Homophily, Popularity and Randomness: Modelling Growth of Online Social Network

Syed Muhammad Ali Abbas
Centre for Policy Modelling,
Manchester Metropolitan University
Business School
All Saints Campus, Oxford Road
M15 6BH, Manchester, UK
ali@cfpm.org

ABSTRACT
This work investigates how local preferences, social structural constraints and randomness might affect the development of the friendship network in Facebook. We do this by analyzing a snapshot Facebook dataset of Princeton University's students, and by building an agent-based simulation for comparison. Several different, but plausible, processes of friendship network development are proposed in which the structural information of the growing network and the student preferences are taken into account and then compared with the data. 'Network formation based on personal preference and social structure with some randomness' matches the data best, and is thus the preferred hypothesis for the way that students add "friends" on Facebook.

Categories and Subject Descriptors
J.4 [Computer Applications]: Social and Behavioral Sciences–Sociology; C.2.4 [Distributed Systems]: Distributed Applications

Keywords
Facebook, Community Structure, Agent-Based Modelling, Social Network Analysis (SNA).

1. INTRODUCTION
Since the advent of online Social Networking Systems (SNS), the internet has become part of everyone's everyday life. A huge number of people have a presence over the internet via a "profile", which is a publicly articulated webpage describing a virtual self. According to [1], there are now over 2.8 billion social media profiles, representing around half of all internet users worldwide. Online SNS present themselves as a platform for such profiles. Not only can people present themselves, but can present their social network as well. Since 2004, when Facebook, currently the most popular SNS, came into being, there has been a lot of research on how people form friendships and interact over it, e.g. [2–4]. Facebook alone has over 910 Million monthly users to its credit [5].

The magnitude of the data present in the online SNS is enormous, and presents itself as a rich source of social information for analysis. According to studies, most of the online social networks act as a representation of the offline, or real social networks [6], [7]. So it could be assumed as an approximation or a proxy of a real world social network. Not only does an SNS capture the social network, but also the activity between users. Mainly due to privacy concerns and also due to its vast commercial value, this data even by the research community is quite difficult to acquire. So we are left with either a snapshot with limited information, or an activity log without any social network. A huge data set of longitudinal nature of Facebook has been collected, but is available with a limited access [2]. The aim of this paper is to reconstruct the development of the social network with the help of an agent-based methodology, so that a possible history of the social network and an understanding of it could be developed.

A lot of social network based models have been proposed. From a general but realistic social network (e.g. see [8], [9]) to a data-driven students' social network [10], but they do not address how such a network might develop within an online environment. This paper attempts to address this concern. First, we simulate some possible strategies of how students meet and develop their social network. Then we compare the obtained results with the underlying dataset we have used and in this way are able to make some inferences as to the probable strategies that the students used.

The main motivation behind this paper is to understand, realize and explain how students interact in their social life and then develop social links with each other. Whether the inherent attributes such as dormitory or the network position plays an important role in friendship development? We relied on agent-based simulation which helps develop realistic environment of students' interaction. It is more of an explanatory model based on a few hypotheses that how students might have developed social links (friends) over a period of time. Students meet and interact not only in their university, such as lecture halls, but also outside it, such as in parties and the dormitories they live in. Keeping such a social life in mind, we drew interaction strategies that are present in students' lives. SNS, such as Facebook, does allow sharing virtual-self with details such as gender, age and school affiliation, but this information alone is not enough to identify the exogenous social settings which resulted in two people's friendship. A mechanism is required to cater for both endogenous and exogenous environment. From the perspective of SNS, such as Facebook, link prediction and recommendation, as is commonly known in computer science, is quite a challenging, and lucrative feature. This model tries to explain through what social interactions people might develop their social network. From information sharing to future business partnerships, the relationships which are developed at the stage of University have

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013),* Ito, Jonker, Gini, and Shehory (eds.), May 6–10, 2013, Saint Paul, Minnesota, USA.

significant impact on one's life [11]. Also, the friendships made at the earlier university years are crucial for not only staying in the university [12] but it remains the most vital place of meeting up for people, where friendships are developed in Facebook [13]. A large analysis of a meme was studied in [13], and was found that the majority of people on Facebook had met their friends in school settings, such as same class/grade. This study also tells that this majority of meeting up for friendship is significant for all ages; and this behaviour goes for all ages regardless of their gender.

The paper is divided in different sections. In Section 2, we define the reference data on which our agent-based model is based – its characteristics and network structure. After that, in Section 3, we define our model and the strategies of interaction it offers. Simulation results and their comparison with the dataset are presented in Section 4. Related work is summarized in Section 5. At the end, in Section 6, we summarize our findings and present the future outlook of our research by concluding the paper.

2. REFERENCE DATASET

The underlying anonymous dataset of Facebook includes both the attributes and social structure for *6575* student of Princeton University. In total there are *293307* links – averaging to *89.2* friends. Each person has four attributes, which are: major course of study (major); their place of living (dorm); year they joined the university, and their high school information. As for the spread of each attributes and their missing values, we have summarized it in Table 1.

Table 1. Attribute spread

Attributes	Dorm	Major	Year	High School
Missing (%)	33.76	24.86	11.77	20.7
Unique	57	41	26	2235
Average	115.72	160.88	244.30	2.95
St. Dev.	293.06	268.68	399.13	29.21

Since it is a relatively bigger university as compared to Caltech covered in [14], we see a very diverse population when we see the number of various high schools. Missing information in the dataset has been coded by *0*. We have dealt it carefully in our model. As for the network structure, this is how it looks like:

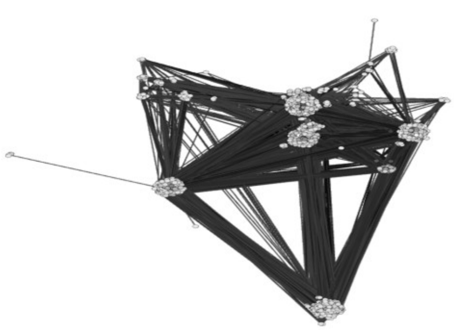

Figure 1. Princeton Social Graph

We can clearly see groups/communities in the network structure.

3. MODEL OUTLINE

In order to understand the dynamics of this social network, we simulate it using an agent-based simulation. The main aim of the paper is to understand the interplay of social processes and their impact on the network structure as a whole. Thus the key focus is on analyzing how students interact and build their social network over time. This work is an extension of the model presented in [14], and is being applied to a larger dataset. We can then see which strategy of interaction seems to produce the best representation of a social network as judged by a comparison with the reference dataset. In this section, the term *agent* will be used to refer to a student. In order to explain our model, we have used a standard protocol, known as Overview, Design concepts, and Details (ODD) [15]. It is specifically developed to describe the details of an agent or an individual based model.

3.1 Purpose

The purpose of this agent-based model is to understand and explore friendship development process in Facebook by the help of four interaction strategies. The scenarios of interactions have been carefully drawn from real life interactions of student; details of which follow in *Process Overview and Scheduling*. The model is used as a *search* mechanism to identify which interaction strategy captures the best representation of a Facebook social network of Princeton University.

Table 2. State Variables and Scales

Variable name	Brief description
Total Links	Total number of links when the simulation is to stop – the total number of links in the reference dataset (16656)
Random Seed	Dynamic random seed for simulation
DormPref.	Preference of inter-dorm homophily
MajorPref.	Preference of inter-major homophily
YearPref	Preference of inter-year homophily
HighSchoolPref	Preference of inter-high School homophily
SimulationMode	Identifying what interaction mode between 1-4 is being set (see Process Overview and Scheduling for details)
ClusterCoeff.	At each simulation step, the overall cluster coefficient (number of triangles) is calculated.
Mean and Standard Dev.	Mean and Standard Deviation in the number of links of each agent is calculated and then recorded in a file.

Table 3. Agent Level Variables

Variable name	Brief description
ID	Identity of Agents – an auto increasing number starting from 1.
Dorm	The dormitory/hostel of an agent – an integer number
Year	The year of joining the university of an agent
Major	The major course of study of an agent
High School	The high school number of an agent
Friends Count	Total friends count

3.2 Process Overview and Scheduling

We have devised four different plausible *strategies* for link formation – each involves matching agents using their attributes, but in different ways. Personal preference is not taken into account when there are missing values for all the four attributes. Hence, in this case, we totally neglect the preference of both the source and the target agent. All of the four interaction strategies use the attribute values defined in Table 6 (see Section 3.4 for more details).

3.2.1 Strategy 1 - Random Strategy

Each source agent selects a randomly chosen target agent after every time step or simulation tick. The target agent is selected using a uniform probability distribution. After the selection, the source agent determines if the target agent satisfies its personal preference – the source also satisfies target's preference. If it does, an undirected link is created among them, which shows that they are friends.

3.2.2 Strategy 2 - Friend of a Friend Strategy

In this strategy, there are two phases for each agent. In the first phase, all agents, on their own, make only limited random friends selected in a uniform distribution. This should satisfy both the source and target agents' preferences. If these are not satisfied, they do not form a link. In other words, the random strategy is employed by ever agent. After this first phase, personal preferences are not taken into account. From then on, in the second phase, new friends are selected in a "friends-of-friends" manner. During this phase, starting from the first friend of a friend – whose degree is consider as the reference point, in chronological order, we search its friends (FOAF) and continue searching till we find a suitable agent. As soon as we find an available FOAF which has a greater degree than the reference FOAF – showing the popularity, we select it and then form a friendship link between the two.

3.2.3 Strategy 3 - Party Strategy

In this strategy the personal preferences are also not taken into account. All students arrange a small party which is held on a regular basis. The number of participants in a party is 100. The selection of the party participants is totally independent and unbiased towards any attribute. At each party, a maximum of *300* new (random) friendships are made. Due to the random selection of party participants, there is a chance of selecting nodes which are already connected to each other. In that case, no new link is established.

3.2.4 Strategy 4 - Hybrid Strategy

This strategy is a combination of the above three strategies. At every simulation time step, a simulation strategy between random and FOAF is chosen on a uniform basis. In order not to overwhelm the randomness, the party strategy is run in every 20^{th} time step. Also, unlike the original model [14], the party mode is executed on a local level, and not on a global level.

For the *random strategy*, all agents have a predefined preference for each of the four attributes we have selected which is known as *"Personal Preference"*. The idea has been inspired from homophily – the love of the similar [16]. It is a probabilistic match of attributes between the source and the target agents. We have shown the illustration in Table 4 for the year attribute. A *chance* out of *100* is randomly selected in a uniform fashion – line 3. If it is under the predefined preference value (80 in case of year preference) and the attribute values of both the source and target agents are known and match with each other, then the year preference is satisfied; and we set the year flag to true – line 6. Also, if the *chance* is greater than the preference value, it is satisfied as well – line 7. We repeat the same process for the remaining attributes. If all the four attributes' conditions are satisfied, we make a friendship link between the source and the target agents – line 11.

Table 4. Algorithm to calculate *"Personal Preference"*

```
1.  Agent Source = getSourceAgent(), Agent Target = getTargetAgent()
2.  Integer YP = getYPValue() // Get Year Pref. value which is fixed as 80
3.  Boolean sameYear = False, Integer chance = get_random_integer(100)
4.  IF (chance < DP){ // 0<chance<=YP
5.  IF (Source.getYear() == Target.getYear()) AND
6.  (Source.getYear() != 0 And Target.getYear() != 0)){
7.  sameYear = True }
8.  }ELSE{ sameYear = True }
9.  ...//repeat the same evaluation for the rest of the attributes - Dorm,Major
    etc
10. IF (sameYear AND sameMajor AND sameDorm AND sameHighSchool)
    // If all conditions satisfy
11. form_a_link(Source, Target) //create a friendship link
```

3.3 Design Concepts

3.3.1 Emergence

The agents are designed to develop their social network by interaction and then selecting the right target as their friend. This interaction results in the emergence of closely linked groups, or communities in other words.

3.3.2 Adaptation

There isn't any learning process designed in the model.

3.3.3 Fitness

Since there are four modes of interaction, there are four fitness algorithms for agents, depending on the mode of interaction being used.

- **Random Mode:** On interaction, the preferences of both the source and the target have to match, in order to form a link.
- **FOAF mode:** If the source has met the target agent via friend (target is a FOAF), then no fitness is required to develop a link between the two. If, however, the target agent is randomly met, then the preferences of both of them have to be satisfied.
- **Party Mode:** No fitness is required here. The links are made randomly among the party participants.

- **Hybrid Mode:** Depending on the mode being run (random, FOAF, and party), the appropriate fitness mechanism applies and then satisfied.

3.3.4 Prediction

Agents do not have any predicting power. They make their decision based on their interactions with others – on the available information.

3.3.5 Sensing

Based on the employed interaction strategy, the agents sense each other, and then evaluate their compatibility.

3.3.6 Interaction

In this section, we discuss how the agents might interact with each other, in terms of making friends in real life. It is assumed that, by and large, these real life social links will then be duplicated within Facebook. We do not claim that we present an exhaustive list of possible strategies; rather the idea is to explore *some* plausible ways that depend on the micro-level preference of agents and then evaluate them. There are four modes, so there are four interaction strategies as well.

- **Random Mode:** Every agent comes across a random agent
- **FOAF mode:** Depending on the number of links an agent has made. If it is lower than *30,* then she will meet a random agent; otherwise she will interact with a FOAF.
- **Party Mode:** In this interaction strategy, agents interact with party attendees
- **Hybrid Mode:** Depending on the mode being run (random, FOAF, and party), the appropriate interaction strategy applies here.

3.4 Submodels

In order to identify the significance of attributes of the four attributes we have considered in social network development, we relied on affinity [17] to guide us. It measures the ratio of the fraction of links between attribute-sharing nodes, relative to what would be expected if attributes were random. It ranges from 0 to infinity. Values greater than 1 indicate positive correlation; whereas values less than but greater than 0 have negative correlation. For an attribute *a*, such as dormitory, we first calculate the fraction of links having the same dormitory, for instance. It is represented by:

$$S_a = \frac{|\{(i,j) \ni E : s.t. a_i = a_j\}|}{(|E|)}$$

where a_i represents the value *a* for a node *i.* In other words, we are identifying the total number of matched nodes with the same attribute values for an attribute *a. E* represents total number of links. And then we calculate E_a which represents the expected value when attributes are randomly assigned. It is calculated by:

$$E_a = \frac{\sum_{i=0}^{k} T_i(T_i - 1)}{|U|(|U|-1)}$$

,where T_i represents the number of nodes with each of the possible *k* attribute values and U is the sum of all T_i nodes, i.e., $U = \sum_{i=0}^{k} T_i$. The ratio of the two is known as *affinity*: $A_a = \frac{S_a}{E_a}$ [17]. Here are the affinity measures of the four attributes in Table 5:

Table 5. Affinity values of the four attributes for all the strategies of interactions

Dorm Affinity	Major Affinity	Year Affinity	High School Affinity
1.48	1.32	4.07	0.89

We see that year is the most important attribute here. This result matches previously published work [18]. Hence, we used this measure as an insight to have a parameter sweep of just the year value. Apart from high school attribute, rest is positively correlated. Further on this will be shown in Section 3.1. Each agent is initialized with the four attributes (major, dorm etc.) of a corresponding individual recorded in the Princeton data set. We have just used these four attributes because of the conformity in the earlier studies done on students. And also, we found them, using the affinity measure, of the utmost importance. The values for each of the four attributes can be seen in Table 6. These values have been found to be the best fitted values when compared with the reference dataset.

Table 6. Values of the four attributes for all the strategies of interactions

Dorm Preference	Major Preference	Year Preference	High School Preference
60	60	80	30

3.4.1 Stochasticity

There is a uniform randomness involved which allows any agent to meet anyone. And also due to a random seed, the order of interaction between agents is completely arbitrary – no ordering is defined.

3.4.2 Observation

The mean and standard deviation in number of friends is calculated at each time step. Cluster coefficient which calculates the number of triangles in a network, is also calculated during each step of the simulation.

3.5 Details

3.5.1 Initialization

The number of agents in all simulation runs is 6575, based on the underlying dataset of Princeton University students. Each individual in the dataset provides the attributes for one agent in the simulation. All agents are created at the start. While initializing a simulation run, the agents are chosen in a random order. Interaction strategy for all the agents is set once in the beginning. It does not change. Each simulation runs until the number of links made is the same as in the reference dataset – 293307. No link is dropped or modified once it is created.

3.5.2 *Input*

Following initialization, environmental conditions remain constant over the course of the simulation run of the model. The pre-simulation calculated preferences for each of the interactions are hard coded in to the model.

4. RESULTS

In this Section, we compare the simulation results with the reference dataset. First we compare the global or overall results in Section 4.1 and then in Section 4.2, we discuss the attribute level comparison.

4.1 Global results

In this Section, we compare the structure based on the overall network of the reference dataset with the various simulation strategies. In Table 7, we have summarized the basic Social Network Analysis (SNA), over the reference dataset and the simulation results of the four interaction strategies.

We have concentrated on a few and important factors of Social Network Analysis (SNA) in order to compare the reference with the simulated network. The factors with their respective values can be seen in Table 7:

Table 7. Reference Dataset and Simulation Output Comparison

Dataset/ Model	St. Dev. Degree	Assort-ativy.	Trans-itivity	Best Fitted Distribution
Ref.	78.55	0.09	0.16	Exponential (Alpha = 1.98)
Random	18.12	0.08	0.019	Normal
FOAF	93.76	0.11	0.09	Exponential (Alpha = 1.84)
Party	19.64	-0.002	0.03	Normal
Hybrid	79.97	0.105	0.07	Exponential (Alpha = 1.97)

Random and Party modes are the most deviant ones when compared with the underlying reference dataset. In terms of standard deviation in number of degrees (# of friends), they are not even close. Also the distribution of degree is *normal* – bell shaped, as opposed to exponential. The FOAF mode has good results in terms of assortativity, transitivity and even the best fitted distribution. In standard deviation, however, the difference is quite large.

Hybrid mode captures the standard deviation, assortativity and connectedness and also best fitted distribution, quite well, when compared with the reference dataset. In terms of transitivity, it is almost half as the reference dataset. In order to align it with that of the reference dataset, we ran a sensitivity analysis over the parameter space. We did find better results when the parameters were changed, but that hampered the standard deviation and assortativity. Hence we focused more on the overall degree fitting and assortativity. The parameter values for the reference dataset, FOAF and Hybrid mode are also mentioned, where Hybrid mode has almost the same alpha value as the reference dataset, for the fitted distribution. The fitting of the degrees have been calculated by setting the minimum degree to 40.

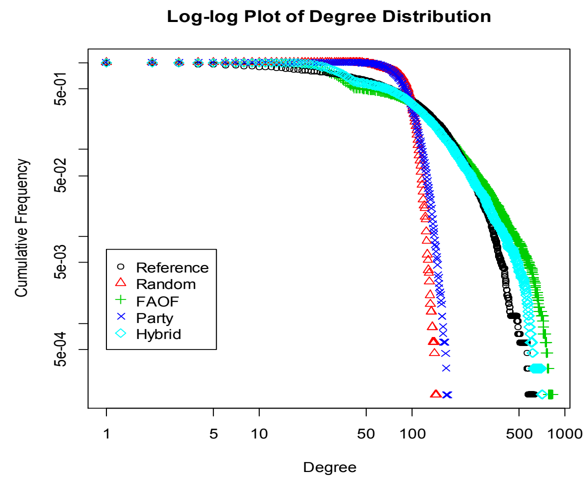

Figure 2. Log-log plot of Total Degree Distribution of all the four simulation strategies and the reference dataset.

We have summarized in Figure 2, the degree distribution of the reference and the four interaction strategies. This only shows the final node degrees after the simulation has been finished. The reference, the FOAF and the hybrid strategy's degree distributions show a power law effect which suggests that most of the nodes have few links while only a few nodes have a lot of links. The other two strategies, random and party seem normally distributed in nature. Their links are more or less uniformly distributed.

If we consider various studies on number of friends in Facebook (see [19–21], mostly all have found that it does have a power law outlook, but there was a seminal work which proved this common belief wrong. According to [22], Facebook has not one but two power-law regimes: one for node degrees less than 300 and one for greater degrees. We found the similar pattern in [23] too. In this case, however, we don't see that for two reasons: firstly, the dataset is too small and secondly the dataset just contains inter-school links. Hence we see just one power-law outlook.

In Table 7 we can clearly identify that Hybrid strategy remains the best candidate when it is compared with the reference dataset. The underlying distribution of both the reference and Hybrid strategy can be identified by such a huge standard deviation; which in turn reflects our earlier finding that both of these are in fact power law distribution.

4.2 Attribute Level Results

In this Section, we compare the results of our simulation runs of all the four strategies for each of the attributes with the reference dataset. We measured the results in terms of the Silo Index. This is an index which identifies the degree of inter-links between nodes with a particular attribute value in a (social) network. If a set of nodes having a value Y for an attribute X, has all the links to itself, and not to any other values of attribute X, that means a very strong community exists, which is totally disconnected from the rest of the network. In short, this index helps us identify how cohesive inter-attribute links are. It ranges from -1 to 1, representing the extreme cases (no in-group links to only in-group links respectively). It can be written as:

$$\frac{I-E}{I+E}$$

where I represents the number of internal links and E the number of external links. In other words, it is the ratio of the difference in internal and external links, to the total links. It is quite similar to E-I index [24], but with the opposite sign. In E-I index we have value 1 when all links are external, while Silo Index has value 1 when all are internal. Hence, the Silo Index could be written as an I-E index. Since the Hybrid strategy has shown the best results, and also due to space limitations, we are presenting Silo Indices comparison of this strategy alone, with that of reference dataset, in Figure 3.

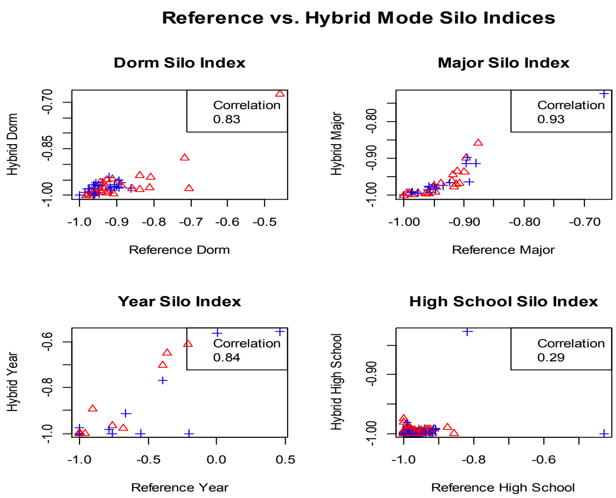

Figure 3. Silo Index for Dorm, Major, Year and High School attributes for Hybrid strategy and the reference network

Apart from High School Silo Index, the rest has quite a high (> 0.83) correlation with that of the reference dataset. Hence it is the preferred mode of interaction. The number of High Schools in the dataset is quite high as shown earlier in Table 1. We could not find better correlation of it with varying parameters values. As for the degree mixing [25] which determines how nodes of similar degrees are connected with each other, we have plotted it in Figure 4.

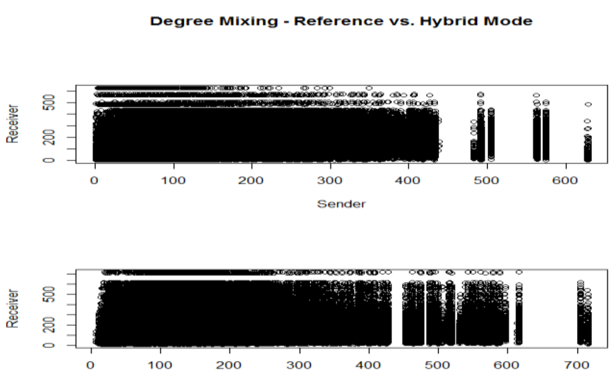

Figure 4. Degree Mixing of Hybrid mode and the reference dataset

In lower degrees (< 400), there is a high rate of similarities between the Hybrid and the reference dataset. In high degrees, the Hybrid mode is slightly different.

After comparing all the four attributes, Hybrid strategy takes the lead when compared with the reference dataset; it presents itself as a good candidate for describing how students might have developed their social network.

5. RELATED WORK

A plethora of research in SNS has been done over the last five years. It is impossible to cover all of it; hence some of the relevant work is being mentioned here. The major focus of such work has been the identification of the static nature of SNS, such as [19]. One of the early works before the popular SNS came into being, was a study done on Club Nexus, a Stanford students online environment in [26] back in 2001. They found that people having similar attributes are more likely to form a friendship link. Based on the various classification of users, an SNS growth model has been presented [27]. Instead of just a snapshot of a social network, but interactions among users, Golder et. el found close social circles in [28] which categorizes the general notion of online "friends" into a broader spectrum. A very detailed quantitative study on students to identify their cultural preferences was done in [2].

To understand the behavior of students' real social network development, a function of contact frequency and shared interests has been used in to make a model. Jackson et el. in [29] developed a model in which a neighbourhood search is done to develop a social network; this can result in many of the characteristics of observed networks.

Adalbert studied Facebook from an economist's point of view [30]. The data which he collected and then studied showed that race plays the most significant role in student friendship development – especially in the case of minorities. In his previous study [11], out of students of Taxes A&M, he found that majority of meeting new friends (26%), were driven by members of the same school organizations. In an another study carried out on students' network [4], race and local proximity, such as dorm were determined to play the most important role, followed by common interests such as major and similar social standing, which in turn were followed by common characteristics such as same year. In our data, however, we could not verify the race factor, as this information is not present in the dataset that we have used.

In case of SNS growth, unlike our model, there are some studies that identify the different classes of users [31]. And also, based on the activity of users, a couple of studies show their social network development [28]. Based on only the structure of an SNS, a couple of exploration techniques have also been devised to predict what new links users are going to make [32], [33], but they usually do not take into account the rich information of attributes of users [34].

In mainstream computer science, there are many "mechanistic and yet tractable" [35] network models, such as Preferential Attachment [36] which specifies an edge creation mechanism, resulting in a network with power-law degree distribution. These models, however, do not take node attributes into account. And in machine learning and social network analysis, where the emphases has been more focused on in the development of statistically sound models that consider the structure of the network as well as the features of nodes and edges in the network [35]. Examples of such models include the Exponential Random Graphs [37] and Stochastic Block Model [38]. These network models are generally intractable and do not offer emergence [35].

6. CONCLUSION AND FUTURE WORK

An agent-based simulation has been described that attempts to explain how students make SNS links, taking into account both endogenous and exogenous factors.

This is an extension of [14] for a larger and diverse university dataset, in which we tried to understand how local preferences and the structural factors might help develop a social network. Unlike the original model, in order to control randomness of Party Mode, we introduced it at an individual level, rather than at the university level. We tailored the model according to the underlying structure of the reference dataset. We have devised and explored a limited number of strategies for student interaction. We compared our simulation results to data gathered from students' Facebook network of Princeton University. We relied on both structural aspects using SNA and semantic using Silo Index for comparison. The strategies of interaction varied from preferential attachment – based on the attribute values, to complete random interactions.

After analyzing the results and comparing them with the reference dataset, we determined that Hybrid strategy, which is a combination of all three strategies: Random, FOAF and Party does the best. It captures the basic essence of the underlying network. From network level measures to the attribute level comparison, it presents itself as a good candidate for the understanding of students' interactions and social network development. Also, FOAF mode captured most of the aspect, apart from the standard deviation in number of friends, which resulted in a different slope for power law outlook. The initial setting of highly similar friends leads to a cohesive community structure and also the friends-of-a-friend process with a power law outlook. Random and Party strategies which are dominated by the random meeting of friends at events did not explain the data well. We do not claim that we presented an exhaustive list of possible social processes, but rather analyzed a few plausible variations. Focusing on personal preference, social structure with some randomness, presents itself as a promising strategy of interaction. While only pre-simulation statistics based on the underlying data, such as correlation, do not necessarily present the best parameter values. For the initial friendship links, the parameter space has to be explored to find the best match.

In future work, we would like to make a more general model, which captures both local and global aspects of a social network. This model will be based on several datasets and on the findings of this model. Also, with the aid of the earlier studies on social network - specifically online social network, we will try to design and understand the processes involved. We will focus both on internal and environmental aspects.

7. ACKNOLWEDGEMENTS

I would like to thank my colleagues Stefano and Bruce at Centre for Policy Modelling for their useful feedback and suggestions. Also I would like to thank the three anonymous reviewers for their comments.

8. REFERENCES

[1] C. Pring, "100 more social media statistics for 2012," 2012. [Online]. Available: http://thesocialskinny.com/100-social-media-statistics-for-2012/. [Accessed: 10-May-2012].

[2] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, "Tastes, ties, and time: A new social network dataset using Facebook.com," *Social Networks*, vol. 30, no. 4, pp. 330–342, Oct. 2008.

[3] A. Dekker, "Realistic social networks for simulation using network rewiring," *International Congress on Modelling and Simulation*, no. i, pp. 677–683, 2007.

[4] D. Marmaros and B. Sacerdote, "How Do Friendships Form?," *The Quarterly Journal of Economics*, vol. 121, no. 1, pp. 79–119, Feb. 2006.

[5] "Facebook Key Facts," *Facebook*, 2012. [Online]. Available: http://newsroom.fb.com/content/default.aspx?NewsAreaId=22. [Accessed: 10-May-2012].

[6] N. B. Ellison, C. Steinfield, and C. Lampe, "The Benefits of Facebook 'Friends:' Social Capital and College Students' Use of Online Social Network Sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, Jul. 2007.

[7] D. M. Boyd and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, Oct. 2007.

[8] L. Hamill and N. Gilbert, "A Simple but More Realistic Agent-based Model of a Social Network," in *Proceedings of European Social Simulation Association Annual Conference, Brescia, Italy.*, 2008.

[9] L. Hamill, "Communications , Travel and Social Networks since 1840 : A Study Using Agent-based Models," *Social Networks*, 2010.

[10] H. M. Singer, I. Singer, and H. J. Herrmann, "Agent-based model for friendship in social networks.," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 80, no. 2 Pt 2, p. 026113, Aug. 2009.

[11] A. Mayer and S. L. Puller, "The old boy (and girl) network: social network formation on university campuses," *Journal of Public Economics*, vol. 92, pp. 329–347, 2008.

[12] V. M. Buote, S. M. Pancer, M. W. Pratt, G. Adams, S. Birnie-Lefcovitch, J. Polivy, and M. G. Wintre, "The Importance of Friends: Friendship and Adjustment Among 1st-Year University Students," *Journal of Adolescent Research*, vol. 22, no. 6, pp. 665–689, Nov. 2007.

[13] L. Adamic, "How You Met Me," *Sixth International AAAI Conference on Weblogs and Social Media*, pp. 371–374, 2012.

[14] S. Abbas, "An agent-based model of the development of friendship links within Facebook," in *7th European Social Simulation Association Conference*, 2011.

[15] V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jørgensen, W. M. Mooij, B. Müller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R. a. Stillman, R. Vabø, U. Visser, and D. L. DeAngelis, "A standard protocol for describing

individual-based and agent-based models," *Ecological Modelling*, vol. 198, no. 1–2, pp. 115–126, Sep. 2006.

[16] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, Aug. 2001.

[17] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You Are Who You Know : Inferring User Profiles in Online Social Networks," in *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 251–260.

[18] A. Traud, E. Kelsic, and P. Mucha, "Community structure in online collegiate social networks," *organization*, pp. 1–15, 2008.

[19] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, S. Bhattacharjee, and B. Bhattacharjee, "Measurement and analysis of online social networks," *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement - IMC '07*, p. 29, 2007.

[20] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," *Proceedings of the fourth ACM european conference on Computer systems - EuroSys '09*, p. 205, 2009.

[21] P. Panzarasa, T. Opsahl, and K. M. Carley, "Patterns and Dynamics of Users ' Behavior and Interaction : Network Analysis of an Online Community," *Journal of the American Society for Information Science*, vol. 60, no. 5, pp. 911–932, 2009.

[22] M. Gjoka, M. Kurant, and C. Butts, "Unbiased sampling of facebook," *Search*, pp. 1–15, 2009.

[23] S. Abbas, "Ethnic diversity in Facebook," *Technical Report Series of Centre for Policy Modeling*, Manchester, 2011.

[24] D. Krackhardt and R. N. Stern, "Informal Networks and Organizational Crises : An Experimental Simulation *," *Social Psychology*, vol. 51, no. 2, pp. 123–140, 2011.

[25] M. Newman, "Mixing patterns in networks," *Physical Review E*, vol. 67, p. 026126, 2003.

[26] L. A. A. and O. B. and E. Adar, "A social network caught in the Web," *First Monday*, vol. 8, pp. 1–22, 2003.

[27] M. Kurant, M. Gjoka, and C. Butts, "Walking on a graph with a magnifying glass: stratified sampling via weighted random walks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 39, pp. 241–252, 2011.

[28] S. A. Golder, D. Wilkinson, and B. A. Huberman, "Rhythms of social interaction: messaging within a massive online network," *In: Steinfield, C., Pentland, B., AckK. Lewis et al. / Social Networks*, vol. 30, pp. 330–342, 2007.

[29] M. O. Jackson and B. W. Rogers, "Meeting Strangers and Friends of Friends: How Random Are Social Networks?," *American Economic Review*, vol. 97, no. 3, pp. 890–915, Jun. 2007.

[30] A. Mayer, "Online social networks in economics," *Decision Support Systems*, vol. 47, no. 3, pp. 169–184, Jun. 2009.

[31] R. Kumar and J. Novak, "Structure and evolution of online social networks," *Link Mining: Models, Algorithms, and Applications*, pp. 337–357, 2010.

[32] L. Backstrom and J. Leskovec, "Supervised random walks: predicting and recommending links in social networks," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 635–644.

[33] A. Agarwal and S. Chakrabarti, "Learning random walks to rank nodes in graphs," *Proceedings of the 24th international conference on Machine learning - ICML '07*, pp. 9–16, 2007.

[34] B. Gao, T. Liu, W. Wei, and T. Wang, "Semi-supervised ranking on very large graphs with rich metadata," *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, no. 49, pp. 96–104, 2011.

[35] M. Kim, "Multiplicative attribute graph model of real-world networks," *Algorithms and Models for the Web-Graph*, vol. 8, no. 1–2, pp. 113–160, Mar. 2010.

[36] A. L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[37] P. Pattison, "Logit models and logistic regressions for social networks," *Psychometrika*, vol. 61, no. 3, pp. 401–425, 1996.

[38] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed Membership Stochastic Blockmodels.," *Journal of machine learning research : JMLR*, vol. 9, pp. 1981–2014, Sep. 2008.