Original Paper

Rahul Gopalsamy, Graduate Student, University at Buffalo
Alexander Semenov, Ph.D., Postdoctoral Researcher, University of Jyväskylä
Eduardo L Pasiliao, Ph.D., Program Director, Air Force Research Laboratory
Scott McIntosh, Ph.D., Associate Professor, University of Rochester
Alexander Nikolaev, Ph.D., Assistant Professor, University at Buffalo, and Docent, University of Jyväskylä

Establishing Engagement as a Driver of Growth of Online Health Communities

Abstract

Background: Emerging research on tracking/nurturing the growth of online communities hypothesizes that growth is driven by network externality effects, wherein every user-generated contribution increases the value of a community for potential new members. The recently introduced concept "engagement value" and metric "engagement capacity" offer a means of quantitatively assessing the ability of an online platform to exploit the network externality effects. If the claim that higher engagement leads to accelerated growth holds true for online health communities, then engagement tracking should become an important tool in the arsenal of managers, enabling them to assess the "power" of their communities, identify the most engaging users – those critical to success, – and design interventions to propel growth in calculated ways**.**

Objective: This study sets to test the hypothesis that engagement *causes* Online Health Forum (OHF) growth over time, by answering the following questions.

Q1: Is there a causal relationship between engagement and the reach of an OHF?

Q2: Does "delurking" tend to occur at a higher rate when engagement increases?

Q3: Is the engagement value delivered by one forum user to another associated with the development of a personal bond (closer virtual relationship) between them?

Q4: Do the successes in engaging peers motivate a user to remain active in an OHF?

Methods: Data were collected from large OHFs between years 1999 and 2016. Longitudinal data analyses were conducted by evaluating forum user *engagement* metrics over time. Two-way causality effects between engagement value and the metrics evaluating OHF growth were analyzed using Granger Causality tests. User activity metrics per week were correlated with engagement metrics.

Results: Using observational data, positive answers were obtained for questions Q1-Q4. A one-way causal relationship was observed between the OHF engagement value and reach ($p=0.02$). A two-way causal relationship was detected between the engagement value and delurking, with further analysis indicating that the former is more likely to cause the latter ($p<0.001$ with lag 2; for the reverse hypothesis, $p=0.01$ with lag 2). The users who engage each other more were found more likely (up to 14 times) to develop personal connections. Finally, it was found that the more

engaging an OHF user was in a given week, the more likely (up to two times) they were to remain active in the following week.
Conclusions: This study supports the hypothesis that engagement is a measurable driver of OHF growth**.** Thus, engagement tracking can be used to gauge OHF success; engagement value fluctuations and changes in the activity of most engaging users – who perhaps can now be *defined* as superusers – can inform OHF managers of upcoming turns of events and enhance their ability to identify, support and rely on superusers in calculated ways.

Keywords: Online Health Community; Online Health Forum; Engagement; Engagement Capacity; Superusers; eHealth; Social Network Analysis

Introduction

Background

Online Heath Forums (OHFs) enable computer-mediated communication on health and health-related issues through the Internet. People use OHFs to seek emotional support, exchange information, ask for help or simply become part of a community of individuals with similar values in life [1]. Tracking and nurturing the growth of OHFs, which have emerged as a means of people-driven health care, have been on the research agenda for many practitioners [2]. It has been observed that typically, only 1% of OHF users generate most of the content; meanwhile, 90% of the users are passive readers who rarely leave traces of participation [3]. By identifying and encouraging the most prolific and engaging content contributors, OHF managers can become more successful at keeping their pro-health online communities growing, and thereby, achieving a greater impact on people's health [4].

The Network Effect Hypothesis

A positive network effect is a phenomenon initially studied in economics [5]. It explains the mechanism behind the process where the value of goods or services tends to increase as more consumers begin to use them: e.g., a sale of each additional unit of goods may increase the value of the goods through positive *network externalities*. On an online platform, such an effect may occur when its users author new content [6]. This proposition motivates one to attempt to quantitatively describe the role that network externalities may play in the growth of online communities. To this end, the term "engagement" has been recently adopted, with cooperative game theory-based methods for measuring it [7].
Adapting the "*reach*" metric (from the RE-AIM program evaluation framework [8]) as the key dimension of impact of any online platform, Nikolaev et al. [7] suggest that reach may be driven by engagement. Web-assisted pro-health interventions, in particular those offered via OHFs, are known to have positive *effectiveness* [9,10]. In the same context, *reach* is defined as the ability of an OHF to serve many users by attracting *new* users and retaining them. In turn, *engagement* is defined as the degree of involvement of *existing* users in the platform. The key premise here is that OHFs grow through their existing users: every contributed post (that is responded to) fosters user "bonding" and enhances the positive network externality effect.

