A Comparative Study of Mathematical Modeling of High School Friendship Networks

Are high school friendship networks better described by the Erdös–Rényi Model or the Exponential Random Graph Models?

A Mathematics Extended Essay

Word Count: 3400

Table of Content

1. Introduction

As the world became interconnected, social network analysis has become an important field of study of the influence of networks' structure and the social interaction among factors in the social networks on human behavior. In the National Longitudinal Study of Adolescent to Adult Health (AddHealth study), Resnick et al. (1997) investigated the factors affecting the U.S. American adolescents' well-being by using the friendship networks constructed by interview data from the AddHealth study. After the study, researchers have been attempting to model the social networks to understand how the high school friendship networks form and evolve and to specify the structure of interaction based on the principle of random graphs (Toivonen et al., 2009; Goodreau et al., 2008; Jackson et al., 2007). Among the models used, Erdös–Rényi Model and Exponential Random Graphs Models (ERGMs) are the most common in high school relationship networks studies (Erdös et al., 1960; Frank et al., 1986; Jackson, 2010). While Erdös–Rényi Model is based on the assumption that networks are purely random, ERGMs assume that in a social network, the emergence of a relationship is influenced by the other relationships and individual attributes (Erdös et al., 1960; Robins et al., 2007). Compared to Erdös–Rényi Model, an ERGM focuses on the interaction between the structures of a social network (such as .... name the statistical components) and individual attributes (such as grade, race, gender) (Robins et al., 2007; Lubbers et al., 2007). Although the two kinds of models are thoroughly studied for eighty years of research on statistical network analysis, there are a limited number of literatures investigating the advantages and disadvantages of each model on a case-by-case basis (Wasserman et al., 1996; Goodreau et al., 2008; Jackson, 2010). Thus, I conduct a comparative study of Erdös–Rényi Model and Exponential Random Graph Models applied to high school friendship networks to answer the research question: "**Are the high school friendship networks better described by the random graph-based model or exponential random graph models?**"

2. Materials and Methods

In this paper, I use four network datasets extracted from the AddHealth study of Resnick et al. (1997), compiled in the *statnet* and *ergm* package of the programming language *R* by Handcock et al. (2003). The datasets I investigate include Faux Mesa High School (Mesa), Faux Magnolia High School (Magnolia), Faux Desert High School (Desert), and Faux Dixon High School (Dixon), which are in-school friendship networks in the U.S. The networks datasets include the number of students, either their friendship nominations or mutual friendships, and each student's grade, sex, and race. The grade levels are from seven to twelve. The races include White (non-Hispanic), Black (non-Hispanic), Hispanic, Asian (non-Hispanic), Native American, and Other (non-Hispanic) (Resnick et al., 1997; Hunter et al., 2008). Using calculation tools of the *statnet* and *ergm* package in *R* developed by Hunter et al. (2008), I attempt to calculate the statistical characteristics of each friendship network, then model the networks using the Erdös–Rényi Model and Exponential Random Graph Model, and evaluate the models' goodness of fit by comparing randomly simulated networks to original networks. From the evaluation of the results, I discuss models' accuracy in describing high school friendship networks.

2. Representing and Measuring High School Social Networks

In this paper, I use the theoretical framework developed by Jackson et al. (2010) to represent and measure high school friendship networks. I choose mathematical representation, including graphs, matrices, and relations, to denote social networks. The nodes and links are also referred to as vertices and edges, respectively. The links between nodes represent the connections between students. The networks I work with are unweighted, as I do not take the weight of the relations (links) into consideration; in other words, all the friendship in this paper is considered equal.

2.1. Representing High School Social Networks (Jackson et al., 2010)

Let $N = \{1, \ldots, n\}$ be the set of n nodes referred to as students in the high school social networks. A graph (N, g) consists of a set of nodes $N = \{1, \ldots, n\}$ and a real-valued $n \times n$ matrix, g, where g_{ij} represents the relation between any two nodes, i and j. The matrix is the adjacency matrix, as it denotes which nodes are linked (adjacent) to each other. g also can be written as the list of links $g = \{ij\}$.

Based on the characteristics of the edges, there are two kinds of social networks: undirected networks and directed networks. In the high school social networks, undirected networks are formed by mutual friendship between students which indicates whether any two students are friends mutually, while directed networks are formed by friendship nominations which means there are students not mutually nominating each other. In this essay, Mesa and Magnolia social networks of friendship are the undirected networks; Desert and Dixon are the directed network.

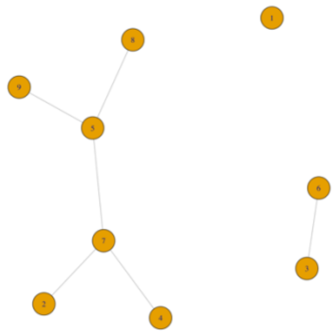

Figure 1.1: Sample undirected social networks

In Fig. 1.1, I present a sample social network $(N_{1.1}, g_{1.1})$ with $N_{1.1} = \{1, \dots, 9\}$, and the matrix $g_{1.1}$, which can be represented by

$$g_{1.1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

or $g_{1.1} = \{27, 36, 47, 57, 63, 72, 74, 75, 85, 95\}$. In matrix $g_{1.1}$, I denote each link by the elements of the matrix. For instance, since node 6 and node 3 are connected to each other, $g_{36} = g_{63} = 1$ ($g_{ij} = 1$ since the network is unweighted). Therefore, in a directed network, $g_{ij} = g_{ji}$.

