**Network Homework5**

**Chenrui Xu**

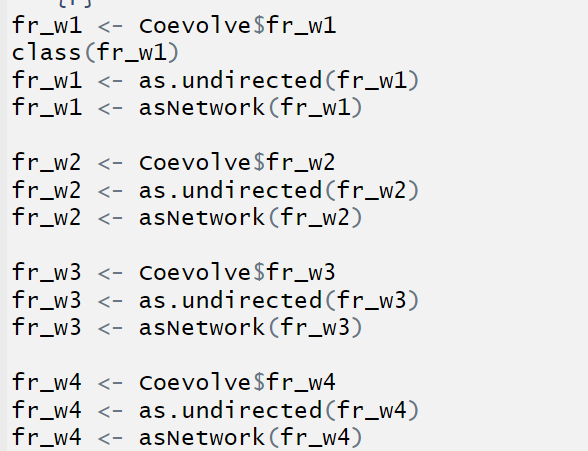
**I Background**

This time we are going to build some EGRMs. The EGRM, which is the short form of exponential random graph model. An EGRM is a true generative statistical model of network structure and characteristics. It is used to handle some complex dependencies of networks. Also, it handles different types of predictors/covariates to predict overall the network characteristics. It is used through well-implemented software. ERGMs are fit using Monte Carlo Markov Chain (MCMC) maximum-likelihood estimation (MLE).

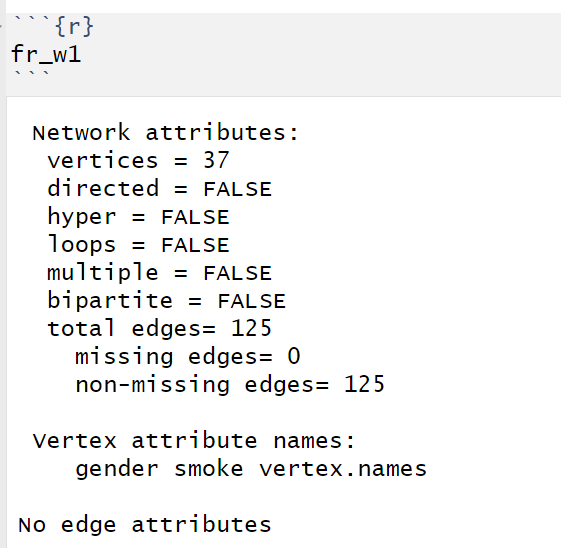
In this project, we are going to use EGRMs in terms of different kinds of data and variables. By comparing the p-values and AIC/BICs, we can get the most significant model.

The data is a network data with four different times. It describes the network of smokers and non-smokers and how those smokers effect non-smokers.

**II Data I**

****

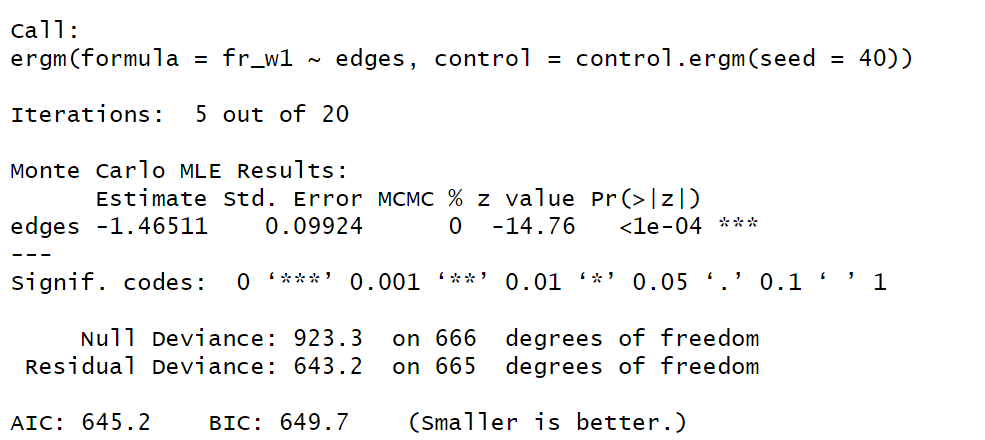
After loading data from package, I turned the data from ‘igraph’ format to network format so that we can continue to the next step.



By picking one dataset, we can find that there are two attributes of vertex. They are gender and smoke. Both of them are factors.

