Summer 2024 Exam

MY461

Candidate ID: 27529

2024-05-06

Contents

Exercise 1

Instructions exercise 1:

Consider the overall network metrics (density, mean degree, average path length, transitivity, and reciprocity) for the spendtime_w2, bestfriend_w2, and conflict_w2 networks. Pick a network model to serve as the most suitable randomized baseline for comparison and justify your choice of model. Compare each of the three networks to its randomized version. What do these comparisons tell you about the nature of socialisation, friendship, and conflict in schools? How do the empirical networks differ from one another, and from their randomized versions? In your answer, make sure to define each of the metrics and give an intuitive interpretation for them.

Next, plot the in- and out- degree distributions of the three empirical networks (consider the most informative plot type for the data, e.g., the degree frequencies or CCDF, linear or log, six distinct plots or one plot with the six distributions overlayed on one another). What do the in- and out-degree distributions suggest about (1) how students answered the survey questions, and (2) relationships of socialisation, friendship, and conflict in the school?

Answer exercise 1:

wordcount: 504

I selected the configuration model for its ability to preserve the original grade distribution while randomizing connections, allowing the impacts of interventions to be evaluated without altering the initial "popularity" or relevance of students.(Hobson, et.al, 2021) To ensure the reliability, the analysis averaged metrics from 100 iterations. Although Barabasi-Albert might be suitable for heavy tailed networks (Pósfai,and Barabási, 2016), the configuration model offers a direct method to examine changes in a static network and effectively replicated density and mean degree, preserving the structural features useful for analyzing the intervention's impact on social dynamics.

As shown in table 1:

Table 1: Summary of the network metrics for the empirical and random networks

	Density	Mean degree	Av.Path Lenght	Transitivity	Reciprocity
Empirical Spend Friend W2	0.017	13.782	4.129	0.267	0.439
Random Spend Time W2	0.017	13.782	3.245	0.043	0.020
Empirical Best Friend Time W2	0.004	3.089	6.563	0.187	0.439
Random Best Friend W2	0.004	3.089	8.612	0.007	0.005
Empirical Conflict W2	0.003	2.431	5.104	0.060	0.159
Random Conflict W2	0.003	2.431	6.743	0.020	0.007

• Density, which measures the proportion of possible connections that actually occur (students are not necessarily connected to all other), remains constant across empirical and random networks, indicating that the model replicates this feature well. This also occurs in the mean degree, which captures the average number of connections where students reported around 13 in "spendtime", 3 "bestfriends", and 2 "conflicts."

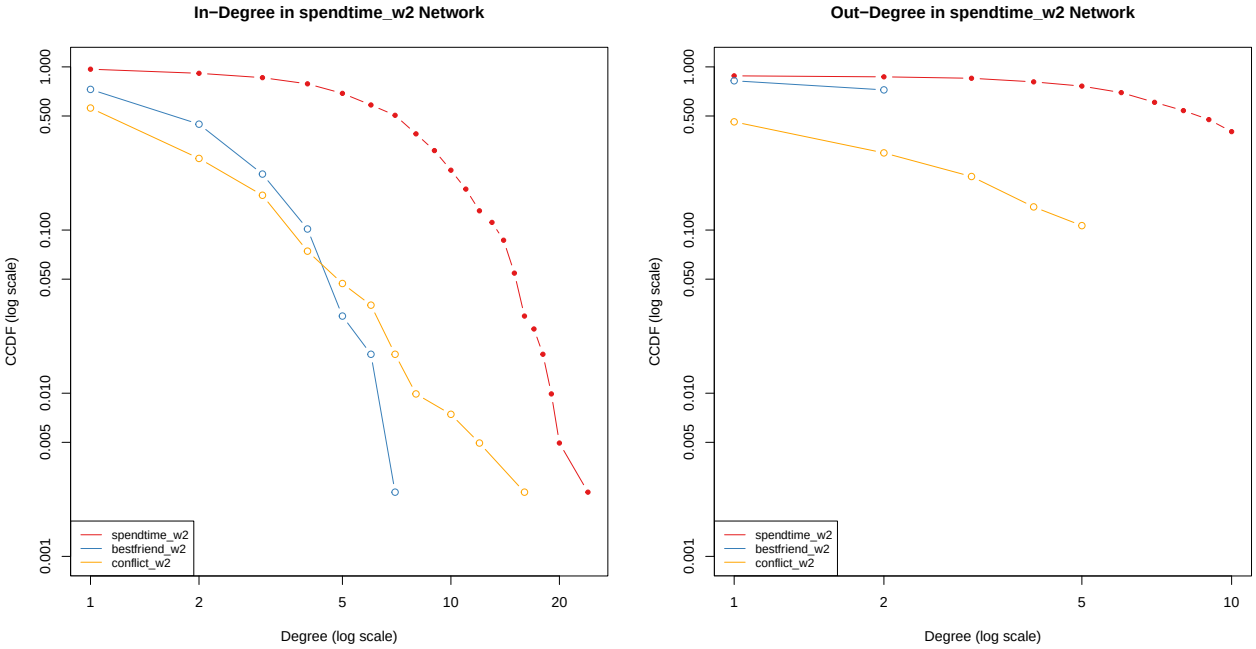
• The average lenght path, which measures the average distance between all pairs of nodes, decreases the network's time when randomized, suggesting that centrally connected nodes, by being redistributed, may facilitate more direct access between students. However, in conflict and friendship networks, the average path increases, this may indicate that both empirical networks have more strategic connections, e.g. clusters of friends or conflict that are dispersed in randomization.

• Transitivity, which measures the probability that two friends of a friend are also friends, is lower in random networks, indicating that the model does not, as expected, capture the tendency to form triangles as in

empirical networks. This suggests that even if a student has a close friend, a group of friends with whom they spend time, or a conflict with someone, this does not necessarily extend to other mutual friends, which in the case of conflict may be a good sign of no "contagion" of conflict.

• Reciprocity, measuring the bidirectionality of connections, decreases in the random model, reflecting the model's inability to capture the mutual interactions evident in time-spending and friendship networks. This is relatively high for spendtime and bestfriend networks, indicating that 44% of relationships or perception of friendship are mutual, in the case of conflict this is only 15% indicating that there is more one-way conflict.

