

FORUM VS. SOCIAL NETWORK: WHAT IS THE BEST FOR COMPANY AND COSTUMERS?

Sara Amini, University of Illinois at Chicago, Department of Computer Science, Room 1120 SEO (M/C 152), Chicago, IL 60607, samini3@uic.edu

Nasim Mobasheri, University of Illinois at Chicago, Department of Computer Science, Room 1120 SEO (M/C 152), Chicago, IL 60607, nmobas2@uic.edu

Chris Kanich, University of Illinois at Chicago, Department of Computer Science, Room 1120 SEO (M/C 152), Chicago, IL 60607, ckanich@uic.edu

Vahid Balali, Department of Civil Engineering and Construction Engineering Management, California State University Long Beach, 1250 Bellflower Blvd, Long Beach, CA 90840, 562-985-1643, yahid.balali@csulb.edu

ABSTRACT

In the today's business world, it is crucial for companies to maintain sustainable growth to connect with both current and prospective customers. Strong connections with customers not only help improve user experiences; but also aid collecting feedback for future product development. Most importantly, these connections help establish long-lasting customer relationships and brand loyalty. This paper explores the network characteristics of two main categories of online communities and social networks from a business perspective for the purpose of facilitating customer engagement. This paper shows which of online communities (forums) and social networks is better and more beneficial for companies to engage with their customers. The network statistics are analyzed by using the data from an online forum and Facebook. This analysis covers degree distribution and investigates various aspects of the networks, such as limited attention and social influence. Information diffusion is also studied based on epidemic models. The results show that social networks such as Facebook will be a better choice for companies for the purposes of connecting with their customers and building relationships.

Keywords: Forum, Social Network, Data Analysis, Costumers Connection.

INTRODUCTION

Nowadays, social media has become an integral part of everyday lives for most people. Facebook, Twitter, Instagram, LinkedIn, Yelp, and etc., are some examples. However, there are actually two different categories under social media. One is the social network and another is the online community. In this paper, the networks of two different kinds of social media are studied from a business perspective to see how companies can utilize them to connect with customers and promote their brands and products.

A company needs to be connected with its customers to maintain sustainable growth and collect feedbacks about the product and incorporate them into future product development. Also, the company can provide more product support and troubleshooting to improve the user experience [1]. All these work can help building a relationship with the customers and promoting the brand and/or products. To achieve this goal, the best platform would be social media due to its ease of access, valuable user information, and large user base. This paper tries to initially answer the research question which one of the social network media is better and more beneficial to the company? Specifically, this research compares Facebook network's structure and characteristics with those of a forum, which is actually a classic example of online

community. Then, the best solution for companies to connect with their customers is proposed. In the following, the definitions of online communities (forums) and social network are discussed more detailed.

Online Community (Forum)

Online communities, such as forums, are designed for discussion about a common subject. This dynamic produces a diverse user's community that is held together by a common interest without any prior knowledge or affiliation or interaction. Forums enable users to interact, collaborate, and learn from each other. The main purpose of participation in these communities are either contributing to a subject of interest or benefiting from discussions. Some of the users with the question or concern tend to visit only once, leaving when their need is addressed or ignored, while others will stay to become regulars to help with other users. Users could never meet each other, but they still communicate with each other without any problems because they have common interests and topics.

Social Network (Facebook)

Social networks, such as Facebook, form and evolve based on pre-established connections between users. Social networks are the online representation of circle of friends and due to their global popularity, they also capture the small-network effect of real-world networks. The connections in these networks are built one at a time and people are held together based on existing relationships. Another interesting aspect of the social network is the uniqueness of social networks for each user. Each user has a distinct social network and therefore less anonymity, harder to create a fake account, and more privacy at risk. Due to social networks nature, sharing information on social networks is more private and less topic-based. The information shared by a user is visible to his/her friend (unless the default privacy setting is changed) and there is no limit or rule for shared topics.

DATA COLLECTION AND CHARACTERISTICS

Two different datasets are used in this study to compare the structure and characteristics of two networks: forum and social networks. A dataset collected from Bleeping Computer forum is used for the forum and Facebook dataset represents the social network. The characteristics of these datasets are described in more detail as follows:

Forum Dataset: Bleeping Computer Forum

Bleeping Computer is a computer help site which was founded in 2004 by Lawrence Abrams [2]. It is a resource site for asking and answering computer-related technical questions. All services on this site are free to the public. It consists of many forums which are organized by different computer-related topics. People can readily register to the site using a valid email address with no fee and ask the computer, security, and technical questions in a topic-related forum. There are professionals who address these questions. Forums within the Bleeping Computer resource site [2] are divided into different topics. Some of these topics are Microsoft Windows support, hardware, security, Internet, and networking. One of the most popular topics is security. Within the security topic, the data was scraped and collected from "Am I infected? What do I do now?" forum.

