Network Science Homework 3

Arghya Kannadaguli (ak5357)

2025-02-21

```
library(igraph)
library(tidyverse)
library(ggplot2)
library(sand)
```

Assignment Description: Visualize a medium-size real-world network with 100 – 1000 nodes, try different layouts and discuss which layout can better highlight the structural features of the network.

Physicians Network Dataset

“This directed network captures innovation spread among 246 physicians in for towns in Illinois, Peoria, Bloomington, Quincy and Galesburg. The data was collected in 1966. A node represents a physician and an edge between two physicians shows that the left physician told that the right physician is his friend or that he turns to the right physician if he needs advice or is interested in a discussion. There always only exists one edge between two nodes even if more than one of the listed conditions are true.”

Data Source: The Konnect Project (http://konect.cc/networks/moreno_innovation/)

Data Import

Find a real-world network.

Here we are pulling in the physicians network described above.

```
physician_edgelist =
  readr::read_delim("data/moreno_innovation/out.moreno_innovation_innovation", skip = 1) |>
  janitor::remove_empty(which = "cols") |>
  janitor::clean_names() |>
  rename(node1 = percent, node2 = x1098)

physician_graph =
  as.matrix(physician_edgelist) |>
  graph_from_edgelist()

rm(physician_edgelist)
```

Here we are pulling in the `aidsblog` network used in class examples, for comparison with visuals made using the physicians network.

```
data(aidsblog)
```

Network Size

How many nodes and edges are in the network?

The physicians network has **436 nodes** and **1098 edges**. In comparison, the aidsblog network has 146 nodes and 187 edges.

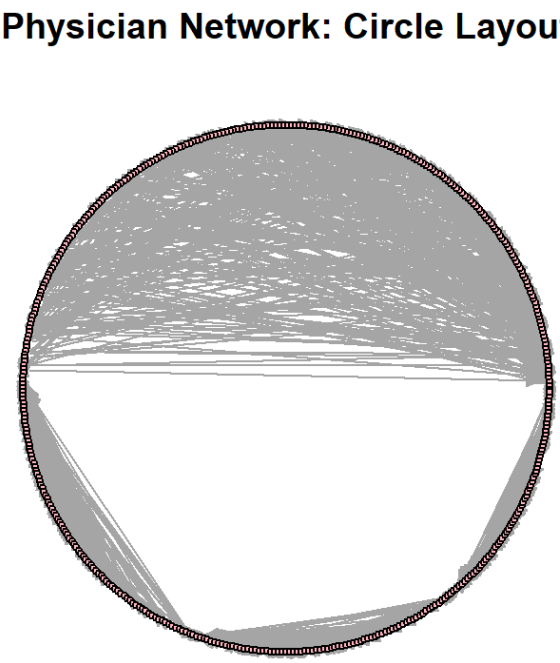
Visualization

Try different layouts for the graph to see which is most informative.

Circular Layout: `layout_in_circle`

This layout does not appear particularly useful. Although the circle looks nice, it is difficult to visually make out how many nodes there might be, this plot does not tell us anything about the organization of the edges.

