**STAT 6289 Network Final**

**Chenrui Xu**

**I Background**

This is the final project of the STAT 6289 Network. The data we will use is based on the flow of transportation between 17 cities. We can name after these cities from A to Q and there are somehow some roads to get them connected. As we should know that all roads here are one-way (directed edges). In the beginning, four cities A, D, F, H are isolated as there were no roads connecting them. And later, two-way roads were build to transport cargo between cities. The flow of cargoes was recorded when they went through one link and also the number of cargoes transported among A, D, F, H (that is what is lost and we are going to estimate).

Some knowledge we will use in the model are:

OD matrix: That is one kind of data frame we will use. It will record where the flow starts (Origin) and where the flow ends (Destination).

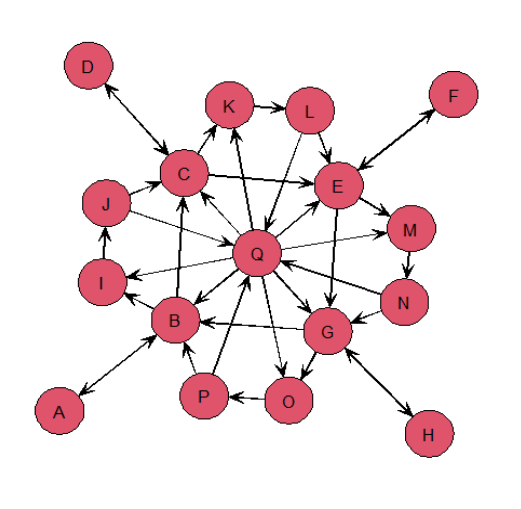
Gravity Model (tomo-gravity model): The regular gravity model will describe how the interaction between nodes. If we only know the amount of flow that goes through links and the timeline, we can use tomo-gravity model to predict the OD matrix. The model is based on the linear formation and the objective function is the loss function with the penalized item.



Here is the formula of the tomo-gravity model.

**II Task I**

In this part, we should build the tomo-gravity model and predict the amount of cargoes among points A, D, F, H.



Some analysis of the data processing:

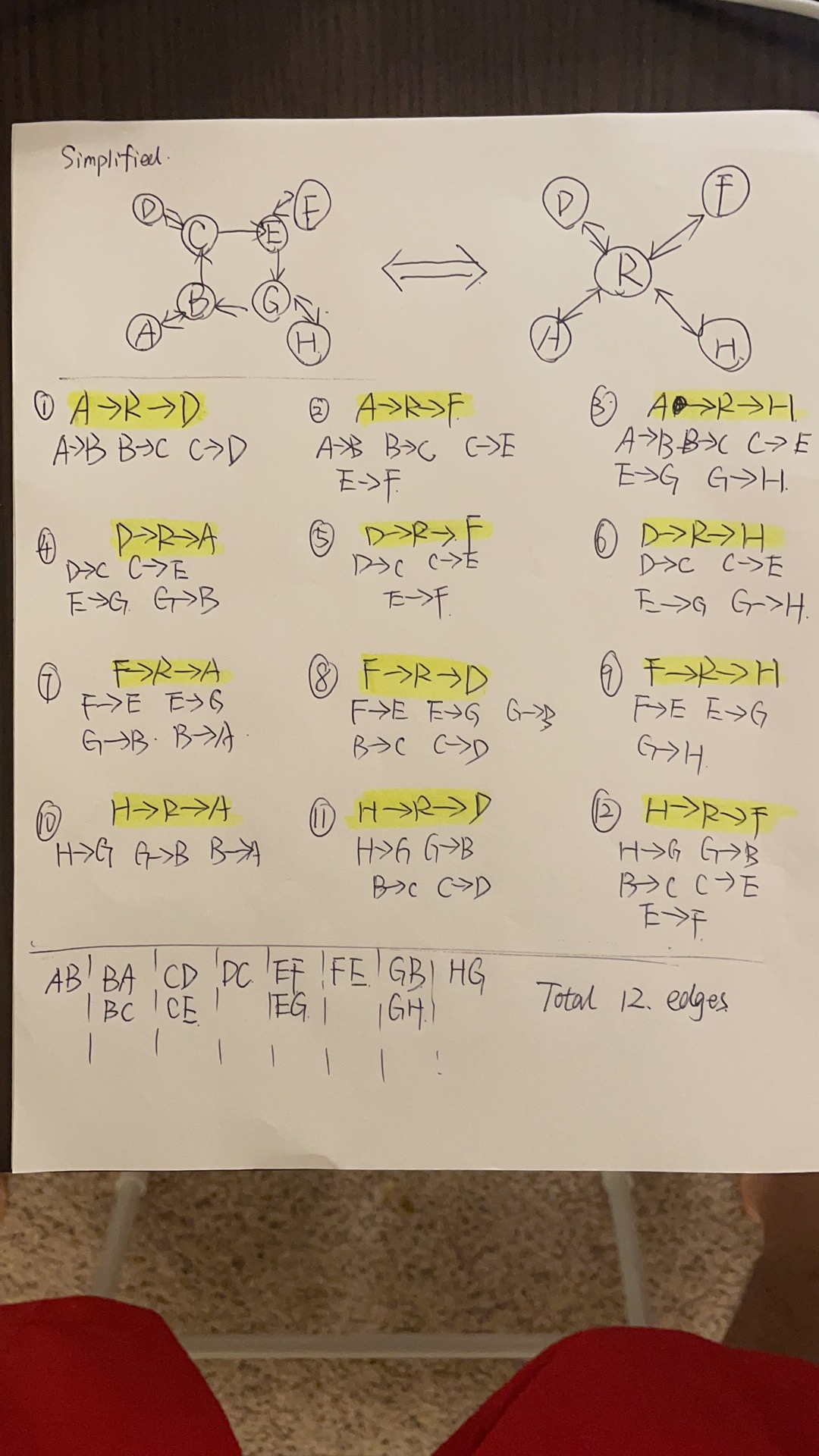
If we want to use the tomo-model, we need to know what are X, B, Z. Z is supplied but that is the OD matrix we need to estimate and compare. X is given as “Road” Matrix, which we will talk about later. B is the rounting matrix.