I chose to use the CCDF on logarithmic scale (see plot 1) because allows frequent and uncommon connections to be visualized, revealing heavy tails in the distributions. When adjusting the basis of comparison, we observed that most students have few peers with whom they interact (low indegree) and few have many connections, which indicates a high concentration of connections in a few students and evidences the existence of "hubs" or very sociable and/or very conflictive students. In the output degree graph, the distribution also shows a decline, less pronounced, indicating that while many students name a few peers, a few name many, reflecting the variability in how students extend their social influence or manage their conflicts, similar to leadership roles or prominence in conflicts, as evidenced in the reference paper.



Plot 1: In-Degree and Out-Degree Distributions of the Empirical Networks

Exercise 2

Instructions exercise 2:

Alongside in-degree, identify two additional potential meanings of "influence" as proxied by different centrality measures, one for the two spendtime networks and one for the two conflict networks (you cannot select degree or out-degree). Present clear interpretations of what each centrality measure (in-degree for the spendtime networks, in-degree for the conflict networks, your chosen measure for the spendtime networks, your chosen measure for the conflict networks) might be capturing about the position of students in the school. For the two centrality measures that you have selected, give a justification as to why you think that particular measure is informative.

Plot the two waves of the spendtime and conflict networks (spendtime_w1, spendtime_w2; conflict_w1, conflict_w2), with nodes sized by your chosen centrality measure, and otherwise presented as you see fit (e.g., with nodes coloured as you think will be informative).Are students' centrality measures stable over time?

Look at the correlation between wave 1 and wave 2 centrality measures (i.e., in-degree in spendtime_w1

versus spendtime_w2, in-degree in conflict_w1 versus conflict_w2, and equivalently for your two chosen measures). Are students' in- degree centrality measures consistent across relationships, and how might this change over time?

Look at the correlation between the spendtime and conflict networks (i.e., in-degree in spendtime_w1 versus conflict_w1 and in-degree in the spendtime_w2 versus conflict_w2). You can do these comparisons with plots, statistical tests, etc., and should then present your conclusions in prose, with a discussion of what these associations imply about socialisation and conflict dynamics in the school.

Answer exercise 2:

wordcount: 506

I chose the betweenes centrality for the conflict network and the page rank for the spendtime network as influence measures because:

- The betweenes measure can shed light on who can act as mediators in the networks. Even though this link is "negative," a student with high betweenness centrality may have a critical role where different conflicting groups intersect. With this, I seek to identify who could influence the dynamics of the conflict, (positively and negatively). The indegree in the conflict network shows how many people are pointing to a student as a person with whom they have conflict.
- Page Rank, help me identify those whose influence derives not only from having many people who mention them as a person they spend time with (high in-degree) but also from being connected with others who are themselves well connected.

Table 2 presents the means of the measures of centrality. Comparing the centrality before and after the intervention, shows that:

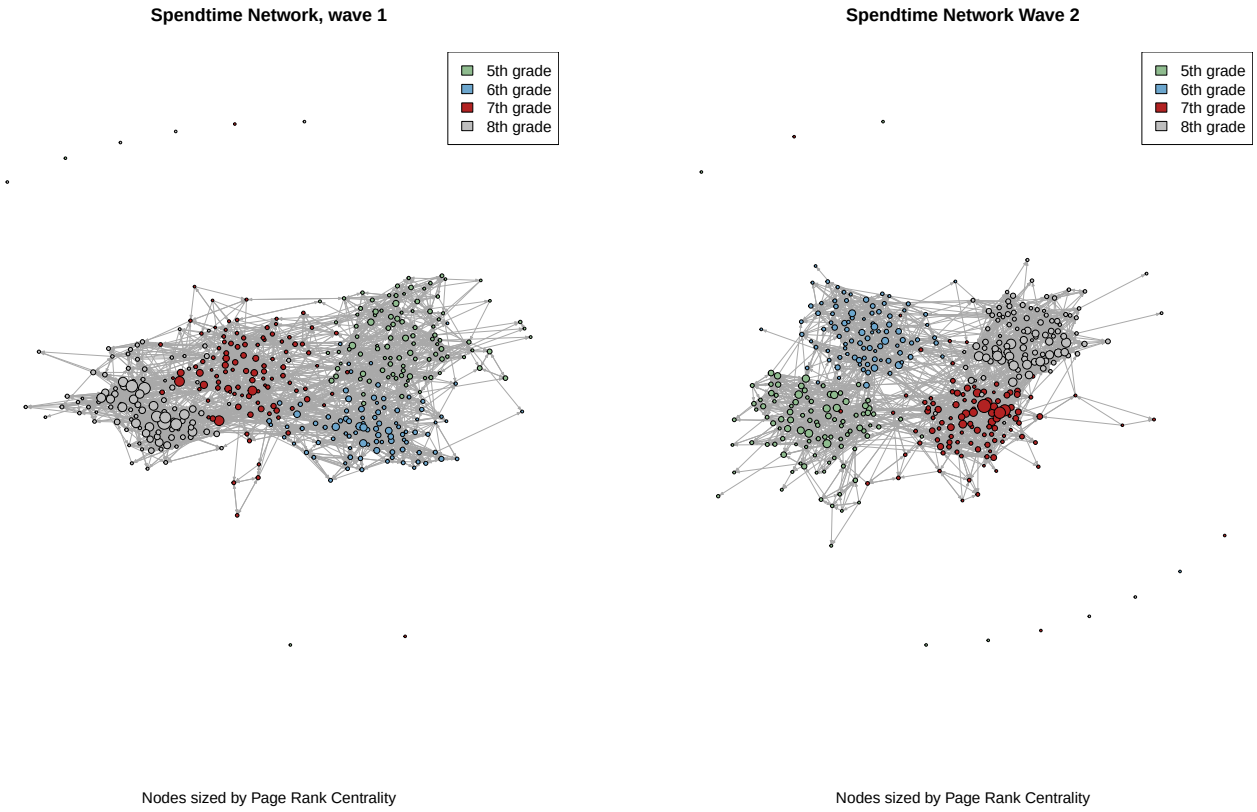
- The average indegree in the conflict network increased slightly, which, since the study indicates that conflicts decreased, may be indicating a greater willingness to report conflicts, rather than an increase.
- The centrality of intermediation increased greatly, suggesting that some students became crucial as mediators in conflicts (for better or worse).
- Pagerank remained stable, implying that the structure of "influent" students did not change.
- There was a slight reduction in the indegree of spendtime, reflecting possible subtle changes in social interactions after the intervention.