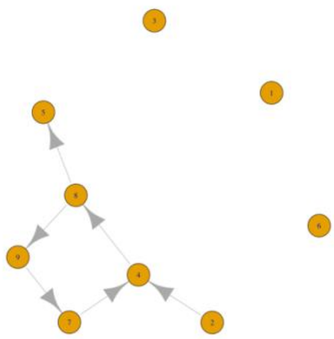

Figure 1.2: Sample directed social networks

In Fig. 1.2, I present a sample social network $(N_{1.2}, g_{1.2})$ with $N_{1.2} = \{1, \dots, 9\}$, and the matrix $g_{1.2}$, which can be represented by

$$g_{1.2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

or $g_{1.2} = \{24, 48, 74, 85, 89, 97\}$. In matrix $g_{1.2}$, I can see that $g_{42} \neq g_{24}$, as $g_{24} = 1$ and $g_{42} = 0$. Hence, in a directed network, $g_{ij} \neq g_{ji}$.

2.2. Summary Statistics of the High School Social Networks (Jackson, 2010; Jackson, 2021; Lubbers et al., 2007; Jiao et al., 2017, Goodreau et al., 2009)

In this section, I investigate the statistical components of social network that are relevant and appropriate for the datasets. The components are degree, number of isolates, number of triangles, number of mutual links, network density, nodal attributes, degree distribution, diameter, average path length, transitivity, and reciprocity. I use these components to evaluate the accuracy of the social network models in the next section.

2.2.1. Degree, In-degree, and Out-degree

In social network analysis, the number of friends of each student are mathematically represented by degree in a network. The degree of a vertex is a number of edges that involve that vertex. Thus, in an undirected network g, a vertex $i's$ degree, written as $d_i(g)$, is $d_i(g) = \#\{j : g_{ij} = 1\}$. In Fig. 1.1, degree of node 7 is 0 as there are 3 edges connect to node 7. I can also calculate the degree of a node by take a sum of its respected column in the matrix, $d_i(g) = \sum g_{ij}$. Node 7 has degree 3 as $\sum g_{7j} = 3$.

In a directed network, the formula $d_i(g) = \#\{j : g_{ij} = 1\}$ is for in-degree. The out-degree of node i is $d_i(g) = \#\{j : g_{ji} = 1\}$. In-degree and out-degree of a node is the number of edges heading towards and away that node respectively. For instance, In Fig 1.2, node 8 has out-degree 2 and out-degree 1. In the matrix, I can calculate the in-degree and out-degree of a node by the equations $d_i(g) = \sum g_{ij}$ and $d_i(g) = \sum g_{ji}$ respectively.

2.2.2. Isolates

In high school social networks, isolates refer to the students no having any friends reported in the dataset. Isolates are the nodes that have degree 0, $d_i(g) = 0$. In Fig. 2.1 below, node 1, 9, and 3 are the isolates.

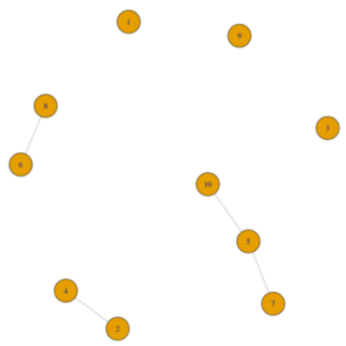

Figure 2.1: Undirected social networks with isolates

2.2.3. Triangles

A triangle is the set of three vertices i, j, and k with three edges: $\{ij, jk, ik\}$. In Fig. 2.2 below, the sets of nodes $\{2, 3, 6\}$ and $\{1, 7, 9\}$ form triangles. The number of triangles helps investigating a group of friends between three students. It is mathematically defined as $\delta(i)$, the number of triangles that node i is involved in a graph g; as each triangle is counted three times, the formula for calculating the number of triangles is as follow:

$$\delta(g) = \frac{1}{3}\sum_{i\in N}\delta(i).$$

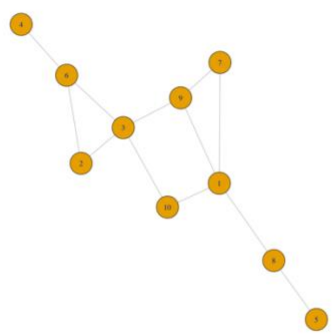

Figure 2.2: Undirected social networks with triangles

2.2.4. Mutual Edges

In a directed social network, I also take the number of mutual edges into consideration. Mutual edges between node i and j are the edges pointing in both directions, $g_{ij} = g_{ji}$. In Fig. 2.3 below, the edge between node 2 and 7 is the mutual edge.

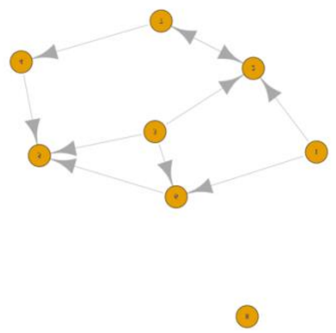

Figure 2.3: Directed social networks with mutual edges

2.2.5. Network Density

When investigating high school social networks, it is important to know if a network is dense or sparse by calculating the network density. The density of a network is the fraction of all possible ties that are present. Density of a network is the average degree divided by $n-1$:

$$D_g = \frac{\overline{d_i(g)}}{n-1} = \frac{\sum g_{ij}}{n(n-1)}$$

with the average degree is the number of edges divided by the number of nodes.

2.2.6. Nodal Attributes

Besides the general components of social networks, I take the sex, grade, and race of the students into consideration. Sex, grade, and race in this case are the nodal attributes of the networks. In this paper, I will compute the number of friendships between students who have the same and different nodal attributes. The figures below represent the nodal attributes in Faux Mesa High School friendship network.

