Social Network Analysis

Final Project

Submitted By. Mishu Dhar

Data:

Cleaning, Preparation

To conduct this analysis I selected data from classroom 1100. The reason behind this selection is I found this classroom is very much consistent with nodes. Compared to the other classrooms, it has very almost similar nodes apart from the node(1126) which exists in the first wave but not in the second and third wave. Node (1161) only exists in the third wave and node (1131) exists in the first and second wave of the network but not in the third wave. So, I dropped these nodes from all three waves to keep simplicity which gives us 30 identical nodes from three waves.

A Few Assumptions

I obtained six matrices from three waves: Wave 1, 2, and 3. Among these matrices, the 'I can ask for help' matrix serves as the dependent network, while the 'good friend' matrix is used as the time-varying (predictor) network for analysis. Unfortunately, I did not find enough information in the network description codebook or on the associated website. To address the proportion of students by gender, I considered the situation in Bangladesh's high schools, a developing country, and distributed the students as 60% female and 40% male, keeping in mind that the data is collected from a less developed region of Hungary. For GPA, I followed the grade point system used in Bangladeshi high schools, as defined by the Bangladesh Board of Education, where the scale ranges from 1 to 6, with 1 being the lowest passing grade and 6 representing an A+. Students who solve disputes in the first wave may not be as influential or active in the subsequent waves, or they may be more effective. This ability to solve disputes between the students changes over time. I considered this skill as leadership quality and performance over time. For example, student 1113 solved 56 disputes in the first wave but not a single one in the second wave. I added all the values of each row and stored them in a new variable as the total of solved disputes by all students. People's ability and motivation to perform in these extra curricular matters in an environment change with time.

Hypotheses

Based on the instructions provided, here are three hypotheses for network dynamics in the chosen networks:

Homophily: Nodes (students) who share similar characteristics or attributes are more likely to form ties for seeking help. This hypothesis is based on the principle of homophily, which suggests that individuals tend to associate with others who are similar to themselves. Therefore, students with similar academic performance levels, interests, or social backgrounds are more likely to seek help from each other.

Reciprocity: There is a tendency for mutual exchange of help-seeking ties between students. Reciprocity in help-seeking behavior implies that if student A seeks help from student B, then student B is likely to seek help from student A in return.

Triad Closure: Triads of students who are indirectly connected through a mutual helper (triadic closure) are more likely to form ties for seeking help. This hypothesis is based on the concept of transitivity in social networks, where if student A seeks help from student B and student B seeks help from student C, then there is an increased likelihood that student A will also seek help from student C. Triadic closure reflects the tendency for interconnectedness and clustering within social networks.

Statistics Wave 2- Classroom 1100

Maximum Eigen Vector Centrality: The maximum eigenvector centrality value is 1, associated with node 1101. This indicates that node 1101 has the highest influence or importance in the network according to the eigenvector centrality measure.

Isolates: There is one isolate in the network, which is node 1114. Isolates are nodes that have no connections to other nodes in the network.

Mean Degree: The mean degree of the network is approximately 7.03. This means that, on average, each node in the network is connected to approximately 7 other nodes.

Key Nodes: The key nodes are identified based on a criterion (e.g., nodes with degrees greater than the mean degree). In this case, nodes 1101, 1112, 1116, 1117, 1118, 1120, 1122, 1124, 1125, 1127, 1129, 1131, and 1132 are considered key nodes. These nodes have a degree greater than the mean degree, indicating they are relatively more connected or influential in the network compared to other nodes.

Descriptive Statistics Wave 3- Classroom 1100

Maximum Eigen Vector Centrality: The maximum eigenvector centrality value is 1, associated with node 1101. This indicates that node 1101 has the highest influence or importance in the network according to the eigenvector centrality measure.

Isolates: There is one isolate in the network, which is node 1130. Isolates are nodes that have no connections to other nodes in the network.

Mean Degree: The mean degree of the network is approximately 7.41. This means that, on average, each node in the network is connected to approximately 7 other nodes.

Key Nodes: The key nodes are identified based on a criterion (e.g., nodes with degrees greater than the mean degree). In this case, nodes 1101 1105 1106 1117 1118 1120 1123 1125 1127 1128 1129 1132 1133 are considered key nodes. These nodes have a degree greater than the mean degree, indicating they are relatively more connected or influential in the network compared to other nodes.

Compared to the statistics of wave 2 we can see that key nodes have been changed over time.

Hamming Distance:

It indicates 126 positions where the values differ between the "I can ask for help" network matrices for wave 1 and wave 3. The Hamming distance of 126 signifies the extent of change in the network structure or the advice-seeking ties between wave 1 and wave 3. It depends on the size of the networks and this value should be considered as high or low. Considering the relatively large Hamming distance compared to the size of the network (30 nodes), it suggests substantial changes in the connections or relationships between the nodes from wave 1 to wave 3.

Jaccard Index:

With a value of approximately 0.265, it suggests that about 26.5% of the edges present in one network (wave 1) are also present in another network (wave 3).