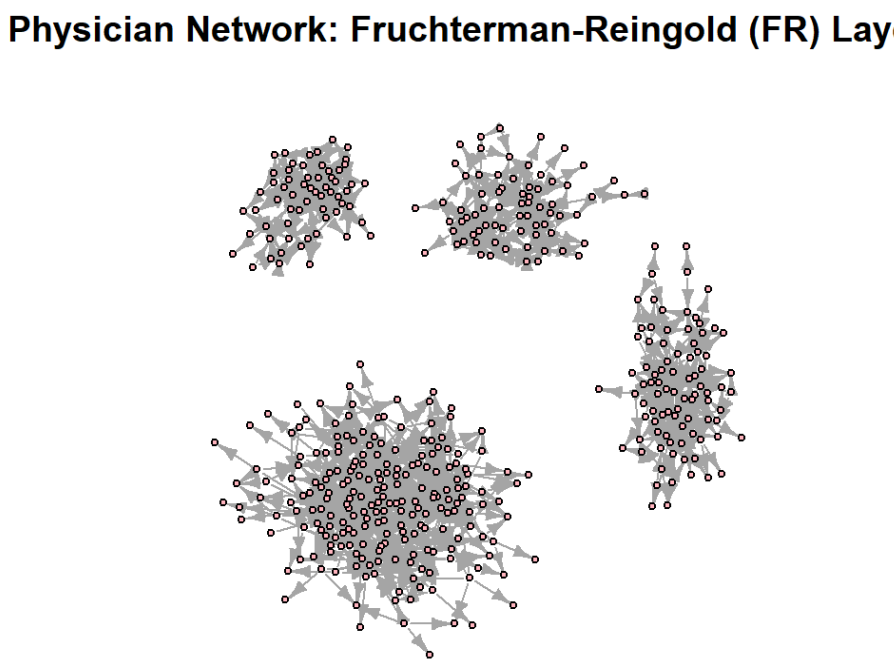
```
igraph_options(
  vertex.size=3,
  vertex.label=NA,
  vertex.color = "lightpink",
  edge.arrow.size=0.5)

plot(physician_graph, layout=layout_in_circle)
title("Physician Network: Circle Layout")
```



Fruchterman-Reingold (FR) Layout: `layout_with_fr`

This layout appears to be more useful than the last. Here we can see that the network has four different components, and we can also visualize the size of each of the components. As for visualizing patterns among the edges, we can now make out some nodes along the edges of each component with low degrees. However, the edges within each component still look somewhat jumbled and we cannot make out any patterns within them (if there are indeed patterns).

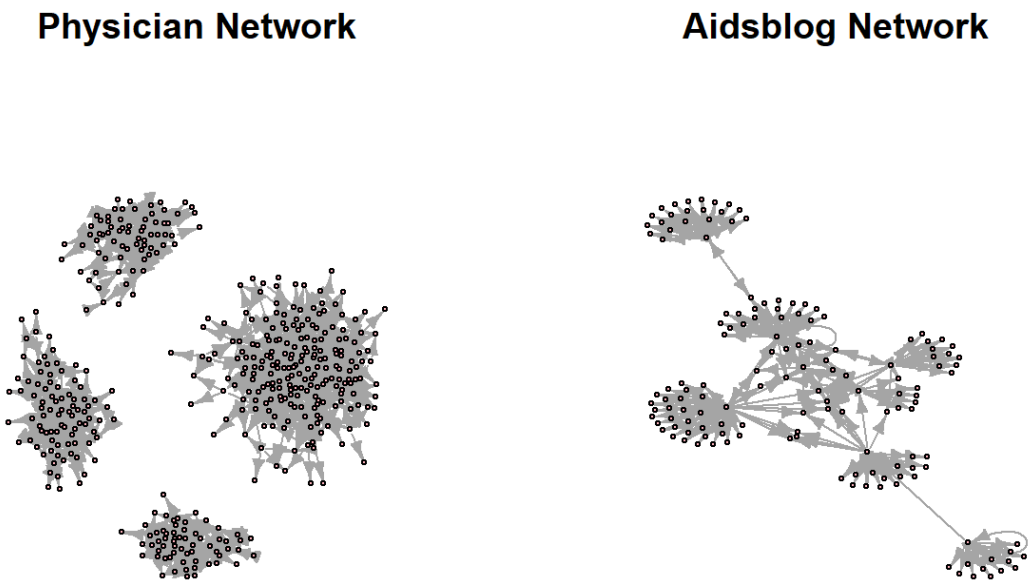
```
plot(physician_graph, layout=layout_with_fr)
title("Physician Network: Fruchterman-Reingold (FR) Layout")
```



However, let's compare this with the example shown in class. When visualizing the aidsblog network using a FR layout, we *do* see some patterns among the edges. Perhaps the directional patterns of the edges in the physician network are more difficult to clearly visualize.

I think this could be because the AIDS disease would spread from one person to one or more other people, whereas in the physicians network many physicians are linked with many other physicians. Perhaps one-to-many relationships are easier to visualize than many-to-many using the FR layout.

