

Cora Graph

Bill

July 2022

We confirm that the symmetric model can recover a well known citation network called the Cora network. This was one of the test networks used in Chanpuriya's paper. We start by reading in this data.

```
# Bill
setwd("/Users/billnunn/Desktop/Project/cora")
cora <- read.csv('cora.cites', sep = '\t' , header = FALSE)
# Adam
# cora = read.csv('./cora/cora.cites', sep = '\t' , header = FALSE)
head(cora)
```

```
##      V1      V2
## 1 35    1033
## 2 35   103482
## 3 35   103515
## 4 35  1050679
## 5 35  1103960
## 6 35  1103985
```

Irritatingly we have to replace each of the numbers with a number between one and the number of vertices to correctly run `graph_from_edgelist`. We first construct a data frame (using it like a python dictionary) in which the row names are the old node names and the entries are the new names.

```
node_names <- union(cora[,1], cora[,2])
length(node_names)
```

```
## [1] 2708
```

```
name_dict <- data.frame(new_names = 1:length(node_names),
                        row.names = node_names)
```

We now parse through the edge list and replace the old names with the new ones.

```
for(i in 1:5429){
  cora[i, 1] = name_dict[as.character(cora[i, 1]),1]
  cora[i, 2] = name_dict[as.character(cora[i, 2]),1]
}
```

And to check this has worked we print the union of entries in the cora data frame.

```
head(union(cora[,1], cora[,2]))
```

```
## [1] 1 2 3 4 5 6
```

Great, the nodes have been renamed to consecutive integers. We can now run `graph_from_edgelist` on the edgelist.

```
G <- graph_from_edgelist(as.matrix(cora))
a <- get.adjacency(G)
```

And we lastly check that the adjacency matrix is of the correct size.

```
dim(a)
```

```
## [1] 2708 2708
```

Blessed.