Visualization

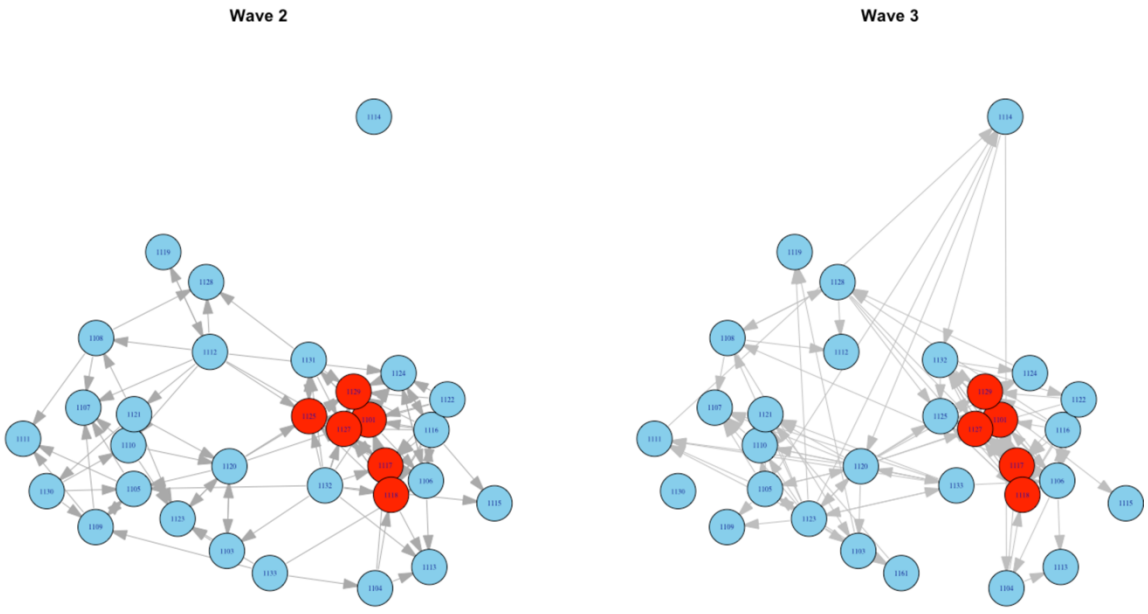


Figure 1 Influential Nodes- In Degree

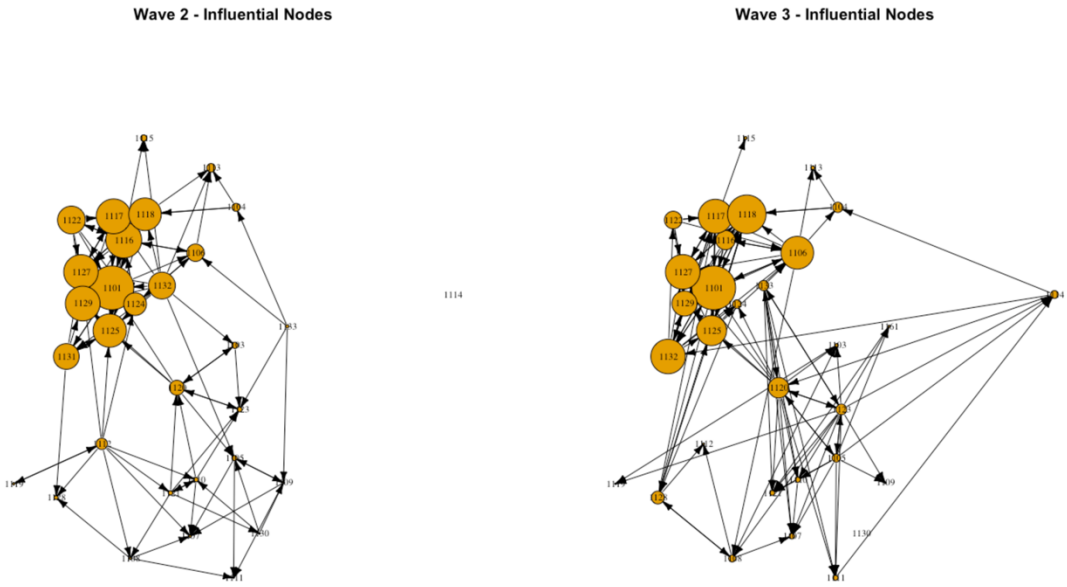


Figure 2 Influential Nodes in Wave 2 & 3. Based on Eigen Centrality

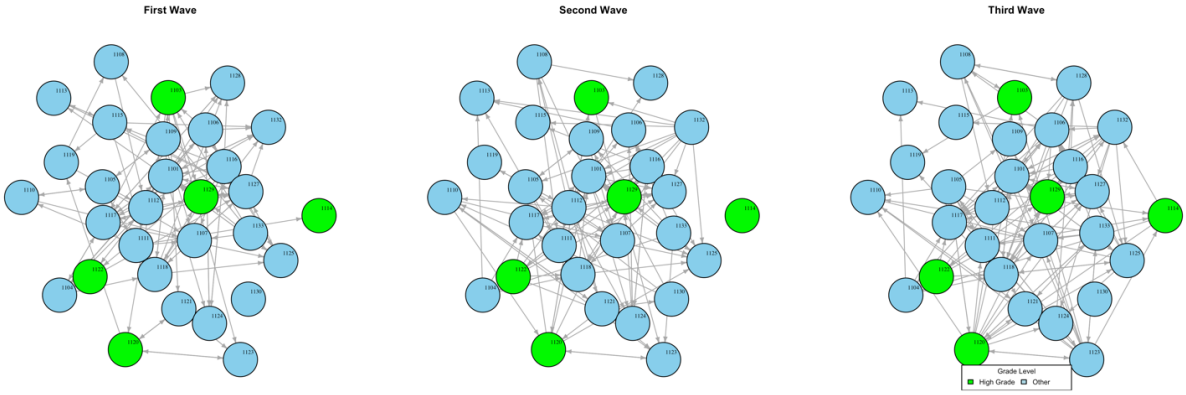


Figure 3 Network in three waves based on the grade points. The figures show, how network dynamics change over time. For example, student 1114 has a good grade but no student asks for advice from him in the second wave, in the first wave only one student. But in the Third wave, many students ask for advice from student 1114.