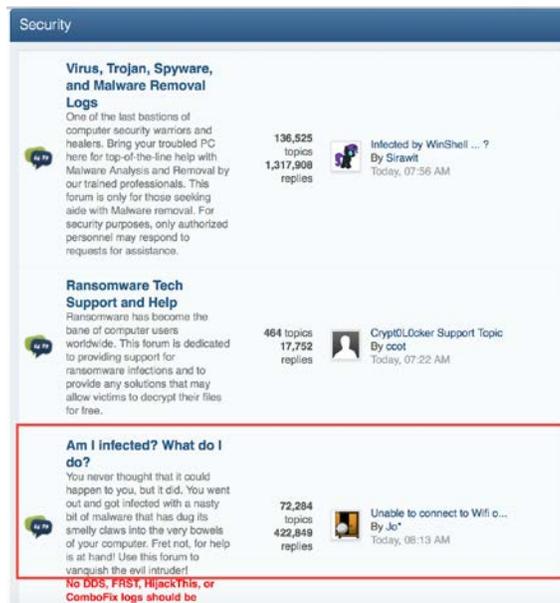


FIGURE 1 - "Am I infected? What do I do?" Forum within Security Topic.

As shown in Figure 1, it includes 72,284 topics and 422,849 replies. A number of topics are selected using random sampling. Two common methods are usually used to create the edges list of a forum. First, edges list is created by connecting every pair of authors who post under the same topic to one another. The second approach is to connect the author of the first post to every author of other posts. After successfully collecting the forum data, the former approach is used to build our edges list in which for each topic every pair of authors is connected to one another. The resulting graph is an undirected one. A list of 7,133 edges is collected using the selected topics which are big enough to get promising results towards the project objectives.

Social Network Dataset: Facebook

The Facebook dataset is obtained from SNAP [3] data collection. This dataset as shown in Figure 2 consists of friends' lists from Facebook which is collected from survey participants using this Facebook app. It includes friends' lists of 10 users which lead to a total of 4,039 users [4]. The edges list consists of 88,234 undirected edges.



FIGURE 2 - Facebook Graph Showing Friendship Network.

ANALYSIS

Several social network experiments were prototyped in R programming language and the results are presented and discussed in this section.

Network Statistics

General Network Comparison

Some experiments were conducted in order to understand the structure of both networks. Table 1 contains the general information about both networks.

TABLE 1 - Network Statistics

Network	Diameter	Average Path Length	Largest Component	Edge Count	Node Count
Facebook	8	3.69	4,039	88,234	4,039
Forums	6	2.96	1,497	7,133	1630

Facebook is a connected graph with small average path length and small diameter relative to its size, whereas, Bleeping Computer forums consists of several connected components and has a larger average path length compare to Facebook.

Network Layout

Online communities are topic based and usually act as information systems. Users may participate in different topics and share their information or express their opinion on a subject with a comfortable amount of anonymity. Each sub-forum/topic/post can be viewed by a community. Compared to social networks where users tend to be spectators, in online communities, members voluntarily contribute by answering questions, requesting answers, offering advice, while taking on various roles, such as leaders, facilitators, and observers. Most importantly, all these provide a sense of belongings, which is an essential need for humans who are social entities. Unlike social networks, online communities preserve the information that was shared and at any time a user can access and participate in different topics without much difficulty. As a result of such dynamics, online communities tend to have a nested layout and often consist of several isolated communities. Figure 3 shows the nested network layout of Bleeping Computer forums and Facebook which depicts the spider-web-like structure of social networks.

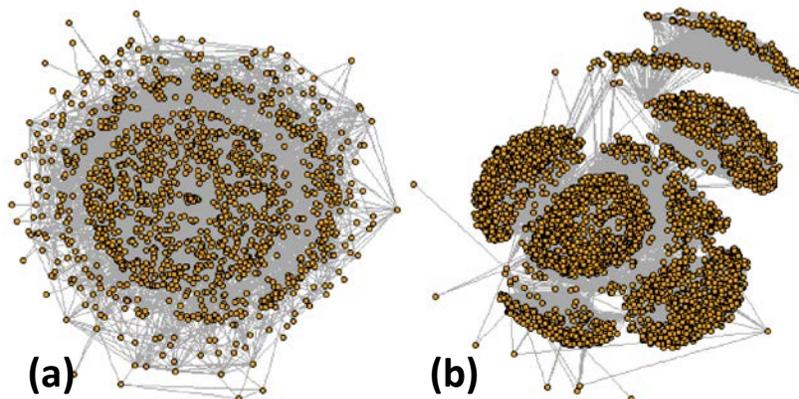


FIGURE 3 – Network Layout for (a) Bleeping Computer Forum; (b) Facebook.

