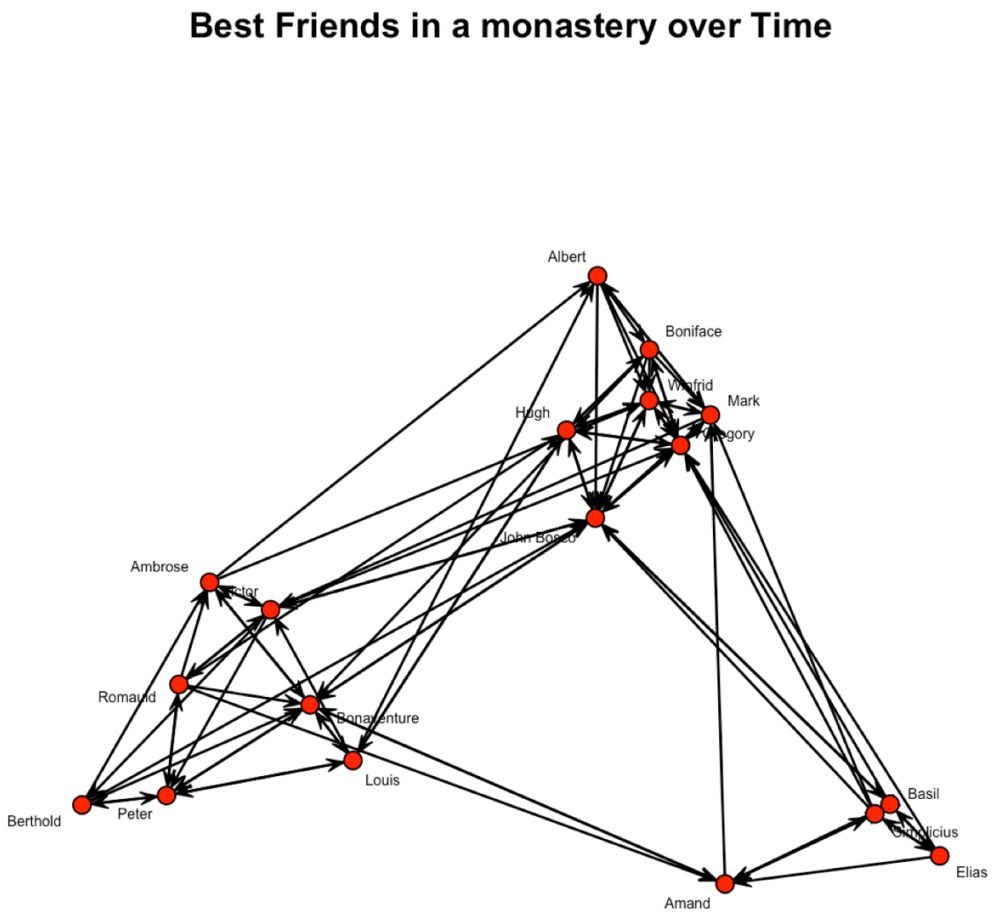
Stats218 Homework 5
Name: Ruchen Zhen
UID: 205036408

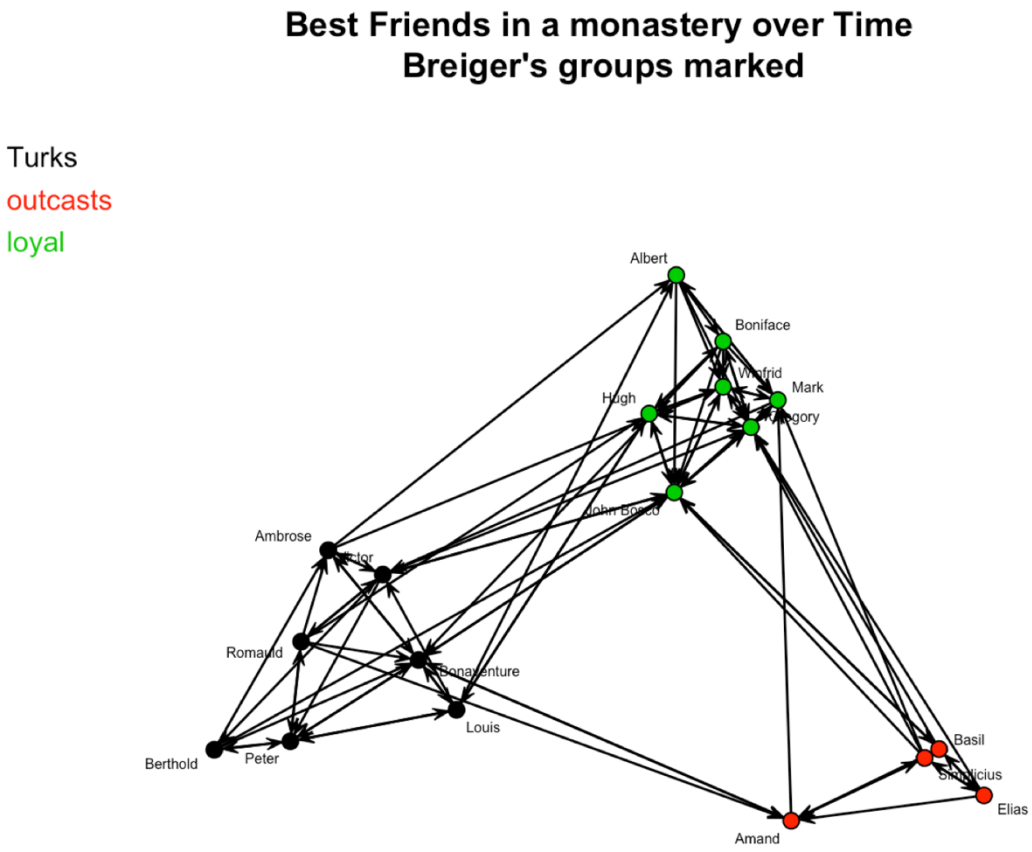
Problem 1: Latent Space Model for the Social Space of Monks

