# Exploring the Effects of Sex and Drink on Friendship Network

KU Leuven - Network Analysis - 18 May 2024

Serkan Shentyurk - ID: r0833419

## Introduction

The data used in the paper is obtained from a paper that conducted longitudinal survey in Hungarian high schools (Vörös and Snijders, 2017). Data is obtained from different cities, schools, and classroom for 4 years. Same cohort of students are followed throughout the high school. Students answered set of questions that measure affective and trust relationships. We are going to utilise wave 1, collected after the start of the first high-school year in October 2010, and wave 2 collected in April 2011. However, we only have a portion of the data. There are 36 students, 10 males and 26 females, in our network and 4 different data is available:

1. **Gender** It is a binary categorical variable. 1 corresponds to males and 2 corresponds to females.
2. **Alcohol consumption** It is an ordinal variable from 1 to 4.
3. **Friendship Network** It is an ordinal relation from -2 to 2, where -2 indicates dislike, 0 indicates neutral, and 2 indicates like.
4. **Trust Network** It is a binary relation.

In this paper we analysed the effects of gender and alcohol consumption on friendship network. We hypothesise that:

$H_1$: Gender is driving force of the evolution of friendship network.
$H_2$: Similar alcohol consumption is driving force of the evolution of friendship network.

We explored the network and applied different methods to test our hypothesis throughout this paper.

## Explaratory Analysis

We first started with visualising the friendship (Figure 1). Blue circles represent males and red squares represent females. Size of the nodes are proportional to drink variable.

The degree, gender, and drink homophily is calculated (Table 1). It is seen that there is a decrease in degree homophily and gender homophily, whereas there is an increase in drink homophily.

**Table 1: Homophily Values.** There is a decrease in degree and gender homophily, wherease an increase in drink homophily over time.

| | Degree Homophily | Gender Homophily | Drink Homophily |
|---|---|---|---|
| Wave 1 | 0.1153 | -0.0557 | 0.0338 |
| Wave 2 | 0.0899 | -0.0809 | 0.1689 |

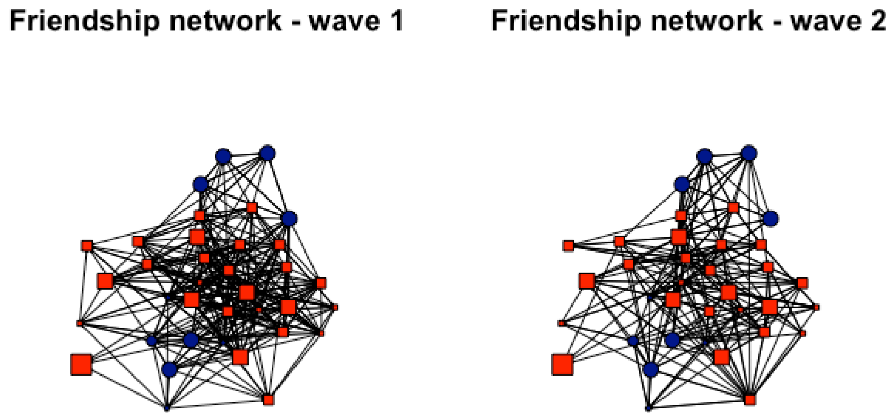


**Figure 1: Friendship Network in Wave 1 and Wave 2**. The colours and shapes represent gender and the size of the shapes represent the drinking pattern.

The density of the wave 1 is 0.2882 and the density of the wave 2 is 0.2045. To observe the gender effect, the gender densities are calculated and normalised (Table 2).

**Table 2: Gender Density.** There is a decrease in female-female over time.

| | Male | | Female | |
|---|---|---|---|---|
| | Wave 1 | Wave 2 | Wave 1 | Wave 2 |
| Male | 1.041 | 1.2498 | 0.3374 | 0.3931 |
| Female | 0.2720 | 0.3033 | 1.1827 | 1.0675 |

In-degree and out-degree frequencies are calculated and plotted in Figure 2. It is observed that as the time progresses the number of edges decreases.

Jaccard index is also calculated to 0.3897 which shows that 40% of the friendships are maintained between two waves.