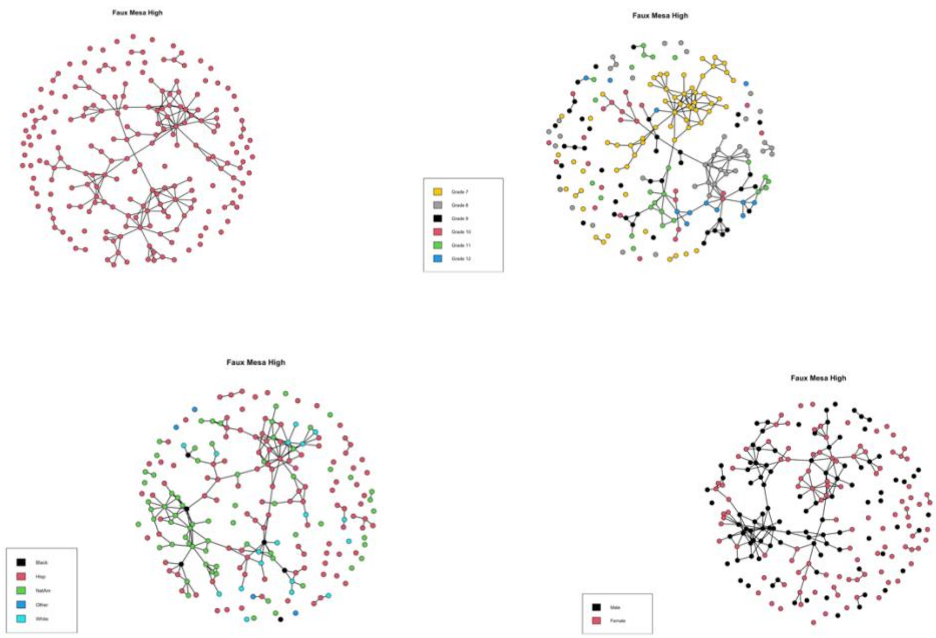

Figure 2.4: Faux Mesa High School Friendship Network attribute-wise

2.2.7. Degree Distribution

As I attempt to investigate how variable is the degree across the nodes of the friendship networks, degree distribution provides greater insight to the network structure than the average degree. Table 2.5 below provides the degrees and their frequency in Mesa friendship network

Degree	0	1	2	3	4	5	6	7	8	9	10
Frequency	57	51	30	28	18	10	2	4	1	2	1

Table 2.5: Degree distribution of Faux Mesa High School friendship network

From the data in the Table 2.5, I plot the degree distribution graph of Mesa friendship network in Fig. 2.6 below.

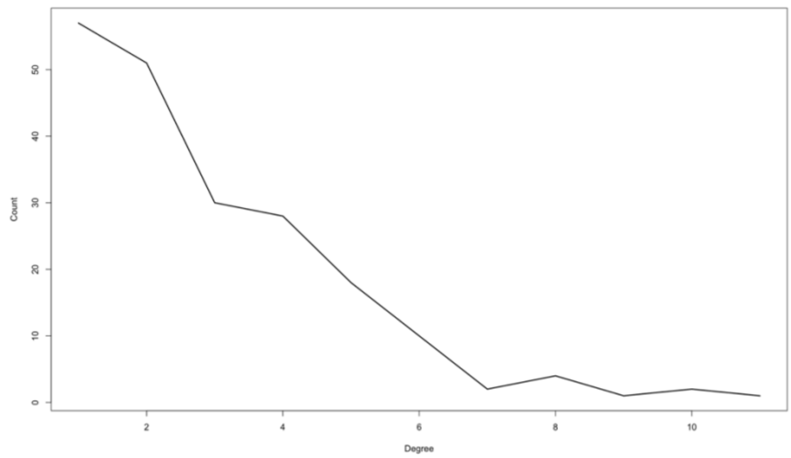

Figure 2.6: Degree distribution of Faux Mesa High School Social Network

2.2.8. Diameters and Average Path Length

Besides degree distribution, diameter and average path length also help us look at the differences in the sets of networks. As the distance between two nodes is the number of edges in the shortest path or geodesic between them, the diameter of a network is the largest distance between any two vertices in the network.

While the diameter only provides the upper limit of the path length, the average path length covers of middle value of the path lengths distribution. Average path length is the average shortest path between two vertices. In this paper, I attempt to compute the diameter and average path length of networks using the computer programming language R.

2.2.9. Transitivity

In friendship networks, it is essential to know to what extent which one of one's friends are friends with each other (Lubbers et al., 2007). This clustered characteristic is measure by taking the value of transitivity. Transitivity refers to the situation where node i has a directed link to j, and j has a directed link to k, and then transitivity checks whether i has a directed link to k. The percentage of times in a network that the answer is “yes” is the value of transitivity represented as follows:

$$T(g) = \frac{\sum_{i;\ j\neq k;\ k\neq j} g_{ij} g_{jk} g_{ik}}{\sum_{i;\ j\neq k;\ k\neq j} g_{ij} g_{jk}}.$$

2.2.10. Reciprocity

Similar to transitivity but dyad-wise, reciprocity is the value of the likelihood of occurring mutual edges between any two nodes. This value is fundamental in learning friendship networks, as reciprocity is the probability of being mutual friends between any two students who are friends. Reciprocity is traditionally determined as follow formula:

$$R(g) = \frac{\sum_{i<j} g_{ij} g_{ji}}{\sum_{i<j} g_{ij} \vee g_{ji}}.$$

3. Modeling High School Social Networks

Based on the theory that the formation of social networks is a random process, which is clearly by chance, in this section, I will attempt to use random graph-based model to model high school friendship networks. The theoretical framework of the models is summarized by Jackson et al. (2010). I will compare the properties including but not limited to degree distribution, diameter, average path length, and the number of isolates and triangles.