Next step: B matrix

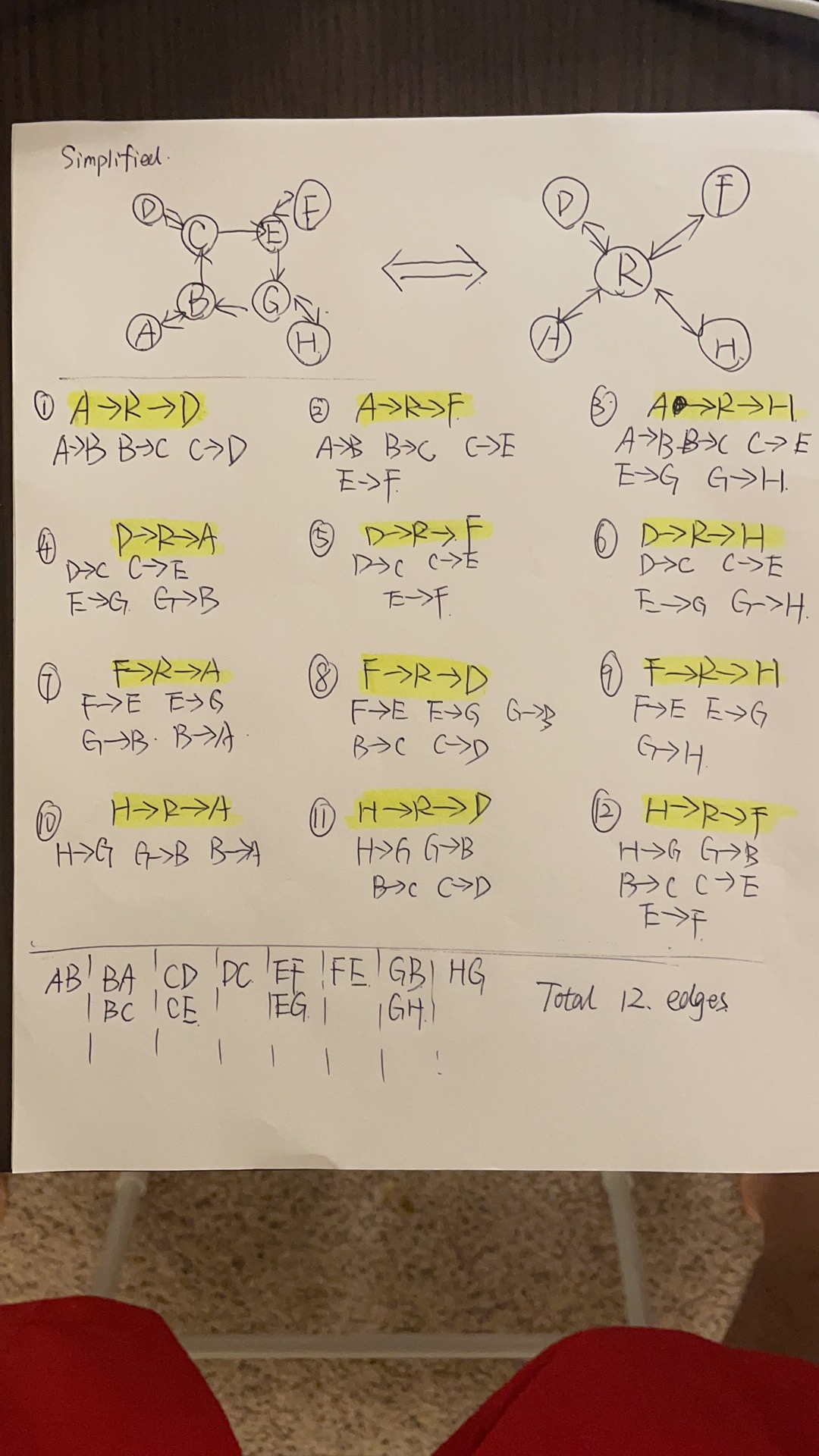
This is the core step during the whole analysis.

**If we want to analysis the way among A, D, F, H, there are four points we have to go through if those ways are the shortest paths. They are B, C, E, G.**