## Structural Analysis

Understanding the structure of the network is important to gather more insights from the network. Dyad and triad census is a easier method to gather information from the network. Dyad and triad census for 2 waves are calculated and presented in Table 3. Number of mutual and asymmetric dyads decrease whereas there is an increase in null dyads. A general trend of decrease in triads except for the triad 003 is also observed. This indicates that there is a general trend of decrease in terms of friendship between 2 waves in this particular network. Random networks are generated with the same structure to see whether the dyad and triad structures can be seen randomly. The distribution of dyad census (Figure 3) and triad census (Figure 4) are plotted and red boxes represent the values of the real network. The counts of both dyadic structures and triad structures are different than randomly generated networks.

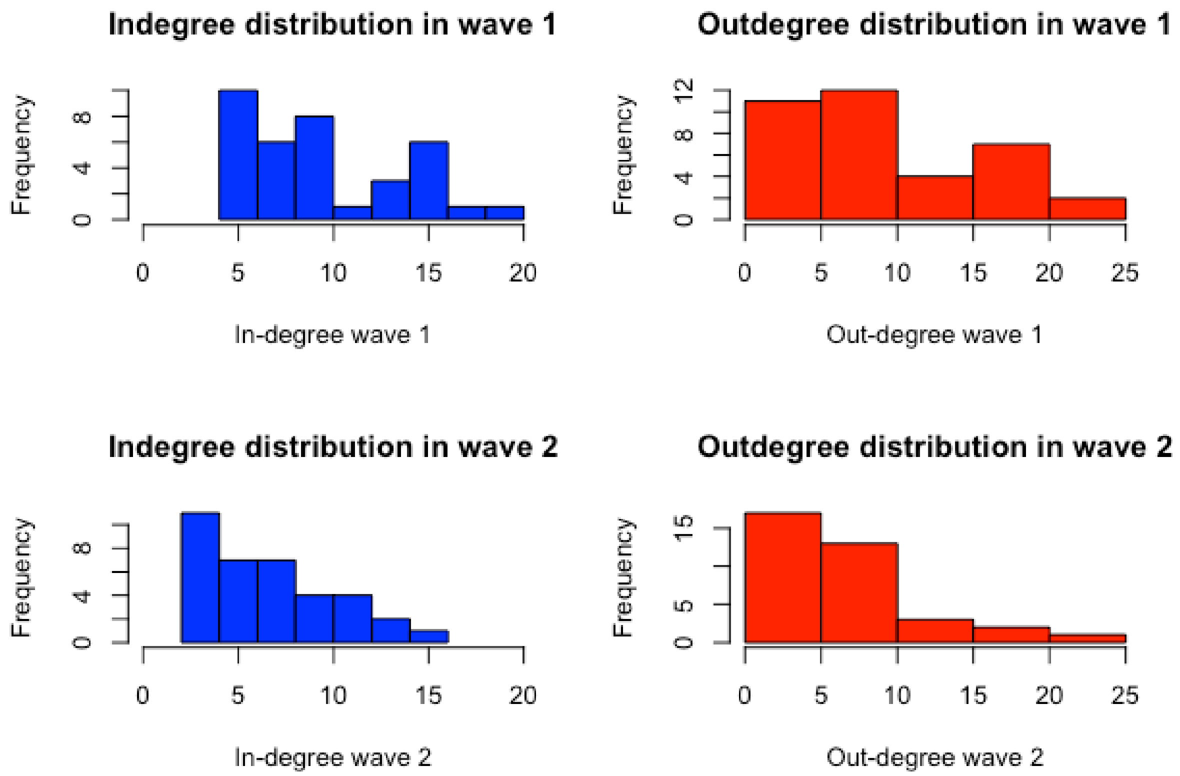


**Figure 2: In-degree and Out-degree Distribution in Wave 1 and Wave 2.** There is a decrease in both in-degree and out-degree distributions over time.

