

Towards evaluating and enhancing the reach of online health forums for smoking cessation

Michael Stearns · Siddhartha Nambiar ·
Alexander Nikolaev · Alexander Semenov ·
Scott McIntosh

Received: 10 December 2013 / Revised: 29 August 2014 / Accepted: 6 September 2014
© Springer-Verlag Wien 2014

Abstract Online pro-health social networks facilitating smoking cessation through web-assisted interventions have flourished in the past decade. In order to properly evaluate and increase the impact of this form of treatment on society, one needs to understand and be able to quantify its reach, as defined within the widely adopted RE-AIM framework. In the online communication context, user engagement is an integral component of reach. This paper quantitatively studies the effect of engagement on the users of the Alt.Support.Stop-Smoking forum that served the needs of an online smoking cessation community for more than 10 years. The paper then demonstrates how online service evaluation and planning by social network analysts can be applied towards strategic interventions targeting increased user engagement in online health forums. To this end, the challenges and opportunities are identified in the development of thread recommendation systems for

effective and efficient spread of healthy behaviors, in particular smoking cessation.

Keywords Social network analysis · Smoking cessation · Online forum communication · RE-AIM framework · Reach · Engagement · Intervention modeling

1 Introduction

Tobacco use is one of several individual modifiable health behaviors including poor diet, alcohol misuse, and physical inactivity, identified by the World Health Organization as leading risk factors for global disease burden (Lim et al. 2012; Narayan et al. 2010; Scarborough et al. 2011). The development of cost-effective public health initiatives, capable of reducing the rate of tobacco use at the population level, is of great importance. Web-assisted tobacco interventions (WATIs) represent one potential solution to this challenge with the expansion of Internet access globally leading an increasing number of individuals to turn to them in place of or as an adjuvant to traditional forms of treatment (Selby et al. 2010). They provide a cost-effective medium for delivering targeted social support to a wide audience (Norman et al. 2008). Models and methods capable of describing the dynamics of social interactions and influence within such communities are primed to become part of the health policy-maker's toolbox.

Intentionally created online networks for smoking cessation have existed for over two decades with newsgroups such as Alt.Support.Stop-Smoking, and websites such as Quitnet, Alt.Support.Stop-Smoking, StopSmokingCenter, and WebCoach, among others. (Cobb et al. 2011). These are dynamic, supervised systems allowing for various modes of communication (e.g., chat rooms, forums, private

M. Stearns · S. Nambiar · A. Nikolaev (✉)
Department of Industrial and Systems Engineering,
University at Buffalo (SUNY), Buffalo, NY, USA
e-mail: anikolae@buffalo.edu

M. Stearns
e-mail: stearns6@buffalo.edu

S. Nambiar
e-mail: s34@buffalo.edu

A. Semenov
Department of Mathematical Information Technology,
University of Jyväskylä, Jyväskylä, Finland
e-mail: alexander.v.semenov@jyu.fi

S. McIntosh
Department of Public Health Sciences, University of Rochester,
Rochester, NY, USA
e-mail: scott_mcintosh@urmc.rochester.edu

messaging), self-representation (e.g., personal profiles, blogs, journals), and affiliations (e.g., friend lists, private groups), ensuring that users can seek for social support of distant friends “like them” in real time (Norman et al. 2008). Recent research has established that modern online health communities for smoking cessation function as a form of treatment for their participants, significantly increasing abstinence rates and exhibiting similar levels of effectiveness as intensive face-to-face counseling (Shahab and McEwen 2009). Web- and computer-based smoking cessation programs for adult smokers were found effective: “in a random-effects meta-analysis of 22 eligible trials (9 web-based, and 13 offline computer-based interventions), the intervention group had a significant effect on cessation (relative risk (RR), 1.44; 95 % confidence interval 1.27–1.64)” (Myung et al. 2009). Similar successes were reported exclusively with web-based interventions for adolescents: RR, 1.40; 95 % CI, 1.13–1.72, where “the intervention group had a significantly larger cessation rate than that of the control group” (Crutzen et al. 2008).