Community Detection

Several simulations were also implemented for community detection. As shown in Figure 3, Facebook contains several communities with interconnections and overlaps while in Bleeping Computer forums, can be observed that several isolated communities which correspond to topics that not many people contributed.

Advanced Network Analysis

Degree Distribution

Analyzing degree distribution in complex networks usually leads to many useful insights. Several experiments were performed on both networks to investigate the degree distribution and the relationship between node degree and its neighbor's degree. Figure 4 shows the degree distribution in both networks.

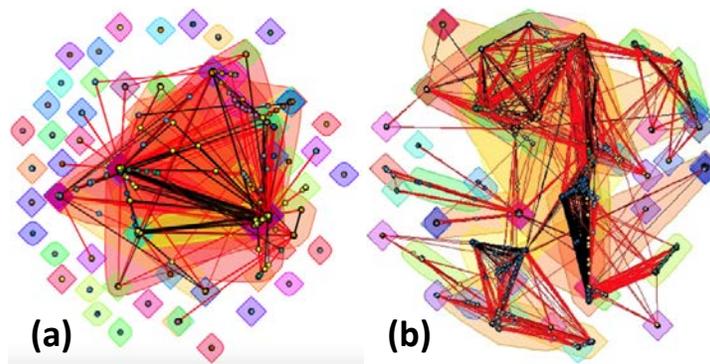


FIGURE 4 - Community Structure for (a) Bleeping Computer Forums and (b) Facebook.

Both networks show scale-free distribution. It is expected to observe few nodes with high degrees and several nodes with lower degrees considering the nature of connections in both online forums and social networks. A number of nodes with very few connections are more in online forums. This outcome is justifiable by understanding the nature of connections. Some people in online forums have few questions and after finding the answer they don't participate in the forums anymore. This behavior is less apparent on Facebook, as most people have a relatively big community of pre-established connections.

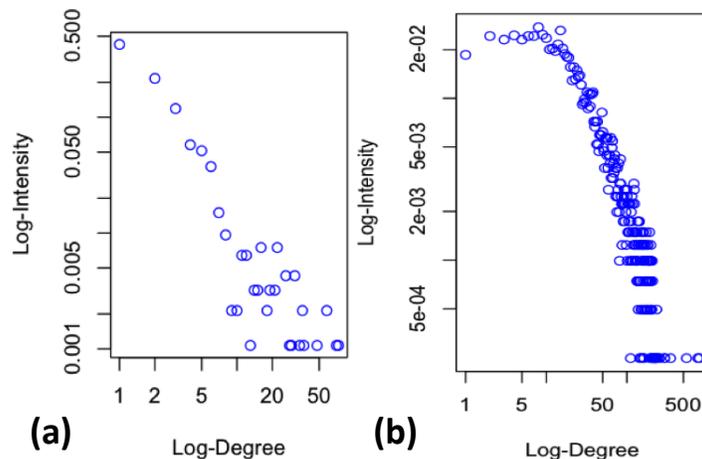


FIGURE 6 – Log-Log Degree Distribution for (a) Bleeping Computer; and (b) Facebook.

Figure 6 depicts the difference between these two networks. In Bleeping Computer network, there are multiple low-degree nodes that are connected to high degree nodes, whereas, in Facebook, the nodes with high degrees tend to form connections with other relatively high-degree nodes. This Figure shows the importance of high-degree node also known as active users in online forums and the positive and influential role that they can play in spreading information and affecting other users' behaviors.

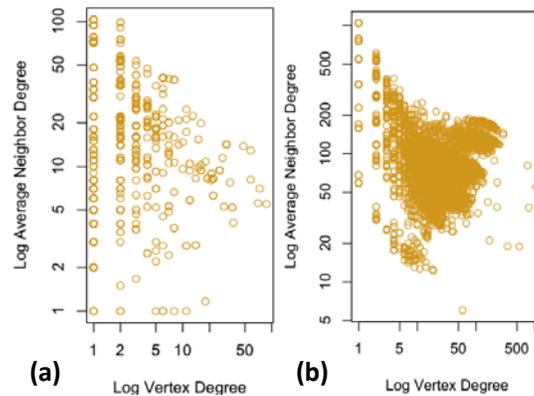


FIGURE 6 – Log Average Neighbor Degree for (a) Bleeping Computer; and (b) Facebook.