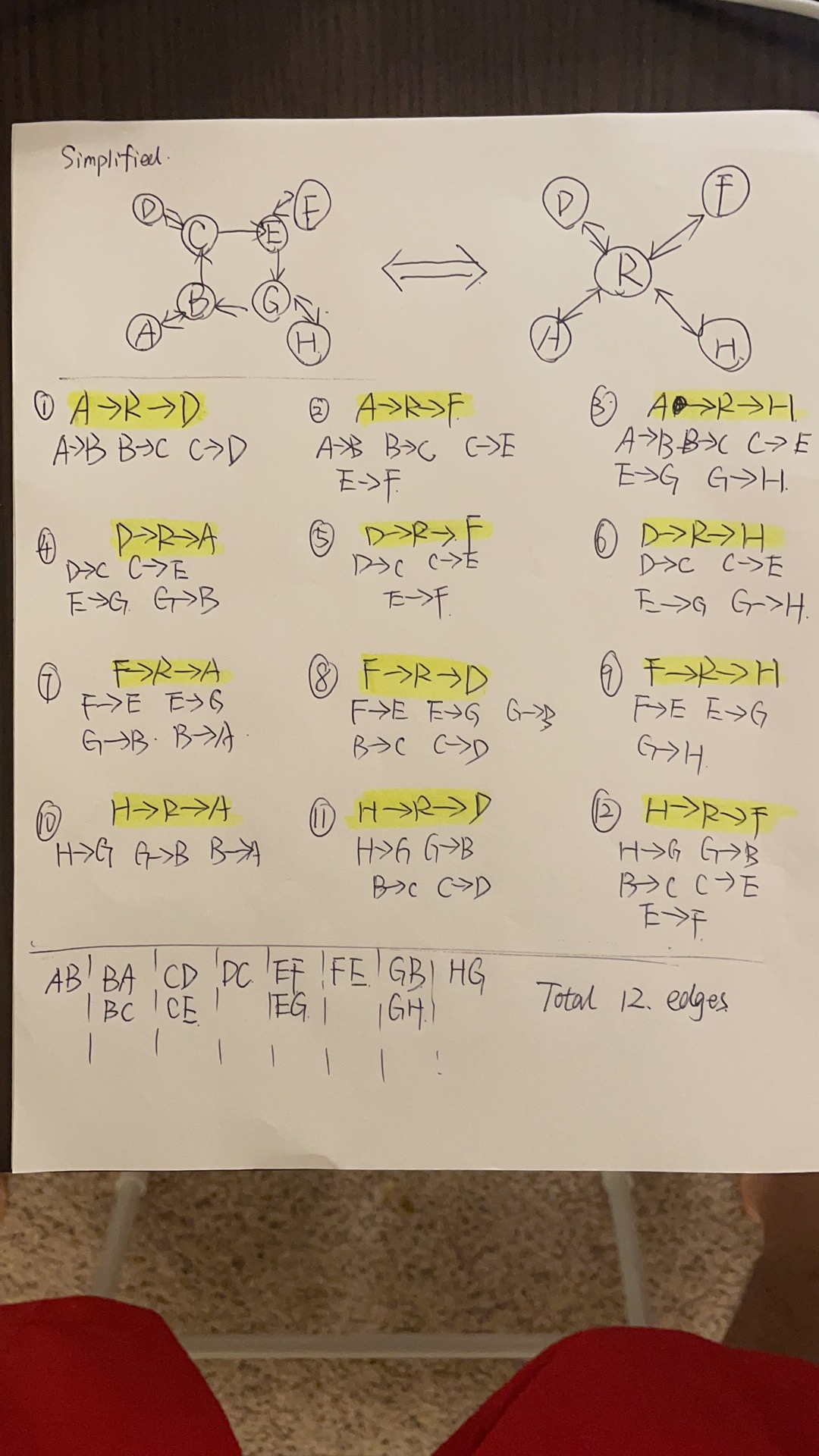
These four points can actually be seen as a system as a router same as the exapmle in the lecture notes. **So the whole 17 points network can be simplified into four points and a router system.**



We do not need to know the data not related to those links, so next step, I will use graphs to show how many links do we need to use here.

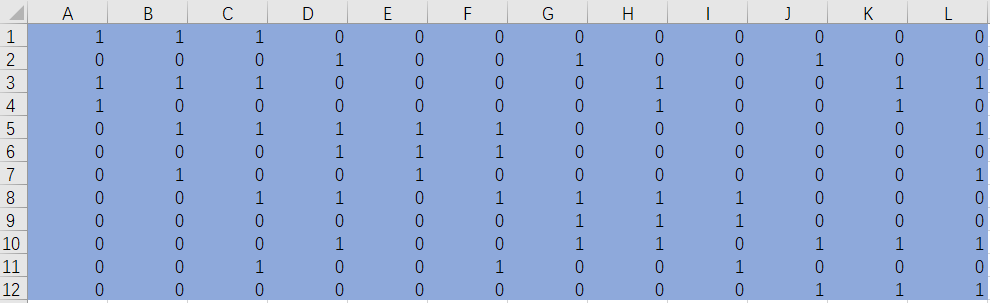


So, there are 12 OD links here (Origin to Destination). Their corresponding composed edges are showing below. So we can know which are useful.



After artificial counting, it is not hard for us to find that there are totally 12 edges will be used into the analysis.