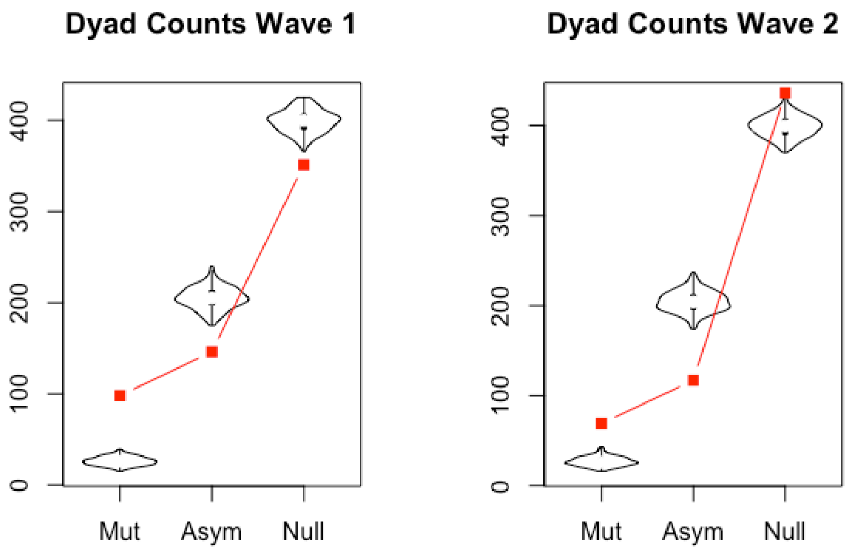


**Figure 3: Dyad Structures of Random Networks.** The real values are different than the randomly generated network.

**Table 3: Dyad and Triad Census.** There is a tendency of decrease in dyad and triad structures.

| | Wave 1 | Wave 2 |
|---|---|---|
| Dyad Mutual | 98 | 69 |
| Dyad Asymmetric | 146 | 117 |
| Dyad Null | 351 | 436 |
| Triad 003 | 1473 | 2465 |
| Triad 012 | 1617 | 1726 |
| Triad 102 | 1137 | 1195 |
| Triad 021D | 279 | 271 |
| Triad 021U | 179 | 159 |
| Triad 021C | 168 | 114 |
| Triad 111D | 348 | 253 |
| Triad 111U | 476 | 308 |
| Triad 030T | 124 | 68 |
| Triad 030C | 1 | 0 |
| Triad 201 | 206 | 110 |
| Triad 120D | 107 | 63 |
| Triad 120U | 127 | 39 |
| Triad 120C | 39 | 17 |
| Triad 210 | 204 | 70 |
| Triad 300 | 60 | 24 |

### Structural Equivalence

Equivalence expresses how similar the actors are in terms of the actors they are connected to. Hamming distance is one of the quantification that measures the equivalence. We clustered the actors in wave 1 (Figure 5) and wave 2 (Figure 6) based on Hamming distance. For the wave 1, we decided to cut the tree around the height of 35, resulting in 6 clusters and for the wave 2, we decided to cut the tree around the height of 25, resulting in 6 clusters.

We applied blockmodelling with 6 clusters, and also applied k-cores and community detection algorithm for wave 1 (Figure 7 and for wave 2 (Figure 8), outputting different results for each method.

## Methods

### Multiple Regression Quadratic Assignment Procedure

To assess how the friendships in wave 1 affect the network in wave 2, multiple regression quadratic assignment procedure (MRQAP) is used. Friendship at wave 2 is regressed against Friendship at wave 1, same sex, sender sex, receiver sex, same drinking pattern, sender drinking pattern, and receiver drinking pattern. The significant parameters are sender sex = -0.4021 ($p < 0.05$) and same drink = 0.1765 ($p < 0.05$). If the sender is female the log odds of a friendship tie existing between them decreases and if the students have similar drinking pattern, the log odds of a friendship tie existing between them increases.