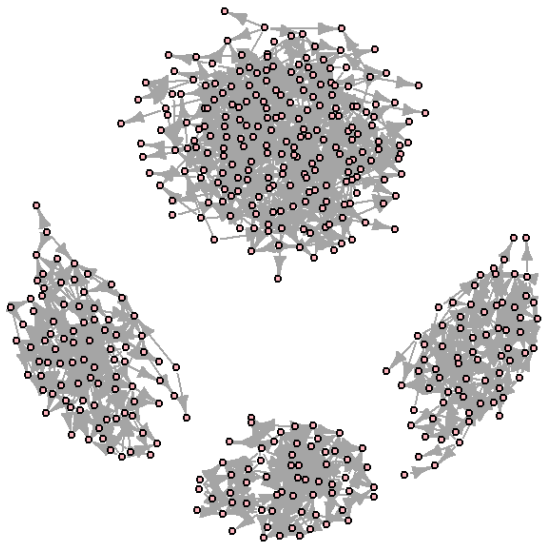
```
par(mfrow=c(1,2)) #set graphical parameters, create multiple plots
plot(physician_graph, layout=layout_with_fr)
title("Physician Network")
plot(aidsblog, layout=layout_with_fr)
title("Aidsblog Network")
```



Kamada-Kawai (KK) Layout: `layout_with_kk`

The KK layout appears better at visually elucidating some of the patterns in the edges and the overall network. Similar to the FR layout, we can clearly see four separate components in the physicians network. Similar to the FR layout, we can see low-degree nodes along the edges of each component, albeit not as clearly. In that sense, perhaps the FR layout is a bit more helpful in visualizing this graph than the KK layout.

```
par(mfrow=c(1,1))
plot(physician_graph, layout=layout_with_kk)
title("Physician Network: Kamada-Kawai (KK) Layout")
```



Other Layouts

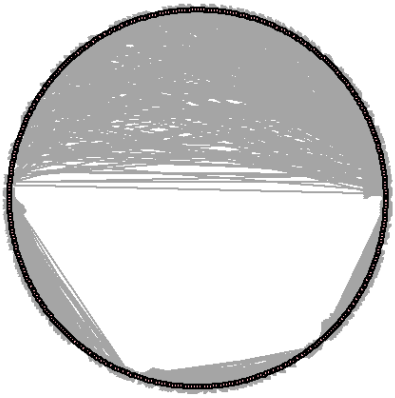
Let's try a few other layouts that RStudio auto-suggests.

Star Layout: `layout_with_star`

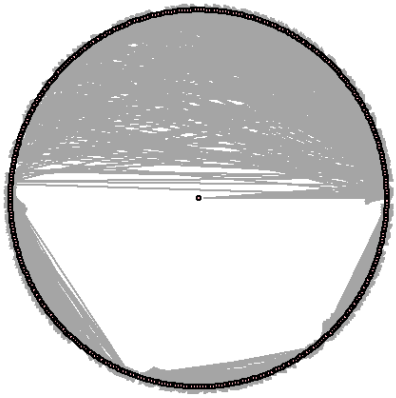
Similar to the circle layout, this one is not useful.

```
par(mfrow=c(1,2))
plot(physician_graph, layout=layout_in_circle)
title("Circle Layout")
plot(physician_graph, layout=layout_as_star)
title("Star Layout")
```