**Model 1**

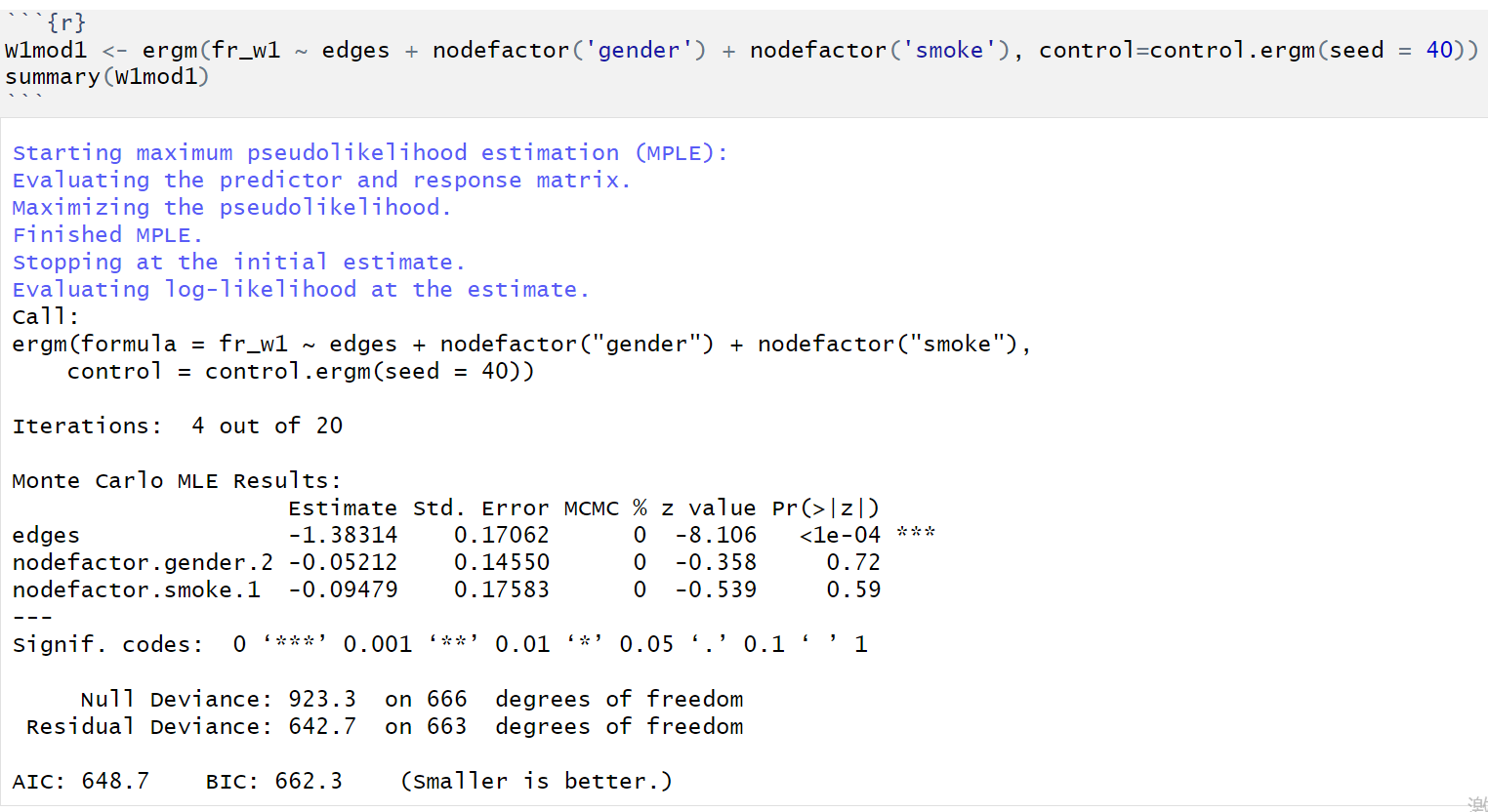
So, at first, we’ll build null model.



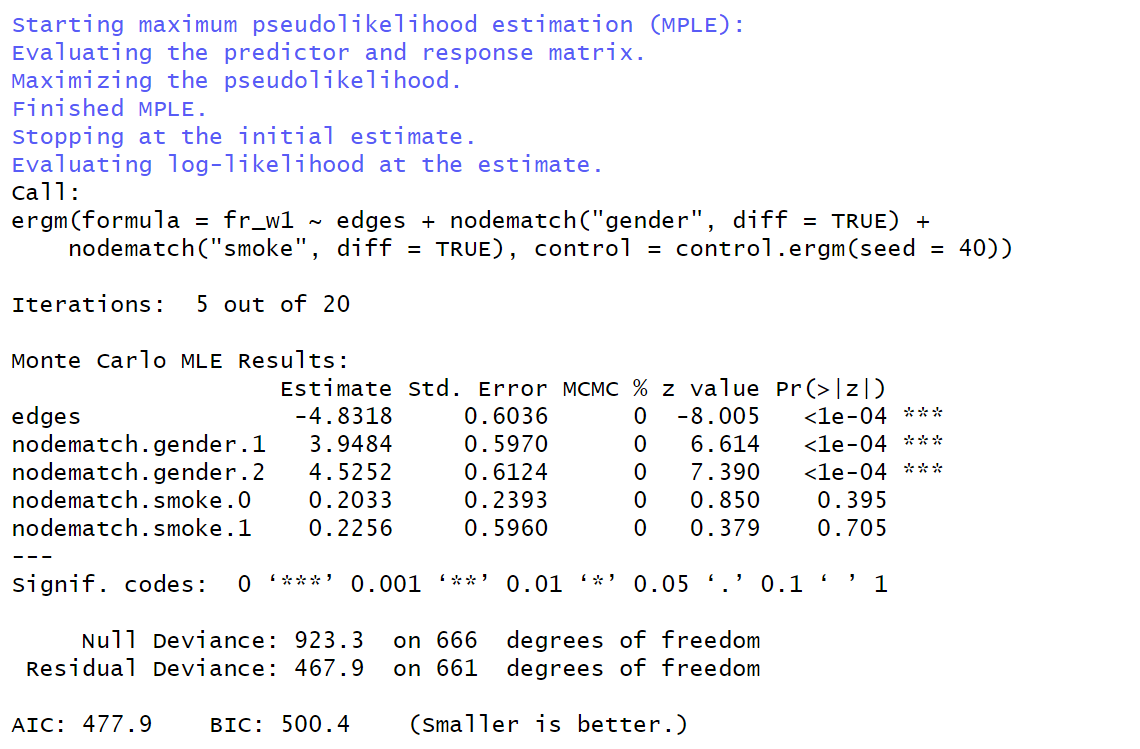
From the model summary, we can see that the AIC is 645 and BIC is 649.

**Model 2**

Then, we will add two factors to the model.



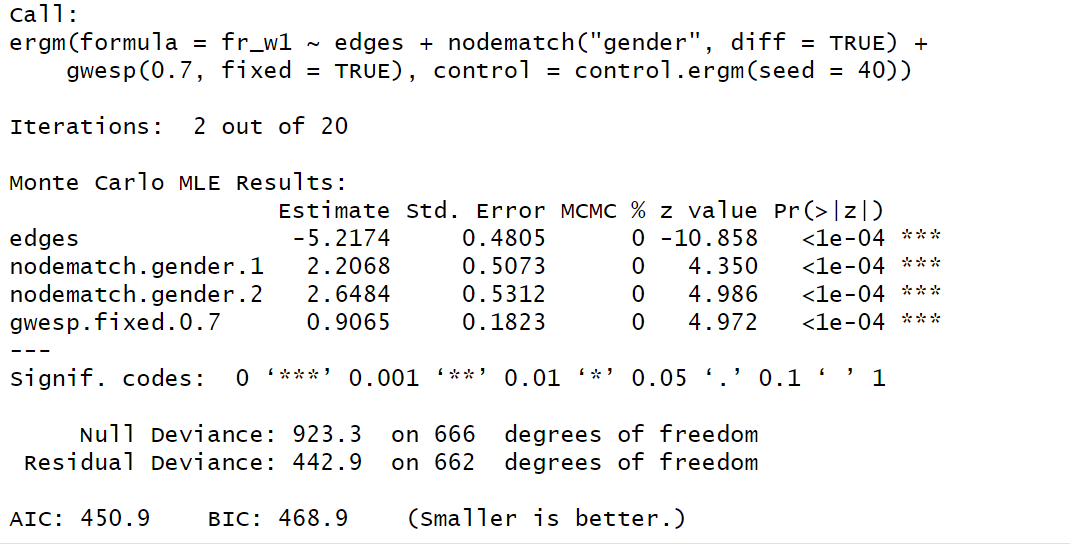
However, from the statistics, it is not hard to find that both of them are not significant, so we do not need to add them. Instead, I’ll use nodematch() function.

