Assignment 2 SNA - Island State Communities

Marc Sparhuber

Table of contents

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.

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.

```
[1] 42
```

The number of reciprocated ties is `r{n_recip(fri)}`.

```
[1] 9
```

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?

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 edgecov 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.

Table 1: blah blah TITLE

	Null Model	Reduc. Model	Full Model
edges	−1.50***	−2.89***	−3.83***
	[−1.68, −1.31]	[−3.48, −2.30]	[−4.44, −3.21]
nodematch.gender		1.81***	0.71***
		[1.34, 2.29]	[0.39, 1.02]
absdiff.school_attitude		−0.40*	−0.18+
		[−0.71, −0.09]	[−0.39, 0.02]
edgecov.hob_t		0.16**	0.06+
		[0.05, 0.28]	[−0.01, 0.13]
mutual			1.86***
			[1.16, 2.56]
gwesp.OTP.fixed.0.5			1.05***
			[0.66, 1.43]
twopath			−0.05+
			[−0.11, 0.01]
Num.Obs.	756	756	756
AIC	720.5	638.7	544.5
BIC	725.2	657.2	576.9

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

Source: GSS data from the socviz R package.

Comments: The reference cateogory for happy is 'Not Too Happy'.

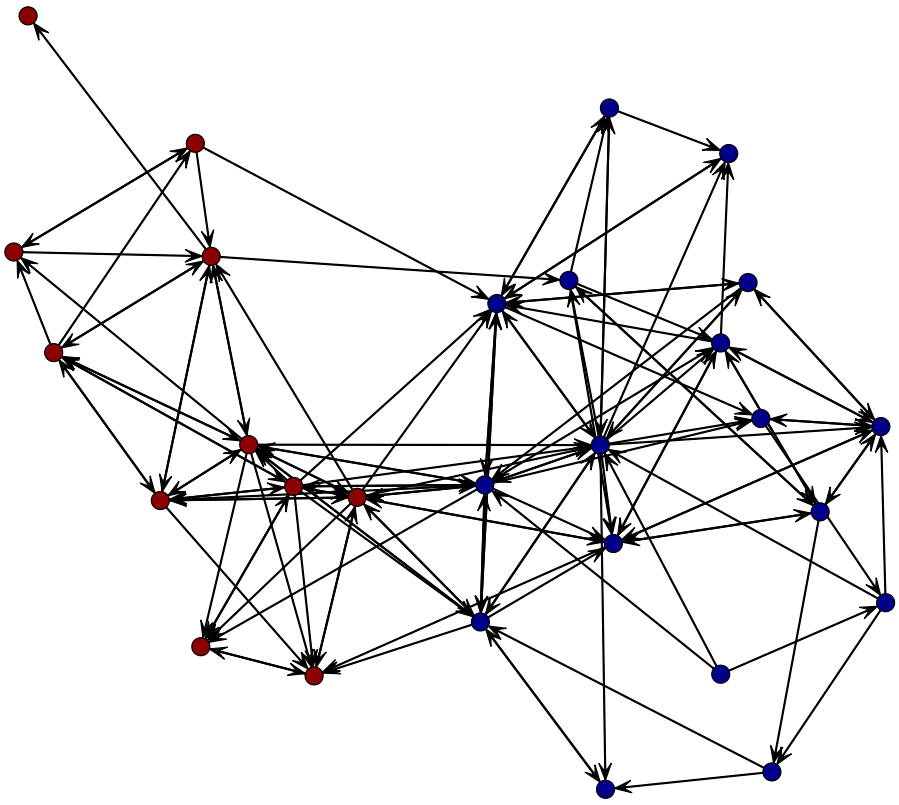
BRIEF INTERPRETATION HERE

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:

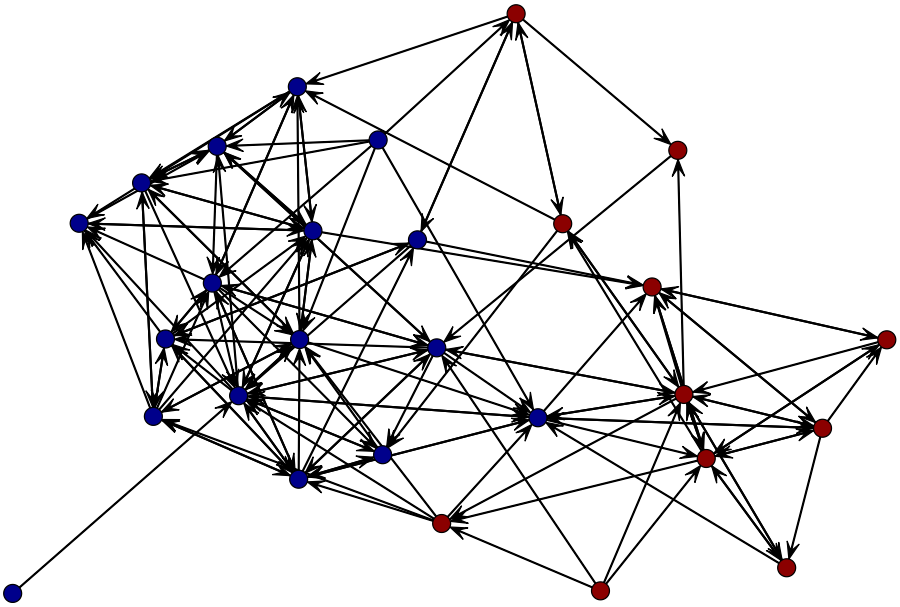
– Make four network visualisations: one example graph from each set of simulations, and the empirical data set, with nodes coloured by gender.