3.1. The Erdös–Rényi Model

Erdös–Rényi model are the model based on principle of random graph as follow. Let $G(n, p)$ contain a set of nodes $N = \{1, \dots, n\}$, a link between any two nodes, i and j, be form with probability p, where $0 < p < 1$. The formation of links is independent. Hence, the link formation follows the binomial model. For instance, if $n = 3$, a complete network with three edges forms with probability p^3 $(= p \times p \times p)$, as the probability of forming each link is p. any given network with two links, three nodes form with probability $p \times p \times (1 - p) = p^2(1 - p)$, as one unformed link has a probability $(1 - p)$. Similarly, any given network with one link and three nodes forms with probability $p(1 - p)^2$, and the empty network which has no link and three nodes forms with probability $(1 - p)^3$. Generally, any given networks which has m edges on n nodes have a probability of forming as follow:

$$p^m(1 - p)^{\frac{n(n-1)}{2} - m}.$$

Next I attempt to find degree distribution of the model. As a degree of a node is the number of link that the node has, the total of link that the node can form is the number of the rest of nodes, $(n - 1)$. Consider the degree distribution of a random network is the probability that any given node will have a degree (number of links) d, the degree distribution is the binomial distribution, which has the probability that any given node i forms exactly d link independently is

$$\binom{n-1}{d} p^d (1-p)^{n-1-d}.$$

3.3. Exponential Random Graph Model/p* (ERGM)

Since social networks are interconnected and the actors are interdependent, there is a need to capture the dependencies of the networks. Since probability of link g_{ij} could depend on presence of g_{jk} and g_{ki}, Exponential Random Graph Model (ERGM) has been studied to capture this phenomenon. Probability of network depends on the number of links, according to Erdös–Rényi model, but probability of network also depends on the other properties like the number of isolates and the statistics of nodal attributes. According to the theorem by Hammersley and Clifford (1971), "any network model can be expressed in the exponential family with counts of graph statistics" as follow:

$$P(g) \sim e^{\beta_A A(g)+\beta_B B(g)}.$$

Hence, let see how the theorem works on Erdös–Rényi model. Let n be the number of vertices, p be the probability of a link, $L(g)$ be the number of links in g.

$$P(g) = p^{L(g)}(1-p)^{\frac{n(n-1)}{2}-L(g)}$$

$$P(g) = \left(\frac{p}{1-p}\right)^{L(g)} (1-p)^{\frac{n(n-1)}{2}}$$

$$P(g) = e^{L(g)log\left(\frac{p}{1-p}\right)-\frac{n(n-1)}{2}log\left(\frac{1}{1-p}\right)}$$

$$P(g) = e^{\beta L(g)-C}$$

Where $\beta = log\left(\frac{p}{1-p}\right)$.

3.3. Modeling the High School Friendship Networks

Building from the established principles, I use programming language *R* to generate a huge number of random networks based on the Erdös–Rényi model, and plot the degree distribution with the box plot to test the goodness-of-fit of the model on the original networks.