Social Web (Web 2.0) technology alone does not guarantee a successful online community where members participate actively and develop lasting relationships (Iriberri and Leroy 2009). Adapting the well-established RE-AIM framework for sustainable interventions to online forums, the following five criteria can be distinguished (Glasgow et al. 2006). *Reach* is an individual-level measure of participation, referring to the percentage and risk characteristics of forum participants. *Effectiveness* is the degree to which the intervention achieved the intended outcomes, e.g., progress towards smoking cessation as a function of social interaction. *Adoption* refers to the proportion and representativeness of the settings that adopt an intervention, e.g., incorporating online community forums as a smoking cessation strategy. *Implementation* describes the extent to which an intervention is delivered as intended, e.g., online social forums have measurable interactions. *Maintenance* is the extent to which an intervention becomes routine, e.g., ongoing utilization and evolution of online forums. The literature focusing on individual-level measures has paid much attention to evaluating the effectiveness of online forums as a treatment of smoking, finding that both intra-treatment and extra-treatment social support are associated with increased rates of smoking cessation (Crutzen et al. 2008). However, little to no research has been reported on measuring reach, which is tantamount to user engagement in the context of WATIs.

The contribution of this paper lies in enhancing our understanding of user engagement as a key component of reach of online treatments, and in particular, social support environments and interventions (e.g., WATIs). The paper illustrates how and why the lack of prescriptive, as opposed to descriptive, models is growing into a serious challenge

in social network analysis today. By distilling the factors that influence user engagement, the present discussion looks for insights that could be applied to adapt thread recommendation research to the context of smoking cessation with the aim of enhancing the reach of online smoking cessation communities. In particular, the paper discusses how targeted thread recommendations can be employed to assist the less experienced health forum users in order to achieve higher levels of user engagement. The paper expands on the argument that social network formation models based on actors’ decisions do not allow for incorporating exogenous interventions, and as a remedy, proposes a strategy to explicitly model weak, acquaintance-type ties which with time can turn into strong, friendship ties. In order to motivate this line of inquiry, posting records from the Alt.Support.Stop-Smoking newsgroup are studied. The members of this online community, which was particularly active in the early 2000s, discussed topics pertaining to smoking cessation in the forum’s threads.

This paper reports the following: Sect. 2 encompasses a study of online health community Alt.Support.Stop-Smoking, and identifies the metrics reflecting the implications of user engagement; Sect. 3 details the challenges and opportunities surrounding the use of prescriptive social network modeling methods within smoking cessation communities; and Sect. 4 concludes the paper and offers directions for future research. It should be noted that the analysis and models presented in this paper are smoking-cessation specific and may not be immediately generalizable to digital health social networks addressing other conditions.

2 Data analysis

The Internet-based Alt.Support.Stop-Smoking forum, used in this study to distill measures to enable the monitoring of engagement patterns, is a Usenet newsgroup. Its structure is similar to other World Wide Web forums in that users can both read and post messages which are stored and available for viewing in a hierarchical tree. Usenet is a distributed system, accessible via Network News Transfer Protocol (NNTP) or, alternatively, using WWW front-ends such as Google Groups. The data analyzed in this paper were downloaded from a Usenet archive via NNTP in September, 2013, and inserted into a PostgreSQL database. Complex data analyses were then conducted using a developed Java code. The de-identified data analyzed in the present study were derived from retrospective publicly available data. Per IRB procedures at the University of Buffalo IRB, submission of a human subject’s research protocol for ethical board review of this type of investigation was not required.

The Alt.Support.Stop-Smoking forum activity examined in this section spans the ten-year period between 8/1/2003 to 9/15/2013. During this time, 438,136 posts were made by 8,236 unique users in 48,518 threads. Each of the 438,136 entries in the dataset corresponds to an individual post made by a user on the forum and comprises the timestamp of the post, the author’s unique forum username, and the thread to which it was submitted. Note that forums user data in the Alt. Support.Stop-Smoking dataset were limited to posting records and therefore the activity of only registered users with at least one post was analyzed. Owing to the difficulty in quantifying the difference in benefits between active posters and passive users (“lurkers”), user records of the latter were not included in the analysis.

