Network Structure, Diffusion and Growth Over Time in a Facebook Application

Hung Truong
University of Michigan
hktruong@umich.edu

1. INTRODUCTION

Notecentric is a note-sharing application that was developed as a standalone social network, and later as an application inside of Facebook using the Facebook Developer's platform. Because the application has a social component (users are encouraged to share notes, discuss class and otherwise connect with their classmates), there is a strong viral growth potential for the application.

This research is concerned with analyzing the growth of the application in order to gain insights about the viral growth and adoption of applications within social networks. As the creator of the application, I have full access to user information such as class rosters, no. of notes taken, user affiliations (universities and regions), user friends, etc. As users added the app, a timestamp was taken,so that a user timeline could be recreated that shows which users may have invited their friends to add the application.

Research questions for this project are: How do network applications grow organically over time? How important are peer recommendations and referrals (versus traditional search and advertising) to the growth of an application? How can applications be built to take advantage of these growth patterns?

2. MOTIVATION

The motivations for studying the Notecentric application network's growth were numerous. First and foremost, I have a vested interest in it because it is my own application. I created the application in hopes that many people would find it useful. If there is a way for the application to see more use, it might be found through analysis of the social network.

Secondly, thanks to the Facebook Developer API, gathering data on an application's users and growth patterns is very easy. Much of the work in network analysis is retrieving good data, and because sufficient data was recorded from the beginning, this project lent itself to analysis.

Finally, there is a lot of talk about so-called "viral growth." Many people, especially marketers and those with products to market, are very interested in the dynamics of viral growth because it is cheap and effective. By studying the application's growth, perhaps a set of best practices for nurturing viral results.

3. RELATED WORK

The paper, "Group Formation in Large Social Networks: Membership, Growth, and Evolution" by Backstrom et alia [1] uses similar techniques for measuring user diffusion into groups. The researchers found that the probability of a user joining a group as a function of the users already in the group increased sub-linearly. One interesting thing to note is that the "networks" studied in the paper did not have the same viral potential as Facebook applications due to user awareness of what their friends are doing (eg Mini Feed, News Feed, etc).

"Structure and evolution of online social networks" by Kumar et alia [2] describes three types of community structures that exist within social networks. One is described as the "singleton" users who do not connect with any other users. Another is a giant component that contains a large group of people who are all connected to each other through direct or indirect links. The middle region consists of all other users; mostly sub-communities that are not connected to the giant component. This is in contrast to the network structure observed in the Notecentric network with two very large components with varying levels of clustering. Perhaps because the network is still in its early stages, the giant component has not yet solidified into one unit.

"The Dynamics of Viral Marketing" by Adamic et alia [3] explore (among other topics) the effectiveness of peer recommendations in convincing a user to purchase a product. The recommendations described in the paper could correlate directly with the "Invite a Friend" functionality found in many applications, including the one being analyzed. The interesting part about recommendations is that as they increase, not only do they not increase the probability of a purchase logarithmically, they actually lower the probability. For example, it was observed that users followed purchasing recommendations with the highest probability at two recommendations. After two, the probability slowly tapered off. This is in contrast to the findings from the Backstrom paper that the higher the number of friends in a group, the more likely a person is to join it. This may be like comparing apples to oranges, however, since membership in a group is not necessarily an endorsement for it.

4. APPROACH AND METHODS

Network data was taken directly from the Notecentric application through the Facebook Platform Application Programming Interface and processed with a Ruby script. The most recent dataset used included information for 648 users of the Notecentric application. Following the terms of use for the Facebook application, only unique user ids and relations between users are stored, along with statistical information like the time of application installation. Because we have timestamp information on user installation, other information can be inferred including the number of friends who had added the application at the time of a user's installation. For example, if I added the application after two of my friends had, the script would record that fact. The number of friends who had not added the application at the time of a user's installation is also recorded, as well as the number of peers in a user's networks that had added the application before the user.

The Ruby script used to process the raw network data contained a few notable algorithms, including a simple implementation of a union-find algorithm used to record component sizes in the growing network. The data from the script was then fed into GUESS, an open source network and graph visualization tool.

The network visualizations that appear in this report are taken from screenshots of the GUESS visualizations.

5. Findings

5.1 Network Structure

In many ways, the visualization of the network follows the structure outlined by Kumar et alia. The network of application users is divided into three categories. There are the singleton users that added the application and invited none of their friends, taking notes by themselves. In addition to these singleton users, there are many users in the middle region. Finally, instead of one large strongly connected component, there are two.

One of these components is quite long (in terms of average shortest path) and contains a number of users from different networks. This component most accurately resembles the kind of growth we'd expect in a viral application. Users from one network slowly propagate the application to other networks. The application grows depth-first through users who adopt and recommend the application to friends.

The other large component is a highly clustered group of fairly homogenous friends. Most likely one person in this group found the application through the search function and invited all of his friends. Because this group had such high clustering, the "peer pressure" to add the application may have encouraged many to add the application. As one saw all of his/her friends adding the app, he/she would feel compelled to add it as well.

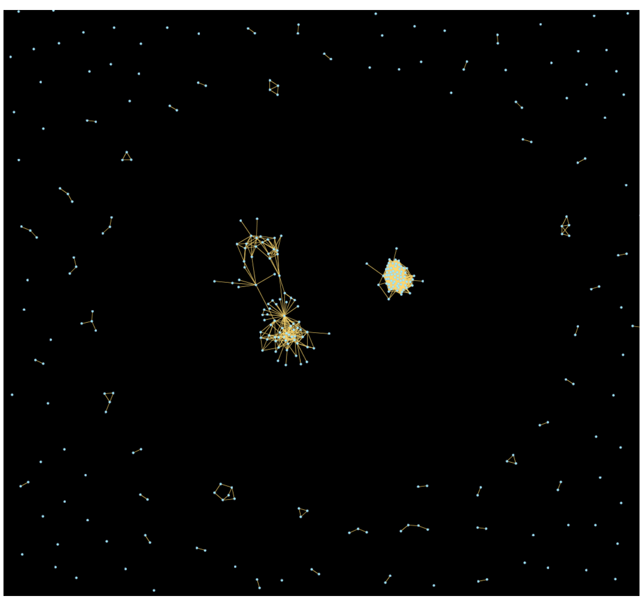
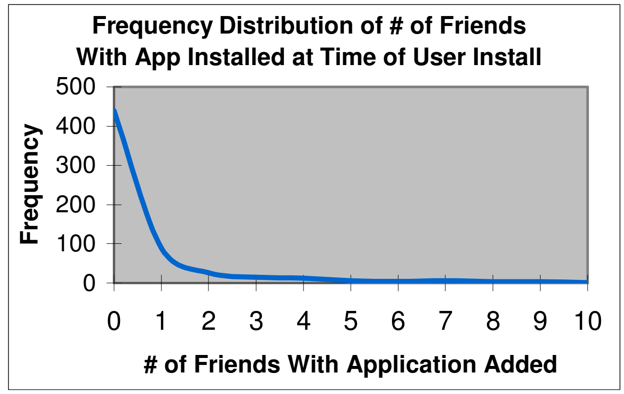


Figure 1. The two giant components are seen surrounded by the outliers.

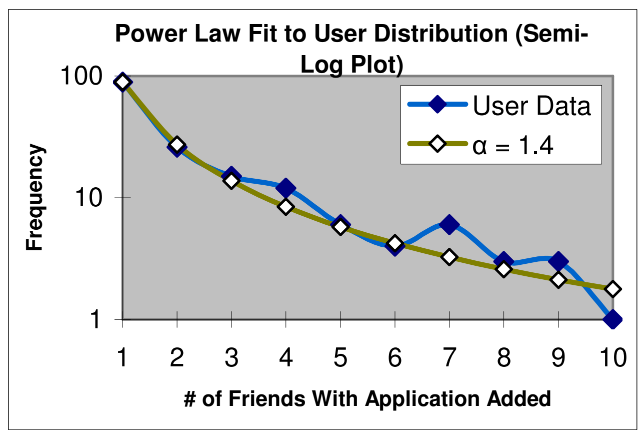
5.2 Adoption Rates and Influence

Out of 648 users, there are 440 users who added the application before a friend added it. This means there were 208 who added after a friend had added the application. We'll assume that the fact that a friend added the application had an effect on the user's own adoption of the application. About 32% of users who added the application did so after at least one friend had added it. This number alone shows the potential for user growth when viral marketing is involved.



I was interested in seeing if the number of friends that had added the app first would have an effect on the probability of adoption for a user to add the app himself. The average number of friends who had added the application before a user added it for himself was 2.04. Despite this, the median was still 0, since the majority of users added the application before any of their friends had.

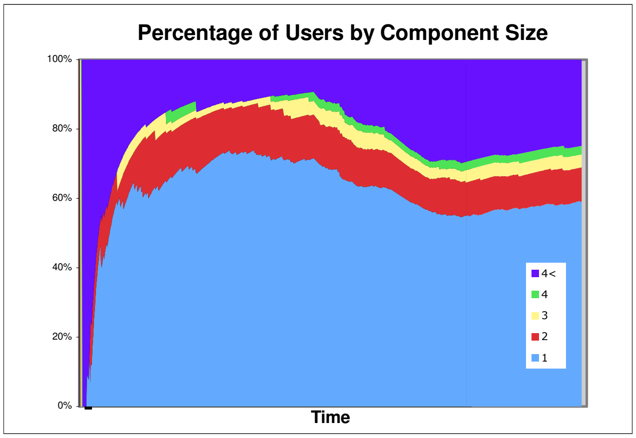
The distribution of the number of friends with the application previously installed at the time of one's own install starts high with 440 users adding the application with no friends having added the application prior, 89 having added the application after one friend had added it, and so on. Not surprisingly, the distribution here follows a power law. The fit with an alpha value of 1.4 is slightly noisy, but can be contributed to the relatively small amount of data gathered thus far.



It would also be interesting to see if there is a correlation between the percentage of friends who have added the application and the likelihood for a user to add the application himself. Among all of the users, the average percentage of friends who had the application added (friends who added/total number of friends) before the user himself added was about 2.7%. The percentages ranged all the way up to about 57% for users with a non-trivially low number of friends (a user with 26 friends, 11 of whom added the app finally decided to add it himself).

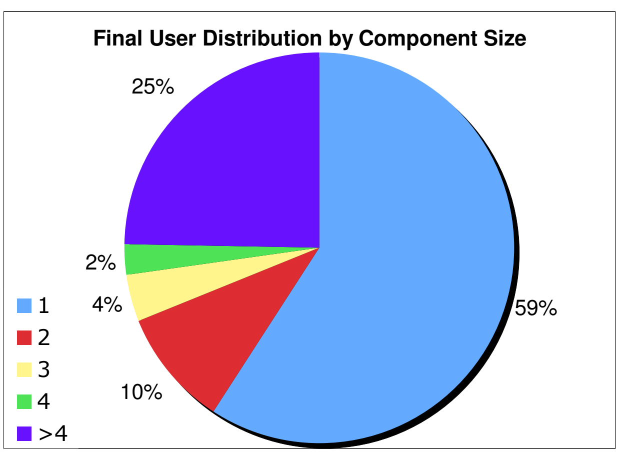
Finally, I thought it would be interesting to see how many peers in a network would have the application added before a user added the application himself. The statistics for the network addition might be slightly inaccurate, because users can be in multiple networks, and therefore be counted multiple times if another user is in the same networks. On average, a user would already have about 11.4 peers in his networks who had added the application before adding it himself. This networks measure is more of an interesting statistic than something that drove user adoption, since users are not given any information about network adoption for applications.

5.3 Component Sizes and Distributions



Because user timestamp data is available, it is possible to see how the network grew. One way to visualize this is to calculate the percentage of users in various sizes of components. Users in components of size 1 are the aforementioned singleton users who do not influence their friends to add the applications. Users in components of size 2-4 are those who lie in the middle region, forming cliques that may be class study groups. Users in components with more than 4 users are assumed to be in one of the giant components.

Each time step occurs at the addition of a user to the Notecentric network, and so there are 648 data points in all. In the first few steps, most users are in small components of the network. Quickly, a giant component emerges but is overtaken by singleton users. Toward the end of the progression, most users are singleton (~59%), with the next most common being in one of the giant components (~25%) and next being in a two-user component (~10%). Interesting to note is that while the giant components are strongly connected, the majority of the users still lie in singleton components. In a randomly formed network with a similar average degree (around 2), the giant component would contain most of the users. The Notecentric network is much more clustered, especially in one of the giant components.



6. CONCLUSIONS

6.1 On Network Structure and Component Sizes

It would appear that the network structure is unusually fragmented. Two giant components and a majority of singletons suggest that the network consists of many sub-networks. This makes sense, since Notecentric is supposed to connect classmates. Users of Notecentric are more likely to share notes with friends in the same class, or at least the same school. The large number of singletons might suggest that the note sharing capabilities of Notecentric are not as useful or easy to use as advertised. It is reassuring to see that there are quite a few users in the middle region, however. As stated earlier, this might be due to users forming online study groups.

The component size distributions also reinforce the idea that the Notecentric network is inherently fragmented. One giant component is the result of organic growth through friends suggesting the application, not necessarily to classmates. The other giant component appears to be an entire class, or a very homogenous group of friends who have decided to use Notecentric together.

6.2 On Adoption Rates and Influence

The adoption rates show that most users find Notecentric by way of an application directory, and not by word of mouth. This is one indicator that the application is probably not viral. Truly viral applications would have a higher number of adoption for users that already have at least one friend with the application added. In order to facilitate this kind of growth, an application should ideally be shown though a user's "news feed" when added. Friends of the user would become aware of the application may add it out of curiosity. This does not always happen since news feed stories are at a premium and are picked using an obfuscated algorithm to determine the most interesting.

7. FUTURE WORK

I intend to continue studying networks in the future. While the data gathered from the Notecentric network was very interesting and yielded very surprising results, I hope to find even more interesting data, perhaps from a truly viral application.

More importantly, the shortcomings from the data provided from Notecentric will be useful in determining what kinds of data should be collected in the future. Information on user application uninstallation, something not captured by Notecentric, would be useful in studying future networks. Future research will build upon the techniques and algorithms used to study the Notecentric network.

8. REFERENCES

[1] L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan. Group Formation in Large Social Networks: Membership, Growth, and Evolution. Proc. 12th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2006.

[2] Kumar, R., Novak, J., and Tomkins, A. 2006. Structure and evolution of online social networks. KDD '06. ACM, New York, NY, 611-617.

[3] Leskovec, J., Adamic, L. A., and Huberman, B. A. 2006. The dynamics of viral marketing. EC '06. ACM, New York, NY, 228-237.