other           1.97033    0.79910      1 0.013678 *
nodematch.role1.otheredprof     0.62078    0.25665      0 0.015576 *
nodematch.role1.specialed       1.92285    0.48269      1  < 1e-04 ***
nodematch.grades.college        1.59627    0.70312      0 0.023195 *
nodematch.grades.generalist    -0.19301    0.09596      0 0.044299 *
nodematch.grades.primary        0.36117    0.14654      0 0.013715 *
nodecov.expert                 -0.06887    0.09605      0 0.473345
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 74294  on 53592  degrees of freedom
 Residual Deviance:  8029  on 53578  degrees of freedom

AIC: 8057    BIC: 8181    (Smaller is better.)
```

b. What is the goodness of fit for this model?


proportion of edges
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8
9
10
11
12
edge-wise shared partners


proportion of dyads
0.5
0.4
0.3
0.2
0.1
0.0
1
2
3
4
5
6
7
8
9
10
11
12
13
NR
minimum geodesic distance





c. Explain the result and general implications of the "<u>Expert</u>" model (~100 words).
For this model, we kept the attributes from the previous one but added 'expert'. This had two results: 'expert' is not statistically significant, and also adding it makes the 'region.South' as not statistically significant either. Even though this model doesn't lead

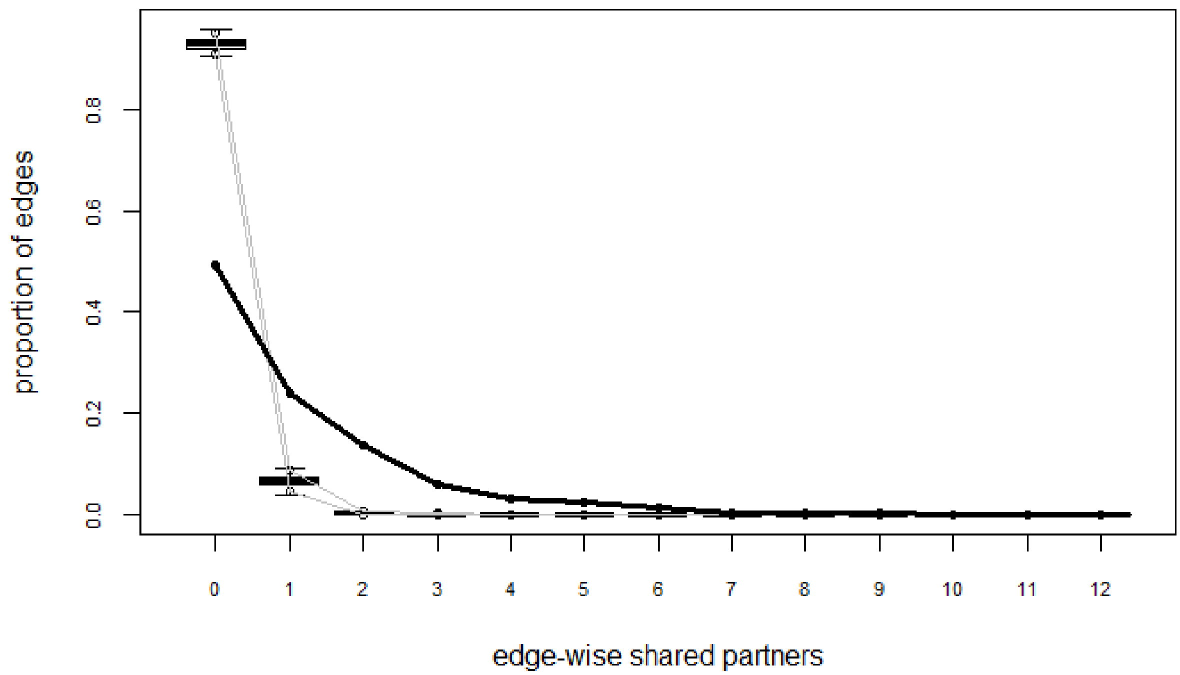
to an improvement, we calculated metrics and goodness-of-fit. AIC remains the same as from model7 at 8057, and BIC goes higher from 8172 to 8181. Goodness-of-fit charts doesn't show any changes from previous models.