The first step in analyzing the Alt.Support.Stop-Smoking data involved extracting and analyzing the aggregate forum metrics as a function of time. Figure 1a, b showcase moving averages for post and thread counts, and new and active user counts over the observed life of the forum, respectively. Users were considered to be active during a period if they were observed to have made one or more posts during the period. As observed in the trends, the initial rapid growth experience during the initial time period is short-lived; January, 2004, marks the relative peak of the forum’s activity, with 12,100 posts, 1,419 new threads, 1,490 active threads, 266 new users, and 439 active users. All of these measures are significantly higher than the overall averages observed in the dataset, where a typical month revealed 3,591 posts, 397 new threads, 445 active threads, 67 new users, and 165 active users.

Over the 9 years following the forum’s popularity peak, a gradual decline is observed in each of the four main aggregate forum metrics. In the last period covered by the dataset, 9/1/2013 to 9/15/2013, there are only five posts submitted to the forum, made by four active users, in three active threads. It is worthwhile to try to understand the factors that precipitated this decline. Moreover, there is a need to study whether the application of calculated external

pressures could enable the forum to reach more users over a longer period of time, thus increasing its cumulative health benefit. Accordingly, Sect. 2.1 offers user-specific analysis, enabling a deeper assessment of a typical forum user’s behavior.

The remainder of Sect. 2 is structured into the following subsections: Sect. 2.1 reports on user-level statistics; Sect. 2.2 studies how the gradually developed strong ties affect user behavior; and Sect. 2.3 classifies users by type and identifies user subgroups that could potentially benefit from engagement-enhancing interventions.

2.1 User-specific analysis

The consideration of historical forum aggregates alone does not fully capture the underlying user activity and engagement patterns. To provide a more complete picture, individual user data must be analyzed. The frequency graphs in Fig. 2a, b indicate that based on the Alt.Support.Stop-Smoking data, the forum content is distributed amongst a relatively small cadre of highly involved, “core” users rather than being distributed evenly throughout a largely homogeneous user base.

In the analyzed data, an average user contributes 53.2 posts during their forum lifetime (defined as the time elapsed between the user’s first post and their last post). The average user contribution is skewed by a small group of users who account for the majority of posts made to the forum. The top 1 % of users ($n = 83$) accounted for 194,498 of the 438,136 total posts (44.39 %), the next 9 % ($n = 741$) accounted for 193,498 posts (44.16 %), and the bottom 90 % ($n = 7,412$) accounted for 50,140 posts (11.44 %). The distribution of thread creators has a similar shape, with the top 1 % of thread creators accounting for 22,707 of the forum’s 48,518 total threads (46.8 %), the next 9 % accounting for 19,014 threads (39.2 %), and the bottom 90 % accounting for 6,797 threads (14.0 %). These measures indicate that the most active users are responsible

Fig. 1 **a** Moving average of posts and threads. **b** Moving average of new and active users

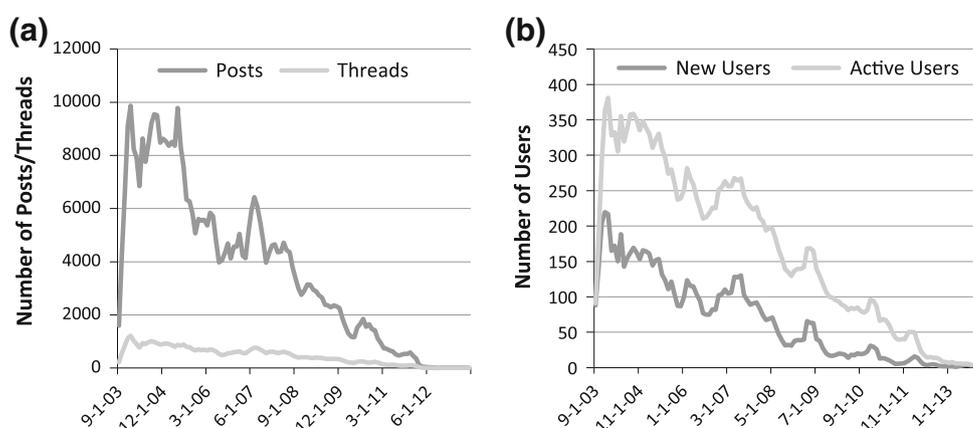


Fig. 2 **a** Distribution of the number of posts per user across the lifetime. **b** Distribution of thread creators for (>5 threads created)