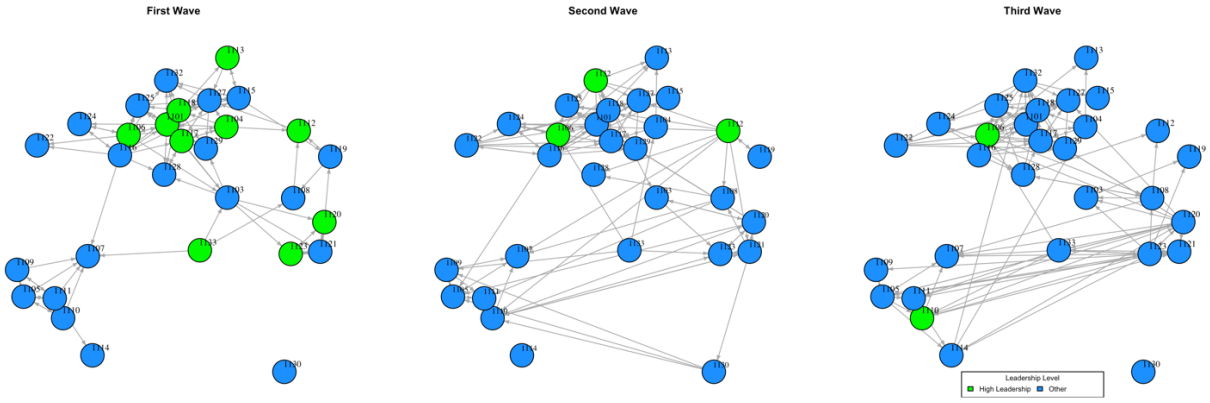


Figure 4 Network dynamics based on how many disputes students solved which I considered as leadership quality. Green nodes are those who solved at least 10 disputes. Student 1112 solved 10 disputes in the first and second waves, but not any in the third wave.

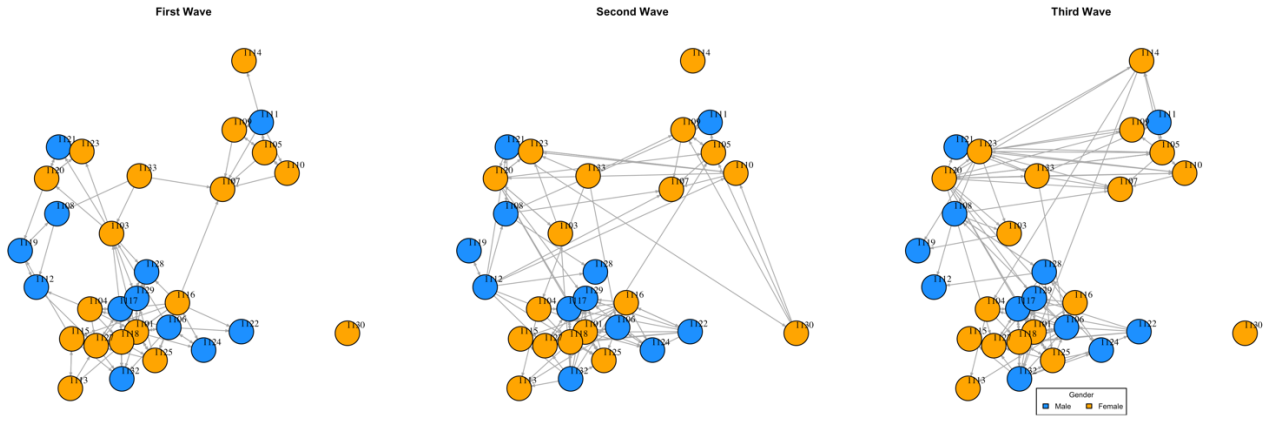
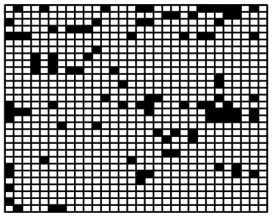


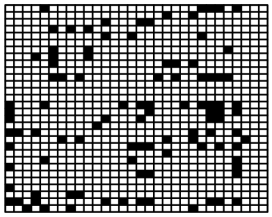
Figure 5 Network Based on Gender. Blue refers to male and Orange refers to female. In the first and third wave, 1130 is an isolate. But in the third wave, the other four students ask for advice from her.

First Wave



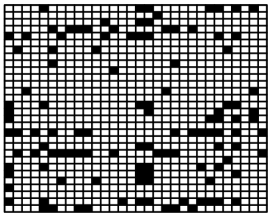
Ask for Help - First Wave

Second Wave



Ask for Help - Second Wave

Third Wave



Ask for Help - Third Wave

Figure 6 Visualization of sociomatrix plots depicting changes in network density across multiple waves.