As the addition of 'expert' is irrelevant, we will remove it from the next model but we will keep 'region.South'.

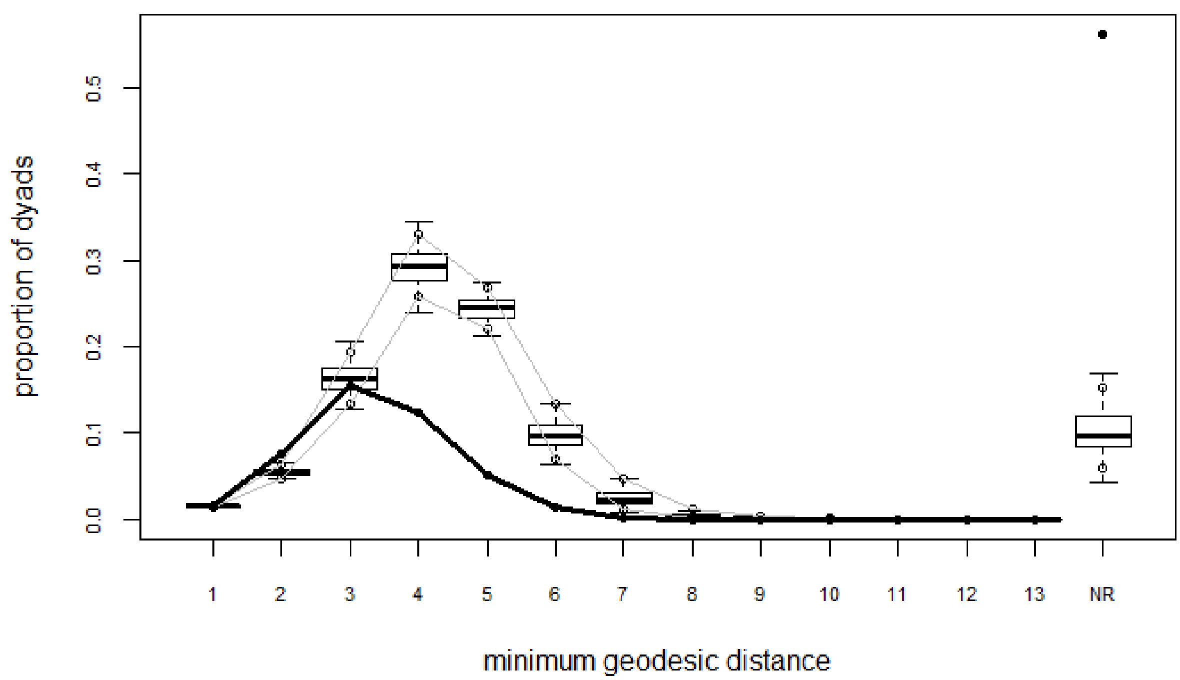
4.