****

From this model, we can get the AIC/BIC equal 477.9/500.4 and only gender node match is significant. So next step we should delete smoke.

**Model 3**

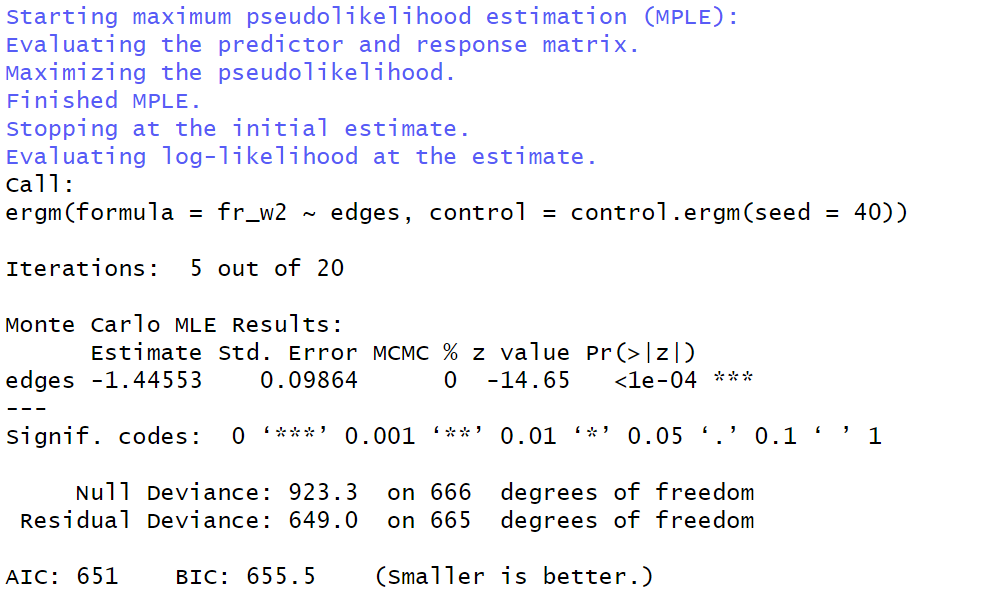
In this model, we will add gwesp.



The AIC/BIC are much smaller this time and all variables are very significant. So, model 3 is the best model given data 1.

