Friendship Network - A Case Study

Yun-Tsz Tsai, Xinyi Jiang

March 07, 2025

Introduction

Many studies have shown that friendship network is highly reciprocative and clustered. However, there are still things that can not be fully explained by reciprocity alone. Sometimes it is might be the case that one considers the other to be a friend while the other individual does not perceive their relationship in the same way. This type of asymmetry could lead to status differences in a network, namely hierarchy.

Aside from this, other mechanisms such as transitivity (namely, befriending the friend of your friend's friend), preferential attachment (befriending someone who have already have many friends) can also play a role in the formation of friendship network. In this article, we use data from a Dutch secondary school in the school year of 2003/04 based on the PhD thesis of Knecht, A. B. (2008) to investigate different mechanisms that shape the friendship network of Dutch teenagers.

We focus on two mechanisms, namely transitive closure and preferential attachment (also known as, *Matthew effect*) in our study and Exponential Random Graph Models (ERGMs) are employed.

Operationalization

1. How transitive closure attributes to hierarchy in a network?

   Imagine a network where there is a node connected to a clique. Transitivity will close up the gap between the node and all the nodes in the clique, and therefore the node that is originally outside of the clique will become the node that have high degree centrality. The differences in degree centrality is then turn into implicit hierarchy in a network.

2. Why preferential attachment does not necessarily lead to transitive closure?

   Preferential attachment, in our case, refers to the tendency to befriend someone who have already had many friends. Suppose attachment tendency is so strong that every individual in the network will only like to befriend someone who is higher in the caste, then it is reasonable that the two individuals will not try to form a tie when both of them are connected to an individual who is more popular (in terms of the number of friends).

Aside from transitive closure and preferential attachment, we control for gender homophily as it is a well-documented mechanism in network formation. If excluding from the models, we risk overestimating the effects of other mechanisms.

Descriptive Information on the Data

Figure 1 shows the visualization of the network. The network consists of 18 individuals, 13 of which are male (in light blue). The size of the node is based on individual score of attitude towards school (with bigger nodes suggesting more positive views on school), and it is normalized for easier comparison. Besides, the network is directed.

As what we can see from the figure, it is clear that there is a strong gender homophily. In other words, girls mostly befriend girls and boys mostly befriend boys with a few exceptions. Overall, the individuals have shown rather positive views ($\mu = 4.77$) on school and it is therefore difficult to assess the impact of attitude towards school on the structure of the network.

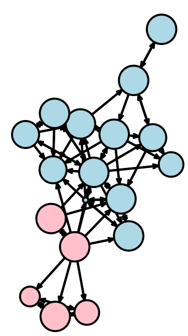
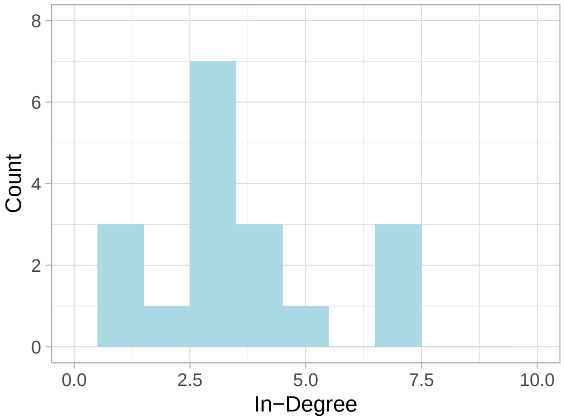


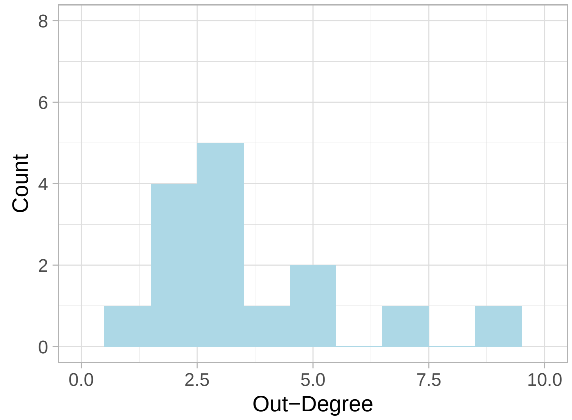
Figure 1. Visualization of Friendship Network by Gender

Figure 2 illustrates the distribution of in-degree ($\alpha^2 = 3.67$) and out-degree ($\alpha^2 = 7.67$) of the network. The distribution shows how the number of incoming ties and that of outgoing ties differ. If the network is perfectly symmetric, the distributions should be identical. In subgraph Figure 2a, we see most individuals receive 3 ties with a few receiving more than twice as many ties. In subgraph Figure 2b, although the number of outgoing ties is still 3, the distribution of ties is slightly more uneven with some receiving more than 9 ties.