Circle Layout



Star Layout

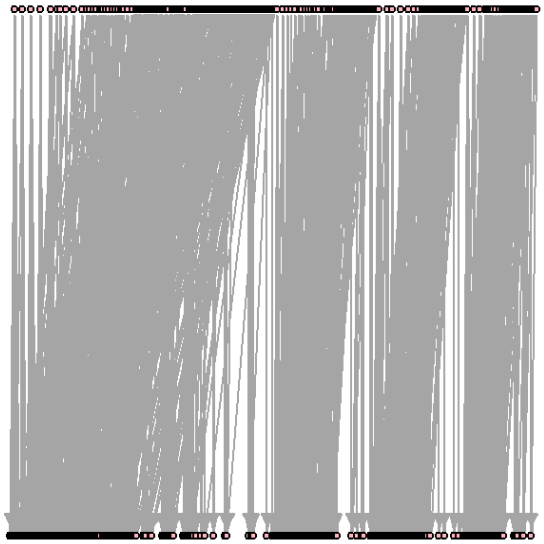


Tree and Sugiyama Layout: `layout_as_tree` and `layout_with_sugiyama`

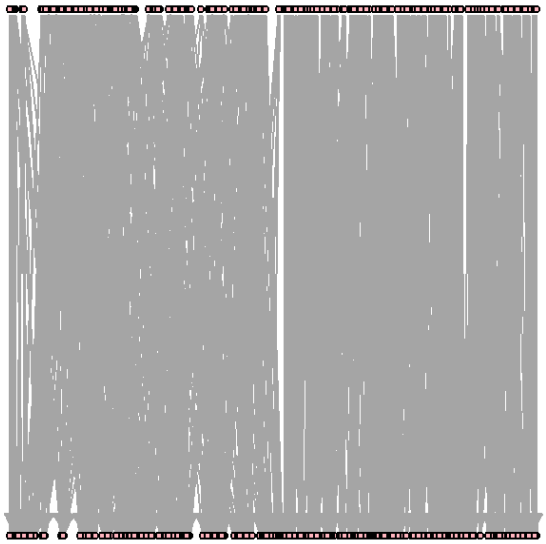
These layouts would be useful for a smaller bivariate network, but are not particularly helpful for the physicians network.

```
plot(physician_graph, layout=layout_as_tree)
title("Tree Layout")
```

Tree Layout



```
plot(physician_graph, layout=layout_with_sugiyama)
title("Sugiyama Layout")
```

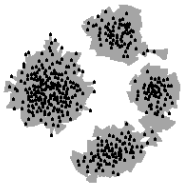


Component-showing Layouts: `layout_nicely` , `layout_components` , `layout_with_mds` , `layout_with_graphopt`

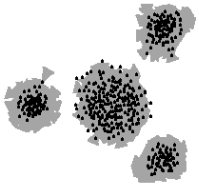
These layouts are all helpful in visualizing the components of the network. From farther away, they all look fairly similar.

```
par(mfrow=c(2,2))
plot(physician_graph, layout=layout_nicely)
title("Layout Nicely")
plot(physician_graph, layout=layout_components)
title("Layout Components")
plot(physician_graph, layout=layout_with_mds)
title("MDS Layout")
plot(physician_graph, layout=layout_with_graphopt)
title("Graphopt Layout")
```