Objective

Nikolaev et al. [7] hypothesized that the internal growth of an online platform (achieved through added user-generated content) increases its utilization leading to the external growth (appealing to potential new users): i.e., *higher engagement leads to higher reach*. However, this theoretical argument still has to be empirically validated, particularly in application to OHFs. This paper reports on an empirical investigation of the temporal dynamics of engagement-based quantities, with the aim of investigating the relationship between engagement and growth of OHFs. More specifically, this study's research questions were:

Q1: Is there a causal relationship between the engagement and reach of OHFs?

Q2: Does "delurking" tend to occur at a higher rate when engagement increases?

Q3: Is the engagement value delivered by one forum user to another associated with the development of a personal bond (closer virtual relationship) between them?

Q4: Do the successes in engaging peers motivate a user to remain active in an OHF?

This paper also seeks to establish if the evaluation of *engagement capacity* can help refine our understanding of the patterns of contributions made to a forum by *superusers* [6]. Answers to these questions can help pave the way for the development of calculated interventions for tactically growing OHFs, informed by user engagement quantities as key indicators of OHF success.

Methods

Foundations of Engagement Tracking

The value of an online platform or forum is generated cooperatively by its users, as they communicate (i.e., they author posts); thus, each user deserves a "credit" for contributing to the overall value of an OHF (i.e., the content it hosts, and consequently, the value it brings to the society) [7]. The metric "engagement capacity" is designed to quantify the ability of users to engage each other in contributing (more) content, under the assumptions that each generated post, submitted in response to prior posts, increments the OHF overall engagement value by one unit and that the credit for this unit value is shared between all the contributors in the thread that attracted the post. By monitoring the total engagement value that an OHF generates, together with the metrics of the OHF growth such as new user count, we set to test if the former causes the latter.

Further, just as OHF users differ by their needs (e.g., in seeking information vs. emotional support [11,12]), some users may be more successful in engaging a certain set of peers over others. The metric "targeted engagement capacity" takes this into account and quantifies the total credit allocated to user *i* for successfully engaging user *j*; note that engagement-based relationships are *directed* and non-symmetric. We thus set to check if targeted engagement can predict the formation of personal connections between users. To this end, the development of personal connections is monitored over time, by tracking the events where users post personal notes on each other's' walls, comment on each other's statuses or follow each other's journals.

Engagement Quantification

The *engagement capacity* of a forum user is computed as their share in the value of all the OHF threads, based on all *engaging subthreads* – sequences of exchanges of OHF posts succeeded by at least one additional post. Per the logic of *cooperative game theory* [13-15], OHF users can be viewed as *players* that play *games* on *k-coalitions* – connected ordered sequences of player *appearances* [7]. In the context of a forum of N users, let $i(T,k)$, $k=1,2,...,K$, denote the position of the k-th appearance of user $i \in H(T)$ in k-coalition $T \in \Omega^K(N)$, with $\Omega^K(N)$ denoting the set of all *threads to which* any user contributes at most K times. With the *value function* $\Delta_v^*(T)$ taken as a total number of post exchanges immediately succeeding such *engaging* subthreads $p \in P$ that have the same membership, size and structure as k-coalition $T \in \Omega^K(N)$, the engagement capacity of user $i \in N$ is computed as

$$\Psi_i^{K\text{-}\alpha}(N,v) = \sum_{T \in \Omega^K(N), i \in H(T), k=1,..,K} \Delta_v^*(T) \frac{\alpha^{|T|-i(T,k)}}{|T|! \sum_{j=0}^{|T|-1} \alpha^j}.$$

The coefficient α, taken to be 0.8, ensures that the author of a post, which directly precedes a responding post, receives more credit for attracting this response than the author(s) of the post(s) higher up in the thread. The calculation of engagement capacity value is fast and can be performed dynamically in real-time; the calculation of *targeted engagement capacity* values is performed similarly [7].