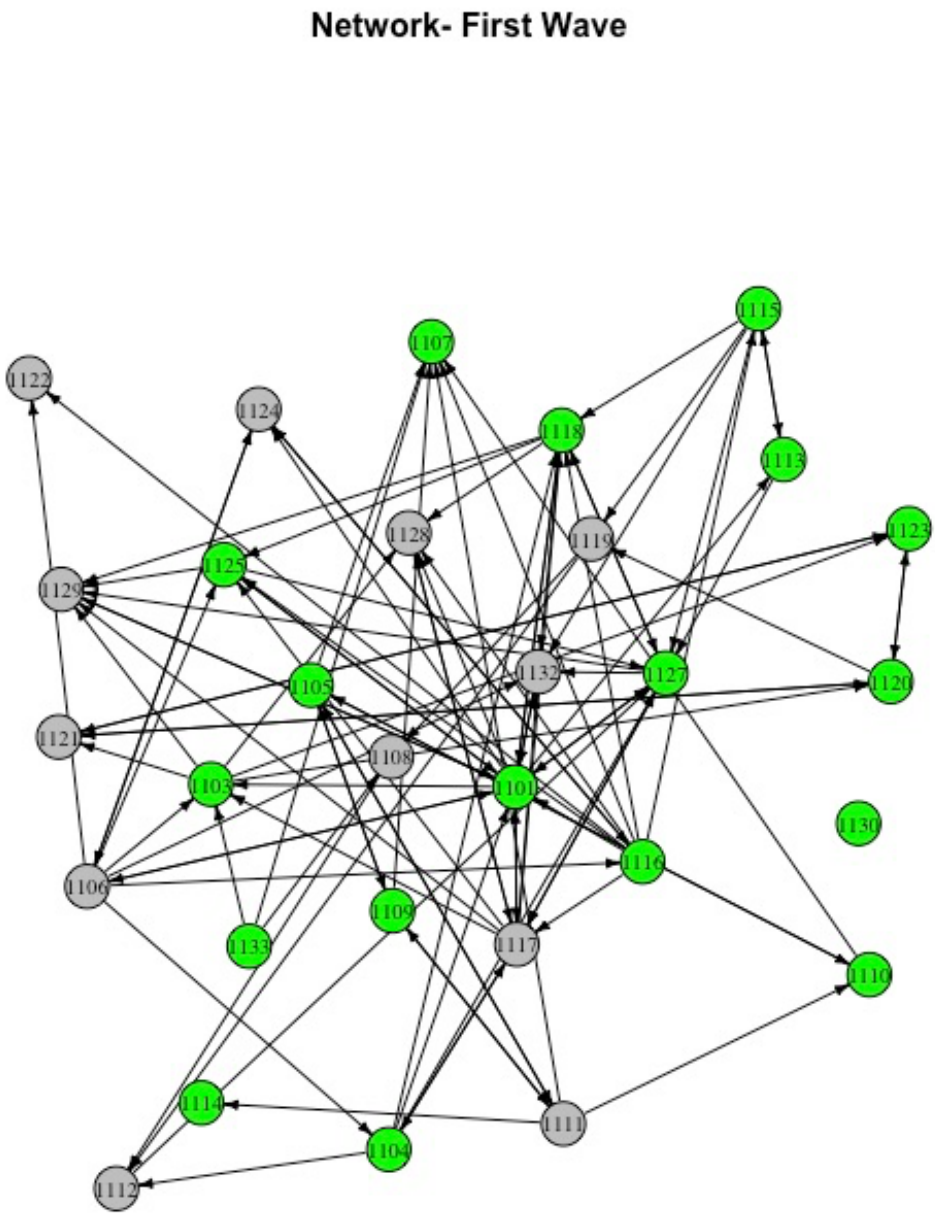


Figure 7 Help Seeking Network First Wave for the sake of Better Visual, Orange vertices are female students.