As discussed previously, transitivity is the tendency to become friend with your friend's friend, or in more technical term, the tendency to close the triad. This empirical network has a transitivity score of ~ 0.21, suggesting that some clusterings occur, but many triad are still open. The variance in degree distribution and rather low transitivity both highlight the need to investigate the roles of transitive closureand preferential attachment.



(a) In



(b) Out

Figure 2. Distribution of Degrees

Model Specification

We estimated four models to examine the key mechanisms shaping the network:

- Full Model, which includes all structural tendencies
- No Closure Model, which removes transitive closure
- No Matthew Model, which excludes preferential attachment
- Null Model, which lacks both closure and preferential attachment. These models allow us to isolate the effects of transitive closure and preferential attachment on tie formation and network structure.

See Table 1 for results for all models.

Across all models, the probability of forming new ties remains relatively low, as indicated by the small but positive coefficients for edge formation. However, the No Closure Model exhibits a slightly lower edge coefficient compared to the Full Model (0.04 vs. 0.07), suggesting that transitive closure facilitates the formation of new ties by reinforcing indirect connections. This aligns with the idea that individuals are more likely to establish ties when they share mutual acquaintances. At the same time, reciprocity emerges as a stronger force in the No Closure and Null Models, indicating that when other structural tendencies are absent, relationships tend to form primarily through direct reciprocation rather than through indirect reinforcement.

Table 1. Results from ERGM Models

Different Model Specification

	Full	No Clo1	No Matt2	Null
	(1)	(2)	(3)	(4)
edges	0.07***	0.04***	0.10***	0.05***
	[0.02, 0.23]	[0.02, 0.11]	[0.04, 0.24]	[0.02, 0.09]
mutual	2.46	5.25***	3.27**	5.35***
	[0.97, 6.23]	[2.01, 13.75]	[1.39, 7.67]	[2.10, 13.60]
homophily*	2.13**	5.17***	2.19***	5.55***
	[1.30, 3.50]	[2.27, 11.79]	[1.38, 3.47]	[2.68, 11.50]
closure tend.*	3.60***		3.36***	
	[2.39, 5.41]		[2.23, 5.06]	
two path	0.73***		0.72***	
	[0.64, 0.84]		[0.63, 0.82]	
preferential tend*	3.04	1.60		
	[0.43, 21.50]	[0.23, 11.16]		
Num.Obs.	306	306	306	306
AIC	233.1	271.4	231.9	269.6
BIC	255.4	286.3	250.5	280.8

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

1 No transitive closure tendencies

2 No Matthew effect

* Refers to *Gender* homophily.

* Closure tendency

* Preferential attachment tendencies

Figures based on authors' calculation

Reference: *Knecht, A. B. (2008). Friendship selection and friends influence. Dynamics of networks and actor attributes in early adolescence. PhD dissertation. Utrecht University*

The Full and No Matthew Models, which include transitive closure, show strong and highly significant coefficients, confirming that individuals are likely to form ties with their friends' friends. The No Closure Model, which lacks this effect, has a noticeably worse model fit (higher AIC/BIC values), indicating that closure is a key factor in network structure. Without it, clustering is weaker, and the network becomes less cohesive.

Preferential attachment, on the other hand, does not emerge as a strong force shaping this network. The No Matthew Model, which excludes preferential attachment, still maintains strong clustering and a good model fit, indicating that transitive closure is the more dominant driver of tie formation. The weak and statistically uncertain estimates for preferential attachment further support this conclusion.

Among the four models, the No Matthew Model achieves the best fit, suggesting that while transitive closure is essential for structuring social networks, preferential attachment does not significantly enhance model performance. Without closure, networks become more fragmented, and individuals rely more on homophily for tie formation.

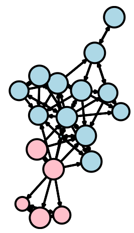
Simulation

Figure 3 shows the simulations of each model. We ran the simulations for each model for 100 times and present one of the 100 simulations for each model.

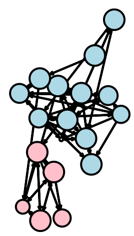
At first glance, all models produce networks that appear structurally similar, with no noticeable differences and all seem to replicate the empirical network fairly well. However, if we take a closer look at the numeric values of model parameters, the differences of models become visible and comparable.

Figure 4 presents the distribution of two key network parameters— density Figure 4a and transitivity Figure 4b across different models. The dashed line represents the observed value of the empirical network.

