

Exercises: Zachary's karate-club network

YSC2210 - DAVis with R

Michael T. Gastner

Introduction

A classic study in social network science was conducted by [Zachary \(1977\)](#). He observed the interactions between members of a karate club at a US university. Each member is represented by a node. An edge between two nodes indicates that these two members also engaged in friendly relations outside the club's lessons and meetings.

By the end of Zachary's study, a conflict arose between the club president and the karate instructor. As a consequence, supporters of the instructor resigned from the club and started a new organisation. Zachary's observational data are often used in models that try to predict how networks split into groups if there is a conflict between members.

Zachary's network is *undirected* (i.e. an edge from i to j is equivalent to an edge from j to i).

Objectives

Using Zachary's data as example, we learn how to obtain quantitative information about networks with the **igraph** and **tidygraph** packages. We also develop more familiarity with **ggplot2** and the add-on packages **ggraph** and **ggforce**.

Data

Download `soc-karate.zip` from <https://networkrepository.com/soc-karate.php> ([Rossi and Ahmed, 2015](#)). The first few lines in `soc-karate.zip` are metadata. Afterwards, there are lines with exactly two integers, starting with the line "2 1". This portion of the file is the edge list.

There is no separate file with node identifiers. You can assume that all nodes in the network appear at least once in the edge list.

Tasks

- (1) Import the data and create an undirected `tbl_graph` called `karate_netw`.
- (2) How many nodes and edges are in the network?
- (3) Are there parallel edges or loops in the network?
- (4) Node 1 is the karate instructor. Node 34 is the club president. Add a node attribute called `role` to the `tbl_graph` with the values `Instructor`, `President` and `Other`.
- (5) Show that the two nodes with the highest degree are the president and the instructor.
- (6) Find all shortest paths between the instructor and the president.
- (7) What is the local clustering coefficient of the instructor? What is the local clustering coefficient of the president? How do you interpret these numbers?
- (8) Which are the largest cliques in the network? Which nodes are in these cliques?

- (9) Confirm that the network is connected (i.e. there is only one component).
- (10) Which communities does the Louvain algorithm detect? Use **tidygraph** to add information about the communities to `karate_netw`.
- (11) Make a plot similar to figure 1.

Manually tweaking the node positions is time-consuming; thus, I already did it for you. Download `karate_node_positions.csv` from our Canvas assignment page. Then use the following hack. Here I assume you used "node" as column name in the node tibble. If not, please change the second argument in `left_join()` below.

```
karate_layout <- create_layout(karate_netw, layout = "stress")
layout_attributes <- attributes(karate_layout)
node_positions <- read_csv("karate_node_positions.csv")
karate_layout <-
  karate_layout |>
  select(-x, -y) |>
  left_join(node_positions, by = "node") |>
  select(x, y, everything())
attributes(karate_layout) <- layout_attributes
ggraph(karate_layout) + ...
```

Here are a few properties that the final figure should have.

- Use `geom_mark_hull()` to show the communities detected by the Louvain algorithm.
- The nodes that stayed in the old club were 10, 15, 16, 19, 21 and 23—34. Indicate the observed split with a Bezier curve (see `?geom_bezier` in the **ggforce** package).
- Add the annotation 'Observed split' near the Bezier curve.
- Draw the edges as thin grey lines and the Bezier curve as thick black line.
- Use colour and shape to identify the node roles. Use a combined legend for colour and shape. Ensure that the legend does not overlap with any nodes or edges.
- Add a title.
- Acknowledge the source in a caption.

Feel free to make more adjustments if you think they improve the quality of the plot.

