

Exercises for Statistical analysis of network data – Sheet 9

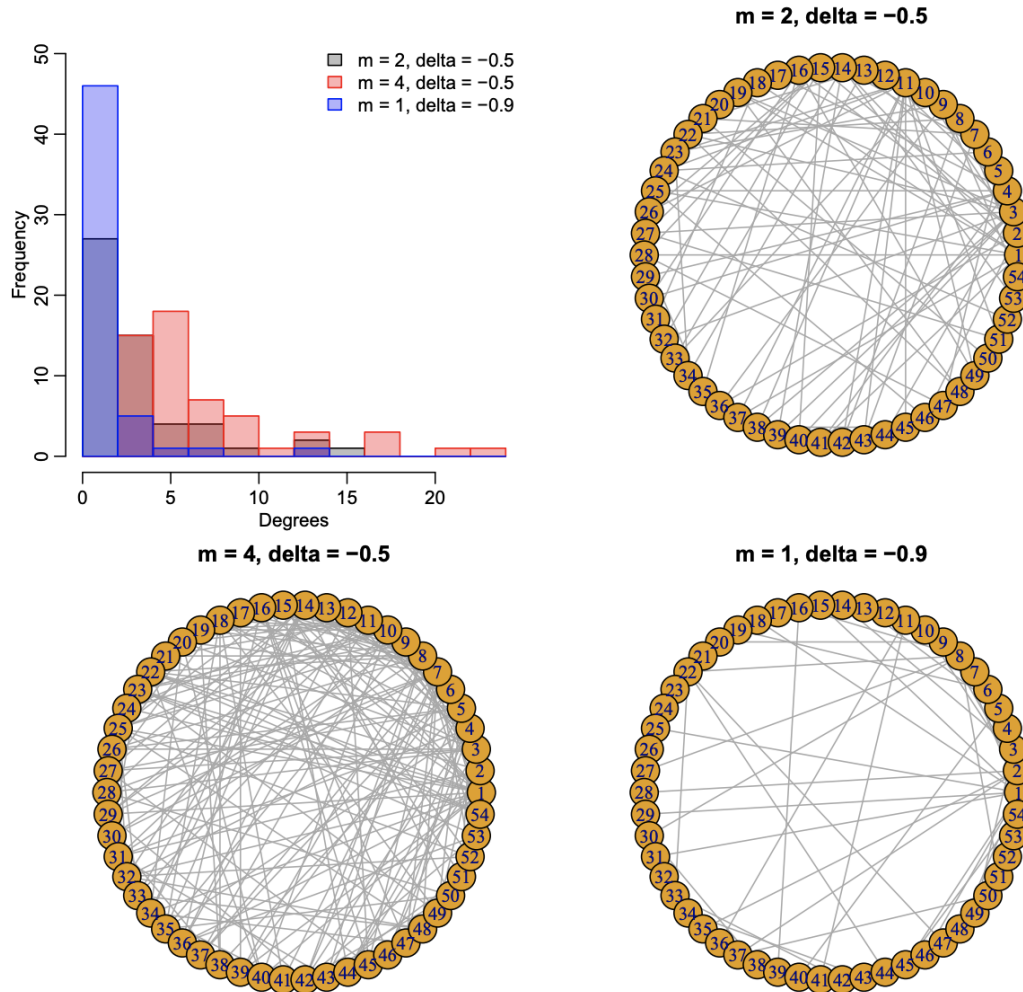


Figure 1:

1. The resulting degree distributions are shown in the top left panel of Figure 1. For illustration, the other three panels show the three graphs corresponding to the histograms. The nodes are numbered in the order of adding them, with 1,2,3 and 4 forming the initial 4-star. It can be observed how the nodes added early have more links than those added late, and how the density of the network depends on the two parameters. A short script in R, producing the above figures, is given here.

```
library("igraph")

s4 <- make_star(4, mode = "undirected")

set.seed(108)
g1 <- sample_pa(54, m = 2, zero.appeal = 0.5, directed = FALSE, start.graph = s4)
d1 <- degree(g1)
g2 <- sample_pa(54, m = 4, zero.appeal = 0.5, directed = FALSE, start.graph = s4)
d2 <- degree(g2)
```

```

g3 <- sample_pa(54, m = 1, zero.appeal = 0.9, directed = FALSE, start.graph = s4)
d3 <- degree(g3)

pdf(file = "sheet9_BAmodel_degredistr.pdf", height = 8, width = 8, pointsize = 12)

par(mfrow = c(2,2), mar = c(2.6,2.6,2,1), mgp = c(1.6,0.6,0))
br <- seq(0, 24, by = 2)
hist(d1, col = "grey", main = "", xlab = "Degrees", breaks = br, ylim = c(0,50))
hist(d2, col = rgb(1,0,0,0.3), border = rgb(1,0,0,1), add = TRUE, breaks = br)
hist(d3, col = rgb(0,0,1,0.3), border = rgb(0,0,1,1), add = TRUE, breaks = br)
legend("topright", legend = c("m = 2, delta = -0.5", "m = 4, delta = -0.5",
  "m = 1, delta = -0.9"), fill = c("grey", rgb(1,0,0,0.3), rgb(0,0,1,0.3)),
  border = c("black",rgb(1,0,0,1),rgb(0,0,1,1)), bty = "n")
plot(g1, layout = layout_in_circle, main = "m = 2, delta = -0.5")
plot(g2, layout = layout_in_circle, main = "m = 4, delta = -0.5")
plot(g3, layout = layout_in_circle, main = "m = 1, delta = -0.9")

dev.off()

```

2. See Figure 1, the blue histogram in the top left panel and the graph in the bottom right panel.
3. The adjacency matrices of the three layers:

$$A_1 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Since the exercise does not specify any cross-layer connections, the set C is empty here. If we take the system as a multiplex network, and identify the nodes by their number (the hub in G_1 and the two endpoints in G_3 have particular roles, but all other nodes are interchangeable within their layers, and node number 6 is an isolate in G_1), we can obtain the projection graph $\tilde{A} = A_2$, as equal to the complete graph.

4. The question is whether we can sample a sufficient number of edges to cover a large enough part of the network for a meaningful inference. Let us suppose that the number N of nodes in the network is very large, and that we start the procedure with d_0 nodes that are “far” from each other, i.e., they have no contacts with each other and no overlapping neighbourhoods. Snowball sampling in an Erdős-Rényi graph with edge probability ρ will result in finding on average $\rho(N-1)d_0$ new edges, and therefore $\rho(N-1)d_0$ new points to check for edges. With a reasoning similar to that applied for the spread of infections, we can say that our sampling process will stop and die out soon if $\rho(N-1)d_0 < 1$, and can be maintained for a sufficient number of steps is $\rho(N-1)d_0 > 1$.

Snowball sampling using a single starting node in a star will cover the whole star: to the centre of the star, all other nodes are adjacent. So, even if we start from one of the spokes, in one step we reach the centre, and then straightforwardly all other nodes. This means if we do snowball sampling using only one step, we will either sample two nodes, one spoke and the centre (if we start from a spoke) or all (if we start from the centre). In two steps, we can cover the whole network.