Data Collection

The data were collected from one of the biggest active and freely accessible online pro-health platforms. Each of its (approximately) 200 predominantly English-speaking communities is devoted to a specific health-related topic. The platform users interact through discussion boards, authors' personal journals, and by posting notes and status updates on personal pages. A web crawler was used to collect the data from "Heart Disease", "Diabetes", "Substance Abuse", "Fitness", "Depression", "Heart Rhythm" and "Anxiety" OHFs. There were approximately 200K threads, although some of the threads did not generate responses (see Table 1). The timeline was broken into fixed intervals for performing longitudinal analyses. Within each time interval, engagement capacity, targeted engagement capacity between users, number of new users, and delurking and friendship-building events were tracked.

Table 1. The statistics (counts) for the observed OHFs, rounded to thousands (K).

OHF Community Discussion Topic	**Total Messages**	**Total Threads**	**Threads of >1 posts**	**Contributing Users**
Heart Disease	104K	32K	23K	31K
Diabetes	14K	4K	3K	5K
Substance Abuse	760K	81K	78K	52K
Fitness	19K	5K	4K	9K
Depression	58K	13K	12K	14K
Heart Rhythm	90K	18K	17K	15K
Anxiety	166K	34K	30K	33K

Data Analyses

The following procedures were carried out to answer the questions Q1-Q4. For Q1, the engagement values were calculated for each day over 17 years based on the observed communication on the platform (all the seven OHFs lumped together). In this calculation, each new post, generated *in response to* any prior post, contributed the engagement value of one to the overall *engagement value*. To quantify the *reach*, the number of new users joining the platform in each day was recorded. Answering Q2 required the tracking of "delurking" – the phenomenon (event) where a user breaks the habit of passive reading by adding content to a forum, thereby extending its reach [16]. The OHFs examined in the present study did not require its visitors to register for passive reading; therefore, for Q2, delurking was taken to occur whenever a registered user contributed a post after at least a month of inactivity. For Q3, a personal virtual connection is assumed to have formed (been initiated) when a user posts a note on a peer's profile front page for the first time. Targeted engagement capacity values for user *i* engaging peer *j* were calculated for all user pairs (*i*, *j*). A binary indicator function was used to track the presence of personal (virtual) connections resulting from user interaction in an open forum – the function took the value of one whenever user *j* posted a note on *i's* wall. In order to answer Q4, the engagement capacity for each user over fixed time periods (one week long) was computed. A binary indicator function was used to track whether the user was active (i.e., made contribution to the forum) in the following week: e.g., the engagement value for user *i* was computed for week *t* and an indicator function was used to record whether user *i* contributed to the forum through posts or comments in week *t*+1.

Granger Causality Test

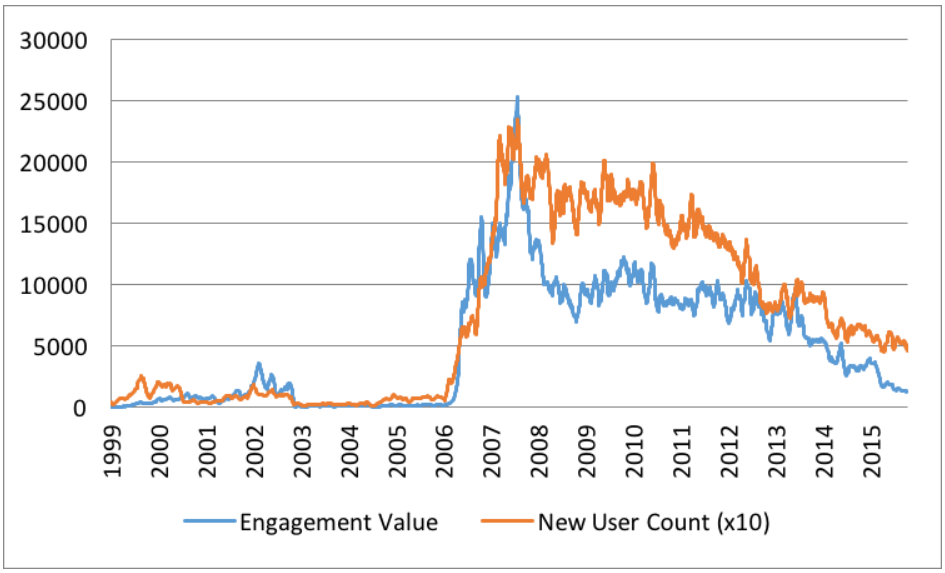
Questions Q1 and Q2 were investigated using Granger causality testing, a widely used tool for the analysis of joint temporal dynamics of multiple observed quantities (here, engagement value, new user count and delurking). Per the theory of Granger causality, one signal (X_1) is said to *Granger-cause* another signal (X_2) if the past value(s) of X_1 contain information that can predict X_2 better than the information contained in the past value(s) of X_2 would do alone [17]. In general, detecting a one-way Granger causation is desired to definitively establish the nature of a cause-and-effect effect between two temporally varying quantities. The two-way Granger causality test is typically conducted to first detect any cause-and-effect relationship in both directions. If one observes that X_2 Granger-causes X_1, while X_1 Granger-causes X_2, then the further analysis is done to determine the direction in which the causal effect is the strongest. In this paper, the OHF engagement value was taken as X_1, and new user count and delurking rate as X_2.