**III Data II**

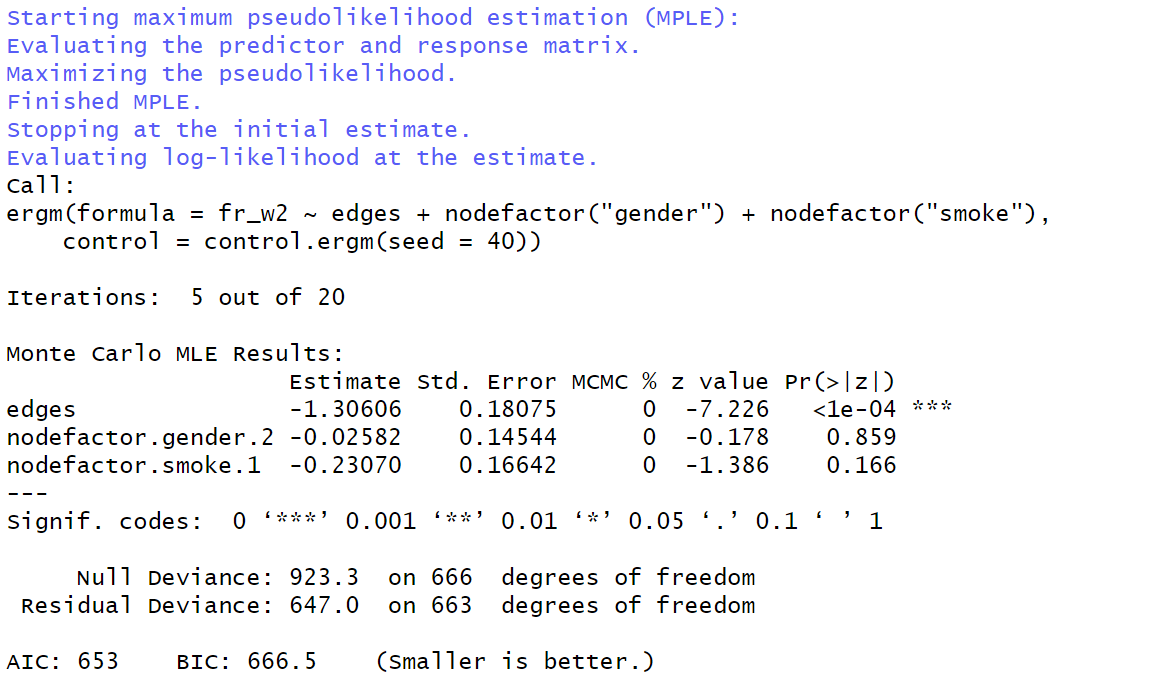
**Model 1**

****

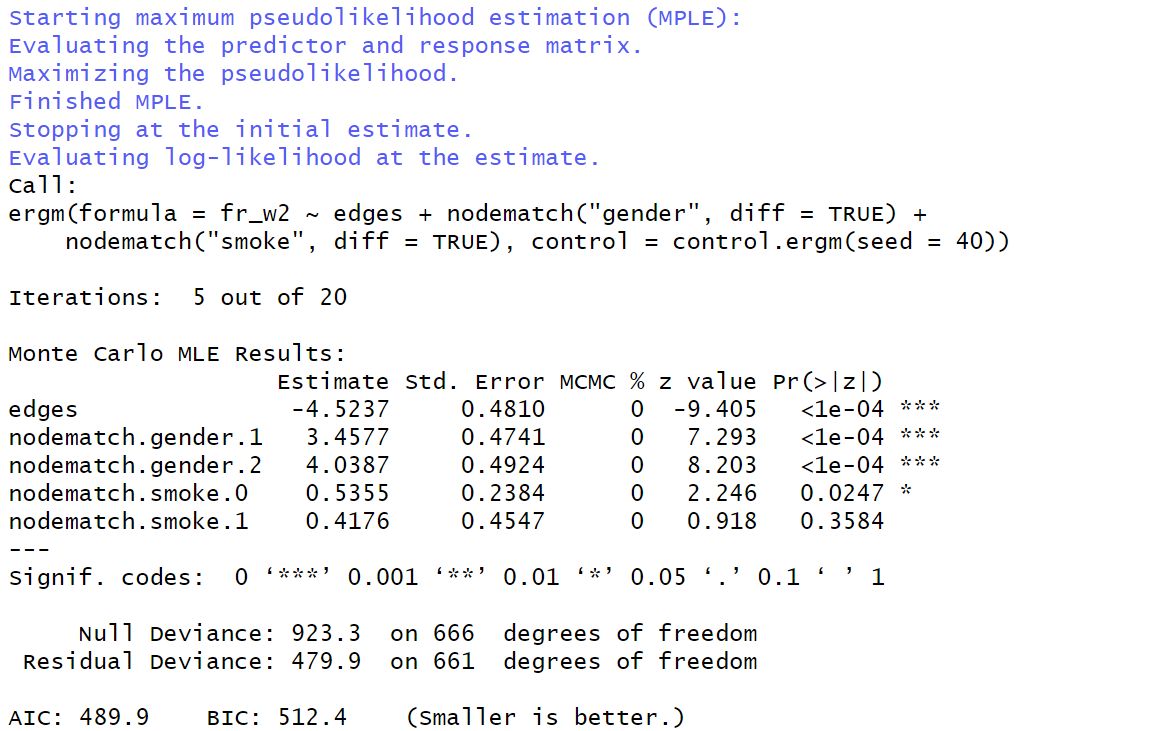
Here is the null model for data 2.

**Model 2**

At first, we take the look at the factor addition.

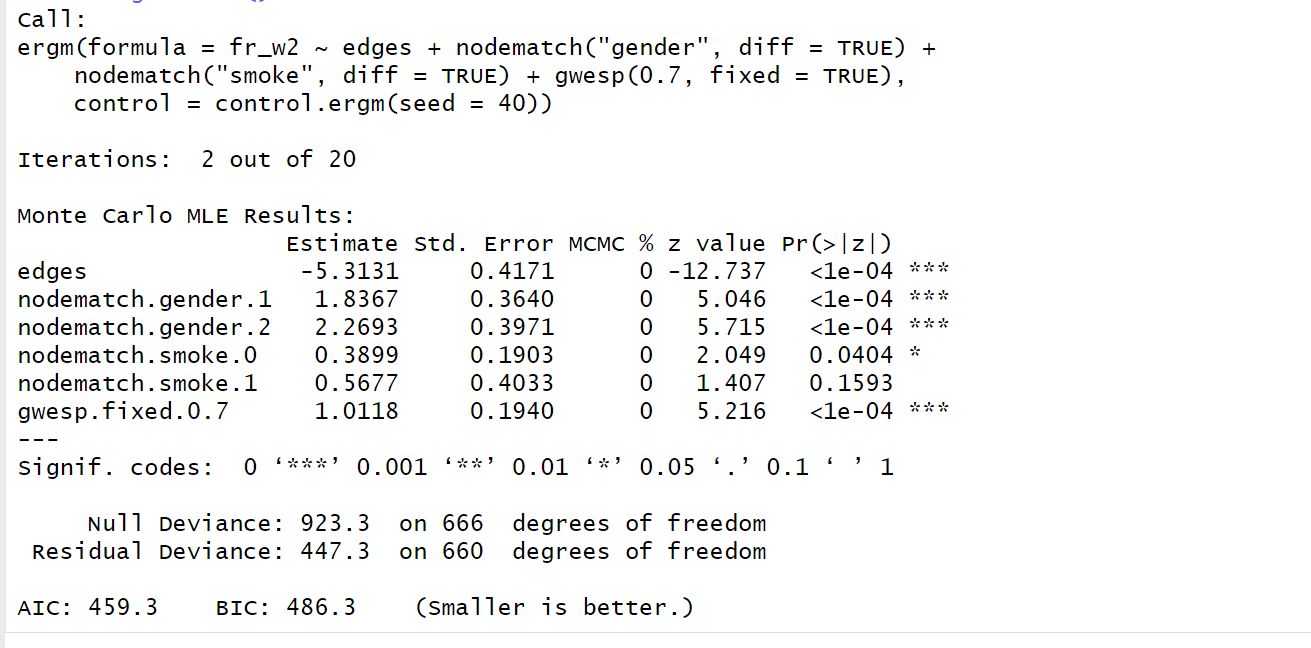


They are not significant.



So, there is the model 2. AIC and BIC are smaller. Half of the smoke match is not significant but the rest is significant.