Information Diffusion

While investigating the best methods for communicating with customers and product marketing, there is a need to consider multiple aspects of human dynamics that can affect information diffusion in a social network. Understanding the factors that affect decision making in the social network can lead to developing a comprehensive model that can predict and evaluate human dynamics accurately. Although with respect to the presented data, some of these factors cannot be measured, their hidden role in the prediction and evaluation needs to be considered.

Limited Attention

Human attention has a cognitive limit on number of social concepts that he/she can sustain [5]. Numerous studies verified the existence of limited attention in social networks. [6, 7] investigated the collective attention on Digg.com and modeled its delay with a single factor. Their model suggested that attention fades on a natural time scale. The emergence of social networks led to rapid grow of information and affected the balance between increasing supply of information and our finite attention. In this context, there is a competition of ideas, information, and a potential decaying process in which the information is not part of the limited attention. The competition between data and limited attention is most effective in Facebook data. Considering the topic-free nature of this social network and increasing amount of information shared/produced by each user per day, the competition between knowledge is very high and the decaying process is much faster. Therefore, for marketing purposes it is needed to take into account the mentioned complexities and come up with strategies to prevent decaying of information. On the other hand, online forums such as Bleeping Computer store the information and knowledge more accessible and organized. The topics guarantee that users participating in them are invested in the topic.

Social Influence

The concept of *Social Influence* is very well known in social network analysis. It is usually referred to popular people or people who are able to persuade others to follow a trend as users with high social

influence. Although there's no unified definition for social influence, degree, information forwarding activity, PageRank scores, and some other network features can be used to measure the social impact. Vertex degree is the best measure to evaluate the social influence of nodes in this study. In Facebook data, high degree nodes are more likely to be popular and persuasive since the connections are based on trust and existing relationships. From a business point of view, if a company can persuade a high-degree node to advertise or share content on behalf of them, not only they reach a broader audience but also it's more likely that other connections accept and adopt the shared information. In Bleeping Computer Forums, high degree indicates an active user that has participated in many discussion across different topics, which also can suggest he/she is a knowledgeable and influential member. The existence of isolated communities in online forums also adds more value to the social influence of high-degree nodes, as they are more likely to form a link with these isolated communities.

Epidemic Models

Communication dynamics can be captured by epidemic spreading. Very similar to a virus or an infection transmission through the population, a piece of information can spread from one individual to another using social connections. Unlike infectious disease spreading, social connections have many free and hidden parameters that need to take into account in order to capture the true process of information diffusion [8]. Hence, it is essential to consider the mechanics of social networks and online forums in order to gain a better understanding. For instance, the forgetting factor should be considered in Facebook. As times passes, the shared information is more likely to be forgotten within the network. On the other hand, online forums preserve the information by organizing them based on specific topics. But accessing that information depends on users' activity and attention focus. Therefore, the prominence of forgetting factor is negligible [9].

The model consists of three compartments— S for the number susceptible, I for the number of infectious, and R for the number recovered (or immune). SIR simulation was implemented on data in order to capture the information diffusion in both networks. Immune nodes were replaced with forgetting factor for Facebook data. For Bleeping Computer Forums, immune and susceptible nodes were removed and the simulation was run. Figure 7 shows the spread of information in the online forum. It is important to notice that although there is no preventive factor (such as forgetting or attention limit) the spread fails to cover all nodes in the network. This is of course because of isolated communities and their lack of communication with other participants. The information passes to an individual also should be considered if and only if that individual is active in the forums. The result of SIR simulation on Facebook data is shown in Figure 8. The idea or information dies down in the network without continuous marketing and advertising because of forgetting factor and limited attention. The excessive amount of information shared on daily bases on social networks like Facebook increase the competition between ideas by filling out the attention limit and increase the forgetting process.

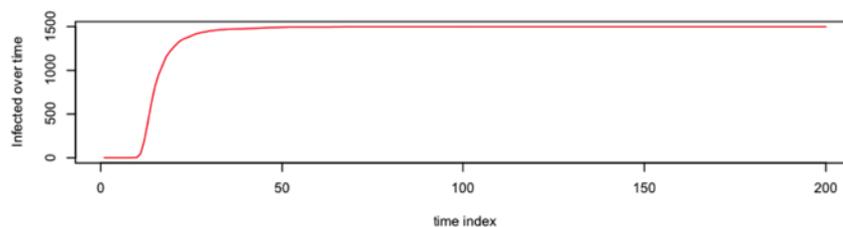


FIGURE 7 - SIR Simulation on Bleeping Computer Forum Data.