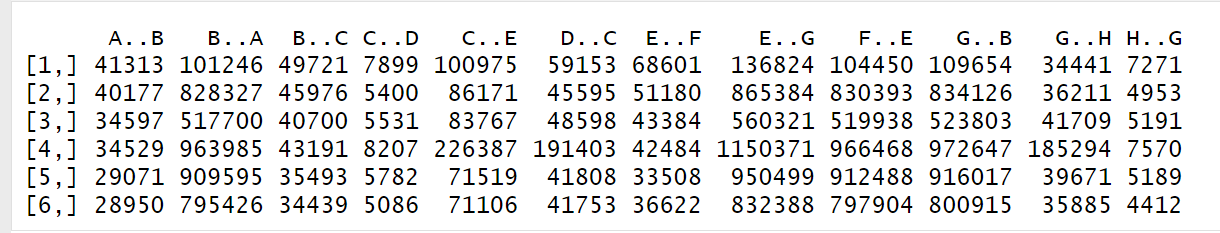
Based on that, we can easily get our B matrix.



I created an Excel file of B so it is convenient to read the file instead of creating it in the RStudio. The column names are OD and the rows are edges we need.

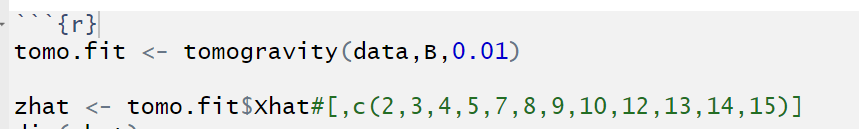
In terms of X:

Actually, it is not hard for us to find that in the file “Road”, if those links are not related, they are all zero. So we can pick those useful 12 columns easily.

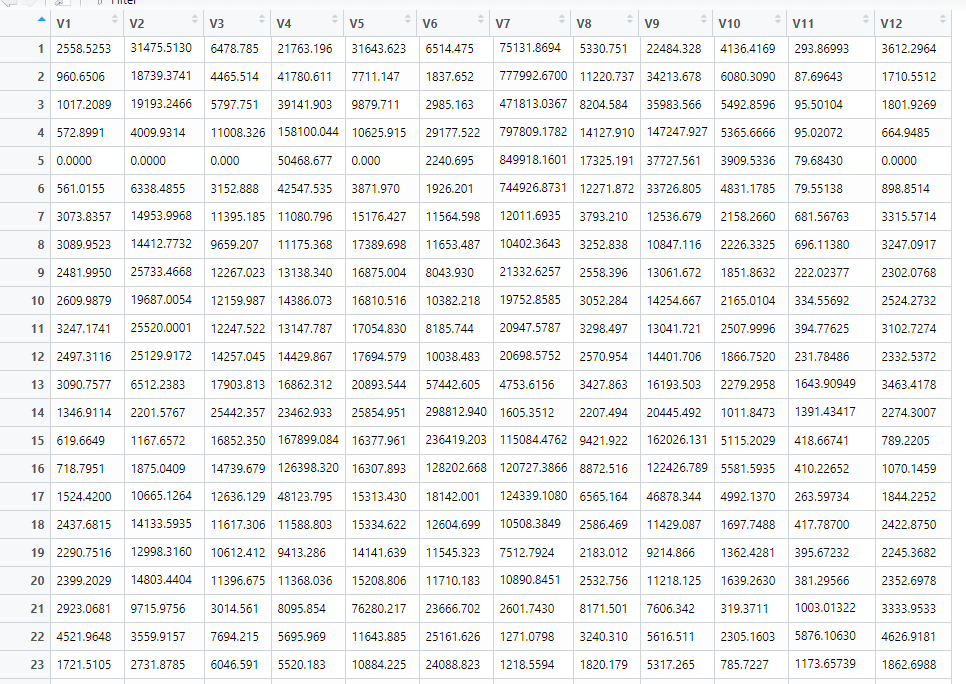


The dimension of the data is 96\*12.

**Fit the model**



Here we set the tuning parameter to be 0.01.