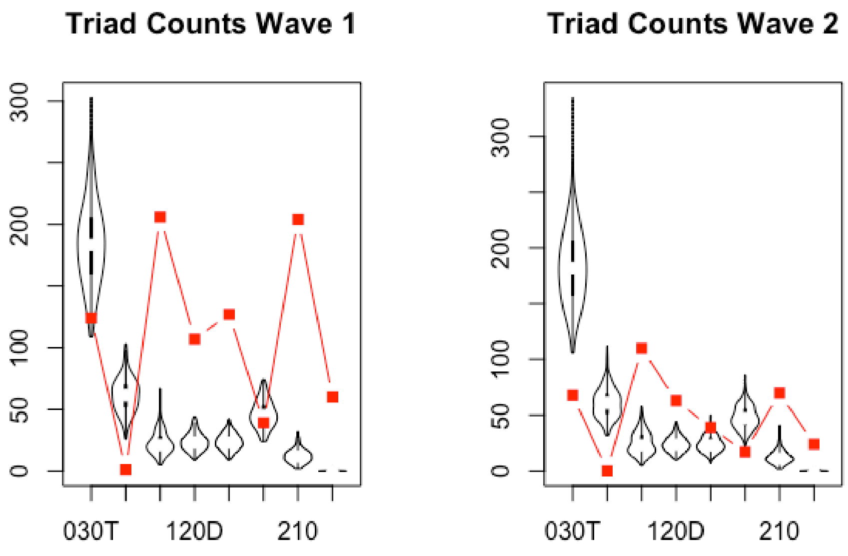


**Figure 4: Triad Structures of Random Networks.** The real values are different than the randomly generated network.

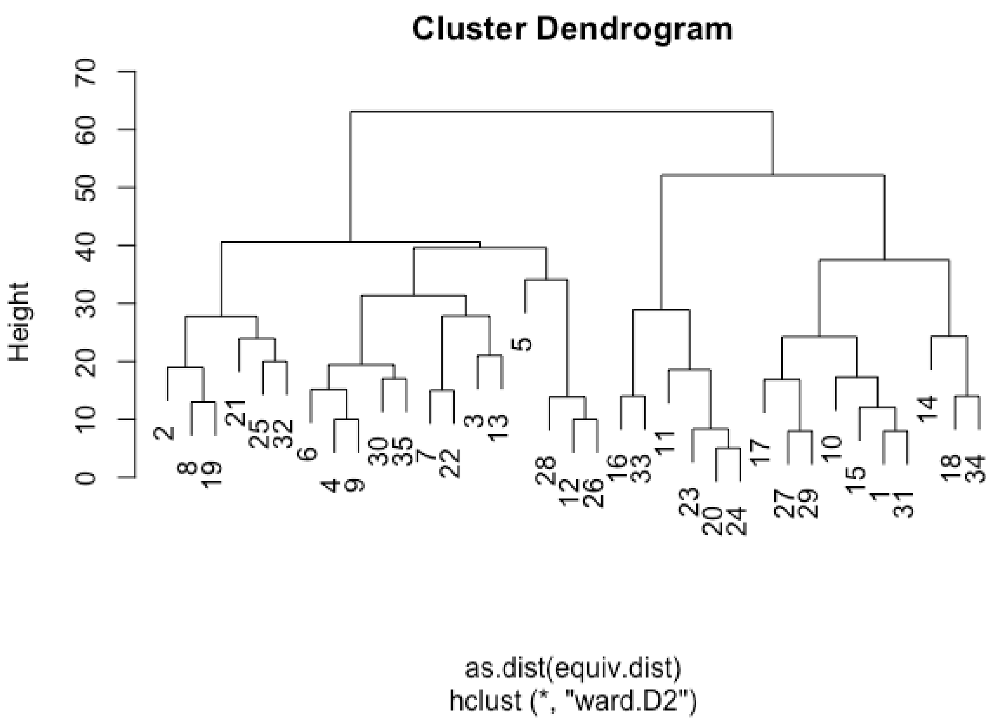


**Figure 5: Dendrogram of Wave 1.** When it is cutted at the height of 35, there are 6 clusters.

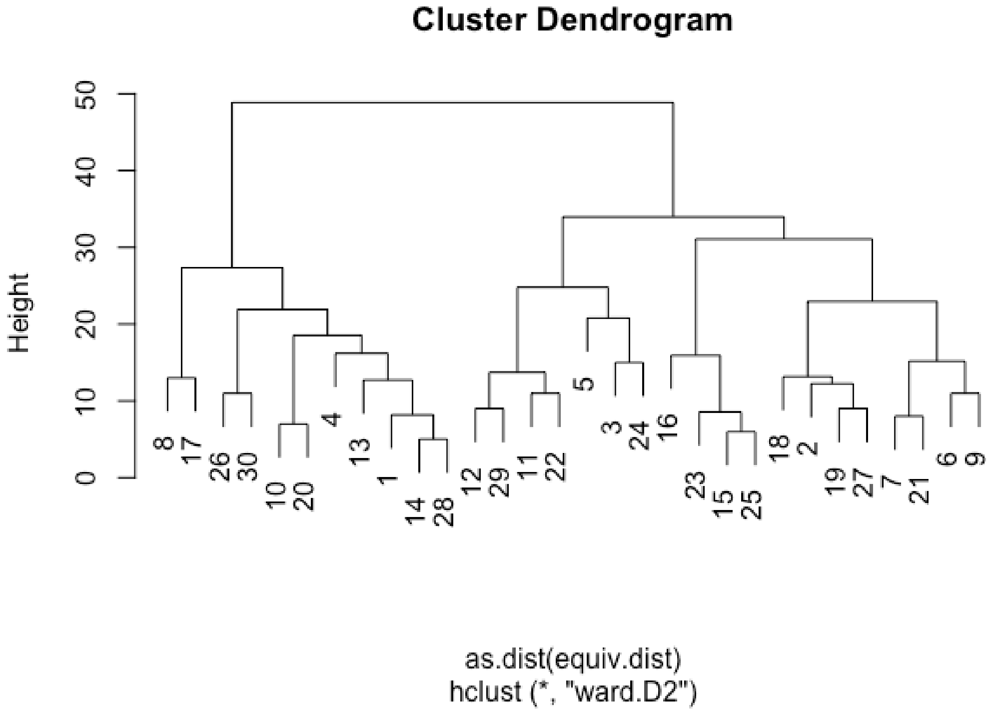


**Figure 6: Dendrogram of Wave 2.** When it is cutted at the height of 25, there are 6 clusters.

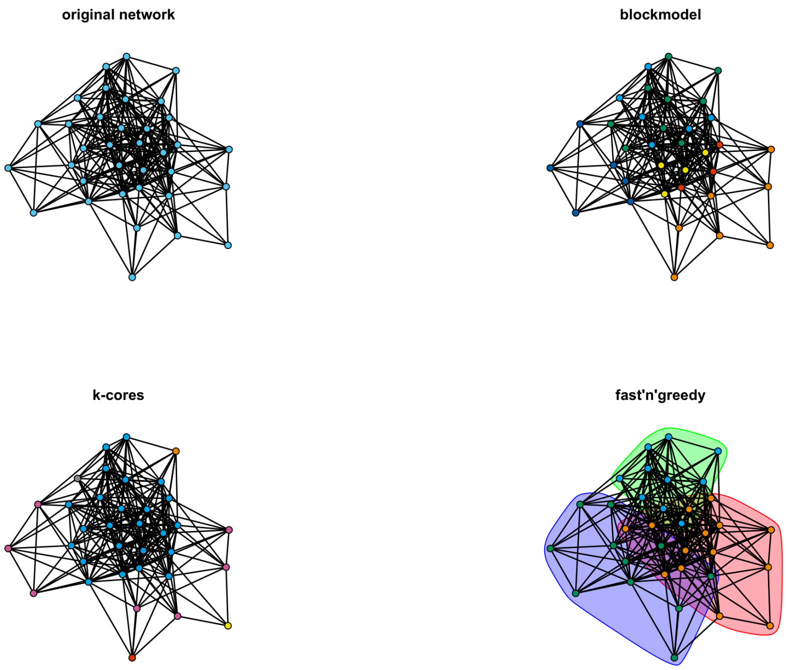


**Figure 7: Different Clustering Methods Applied to Wave 1.** Different methods outputted different clusters.

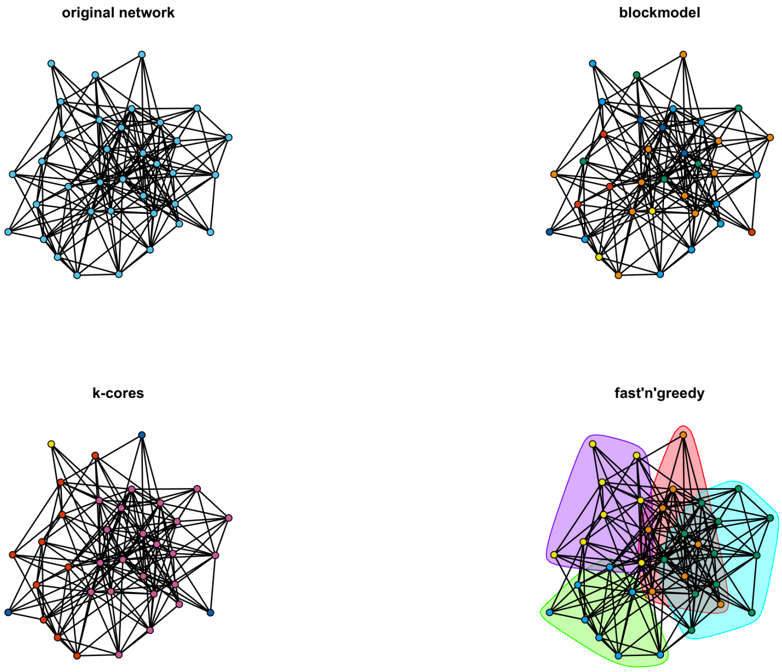


**Figure 8: Different Clustering Methods Applied to Wave 2.** Different methods outputted different clusters.

## Exponential Random Graph Models

Exponential random graph models (ERGMs) are statistical models for investigating network structure at one particular instant. We are interested in whether drinking pattern similarity and gender similarity affect the pattern formation.

We started applying the basic model, and we added other effects. The initial models are omitted, only the final model and results are presented. The final model assesses the effects of mutual friendship, triangle formation, same sex, sender's sex, receiver's sex, same