Table 2: Mean centrality measures for the networks

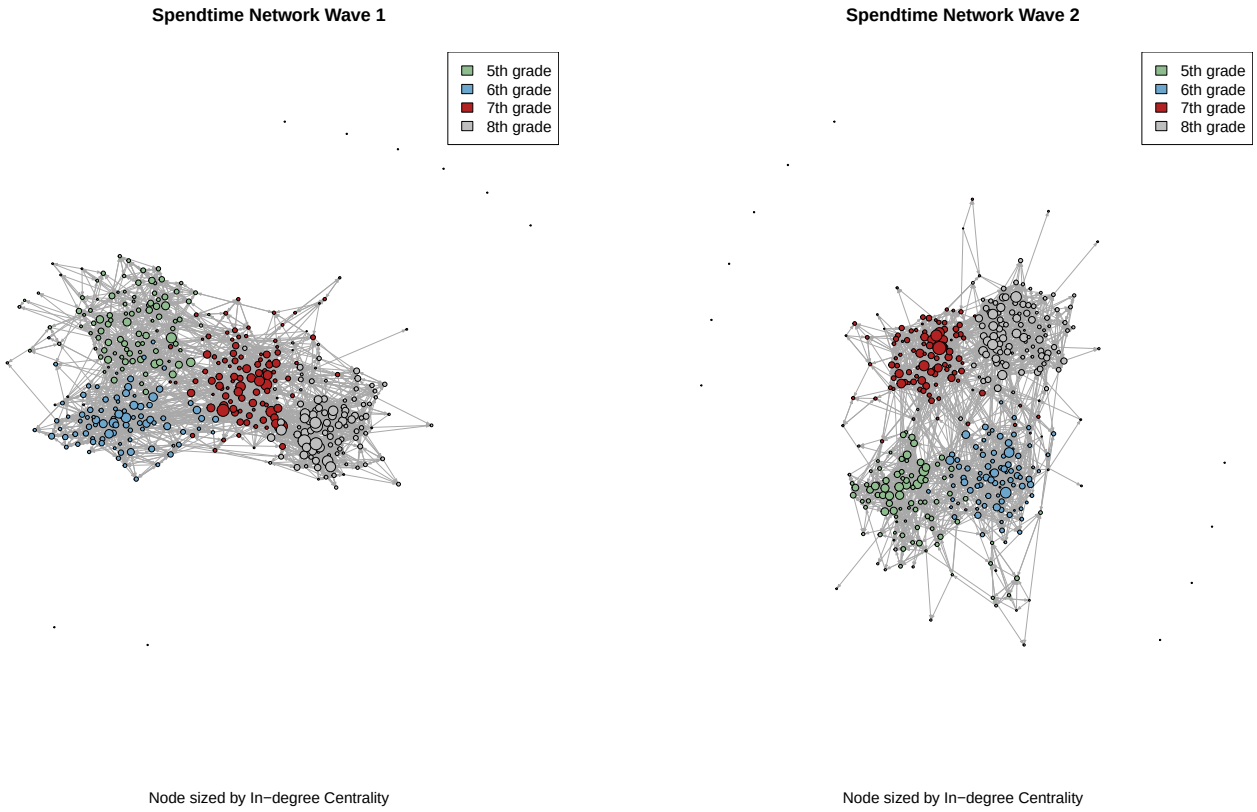
Centrality Measure	Mean Centrality
ind_spendtime_w1	7.1386139
ind_spendtime_w2	6.8910891
ind_conflict_w1	1.0569307
ind_conflict_w2	1.2153465
betw_conflict_w1	21.4232673
betw_conflict_w2	71.7252475
pr_spendtime_w1	0.0024752
pr_spendtime_w2	0.0024752

This intuition also reflects in the plots,(see plots 2 to 5) taking care to maintain the layout for positional consistency that allows a comparative analysis. The grade was chosen to show the natural clustering that occurs in schools by academic grade. We again see Page Rank in spendtime remains relatively stable. In the conflict network, with betweenes centrality increases markedly in the second wave, especially in 6th

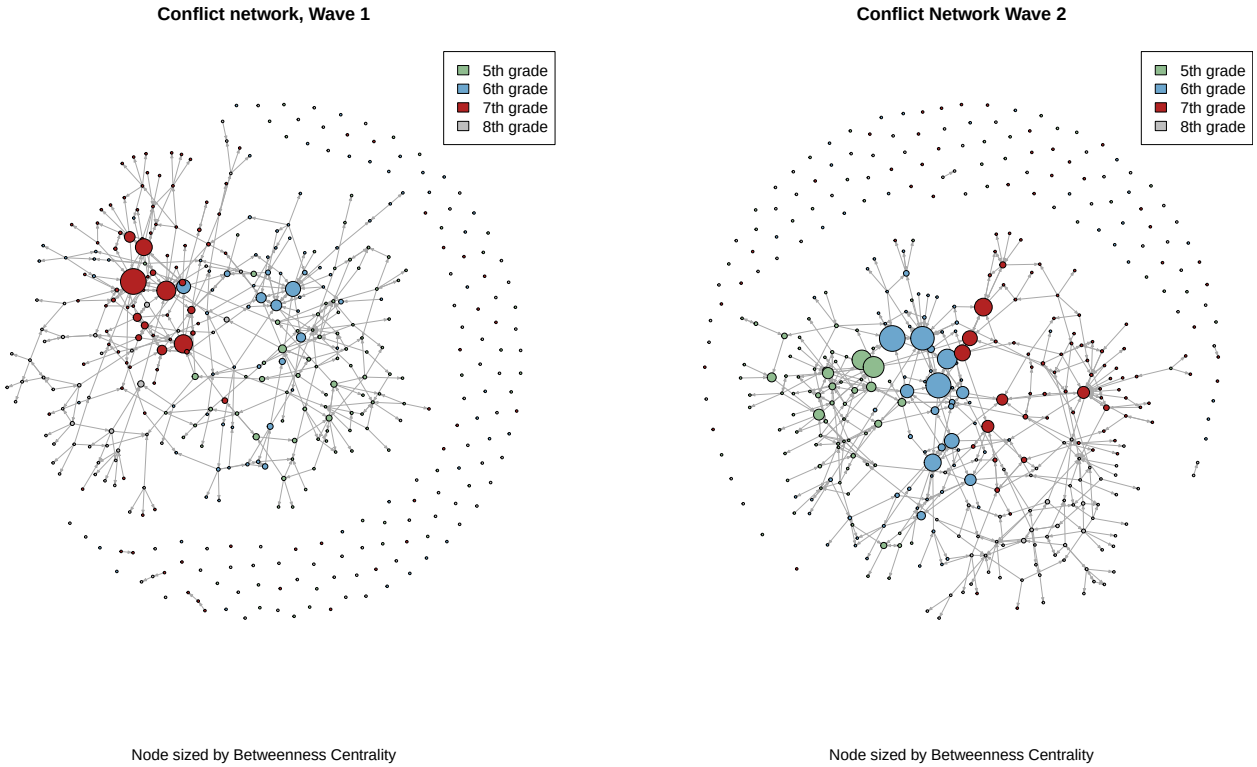
grade, perhaps indicating that more students may be being more significant intermediaries in conflicts over time, potentially as a result of the paper's intervention. Conflict by in-degree may be manifesting a more homogeneous distribution of conflicts with less prominent nodes in the second wave which is a really good sign, while spendtime by indegree remains very stable in both waves.