(a)



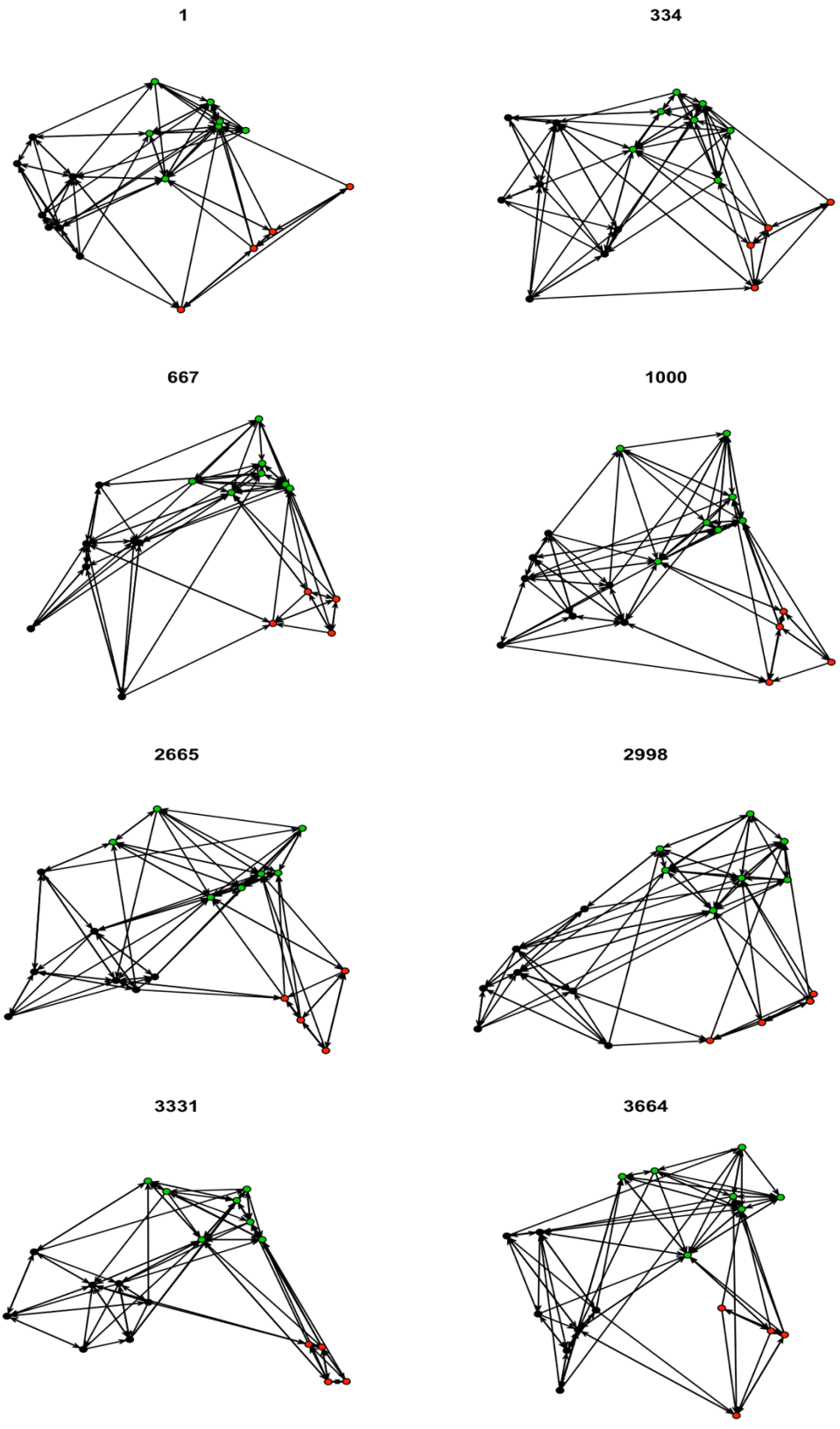
Above is the plot for the resulting positions. From the plot, we can see that there are three groups in the plot according to their positions, and no monks are isolated.

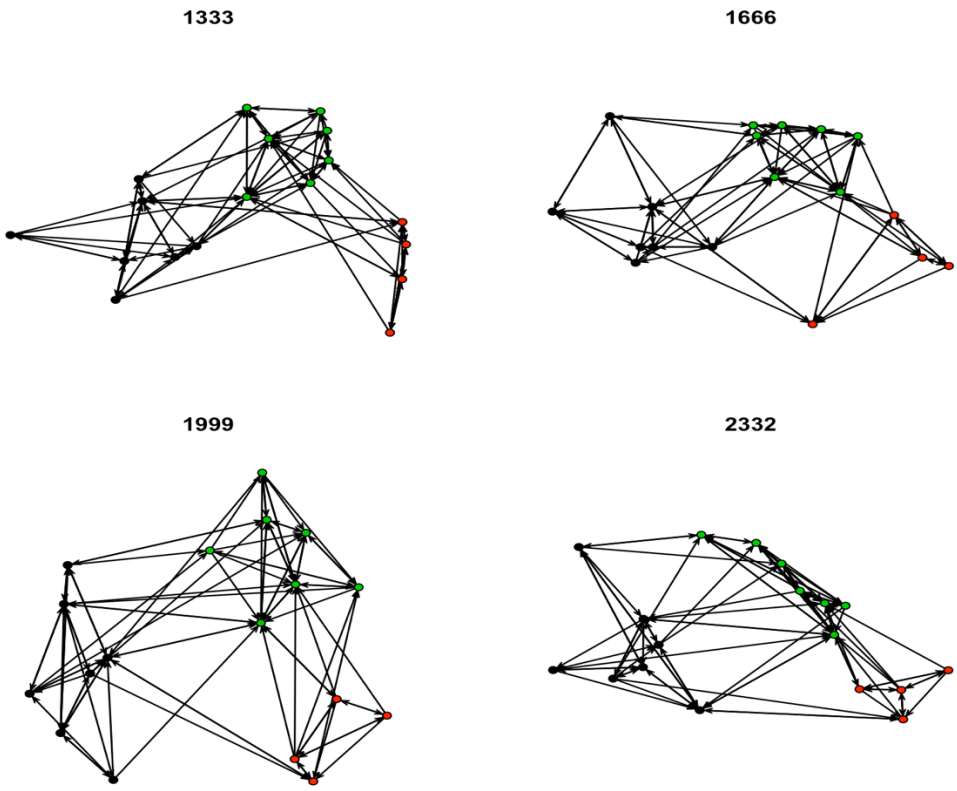
(b)