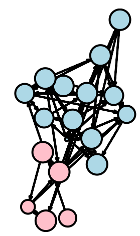
All models generated networks whose density is not far away from the empirical value, while if we look at transitivity, only the full model and no-Matthew managed to simulate networks with roughly the same number of triads as the empirical network. Given results from AIC and BIC (see Table 1), we can safely say that no-Matthew model is the model that reproduce our empirical network the best.



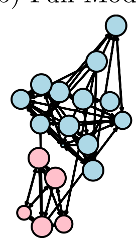
(a) Empirical Network



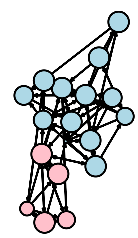
(b) Full Model



(c) Without Closure

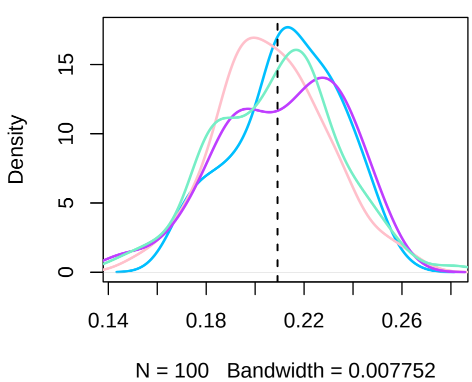


(d) Without Matthew Effect

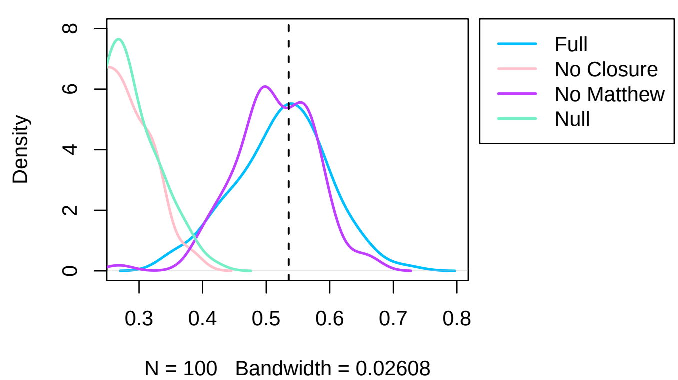


(e) Null Model

Figure 3. Visualization of Simulation



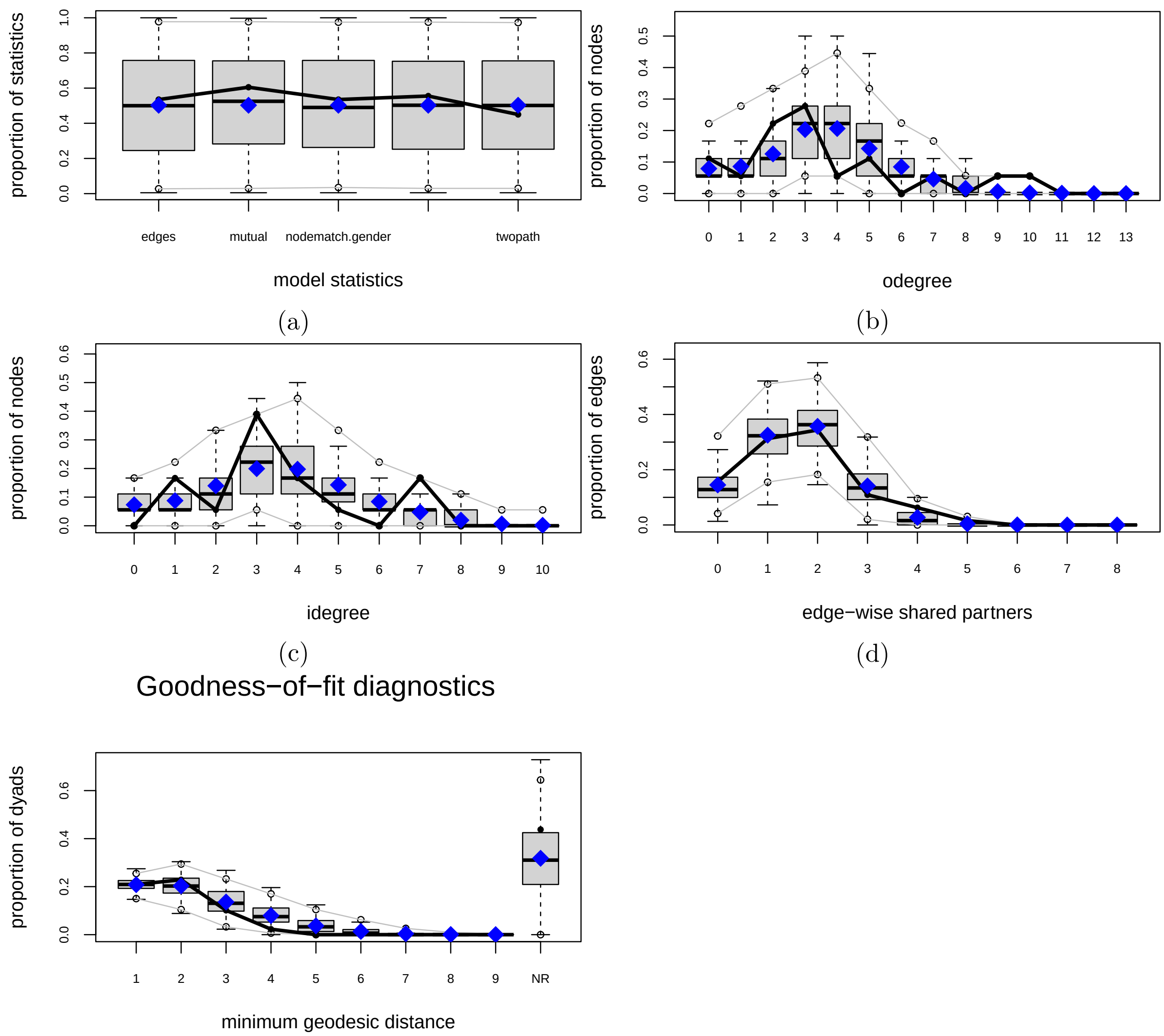
(a) Density



(b) Transitivity

Figure 4. Model Parameters

Goodness of Model Fit



(e)

Figure 5. Overall Evaluation on No-Matthew-effect Model

(a) Model Statistics: This plot compares observed and simulated values for key network statistics: edge count, mutual ties, gender homophily, and two-path structures. The black line represents observed statistics, while blue diamonds indicate simulated means with box plots capturing variation across simulations.For edges and mutual ties, the ob- served values are slightly higher than the simulated means, suggesting that the model underestimates the total number of ties and reciprocal connections. Gender homophily is reasonably well captured, though there is a slight tendency to underestimate the effect. The two-path structure aligns closely with the simulation, indicating that the model accurately captures indirect connections.

(b) Out-degree Distribution: The No-Matthew-effect Model fits well for lower-degree nodes (0–3), where the observed values align with the predicted means within the interquartile range. However, for nodes with an out-degree of 4–6, the model slightly overestimates their proportion, as the observed values fall below the simulated means. Additionally, for high-degree nodes (7 and above),the pattern is mixed. Some degree values, like 9 and 10, show an underestimation, certain point like degree 8, the observed values fall below the model's predictions. underestimating others. Overall, the model does a decent job of capturing the out-degree distribution, as most observed values fall within the simulated range. While there are some fluctuations, the fit is generally reasonable.

