Big Data HW5

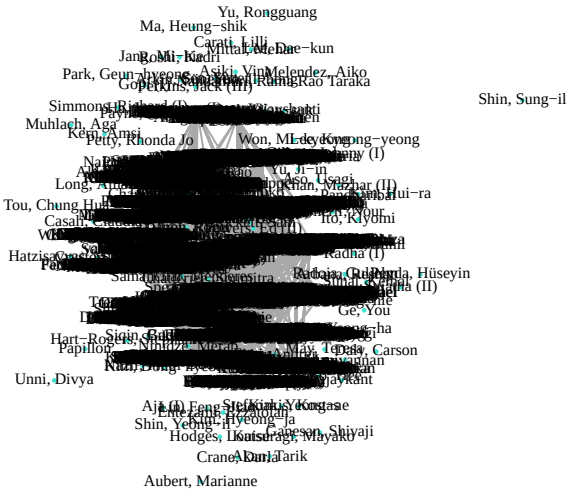
Elena Li

2024-04-29

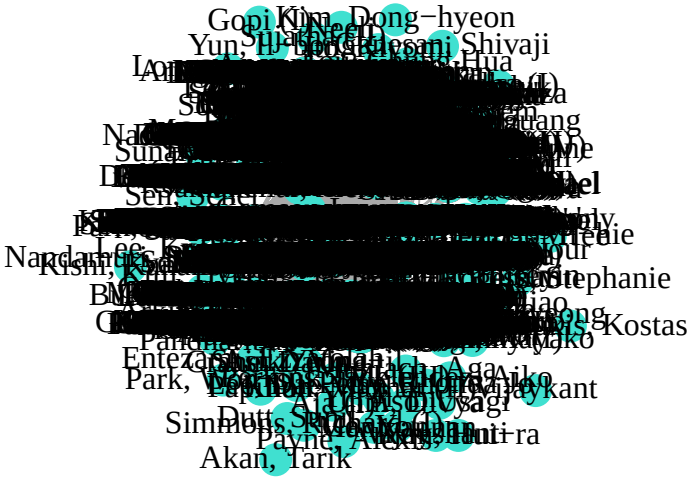
We'll explore casts for 'drama' movies from 1980-1999. See actors example code and data. I've limited the data to actors in more than ten productions over this time period (and to movies with more than ten actors).

[1] The actors network has an edge if the two actors were in the same movie. Plot the entire actors network.

```
plot(actnet, edge.arrow.size = 0.5, edge.curved = FALSE, vertex.size = 2, vertex.label.cex = 0.5)
```




```
plot(actnet)
```

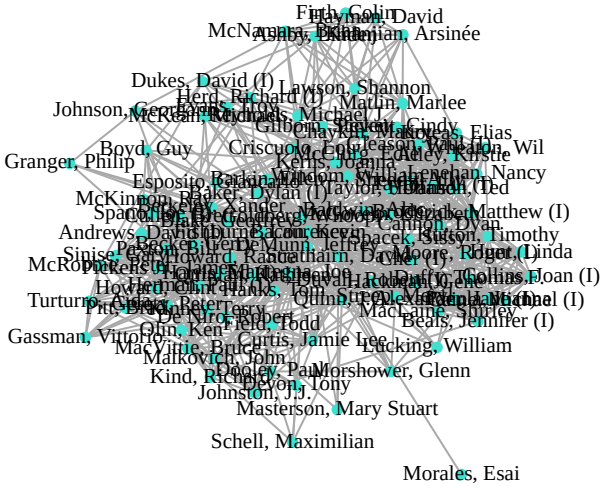


[2] Plot the neighborhoods for "Bacon, Kevin" at orders 1-3. How does the size of the network change with order?

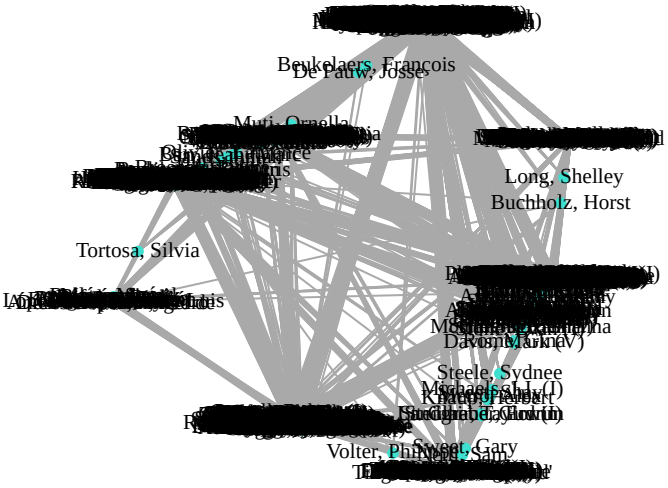
```
bacon <- which(V(actnet)$name == "Bacon, Kevin")

for (i in 1:3) {
  ind_actnet <- induced_subgraph(actnet, unlist(neighborhood(actnet, order=i, nodes=bacon)))
  plot(ind_actnet, main=paste ("Kevin's", i,"-order Neighborhood"), vertex.size=5,vertex.label.cex=0.7)
}
```