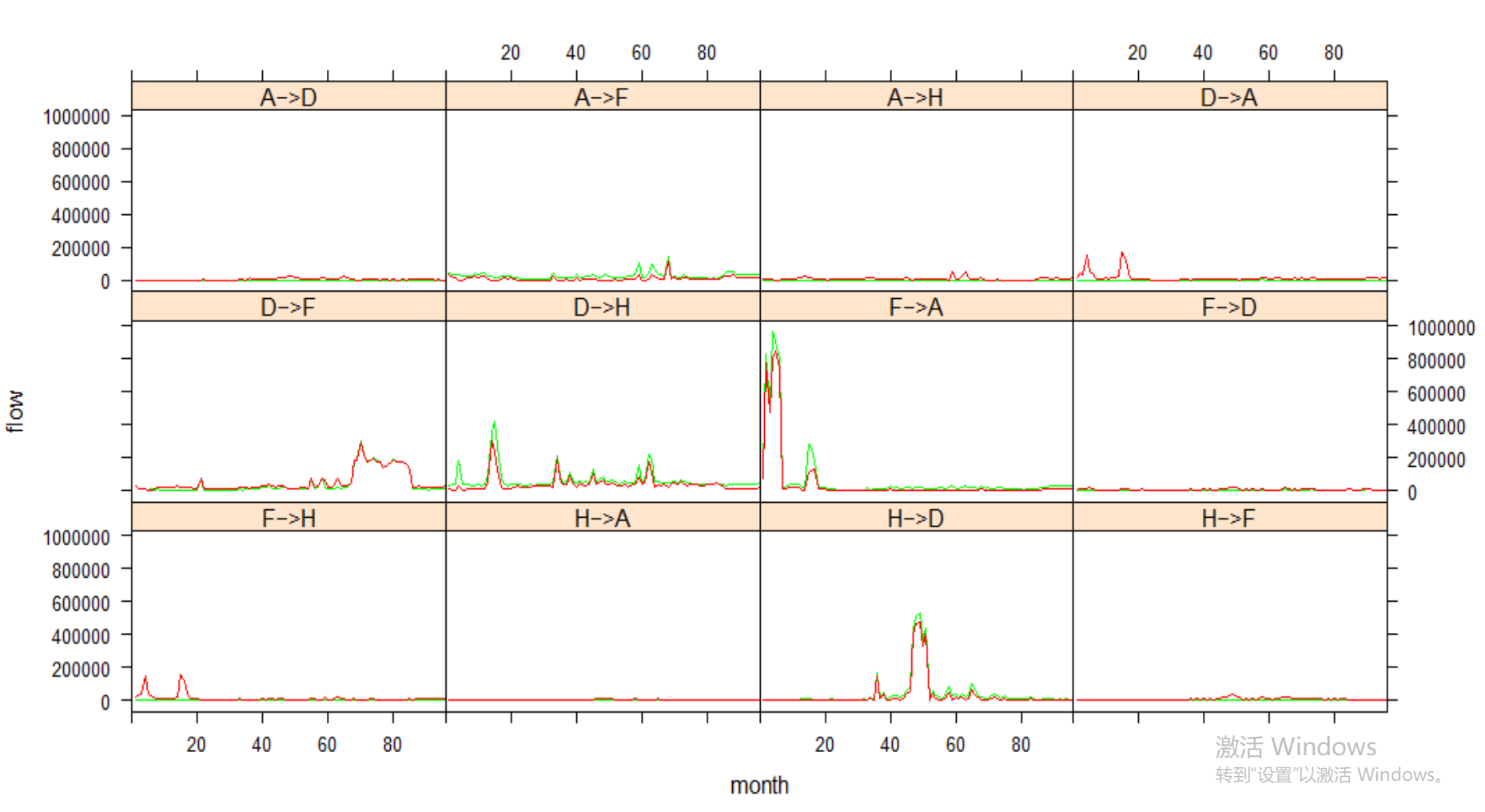


Here is the screenshot of the prediction. And in the next step, I will put the prediction and the actual value together and compare them.