We then plot the network again, with different colors representing the groups they belong to. As can be seen in the plot, the three groups are perfected discovered by the latent space model.

(c)





Above we have generated 12 samples from the model we fit. From these plots, we can see that we are quite certain about the positions of the monks. The 4 red nodes are at the right down side, the 7 black nodes are at the left down side, while the 7 green nodes always locate at the middle up side. They are well separated. Nodes with same color gathers together, while they are much far away from the nodes with different color.

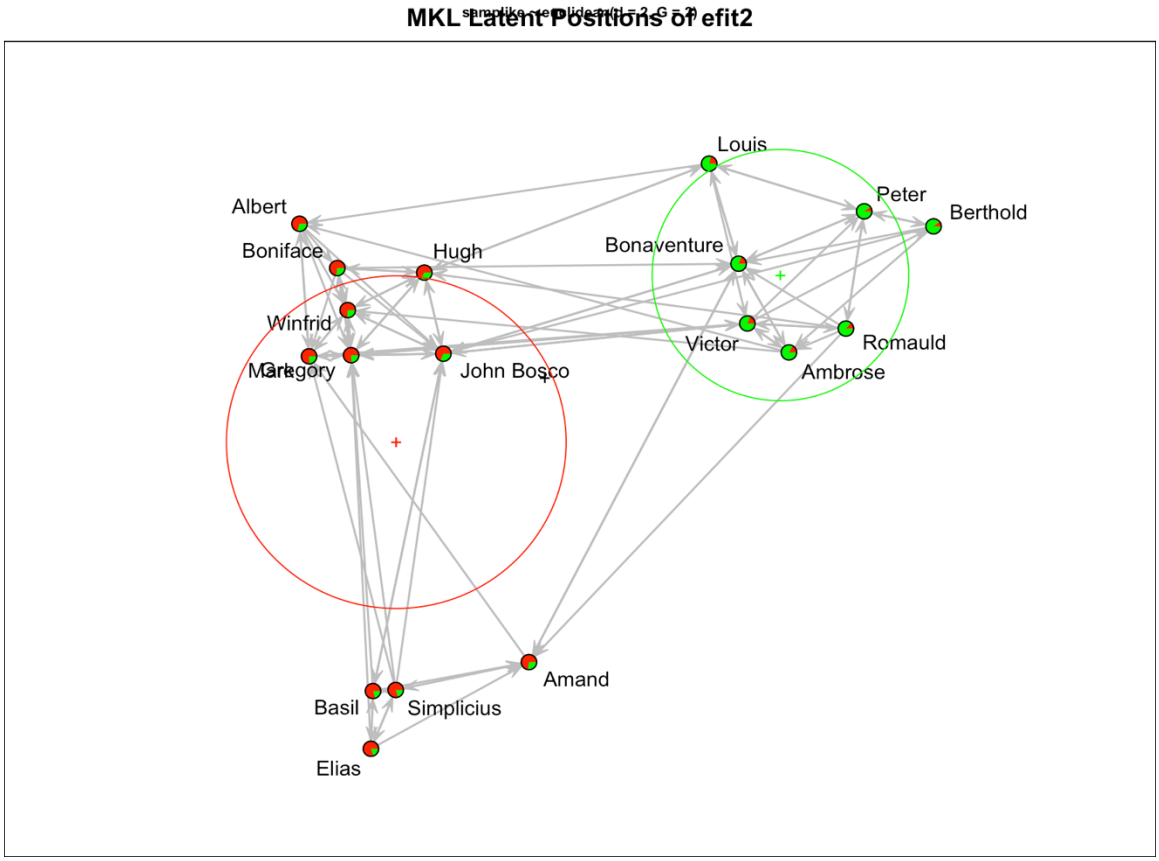
Problem 2: Joint Modeling of the Sub-groups and Social Space of Monks