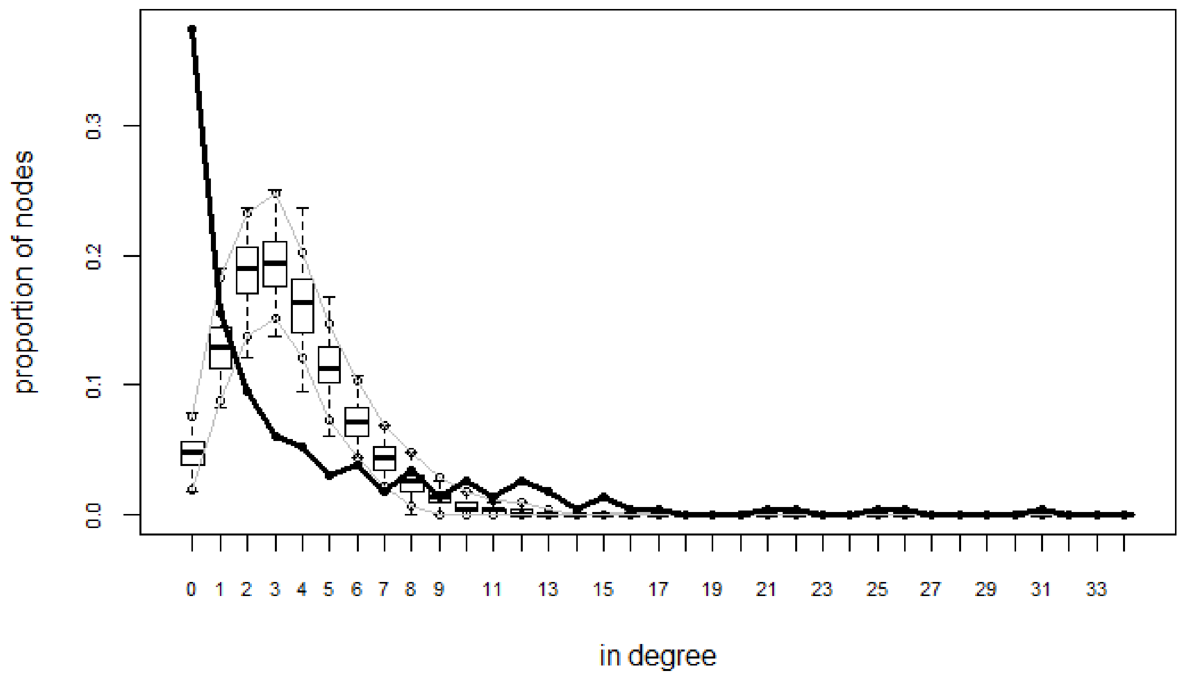
a. Include the ERGM output for the "Connect" model.

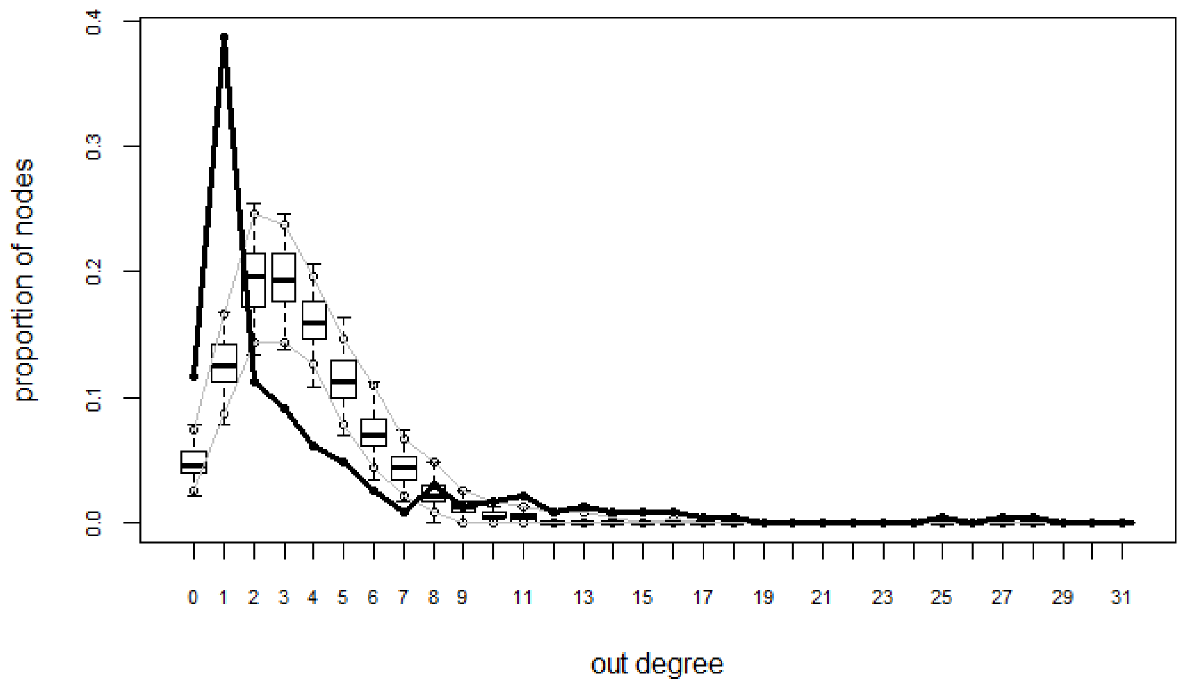
```
> summary(model9)

==========================
Summary of model fit
==========================

Formula:   net ~ edges + nodematch("gender", diff = TRUE, keep = c(1)) +
    nodecov("experience") + nodematch("region", diff = TRUE,
    keep = c(1, 4)) + mutual + nodematch("role1", diff = TRUE,
    keep = c(5, 7, 8, 11)) + nodematch("grades", diff = TRUE,
    keep = c(1, 2, 4)) + nodecov("connect")

Iterations:  2 out of 20

Monte Carlo MLE Results:
                               Estimate Std. Error MCMC % p-value
edges                          -4.99493    0.13699      0 < 1e-04 ***
nodematch.gender.female         0.21306    0.06629      0 0.00131 **
nodecov.experience              0.05482    0.02967      0 0.06466 .