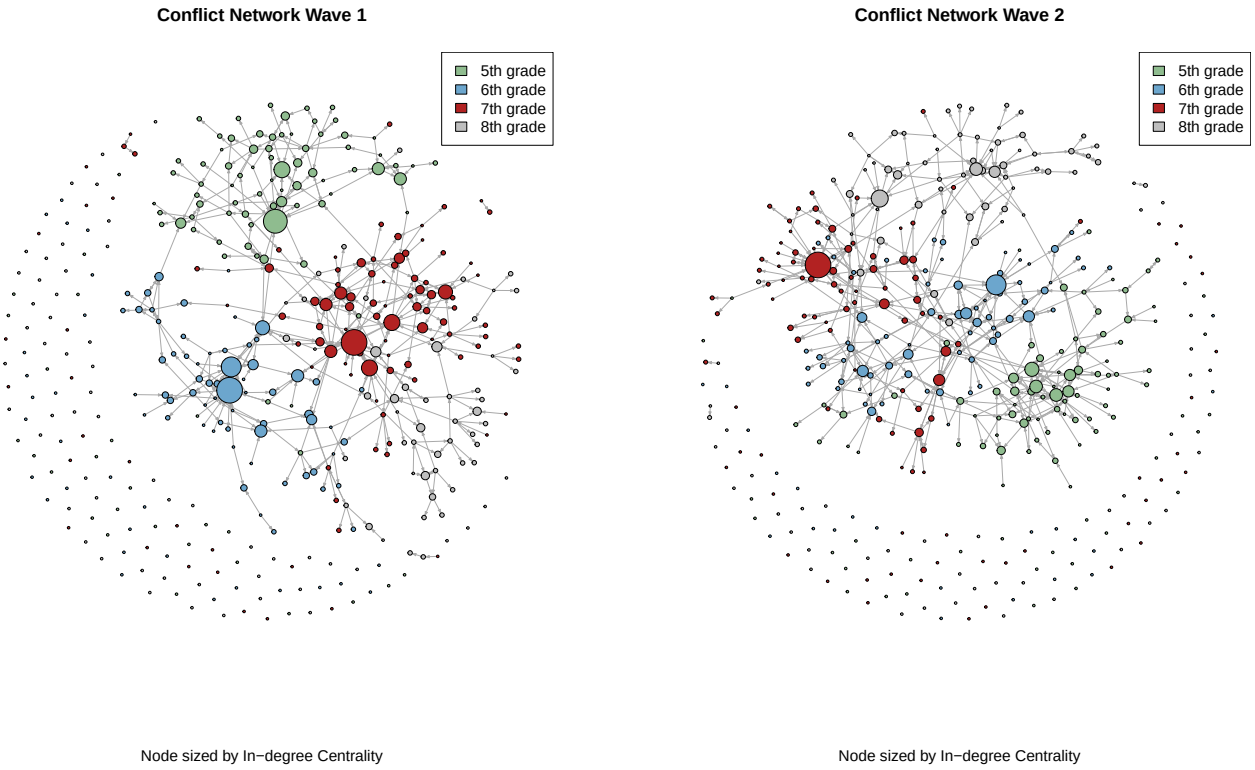
Plot 2: Spendtime network waves 1 and 2 with nodes sized by Page Rank centrality



Plot 3: Spendtime network waves 1 and 2 with nodes sized by in-degree centrality



Plot 4: Conflict network waves 1 and 2 with nodes sized by Betweenness centrality



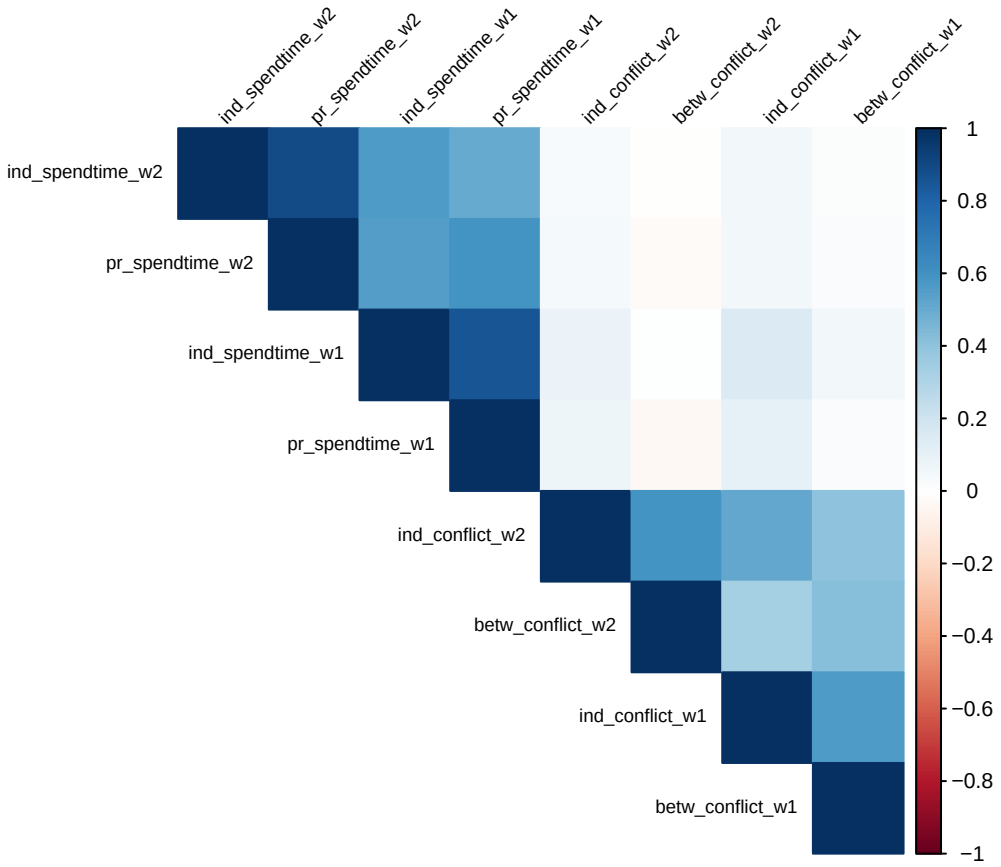
Plot 5: Conflict network waves 1 and 2 with nodes sized by In-degree centrality