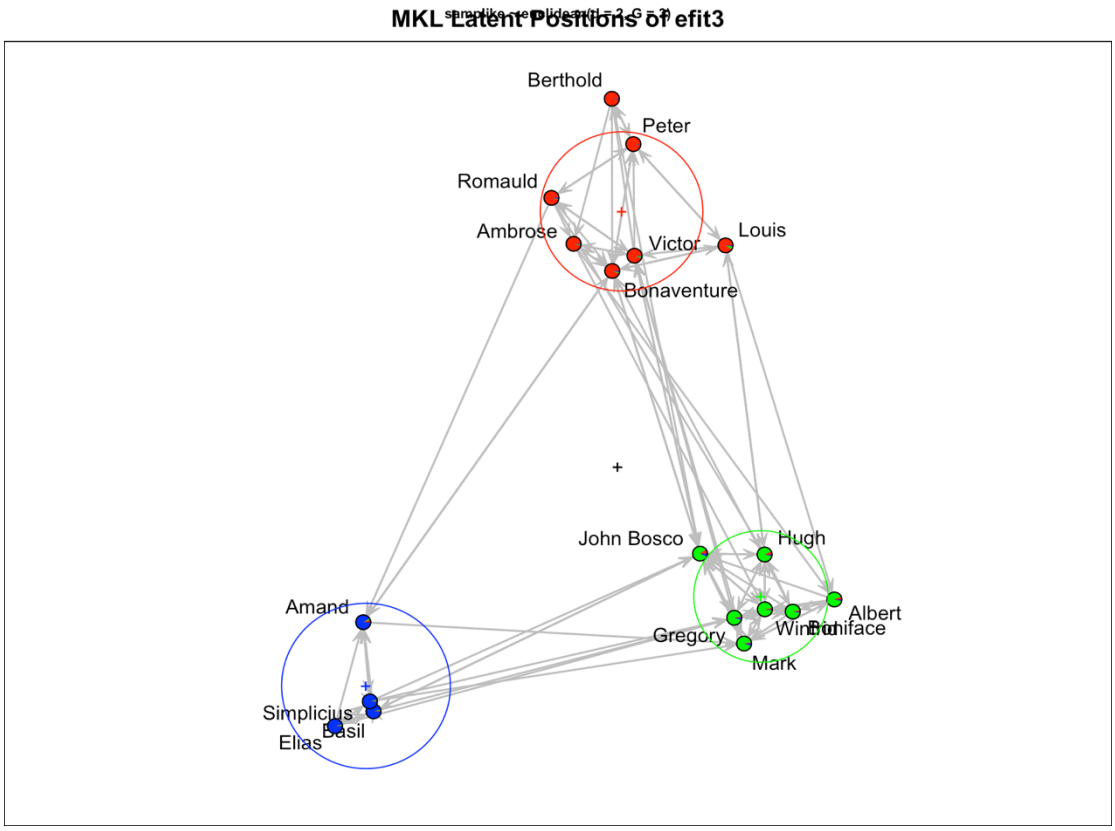
(a)
We know that there are actually 3 groups, thus, we will fit three models with group numbers of 2, 3, 4 and compare them. We expect that the model with 3 groups will perform the best. First is their BICs:

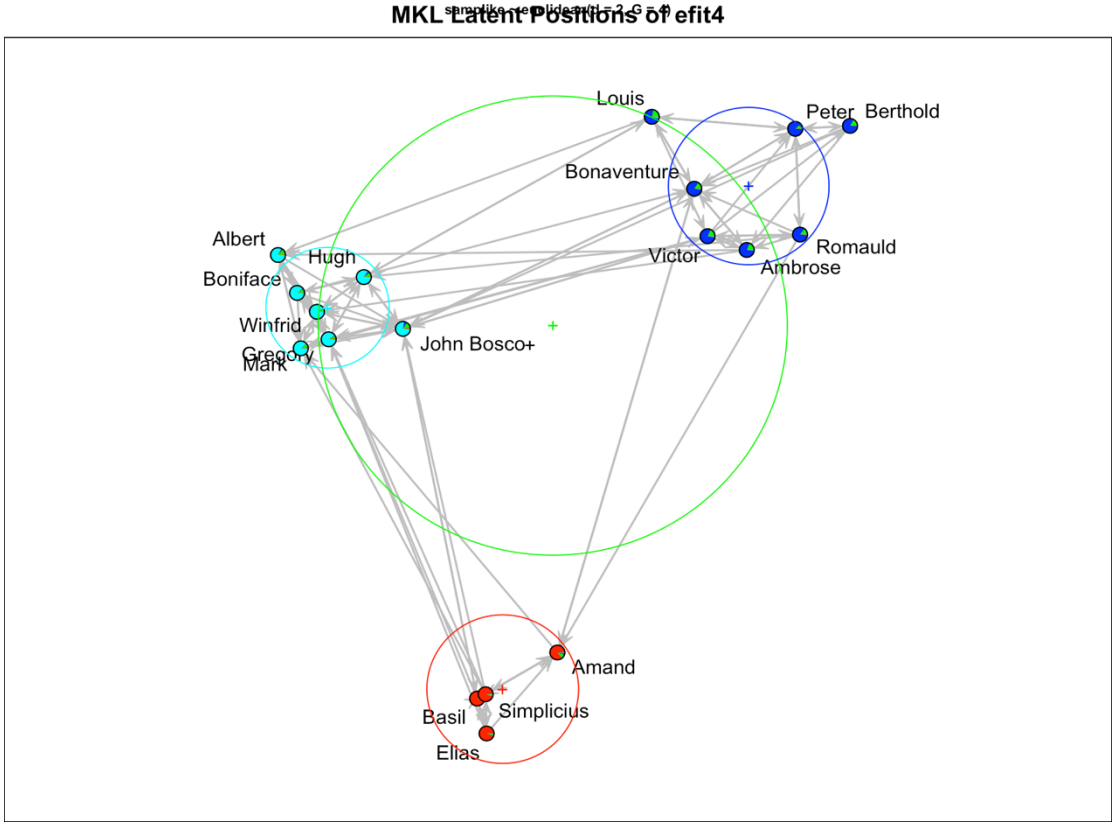
Number of Groups	Overall BIC	Latent space/clustering BIC
2	373.201	119.5765
3	337.1682	82.7319
4	342.7474	88.77115

As can be seen in the table, the BIC decreases when goes from 2-groups model to group-3 model, then increases when goes to group-4 model. We can see that group of 3 is strongly supported.