nodematch.region.International  0.85960    0.29919      0 0.00407 **
nodematch.region.South          0.08481    0.08028      0 0.29077
mutual                          2.85178    0.13899      0 < 1e-04 ***
nodematch.role1.libmedia        0.86248    0.31618      0 0.00638 **
nodematch.role1.other           2.41555    1.33461      0 0.07031 .
nodematch.role1.otheredprof     0.48721    0.25422      0 0.05530 .
nodematch.role1.specialed       2.09265    0.53109      1 < 1e-04 ***
nodematch.grades.college        1.47151    0.55544      1 0.00807 **
nodematch.grades.generalist    -0.25996    0.09472      0 0.00607 **
nodematch.grades.primary        0.60178    0.15408      0 < 1e-04 ***
nodecov.connect                 0.53176    0.05258      0 < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 74294  on 53592  degrees of freedom
 Residual Deviance:  7931  on 53578  degrees of freedom

AIC: 7959    BIC: 8084    (Smaller is better.)
```

b. What is the goodness of fit for this model?


proportion of edges
0.0
0.2
0.4
0.6
0.8
0 1 2 3 4 5 6 7 8 9 10 11 12
edge-wise shared partners


proportion of dyads
0.0
0.1
0.2
0.3
0.4
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 NR
minimum geodesic distance



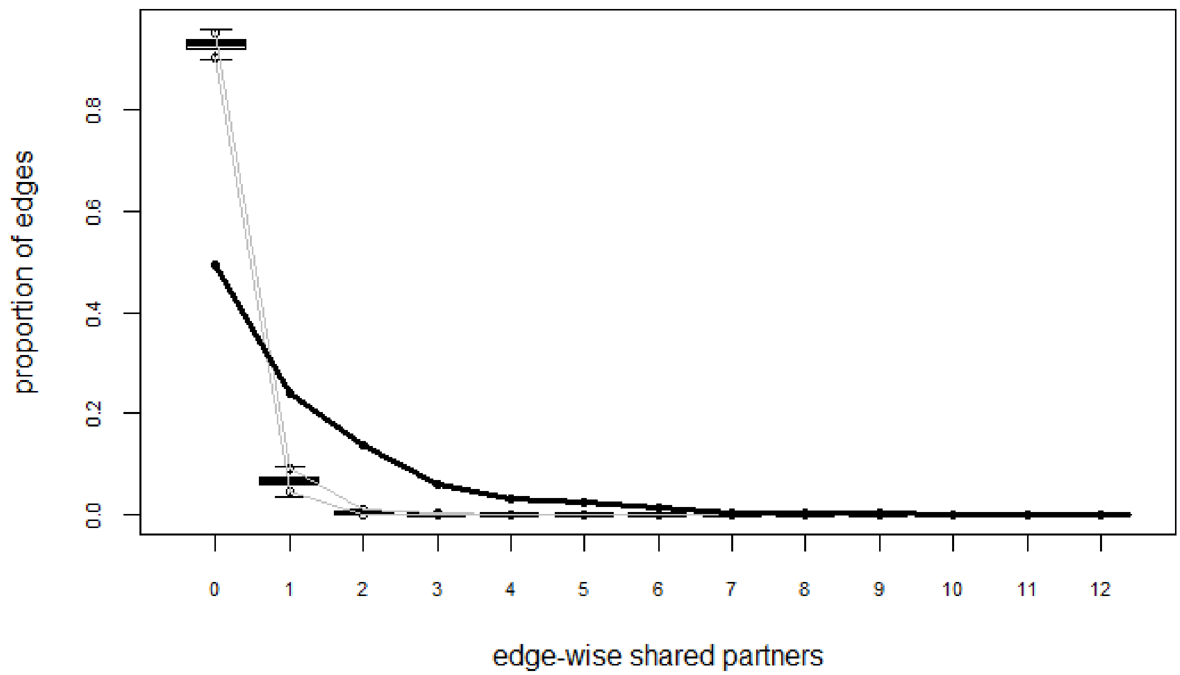


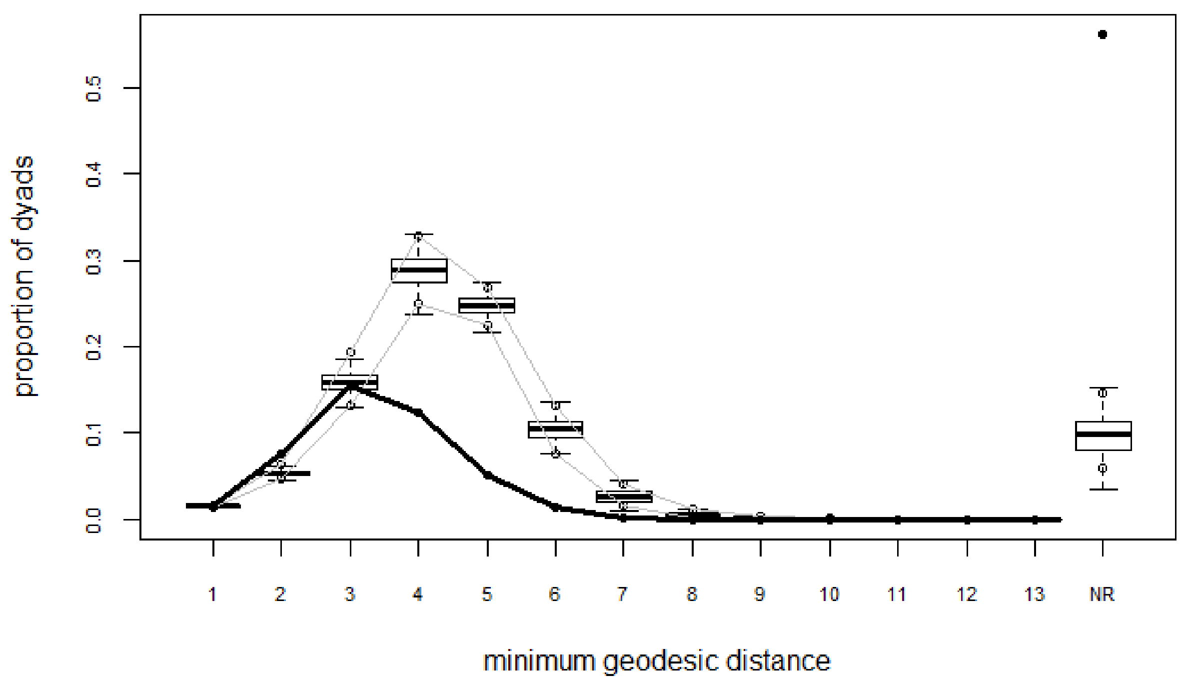
c. Explain the result and general implications of the "<u>Connect</u>" model (~100 words).