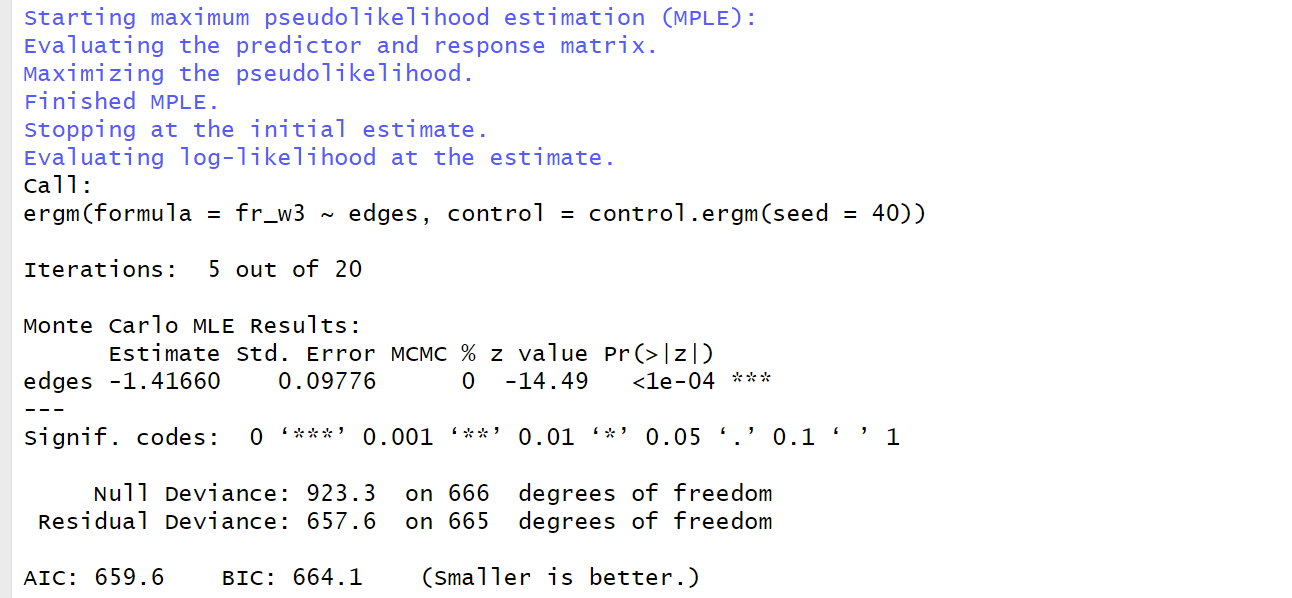
**Model 3**

****

Here is the result if we keep all match variables. The AIC/BIC is the smallest one. So, the model 3 is the best one.