We then plot the resulting groups:

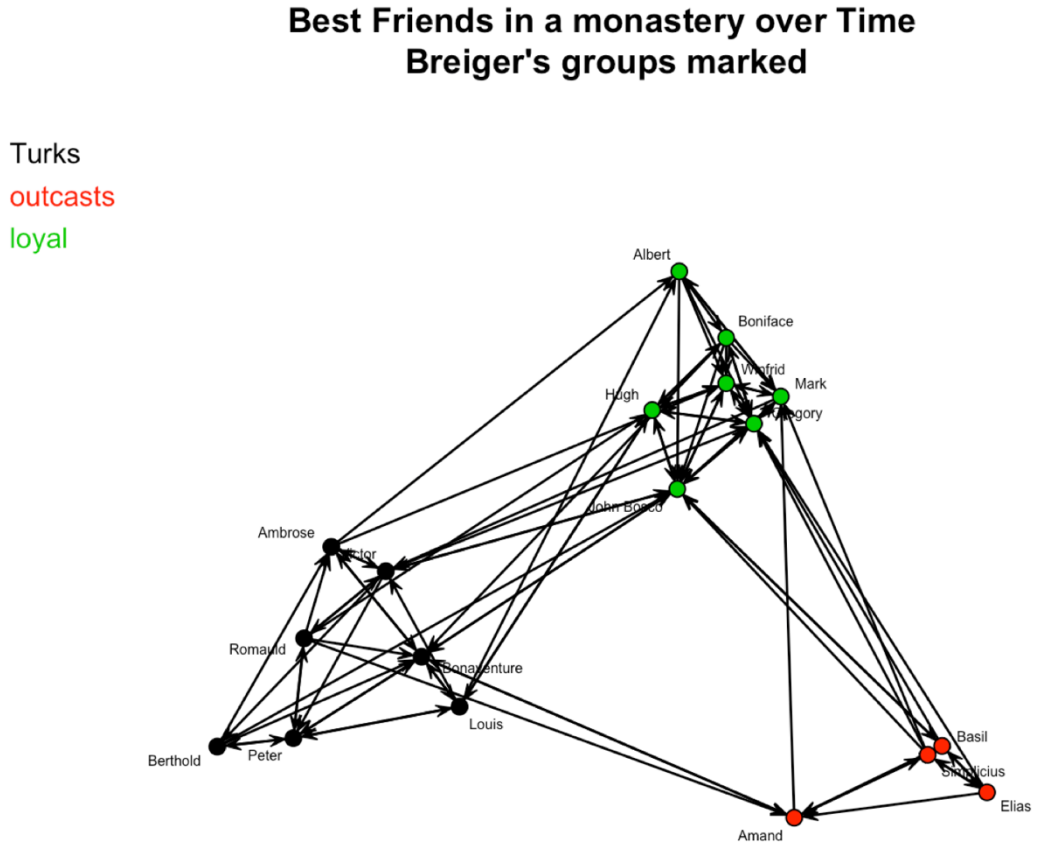



MKL Latent Positions of efit3
Berthold
Peter
Romauld
Ambrose
Victor
Louis
Bonaventure
John Bosco
Hugh
Amand
Albert
Gregory
Mark
Simplicius
Basil
Elias


MKL Latent Positions of efit4
Louis
Peter
Berthold
Bonaventure
Albert
Hugh
Boniface
Winfrid
Gregory
Mark
John Bosco
Victor
Romauld
Ambrose
Amand
Basil
Simplicius
Elias

Comparing the three plots, we can see that: 1. In the 2-groups model, there is a gap in the red group: with large distance between the upper nodes and lower nodes. 2. The 3-groups model performs well. 3. In the 4-groups model, the green group is meaningless. It only covers some boundary nodes of the other 3 groups. We may conclude that it is pretty clear that there should be three groups.

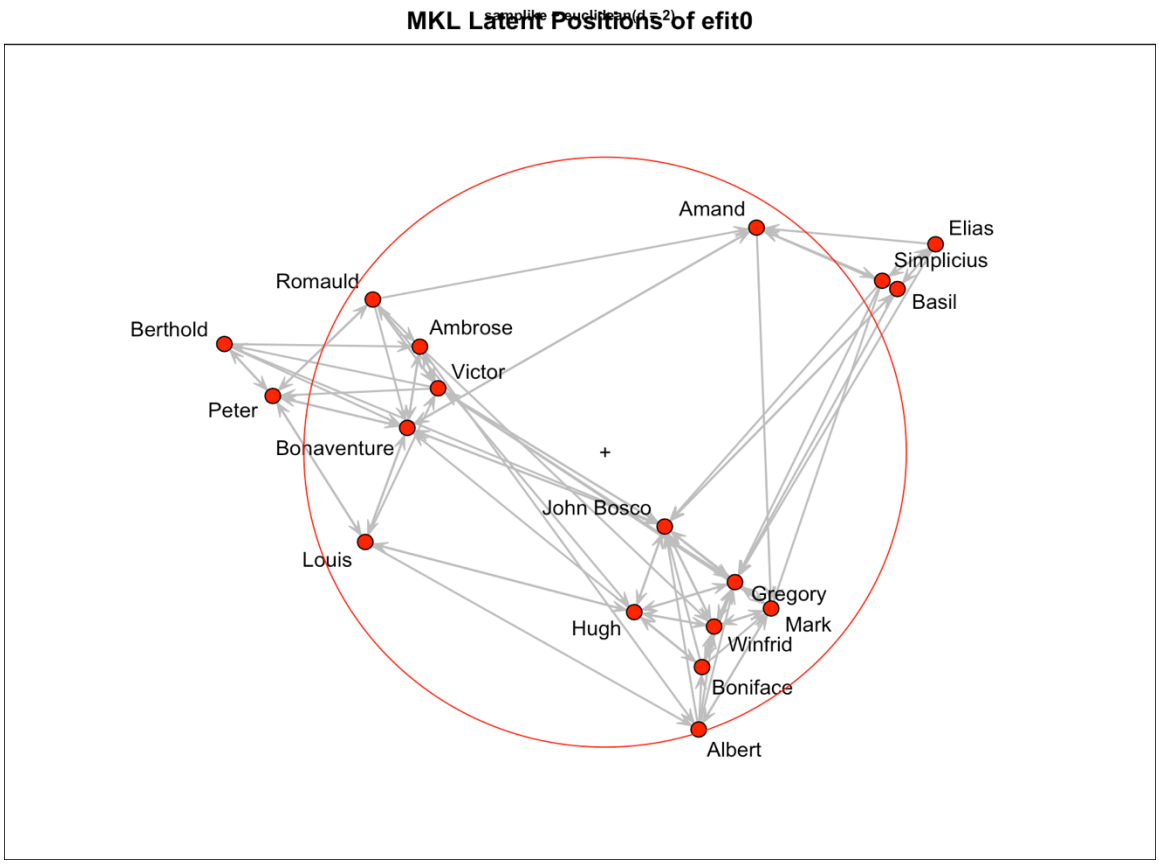
(b)
Consider again the original groups plot. It matches the 3-groups model well. This also supports that there should be 3 groups.



