BUS 41201 Homework 5 Assignment

Group 24: Shihan Ban, Yi Cao, Shri Lekkala, Ningxin Zhang

30 April 2024

Setup

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).

```
## actors network example

library(igraph)
```

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```
### GRAPH
## read in a graph in the `graphml' format: xml for graphs.
## it warns about pre-specified ids, but we want this here
## (these ids match up with the castlists in movies.txt)

actnet <- read.graph("actors.graphml",format="graphml")

### TRANSACTION
## read in the table of actor ids for movies
## this is a bit complex, because the movie names
## contain all sorts of special characters.

movies <- read.table("movies.txt", sep="\t",
    row.names=1, as.is=TRUE, comment.char="", quote="")

## it's a 1 column matrix.  treat it like a vector

movies <- drop(as.matrix(movies))

## each element is a comma-separated set of actor ids.
## use `strsplit' to break these out

movies <- strsplit(movies,",")
```

```
## and finally, match ids to names from actnet

casts <- lapply(movies,
    function(m) V(actnet)$name[match(m,V(actnet)$id)])

## check it

casts['True Romance']
```

```
## $`True Romance`
##  [1] "Arquette, Patricia"    "Ferrell, Conchata"     "Levine, Anna (I)"
##  [4] "Argo, Victor"          "Beach, Michael"        "Corrigan, Kevin (I)"
##  [7] "D'Angerio, Joe"        "Hopper, Dennis"        "Jackson, Samuel L."
## [10] "Lauter, Ed"            "Oldman, Gary"          "Penn, Chris (I)"
## [13] "Pitt, Brad"            "Rapaport, Michael (I)" "Rubinek, Saul"
## [16] "Walken, Christopher"
```

```
## format as arules transaction baskets

library(arules)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'arules'
```

```
## The following objects are masked from 'package:base':
##
##     abbreviate, write
```

```
casttrans <- as(casts, "transactions")


## Set up STM information

castsize <- unlist(lapply(casts, function(m) length(m)))

## see ?rep.int: we're just repeating movie names for each cast member

acti <- factor(rep.int(names(casts),times=castsize))

## actors

actj <- factor(unlist(casts), levels=V(actnet)$name)

## format as STM (if you specify without `x', its binary 0/1)

actmat <- sparseMatrix(i=as.numeric(acti),j=as.numeric(actj),
        dimnames=list(movie=levels(acti),actor=levels(actj)))

## count the number of appearences by actor
```

```
nroles <- colSums(actmat)

names(nroles) <- colnames(actmat)
```

Question 1

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

Question 2

Plot the neighborhoods for "Bacon, Kevin" at orders 1-3.

How does the size of the network change with order?

Question 3

Who were the most common actors?

Who were the most common actors?

Who were most connected?

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

Question 4

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

[Bonus]

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