**III Task II**

Below is the comparation between the actual value and estimated value flow. The green one is the actual value and the estimated value is marked as red one.

From the pictures below, it is not hard for us to find that the estimation is very good if ignoring some outliers. Most waves share the similar shapes. So, it is a good model.



**IV Appendix**

---

title: "Network\_Final"

author: "Chenrui Xu"

date: "2021/5/8"

output: html\_document

---

```{r}

library(sand)

library(networkTomography)

library(igraph)

library(statnet)

```

```{r}

netmat <- rbind(c(1,2),

c(2,1),

c(2,3),

c(2,9),

c(3,4),

c(3,5),

c(3,11),

c(4,3),

c(5,6),

c(5,7),

c(5,13),

c(6,5),

c(7,2),

c(7,8),

c(7,15),

c(8,7),

c(9,10),

c(10,3),

c(10,17),

c(11,12),

c(12,5),

c(12,17),

c(13,14),

c(14,7),

c(14,17),

c(15,16),

c(16,2),

c(16,17),

c(17,2),

c(17,3),

c(17,5),

c(17,7),

c(17,9),

c(17,11),

c(17,13),

c(17,15)

)

net <- network(netmat, matrix.type = "edgelist", directed = TRUE)

network.vertex.names(net) <- c("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O","P","Q")

gplot(net, vertex.cex = 3,

label.pos = 5, arrowhead.cex = 1,

displaylabels = TRUE, mode='fruchtermanreingold')

```

```{R}

path="C:/Users/StevenRui/Desktop/Network/Final"

Road=read.csv("C:/Users/StevenRui/Desktop/Network/Final/FinalProjectRoad.csv")

Transportation=read.csv("C:/Users/StevenRui/Desktop/Network/Final/FinalProjectTransportation.csv")

```

```{r}

head(Road)

```

```{r}

data=data.frame(Road[,c(2,3,4,6,7,9,10,11,13,14,15,17)])

data=as.matrix(data)

head(data)

dim(data)

```

```{r}

dim(data)

```

```{r}

B=read.csv("C:/Users/StevenRui/Desktop/Network/Final/B.csv",head=F)

B=as.matrix(B)

dim(B)

# B2=rbind(

# c(1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0),

# c(0,0,0,0,1,1,1,1,0,0,0,0,0,0,0,0),

# c(0,0,0,0,0,0,0,0,1,1,1,1,0,0,0,0),

# c(0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1),

# c(1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0),

# c(0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0),

# c(0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0),

# c(0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1)

# )

# dim(B2)

```

```{r}

head(B)

```

```{r}

tomo.fit <- tomogravity(data,B,0.01)

zhat <- tomo.fit$Xhat#[,c(2,3,4,5,7,8,9,10,12,13,14,15)]

dim(zhat)

head(zhat)

View(zhat)

```

```{r}

Z=Transportation[,2:13]

nt=nrow(Z)

nf=ncol(Z)

t.dat=data.frame(z=as.vector(c(Z)),

zhat=as.vector(c(zhat)),t=c(rep(as.vector(Road$month),nf)))

od.names=c (

rep('A−>D',nt),

rep('A−>F',nt),

rep('A−>H',nt),

rep('D−>A',nt),

rep('D−>F',nt),

rep('D−>H',nt),

rep('F−>A',nt),

rep('F−>D',nt),

rep('F−>H',nt),

rep('H−>A',nt),

rep('H−>D',nt),

rep('H−>F',nt))

t.dat=transform(t.dat,OD=od.names)

z=as.vector(c(Z))

z=sapply(z,as.numeric)

xyplot(z~t|OD, data=t.dat,

panel=function(x,y,subscripts){

panel.xyplot(x,y,type='l',col.line='green')

panel.xyplot(t.dat$t[subscripts],

t.dat$zhat[subscripts],

type='l',col.line='red')

},as.table=T,subscripts=T,xlim=c(0,96),

xlab='month',ylab='flow')

```

```{r}

View(Transportation)

View(zhat)

```