Results

Causal Relationship between Engagement and Reach

The Granger causality test performed between the platform's engagement and new user count indicates that engagement Granger-causes reach (p-value of 0.02 with lag 2); Figure 1 shows the temporal dynamics of both.

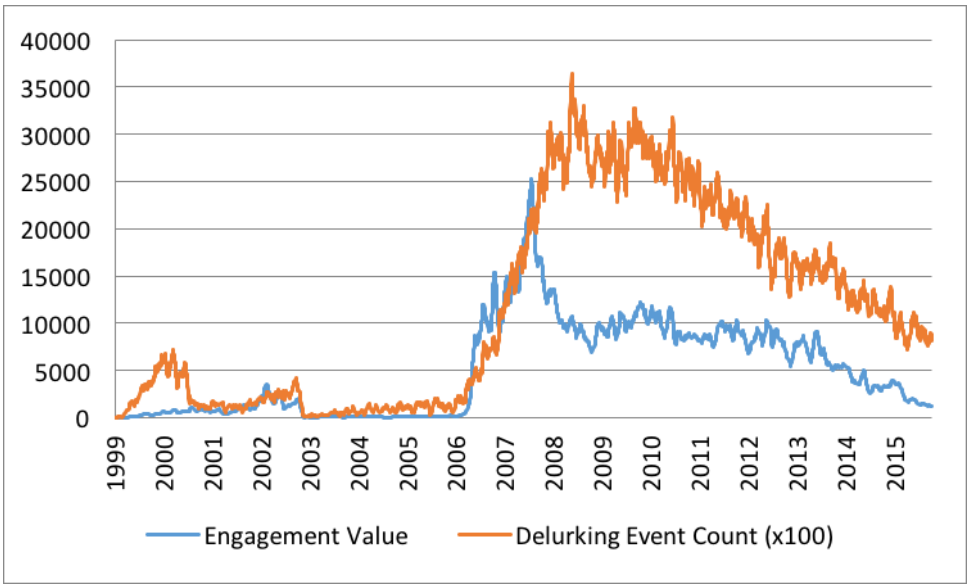
Figure 1. Temporal dynamics of the platform's engagement value and new user counts (with the latter scaled up by factor 10).



Causal Relationship between Engagement and Delurking

Figure 2 shows that the fluctuations in the rate of delurking follow the fluctuations in engagement value. The Granger causality analysis reveals a two-way causation, with the hypothesis of engagement Granger-causing delurking supported with the p-value of <0.001 with lag 2, and the hypothesis of delurking Granger-causing engagement supported with the p-value of 0.01 with lag 2. It is thus concluded that we have a stronger support of the claim that engagement causes delurking (than the other way around).

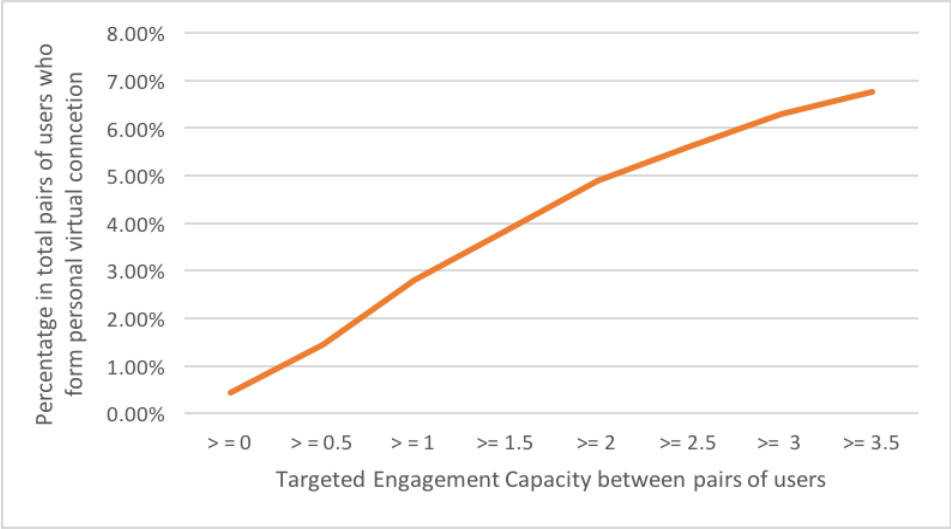
Figure 2. Temporal dynamics of the platform's engagement value and delurking event counts (with the latter scaled up by factor 100).