Layout Nicely

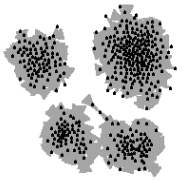


Layout Components



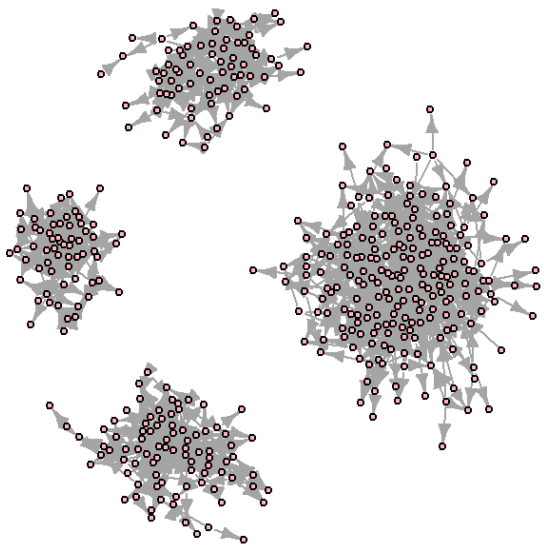
MDS Layout

Graphopt Layout

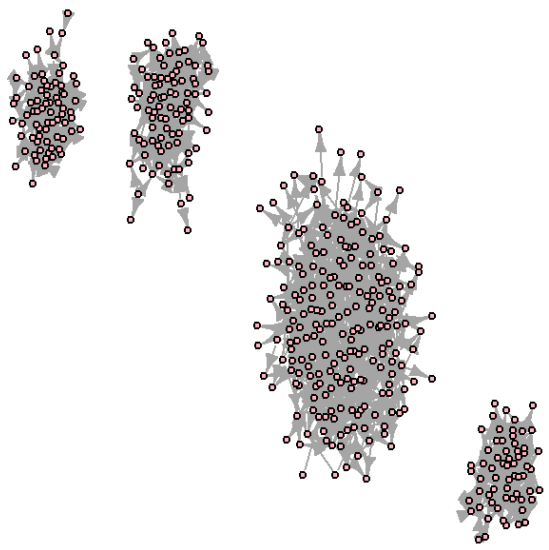


Upon closer inspection, I think that `layout_with_graphopt` produces the best and visually clearest result.

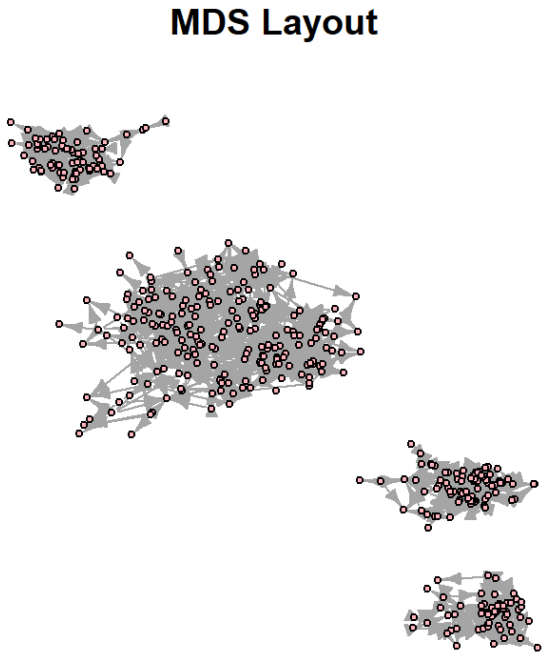
```
par(mfrow=c(1,1))
plot(physician_graph, layout=layout_nicely)
title("Layout Nicely")
```



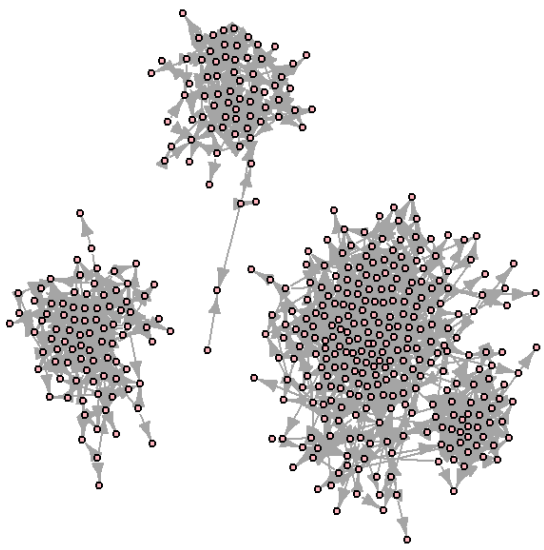

```
plot(physician_graph, layout=layout_components)
title("Layout Components")
```




```
plot(physician_graph, layout=layout_with_mds)
title("MDS Layout")
```



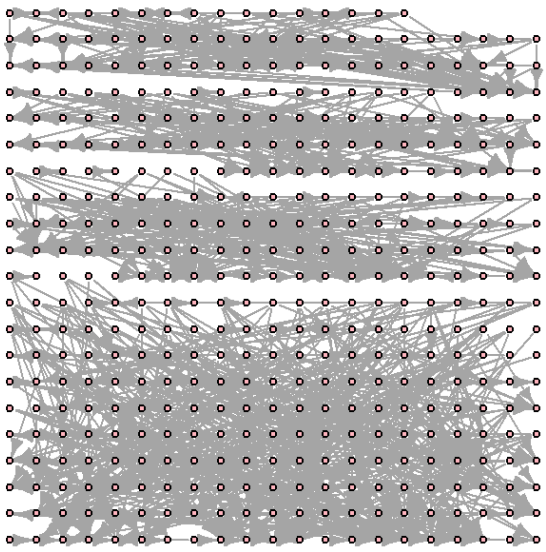

```
plot(physician_graph, layout=layout_with_graphopt)
title("Graphopt Layout")
```