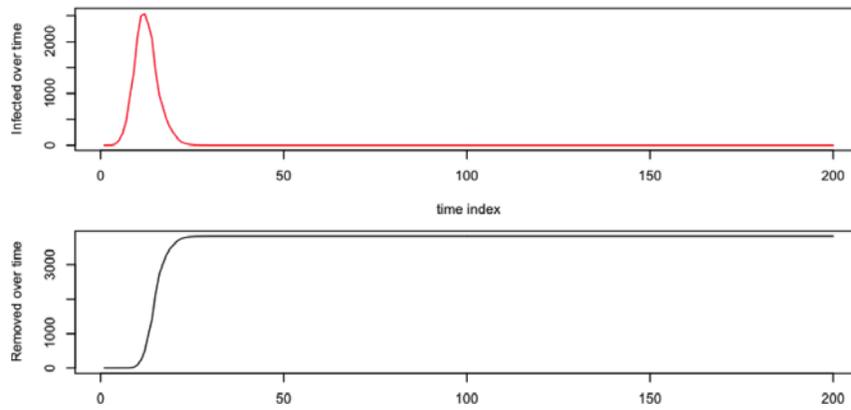


FIGURE 8 - SIR Simulation on Facebook Data.

CONCLUSION

These communities are segmented groups from a business perspective, with focused topics, such as user experience, improvement recommendations, complaints, and technical problems. It is always more productive to communicate with a small segmented audience than to a large group of people where most of the discussions and comments are noise and therefore with little value. However, the downside of online communities, such forums, is sufficiently large user basis. Therefore, it is not uncommon to see in Figure 5(b) that there can be many small posts with only a handful of members. Facebook has become almost essential for everyone, and most people are on Facebook today. So can we leverage such a massive user basis of a face, and incorporate that into the online communities? The answer is YES. Facebook has a feature called *Groups* where users can create groups around certain interests, topics, theme, products, and etc. Similar to the forums, group members can post questions, answer questions, share links, and updates. Another similar feature from Facebook is called Community Pages, where contents are populated with links from Wikipedia and related post from users. Since its launch in April 2010, people have been debating its functionality. Some criticize its privacy issue, while others praise it being one stop to collect all the user feedbacks. However, the downside of this, as with all the information obtained from the social network, is that large volume also comes with large amount of noises, where it would require a huge amount of resources to filter and monitor information. This is where online communities shine for being segmented and focused on specific topics.

The future direction of this research can be summarized as the following tasks:

- *Topic Detection in Clusters* - Within each community (cluster) in a forum, LDA topic detection technique can be used to detect the dominant topic in hand. This could be significantly beneficial to companies since they can find the most popular topics and advertise their topic-related products specifically within that community/forum. Moreover, this may lead to innovation to build a product to address the common concerns/issues of the users once they are revealed within the forum.
- *PageRank Score and Social Influence* - Undirected edges list for the forum dataset was created in this study. Directed edges (e.g. connecting the author of the first post under each topic to other authors on the same topic) can also be created and used to measure PageRank scores for each node/author. This leads to the identification of the most influential members more reliably. From a business point of view, convincing these influential and prominent members to advertise a specific product could lead to better spreading the idea/product information. Eventually, this results in an increase in a number of the customers.

REFERENCES

1. Easley, D. and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. 2010: Cambridge University Press.
2. computer, B., *Bleeping computer*, 2016: <http://www.bleepingcomputer.com/forums/>.
3. Leskovec, J. and A. Krevl, *{SNAP Datasets}:{Stanford} Large Network Dataset Collection*. 2015.
4. Leskovec, J. and J.J. Mcauley. *Learning to discover social circles in ego networks*. in *Advances in neural information processing systems*. 2012.
5. Dunbar, R.I., *The social brain hypothesis and its relevance to social psychology*. *Evolution and the social mind: Evolutionary psychology and social cognition*, 2007: p. 21-31.
6. Huberman, B.A., D.M. Romero, and F. Wu, *Social networks that matter: Twitter under the microscope*. 2008.
7. Huberman, B.A. and F. Wu, *The economics of attention: maximizing user value in information-rich environments*. *Advances in Complex Systems*, 2008. **11**(04): p. 487-496.
8. Anderson, R.M., R.M. May, and B. Anderson, *Infectious diseases of humans: dynamics and control*. Vol. 28. 1992: Wiley Online Library.
9. Bailey, N.T., *The mathematical theory of infectious diseases and its applications*. 1975: Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.