By removing 'expert' and adding 'connect', we see the first somewhat significant drop in AIC in the last couple iterations (from 8057 to 7959) and a small drop in BIC (from 818 to

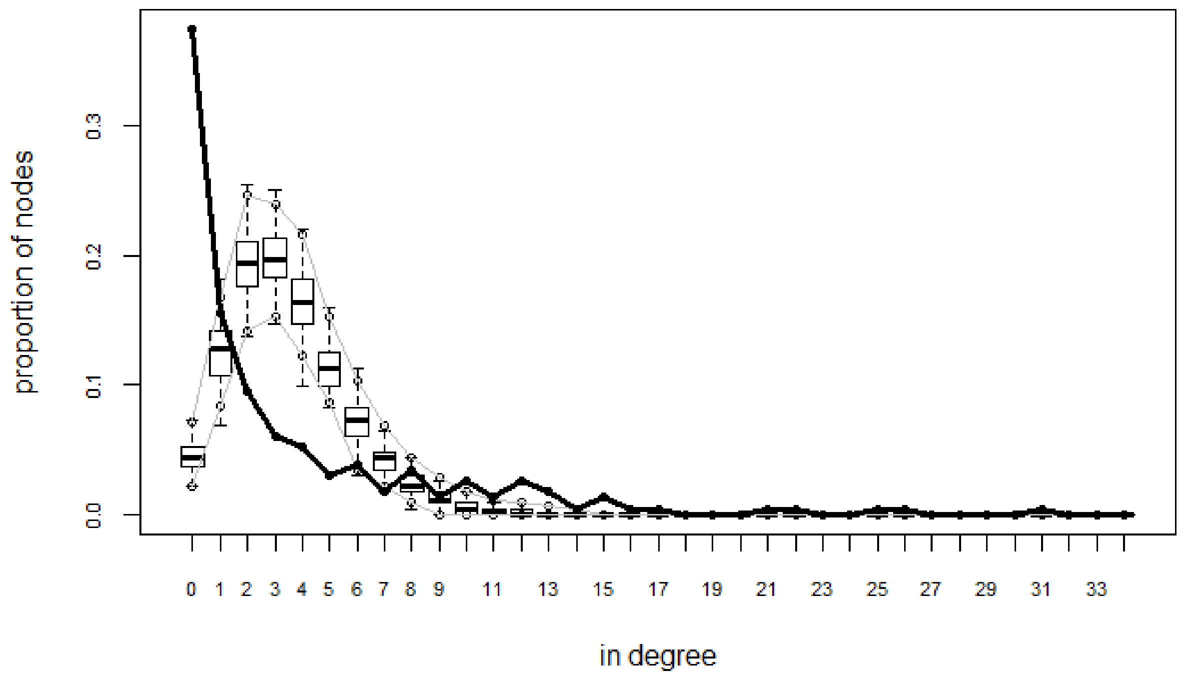
8084). As in the previous model, now we see that “region.South” it is not longer statistically significant in this model, which means it can be removed from the model. I created an additional model that removes the region (results in the next pages) and see slight changes in AIC (from 7959 to 7960) and BIC (from 8084 to 8086).

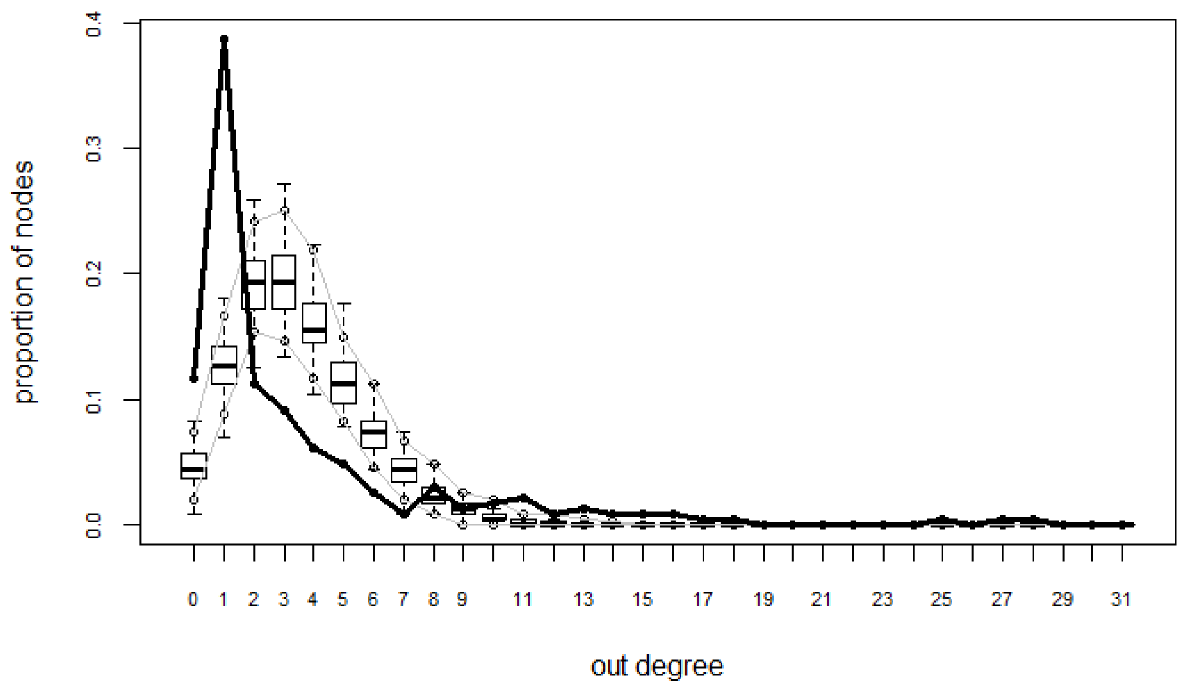
Same as with previous cases, the charts for goodness-of-fit don’t show changes and remain like all the previous cases.

```
> summary(model10)

==========================
Summary of model fit
==========================

Formula:   net ~ edges + nodematch("gender", diff = TRUE, keep = c(1)) +
    nodecov("experience") + nodematch("region", diff = TRUE,
    keep = c(1)) + mutual + nodematch("role1", diff = TRUE, keep = c(5,
    7, 8, 11)) + nodematch("grades", diff = TRUE, keep = c(1,
    2, 4)) + nodecov("connect")

Iterations:  2 out of 20

Monte Carlo MLE Results:
                               Estimate Std. Error MCMC %  p-value
edges                          -4.97341    0.13394      0  < 1e-04 ***
nodematch.gender.female         0.21811    0.06445      0 0.000714 ***
nodecov.experience              0.05351    0.02914      0 0.066258 .
nodematch.region.International  0.79989    0.29924      0 0.007519 **