drinking pattern, sender's drinking pattern, and receiver's drinking pattern (Table 4). We were not able to include the lower patterns of triangle formation, such as two-star configuration, alternating two-star configuration, and two star configuration without mutual ties, since the algorithm fails to converge and outputs the results. We encountered the same problem when ERGM is fitted to the wave 2, therefore, we only present the result from the wave 1.

It is seen that homophily, gender homophily, triangle formation, and the sender's drinking pattern are important factors in friendship formation ($p < 0.05$) (Table 4) and the estimates are 1.79165, 0.37321, 1.09400, -0.18112, respectively. The results show that there is a reciprocity and gender reciprocity: If X likes Y, Y has higher probability to like X. It is also seen that there is a strong positive association between the presence of triangles and the likelihood of edges. Additionally, as the person drinks more, the person tends to form less friendship.

Goodness of fit tests across the models output similar AIC and BIC results, which shows that increasing the parameters do not substantially increase the model. The performance of the other parameters the model does not explicitly estimate is shown in Figure 9.

**Table 4: Results of ERGM Model.** Dyad structures, same sex, and sender actor's drinking pattern are significant.

| | Estimate | Std. Error | MCMC % | z value | Pr(>\|z\|) |
|---|---|---|---|---|---|
| edges | -2.81702 | 0.61572 | 0 | -4.575 | < 1e-04 *** |
| mutual | 1.79165 | 0.20234 | 0 | 8.854 | < 1e-04 *** |
| gwesp.OTP.fixed.0 | 1.09400 | 0.48991 | 0 | 2.233 | 0.02554 * |
| nodematch.sex | 0.37321 | 0.13669 | 0 | 2.730 | 0.00633 ** |
| nodeicov.sex | 0.27976 | 0.17096 | 0 | 1.636 | 0.10175 |
| nodeocov.sex | -0.15031 | 0.17586 | 0 | -0.855 | 0.39271 |
| nodematch.drink | 0.01371 | 0.11715 | 0 | 0.117 | 0.90686 |
| nodeicov.drink | 0.03250 | 0.09325 | 0 | 0.349 | 0.72744 |
| nodeocov.drink | -0.18112 | 0.08629 | 0 | -2.099 | 0.03583 * |

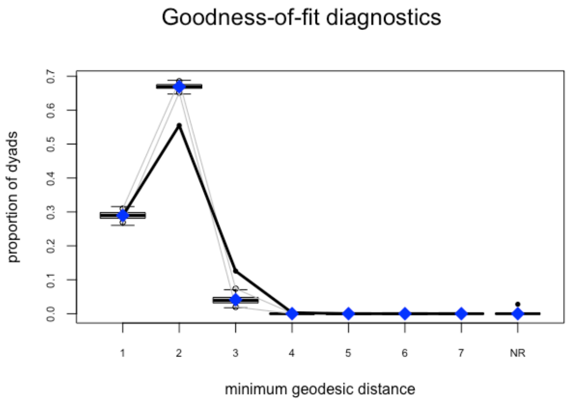


**Figure 9: Goodness-of-fit Diagnostics of ERGM.** The model performs differently depending on the geodesic distance.

## Stochastic Actor Oriented Models (SAOMs)

Stochastic Actor Oriented Models (SAOMs) are another type of statistical model that allow investigating the evolution of the network and estimating the drivers of the change over time. In our network (Figure 1), it is seen that there is a decrease in friendship (Table 3) and here, we are aiming to identify the drivers of this change.