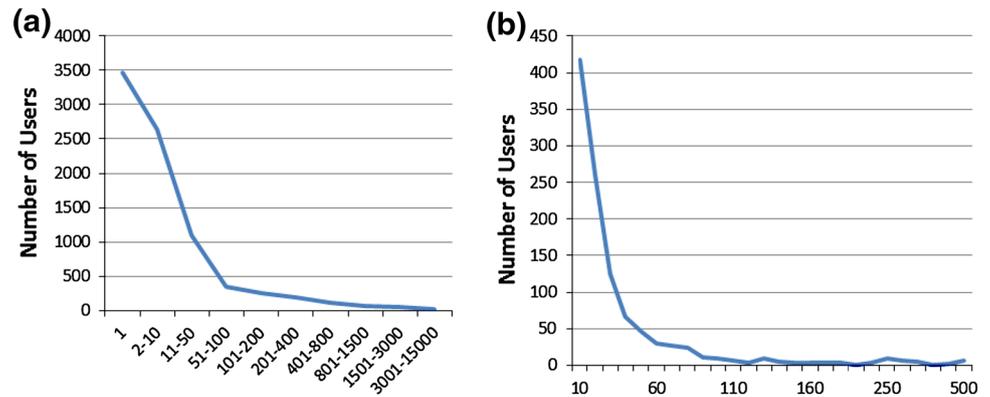
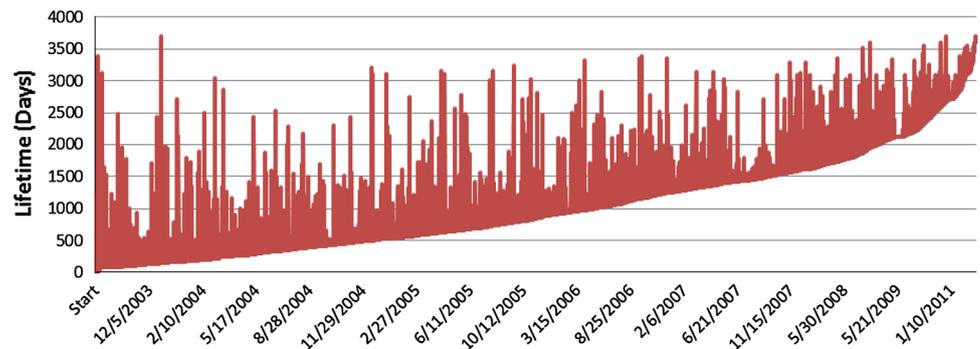


Fig. 3 User active lifetime bars



for a disproportionate amount of the forum’s overall content. Previous analyses of online communities have observed a similar phenomenon, referred to as the 1 % rule or the 90–9–1 principle, in which 90 % of actors observe and do not participate, 9 % participate sparingly, and 1 % create the vast majority of new content (van Mierlo 2014).

Overall, the majority of the users have short active forum lifetimes, with 4,557 users (55.33 % of the user base) having a lifetime of 1 day and only 634 users (7.7 % of the user base) having an observed lifetime over 1 year. Amongst the 100 most active posters, the observed average lifetime is 936.45 days. These observations imply there is a largely transient user base that enters and exits before having any opportunity for engagement. It is worthwhile to note that some short-term users were likely “bots” (automatic programs posting commercial ads) that must have been banned by the server’s administration.

Figure 3 indicates that, as the forum grew older and its active user base became more static, fewer new members joined and even fewer elected to remain engaged. A plausible explanation for this phenomenon lies in the increased difficulty faced by new users in trying to integrate themselves into an established community, with the majority of active members enjoying already established friendship relationships. Young (2013) indicated that when users start to think that they can no longer influence the

community, they will disengage. Failure to reverse such patterns of user-disengagement and barrier to entry, can lead to the death of the forum as the established user base dwindles and fewer new users join to take their place. Thus, it is necessary to determine how these friendship networks that initially served as a barrier to new users could instead be leveraged to engage them. To do so, the concept of friendship between users (how it arises and the influence it exerts on user behavior) must be defined.