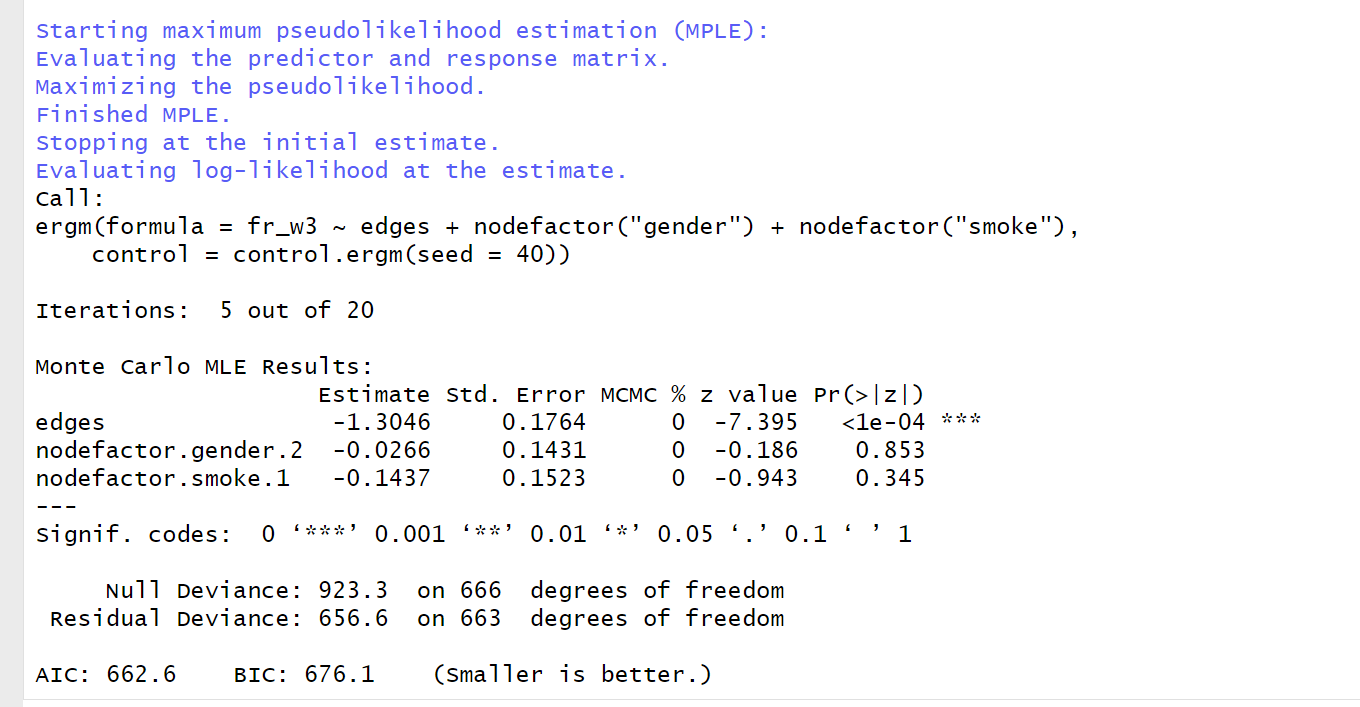
**IV Data III**

**Model 1**

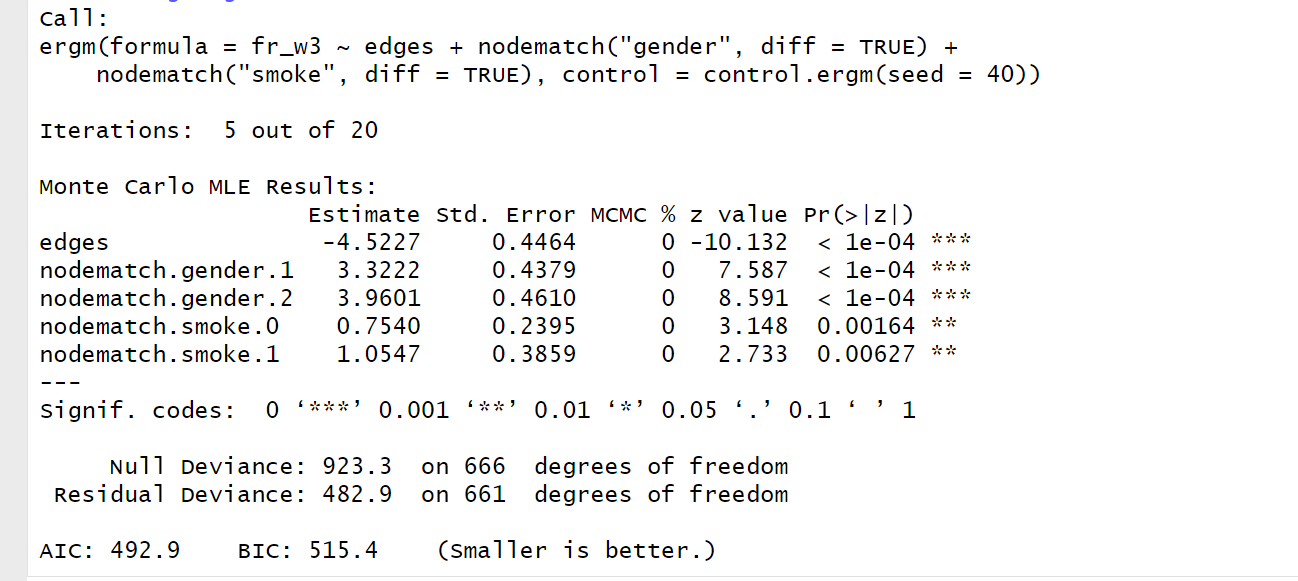
****

Here is the null model. It is not hard to find that all null models are similar according to AIC/BIC.

**Model 2**

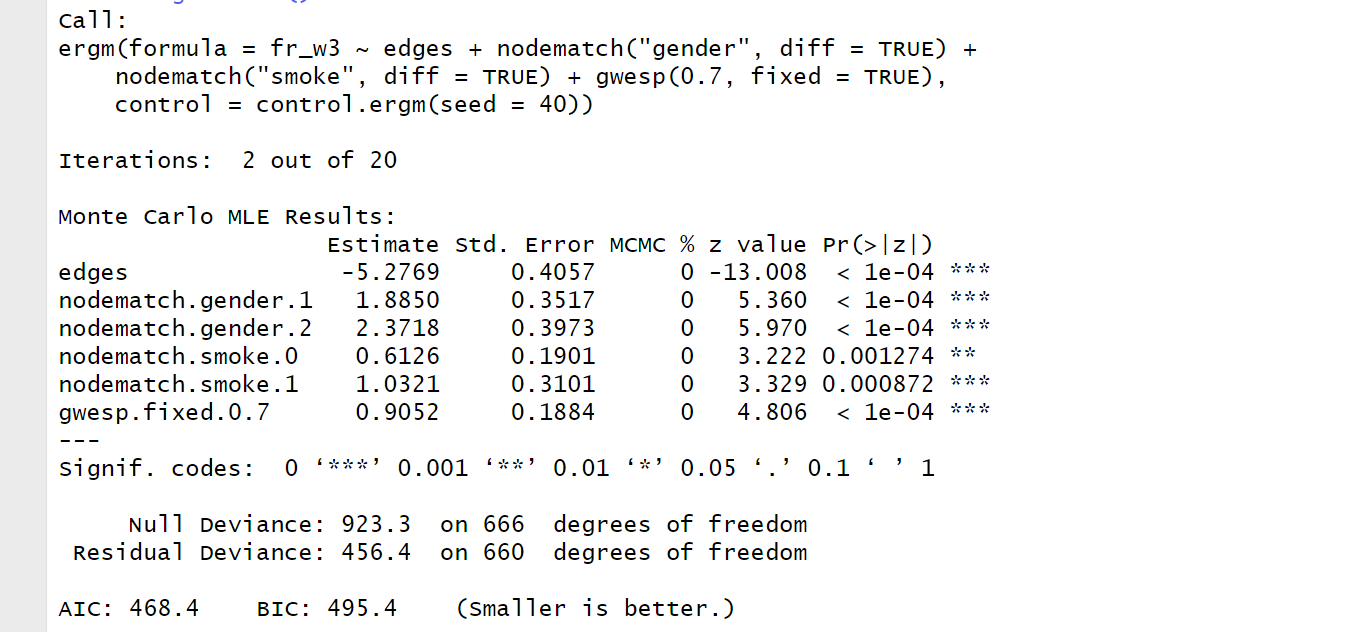


Still not significant in terms of two factors.



If using match function, the model can fit well as all variables are all significant. And there is a great reduce for AIC/BIC.

