

Assignment 3 SNA - School Networks: ERGM

Marc Sparhuber

Table of contents

Task 1	2
Task 2	2
Task 4	2
Task 5	3
Task 6	3
Task 7	5
Task 8	9
Task 9	10

Task 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.

When individuals chose one another as friends as in the case of homophily based on shared characteristics, then this can inadvertently lead to an increase in reciprocated ties. This is due to a shared characteristic in the nodes i and j leading to the nomination of j as a friend of i 's. Since both nodes, however, possess the same characteristic, it is entirely feasible that j also nominates i as a friend because of the shared characteristic. This then results in the same outcomes as reciprocity (i and j nominating each other as friends) but is simply the result of similarity of node characteristics rather than due to having been chosen as a friend first.

Regarding transitive closure, the tendency to have an increased number of transitively closed triads is due to the fact that nodes i and j have both been chosen as friends by a third node k . Homophily assumes that when k made the decision to nominate them as friends that this was due to shared characteristics. This extends to a potential tie between i and j as both of them can be assumed to have shared characteristics due to k having chosen both of them as friends. Therefore, although the result may seem to be the result of an explicit transitive closure mechanism, it may just be the outcome of homophily.

Task 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.

The number of reciprocated ties is 42.

The number of transitively closed triads is 9.

Task 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?

The average amount of hobbies that friends have in common is 3.44. The average amount of hobbies that non-friends have in common is 2.90.

These results indicate slightly higher hobby-homophily among friends than among non-friends. Importantly, the way this is measured only takes into account whether i nominates j as a friend and not whether it is reciprocal. Therefore, a friend, in this conceptualization is someone that an individual nominated as such.

Task 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 `edg cov` 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.

Task 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.

Table 1: School network: ERGM results

	Null Model	Reduc. Model	Full Model
Edges	0.22*** [0.19, 0.27]	0.06*** [0.03, 0.10]	0.02*** [0.01, 0.04]
Reciprocity			6.43*** [3.18, 12.98]
Transitive Closure			2.85*** [1.94, 4.18]
Gender		6.14*** [3.83, 9.84]	2.03*** [1.48, 2.79]
School Attitude		0.67* [0.49, 0.92]	0.83+ [0.68, 1.02]
Shared Hobbies		1.18** [1.05, 1.32]	1.07+ [0.99, 1.14]
Twopath			0.95+ [0.90, 1.01]
Num.Obs.	756	756	756
AIC	720.5	638.7	544.5
BIC	725.2	657.2	576.9

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The above Table 1 shows the results of the three ERGM models previously calculated. The decreasingly smaller edges coefficient indicates that the ties are very likely not formed at random. The reduced model takes into account the vertex attributes gender, school attitude and shared hobbies. By far the greatest of these effects is attributed to Gender, the coefficient of which has conditional log odds of 6.14 and is significant. While surprisingly, the school attitude is significantly negatively associated with tie formation, shared hobbies increase their likelihood slightly at a significance level of $p < 0.01$. Adding the twopath for model stability, Reciprocity and Transitive Closure results in the effects of some of the vertex attributes to become non-significant and lower, yet retain the direction of their effects, as can be seen with the effect of gender, going from conditional log odds of 6.14 to 2.03. Instead, reciprocity and transitive closure are highly significantly positive, taking large shares of the effects of the vertex attributes. Across models, the more parameters were added the lower the goodness of fit measures AIC and BIC, indicating a better fit for the full model.

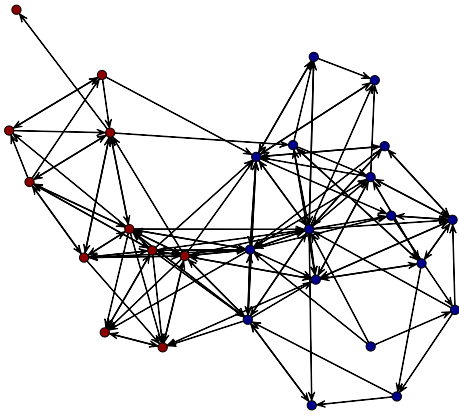
Task 7

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:

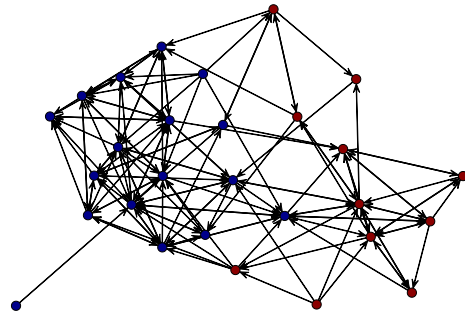
- Make four network visualisations: one example graph from each set of simulations, and the empirical data set, with nodes coloured by gender.
- Make two plots, one for each function programmed in Exercise 2), each of which contains the three simulated distributions and the empirical value of the index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

The following network visualizations aid in comparing the models. While the full model and the Empirical Network look very similar, as backed up by goodness of fit measures (not presented due to not being able to add them to the end of the modelsummary table). The reduced model still retains much of the visual similarity to the empirical network. This is despite the reduced number and thereby differentially strong effects compared to the full model. One way to explain this similarity is to recall the fact that homophily, which is included in the reduced model and the full model is expected to be heavily confounded with reciprocity and transitive triads which are only added in the full model. This, then hints toward why despite the dramatic shift in coefficient sizes the model looks quite similar: different process (homophily instead of reciprocity and transitive closure), similar results. Finally, the removal of both homophily as well as reciprocity and transitive closure in the null model takes away the effects of homophily completely. Whereas in the other models girls and boys are clearly more interconnected (be that due to reciprocity or homophily), the null model does not take tendencies of like this into account, only focusing on the number of ties between nodes.

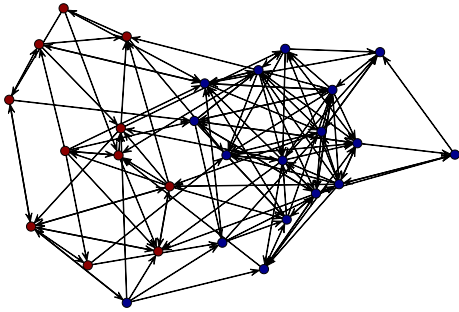
Empirical Network



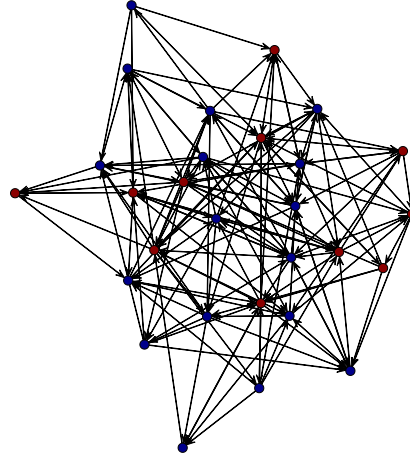
Full Model



Reduced Model



Null Model

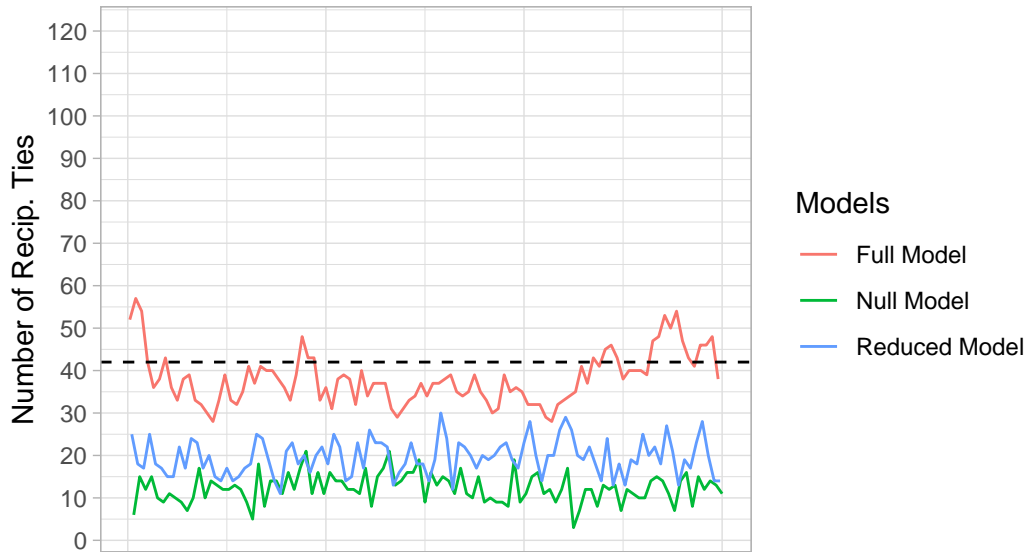


The following two graphs present the difference between the observed and simulated data in regards to the number of reciprocated ties and number of transitively closed triads. The dashed line indicating the respective empirical values is ideally closely traced by the red line signifying the full model. In both cases, the full model seems to be closer to the observed data than the reduced and null models, although the results for the number of transitively closed triads seems to be better simulated than reciprocity. Interestingly, the values in the second graph for the reduced and null model vary highly (and near equally badly compare to the observed data) while the full model captures the observed data without much variation. This is likely