Mesa *Magnolia*

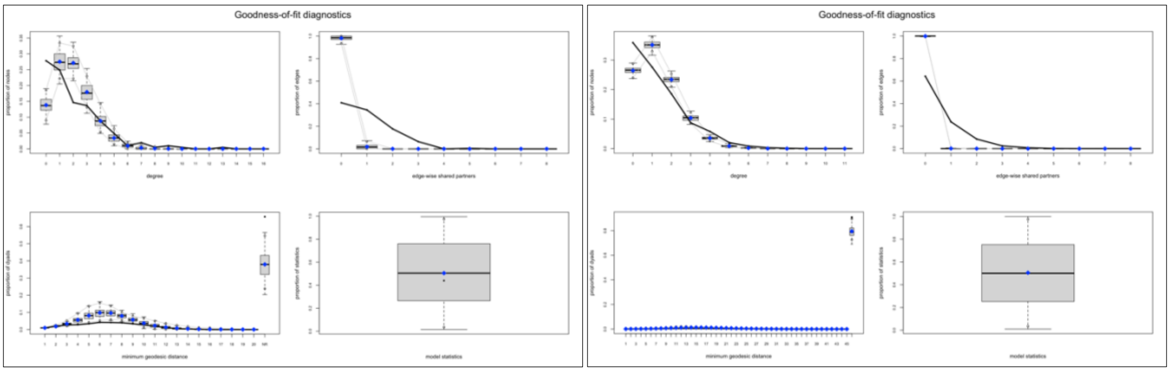

Figure 3.1: Goodness-of-fit of Erdös–Rényi Model on Mesa and Magnolia

Desert *Dixon*

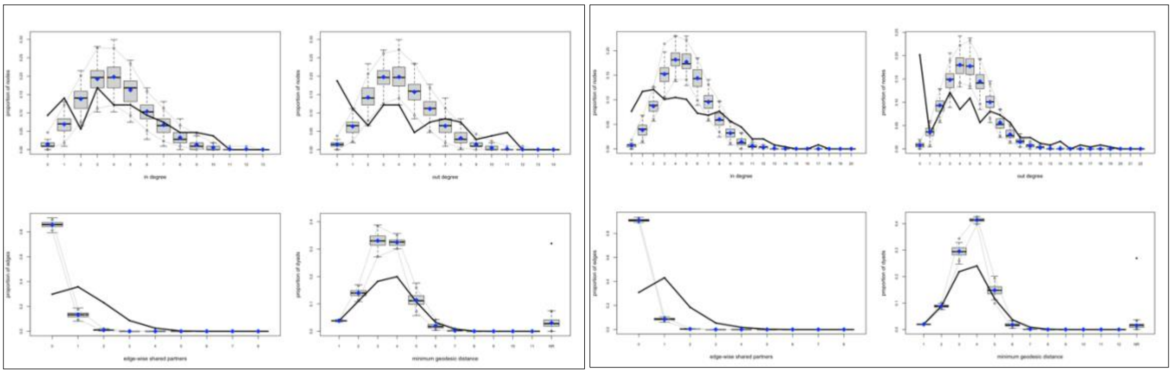

Figure 3.2: Goodness-of-fit of Erdös–Rényi Model on Desert & Dixon

I attempt to investigate two ERGM models, which I call ERGM 1 and ERGM 2. In the ERGM 1 model, besides probability of links, I take probability of isolates into consideration for undirect networks and probability of mutual edges into consideration for directed networks. In ERGM 2 model, I include the properties of the ERGM 1 model, and add the probability related to nodal attributes: the probability of links with opposite sex, same race, or same grade. The process of modeling with ERGM is similar to the one with Erdös–Rényi model.