2.2 Engagement related analysis

As the Alt.Support.Quit-Smoking forum does not explicitly report on friendship ties between its members, they must be inferred heuristically. Online friendships capture the emergence of mutual recognition between two persons, arising from their repeated interactions. In this vein, Rheingold (2000) describes online communities as “cultural aggregations that emerge when enough people bump into each other often enough in cyberspace”, while Preece (2001) defines them as “any virtual social space where people come together to get and give information or support, to learn, or to find company”.

Interaction instances, termed “weak ties” hereafter, between users were derived by analyzing posting patterns within threads. If a certain number of interactions or weak

ties are observed between a pair of users, it can be surmised that a strong tie is formed between them, i.e., that they have become friends. When User#1 submits a post to a thread within 2 days of User#2, it is (by assumption) interpreted as an instance of interaction between them. The gain in recognition arising from such interactions is divided into two sub-components: User#1 adds a weak in-tie from User#2 while User#2 simultaneously adds a weak out-tie to User#1. The reasoning for this division is based on an interpretation of how friendship germinates, being restricted to those pairs of users that demonstrate equitable and balanced interaction patterns. When the number of recorded in-ties/out-ties between a pair of users exceeds a specified threshold (10 of each in the present analysis), those users are assumed to have become friends in the sense that they can distinguish each other from the general user body and such recognition prompts them to communicate more. Following this logic, the analysis of the dataset’s posting patterns reveals the distribution of friendship ties between forum users (see Fig. 4).

As evidenced by user-based metrics, the forum’s user base is highly segmented. The user with the largest friendship network has 395 friends, with only four other users exceeding the friend count of 300 or more, and only 29 exceeding the count of 100. Unconnected, i.e., friendless users, comprise the largest segment of the forum’s user base ($n = 7,206$ users). These users were not reached, and therefore were not affected by the forum to the extent where their experience/thoughts/social support could be helpful to others.

Having defined strong (friendship) ties, the assessment of the influence that such ties might exert on user behavior can proceed. To this end, two research questions were explored: (1) Do friendship ties (or lack thereof) influence a user’s propensity to abandon the forum? and (2) Do friendship ties influence a user’s posting behavior in that the users are more likely to post in threads created by their friends as compared to those created by non-friends?

In order to answer the first question, active users during each time period (month) were divided into two groups:

those who elected to leave during that period and those who elected to remain active. Analysis revealed 12,064 instances of user “survival” and 8,236 instances of user “death” (forum abandonment). The average number of active friends for each user in these two groups was then determined. When the entire duration of the dataset was examined it is discovered that, on average, surviving users have 8.244 active friends while outgoing users had 1.165 active friends. These results indicate that the presence of an active friendship network is highly correlated with a user’s decision of whether to stay or leave, with the users having comparatively larger active networks being more likely to remain.

In order to answer the second question, data consisting of active threads and active users, for whom there existed at least one active thread created by a friend, were collected for each time period (month). The number of friend- and non-friend-threads to which users responded (among the active threads), and the number of posts made in each was then obtained for each user. Of 998,884 opportunities to post in a friends thread, 135,640 were used, with contributions submitted to 87,900 distinct threads. Conversely, of 3,487,854 opportunities to post to a non-friends thread, 239,150 were used, with 160,652 distinct threads receiving contributions. This corresponds to an 8.8 % probability of a user posting in a friend-created thread and a 4.6 % probability of posting in a non-friend-created thread. When the gross number of posts made within these threads is considered, the observed counts correspond to an average of 0.1358 posts made by an average user per friend-thread and 0.0685 posts made per non-friend-thread. This effect is not only statistically significant (which is not surprising, given the sample sizes), but it is, more importantly, practically significant.

2.3 Analysis of user engagement needs

In order to summarize and simultaneously provide a more in-depth analysis of user behavior patterns and engagement needs, the forum’s user base is provisionally divided into

Fig. 4 Distribution of friendship network sizes (>0 friends)