mutual                          2.84445    0.13831      0  < 1e-04 ***
nodematch.role1.libmedia        0.80092    0.32620      0 0.014079 *
nodematch.role1.other           2.32044    1.07456      0 0.030821 *
nodematch.role1.otheredprof     0.54949    0.26499      0 0.038117 *
nodematch.role1.specialed       2.84340    0.82868      0 0.000601 ***
nodematch.grades.college        1.94684    0.71214      0 0.006263 **
nodematch.grades.generalist    -0.26464    0.09287      0 0.004382 **
nodematch.grades.primary        0.61739    0.15792      0  < 1e-04 ***
nodecov.connect                 0.52672    0.05263      0  < 1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 74294  on 53592  degrees of freedom
 Residual Deviance:  7934  on 53579  degrees of freedom

AIC: 7960    BIC: 8076    (Smaller is better.)
```


proportion of edges
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
12
edge-wise shared partners


proportion of dyads
0.0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
5
6
7
8
9
10
11
12
13
NR
minimum geodesic distance


proportion of nodes
0.0
0.1
0.2
0.3
0 1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 23 25 27 29 31 33
in degree


proportion of nodes
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 23 25 27 29 31
out degree

Here is a chart with the changes of AIC and BIC for all the models, and in the next page a table that summarizes all models, their characteristics and their values.