Targeted Engagement Capacity as a Predictor of the Development of Personal Connections between Users

A close-to-linear relationship is observed between the propensity of building personal (virtual) connections among OHF user pairs and the targeted engagement capacity delivered by one user in a pair to another user in this pair (see Figure 3): the more frequently User *i* manages to engage User *j* (prompting *j* to respond), the more likely *i* is to receive a posted personal note from *j* for the first time.

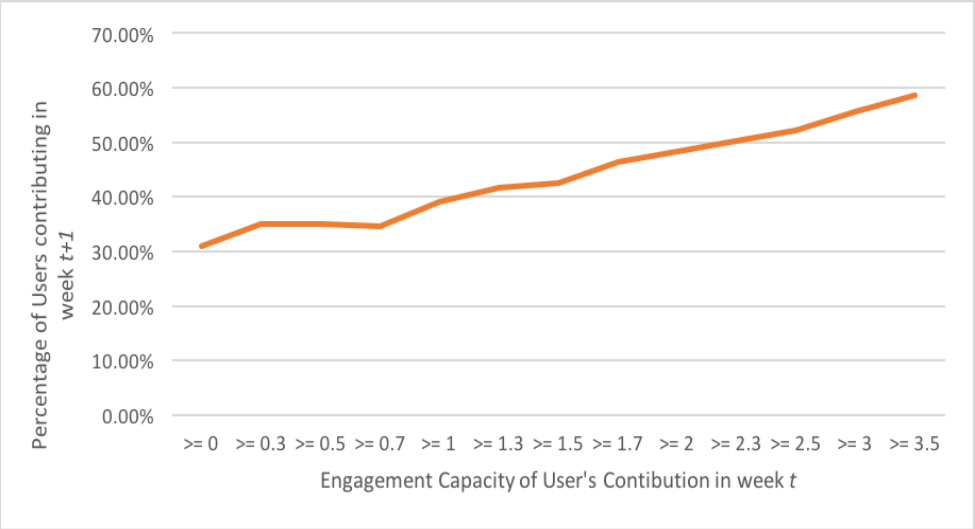
Figure 3. The probability of building a personal connection as a function of targeted engagement capacity between OHF users.



Engagement Capacity as a Predictor of Future Activity

A close-to-linear relationship is observed between the propensity of a user to stay active in an OHF and the engagement capacity of this user earned over the past week (see Figure 4): the users who manage to engage many peers are almost twice as likely to remain active contributors compared to the least engaging users.

Figure 4. The propensity of a user staying active in an OHF as a function of their engagement capacity earned over time.



Discussion

Principal Findings

The seven OHFs observed in this study were diverse, with a varied number of posts and users. They were all created at different points in time. Note that some of the users were simultaneously active in more than one forum. Based on these observational data, the hypothesis that higher engagement causes higher reach was supported, per the performed Granger Causality testing. Targeted engagement, as the measure of the amount of directed communication between a pair of users, was found to inform the development of a virtual connection between them. In the future, this metric can become an integral component of OHF thread recommender systems, where suitable threads can be recommended to those users who are more likely to develop personal connections through communication. Engagement capacity was also found to be a reliable metric for predicting the propensity of a user for contributing to the OHF. This suggests that OHF managers / moderators can focus on developing intervention mechanisms to increase engagement in calculated ways. Finally, the engagement capacity analyses can help refine our understanding of the patterns of contributions made to a forum by superusers, over a period of time [18]; Figure 5 shows that some of the most engaging users can be active for prolonged time periods, while others for shorter time periods. If the users with an innate ability to engage others, who can now be *defined* as superusers, can be motivated to sustain their activity, e.g., with suitable incentives, then OHFs can avoid the periods of low activity and maintain growth through user retention and addition.