In the correlation analysis (see plot 6, table 3) its observed consistent social interaction patterns over time, evidenced by moderate positive correlations across network centrality measures. Specifically, in the "spendtime" network, the in-degree centrality showed a correlation of 0.561 between waves, closely followed by PageRank centrality at 0.59. In the "conflict" network, the in-degree centrality correlation was slightly lower at 0.512, and the betweenness centrality, indicative of mediator roles, registered at 0.425. These correlations suggest a notable stability in social roles and influence across different periods within these networks. This may be indicating enduring social prominence, influence, and mediation roles among students.

However, correlations between in-degrees of "spendtime" and "conflict" are low in both waves (0.159 and 0.035), highlighting that students involved in social interactions are not the same as those involved in conflicts.

Table 3: Correlation matrix of centrality measures

	ind_spendtime_w1	ind_spendtime_w2	ind_conflict_w1	ind_conflict_w2	betw_conflict_w1	betw_conflict_w2	pr_spendtime_w1	pr_spendtime_w2
ind_spendtime_w1	1.0000000	0.5617583	0.1591507	0.0824664	0.0502148	0.0097084	0.8536205	0.5567842
ind_spendtime_w2	0.5617583	1.0000000	0.0583464	0.0356618	0.0146977	-0.0014141	0.5033773	0.8997005
ind_conflict_w1	0.1591507	0.0583464	1.0000000	0.5127793	0.5631226	0.3399505	0.1073462	0.0507263
ind_conflict_w2	0.0824664	0.0356618	0.5127793	1.0000000	0.4033249	0.5940544	0.0714215	0.0462146
betw_conflict_w1	0.0502148	0.0146977	0.5631226	0.4033249	1.0000000	0.4250521	0.0222031	0.0243399
betw_conflict_w2	0.0097084	-0.0014141	0.3399505	0.5940544	0.4250521	1.0000000	-0.0325595	-0.0271899
pr_spendtime_w1	0.8536205	0.5033773	0.1073462	0.0714215	0.0222031	-0.0325595	1.0000000	0.5929046
pr_spendtime_w2	0.5567842	0.8997005	0.0507263	0.0462146	0.0243399	-0.0271899	0.5929046	1.0000000



Plot 6: Correlation matrix of centrality measures

Exercise 3

Instructions exercise 3:

How do gender, grade, and age influence socialisation and conflict? Calculate assortativity by gender, grade, and age on both waves of the spendtime, bestfriend, and conflict networks (spendtime_w1, spendtime_w2; bestfriend_w1, bestfriend_w2; conflict_w1, conflict_w2). Compare the results across the two waves and across the three networks. What does this imply about socialisation and conflict, and how these might change over the course of the school year?

For simplicity, construct new versions of the spendtime_w2 and conflict_w2 networks that remove isolates, using delete.vertices(network, which(degree(network) == 0)). Use the leading_eigen() community detection algorithm to divide the no-isolate wave-2 spendtime and conflict networks into communities (note that you should get a warning saying "This method was developed for undirected graphs" – we are ignoring this). Plot the no-isolate wave-2 spendtime and conflict networks with nodes coloured by their community membership. Explore and discuss the extent to which the resulting communities align with gender, grade, and any other attributes you may wish to consider.

Answer exercise 3:

wordcount: 491

According to table 4:

- There is a trend of students preferring to form connections with peers of same grade, especially noticeable in the bestfriends network, where assortativity per grade is highest and increases in the second wave.