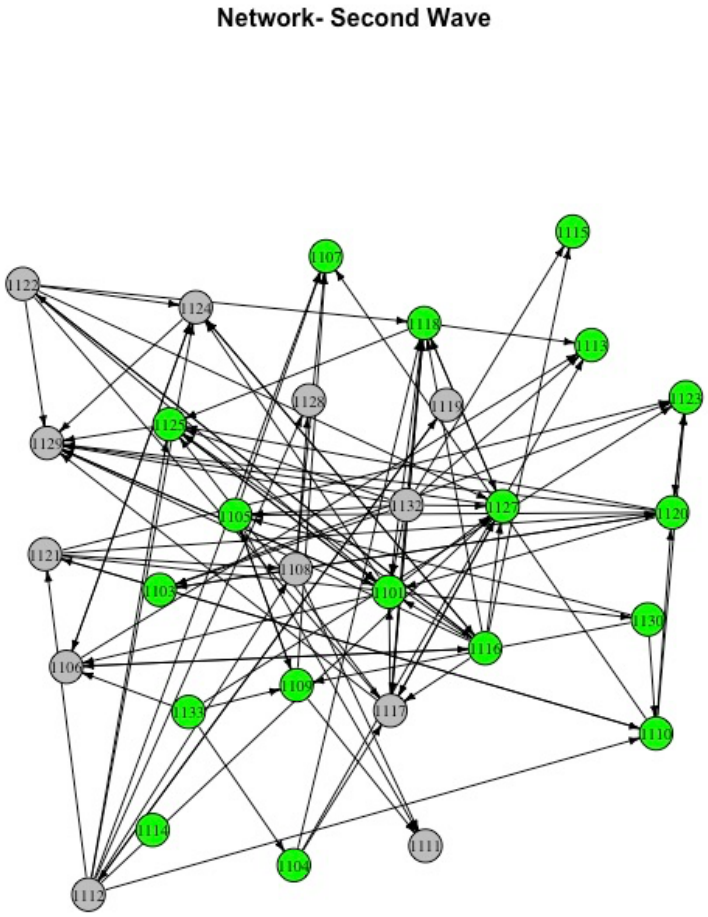


Figure 8 Network Second Wave

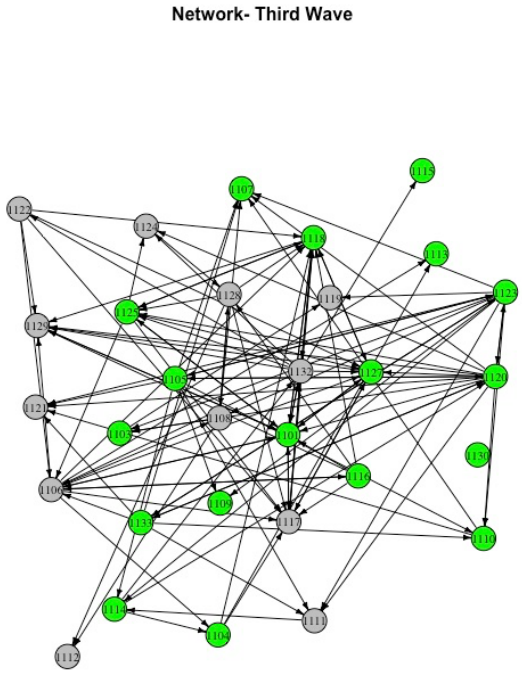


Figure 9 Network Third Wave

For the Network- first wave

Density: The density of the network is approximately 0.116, which indicates that about 11.6% of all possible ties are present in the network.

Reciprocity: The proportion of dyads that are symmetric (mutually connected) in the network is 0.864, indicating a high level of reciprocity where most dyads have mutual connections.

Reciprocity ignoring null dyads: When ignoring null dyads (dyads with no ties), the reciprocity is 0.262, referring that about 26.2% of the non-null dyads have mutual connections.

Transitivity: The transitivity of the network is approximately 0.455, which means that the extent to which ties in the network tend to cluster together.

Comparison with the Third Wave:

Comparing the network metrics between the third wave and the first wave:

The density of the third wave network (0.134) is slightly higher than that of the first wave network (0.116), indicating a greater proportion of ties in the third wave.

Reciprocity: The proportion of dyads that are symmetric (mutually connected) in the third wave network (0.818) is slightly lower than that of the first wave network (0.864), suggesting a slight decrease in reciprocal connections between nodes.

Reciprocity (excluding null dyads): When ignoring null dyads, the reciprocity of the third wave network (0.194) is notably lower than that of the first wave network (0.262), indicating a decrease in reciprocal connections when considering only non-null dyads.

Transitivity: The transitivity of the third wave network (0.524) is slightly higher than that of the first wave network (0.455), indicating a greater tendency for ties to cluster together in the third wave.

Dyad Census:
In the first wave, there are 21 mutual dyads, 59 asymmetric dyads, and 355 null dyads.

In the third wave, there are 19 mutual dyads, 79 asymmetric dyads, and 337 null dyads.

Overall, the number of mutual dyads is slightly lower in the third wave compared to the first wave indicating a minor decrease in reciprocated connections. Conversely, both the number of asymmetric and null dyads is slightly higher in thirdwave suggesting a slight increase in one-sided and non-existent connections in the network.

Triad Census:
In the first wave, there are 2229 triads with no ties among the three nodes, indicating disconnected groups. Additionally, there are 1088 triads with one mutual and one asymmetric tie, suggesting mixed reciprocated and unreciprocated relationships.
In the third wave, similar patterns are observed with 1921 triads of type 003 and variations in other triad types, reflecting potential changes in the network structure over time.

SAOMs (Stochastic Actor Oriented Models):

```
# Prepare data in "RSiena" format:
# DEPENDENT NETWORK:
ask <- sienaDependent(array(c(askhelp_wave1_mat, askhelp_wave2_mat, askhelp_wave3_mat),
            dim=c(30, 30 ,3)))
# PREDICTOR NETWORK (time-varying):
friendship <- varDyadCovar(array(c(friendship_wave1_mat, friendship_wave2_mat),
            dim=c(30, 30, 2)))
# CONSTANT ACTOR ATTRIBUTES:
# CONSTANT ACTOR ATTRIBUTES:
gender <- coCovar(df$sex)
gpa <- coCovar(df$gradePoint)
# TIME-VARYING ACTOR ATTRIBUTE:
dispute_wave <- varCovar(ddd_mat[, 1:2])
# JOIN THEM ALL INTO DATA OBJECT FOR SAOM:
Thedata <- sienaDataCreate(ask, friendship, gender, gpa, dispute_wave)
```

```
# Specify model for the dynamics of advice seeking:
(Themodel <- getEffects(Thedata)) # default: rate, outdegree, reciprocity
# ADD EFFECTS TO OBJECTIVE FUNCTION:
Themodel <- includeEffects(Themodel,
        transTrip, transRecTrip) # clustering
Themodel <- includeEffects(Themodel,
        X, interaction1='friendship') # friendship main effect
Themodel <- includeEffects(Themodel,
        egoX,altX, # sender, receiver, and ..
        sameX,interaction1='gender') # .. homophily effects on gender
Themodel <- includeEffects(Themodel,
        sameX, interaction1='gpa') # educational background homophily
Themodel <- includeEffects(Themodel,
        egoX, altX, # sender, receiver, and ..
        simX, interaction1='dispute_wave') # .. homophily effects on performance
# INSPECT MODEL SPECIFICATION:
Themodel
```

Effect	par.	(s.e.)	t stat.
Rate 1	9.77	(2.03)	.
Rate 2	10.34	(2.17)	.
outdegree (density)	–1.95***	(0.14)	–13.47
reciprocity	0.52†	(0.28)	1.87
transitive triplets	0.48***	(0.08)	6.22
transitive recipr. triplets	–0.43*	(0.18)	–2.30
friendship	0.90***	(0.14)	6.24
gender alter	0.05	(0.13)	0.43
gender ego	–0.06	(0.13)	–0.45
same gender	0.16	(0.13)	1.20
same gpa	–0.02	(0.13)	–0.15
dispute-wave alter	0.01	(0.01)	0.74
dispute-wave ego	–0.00	(0.01)	–0.36
dispute-wave similarity	0.68	(0.94)	0.72

† p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001;
all convergence t ratios < 0.05.
Overall maximum convergence ratio 0.11.