Limitations

An engagement value of one was assigned to all the posts that generated responses. However, certain posts may be controversial, generating mixed responses, and may differ by topic and overall value. Not all posts might necessarily foster user bonding. Also, some users might occasionally spam the system by posting irrelevant content. Proper care should be taken to avoid or account for such threads (e.g., obvious outliers) before performing OHF engagement analyses. The latter issue, however, might not be acute with OHFs, since pro-health online resources are typically well moderated.

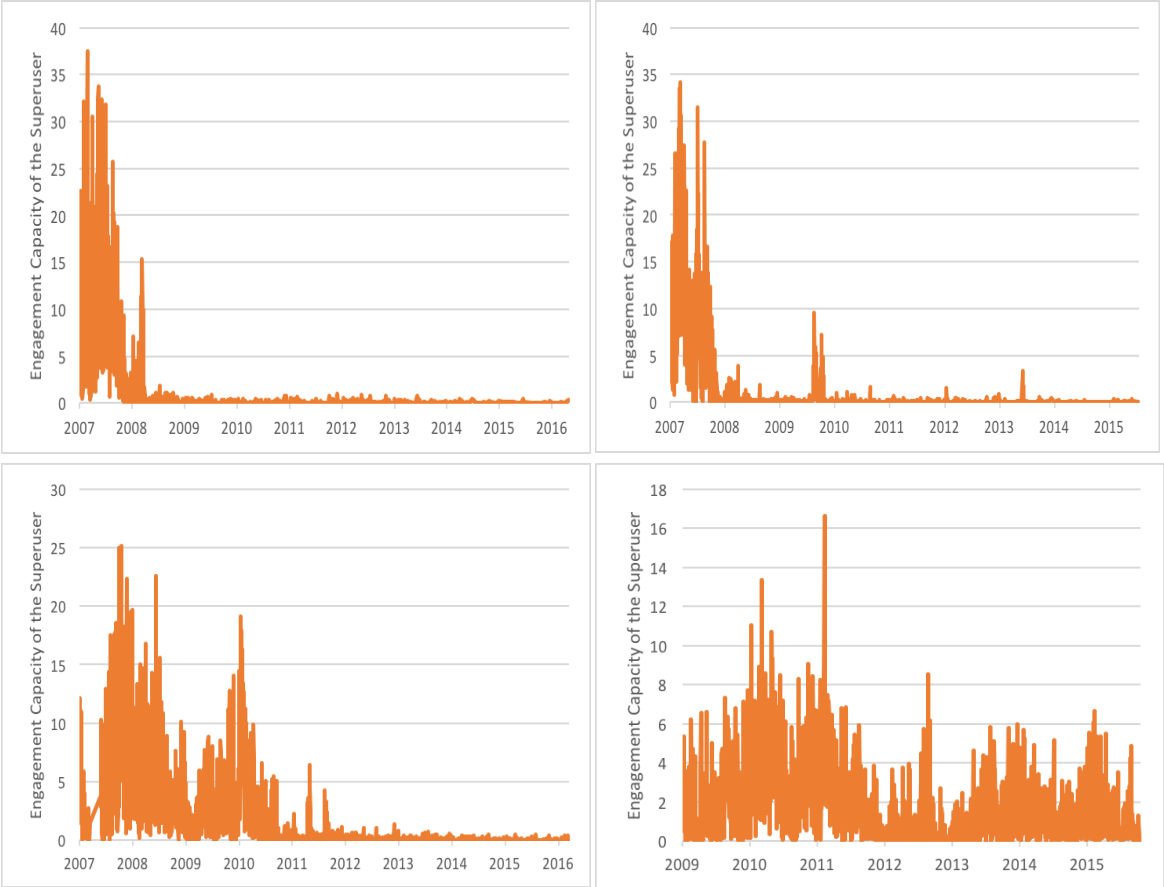
Comparison with Prior Work

Van Mierlo et al. [2] presented a five-step procedure for identifying superusers based on the number of posts initiated by users in OHFs; however, their method does not account for the variations in the number of responses generated by each thread. The present work meets this challenge by offering a new, quantitative way to identify superusers – as consistently most engaging users.