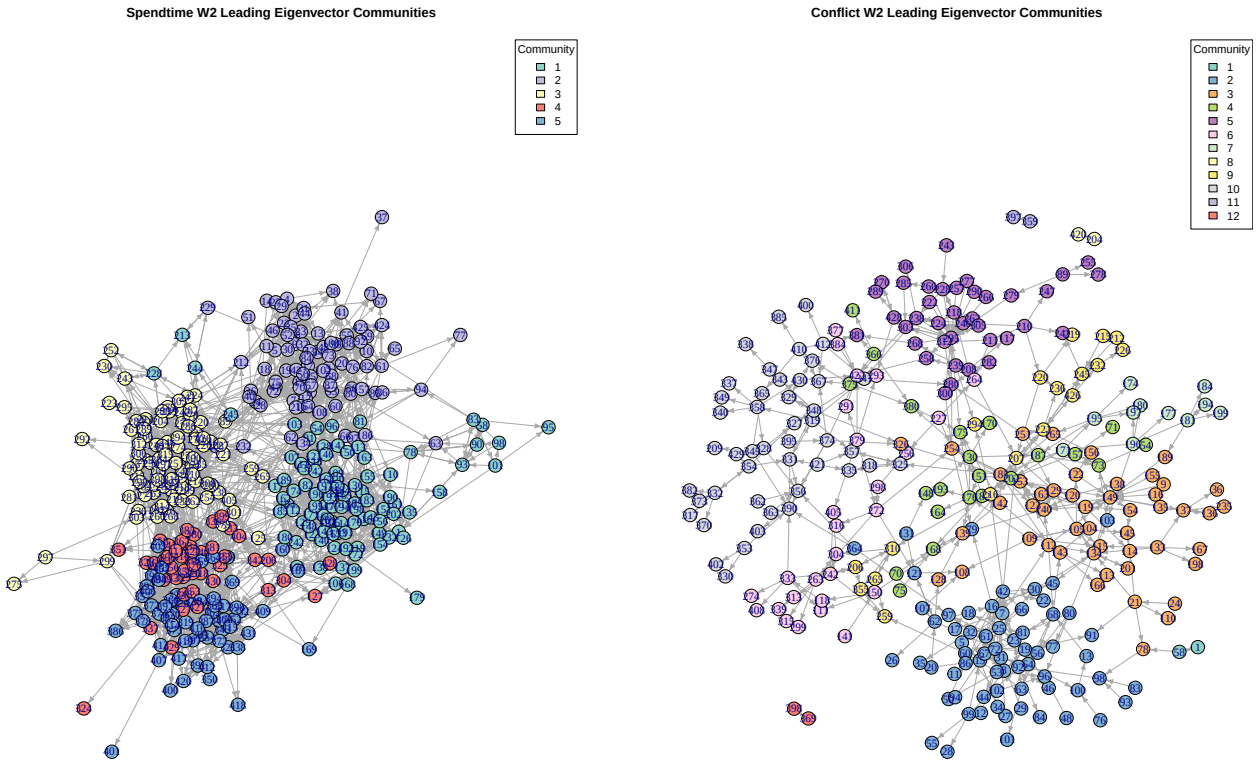
Suggesting that as the school year progresses, some students strengthen relationships within their own grade, indicating group cohesion.

- In contrast, gender shows less influence on the formation of networks, particularly low in the conflict network during the second wave, implying gender is less determinant in the formation of conflict groups and spending time groups compared to academic degree, although for bestfriend the asortivity by gender is very high.
- Age shows an intermediate assortativity between gender and the grade degree on spendtime and conflict networks, being its lowest in the spendtime and it decreases towards the end of the year. That's counterintuitive because age and grade usually go hand in hand. One possible explanation may be that students differ in social maturity or interests, and as the year progresses, students may become more comfortable expanding their circles. Also that the effect of the experiment facilitated more conections between students of different ages. In the conflict network, age assortativity increases in the second wave, which could reflect a maturation of conflicts as students get to know each other more and tensions increase or become more evident over time, which is very intuitive and occurs for the grade as well.
- These patterns suggest grade is a strong predictor of socialization and network formation, while gender and age have more nuanced and contextual roles, but there also appear to be more complex relationships due to the discrepancy between age and grade assortivity.

Table 4: Assortativity Summary by Network

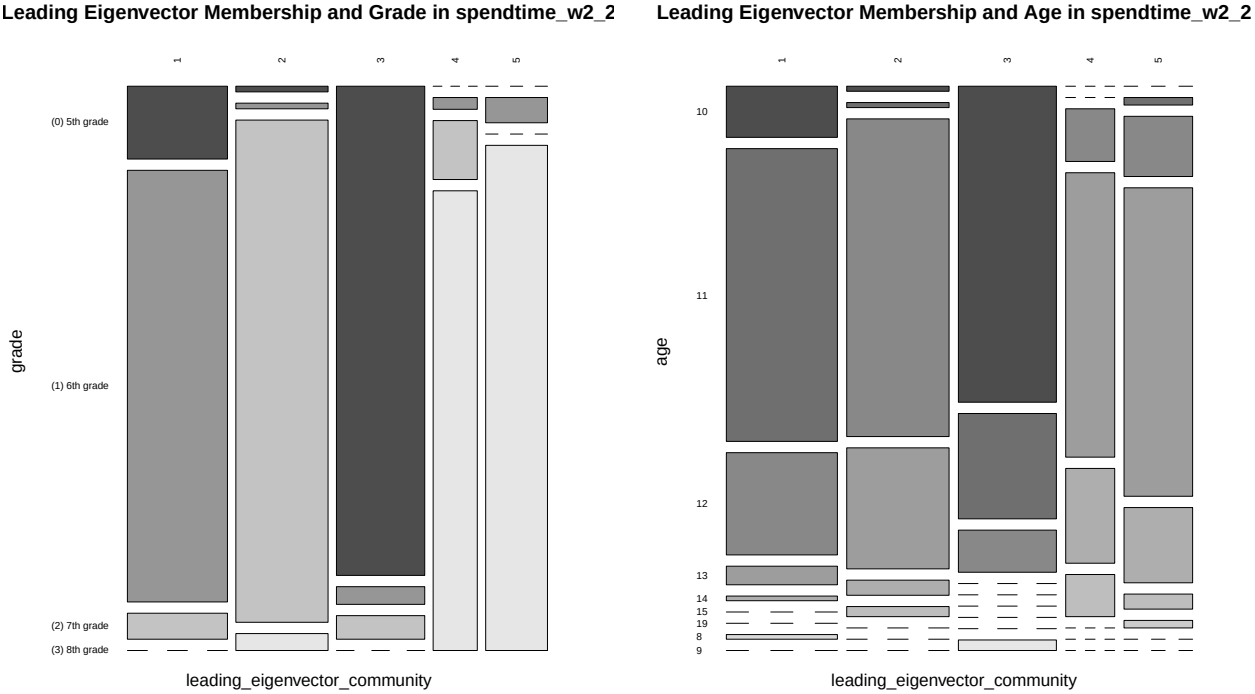
Network	Gender	Age	Grade
Spendtime W1	0.6143365	0.6909627	0.7955270
Spendtime W2	0.5624006	0.6875355	0.8337447
Conflict W1	0.5307948	0.7063587	0.7573465
Conflict W2	0.4741985	0.7279939	0.8254659
Bestfriend W1	0.7960722	0.7323783	0.8440458
Bestfriend W2	0.7626645	0.7477626	0.8887121

