

hw2 - Network Analysis

Qian LI

4/23/2018

Intro

Below, I'm going to generate two graph models for iris data where the nodes represent the 150 flower observations. The first model is association model and the second is gaussian model.

Association model

Below, we load the iris data, and construct an unweighted graph where an edge represents whether or not the correlation between two observation is statistically significant at a 0.01 level.

(a)

There are 1846 correlation that is statistically significant at a 0.01 level, which is far less than 150×150 . The adjacent matrix is sparse according to the image, which is good.

```
iris.t <- t(iris[, -5])

m <- dim(iris.t)[1] # 4 features
n <- dim(iris.t)[2] # 150 observations

# correlation matrix
iris.cor <- cor(iris.t) # 150 * 150 matrix

# Fisher transform
z <- 0.5 * log((1 + iris.cor) / (1 - iris.cor))

# calculate p-values from Normal Distribution
z.vec <- z[upper.tri(z)]
corr.pvals <- 2 * pnorm(abs(z.vec), 0, sqrt(1 / (m-3)), lower.tail = FALSE)

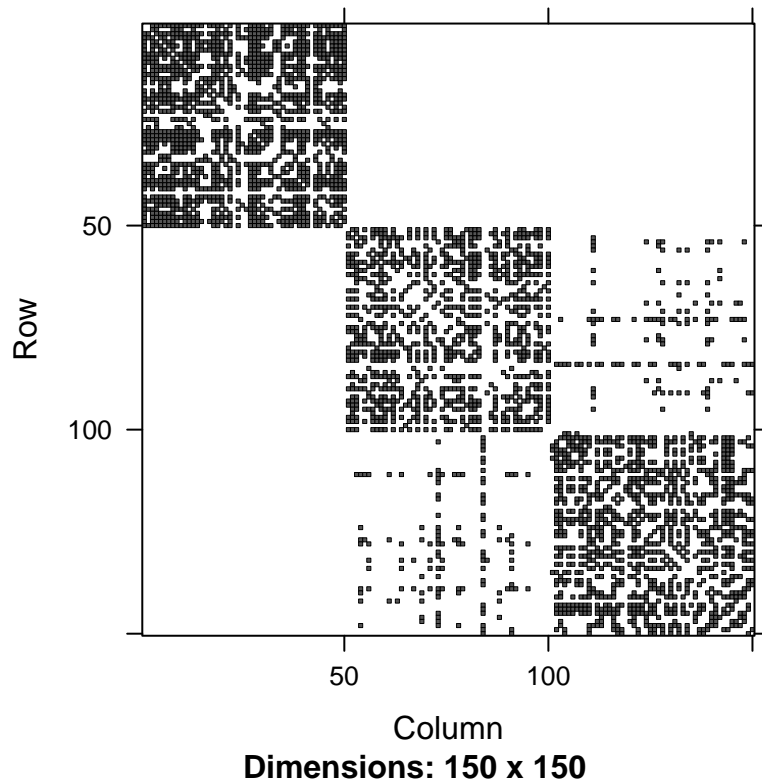
# adjust p-values using Benjamin-Hochberg multiple
corr.pvals.adjusted <- p.adjust(corr.pvals, "BH")

# how many values are significant at 0.01 level
length(corr.pvals.adjusted[corr.pvals.adjusted < 0.01])

## [1] 1846

# create a network out of this
iris.adjacency <- matrix(0, n, n)
# first fill in the upper triangle
iris.adjacency[upper.tri(iris.adjacency)] <- corr.pvals.adjusted < 0.01
# now fill the lower triangle with the transpose of the upper
iris.adjacency <- iris.adjacency + t(iris.adjacency)
```

```
# image of the adjacency matrix
image(Matrix(iris.adjacency))
```



(b)

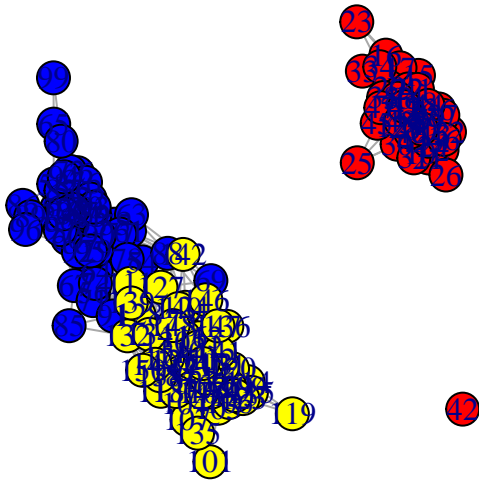
Below is the plot of the network with vertices colored based on different species names.

```
# convert the adjacency matrix to an igraph object
iris.g <- graph.adjacency(iris.adjacency, mode = "undirected")

# color the vertices based on species name
V(iris.g)[iris$Species == "setosa"]$color <- "red"
V(iris.g)[iris$Species == "versicolor"]$color <- "blue"
V(iris.g)[iris$Species == "virginica"]$color <- "yellow"
igraph_options(label=NA)

plot(iris.g)
title("Iris Network - Association Model")
```