Conclusions

The present analysis supports the hypothesis that engagement can be viewed as a driver of growth of OHFs; the theory-supported engagement quantification provides a framework for systematically assessing a platform's health. The engagement value

Figure 5. Engagement capacity dynamics of four most engaging OHF users. The bottom two plots are for the users who have been active for longer periods of time, whereas the top two are for the users whose initially high activity then faded.



of an OHF can be used as an indicator of its (near-future) reach. The targeted engagement capacity metric can help forum managers / moderators predict, and perhaps, direct the development of virtual bonds between users. Engagement capacity analyses can reveal trends in superusers' activity over time, which can be useful for designing suitable intervention mechanisms targeting calculated growth: possible examples of such interventions include badge allocation, thread recommendation, targeted requests for users to contribute "featured posts", and perhaps, monetary incentives to superusers. Future research should aim to identify different types of superusers in a platform by assessing and understanding the nature of their contributions, in particular developing methods to predict user engagement capacity and explain what makes superusers engaging [19].

Acknowledgements

This work was funded in part by the NIH NCI Grant R01CA152093-01, NSF Award 1635611, Academy of Finland Grant 268078 (MineSocMed), University of Florida MMO Institute, and 2016 U.S. Air Force SFFP Fellowship.

Conflicts of Interest

None Declared

Reference

1. Swan, M. (2009). Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *International Journal of Environmental Research & Public Health*, *6*(2), 492-525.
2. Balatsoukas, P., Kennedy, C. M., Buchan, I., Powell, J., & Ainsworth, J. (2015). The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. *Journal of Medical Internet Research*, *17*(6).
3. van Mierlo, T. (2014). The 1% Rule in Four Digital Health Social Networks: An Observational Study. *Journal of Medical Internet Research*, *16*(2).
4. Stearns, M., Nambiar, S., Nikolaev, A., Semenov, A., & McIntosh, S. (2014). Towards evaluating and enhancing the reach of online health forums for smoking cessation. *Network Modeling Analysis in Health Informatics & Bioinformatics*, *3*(1), 1-11.
5. Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review*, *75*(3), 424-440.
6. van Mierlo, T., Hyatt, D., & Ching, A. T. (2015). Mapping power law distributions in digital health social networks: methods, interpretations, and practical implications. *Journal of Medical Internet Research*, *17*(6).
7. Nikolaev, A., Gore, S., Khopkar, S., & Govindaraju, V. (2016). Engagement capacity and engaging team formation for reach maximization of online social media platforms. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 225-234
8. Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *American Journal of Public Health*, *89*(9), 1322-1327.
9. Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*, *15*(4), e85.
10. White, M., & Dorman, S. M. (2001). Receiving social support online: implications for health education. *Health Education Research*, *16*(6), 693-707.
11. Wu, B., Jiang, S., & Chen, H. (2016). Effects of individuals' motivations on communications in online health forums. *Social Behavior and Personality: an International Journal*, *44*(2), 299-312
12. Lin, K. Y., & Lu, H. P. (2011). Why people use social networking sites: an empirical study integrating network externalities and motivation theory. *Computers in Human Behavior*, *27*(3), 1152-1161.
13. Shapley, L. S. (2016). A value for n-person games. *Contributions to the Theory of Games (AM-28)*, *2*, 307.
14. Sanchez, E., & Bergantinos, G. (1997). On values for generalized characteristic functions. *Operations-Research-Spektrum*, *19*(3), 229-234.
15. Nowak, A. S., & Radzik, T. (1994) The Shapley value for n-person games in generalized characteristic function form. *Games & Economic Behavior,* 6(1), 150-161.
16. Rafaeli, S., Ravid, G., & Soroka, V. (2004, January). De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on* (pp. 10-pp). IEEE.
17. Granger, C. (1988) Some recent development in a concept of causality. *Journal of Econometrics,* 39(1), 199-211.
18. van Mierlo, T., Voci, S., Lee, S., Fournier, R., & Selby, P. (2012). Superusers in social networks for smoking cessation: analysis of demographic characteristics and posting behavior from the Canadian Cancer Society's smokers' helpline online and StopSmokingCenter.net. *Journal of Medical Internet Research*, *14*(3), e66.
19. Wang, X., Zuo, Z., Zhao, K., & Street, N. (2015, October). Predicting User Participation and Detecting User Role Diffusion in Online Health Communities. In *Healthcare Informatics (ICHI), 2015 International Conference on* (pp. 483-483). IEEE.