

# Problem set 1 (CSS)

## Solutions

**1** In your own words, briefly describe what a social network is and why sociologists are interested in looking at those. **(2 points)**

A set of actors and the connections between those. Social scientists are interested in those because the placement in a network has different implications for social phenomena, like homophily or social capital.

**2** Examine the two networks shown below.



Figure 1: Networkgraphs

**2.1** Formally describe the networks. **(3 points)**

a): Ein ungerichtetes Netzwerk mit 9 Knoten und 13 Kanten

$G_a = \{V_a, E_a\}$

$V_a = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$E_a = \{\{1, 3\}, \{1, 7\}, \{1, 9\}, \{2, 7\}, \{2, 8\}, \{2, 9\}, \{3, 5\}, \{4, 5\}, \{4, 6\}, \{4, 7\}, \{4, 8\}, \{6, 7\}, \{7, 8\}\}$

b): Ein gerichtetes Netzwerk mit 9 Knoten und 15 Kanten

$G_b = \{V_b, E_b\}$

$Vb = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$Eb = \{(1, 3), (1, 7), (2, 8), (2, 9), (3, 6), (6, 3), (6, 5), (6, 8), (6, 9), (8, 4), (8, 6), (8, 7), (9, 2), (9, 6), (9, 7)\}$

**2.2** Provide one real-world example for each type of network shown in the figures. **(2 points)**

- a): Siblings in a family
- b): Mother to her children

**2.3** Based on the directed network (*figure b*), write down the corresponding adjacency matrix. **(3 points)**

```
M <- matrix(c(0,0,1,0,0,0,1,0,0,
              0,0,0,0,0,0,0,1,1,
              0,0,0,0,0,1,0,0,0,
              0,0,0,0,0,0,0,0,0,
              0,0,0,0,0,0,0,0,0,
              0,0,1,0,1,0,0,1,1,
              0,0,0,0,0,0,0,0,0,
              0,0,0,1,0,1,1,0,0,
              0,1,0,0,0,1,1,0,0),
            nrow = 9,
            byrow = TRUE)

p <- graph_from_adjacency_matrix(
  M,
  mode = "directed"
)

plot(p,
      edge.arrow.size = .2)

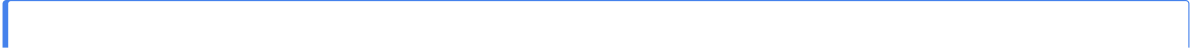
ecount(p)
```

---

**3** A city partnership (also known as a twin city or sister city relationship) is a formal agreement between two cities (most of the time from two different countries) to promote cultural and economic collaboration. The city of Leipzig has 15 such twin cities. On Moodle, you will find an R data file named `twin_cities.rds`, which contains a list of Leipzig's partner cities and their own partner cities. You can load the file into R using the following command:

```
twin_cities <- readRDS("your/path/twin_cities.rds")
```

**3.1** Using this data, create an igraph object where each node represents a city. Add an edge between two cities if they are partnered and include the country of each city as a vertex attribute. **(5 points)**



```

library(stringr)
twin_cities <- readRDS("Data/twin_cities.rds")

# Fix spelling differences
name_fixes <- c(
  "Addis Ababa" = "Addis_Abeba",
  "Frankfurt am Main" = "Frankfurt_am_Main",
  "Ho Chi Minh City" = "Ho_Chi_Minhtadt",
  "Kyiv" = "Kyjiw",
  "Krak w" = "Krakau"
)

# Apply fixes
names(twin_cities) <- str_replace_all(names(twin_cities),
                                     name_fixes
                                   )

# apply fixes to each twin city entry in the list
for (city in names(twin_cities)) {
  twin_cities[[city]]$cities <- str_replace_all(twin_cities[[city]]$cities,
                                               name_fixes
                                             )
}

#-----

# 1. Create a data frame of edges
edges <- data.frame(from=character(),
                    to=character(),
                    stringsAsFactors = FALSE
                  )

# 2. Create an empty vector for the countries
countries <- vector("list", length = length(twin_cities)
                  )

# 3. Fill the edges and countries list
for (city in names(twin_cities)) {
  for (twin in twin_cities[[city]]$cities) {
    edge <- sort(c(city, twin)) # Sort to avoid duplicate directions
    edges <- rbind(edges,
                  data.frame(from=edge[1],
                              to=edge[2],
                              stringsAsFactors = FALSE
                            )
                )
  }
}

# 4. Remove duplicate edges
edges <- unique(edges)

#unique(edges$from)

```

## Leipzig's twin-city network



Figure 2: Twin city-network

### 3.2 Categorize the form of data collection. (1 point)

The network is a egocentric network (Locally collected). Starting from Leipzig we collected all nodes that have a path  $p \leq 2$ .

### 3.3 Calculate the density of the network. Then explain in one or two sentences what this value tells you about the structure of the network. (2 points)

```
edge_density(g)
```

The density of the network is 0.0116. This means that just over 1 percent of all possible connections (if every city were connected to every other city) are actually realized in the network.

The low density can be attributed to the fact that only a few of the partner cities of Leipzig's partner cities are also connected among themselves.

**3.4** Determine the diameter of the network. What does this diameter represent in the context of the city partnership network? **(1 point)**

```
diameter(g)
```

The longest distance between two cities in the network is 4. This is due to the way the data was collected—it only includes cities that are at most two steps away from Leipzig (i.e., with a distance of  $d = 2$  from Leipzig).

**3.5** Find the shortest path between the cities Budapest and Wuhan. List all cities that lie along this path. **(1 point)**

```
# Find the shortest path between Hamburg and Tel Aviv
shortest_path <- shortest_paths(g, from = "Budapest", to = "Wuhan", output = "both")

# Print the shortest path
print(shortest_path)

length(shortest_path$vpath[[1]]) - 1

# Extract the vertices in the shortest path
path_vertices <- shortest_path$vpath[[1]]

# Print the cities (vertices) in the shortest path
print(V(g)[path_vertices]$name)
```

The shortest path goes through the cities Krakau and Kyjiw (or possibly through other cities (there can be short paths of the same length via different nodes.))

**3.6** Describe the distribution of shortest paths in the network. Then interpret what this distribution tells you about the overall connectivity of the cities. **(\*2 points)**

The distribution of the shortest paths in the network is left-skewed and shows that only a few cities are connected to each other through one, two, or three edges. Most cities have a distance of 4.

The distribution of the edges indicates exponential growth in the number of connections: the further one moves away from Leipzig (the “ego”), the more connections there are to other cities, since each newly connected city brings its own network of partner cities into the overall network. At the same time, we can assume that the network reaches a saturation point at some stage—if we were to include all the cities in the world that participate in a sister city program in our dataset.