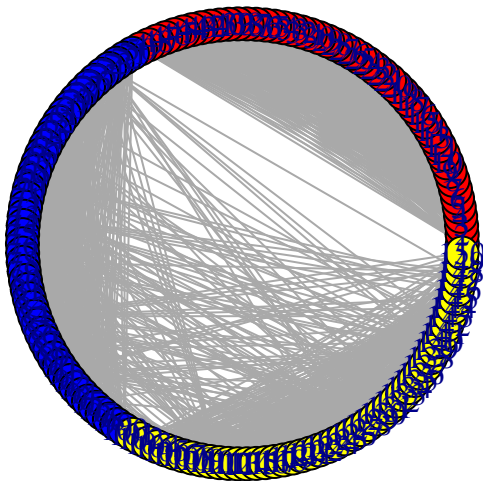
Iris Network – Association Model



(c)

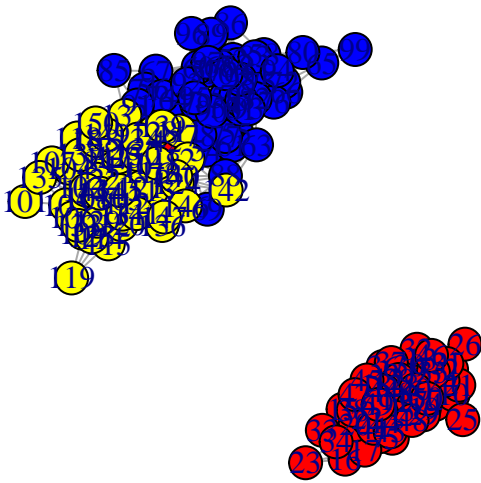
After experimenting with different layouts, Kamada-Kawai best represents the inherent structure. So below, I'm displaying the network using Kamada-Kawai layout.

```
plot(iris.g, layout=layout.circle)
```

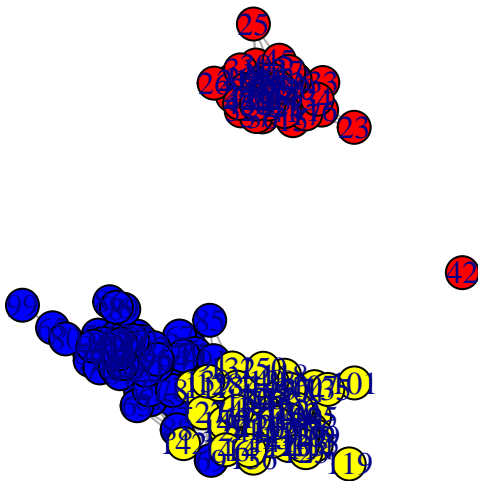


```
plot(iris.g, layout=layout.kamada.kawai)  
title("Iris Network: \nKamada-Kawai Layout")
```

Iris Network: Kamada–Kawai Layout



```
plot(iris.g, layout=layout.fruchterman.reingold)
```



Gaussian graphical model

Below, we provide fit a Gaussian graphical model to the iris data.

(a)

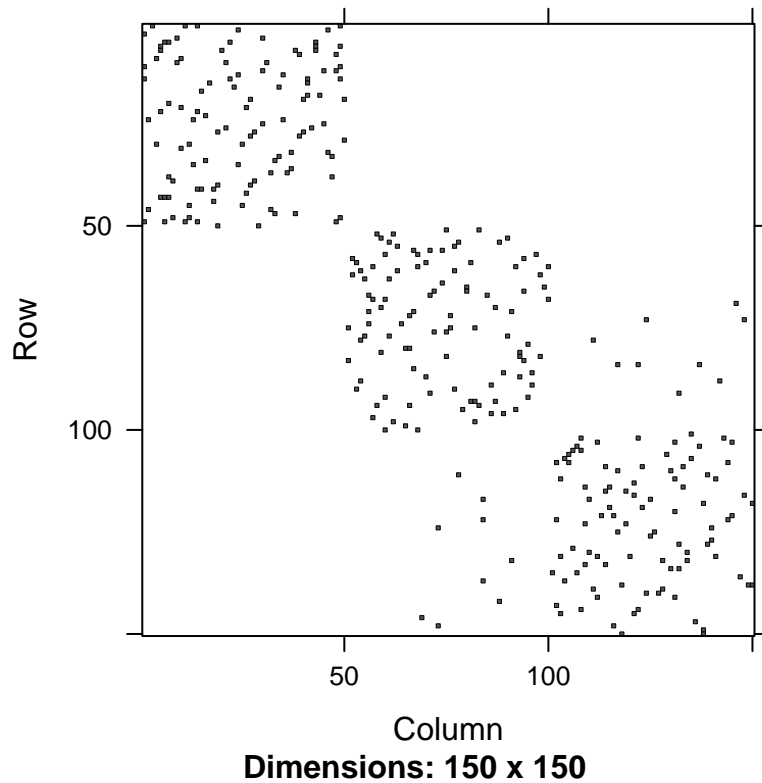
```
# create a huge object for analysis from the observed data matrix  
# this is fitting the Gaussian graphical model to this data  
huge.out <- huge(scale(iris.t))
```

```
## Conducting Meinshausen & Buhlmann graph estimation (mb)...done  
# Now, we need to select which partial correlation values are  
# statistical significant. There are two primary ways of doing this  
# The first is known to be prone to under-selection. This is seen
```

```
# in the empty graph formed using the "ric" criterion below
huge.opt1 <- huge.select(huge.out, criterion = "ric")
```

```
## Conducting rotation information criterion (ric) selection....done
## Computing the optimal graph....done
```

```
# Plot matrix
image(huge.opt1$refit)
```



