**OCR Supplement Steps**

1. Extract files from SEAL Lab site.

a) directors\_trustees-02.csv (data file)

b) merged07.csv (data file)

c) codebook 03\_17\_V3.pdf will also be helpful

2. Open both data files in R, create smaller files that can be opened in excel.

a) dirsub <- subset(directors\_trustees.02, year==2017)

merge1 <- subset(merged07, year==2017)

mergesub <- subset(merge1, province==ON)

write.csv(dirsub, ‘dirsub.csv’)

write.csv(mergesub, ‘mergesub.csv’)

3. Open both shortened files now as a csv to sort further and combine.

1. Sort mergesub.csv by city (for Ottawa, we included Gloucester, Kanata, Nepean, Osgoode, Ottawa, Rockcliffe, and Vanier), delete any unneeded rows
2. Sort both files by the column labeled ‘bn’
3. Copy the ‘bn’ column from “mergesub.csv” beside the data in “dirsub.csv”
4. Create a .xlsx file labeled “matching”
5. Go back to dirsub.csv and use the search and find function to find the matching bns (dirub.csv will have many more than mergesub.csv because it cannot be sorted by location). Copy the directors from dirsub.csv and the corresponding data in mergesub.csv into matching.xlsx.

5. Cleaning matching.xlsx

1. Create a new column beside the ‘last name’ ‘first name’ and ‘middle name/initial’ columns and call it ‘full name’
2. Highlight the ‘last name’ column and use the Conditional Formatting function to search for duplicates
3. For any last name that is duplicated, combine the first and last name. Either use all caps or no caps, and remove any symbols. Some names will have special characters representing letters with accents. Use this website (xxx) to confirm which letters should be replaced, and do not include an accent. Do not include the middle name or initial, but flag names that do have them. Flag anything that looks like it could be a typo or nickname to be manually checked later.
   1. E.g. JOHN SMITH-JONES -> johnsmithjones

6. Creating a node list

1. With a different colour, search for duplicates in the ‘full name’ column of matching.xlsx
2. Create a .csv file called “nodelist” with two columns: ‘id’ on column A and ‘label’ on column B
3. From matching.xlsx, copy the names of all organizations that contain duplicated directors into the ‘label’ column
4. In the ‘id’ column, call the first organization o001, the second o002, and so forth.
5. From matching.xlsx, copy the ‘full name’ of all duplicated directors into the ‘label’ column below the organizations. Ensure that each duplicated director is present but only listed once, regardless of how many boards they serve on. In the id column, call the first director p001, the second p002, and so forth.
6. Create a .csv file called “Nodes Legend”. Copy all of the organization names into column A. In column B, create a short form for each organization so that the labels on the network map will not be too long. Save this file, and replace the organization names in “nodeslist” with the shortforms you just created.

7. Creating an edge list

1. Create a .csv file called “edgelist” with two columns: ‘Director’ and ‘Org1’
2. For any organization that has a duplicate, copy the org name duplicate director name(s) from matching.xlsx into edgelist.csv. Be sure that if an org has more than one duplicated director, you put the org name beside each one, as shown below.

|  |  |
| --- | --- |
| **Director** | **Org1** |
| johnsmithjones | abc |
| sallysmith | jkl |
| sallysmith | ghi |
| johnsmithjones | def |
| sallysmith | mno |

|  |  |  |
| --- | --- | --- |
| **Director** | **Org1** | **Org2** |
| johnsmithjones | abc | def |
| sallysmith | ghi | jkl |
| sallysmith | ghi | mno |
| sallysmith | jkl | mno |

1. After this, use the search and find function to figure out which directors match with each other. Create a third column, labeled org 2. When a director is serving on two boards, their name should show up in one row with two orgs beside them (see johnsmithjones in the example below). When a director is serving on three boards, their name should show up in three rows so that each org can be matched with each other (see sallysmith).

8. Creating an edge matrix.

1. Create another .csv file and call it “edgematrix”
2. On the first row, starting with column B, insert the organization ids (o001, o002, o003, etc.)
3. On the first column, starting with row 2, insert the directors’ ids (p001, p002, p003, etc.)
4. Where an org and a person are connected, put a ‘1’. Everywhere else, put a ‘0’.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | o001 (abc) | o002 (def) | o003 (ghi) | o004 (jkl) | o005(mno) |
| p1 (john) | 1 | 1 | 0 | 0 | 0 |
| p2 (sally) | 0 | 0 | 1 | 1 | 1 |

9. Code for the map.

1. In R, use the following code to create your network map! It’s a good idea to first use a sample dataset to make sure that everything happens as you expect it should and that your nodelist and edgematrix are formatted correctly.

library(igraph)

nodes2 <- read.csv("nodelist.csv", header=T, as.is=T)

edges2 <- read.csv("edgematrix.csv", header=T, row.names = 1)

links2 <- as.matrix(edges2)

net2 <- graph\_from\_incidence\_matrix(links2)

# Modify nodes – directors will show up as an orange circle with no label, organizations have no shape, just the label

V(net2)$color <- c("NA", "orange")[V(net2)$type+1]

V(net2)$shape <- c("none", "circle")[V(net2)$type+1]

V(net2)$label <- ""

V(net2)$label[V(net2)$type==F] <- nodes2$media[V(net2)$type==F]

V(net2)$label.cex=.85

V(net2)$label.font=2

plot(net2, edge.width=1.5, edge.color="dimgray",vertex.label.color="black", vertex.frame.color=NA, vertex.size=2.5)

#Export the image as a 1500x1500 tiff file

1. Spot check your data to make sure there are no mistakes. Use “table(V(net2)$type)” to double check that you have the right number of nodes and edges.