(c)
Consider again the three plots in (a). The probabilities of a node, belonging into the three groups, are showing as pie chart when plotting the node color. For the 3-groups plot, the colors of the group the nodes belong to almost dominate. However, for the 2-groups model and 4-groups model, ambiguity occurs. But the main color still occupies around 70% of the circle's area.

(d)

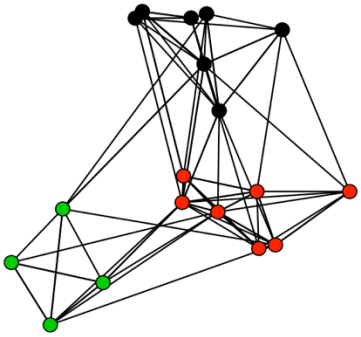
The plots are still in (a). For the 4-groups model, we can hardly see the color of the forth group, green, in the plot. There is only a little bit green in the blue group, and quite a little in the light blue group. We plot the latent space model fit when setting the group number to be 1, as follows. We can see that this group is very similar to the green group in the 4-groups plot. This is also a strong support that there should actually be 3 groups, given that the forth group seems to be the "combining group" of the three existing groups.



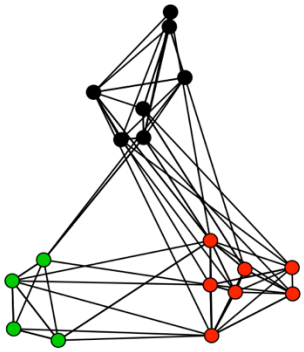
(e)

We plot 24 samples, simulated with our best model, the 3-groups one, as shown in the following plots. In the 24 plots, we have actually 5 plots with uncertainty, as pointed by the arrow. In these plots, there are nodes that locate at the position of other groups.