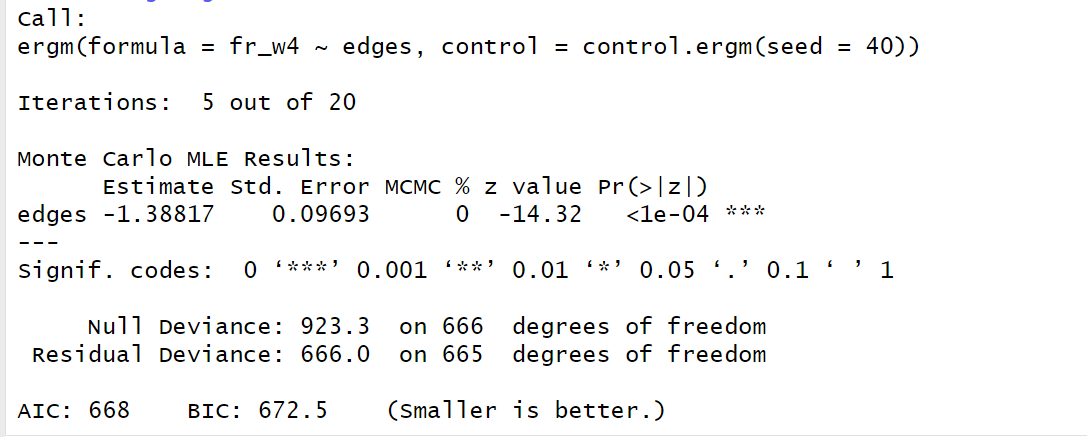
**Model 3**

****

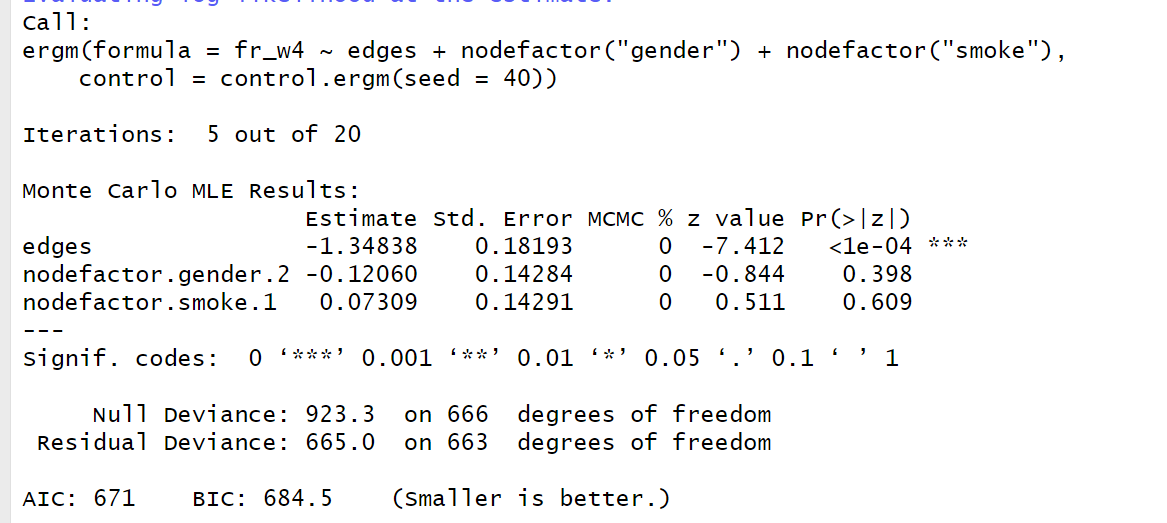
By adding gwesp, AIC/BIC still going down. So the model 3 is always the best model.

**V Data IV**

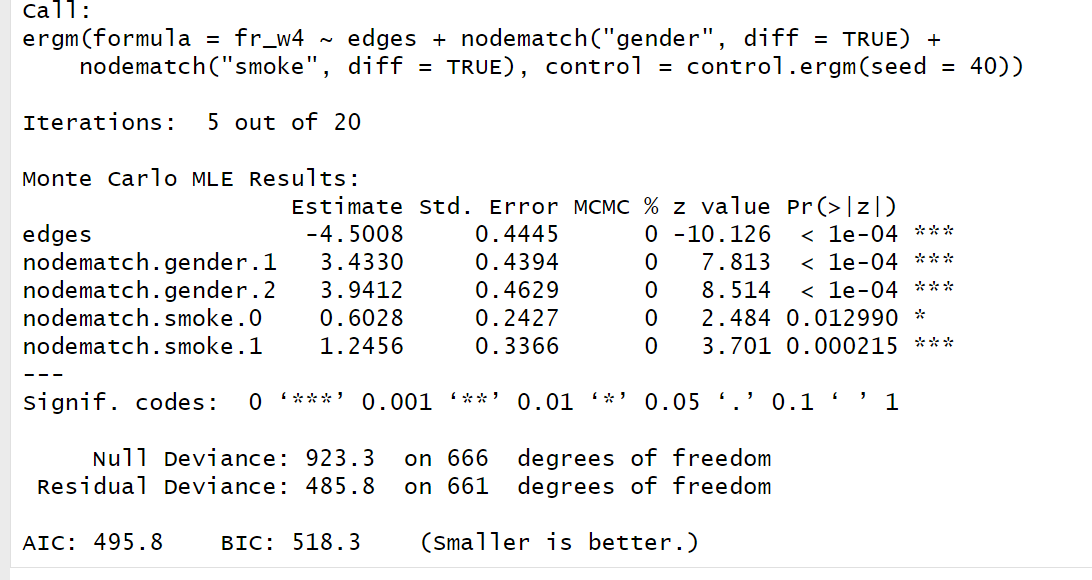
**Model 1**

****

**Model 2**

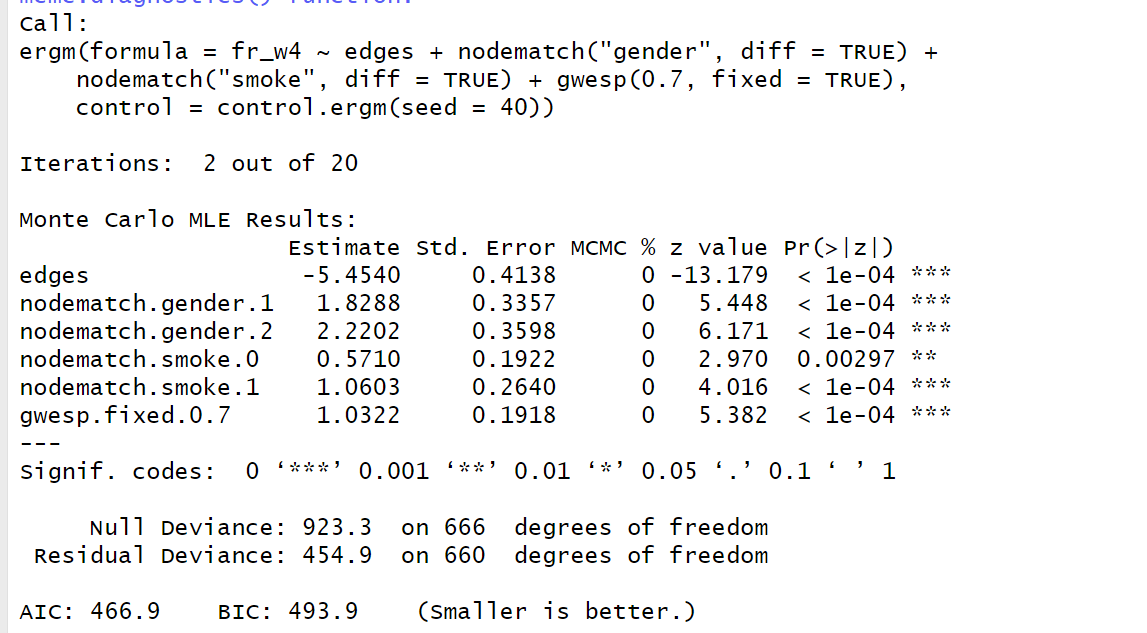
****

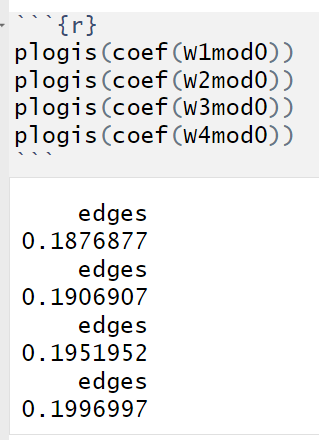
Not possible for factors to be significant in this data…..lol



All good, without any hesitation.

