

CSSS 567 Homework 4

Jess Kunke, due 11/19/2021

Specifying, fitting, and interpreting SBMs

For this homework, we return to the `lazega` data set from the `sand` package:

```
library(igraph)
library(sand)
library(blockmodels)
data(lazega)
```

We're working with a data set we've used before, so you don't need to describe it.

Using the `blockmodels` package, fit an SBM without covariates.

- Report how many groups you found, plot the graph with nodes colored by cluster (and use `set.seed` as in lab), and provide a table of the estimated edge probabilities.
- In just 2-3 sentences, summarize the table: do the groups have roughly equal intragroup edge probabilities, or if not, how would you summarize what you see? Which intergroup edge probabilities are the largest or smallest, or are they all comparable?
- Look at the node degree and node attributes by cluster. Do you notice differences between the clusters the model found? How would you characterize the clusters, or what underlying structure do you think the clusters might be picking up in terms of degree and/or specific node attributes?

Now fit the same model except include a single covariate: a matching effect on one of the covariates, as in lab. For example, you might choose the indicator function for whether two lawyers practice the same type of law, in which case your model is

$$\text{logit}(P(Y_{ij} | c_i, c_j, \text{Practice}_i, \text{Practice}_j)) = m_{c_i, c_j} + \beta I\{\text{Practice}_i = \text{Practice}_j\}.$$

- Report how many groups you found, plot the graph with nodes colored by cluster (and use `set.seed` as in lab), and provide a table of the estimated edge probabilities.
- What do the results of the two models have in common, and in what ways do they differ?

Instructions: please read

Let's use this assignment as practice for future work:

- First do your own exploration and analysis, which will likely include code that won't end up in your final Rmd file.
- As you work, clean up your Rmd file to include only the code that's relevant for what you want to present in your report. Don't include things like `?lazega` which is a call to a help file, and leave out code that doesn't lead to results you present in your report. Knit your file/run your code frequently so that your errors don't accumulate and become more complicated to debug. Your final code in the Rmd file should be neatly organized and commented, and I should be able to run it from start to finish to reproduce your results.
- Your knitted file should look polished and presentable; it doesn't have to be fancy or have references, but don't print package-loading messages or model-fitting messages, and don't show your code or raw output. Don't include the instructions or number your answers; instead, incorporate your responses into a cohesive and concise report. If you want to present model estimates or other results, do so in sentences in the body of your text or in organized tables. Only include results that you comment on in the body of your text.

Please submit both your knitted html or pdf document and your Rmd file, just as you would provide both to a colleague or client if you wanted to allow them to reproduce your report/results. If you prefer

to write your report in a text editor instead of in your Rmd file, that's fine; in that case, please submit the report as a pdf and also submit an R script or Rmd file that meets the requirements above and reproduces your model results and any other results you refer to in the report. As before, please title all files you submit with "HW1_" and your last name. For example, I would submit my homework as HW1_Kunke.pdf (and HW1_Kunke.R if I had a separate R file with my code).