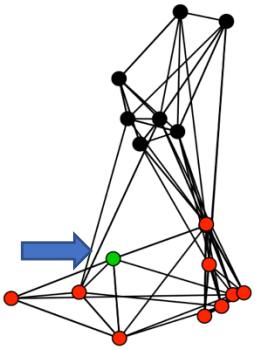
Sample number 281



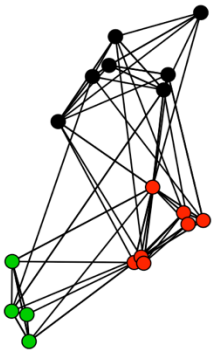
Sample number 3664



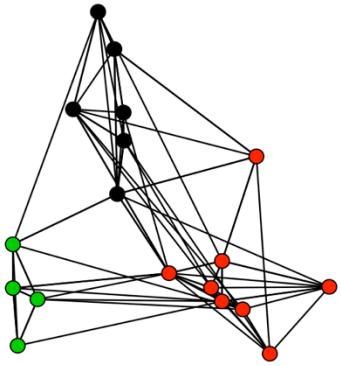
Sample number 2920



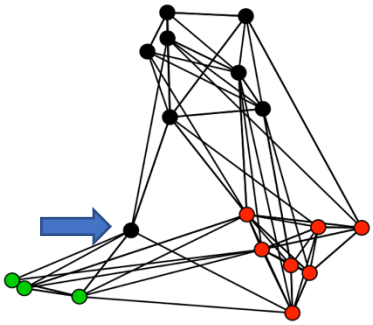
Sample number 3323



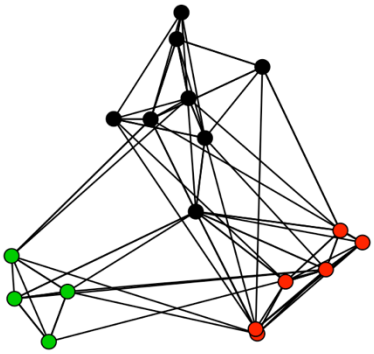
Sample number 1309



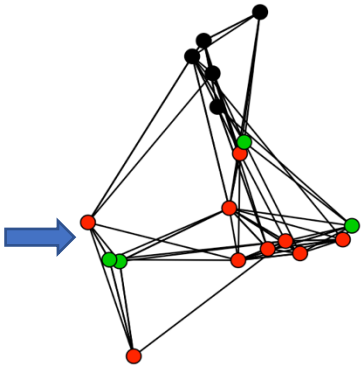
Sample number 1121



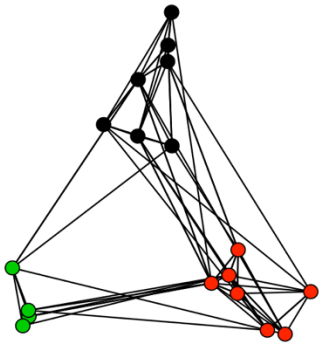
Sample number 3787



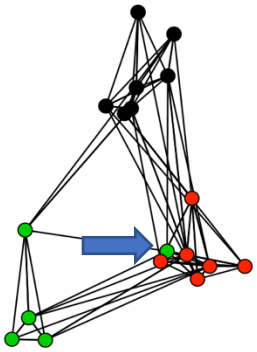
Sample number 1902



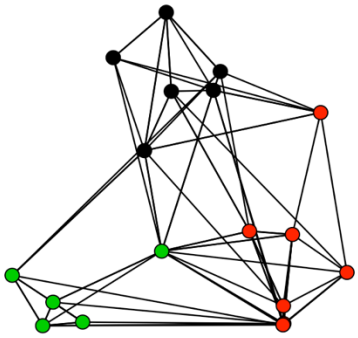
Sample number 3627



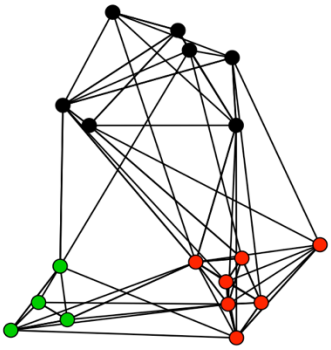
Sample number 601



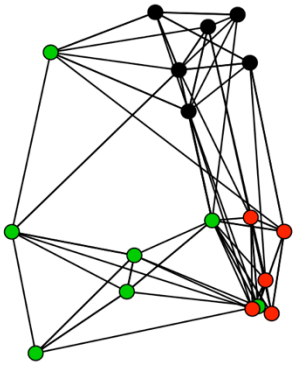
Sample number 192



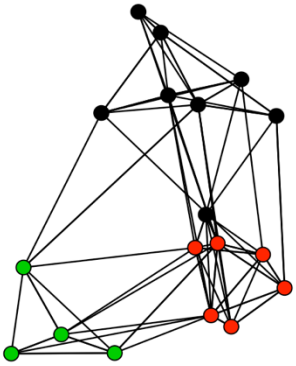
Sample number 1854



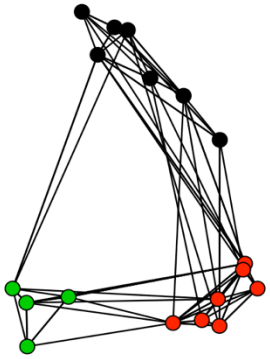
Sample number 1222



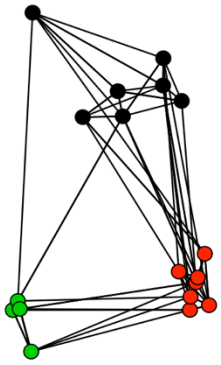
Sample number 677



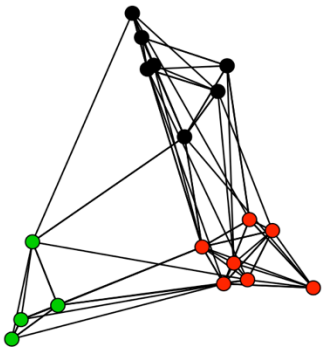
Sample number 3291



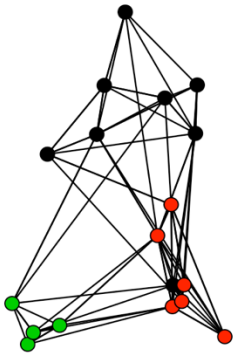
Sample number 3085



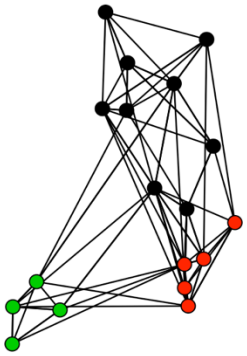
Sample number 2574



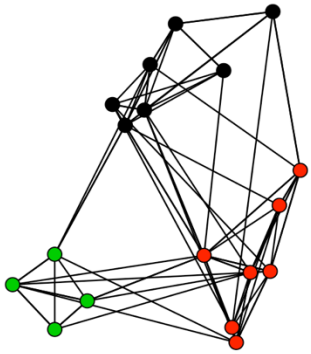
Sample number 1495



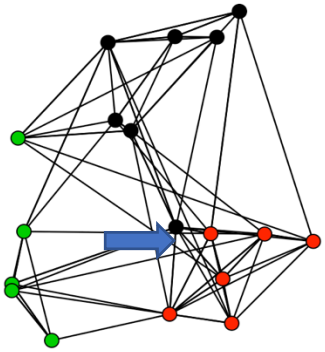
Sample number 2183



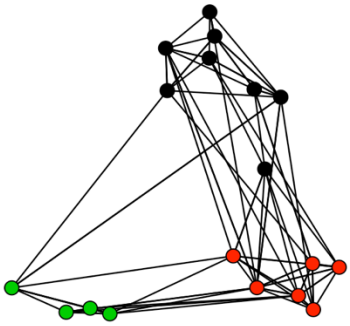
Sample number 3078



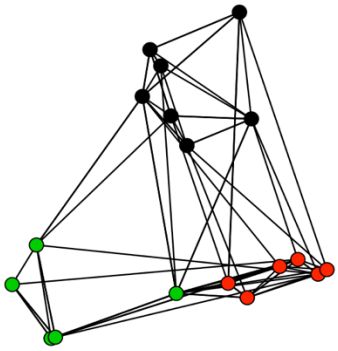
Sample number 1972



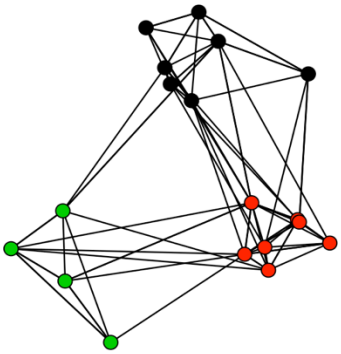
Sample number 2611



Sample number 1265



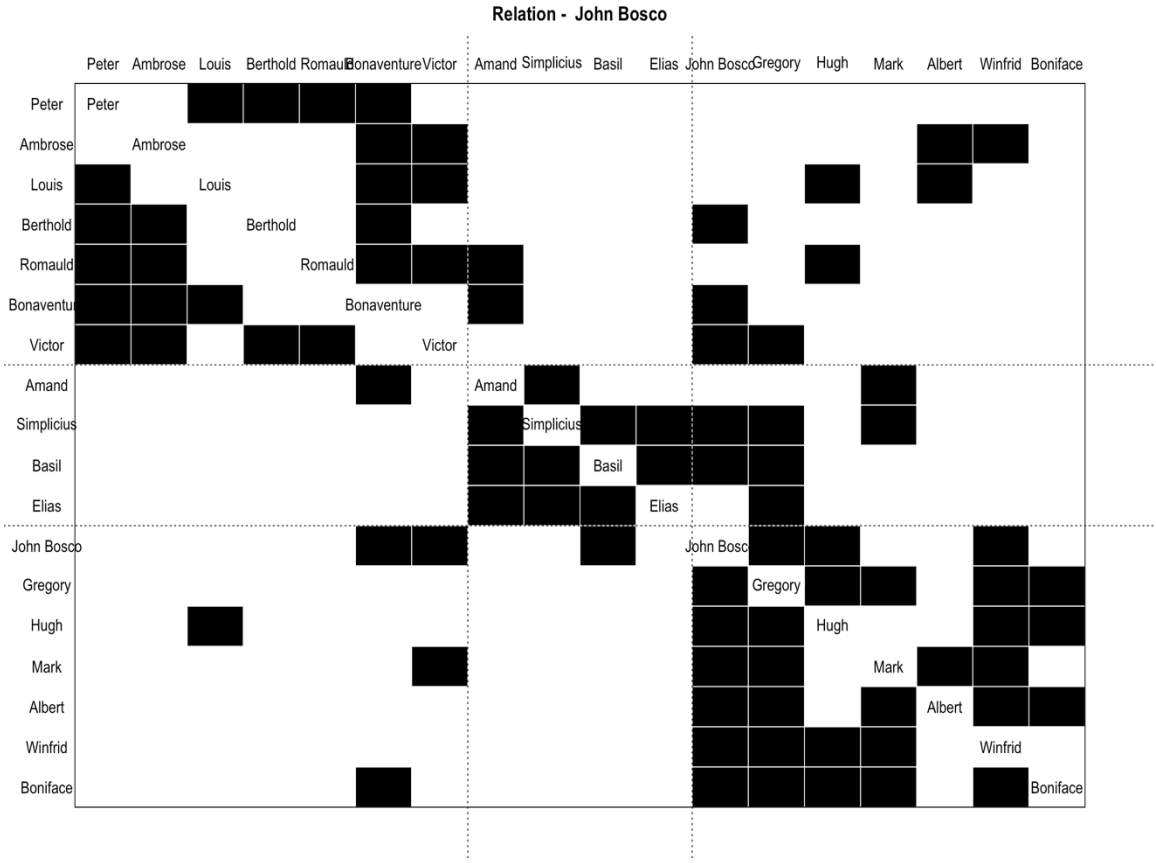
Sample number 1530



Problem 3: "Block" models for the Social Relations of Monks

(a)

In this question, we fit a hierarchical clustering model using sna. Below is the result and the actual members for the three groups:




```
> names[turks]