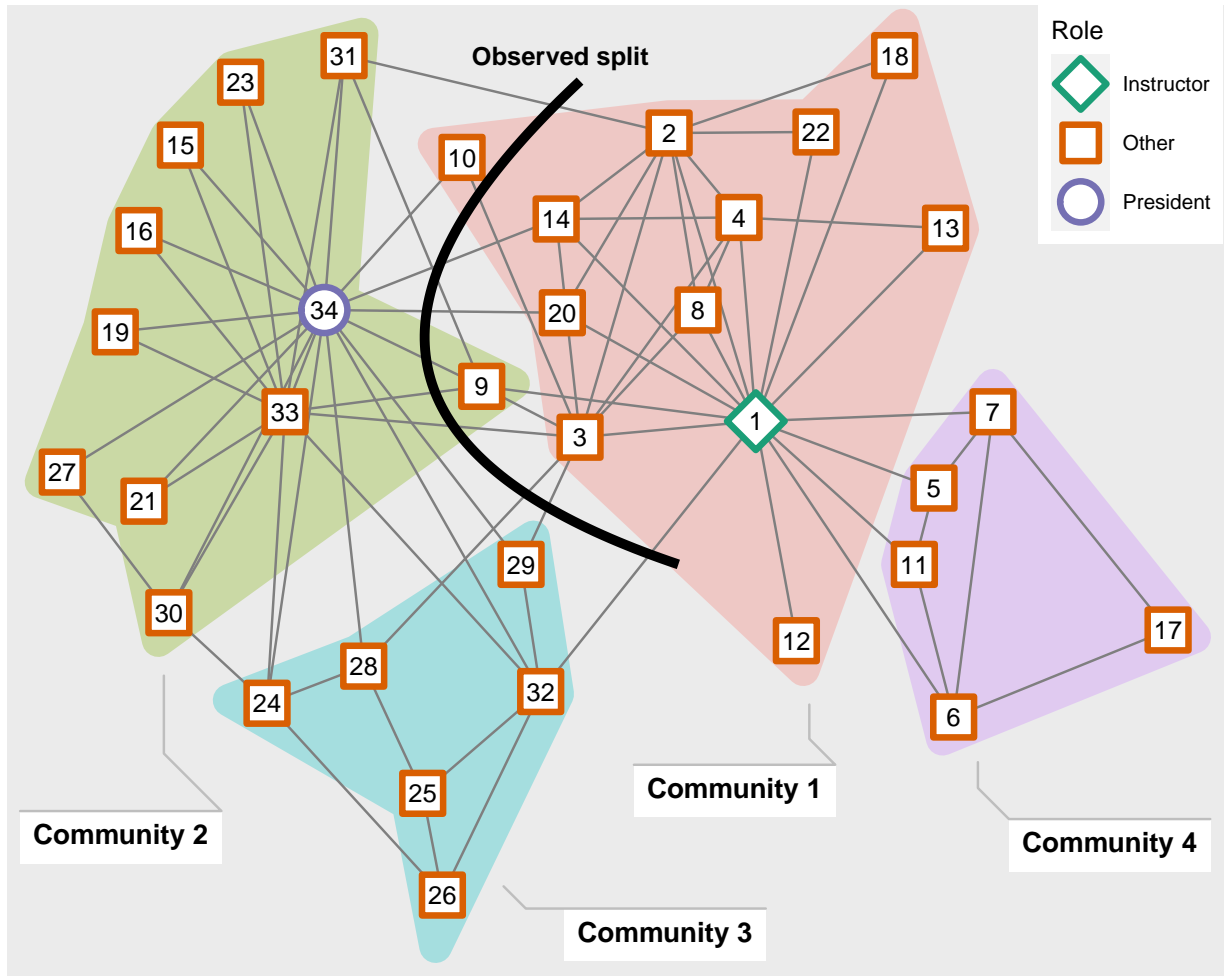
- (12) Briefly explain to a reader what the figure reveals about the karate-club network.

References

- Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 4292–4293.
- Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, **33**(4), 452–473.

Zachary's Karate Club

Communities assigned by Louvain algorithm



Source: <https://networkrepository.com>

Figure 1: By the end of this exercise, we want to produce a plot similar to the plot shown above.