Figure 10 Model 1

According to RSiena manual, convergence is excellent when the overall maximum convergence ratio is less than 0.2, and for all the individual parameters the t-ratios for convergence all are less than 0.1 in absolute value; convergence is reasonable when the former is less than 0.30.

Based on that, we can tell that the model has reached to convergence.

Interpretation:
The rate goes higher.

Outdegree (Density): The estimated coefficient is -1.93 with a standard error of 0.13. This effect is statistically significant ($p < 0.001$), referring that as the outdegree (density) of the network increases, the rate of help-seeking behavior decreases. In other words, individuals with more connections (higher outdegree) are less likely to seek help.

Reciprocity: The estimated coefficient is 0.50 with a standard error of 0.24. This effect is marginally statistically significant ($p < 0.1$) at a marginal level, implying that reciprocity in the network has a positive effect on help-seeking behavior.

Transitive Triplets: The estimated coefficient is 0.47 with a standard error of 0.07. This effect is statistically significant ($p < 0.001$), indicating that transitive triplets (clusters of interconnected nodes) have a positive effect on advice-seeking behavior.
From my point of view, when the network exhibits transitive triplets, where node A is connected to node B, node B is connected to node C, and consequently, node A is also connected to node C, it suggests a higher level of cohesion or interconnectedness within the network. In simpler terms, when clusters of nodes are interconnected, students are more likely to seek help within these clusters. This positive effect implies that cohesive subgroups within the network play a role in facilitating help-seeking behavior.

Transitive Reciprocal Triplets: The estimated coefficient is -0.41 with a standard error of 0.16. This effect is statistically significant ($p < 0.05$), suggesting that transitive reciprocal triplets have a negative effect on help-seeking behavior.

Friendship: The estimated coefficient is 0.90 with a standard error of 0.14. This effect is statistically significant ($p < 0.001$), indicating that friendship ties have a positive effect on help-seeking behavior.

The effects of the same gender, same types of GPA, and performance in dispute solving are not statistically significant.

Now, enriching the model by endogenous, degree.related terms:

Effect	par.	(s.e.)	t stat.
Rate 1	7.00	(1.07)	.
Rate 2	7.55	(1.27)	.
outdegree (density)	–1.14†	(0.64)	–1.77
reciprocity	0.35	(0.61)	0.58
transitive triplets	0.71***	(0.19)	3.76
transitive recipr. triplets	0.19	(0.57)	0.34
indegree - popularity	0.07	(0.05)	1.34
outdegree - activity	0.03	(0.03)	1.08
indegree - activity	–0.68†	(0.40)	–1.73
friendship	1.84***	(0.53)	3.48
gender alter	0.08	(0.19)	0.41
gender ego	–0.18	(0.27)	–0.67
same gender	0.11	(0.21)	0.53
same gpa	–0.43	(0.29)	–1.48
dispute-wave alter	0.02	(0.02)	0.97
dispute-wave ego	–0.01	(0.02)	–0.41
dispute-wave similarity	1.67	(1.83)	0.91

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;
all convergence t ratios < 0.1.
Overall maximum convergence ratio 0.18.

Figure 11 Model 2- Matthew Effect

There is no evidence for the Matthew effect (indegree popularity).

Goodness of Fit:

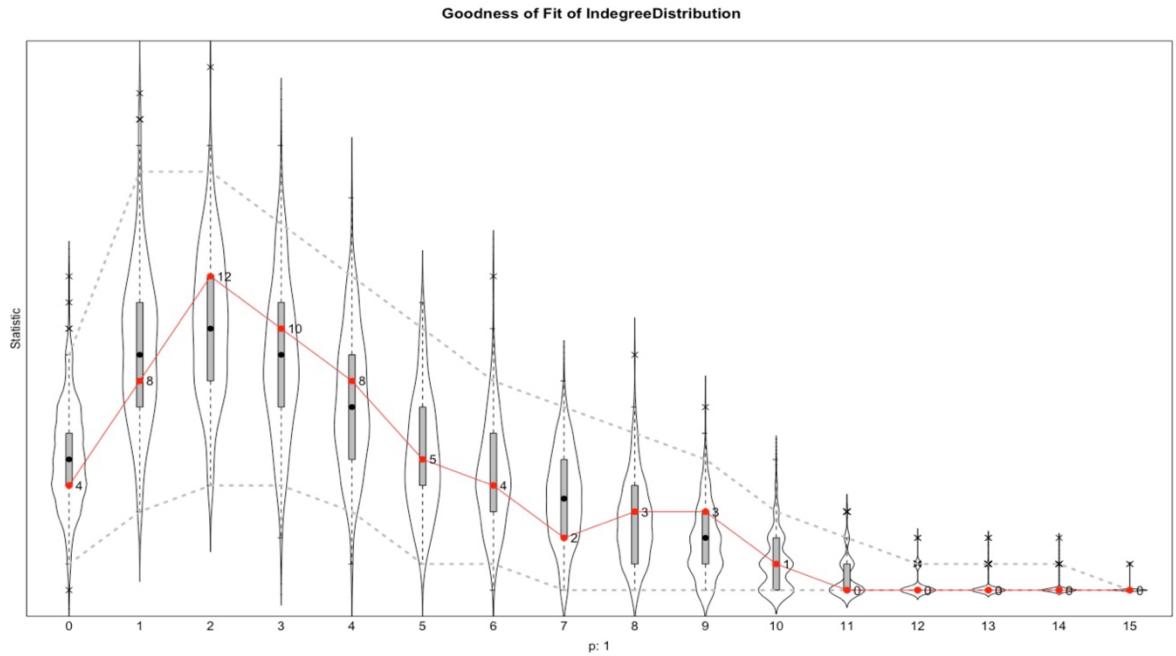


Figure 12 Goodness of Fit Indegree Distribution

P value is in a really non-significant region, fit is acceptable.

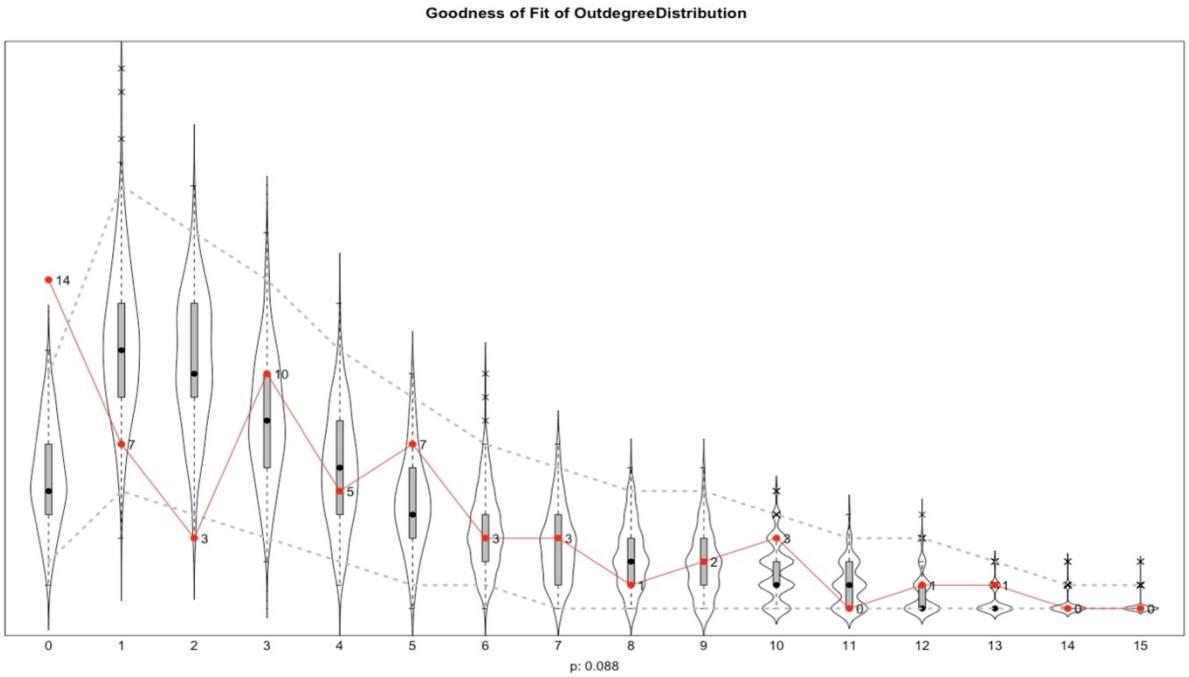


Figure 13 Goodness of Fit- OutDegree Distribution.

P value is in a non-significant region, fit is acceptable. Not as good as out-degree distribution.