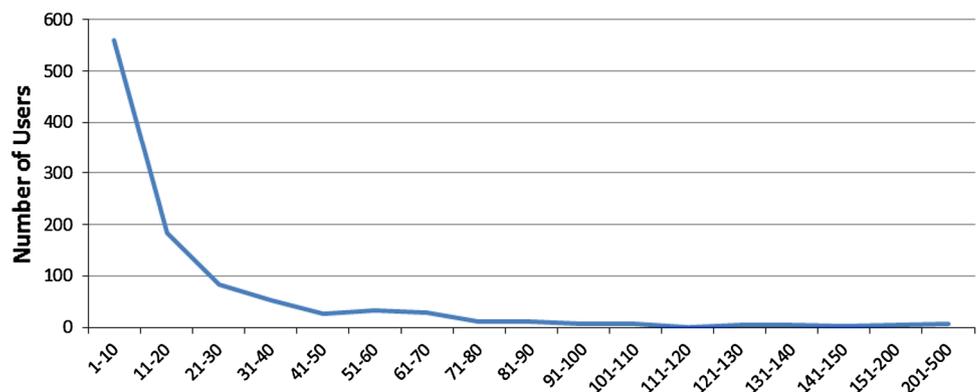
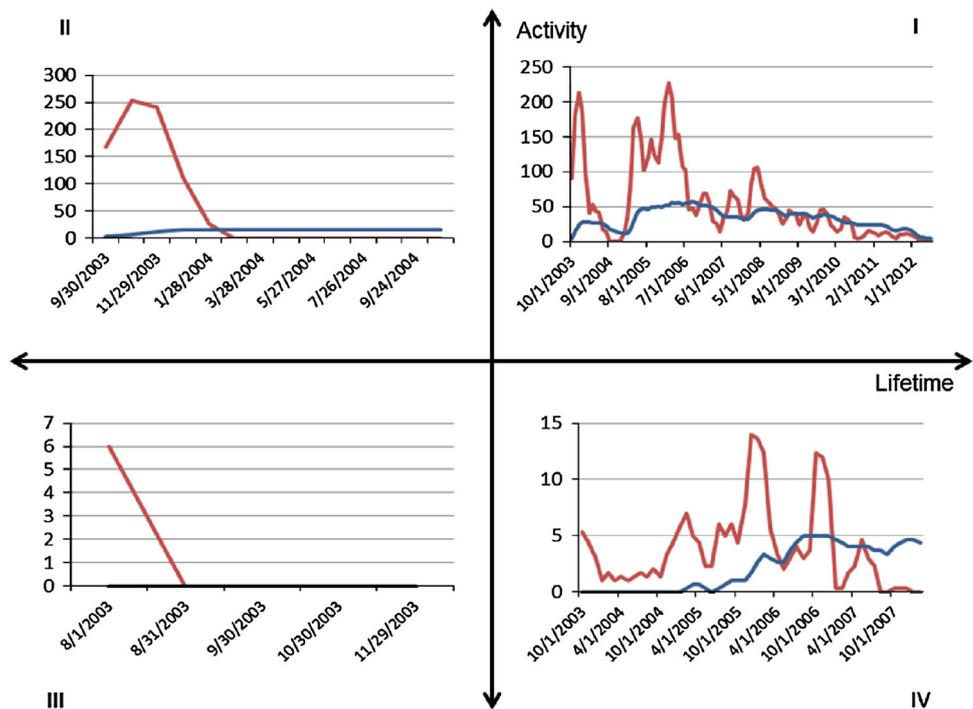


Fig. 5 Representative examples of different user types



four distinct groups. The drivers of the division are users' forum lifetimes and the observed levels of activity. Users may be generally divided into short-term users and long-term users, with each having two distinct subgroups. The four distinguished user types along with representative examples of engagement-specific activity patterns are shown in Fig. 5.

2.3.1 Short-term users

Quadrants II and III in Fig. 5 comprise those users having relatively short lifespans—frequently a week or less. Quadrant II and III users are differentiated from each other by their respective activity levels. Quadrant III users are those who join the forum, make a small number of initial contributions, and then leave for good. As shown in Sect. 2.1, such users make up a significant proportion of the forum's user base. Conversely, Quadrant II users post heavily immediately upon joining the forum, only to leave soon after.

Although it is impossible to definitively conclude the primary motivators for short-term users, research has suggested that they are composed largely of recent quitters seeking support while struggling with their quit attempt. In the analysis of an online smoking cessation community, Selby et al. (2010) found that seeking support and advice was the most common theme identified in first posts among both recent and longer term quitters. In their analysis of 2,562 first posts to an online smoking cessation support

group, approximately 54.7 % were made by individuals who had quit smoking within the past month, 8.9 % by those who had quit more than 1 month prior, and 24.9 % by those who had not yet quit smoking (Selby et al. 2010).