due to the fact that geometrically weighted edgewise shared partners were added to the full model to account for transitivity. These graphs also highlight that “just homophily” as specified in the reduced model cannot account for the complexity of the observed network. In fact, looking at the differences in the second graph, there is barely a difference to not having specified homophilic parameters at all, while the first graph distinguishes itself from the null model but does not reach the precision of the full model.

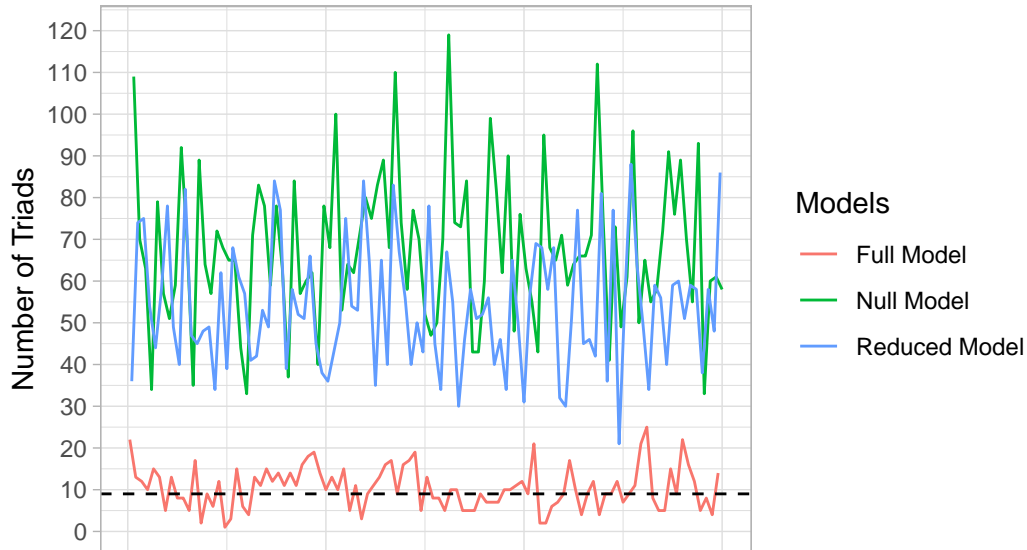
School Network: Reciprocated Ties

Dashed line indicates empirical result



School Network: Transitively Closed Triads

Dashed line indicates empirical result



Task 8

Try to substantiate your reasoning in Exercise 1) based on these results.

The points stated in Task 1 echo throughout the results table and later visualizations and plots. In the results section it is evident that going from the reduced to the full model the homophily coefficients get reduced in their effect strength while the reciprocity and transitive closure effects gain significantly. This shows how intertwined the two are. When visualizing the networks, similar results can be observed: while the reduced and full model are both relatively similar to the empirical network, the null model is vastly different, showing how the null model with neither homophily nor reciprocity and transitive closure is the only model that clearly visually differs from the others due to have neither specified. This goes hand in hand with my suspicion that if one were to specify a second reduced model without the homophily coefficients it would also look similar to the current reduced and full model due to the confounding of these mechanisms. Finally, the last two plots show “just homophily” fails at capturing the formation of transitively closed triads, vastly overestimating them while being slightly better than the null model at arriving at the correct number of reciprocated ties. This highlights the inability of homophily coefficients to capture transitive closure. This extends on the answer provided in task 1, as there was no discussion there of how strongly homophily would confound reciprocity compared to transitive closure, with the

last two plots indicating that there is “less homophily in transitive closure than in reciprocity”.

Task 9

Please indicate whom you collaborated with when answering the above questions, and whose input you considered particularly helpful.

I was inspired by Vincent’s solution to Task 4. Otherwise, ChatGPT helped a bit with adding colors to the network and polishing the final two plots.