– 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).

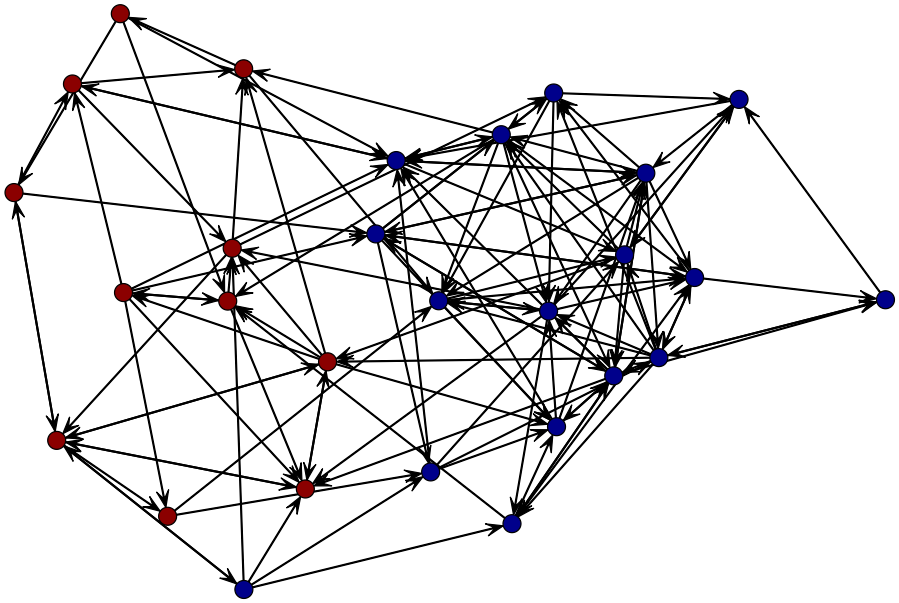
empirical



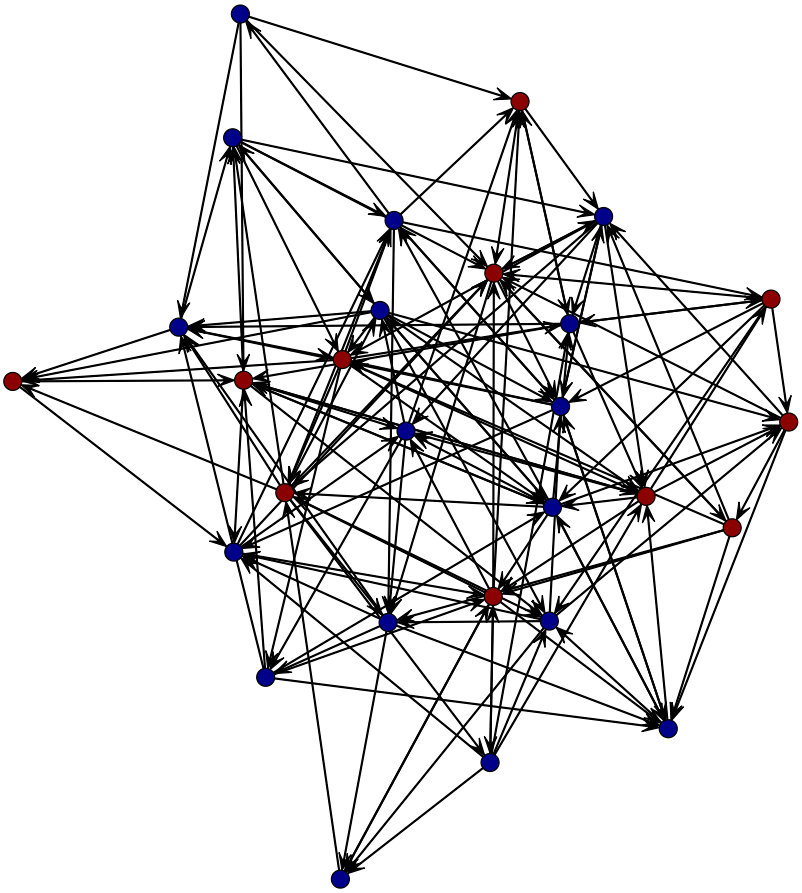
Sim Full