Kevin's 1 −order Neighborhood



Kevin's 2 −order Neighborhood



[3] Who were the most common actors? Who were most connected? Pick a pair of actors and describe the shortest path between them.

```
most_common_actors <- names(sort(nroles, decreasing = TRUE)[1:10])
most_connected_actors <- names(which.max(degree(actnet)))
cat("Most common actor:", most_common_actors, "\n")
```

```
## Most common actor: Zivojinovic, Velimir 'Bata' Jeremy, Ron Doll, Dora Dobtcheff, Vernon Berléand, Fra
```

```
cat("Most connect actor:", most_connected_actors)
```

```
## Most connect actor: Dobtcheff, Vernon
```

Pick a pair of actors and describe the shortest path between them.

We will look at the path between Gary Oldman and Moon Moon Sen.

```
shortest_path <- shortest_paths(actnet, from = "Oldman, Gary", to = "Sen, Moon Moon", mode = "all")$vpa
shortest_path
```

```
## + 4/7015 vertices, named, from c32038e:
## [1] Oldman, Gary    Gielgud, John   Lagoo, Shreeram Sen, Moon Moon
```

This means that the shortest path between Gary Oldman and Moon Moon Sen in the actor network involves four actors: Gary Oldman, John Gielgud, Shreeram Lagoo, and Moon Moon Sen. This path represents the shortest sequence of connections (movies they appeared in together) between the two actors in the network.

[4] Find pairwise actor-cast association rules with at least 0.01% support and 10% confidence. Describe what you find.

```
ts <- as(casttrans, "transactions")
rules <- apriori(ts, parameter = list(support = 0.0001, confidence = 0.1, maxlen=3))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.1    0.1    1 none FALSE            TRUE       5   1e-04      1
##  maxlen target  ext
##       3  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 1
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[6953 item(s), 14326 transaction(s)] done [0.01s].
## sorting and recoding items ... [6874 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.21s].
## writing ... [387929 rule(s)] done [0.10s].
## creating S4 object  ... done [0.04s].
```

```
inspect(head(rules, 10))
```

```
##      lhs                     rhs                   support      confidence
## [1]  {Mizutani, Kei}      => {Jô, Asami}           0.0001396063 1.0000000
## [2]  {Jô, Asami}          => {Mizutani, Kei}       0.0001396063 0.5000000
## [3]  {Magdalena, Rita}    => {Garcia, Emilio}      0.0001396063 1.0000000
## [4]  {Garcia, Emilio}     => {Magdalena, Rita}     0.0001396063 0.3333333
## [5]  {Illiopoulos, Dinos} => {Logothetis, Ilias}   0.0001396063 1.0000000
## [6]  {Logothetis, Ilias}  => {Illiopoulos, Dinos}  0.0001396063 0.2500000
## [7]  {Murali (II)}        => {Mammootty}           0.0001396063 1.0000000
## [8]  {Mammootty}          => {Murali (II)}         0.0001396063 0.1111111
## [9]  {Anisa}              => {Ashley, Brooke}      0.0001396063 1.0000000
## [10] {Ashley, Brooke}     => {Anisa}               0.0001396063 0.2857143
##      coverage     lift      count
## [1]  0.0001396063 3581.5000 2
## [2]  0.0002792126 3581.5000 2
## [3]  0.0001396063 2387.6667 2
## [4]  0.0004188189 2387.6667 2
## [5]  0.0001396063 1790.7500 2
```

```
## [6]  0.0005584252 1790.7500 2
## [7]  0.0001396063  795.8889 2
## [8]  0.0012564568  795.8889 2
## [9]  0.0001396063 2046.5714 2
## [10] 0.0004886221 2046.5714 2
```

For example, Rule [4] indicates that {Garcia, Emilio} is associated with {Magdalena, Rita} with a support of 0.0001396063 and a confidence of 0.3333333. This suggests that in the transactions dataset, when Garcia, Emilio appears, Magdalena, Rita also appears 33.33% of the time. The support value indicates that this combination appears in roughly 0.014% of the transactions.

[+] What would be a regression based alternative to ARules? Execute it for a single RHS actor.