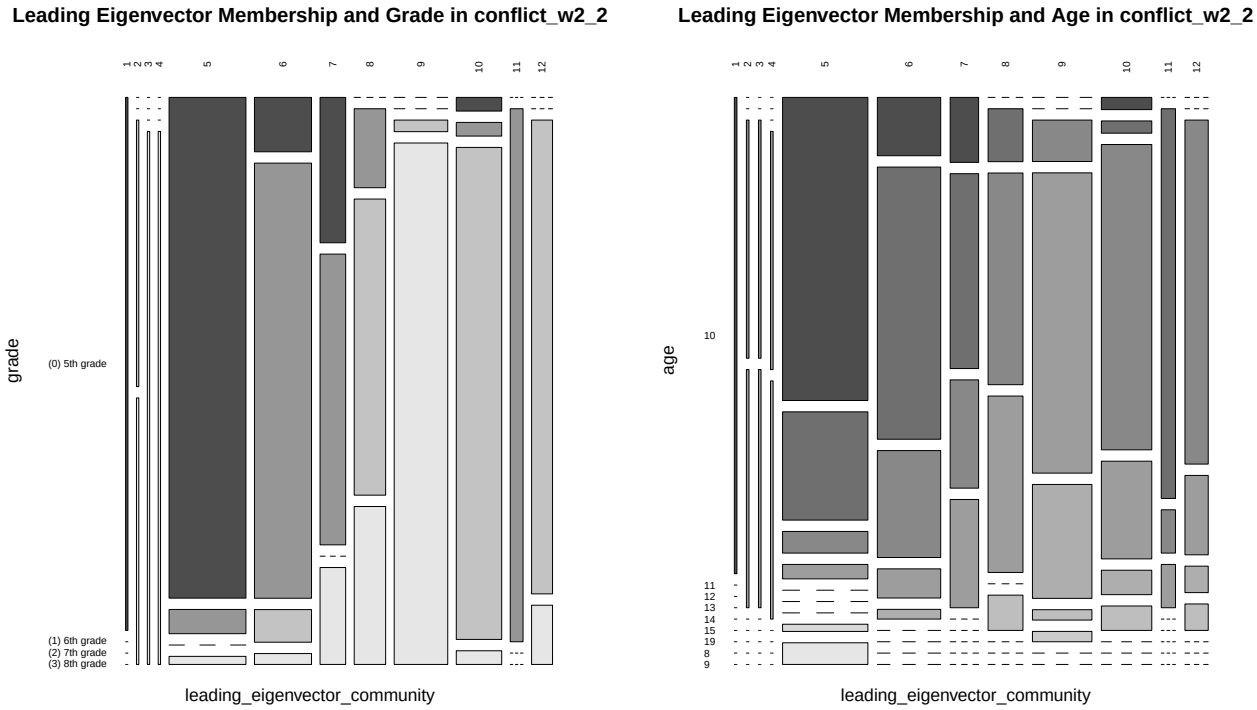
Plot 7 shows the detecting communities using Leading Eigenvector algorithm, resulting in 5 communities in the "spendtime" network and 12 in "conflict", showing the distribution of nodes per community.



Plot 7: Leading Eigenvector Communities in spendtime_w2_2 and conflict_w2_2

A more detailed analysis (plot 8) shows a clear correlation between the communities detected in "spendtime" and the academic grade. E.g. in community 1, 6th grade students predominate with 83 members, in community 2, 7th grade with 89 students. Community 3 – remarkably large – is almost exclusively composed of 5th graders with 83 members, and 4th and 5th graders are dominated by 8th graders.





Plot 8: Mosaic plots of Leading Eigenvector Membership of each network with Grade and Age

The Normalized Mutual Information (NMI) metric (table 5) of 0.686 for grade level in the "spendtime" network confirms a high correlation between community membership and grade, indicating not a "random" formation while gender (NMI of 0.018) and age (NMI of 0.399) show much lower correlations. In the conflict network, communities also appear to be influenced by grade level, although with a somewhat lower correlation than in the spendtime. The NMI between community membership and school grades is 0.553, indicating a moderate association. E.g large portion of 5th graders are clustered in community 5, while 7th graders are scattered but predominate in community 10. Differences in the number of communities (12 in "conflict" versus 5 in "spendtime") could reflect a greater complexity in negative interactions

Table 5: NMI for Grade, Age, Gender per network

Network	Attribute	NMI
Spendtime	Grade	0.6865115
Spendtime	Gender	0.0177694
Spendtime	Age	0.3992710
Conflict	Grade	0.5526220
Conflict	Gender	0.0853479
Conflict	Age	0.3245757

Exercise 4

Instructions exercise 4

What helps predict whether students report spending time with one another? Consider the results of the exponential random graph model (ERGM) below, fit to the spendtime_w2 network

Interpret each term in the ERGM (except edges) – how does each term influence whether students report spending time with one another? Use odds ratios in your substantive interpretation of each term. Calculate and report the fitted probabilities of i naming j as someone with whom they spend time, where:

```
Call:
ergm(formula = spendtime_w2 ~ edges + nodematch("Gender") + nodefactor("Gender") +
    nodematch("Grade") + nodefactor("Grade") + edgecov(bestfriend_w2) +
    edgecov(conflict_w2) + mutual)

Monte Carlo Maximum Likelihood Results:

                             Estimate Std. Error MCMC % z value Pr(>|z|)
edges                        -6.56842    0.07561      0 -86.874  < 1e-04 ***
nodematch.Gender              0.88070    0.04600      0  19.144  < 1e-04 ***
nodefactor.Gender.(1) Boy    -0.03562    0.02270      0  -1.569  0.11661
nodematch.Grade               2.47710    0.06051      0  40.935  < 1e-04 ***
nodefactor.Grade.(1) 6th grade  0.03551    0.03074      0   1.155  0.24805
nodefactor.Grade.(2) 7th grade  0.02145    0.02767      0   0.775  0.43831
nodefactor.Grade.(3) 8th grade  0.07379    0.02801      0   2.634  0.00843 **
edgecov.bestfriend_w2         5.48828    0.17830      0  30.781  < 1e-04 ***
edgecov.conflict_w2           0.94391    0.14187      0   6.653  < 1e-04 ***
mutual                        2.92381    0.07394      0  39.542  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 225705  on 162812  degrees of freedom
 Residual Deviance:  17806  on 162802  degrees of freedom

AIC: 17826  BIC: 17926  (Smaller is better. MC Std. Err. = 1.497)
```

Figure 1: Reference Image From Instructions

1. i is a girl and j is a boy, both are in 7th grade, neither named the other as a 'best friend' or as someone with whom they have conflict, and j did not name i as someone with whom they spend time.
2. i is a girl and j is a boy, both are in 7th grade, neither named the other as a 'best friend', i named j as someone with whom they have conflict, and j did not name i as someone with whom they spend time.
3. i is a girl and j is a boy, both are in 7th grade, i named j as a 'best friend', neither named the other as someone with whom they have conflict, and j did not name as someone with whom they spend time.
4. i is a girl and j is a boy, both are in 7th grade, i named j as a 'best friend', neither named the other as someone with whom they have conflict, and j named i as someone with whom they spend time.