Model Name	Model Description	Attribute Added	R Code	AIC	BIC
model0	'Null' model	Null	`ergm(net ~edges)`	8418	8427
model1	Differential Homophily based on Gender	Gender	`ergm(net ~edges +  nodematch('gender',diff=T))`	8402	8428
model2	Differential Homophily based on Gender (just for "female") and experience	Experience	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience"))`	8391	8418
model3	Differential Homophily based on Gender (just for "female"), experience and region.	All Regions	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUE))`	8337	8448
model3	Differential Homophily based on Gender (just for "female"), experience and region (just 'International' and 'South').	Region	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUEkeep=c(1,4))`	8374	8419
model4	Differential Homophily based on Gender (just for "female"), experience and region (just 'International' and 'South') and mutuality.	Mutual	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUEkeep=c(1,4) + mutual)`	8076	8129
model5	Differential Homophily based on Gender (just for "female"), experience and region (just 'International' and 'South'), mutuality and and roles.	All Roles	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUE, keep=c(1,4)) + mutual + nodematch("role1", diff=TRUE)`	N/A	N/A
model5	Differential Homophily based on Gender (just for "female"), experience and region (just 'International' and 'South'), mutuality and roles (libmedia, other, otheredprof, specialed).	Role	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUE, keep=c(1,4)) + mutual + nodematch("role1", diff=TRUE, keep=c(5,7,8,11))`	8064	8153

model7	Differential Homophily based on Gender (just for "female"), experience and region (just 'International' and 'South'), mutuality, roles (libmedia, other, otheredprof, specialed) and all grades.	All Grades	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUE, keep=c(1,4)) + mutual + nodematch("role1", diff=TRUE, keep=c(5,7,8,11) + nodematch("grades", diff=TRUE))`	8057	8191
model7	Differential Homophily based on Gender (just for "female"), experience and region (just 'International' and 'South'), mutuality, roles (libmedia, other, otheredprof, specialed) and grades (college, generalist, primary).	Grades	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUE, keep=c(1,4)) + mutual + nodematch("role1", diff=TRUE, keep=c(5,7,8,11) + nodematch("grades", diff=TRUE, keep=c(1,2,4))`	8057	8172
model8	Differential Homophily based on Gender (just for "female"), experience and region (just 'International' and 'South'), mutuality, roles (libmedia, other, otheredprof, specialed), grades (college, generalist, primary) and expert.	Expert	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUE, keep=c(1,4)) + mutual + nodematch("role1", diff=TRUE, keep=c(5,7,8,11) + nodematch("grades", diff=TRUE, keep=c(1,2,4) + nodecov("expert"))`	8057	8181
model9	Differential Homophily based on Gender (just for "female"), experience and region (just 'International' and 'South'), mutuality, roles (libmedia, other, otheredprof, specialed), grades (college, generalist, primary), no expert and connect.	Connect	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUE, keep=c(1,4)) + mutual + nodematch("role1", diff=TRUE, keep=c(5,7,8,11) + nodematch("grades", diff=TRUE, keep=c(1,2,4) + nodecov("connect"))`	7959	8084

model10	Differential Homophily based on Gender (just for "female"), experience and region (just 'International'), mutuality, roles (libmedia, other, otheredprof, specialed), grades (college, generalist, primary), no expert and connect.	Connect with no South region	`ergm(net ~ edges + nodematch("gender", diff=TRUE, keep=c(1)) + nodecov("experience") + nodematch("region", diff=TRUE, keep=c(1)) + mutual + nodematch("role1", diff=TRUE, keep=c(5,7,8,11) + nodematch("grades", diff=TRUE, keep=c(1,2,4) + nodecov("connect"))`	7960	8086

5. What was the most challenging part of this lab? And what did you learn? (Your answer to this question should be at least 150 words).

The most challenging part of the lab was figuring out what to do with the model that generated -Inf values for the coefficient estimates, what to do when adding variables renders other variable to be non statistically significant, and also how to understand and interpret changes in the goodness-of-fit charts generates. Both items were addressed in the reference documentation provided in class, so it was just a matter of researching and reading to solve these.

In terms of learning, I learned about the concept of ergm models and how it can be used for networks in general. Also, I learned about goodness-of-fit and how to use it to compare networks with simulated runs from models. The whole concept of ergms and how they can be used to model and simulate different types of networks in different situations really sparked my intereste and I kept reading about this topic and its application for many other scenarios that are not just in the context of social networks.