We defined the model by including out-degree, reciprocity, transitive triplets, 3-cycles, in-degree, drinking pattern of the ego (sender), drinking pattern of the alter (receiver), matching drinking pattern, gender of the ego, gender of the alter, and matching gender (Table 5). The overall maximum convergence ratio of the model is 0.1357, the model converged well. The effects of out-degree, reciprocity, transitive triplets, 3-cycles, matching drinking pattern are significant and the estimates are -1.9171, 1.6395, 0.1591, -0.2498, 0.2312. The negative out-degree indicate that actors are less likely to initiate new connections. The positive reciprocity indicates that if one actor forms a tie with another, the other actor is likely to form a tie back. The positive transitive triplets means that if actor A likes actor B, and actor B likes actor C, then actor A is more likely to like actor C. The positive estimate for matching drinking pattern indicates that actors that have similar drinking patterns tend to form friendship more.

**Table 5: Results of SAOM Model.** Out-degree, reciprocity, transitive triplets, 3-cycles, matching drinking pattern are significant.

| | Estimate | Std. Error | Convergence t-ratio |
|---|---|---|---|
| 1. outdegree (density) | -1.9171 | (0.2403) | 0.0944 |
| 2. reciprocity | 1.6395 | (0.1935) | 0.0914 |
| 3. transitive triplets | 0.1591 | (0.0222) | 0.0928 |
| 4. 3-cycles | -0.2498 | (0.0549) | 0.1036 |
| 5. indegree - popularity | 0.0075 | (0.0195) | 0.0840 |
| 6. drink.coCovar alter | 0.1105 | (0.0767) | 0.0545 |
| 7. drink.coCovar ego | -0.0581 | (0.0742) | 0.0053 |
| 8. same drink.coCovar | 0.2312 | (0.1179) | 0.0786 |
| 9. gender.coCovar alter | 0.0351 | (0.1605) | 0.0110 |
| 10. gender.coCovar ego | -0.2958 | (0.1437) | -0.0268 |
| 11. same gender.coCovar | 0.0701 | (0.1445) | 0.0881 |

## Conclusion

The friendship network evolved and both in-degree and out-degree frequencies decreased over time (Figure 2), resulting a decrease in dyad and triads (Figure 3). Community detection algorithm found 3 different community in wave 1 (Figure 7) and 4 different community in wave 2 (Figure 8), indicating that the students tend to group into smaller groups, resulting in decrease in overall interaction between students. This is inline with the previous finding.

ERGM model showed that homophily, gender homophily, and sender's drinking pattern are the drivers of the network pattern in wave 1 (Table 4). Considering the first wave is conducted in the beginning of the high school, it is expected that students that have similar sex form friendship with each other.

SAOM model showed that out-degree, reciprocity, transitive triplets, 3-cycles, matching drinking pattern are the divers of the network evolution (Table 5). There is a tendency to decrease in outgoing ties, in other words, students tend to see less people as their friends. There is a tendency to form a friendship if the other person sees them as a friend. It is also seen that friend of a friend tend to be friend over time. Also, even though the matching drinking pattern was not significant in ERGM mode, it is seen that matching drinking pattern is important in the evolution of the network. Students who have similar drinking patterns tend to from friendship.

It is seen that the gender homophily is important in the formation of the friendship network, however, it is not important for the evolution. Drinking homophiliy is not important in the formation of the friendship network, but it is one of the drivers of the evolution of the network. Therefore, there is not enough information to accept the first hypothesis $H_1$ and we accept the second hypothesis $H_2$.

## References

Vörös, A., & Snijders, T. A. (2017). Cluster analysis of multiplex networks: Defining composite network measures. *Social Networks*, *49*. https://doi.org/10.1016/j.socnet.2017.01.002