Propose two additional model terms to add to the model and provide a justification for them. You can look to the many other attributes in the questionnaire and codebook to see what could be used. What do you think your new terms would capture that is currently missing from the ERGM? [You do not need to run this new ERGM].

Answer exercise 4:

wordcount: 509 words

From the call of the estimated model:

- Nodematch("Gender"): The positive coefficient (0.88070) indicates that students of same gender are significantly more likely to spend time together. In this case same-gender pairs have about 2.41 times higher odds of spending time together compared to mixed-gender pairs.
- Nodefactor("Gender.1") Boy: The small negative coefficient (-0.03562) suggests a slightly lower likelihood for boys, compared to the baseline (girls), to form connections. However, this effect is not statistically significant according to the pvalue.

- Nodematch("Grade") similar as in gender indicates that students of the same grade are more likely to spend time together. If they are in the same grade, they are 12 times more likely to spend time together.
- Nodefactor for "Grade 1, 2 and 3" have in all three cases a positive probability greater than 1 of spending time together (1.036, 1.022 and 1.077 respectively and with respect to the base category which is the fifth degree. This makes sense in practice because new students or students at the lowest level do not yet consolidate relationships as much as in other grades, although the difference for all levels is very small.
- Edgecov("bestfriend_w2") and edgecov("conflict_w2"): The coefficients (5.48828 for best friends and 0.94391 for conflict) indicate existing relationships significantly influence spending time together. Students who are best friends are far more likely to spend time together, with an odds ratio of 241 times more likely, indicating a very strong effect. For conflict, the odds ratio is 2.57, suggesting that even students with conflicts have more than twice the odds of spending time together compared to those without reported conflicts.
- Finally the mutual term has positive coefficient (2.92381) indicating mutual tendency in the relationships, meaning if one student reports spending time with another, the other is also likely to report it. The odds ratio is exp(2.92381) = 18.61, showing very high likelihood of reciprocity of declaring spending time together

The fitted probabilities for the four scenarios are as follows (estimated the ergm model so slightly different from the image):

```
summary(ex4_ergm)
```

```
## Call:
## ergm(formula = spendtime_w2 ~ edges + nodematch("Gender") + nodefactor("Gender") +
##     nodematch("Grade") + nodefactor("Grade") + edgecov(bestfriend_w2) +
##     edgecov(conflict_w2) + mutual)
##
## Monte Carlo Maximum Likelihood Results:
##
##                               Estimate Std. Error MCMC % z value Pr(>|z|)
## edges                         -6.56999    0.07610      0 -86.329  < 1e-04 ***
## nodematch.Gender               0.87938    0.04582      0  19.193  < 1e-04 ***
## nodefactor.Gender.(1) Boy     -0.03366    0.02235      0  -1.506  0.13205
## nodematch.Grade                2.47692    0.06067      0  40.823  < 1e-04 ***
## nodefactor.Grade.(1) 6th grade 0.03624    0.03080      0   1.177  0.23931
## nodefactor.Grade.(2) 7th grade 0.02137    0.02868      0   0.745  0.45615
## nodefactor.Grade.(3) 8th grade 0.07421    0.02838      0   2.615  0.00892 **
## edgecov.bestfriend_w2          5.48580    0.17780      0  30.853  < 1e-04 ***
## edgecov.conflict_w2            0.95037    0.14149      0   6.717  < 1e-04 ***
## mutual                         2.92136    0.07275      0  40.158  < 1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 225705  on 162812  degrees of freedom
##  Residual Deviance:  17806  on 162802  degrees of freedom
##
## AIC: 17826  BIC: 17926  (Smaller is better. MC Std. Err. = 1.823)
```

Table 6 shows the fitted probabilities for the fourth escenarios. In situations where there are no bonds of friendship or conflict, the probability is minimal (1.622% and 4.090% for the first two scenarios). However, this probability shoots up to 79.906% when one student names the other as a bestfriend, and reaches a maximum of 98.664% if the feeling is reciprocated (also evidenced by edgecovariates and mutual coefficients).

Table 6: Fitted probabilities of i naming j as someone with whom they spend time for each scenario

Scenario	Probability.of.occurence
First	1.622
Second	4.090
Third	79.906
Fourth	98.664

I think integrating the term GWESP would shed light on how the decision to spend time with a student is influenced by my friends decision to spend time with someone. This identification of transitivity could shed light on the creation of relationships that at first glance might not be so obvious their formation.

Finally, due to the widely researched influence of ethnicity on friendship formation (see for example Currarini, Jackson and Pin, 2010; Kogachi and Graham 2021), I consider that ethnic homophily can be capture using the nodematch of ethnicity (currently divided into multiple variables, so it require the creation of a categorical variable).

References

- Currarini, S., Jackson, M. O., & Pin, P. (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. Proceedings of the National Academy of Sciences, 107(11), 4857-4861. https://doi.org/10.1037/dev0001263
- Hobson, E. A., Silk, M. J., Fefferman, N. H., Larremore, D. B., Rombach, P., Shai, S., & Pinter-Wollman, N. (2021). A guide to choosing and implementing reference models for social network analysis. Biological Reviews, 96(6), 2716-2734. https://doi.org/10.1111/brv.12775
- Kogachi, K., & Graham, S. (2021). Same-race friendship preference across the middle school years: The role of school racial/ethnic context. Developmental psychology, 57(12), 2134. https://doi.org/10.1073/pnas.0911793107
- Paluck, E. L., Shepherd, H., & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. Proceedings of the National Academy of Sciences, 113(3), 566–571. https://doi.org/10.1073/pnas.1514483113
- Pósfai, M., & Barabási, A. L. (2016). Network science. Cambridge, UK:: Cambridge University Press. http://networksciencebook.com/