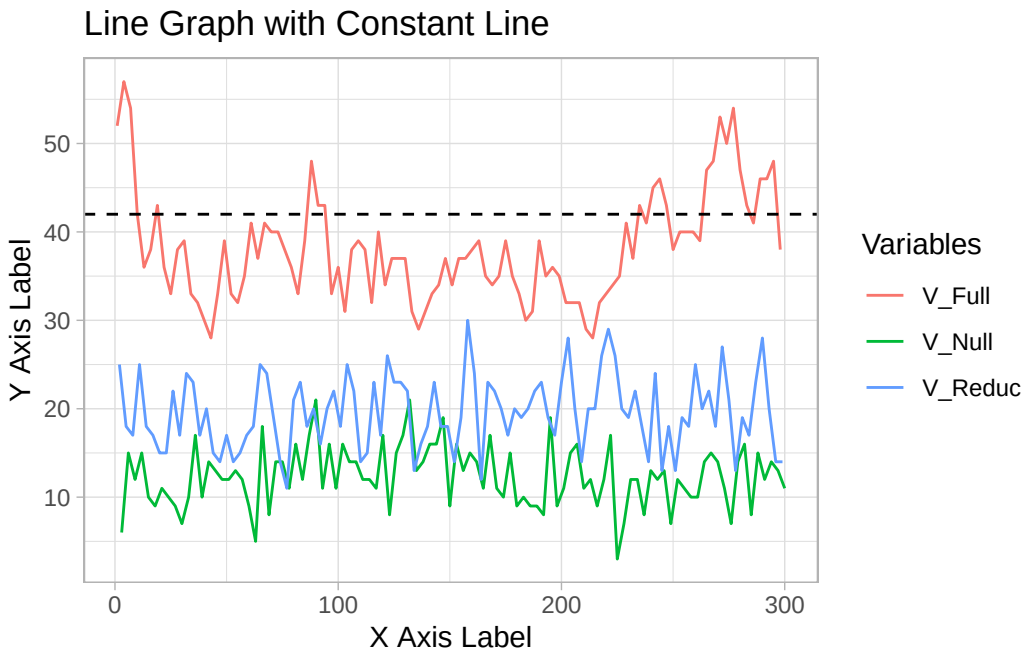


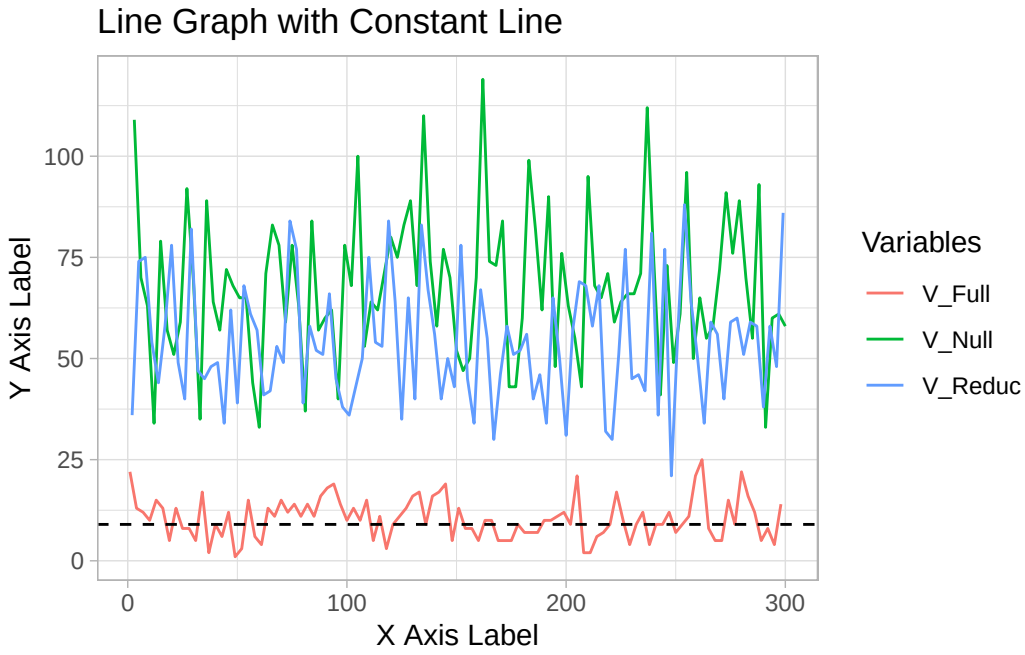
Sim Reduc



Sim Null







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.