(b)

```
#Create a graph object and plot the result
igraph <- graph.adjacency(huge.opt1$refit, mode = "undirected")
summary(igraph) #150 nodes and 162 edges
```

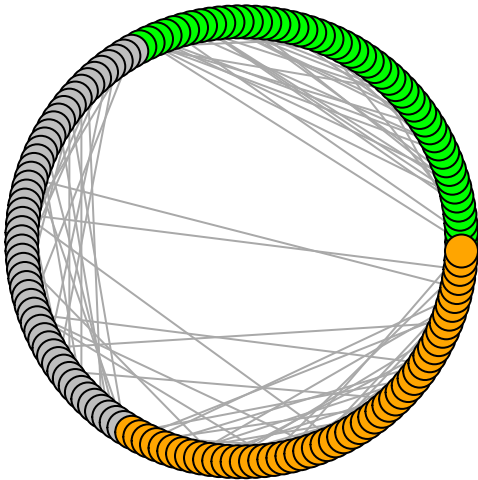
```
## IGRAPH 52d1732 U--- 150 162 --
```

```
# color the vertices based on species name
V(igraph)[iris$Species == "setosa"]$color <- "green"
V(igraph)[iris$Species == "versicolor"]$color <- "grey"
V(igraph)[iris$Species == "virginica"]$color <- "orange"
```

(c)

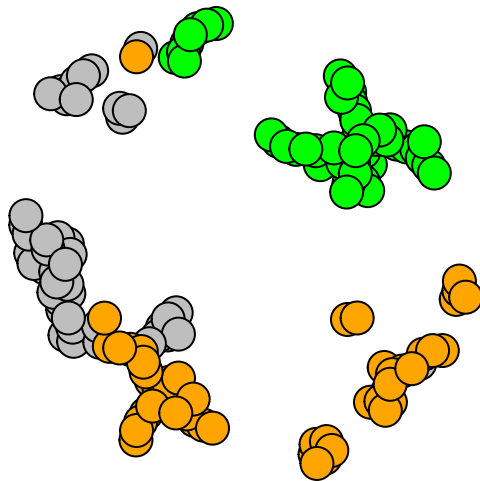
```
igraph.options(vertex.label=NA)
plot(igraph, layout = layout.circle)
title("Iris Network - Gaussian Model (Circle)")
```

Iris Network – Gaussian Model (Circle)



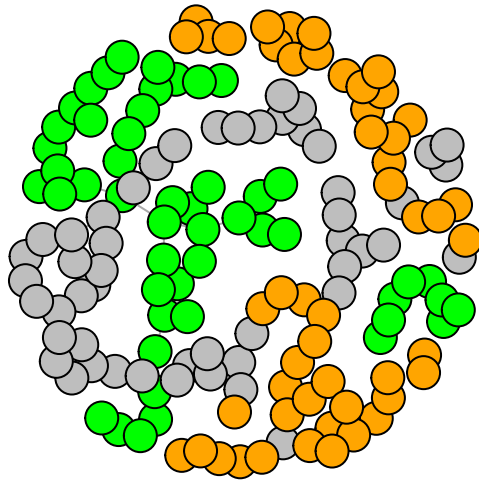
```
plot(igraph, layout = layout.kamada.kawai)  
title("Iris Network – Gaussian Model (Kamada Kawai)")
```

Iris Network – Gaussian Model (Kamada Kawai)



```
plot(igraph, layout = layout.fruchterman.reingold)  
title("Iris Network – Gaussian Model (Fruchterman Reingold)")
```

Iris Network – Gaussian Model (Fruchterman Reingold)



(d)

Similarities: - Adjacency matrix: from both methods show similar pattern; - Circle layout: - Kamada Kawai layout: - Fruchterman Reingold:

Differences: - Adjacency matrix: - Circle layout: - Kamada Kawai layout: - Fruchterman Reingold: