Seminar Paper

Advanced Topics in Social Network Analysis

Notification as a mean to increase user
activity in newborn online
communities.

Submitted by:

Ezra Fuks

David S. Brima

October 1th, 2015

Table of content

1. Abstract	3
2. Introduction	4
3. Related Work	5
3.1 Theory	5
3.2 Prior Research	6
4. Research Question and Hypotheses	9
5. Methodology	10
5.1 Measures	10
5.2 Sample and Population	10
5.3 Instruments and process	10
6. Results	12
6.1 Descriptive results	12
6.2 Correlational test and hypothesis testing results	12
6.3 Graph network analysis with NodeXL	13
7. Discussion	14
7.1 Limitations	14
7.2 Future work	15
8. Conclusion	16
9. References	17

1. Abstract

There is a considerable challenge in getting online communities of the ground; founders of such platforms employ many strategies to manage this issue and go over the tipping point; this study examines the influence of notifications on user activity in general, and in context of admin activity. An online platform for co-creation of playlists called Echo was studied; Both classical statistic and network analysis was employed in order to establish the relationship between these variables and the structure of the network. A weak to moderate relationship was found, while the main carriers of activity were admin users, as key carriers of user engagement in the early days of Echo.

2. Introduction

Online communities whose main focus is user generated content by definition need contribution from the user in order to fulfill this purpose. The following research will focus on how push notifications influence user activity and contribution within an online community, based on a mobile app called Echo. It is a recently launched iOS app for co-creation of playlists; the platform employs third-party content providers that the users can use as the source for their music; users choose individual tracks, decide on the order and invite other users to add tracks. The platform uses different features to encourage participation, such as letting users follow each other and receive notifications based the activities of people following. Push notification are made to physically grab the user's short attention span and bring them into the app; such features are called “hooks” and their effectiveness has been shown in many studies of social networks and online communities.

As Echo is two months from its launch on the iOS App Store, it is essential to analyze how social features influence user's activity, in the early period of user acquisition; This is important as rolling out features is costly to early stage startups with no source of revenue, involving different people to commit to building them, increasing labor costs. Furthermore, if they prove to be ineffective or counterproductive, the product will be equally or less sticky for existing users, costing the company time, stamina and funds. On the other hand, if they prove to be effective in the sensitive early days of the launch, tweaking details could lead to users coming back more often and becoming more proactive and lead to a tipping point for the growth of the community.

The study will focus on finding the relationship between users receiving notifications and their activity on the app, listening and creating playlists. Furthermore, the study analyzed this relationship from a network perspective, by looking into the networks users form based on activity mutual activity, identifying key insights from network properties and heat points. By studying the community on a network level, it is possible to identify overall trends and visualize important connections.

3. Related Work

3.1 Theory

Different theories were used in order to better study the data and discover how the user's activity is connected to the user's initial acquisition.

The main theories used to describe the user motivation in participating in online communities are the uses and gratifications theory (UGT) and the organizational commitment theory (OCT).

The UGT analyze what will make users to consume media based on their personal specific needs. This theory falls under the media consumption theories but UGT differs from the others since gives the power to the consumer, and allows him to choose which media they might consume with the assumption that consumers has a clear intent and use.. The UGT tries to investigate on why and how users seek specific media to satisfy their needs. It assumes that registered users are not passive consumers of the media. The Echo application is inside the music segment and since the registered user are able to create their own content and share it with their followers, the users has a clear intent in sharing music only and only the genres that they think are appropriate to share so it will also assume an active role in interpreting and integrating media into their own lives. This theory would then imply that the media compete against other information sources for viewers' gratification. So that Echo's users are interested in being part of the online community. (Lampe, Wash, Velasquez, & Ozkaya, 2010, p. 02)

The OCT predicts that a larger group of users will identify more with an organization and then they will start to contribute to the online community. This mean that in an online network a group of people who is very active can be defined as "organization" since they are the one that has more connection and influence on others. This theory is part of the organizational behavior segment, which analyze the psychological attachment of the user or members of a physical or digital community. In Echo there are admins who posted and shared playlist and create a music activity, and then share this with different segments of the users. According to this theory thanks to this process the newer user will contribute more to the online community since they are engaged by the admins account. (Lampe, Wash, Velasquez, & Ozkaya, 2010, Hayat, Lesser & Samuel-Azran, Forthcoming).

The social learning theory is used to identify several steps for learning process such as attention, retention, reproduction and motivation. Some of this processes can be found in social networking and online communities, since the feature of a content feed user are able to view their friends activity (attention) recall the activities of the users (retention) and

this will encourage them to participate to reproduce content. So this will motivate users to participate (singling out). (Burke, Marlow, & Lento, 2009)

3.2 Prior Research

Many researches has been done in order to study the behavior of the member in an online communities (e.g., Hayat & Mo, 2015; Hayat, Samuel-Azran & Galily, 2016), and those have been used as the base for this current work, since no research has been done with research question and by using the Echo platform.

The paper “Motivation to Participate in Online Communities” by Cliff Lampe, Rick Wash, Alcides Velasquez and Elif Ozkaya, discuss the problem with designing a socio/technical system to motivate user to contribute content. Their study was designed for an user-generated encyclopedia online called Everything2.com (Lampe, Wash, Velasquez, & Ozkaya, 2010, p. 02), which has similar feature to Echo, they are both based on a user generated content, and there in both is given a lot of control to its users, they can create content and only if users wants to interact with each other then they need to register to the platform. Thanks to all of this similarities is it possible to determine how the user will behave and as resulted in the research that at the beginning users came to get information to but some of them were also motivated to use to site to get or provide entertainment.

Sites based on user interactions need to accommodate multiple motivations (Mo, Hayat & Wellman, 2015), and not assume that they know what all the users will create from the beginning during the research it became evident that this was not necessarily to get motivation since in order to have a greater motivation the administrator need to interact a lot with its users.

This research also express the difference between the active users and the lurkers. In the paper written by Blair Nonnecke and Jenny Preece “Lurker demographics: Counting the silent” is discussed the importance of the lurkers in an online community since the lurking levels in online communities is very high, there are almost 90% of the user who are lurkers (Nonnecke & Preece, 2000). A lurking user was defined as someone who did not post at all in the time that the research has been done. The study also discuss that in smaller discussion list or communities the importance of having lurkers is really a key point in developing the community. In Echo and in his newborn users have a good probability of a high percentage of users who are lurkers and is important to know the relationship that they have with active users.

In the paper written by Cliff Lampe, Rick Wash, Alcides Velasquez and Elif Ozkaya it express the importance of the stickiness of the user to a developing community. (Lampe, Wash, Velasquez, & Ozkaya, 2010, p. 02) Such as Jill Freyne, Michal Jacovi, Ido Guy and Werner Geyer, authors of the paper “Increasing Engagement through Early Recommender Intervention” discuss as the important of the participant who after the first sign up fall away and are not sticky to the new communities while in order to succeed the platform need to create a critical mass of user, which in Echo are the admins, and make new user aware of them. (Freyne, Jacovi, Guy, & Geyer, 2009, p. xx) The main issue in losing users is that they cannot find others connected to them (Hayat & Lyons, 2010), or the platform is not about something he is passionate about which is strictly connected with the issue of the lack of information given by the user, because the more information the platform has the better connection can be done between users, in the research paper in fact is found that the users who are connected with more familiar users will tend to have a higher activity after sign up (Hayat, Brainin & Neter, Forthcoming). The early recommender will influence new members and in order to have the better outcome in create a successful community is to sort users based on their personal activity.

Moira Burke, Cameron Marlow and Thomas Lento wrote in “Feed Me: Motivating Newcomer Contribution in Social Network Sites” that friends has a modest impact in newcomers. in the Facebook platform it was discovered that “newcomers who uploaded more than 1 photo had a mean of 27 friends who uploaded an average total of 220.2 photos” (Burke, Marlow, & Lento, 2009).

In this research the authors describe that the implications for designing an online community must include highlighting opportunities for others to leave feedback and increase distribution for the active newcomers and encouraging friends to pay more attention through singling out for who’s not active.

The research who has most relevance with this paper is “A New Framework for Increasing User Engagement in Mobile Applications Using Machine Learning Techniques” by Merve Gençer, Gökhhan Bilgin, Özgür Zan and Tansel Voyvodaog̃lu which study the user engagement in mobile application (Gençer, Bilgin, Zan, & Voyvodaog̃lu, n.d.). They analyze that the smartphone adoption has potentially increased in the last few year as never before. This increase gave the possibility to many developer to create mobile application. As of June 2012, Apple App Store operates in 155 countries with more than 650.000 applications, and as of March 2012 Google Play operates in 129 countries with more than 500.000 applications. Echo is only available on the Apple App

Store. The authors define the engagement and the stickiness as key point to succeed in a mobile application. In the paper the authors define several methods in order to increase to engagement such as rewards and material incentives provide a higher motivation to the newcomers. But the most relevant method to create a reengagement is sending push notification to mobile phones which trigger usage and the power of the push notification are miscalculated since through it is possible to extract contextual information personal tendencies and personal motivation, based on Bartle's research. (Gençer, Bilgin, Zan, & Voyvodoğlu, n.d.). Is it also possible to analyze that survey based data provide quick database building but most of the time the survey's data do not follow the truth, so this process will not be used in this research, but the data will be taken from the account of the users. In order to generate the data for the database it was used a similar process that was used by the authors of "A New Framework for Increasing User Engagement in Mobile Applications Using Machine Learning Techniques". But for Echo only the user action information has been used. Which covers the most important features of the database. All the actions of the user in the application are tracked and recorded to the database.

4. Research Question and Hypotheses

The study focuses on two main points: the relationship between the push notifications of the users in a newborn online community and their activity in the online community and the early role of admins in pushing users to be more activity; these has been chosen as a point of interest as these were predicted to be necessary by findings and theories found in the previous research; Uses and Gratification theory predicts that in order for online media to succeed, it needs to fulfill user needs; the other main theory found in the previous research was Organizational Commitment theory, predicting that in order for users to begin contributing to an online community, they need to see an example of contribution and activity from organization members. It was also analyzed that there is a considerable increment in traffic usage of the mobile application when admins push user to be more active. All of this point are the key to answer the research question: "What is the relationship between users receiving notification and their activity on a mobile-based online community?"

The H1 for the research is that the notifications should have a positive relationship with user activity; H2 would be that there is a negative relationship between notifications and user activity and H0 is that the there is no relationship between the two variables. Additionally, this relationship occurs within a network, meaning a classic statistical analysis would reveal only a part of the picture; as the platform is in its infancy, an additional inquiry is the role of admins in influencing activity through notifications sent to the users following them; it is also an interesting point since members of the Echo team have done this on purpose with the assumption that it is necessary for users to come back; as predicted by OCT, the hypothesis for is that admins are a necessary link in an online community, as without them the network will regress to complete shut down. According to the prior research there are possibility to increase users activity by using admins and thanks to this hypothesis it will be easier to find the relationship between push notification and the activity of the users.

5. Methodology

5.1 Measures

The main purpose of the approach is to establish if there is a correlation between the number of notifications users on Echo receive and the amount of crucial activity they performed, operationalized as number of playlists created and the number of time users listened to a playlist. These variables were picked for several reasons; the project managers of the startup used this feature as a strategy to induce activity from users, which is a prescribed strategy according to the Uses and Gratifications theory as it fulfills certain needs users seek within the context of mobile use, while listening and creating playlists are the main activity. Listening is a form of lurking (users have engagement with content on the platform without contributing) but as previous research has shown, such activity is crucial for understanding and project management decision making, as it can contribute to more than 90% of total community activity. Such features are also relevant to increase stickiness of the app, as prescribed in the Lamp et al study (2010). Another aspect of the study was testing the effect of notifications about admin posts on users activity within the community. Furthermore, the data had been analyzed from a network perspective, identifying influential nodes and dense clusters, thus giving us insight into the networks arising as a result of activity and context to the data.

5.2 Sample and Population

The company behind Echo agreed to provide user data for all their users since Echo's launch on the App Store, having obtained the the agreement from their users to do so; the database provided has been gathered on a spreadsheet, with the ID of each user and the number of notifications, listens made and playlists created for each user; the database was further formatted into an edge list, where each activity event was operationalized as tie between two User IDs; the user database was further filtered under countries, as different cultural/economic/infrastructural factors differ significantly from country to country, and are of valuable insight for the company.

5.3 Instruments and process

Upon entering the Activity Count Data into SPSS, the company workers IDs were taken out as they were creating playlists, as their activity was the part of community building strategy, and have no connection to notifications. Pearson's Correlations tests were used to test the relationships between: the number of notifications and. the number of playlists

created, the number of playlists listened to and the number of playlists created, the number of notifications and the number of playlists created and listened to. After completing statistical analysis of the complete data set, the edge list for the listening events were entered into NodeXL, an open source add-on for Microsoft Excel for graph building and calculations of network measures. Notifications were operationalized as an edge lists of User ID activity and User ID notifications, Listeners were operationalized as User ID of the listener and User ID of the playlist creator. The whole database was entered into NodeXL and the graph metrics were calculating. The following step was identifying active clusters and taking their network properties. Within these clusters, admin nodes were identified and receivers of their activity notifications. As total amount of listeners came from the playlist created by a small amount of users, so the IDs in the playlists created were taken ordered from smallest to largest ID number, to see the most common nodes. The dataset was then visualized using the Fruchterman-Reingold graph algorithm and using the Dynamic Filter option on NodeXL, they were filtered down to the values in Vertex of the IDs that created playlists, in order to reduce clutter and reveal main activity. The Harel-Koren Fast Multiscale graph algorithm was then ran to provide for more visual clarity. The IDs of listeners and playlist creators were then copied to a separate sheet for further analysis. After registering the more active parts of the graph through graphic analysis and extracting the user list they contain, overall graph metrics were calculated including: Number of vertices in the network, total number of unique edges, number of edges with duplicates, total number of edges, self-loops, when a vertex maps into itself, betweenness centrality, a measure of importance for a node within a network, eigenvector centrality, a measure of the degree of the nodes to which the observed node is connected to, clustering coefficient, a measure of the tendency of nodes in a graph to cluster together. The most relevant nodes were found and their individual measures were taken.

6. Results

Descriptive and correlation test results for the total population are presented in the first part of the results section; in the second part, results for the network analysis are presented, with descriptive and correlation test results for relevant user groups identified through graph analysis and calculated network properties.

6.1 Descriptive results

The total number of users the app had in its database was $N=5763$, with three admin users being taken out from the data set, meaning $n=5760$. The number of notifications received were $n=7426$, the number of playlists listened to were $n=$ and the number of playlists created were $n=387$; the average number of notifications received per user was $n=1.31$, while the average listening was and the average number of playlists created per user was $n=0.08$, while the average of playlist listened per user was 1.69; 78% of user come from China, 12% from the United States and the rest from the remaining 67 countries (according to Google Analytics data of the iOS app). The number of excluded admin user IDs was three.

6.2 Correlational test and hypothesis testing results

Within the overall user base, a statistical significance has been found to show notifications on Echo have made influence on both listening and playlist creation activities; the Pearson's Correlation coefficient on the relationship between the number of playlists created and the number of notifications received is $r=0.866$, indicating strong positive relationship, while the coefficient for listening was $r=0.259$, a weak positive relationship (Table 1).

As for the correlation coefficients for different territories, the more relevant countries with significant portion of the user base and/or activity are presented here; for users coming from the United States, the Pearson's correlation coefficient between creating playlists and receiving notifications was $r=0.284$, while for listening to playlists it was $r=0.149$; in China, the correlation coefficient for the relationship between notifications received and playlist created was $r=0.108$, while the coefficient for the relationship between notifications received and playlist listened to was $r=0.199$. For users that had no country of belonging, accounting to 41% of the total user base, the correlation coefficient for the relationship between notifications received and playlist creation was $r=0.342$ while for listening it was $r=0.101$ (Table 2).

6.3 Graph network analysis with NodeXL

Three graphs were produced during the process: a graph that includes all users, a graph including the portion of the user base (including admin accounts) with above average activity, and the user base only accounting for the biggest amount of edges (Table 3). Through this analysis, it has been found that around 3338 users engage in some form of activity and are integrated in the community through their activity. Around 50 nodes account for almost all events occurring in the community, while admin accounts represent the 70% of these activities. The overall network measures are as following: the number of vertices $N=3395$, the number of unique edges is $N=5163$, edges with duplicates add up to $N=4466$ and the total number of edges is $N=9629$. The number of self-loops in the data set is $N=171$. The maximum degree is $n=2644$, the minimum degree is $n=1$, average degree is $n=4.025$ while the median degree is $n=2$. The minimum betweenness centrality is $n=0$, the maximum is $n=4419058.912$, the average is $n=2309$, while the median is $n=0$. Maximum Eigenvector centrality is $n=0.016$ while the average, median and minimum is $n=0$ (Table 4). It is interesting to add that 10 nodes have a overwhelmingly higher betweenness centrality over the rest of the data set, the admin IDs: 35 with $n=4419058$, 34 with $n=973119$, 43 with $n=735833$.

7. Discussion

The results of both the classical statistic analysis and the network approach analysis are providing a rare view into a community in its infancy; results of statistical analysis are in line with the prediction of both the user gratification theory and organizational commitment theory; notifications are supposed to satisfy socialization user needs and are giving a positive relationship ranging from country to country; for Chinese users this has been somewhat weaker than for US users, while it has been shown to be a more significant influence for creating playlist than for getting users to listen to playlists; the different receptions of notifications across territories can be accounted by cultural differences and the context such places have across cultures, but the language barrier can also present itself as a problem, since Chinese users will have a harder time comprehending notifications they receive in English than English speaking Western users; in spite of the relatively weak relationships shown in the statistical analysis, the community is developing in line with the predictions of OGT; founders of the community create activity and recommend new users their profiles for following, in combination with notifications on their activity works; admins account for the biggest amount of contribution to the platform, as a result they created a network followers that come back into the app through push notifications; eventually, they began transitioning from lurkers to contributing users. The relatively weak relationships can also be a result of poor calibration and design of notifications, rather than predicting the influence of notifications in general. The second part of the research that looks into the network structure and metrics has provided for a detailed picture for the previous findings; an overwhelming level of clustering around the admin accounts and a certain number of accounts around them have shown that the community in its current level cannot survive without admin contribution and activity. Even so, there are signs of independent users that map into each other without the involvement of admins in the process.

7.1 Limitations

The data gathered is from Echo comes from a relatively short timespan; the app is online since July 11th, making assumptions about the general influence of notifications on user activity in online communities limited; the definition of a community in infancy is problematic, as it can mean different things for different communities. On the other hand, the data is not temporal, so it is not possible establish any kind of causation. Furthermore, tracking user activity data including measuring the activity in the user interface and mapping it into the wider picture would enable better understanding of user behavior on the platform in its infancy.

7.2 Future work

For future studies, it would be recommendable to develop a system that would track and assign actions to each individual user, tracking his actions within the app, the overall actions she makes and the connections she makes with other users. Longitudinal statistical studies of the correlations would provide for more in depth information regarding user activity. Furthermore, it would be recommendable to perform some qualitative studies of users using surveys, questionnaires and user interviews in order to provide some context to the mined data.

8. Conclusion

The study had been intended to understand notification and admin contribution strategies in user engagement within online communities; the data for the research came from Echo, a social platform for co-creation of playlists; the theoretical approach taken has been based on the Uses and gratification theory and Organizational Commitment theory; overall statistical test were performed on the complete database of users and a moderate positive relationship has been found between receiving notifications and users creating and listening playlists in line with the prediction of Uses and Gratification theory. Through network analysis, it has been shown that in the infancy of an online community, admin contribution is necessary for pushing new users to contribute and become more active as predicted by Organizational Commitment theory. Without these admins, according to graph metrics, the network would fall apart.

9. References

- Burke, M., Marlow, C., & Lento, T. M. (2009). Feed me: motivating newcomer contribution in social network sites. doi:10.1145/1518701.1518847
- Freyne, J., Jacovi, M., Guy, I., & Geyer, W. (2009). Increasing engagement through early recommender intervention. doi:10.1145/1639714.1639730
- Gençer, M., Bilgin, G., Zan, Ö., & Voyvodaoglu, T. (n.d.). A New Framework for Increasing User Engagement in Mobile Applications Using Machine Learning Techniques.
- Hayat, T. Z., Brainin, E., & Neter, E. (Forthcoming). With some help from my network: supplementing eHealth literacy with social ties. *Journal of medical Internet research*, 19(3).
- Hayat, T. Z., Lesser, O., & Samuel-Azran, T. (Forthcoming). Gendered discourse patterns on online social networks: A social network analysis perspective. *Computers in Human Behavior*, 77, 132-139.
- Hayat, Z., & Lyons, K. (2010). The Evolution of the CASCON Community: A Social Network Analysis. CASCON 20th Annual International Conference. Toronto, ON, Canada.
- Hayat, Z., & Mo, G. Y. (2015). Advice giving and receiving within research networks. *American Behavioral Scientist*, 59 (5), 582-598.
- Hayat, T., Samuel-Azran, T., & Galily, Y. (2016). Al-Jazeera Sport' s US Twitter followers: Sport-politics nexus?. *Online Information Review*, 40(6), 785-797.
- Lampe, C., Wash, R., Velasquez, A., & Ozkaya, E. (2010). Motivations to participate in online communities. doi:10.1145/1753326.1753616
- Mo, G. Y., Hayat, Z., Wellman, B. (2015). How Far Can Scholarly Networks Go? Examining the Relationships between Distance, Disciplines, Motivations, and Clusters. *Studies in Media and Communications*, 9, 107 - 133.
- Nonneke, B., & Preeee, J. (2000). Lurker demographics: Counting the silent.