ERGM 1 *ERGM 2*

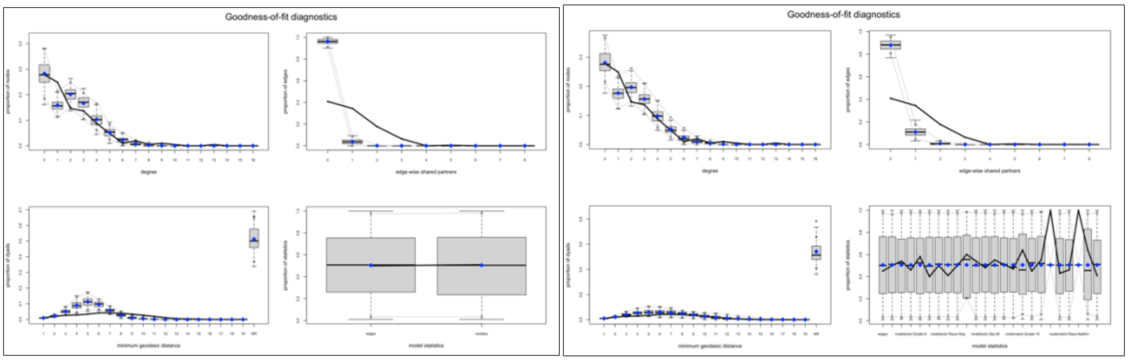

Figure 3.3: Goodness-of-fit of ERGM 1 & 2 of Mesa

ERGM 1 *ERGM 2*

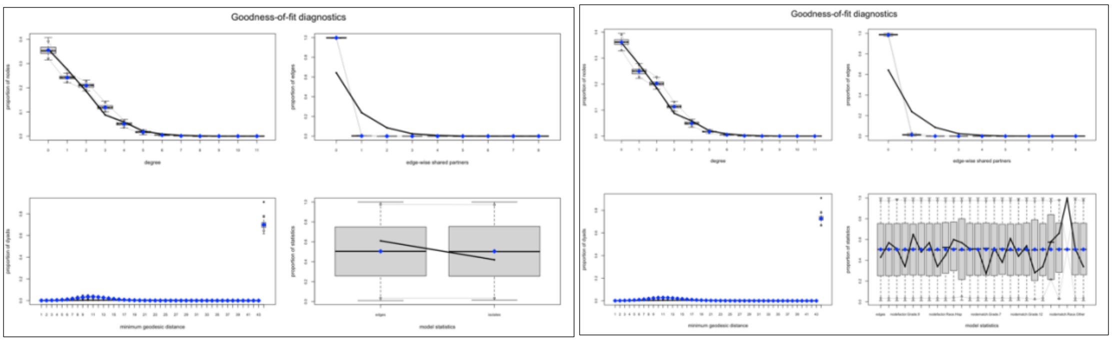

Figure 3.4: Goodness-of-fit of ERGM 1 & 2 of Magnolia

ERGM 1 *ERGM 2*

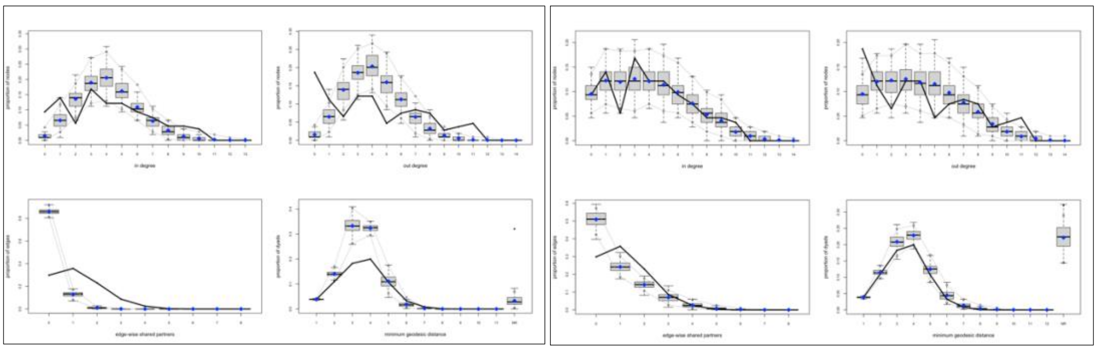

Figure 3.5: Goodness-of-fit of ERGM 1 & 2 of Desert

ERGM 1 *ERGM 2*

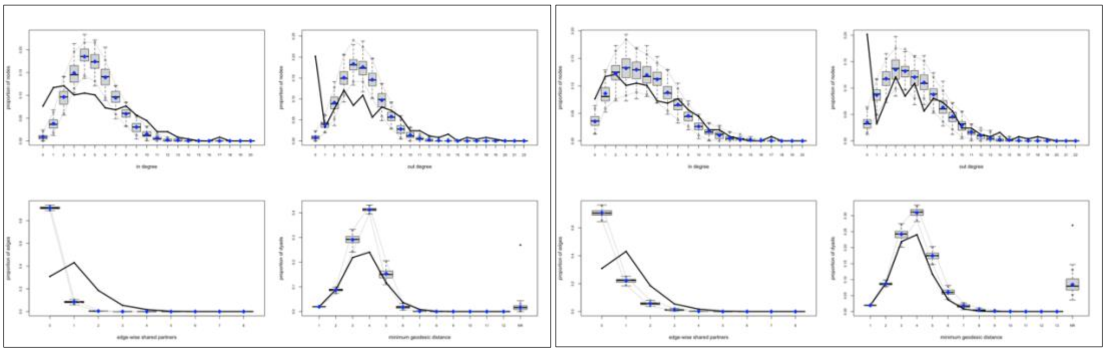

Figure 3.6: Goodness-of-fit of ERGM 1 & 2 of Dixon

3.4. Calculating the Statistical Components of Original and Simulated Social Networks

Using the *statnet*, *ergm*, *igraph*, and *sna* package in *R* which are based on the theoretical framework in section 2.2, I calculate the statistics components of the modelled networks for each original networks.

	Magnolia	Erdös–Rényi	ERGM 1	ERGM 2
# Nodes	1461	1461	1461	1461
# Edges	974	1007	1036	1031
# Isolates	524	389	479	497
# Triangles	169	0	2	3
Average degree	1.333333	1.378508	1.418207	1.411362
Average path length	16.69695	15.36838	11.19163	11.39835
Diameter	40	41	27	28
Density	0.000913242	0.000944184	0.000971375	0.000966686
Transitivity	0.2784185	0	0.003350084	0.00483871
Opposite sex	689	502	514	721
Same grade	820	174	163	845
Same race	787	558	560	832

	Mesa	Erdös–Rényi	ERGM 1	ERGM 2
# Nodes	205	205	205	205
# Edges	203	204	245	234
# Isolates	57	27	48	60
# Triangles	62	2	2	11
Average degree	1.980488	1.990244	2.3902439	2.28292683
Average path length	6.809849	6.932239	4.545239	6.534849
Diameter	16	17	10	15
Density	0.009708274	0.009756098	0.01171688	0.01119082
Transitivity	0.2822458	0.01401869	0.008510638	0.04782609
Opposite sex	132	107	131	160
Same rade	163	36	40	188
Same race	103	77	112	114

Table 3.7: Statistical components of observed and simulated networks of Magnolia and Mesa

	Dixon	Erdös–Rényi	ERGM 1	ERGM 2			Desert	Erdös–Rényi	ERGM 1	ERGM 2
# Nodes	248	248	248	248		# Nodes	107	107	107	107
# Edges	1197	1264	1174	1153		# Edges	439	466	475	453
# Isolates	9	0	0	4		# Isolates	7	0	0	6
# Triangles	1595	180	191	472		# Triangles	677	98	125	732
Average degree	4.826613	5.096774	4.733871	4.649194		Average degree	4.1028037	4.35514019	4.43925234	4.23364486
Average path length	3.680229	3.554912	3.669437	3.882716		Average path length	3.529847	6.932239	4.545239	6.534849
Diameter	9	7	8	10		Diameter	8	17	10	15
Density	0.01954094	0.02063471	0.01916547	0.01882265		Density	0.0387057	0.04108623	0.04187974	0.03994005
Transitivity	0.1566325	0.02065217	0.02457703	0.05716131		Transitivity	0.2305654	0.03713123	0.04567079	0.226461
Reciprocity	0.9752188	0.9596448	0.9751861	0.9748923		Reciprocity	0.9546817	0.9217069	0.9525657	0.9553871
# Mutual edges	219	14	207	192		# Mutual edges	91	11	103	100
Opposite sex	681	622	594	662		Opposite sex	250	237	243	259
Same grade	785	207	183	724		Same grade	321	93	78	352
Same race	912	533	490	889		Same race	367	351	369	382

Table 3.8: Statistical components of observed and simulated networks of Dixon and Desert

4. Evaluation and Conclusion

4.1. The Degree Distribution

Regarding the degree distribution, in Fig. 3.1, the Erdös–Rényi Model describes the undirected networks, Mesa and Magnolia, fairly accurate with some underprediction and overprediction the number of nodes with low degree. Fig. 3.3 and 3.4 show that the ERGMs describe the degree distribution of undirected network better than the Erdös–Rényi Model since the plot fit the original graph better than the one in Fig. 3.1. However, for directed networks, in Fig. both kinds of random graph-based model fail to accurately describe the original networks. The reason might be the chosen statistical components do not fully capture the characteristics of directed networks, and this is the limitation of this paper.

4.2. The Statistical Components

Regarding the statistical components, in Table 3.7 and 3.8, both types of the random graph-based model capture the number of edges and average degree fairly close to the original networks. The Erdös–Rényi Model describes the dimeter and average path length of undirected networks better than the ERGMs. However, the ERGMs capture accurately most of the statistical components of the original graph, except transitivity, the number of triangles, diameter, and average path length.

4.3. Conclusion

The random graph-based model can fit the original social networks to some extent, mainly degree-wise properties like degree distribution, density, average degree, and the number of isolates, and edge-wise properties like the number of edges, average path length, and diameter. Surprisingly, the Exponential Random Graph Models can accurately capture the nodal attributes of the original networks, and better describe the friendship networks in almost every chosen statistical components, except the diameter and average path length, than the Erdös–Rényi model. No model can replicate the clustering of the originals like transitivity and the number of triangles. During the process, I face some difficulties using the number of triangles as a property for random graph-based model since it is hard to count triangles in a complex network. In the future, I will continue work on modeling the clustering of the model by attempting to accurately count and model the number of triangles, and on identifying the statistical components that can capture the characteristics of directed networks.