**Model 3**

****

****

**Conclusion**

For all four datasets, null model always the worst one while model with match and gwesp is always the best one. Factors in these models are always non-significant. Match function is a very interesting method to reduce AIC/BIC. Gwesp can reduce the AIC/BIC based on the match function.

**VI Appendix**

**---**

**title: "Network\_HW5"**

**author: "Chenrui Xu"**

**date: "2021/3/21"**

**output: html\_document**

**---**

**```{r}**

**library(igraph)**

**library(UserNetR)**

**library(ergm)**

**library(intergraph)**

**```**

**```{r}**

**data("Coevolve")**

**Coevolve**

**```**

**```{r}**

**fr\_w1 <- Coevolve$fr\_w1**

**class(fr\_w1)**

**fr\_w1 <- as.undirected(fr\_w1)**

**fr\_w1 <- asNetwork(fr\_w1)**

**fr\_w2 <- Coevolve$fr\_w2**

**fr\_w2 <- as.undirected(fr\_w2)**

**fr\_w2 <- asNetwork(fr\_w2)**

**fr\_w3 <- Coevolve$fr\_w3**

**fr\_w3 <- as.undirected(fr\_w3)**

**fr\_w3 <- asNetwork(fr\_w3)**

**fr\_w4 <- Coevolve$fr\_w4**

**fr\_w4 <- as.undirected(fr\_w4)**

**fr\_w4 <- asNetwork(fr\_w4)**

**```**

**```{r}**

**fr\_w1**

**```**

**```{r}**

**W1mod0 <- ergm(fr\_w1 ~ edges, control=control.ergm(seed = 40))**

**summary(W1mod0)**

**```**

**```{r}**

**W1mod1 <- ergm(fr\_w1 ~ edges + nodefactor('gender') + nodefactor('smoke'), control=control.ergm(seed = 40))**

**summary(W1mod1)**

**```**

**```{r}**

**W1mod2 <- ergm(fr\_w1 ~ edges**

**+nodefactor('gender')+nodefactor('smoke')**

**+ nodematch('gender') +**

**nodematch('smoke'), control=control.ergm(seed = 40))**

**summary(W1mod2)**

**```**

**```{r}**

**W1mod3 <- ergm(fr\_w1 ~ edges**

**#+ nodefactor('gender') + nodefactor('smoke')**

**+ nodematch('gender', diff=TRUE)+ nodematch('smoke', diff=TRUE),**

**control=control.ergm(seed = 40))**

**summary(W1mod3)**

**```**

**```{r}**

**W1mod4 <- ergm(fr\_w1 ~ edges**

**+ nodematch('gender', diff=TRUE)+ gwesp(0.7, fixed=TRUE), control=control.ergm(seed = 40))**

**summary(W1mod4)**

**```**

**```{r}**

**W2mod0 <- ergm(fr\_w2 ~ edges, control=control.ergm(seed = 40))**

**summary(W2mod0)**

**```**

**```{r}**

**W2mod1 <- ergm(fr\_w2 ~ edges + nodefactor('gender') + nodefactor('smoke'), control=control.ergm(seed = 40))**

**summary(W2mod1)**

**```**

**```{r}**

**W2mod2 <- ergm(fr\_w2 ~ edges**

**+ nodematch('gender', diff=TRUE) +**

**nodematch('smoke', diff=TRUE), control=control.ergm(seed = 40))**

**summary(W2mod2)**

**```**

**```{r}**

**W2mod3 <- ergm(fr\_w2 ~ edges + nodematch('gender', diff=TRUE)**

**+ nodematch('smoke', diff=TRUE)**

**+ gwesp(0.7, fixed=TRUE), control=control.ergm(seed = 40))**

**summary(W2mod3)**

**```**

**```{r}**

**W3mod0 <- ergm(fr\_w3 ~ edges, control=control.ergm(seed = 40))**

**summary(W3mod0)**

**```**

**```{r}**

**W3mod1 <- ergm(fr\_w3 ~ edges + nodefactor('gender') + nodefactor('smoke'), control=control.ergm(seed = 40))**

**summary(W3mod1)**

**```**

**```{r}**

**W3mod2 <- ergm(fr\_w3 ~ edges + nodematch('gender', diff=TRUE) +**

**nodematch('smoke', diff=TRUE), control=control.ergm(seed = 40))**

**summary(W3mod2)**

**```**

**```{r}**

**W3mod3 <- ergm(fr\_w3 ~ edges + nodematch('gender', diff=TRUE) +**

**nodematch('smoke', diff=TRUE) + gwesp(0.7, fixed=TRUE), control=control.ergm(seed = 40))**

**summary(W3mod3)**

**```**

**```{r}**

**W4mod0 <- ergm(fr\_w4 ~ edges, control=control.ergm(seed = 40))**

**summary(W4mod0)**

**```**

**```{r}**

**W4mod1 <- ergm(fr\_w4 ~ edges + nodefactor('gender') + nodefactor('smoke'), control=control.ergm(seed = 40))**

**summary(W4mod1)**

**```**

**```{r}**

**W4mod2 <- ergm(fr\_w4 ~ edges + nodematch('gender', diff=TRUE) +**

**nodematch('smoke', diff=TRUE), control=control.ergm(seed = 40))**

**summary(W4mod2)**

**```**

**```{r}**

**W4mod3 <- ergm(fr\_w4 ~ edges + nodematch('gender', diff=TRUE) +**

**nodematch('smoke', diff=TRUE) + gwesp(0.7, fixed=TRUE), control=control.ergm(seed = 40))**

**summary(W4mod3)**

**```**

**```{r}**

**plogis(coef(W1mod0))**

**plogis(coef(W2mod0))**

**plogis(coef(W3mod0))**

**plogis(coef(W4mod0))**

**```**