(c) In-degree Distribution: The in-degree distribution is generally captured by the model, as most observed values fall within the expected range of the simulations. However, there is a notable deviation at in-degree3 and 7, where the model appears to underestimate the value. This suggests that while the model overall provides a good fit, it may not fully account for certain structural patterns in how individuals receive ties.

(d) Edge-wise Shared Partners: Overall, the model fits well, with most points aligning closely with the boxplots and blue diamonds. The only noticeable deviation is at 4 shared partners, where the black line is slightly above the blue diamond, indicating a slight overestimation. Beyond this, the model captures the distribution effectively.

(e) Minimum Geodesic Distance: The model captures the distribution of shortest path lengths between dyads well, with most points falling within the boxplots. The black line (observed values) and blue diamonds (simulated means) are closely aligne

Conclusion

The results confirm that transitive closure alone is enough to create hierarchical structures in networks, even without explicit preferential attachment. It might seems counter-intuitive but is actually not.

Hierarchy in a network could simply happen in a case where A is connected to B, and B is connected to C, A is also likely to connect to C as a result of the connections. Over time, certain nodes repeatedly participate in these triadic formations, which makes them more central in the network. This process naturally leads to a core-periphery structure, where well-connected nodes form the core, while those with fewer ties remain on the periphery.

In our case, this finding appears fairly reasonably. Transitive closure is important in the formation of teenage friendship network while preferential attachment is not. Teenagers probably do not care as much about befriending popular kids in their class as befriending their friends' friends.

Reference

- Knecht, A. B. (2008). Friendship selection and friends influence. Dynamics of networks and actor attributes in early adolescence. PhD dissertation. Utrecht University.