[1] "John Bosco" "Gregory"    "Mark"       "Winfrid"    "Hugh"       "Boniface"
[7] "Albert"
> turks <- samplike %v% "group" =="Turks"
> names[turks]
[1] "John Bosco" "Gregory"    "Mark"       "Winfrid"    "Hugh"       "Boniface"
[7] "Albert"
> outcasts <- samplike %v% "group" =="Outcasts"
> names[outcasts]
[1] "Basil"      "Amand"      "Elias"      "Simplicius"
> loyal <- samplike %v% "group" =="Loyal"
> names[loyal]
[1] "Peter"       "Bonaventure" "Berthold"    "Victor"      "Ambrose"
[6] "Romauld"     "Louis"
```

We can see that the block model successfully figures out all the three groups. Seems that it is a good fit.

(b)

```
b1 <- blockmodels::BM_gaussian("SBM",g2)
b1$estimate()
which.max(b1$ICL)

b2 <- blockmodels::BM_bernoulli("SBM",g2)
b2$estimate()
which.max(b2$ICL)

b3 <- blockmodels::BM_poisson("SBM",g2)
b3$estimate()
which.max(b3$ICL)
```

```
> which.max(b1$ICL)
[1] 3
> which.max(b2$ICL)
[1] 2
> which.max(b3$ICL)
[1] 1
```

Using the blockmodels package, we fit three stochastic block models: Gaussian model, Bernoulli model and Poisson model. The b1$estimate shows the detailed ICL checking process, while here we only present the number of groups each model predicts. As can be seen in the picture, the Gaussian model obtained the correct result, while Bernoulli model and Poisson model do not. Poisson block model actually predict that there is only one group.

(c)

Compared all models, the latent space model obtains the correct result that there is 2 groups in total, while there does exist some uncertainty when simulating. The blockmodel in sna package does obtain the correction division. As for the blockmodels package, Gaussian model get the correct group number while Bernoulli model and Poisson model do not.