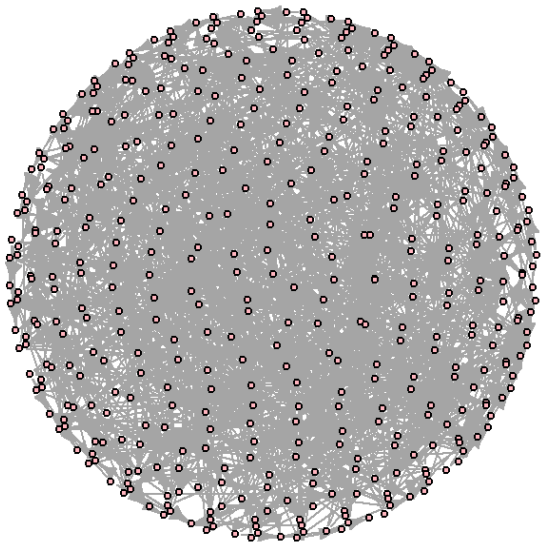
Shape-based Layouts: `layout_on_grid` and `layout_on_sphere`

These layouts look cool but are not useful.

```
plot(physician_graph, layout=layout_on_grid)
title("Layout on Grid")
```



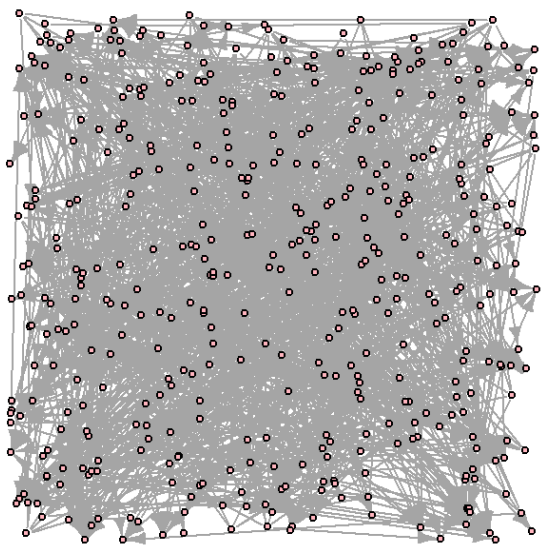

```
plot(physician_graph, layout=layout_on_sphere)
title("Layot on Sphere")
```



Random Layout: `layout_randomly`

Unsurprisingly, this layout does not tell us anything about the network other than giving a visual estimate of how many nodes there are, which most of the other layouts do as well.

```
plot(physician_graph, layout=layout_randomly)
title("Random Layout")
```

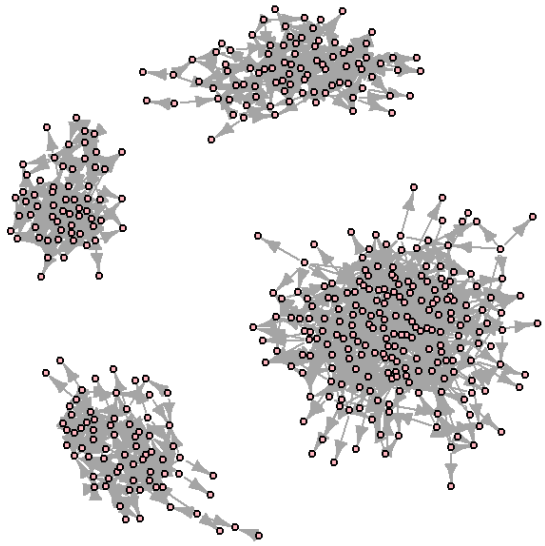


Conclusion

In conclusion, I think the FR layout is most useful in visualizing the physicians network. Here it is again, in all its beautiful glory.

```
plot(physician_graph, layout=layout_with_fr)
title("Physician Network: Fruchterman-Reingold (FR) Layout")
```

Physician Network: Fruchterman-Reingold (FR) Layout



The End :)