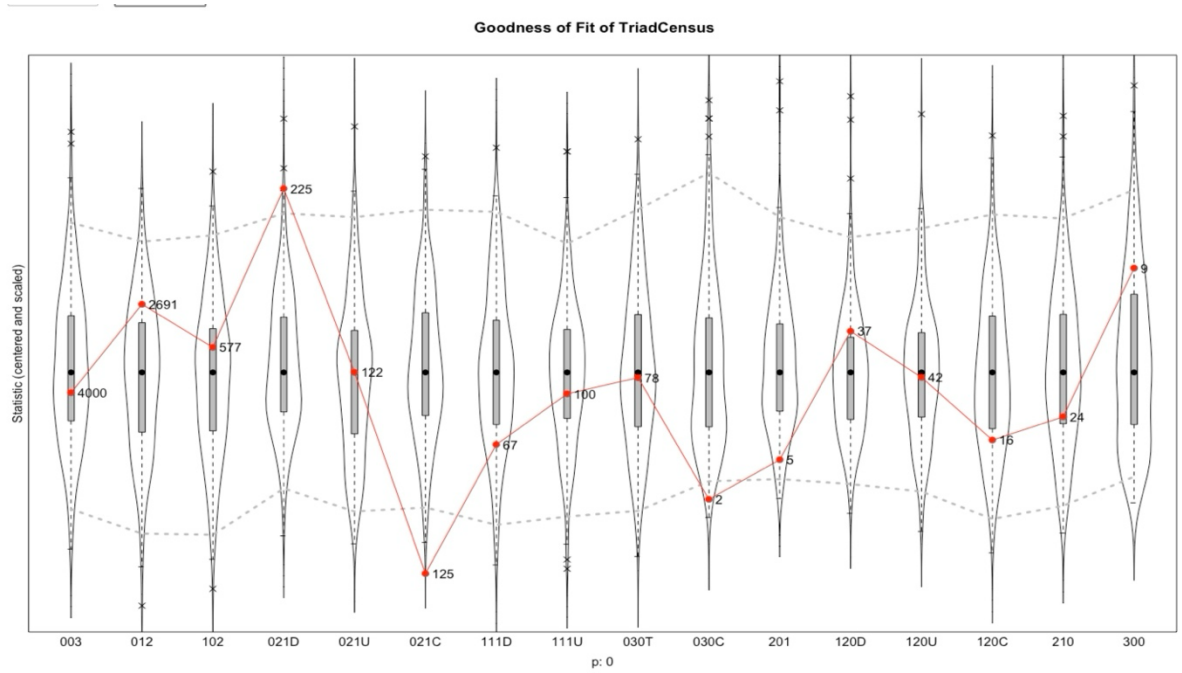


Figure 14 Goodness of Fit - Triad Census

Seems ok to me. The p value is 0, this is a Monte-Carlo p-value for 250 simulated datasets. One could improve the model by including inPop (indegree popularity) and reciAct (reciprocated degree activity).
After adding these effects and running the model

Effect	par.	(s.e.)	t stat.
Rate 1	9.47	(1.78)	.
Rate 2	10.14	(1.96)	.
outdegree (density)	–1.70***	(0.21)	–8.25
reciprocity	0.85*	(0.34)	2.53
transitive triplets	0.56***	(0.09)	6.49
transitive recipr. triplets	–0.32†	(0.19)	–1.73
indegree - popularity	–0.05	(0.03)	–1.47
rec.degree - activity	–0.12	(0.07)	–1.63
friendship	0.88***	(0.16)	5.55
gender alter	0.06	(0.13)	0.45
gender ego	–0.06	(0.13)	–0.43
same gender	0.15	(0.13)	1.15
same gpa	–0.04	(0.14)	–0.28
dispute-wave alter	0.01	(0.01)	0.72
dispute-wave ego	–0.00	(0.01)	–0.36
dispute-wave similarity	0.67	(0.99)	0.68

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;
all convergence t ratios < 0.12.
Overall maximum convergence ratio 0.2.

Figure 15 Model 3- Triad Census

While the model converged successfully, some newly introduced effects like indegree-popularity and reciprocated degree activity did not show statistical significance. This suggests that these factors may not significantly contribute to the stability of network modeling. Other effects observed in the model remained consistent with our initial findings. Although there were differences in the rate of effects, no new factors emerged to impact help-seeking behavior. Overall, these results offer valuable insights into understanding the dynamics of help-seeking within the network.

Findings:

Our analysis reveals several key insights into the dynamics of help-seeking behavior within the school network of Hungary. Notably, I found that as the network's outdegree (density) increases, the rate of help-seeking behavior decreases, implying that individuals are less likely to seek help in denser networks. Additionally, reciprocity within the network positively influences help-seeking behavior with a marginal p-value, referring that people tend to seek help when there is a sense of mutual aid. Moreover,

the presence of transitive triplets (clusters of interconnected nodes) was associated with a higher likelihood of seeking help. Transitive reciprocated triplets negatively influence help-seeking with a p-value (< 0.05). Friendship positively influences help-seeking within the network with a highly statistically significant p-value ($p < 0.001$). However, the significance of certain factors, such as indegree-popularity, gender, GPA, and leadership skills (dispute solving) was not observed, indicating that these factors may not significantly influence help-seeking behavior within the network.

Comparison with the Hypotheses and Conclusion:

Homophily: I assumed that people with similar attributes such as gender, higher or lower GPA, friends, and leadership skills are more likely to seek help from each other. But among these factors, only friendship has a notable impact on help-seeking.

Reciprocity: I expected a larger impact of reciprocity on help-seeking, but it is not. Reciprocity impacts help-seeking but with a marginal p-value ($p < 0.1$). The reason behind my assumption was that helping each other is most of the time mutual, especially in educational institutions where the targets, and challenges are almost the same among the students.

Triad: The interconnected clusters, known as transitive triplets, have a positive influence on help-seeking behavior. This means that when people are part of a very closed group, they are more likely to seek help. However, it is also found that reciprocal ties within these clusters, called transitive reciprocal triplets, actually have a negative effect on seeking advice that seems contradictory to each other. One explanation of this issue from my point of view is while being part of a close-tied group encourages seeking help, too much exchange within these groups could actually discourage it. So, whether seeking advice is encouraged or discouraged depends on how balanced the network is between close connections and reciprocal interactions.

Limitations:

1. Despite the similarity in results between the first and third models and the convergence of all three models, I remain skeptical about the stability of these models. A deeper understanding is required to conduct a more nuanced analysis.

2. The explanation for the contradictory effects between transitive triplets and transitive reciprocal triplets is based solely on my assumptions. I lack scientific or researched evidence to support this interpretation.

Note: I did not all the figures that I plotted in the R script. In this report, I only included those figures that were nice and I thought relevant and important.

Reference:
[1] Snijders, T. A. B. (2023). Goodness of fit testing in RSiena. University of Groningen, University of Oxford.

[2]Snijders, T. A. B. (2023). Descriptive statistics and Visualization. R Script.
[3] Ripley, R. M., Snijders, T. A. B., Boda, Z., Vo ŕ ös, A., & Preciado, P. (2024). Manual for RSiena (pp. 71-75). University of Oxford: Department of Statistics; Nuffield College University of Groningen: Department of Sociology.