

Social Network Analysis

Assignment 3 Modelling network structure 21 February 2024

The purpose of this assignment is to get experience with exponential random graph modelling of network structure, both in *thinking* and in *practice*. You are asked to reason about network mechanisms and assess the validity of your reasoning based on ERGM analyses and ERGM simulations.

The personal school class network data set referenced in the exercises below is distributed via LISAM together with this assignment on 21 February, please download it and start working on this assignment in the lab class of 23 February ← Best come prepared by already having run the lab example!

Exercises

Friendship networks in school are typically highly gender segregated, highly clustered, and highly reciprocated. What is more, the corresponding *network mechanisms* of *homophily*, *reciprocation* and *transitive closure* are confounded. Here are their definitions.

Mechanism	Verbalisation
<i>homophily</i>	<i>i</i> chooses <i>j</i> as friend when <i>i</i> and <i>j</i> are similar on some characteristic
<i>reciprocation</i>	<i>i</i> chooses <i>j</i> as friend when <i>j</i> has chosen <i>i</i> as friend
<i>transitive closure</i>	<i>i</i> chooses <i>j</i> as friend when <i>j</i> is a friend of a friend of <i>i</i> 's, which means: when there is a third person <i>k</i> who was chosen as friend by <i>i</i> and who has chosen <i>j</i> as friend

Please answer the following conceptual question (independent of any particular data):

- 1) Explain why the *homophily mechanism* might create networks in which there is a substantial number of *reciprocated ties* and *transitively closed triads*, even if no explicit *mechanisms* of reciprocation and transitive closure operate.

Now let us move to network analysis. In your personalised school class data set, there is a friendship network (given as R matrix variable `fri`) which you are first asked to statistically describe a bit.

- 2) Please program two R-functions for evaluating a given network:
 - one that calculates the number of reciprocated ties,
 - another one that calculates the number of transitively closed triads.

Then evaluate your personalised school class network data set on these two dimensions by applying these functions.

- 3) Bonus question: Explain for each function what its expected value is for random networks with the same number of nodes and the same probability of ties as your personalised school class network (*hint*: you need not generate random networks for this, it can be calculated analytically). Is the empirical value in your data set (calculated in Exercise 2) above this chance-expectation?

Furthermore, there is a two-mode network of students by the hobbies they have (given as R matrix variable `hob`).

- 4) Please calculate a student-by-student matrix that indicates how many hobbies two students have in common.

What is the average amount of hobbies that friends have in common in your data set? What is the average amount of hobbies that non-friends have in common? Do these numbers indicate hobby-homophily of friendship?

(Note: make sure that you handle the matrix diagonal correctly in these calculations! Nobody is their own friend, and nobody is a “non-friend” to themselves either!)

Finally, let us move to the modelling of your data set, in which your reasoning from Exercise 1 can be quantitatively assessed, combining empirical analysis with simulation.

- 5) Please fit three exponential random graph models to the data:

The “full model”: A model that simultaneously assesses overall tie creation tendencies, reciprocation tendencies, transitive closure tendencies (*hint: use `gwesp`*), and three homophily tendencies: for gender (*hint: use `nodematch`*), school attitude (*hint: use `absdiff`*) and shared hobbies (*hint: use `edgescov`* applied to the matrix obtained in Exercise 4). For statistical reasons (*hierarchy principle*) please also include the `twopath` effect; this will help avoid estimation problems.

The “reduced model”: A model where reciprocation, transitive closure and the `twopath` effect are dropped from the full model, whereas the three homophily terms and the general tie creation tendency are retained.

The “null model”: A model in which also the homophily terms are dropped, and only the overall tie creation tendency is retained.

- 6) Report the results in one(!) table and interpret them in a brief text. Make sure to include information criteria in your table and address them in a model comparison.

Bonus addition: If you have time, you can also assess *goodness of fit* and include summary information (e.g., p-values) in the table and discuss them next to the information criteria.

- 7) Please simulate 100 networks from each model. Evaluate the simulated data in terms of the functions you programmed for Exercise 2) above. Compare the models in terms of their simulations to each other graphically and to the observed data. Concretely:

- Make four network visualisations: one example graph from each set of simulations, and the empirical data set, with nodes coloured by gender.
- Make two plots, one for each function programmed in Exercise 2), each of which contains the three simulated distributions and the empirical value of the index as reference (i.e., the value you calculated in Exercise 2).

- 8) Try to substantiate your reasoning in Exercise 1) based on these results.
- 9) Please indicate whom you collaborated with when answering the above questions, and whose input you considered particularly helpful.

Note: this will not be counted against any of you. I realise that collaboration is a good thing when mastering complicated tasks. I ask for this network information only because it helps me giving feedback efficiently when correcting the assignments you will hand in.

Consultation hour

You should start working on the assignment in the lab class on February 23rd. Halfway before you submit your work, I (Christian Steglich) offer a max. 30 minutes online consultation. It will take place on **Wednesday 28 February 15:15-15:45**, under this Zoom link:

<https://liu-se.zoom.us/j/64017312422?pwd=TnRGUVh4a1FkWGgrWXNLNmQraU82UT09>

If you join without link, the Meeting ID is 640 1731 2422 and the Passcode is 189883.

Deadline

Please hand in your answers to the exercises by uploading two documents to our course's Lissam environment: One is the R- or Rmd-script you used, the other is the MS Word or pdf document containing the answers to the exercises above. There are separate submission portals for both. Note that by submitting online, there will routinely be a plagiarism check of the submitted documents. Deadline for handing in is **Wednesday 6 March 2023, 10:00** before the SAOM course lecture. Then, the online submission portals will close.

Literature

Feld, S.L. (1981). The Focused Organization of Social Ties. *American Journal of Sociology* 86, 1015-1035.

Goodreau, S.M., Kitts, J.A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* 46, 103-125.