Bibliography

Butts, Carter T. Sna: Tools for Social Network Analysis (version R package version 2.6), 2020. https://CRAN.R-project.org/package=sna.

Csardi, Gabor, and Tamas Nepusz. "The Igraph Software Package for Complex Network Research." InterJournal Complex Systems (n.d.): 1695.

Erdős, P., and A. Rényi. "On the Evolution of Random Graphs." In Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 17–61, 1960.

Frank, Ove, and David Strauss. "Markov Graphs." Journal of the American Statistical Association 81, no. 395 (September 1, 1986): 832–42. https://doi.org/10.1080/01621459.1986.10478342.

Goodreau, Steven M., James A. Kitts, and Martina Morris. "Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks*." Demography 46, no. 1 (February 1, 2009): 103–25. https://doi.org/10.1353/dem.0.0045.

Hammersley, J. M., and P. Clifford. "Markov Fields on Finite Graphs and Lattices," 1971.

Handcock, Mark S. "Assessing Degeneracy in Statistical Models of Social Networks." In Working Paper No. 39. University of Washington, Seattle: Center for Statistics and the Social Sciences, University of Washington, 2003. https://www.csss.washington.edu/research/working-papers/assessing-degeneracy-statistical-models-social-networks.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. "Statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data." Journal of Statistical Software 24, no. 1 (May 5, 2008): 1–11. https://doi.org/10.18637/jss.v024.i01.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. “Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks.” Journal of Statistical Software 24, no. 1 (May 5, 2008): 1–29. https://doi.org/10.18637/jss.v024.i03.

Hunter, David R., Steven M. Goodreau, and Mark S. Handcock. “Goodness of Fit of Social Network Models.” Journal of the American Statistical Association 103, no. 481 (March 1, 2008): 248–58. https://doi.org/10.1198/016214507000000446.

Jackson, Matthew O. “Social and Economic Networks: Models and Analysis.” Stanford University, Coursera. Accessed January 17, 2021. https://www.coursera.org/learn/social-economic-networks.

Jackson, Matthew O. Social and Economic Networks. United States: Princeton University Press, 2010.

Jackson, Matthew O., and Brian W. Rogers. “Meeting Strangers and Friends of Friends: How Random Are Social Networks?” American Economic Review 97, no. 3 (June 2007): 890–915. https://doi.org/10.1257/aer.97.3.890.

Jiao, Can, Ting Wang, Jianxin Liu, Huanjie Wu, Fang Cui, and Xiaozhe Peng. “Using Exponential Random Graph Models to Analyze the Character of Peer Relationship Networks and Their Effects on the Subjective Well-Being of Adolescents.” Frontiers in Psychology 8 (2017). https://doi.org/10.3389/fpsyg.2017.00583.

Lubbers, Miranda J., and Tom A. B. Snijders. “A Comparison of Various Approaches to the Exponential Random Graph Model: A Reanalysis of 102 Student Networks in School Classes.” Social Networks 29, no. 4 (October 1, 2007): 489–507. https://doi.org/10.1016/j.socnet.2007.03.002.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2020. https://www.R-project.org/.

Resnick, Michael D., Peter S. Bearman, Robert Wm. Blum, Karl E. Bauman, Kathleen M. Harris, Jo Jones, Joyce Tabor, et al. “Protecting Adolescents From Harm: Findings From the National Longitudinal Study on Adolescent Health.” JAMA 278, no. 10 (September 10, 1997): 823–32. https://doi.org/10.1001/jama.1997.03550100049038.

Toivonen, Riitta, Lauri Kovanen, Mikko Kivelä, Jukka-Pekka Onnela, Jari Saramäki, and Kimmo Kaski. “A Comparative Study of Social Network Models: Network Evolution Models and Nodal Attribute Models.” Social Networks 31, no. 4 (October 1, 2009): 240–54. https://doi.org/10.1016/j.socnet.2009.06.004.

Wasserman, Stanley, and Philippa Pattison. “Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs Andp.” Psychometrika 61, no. 3 (September 1, 1996): 401–25. https://doi.org/10.1007/BF02294547.

Appendix

Code for social network visualization and analysis

I use the *statnet*, *ergm*, *sna*, and *igraph* packages of the programming language *R* to generate graphs and figures, and to conduct calculation in this paper.

```
#Loading the installed packages

library(statnet)
library(sna)
library(ergm)
library(igraph)


#Loading the data form the package ergm
set.seed(1)
data("faux.mesa.high")
data("faux.magnolia.high")
data("faux.desert.high")
data("faux.dixon.high")
mesa <- faux.mesa.high
magnolia <- faux.magnolia.high
desert <- faux.desert.high
dixon <- faux.dixon.high

#Converting the network project into graph project
ms <- intergraph::asIgraph(mesa)
mg <- intergraph::asIgraph(magnolia)
ds <- intergraph::asIgraph(desert)
dx <- intergraph::asIgraph(dixon)



#The lines below are the code I use to model and calculate the Mesa

network dataset. The analysis of Magnolia is similar to the Mesa's.


#Plotting the Faux Mesa High friendship network
plot(mesa, main = "Faux Mesa High", vertex.cex = 0.7)
par(mfrow=c(1,1))
plot(mesa, vertex.col='Grade', main = "Faux Mesa High", vertex.cex = 0.7)
legend('bottomleft',fill=7:12,
       legend=paste('Grade',7:12),cex=0.75)

plot(mesa, vertex.col='Race', main = "Faux Mesa High", vertex.cex = 0.7)
legend('bottomleft',fill=1:5,
       legend=c("Black", "Hisp", "NatAm", "Other", "White"),cex=0.75)

plot(mesa, vertex.col='Sex', main = "Faux Mesa High")
```

```
legend('bottomleft',fill=1:5,
       legend=c("Male", "Female"),cex=0.75)



#Erdos-Renyi Model

set.seed(1)
mesa.b <- ergm(mesa ~ edges)
summary(mesa.b)  #Finding coefficients (the probability of t


#Examining the quality of model fit
mesa.b.gof <- gof(mesa.b)
mesa.b.gof
par(mfrow=c(2,2))
plot(mesa.b.gof)


#Calculation & summary of structural effects of observed and simulated net
works of Mesa High
set.seed(1)
mesa.b.sim <- simulate((mesa.b))
summary(mesa ~ edges + triangles +
          density + isolates + meandeg +
          nodematch('Sex') + nodematch('Grade') + nodematch('Race'))
summary(mesa.b.sim ~ edges + triangles +
          density + isolates + meandeg +
          nodematch('Sex') + nodematch('Grade') + nodematch('Race'))


par(mfrow=c(1,2))
plot(mesa, vertex.col='Grade', main = "real mesa")
plot(simulate(fauxmodel.01), vertex.col='Grade', main = "simulated mesa")
legend('bottomleft',fill=7:12,
       legend=paste('Grade',7:12),cex=0.75)

#ERGM 1

set.seed(1)
mesa.e1 <- ergm(mesa ~ edges + isolates)
summary(mesa.e1)

#Examining the quality of model fit
mesa.e1.gof <- gof(mesa.e1)
mesa.e1.gof
par(mfrow=c(2,2))
plot(mesa.e1.gof)

set.seed(1)
mesa.e1.sim <- simulate(mesa.e1)
summary(mesa.e1.sim ~ edges + triangles +
          density + isolates + meandeg +
          nodematch('Sex') + nodematch('Grade') + nodematch('Race'))
```

```
#ERGM 2

set.seed(1)
mesa.e2 <- ergm(mesa ~ edges + isolates + nodefactor("Grade") + nodefactor
("Race") +                  nodefactor("Sex") + nodematch("Grade",diff=TRU
E) +
                  nodematch("Race",diff=TRUE) + nodematch("Sex",diff=FALSE
))
summary(mesa.e2)

mesa.e2.gof <- gof(mesa.e2)
mesa.e2.gof
par(mfrow=c(2,2))
plot(mesa.e2.gof)


set.seed(1)
mesa.e2.sim <- simulate(mesa.e2)
summary(mesa.e2.sim ~ edges + triangles +
          density + isolates + meandeg +
          nodematch('Sex') + nodematch('Grade') + nodematch('Race'))


#Converting the simulated networks into graphs
ms.b <- intergraph::asIgraph(mesa.b.sim)
ms.e1 <- intergraph::asIgraph(mesa.e1.sim)
ms.e2 <- intergraph::asIgraph(mesa.e2.sim)
ms.e3 <- intergraph::asIgraph(mesa.e3.sim)



###########################################################################

#DESERT
#The lines below are the code I use to model and calculate the Desert

network dataset. The analysis of Dixon is similar to the Desert's.



#Erdos-Renyi
set.seed(1)
des.b <- ergm(desert ~ edges)
summary(des.b)  #Finding coefficients

#Examining the quality of model fit
des.b.gof <- gof(des.b)
des.b.gof
par(mfrow=c(2,2))
plot(des.b.gof)


#Calculation & summary of structural effects of observed and simulated net
```

```
works of Desert High
set.seed(1)
des.b.sim <- simulate((des.b))
summary(desert ~ edges + triangles + mutual +
          density + isolates + meandeg +
          nodematch('sex') + nodematch('grade') + nodematch('race'))
summary(des.b.sim ~ edges + triangles + mutual +
          density + isolates + meandeg +
          nodematch('sex') + nodematch('grade') + nodematch('race'))


set.seed(1)
des.e1 <- ergm(desert ~ edges + mutual)
summary(des.e1)


#Examining the quality of model fit
des.e1.gof <- gof(des.e1)
des.e1.gof
par(mfrow=c(2,2))
plot(des.e1.gof)

set.seed(1)
des.e1.sim <- simulate(des.e1)
summary(des.e1.sim ~ edges + triangles + mutual +
          density + isolates + meandeg +
          nodematch('sex') + nodematch('grade') + nodematch('race'))

set.seed(1)
des.e2 <- ergm(desert ~ edges + mutual + nodefactor("grade") + nodefactor(
"race") +
                  nodefactor("sex") + nodematch("grade",diff=TRUE) +
                  nodematch("race",diff=TRUE) + nodematch("sex",diff=FALSE
))
summary(des.e2)

des.e2.gof <- gof(des.e2)
des.e2.gof
par(mfrow=c(2,2))
plot(des.e2.gof)


set.seed(1)
des.e2.sim <- simulate(des.e2)
summary(des.e2.sim ~ edges + triangles + mutual +
          density + isolates + meandeg +
          nodematch('sex') + nodematch('grade') + nodematch('race'))


#Convert the simulated networks into graph project
ds.b <- intergraph::asIgraph(mesa.b.sim)
ds.e1 <- intergraph::asIgraph(mesa.e1.sim)
ds.e2 <- intergraph::asIgraph(mesa.e2.sim)
ds.e3 <- intergraph::asIgraph(mesa.e3.sim)
```