Modeling the RIT Facebook Social Network

David J. LONGO*, Bernard P BROOKS

Rochester Institute of Technology, School of Mathematical Sciences

Rochester, NY 14623-5603

***Corresponding author e-mail** bpbsma@rit.edu

Keywords: Scale-free, assortativity, degree, geodesic distance, modularity, dialogue, rumor, rumour, clustering, co-authorship

The friendship information of Rochester Institute of Technology students was collected through a custom Facebook survey. A social network was then constructed using that friendship data. The RIT Facebook network will provide a target network to which randomly simulated networks will be calibrated in order to repeatedly conduct Monte Carlo experiments on rumor propagation through a social network. Properties of the RIT Facebook network were analyzed with a specific focus on assortative mixing patterns of degree and similarities to the neuroscience co-authorship network studied by Barabàsi. Community structure was found through the Newman modularity maximization algorithm. The flow over the RIT Facebook network of the GBN-Dialogue model of rumor transmission was explored. The RIT Facebook network was found to be best replicated by an evolving neighborhood model.

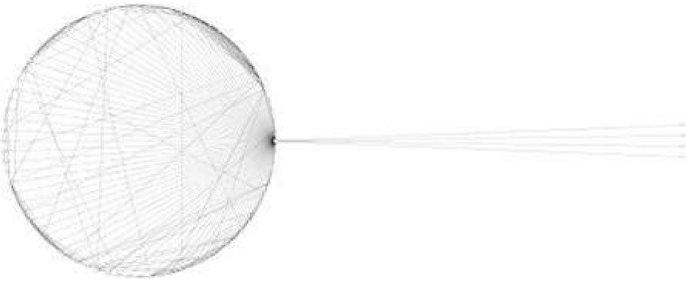

Figure 11. An example of a local network returned to our survey participants. Occasionally a student's network displayed two distinct groups: typically their freshman floor and the remainder of their friends.

Introduction

The most familiar example of a social network is Facebook. A social networking community consists of both the people and the connections between them. These communities provide virtual sanctuaries for people to meet and share information. Facebook gives students a virtual life in which they may make friends, announce parties, and share photos. Acquaintances can be made and quickly turned into real relationships. When mathematicians think about Facebook they see it as a graph or network where the people are the vertices and the friendships between them the edges. The number of friends (number of connections) a person has is called the degree of that person (vertex). When the network consists of only a few people we can visually inspect it and easily determine who seems the most central person, who has the most friends and what is the typical number of friends. The Facebook social network at Rochester Institute of Technology has currently well over 20000 pages and so getting any information by visual inspection is impossible. Thus we need to be able to describe the RIT Facebook network and its members' pages quantitatively. Our goal was to be able to quantitatively describe the RIT Facebook network in order to provide a target for our algorithms whose purpose is to generate realistic artificial social networks. The closer the quantitative measures of the artificial networks are to those of the real network, the more confidence we will have in using the artificial networks in simulations of rumor propagation.

In our study we created a Facebook application which collected user data and presented the user with a visual representation of their local social network as shown in Figure 1. The methodology involved in our data collection is discussed in Section 2.

From postal routes to terrorist cells, network analysis proves a powerful tool. In the world of sociology, social network analysis is providing insight into the structure behind human interaction. The network perspective presents a new method of answering behavioral questions through formalization of relationships and characteristics.

Facebook is an interesting network to study because it should be the synthesis of two completely different concepts. On one hand, one would expect a college campus to exhibit tremendous small-world characteristics. The six degrees of separation is sometimes used as an example of the small world characteristic. Social networks exhibit the small world characteristic if the path of connections between any two members is relatively small, as compared to a random network, given how clustered are the members. Students cluster together by year, major, club, sports team, Greek system, etc. However, Facebook is an online application and thus should exhibit scale-free characteristics. A scale free network has a few very popular people/pages with most people having less than the average number of friends.

There are several established methods of randomly generating a network. A preferential attachment method is one where new people are connected to existing network members with a probability weighted by the number of friends the existing members already have. This results in a rich-get-richer network with a scale-free degree probability distribution which remains constant as the network grows. In contrast, a binomial random network where every possible connection is included with a constant probability, p, exhibits a normal degree probability distribution with mean degree of np. Thus the average number of friends grows as the number of people, n, in the network increases.

In this paper we will attempt to determine which model best describes the RIT Facebook network. Once we have determined which model is the closet fit we can use that model to generate artificial networks with similar properties in order to conduct simulations of rumor flow. Not only would using the real RIT Facebook network be unethical it would not allow for experimental repeatability.

Data Collection

Our data was collected using the Facebook Developer's Application Programming Interface (API). This API gives developers access to a client's local network information. Our Facebook application builds a set of nodes from a client's friends and a set of edges from the friendships between them. For security reasons, developers only have access to a client's immediate friend network. Because of this limited access, we collected multiple samples, which were later overlaid to infer properties of the larger RIT Facebook network.

From the students' perspective, the survey was a simple website that provided them with a picture of their local social network. A student would navigate to a login page where they would login to Facebook. They would then be sent back to the survey page which used the Developer API to collect a list of the student's friends and then check if each pair within the set were connected using the "are_friends" method.

In our experiment, we reached a practical capacity at 5,222 vertices from 139 students. In total the RIT Facebook network contained 16,809 members. Thus, we collected a very strong sample (approximately 1/4 of the total network after correction) and the data collected reliably approximates total network statistics. However, Facebook clients are able to block certain functions of the Developer API. Thus, we found some inconsistencies in our data as certain vertices showed degree 0. Of course this is impossible. If a student has 100 friends, then our survey will get 100 vertices. These 100 vertices may be replications of already collected data, but each will have at least one in-edge as they are friends with our current student. The vertices with zero degree can be attributed to Facebook's security functions. Thus, if we discard the inconsistent vertices, the result is a network of order 4160 with 46936 edges.

RIT Facebook Network Characteristics

Degree Distribution

Social networks are made up of sets of vertices which typically represent people, and edges which represent relationships. The degree of a vertex, $d(n_i)$, is the number of edges incident with vertex n_i, or in other words, the number of friends on a person's Facebook account. The *degree distribution* of a network is the probability distribution of vertices' degrees. By looking at a network's degree distribution, we can infer the network structure. For example, a binomial random network exhibits a normal degree distribution centered at its mean degree, *np* (Erdos and Renyi, 1959). A scale-free network follows a power-law degree distribution. *Scale-free networks* are networks whose structure and dynamics are independent of their size. The models we will consider for comparison which demonstrates similar properties to that of the Facebook network, are the Barabasi-Albert preferential attachment model and the neuroscience co-authorship network (Freeman, 1977; Barbasi, Albert and Jeong, 2000; Barbasi et al., 2000). A power-law distribution is given by $P(k)=k^{-\gamma}$ where k is the degree and γ is a positive constant.

Scale-free networks are not normally seen in real-life friend networks; instead one sees collaboration networks. Examples of collaboration networks include scientific collaboration networks (Barbasi et al., 2000) and the World Wide Web. The reason scale-free networks do not typically represent real-life friendship networks is that there is a cost to having a friend. Let's say Joe has ten hours of leisure time each week to spend with his friends. To be a good friend, Joe has to spend at least one hour with each friend. Thus, Joe can only have at most ten real friends. A scale-free network has a few people with a huge number of friends, hubs, and most people with only relatively few friends, spokes. Typical scale-free structures such as hubs most likely do not exist in real friendship networks because there is a practical limit as to the maximum number of real-life friends you can have. Collaboration networks typically exhibit an exponential degree distribution.

In the case of the degree distribution of the RIT Facebook network, an exponential fit worked best with an R^2 value of 0.97. The same result was seen in the neuroscience co-authorship network. Equation 3.1 shows the exponential fit of $P(k)$ vs. k in the RIT Facebook network. The exponential shape of the degree distribution demonstrates the ineffectiveness of preferential attachment as the network ages as discussed in the neighborhood evolving network model (Cao et al., 2006), thus further showing a neutral assortativity as explained later in the paper.

$$P(k) = 0.1909e^{-0.402(k)} \qquad (3.1)$$

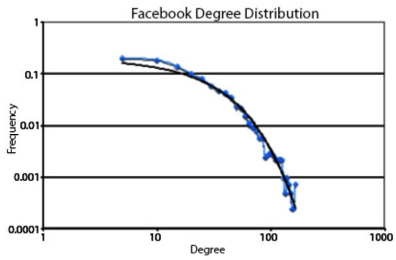

Figure 2. The degree distribution of the RIT Facebook network shows an exponential relationship as demonstrated by the best fit line shown in black. This data was analyzed using Ucinet (Borgatti, Everett and Freeman, 2002).

Therefore in order to simulate the RIT Facebook network we might use a neighborhood evolving network method (Cao et al., 2006). These methods of randomly generating a social network are similar to the preferential attachments but the probabilities of a new joining person being connected to an already connected person are not just proportional to the existing person's degree but also to which neighbors they are already have. So it seems that when a new student joins the RIT Facebook network they are more likely to connect to people that are already popular (high degree vertices) as would be the case in a preferential attachment method but they also tend to favor connecting to people with which they already share mutual friends. We infer this from the degree distribution of the RIT Facebook network matching those simulated by a neighborhood evolving method.

Degree distribution is one quantitative metric but there are several others that must also be considered in order to better understand the RIT Facebook network.

Clustering Coefficient

The *clustering coefficient* is an important measure of "cliquishness," which was first introduced by Watts and Strogatz (1988) in their paper "Collective Dynamics of 'Small-World' Networks." The clustering coefficient tells us how many of our friends are themselves friends with each other, which was one of the major focuses of our Facebook study. To calculate a person's clustering coefficient we count the number of relationships in their set of friends and divide by the number of total possible relationships. If the person has *n* friends then the total number of possible relationships that could exist amongst their friends is *n(n-1)/2*. For example if I have 4 friends but only 2 of them know each other my clustering coefficient is C_i = 1/3. The global clustering coefficient is simply the average of clustering coefficients of the people in the network.

The global clustering coefficient for a network generated with the Barabasi-Albert (BA) model scales with network size according to $C \sim N^{0.75}$ where N is the number of people in the network. In the RIT network we found a much higher global clustering coefficient with C = 0.534 than that predicted (C = 0.0019) if the RIT Facebook network had been created using a BA type preferential attachment model. However, this clustering coefficient of C = 0.534 was similarly found in the neuroscience - network studied by Barabasi et al. (2000) further implying that the RIT Facebook network was created using a neighborhood evolving method.

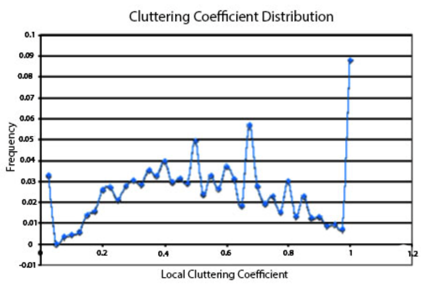

Figure 3. The plot shows the probability distribution of the local clustering coefficients of the RIT Facebook network.

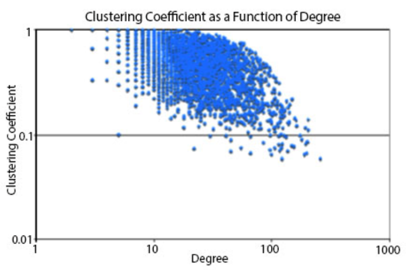

Figure 4. If we plot clustering coefficient as a function of degree we see a plot similar to the co-authorship network in the field of neuroscience (assortative) as discussed by Barabasi et al. (2000).

Geodesic Distance

The geodesic distance between two vertices is the length of the shortest path between them. The characteristic path length is the average of all geodesic distances in the network. Stanley Milgram conducted an experiment, called the "small-world experiment," in which he sent letters to starter persons who were asked to pass them along to a remote target person across the country. Out of 96 starters, 18 letters reached their final target and Milgram established that "we live in a small world, and are only six steps apart from each other" (Csermely, 2006). That is the origin of the oft quoted "six degrees of separation". Watts-Strogatz formally proved that there is a characteristic path length of around six between any two people in the world and defined the Watts-Strogatz (WS) small-world network model in which a 2-regular lattice is rewired with probability *p* and exhibits this small-world property. The characteristic path length of the BA scale-free network model increases approximately logarithmically with system size, N, according to the relation $l \sim ln(N)/ln(ln(N))$ (Barbasi and Albert, 2002).

For a BA network with size N = 4160, the characteristic path length l should be approximately 3.930. At RIT, we found a characteristic path length of 3.373. The low characteristic path length also indicates the small-world phenomenon in the network. Thus people at RIT are in general more closely connected than they would be if the RIT Facebook network had been created solely through a preferential attachment method.

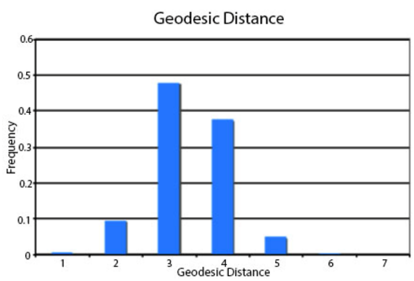

Figure 5. The frequency distribution of geodesic distance shows an average of 3.373. The majority of shortest paths between people are less than 4 steps with 3 steps being the most common.

Assortativity

Assortative mixing in networks is a measure of the "tendency for people (vertices) in networks to be connected to other people (vertices) that are like them in some way" (Newman, 2003). For example, humans tend to choose sexual partners of the same age or ethnicity. This mixing characteristic influences a network's ability to spread rumors and disease effectively. Disassortative networks (such as certain scale-free networks) will spread rumours or disease more quickly as they contain "hubs" that reach far into the network (Goh et al., 2003). We measured assortativity with the assortativity coefficient which is a metric of the tendency for popular people to be connected with popular people and less popular people to be connected with less popular people. The assortativity coefficient is simply the Pearson correlation coefficient of degree between pairs of linked vertices. The Facebook network is assortatively neutral with an assortative mixing coefficient of 0.01781. This neutrality implies that people at RIT are connecting together in clusters where popularity is not an important group criterion. The neutrality of the RIT Facebook network further supports its similarities to the co-authorship networks previously mentioned.

Algorithms

Modularity Maximization

The neutral assortativity indicates that the number of friends is not driving the clustering of people in the RIT Facebook network. There are existing algorithms that can be used to divide up the network into communities based on the network structure. M.E.J. Newman created an algorithm that works well in defining community structure without initial constraints for group size. The *modularity* is, "up to a multiplicative constant, the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random"(Newman, 2006). A community should have many relationships between its members, much more than the network average and certainly more than you would expect to find by chance. Thus, the algorithm divides a network into two communities by maximizing the difference between how many friendships between community members actually exist as compared to how many you would expect by chance; that is, if the people were connected without regard to group membership. The adjacency matrix, *A*, is a matrix with elements of 1 or 0 coding whether a relationship exists between any two people in the network.

Given a vertex with degree k_i and a vertex with degree k_j in a network with mean degree *m* the probability of there being a connection between the two vertices by chance is $p_{ij} = (k_i k_j)/2m$. The algorithm used to divide the network into two communities maximizes the difference matrix = $A - [p_{ij}]$. The sign of the elements of the leading eigenvector of *B* are used to assign each person into one of the two groups. An iteration of the algorithm splits a network into two communities. The algorithm is then applied to each of the resulting sub-networks stopping when the result is a dominant eigenvector whose elements are all the same sign.

The algorithm split the RIT Facebook network into seven groups with the following sizes: 1751, 942, 675, 330, 266, 128, and 68. By inspection of the larger network, we hypothesize that the group of size 266 consists of gay males at RIT. By further inspection, we also hypothesize that the group of size 68 consists of members of the RIT Singers, which is a choral group on campus.

GBN-Dialogue Model of Rumor Propagation

Mathematical models of rumor propagation are rarely based on sociological and psychological research of real world transmission behavior. As a result, these models tend not to incorporate the connections between people in social networks. The GBN-Dialogue model of rumor transmission was developed based on empirical rumor transmission data (Brooks, DiFonzo and Ross, 2013). The GBN-Dialogue model is based on three factors of transmission: group membership, belief in the rumor, and the novelty of the rumor. This agent-based model is used to simulate the spread of an out-group negative rumor over the RIT Facebook network.

Because the simulated rumor is derogatory about the out-group, the RIT Facebook network was split into two groups using the Newman modularity maximization algorithm. The two resulting groups (in-group and out-group) are homogeneous with respect to a person's anxiety, belief, and novelty. The logic here is that a person in the in-group will more readily believe the derogatory rumor about the out-group, be more anxious to talk about it, and will be less likely to get tired of the rumor. People in the out-group will be less likely to believe the rumor and will actually spread anti-rumors, that is, rebut the derogatory claims.

After the people in the RIT Facebook network were categorized as either in-group or out-group members one member of the in-group was chosen as the initial person to first spread the rumor. Each iteration of the GBN-Dialogue model begins by randomly choosing an edge from the set of all relationships in the network. The probability of rumor transmission is then calculated as a function of the two interlocutors' group statuses, their belief in the rumor, their perception of the novelty of the rumor, and their anxiety and uncertainty about the situation (Brooks, DiFonzo and Ross, 2013). If rumor is transmitted, the belief levels of both interlocutors are altered. Hearing a rebuttal lowers your belief in the rumor whereas hearing agreement increases the belief.

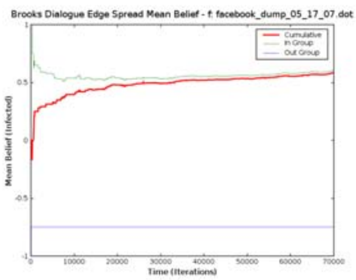

Figure 6. The plot shows the in-group's mean belief and the mean belief of the entire network as the simulated rumor spreads through the RIT Facebook network. The trace for the out-group shows that in this simulation they never believe the derogatory rumor about themselves.

The collection of the RIT Facebook network provides a real world medium for algorithm testing. The network was the first real world example test of the GBN-Dialogue model of rumor transmission. Previous tests were done on artificially created classic networks such random, ribbon, and WS small-world (DiFonzo et al., 2013). More realistic artificial networks can now be calibrated based on the above real network statistics of the RIT Facebook network.

As well, one can test whether a minority group embedded in a simulated RIT Facebook network can resist a minority derogatory rumor. The more the minority is integrated into the network as a whole, the less resistance it has to the minority derogatory rumor (Brooks, 2013). The degree of integration can be quantified by calculating the Minority Integration Metric (MIM); if the MIM is above threshold, the rumor will propagate through the minority sub-network.

Conclusion

By inspection of degrees, clustering coefficients, degree and assortativity coefficients, and geodesic distances, we can confidently conclude that the Facebook network has a strong correlation to the neuroscience collaboration network. The striking similarity might be attributed to the apparent necessity for collaboration in academia at all levels. RIT students connect through Facebook not only due to conventional social motivations but also for academic collaboration. Along the same lines, the RIT Facebook network exhibits a neutral assortativity coefficient due to the aforementioned mixture of connection methods. Socially, a popular person may fraternize solely with other popular people. Academically, however, that same popular person may come in contact with less popular people and thus break down the invisible social barrier between them. The weak connections formed by collaboration are beneficial to marketing firms as they allow information to be passed from group to group with ease. Marketers should seek out networks, such as Facebook, with neutral assortativity.

In order to have experimental repeatability for rumour propagation experiments the neighbourhood evolving method of randomly generating an artificial network seems to produce the most realistic reproduction of the RIT Facebook network. That is the method that will be used to create the artificial test networks over which simulated rumors will be propagated by means of the GBN-Dialogue model.

Acknowledgements

This work is based upon work supported by the National Science Foundation under Award No. BCS-0527371

References

Barabàsi, A., Albert, R., and Jeong, H. (2000) Scale-free characteristics of random networks: The topology of the World Wide Web. *Physica*, A, 69–77.

Barabasi, A.L. and Albert, R. (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics,* **74**, 47–97.

Barabasi, A.L. et al. (2002) Evolution of the social network of scientific collaborations. *Physica*, A **311**, 590–614.

Borgatti, S.P., Everett, M.G., and Freeman, L.C. (2002) *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Brooks, B.P., DiFonzo, N., and Ross, D.S. (2013) GBN-Dialogue model of outgroup-negative rumor transmission: Group membership, belief, and novelty. *Nonlinear Dynamics, Psychology and Life Sciences*, **17**(2), 269-293.

Cao, Y.J., Wang, G.Z., Jiang, Q.Y., and Han, Z.X. (2006) A neighbourhood evolving network model. *Physics Letters,* A **349**, 462–466.

Csermely, P. (2006) Small-Worldness. *Weak Links*, Springer, New York, 7–13.

DiFonzo, N., Bourgeois, M.J., Suls, J.M., Homan, C., Stupak, N., Brooks, B., Ross, D. S., & Bordia, P. (2013) Rumor clustering, consensus, and polarization: Dynamic social impact and self-organization of hearsay. *Journal of Experimental Social Psychology*, **49**(3), 378-399.

Erdos, P. and Renyi, A. (1959) On random graphs. *Publ. Math. Debrecen,* **6**, 290–297.

Freeman, L.C. (1977) A set of measures of centrality based on betweenness. *Sociometry,* **40** , 35.

Goh, K.I., Oh, E., Kahng, B., and Kim, D. (2003) Betweenness centrality correlation in social networks. *Physical Review,* **67**, 1-4.

Newman, M.E. (2003) Mixing patterns in networks. *Physical Review*, **E 67**, 1–13.

Newman, M.E. (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences,* **103**, 8577–8582.

Watts, D.J. and Strogatz, S.H. (1998) Collective-Dynamics of ‘small-world’ networks. *Nature,* **393**, 440-442.

Brooks, B.P. (2013) Rumour propagation on social networks as a function of diversity, advanced dynamic modeling of economic and social systems. Studies in Computational Intelligence, 448, 49-60.