The analysis of posting patterns within the Alt.Support.Stop-Smoking community indicates that the typical user is narrowly focused, limiting their posting activity to a few number of threads, oftentimes their own. Of 8,236 unique users, 51.2 % (4,219/8,236) limited their posting activity to a single thread, and 73 % (6,009/8,236) to five or fewer threads. Considering the 4,219 users whose activity was confined to a single thread, 43.6 % (1,839/4,219) posted solely to the threads that they themselves had created, indicating that their primary motivator for participation is personal benefit.

Although the short-term users may have received the benefit of social support during their time on the forum, the failure to retain them as contributing members can be considered an overall community loss. By leaving the forum soon after having joined, such users will not "return the favor" by providing social support to other users in the future. This behavioral pattern may not necessarily be considered a disservice to the exiting short-term user. The literature is split on the significance of continued and active participation during quit attempts (Preece et al. 2004; An et al. 2008), although it could be viewed as a disservice to other current and future members who will not benefit from the user's experience and insights.

2.3.2 Long-term users

Quadrants I and IV comprise users with long lifespans, often many years. Quadrant IV users demonstrate relatively low activity levels and small friendship networks, but nonetheless chose to remain active for an extended period of time. These users are likely heavily topic-driven, primarily posting to threads that serve their immediate needs or pertain to a personal interest. Quadrant I users demonstrate sustained high activity over long lifetimes. They are sometimes referred to in the literature as “core-users” or “super-users” (Young 2013).

Core-users are responsible for much of the forum content, as illustrated in Sect. 2.1, and form the backbone of the community—exerting a disproportionate level of influence relative to their overall numbers (O’Neill et al. 2014; van Mierlo et al. 2012). Participation of core-users is motivated more by community factors than personal interest in specific topics. It has been found that many core-users are altruistic and truly serve the community: they are the first to greet newcomers and provide social support to other users. They may have benefited from the forum in the past, and are motivated to “pay it forward”. In previous analysis of an online smoking cessation forum, it was found that the majority of responses (>50 %) to new users’ first posts were made by members who had quit for a month or more, with only 1 % of first replies being made by members who had not yet quit (Selby et al. 2010).

Posting patterns observed in the Alt.Support.Stop-Smoking reflect the role played by core-users in the community. Lurkers are often hesitant to ask a question or seek support within an inactive community where they perceive the likelihood of a response to be low, with core-user activity giving lurkers the confidence to join in the conversation (Bishop 2007). As seen in Sect. 2.1, the rate at which new members posted to the community is positively correlated with the number of posts and active threads during that time—content for which core-users were largely responsible.

Core-users’ role as community ambassadors, typically being among the first to respond to newcomers, is another essential function. In the Alt.Support.Stop-Smoking dataset, 3,743 users started a new thread within 2 days of having joined the forum, with 45.0 % (1,686/3,743) receiving a prompt reply (within 3 h of their initial post). Initial responses to these threads were typically made by core-users, with the average post count and number of friends of first responders being equal to 1,681.18 and 71.39, respectively. Both of these values were significantly higher than the overall community averages of 53.2 posts and 2.505 friends, respectively ($p < 0.001$, $p < 0.001$).

Following an initial post, the prompt engagement of newcomers by core-users was found to have a significant

correlation with their future activity patterns. New users receiving a prompt reply to their first thread had an average lifespan of 114.52 days and an average post count over their lifespan of 93.36. Conversely, new users who did not receive a prompt reply to their first thread were found to have an average lifespan of 61.36 days and an average lifespan post count of 49.00. These differences are both statistically ($p < 0.001$) and practically significant, having 95 % confidence intervals of 22.4–66.3 days and 35.79–70.53 posts, respectively.

3 Directions and future considerations for increasing engagement in smoking cessation communities

This section builds upon the insights offered by Sect. 2, demonstrating how online service evaluation and planning by social network analysts can be applied towards strategic interventions targeting increased user engagement in online health forums. Calculated strategic management is essential for maintaining successful online communities where members actively participate and develop lasting relationships (Iriberry and Leroy 2009). Modeling the dynamics of interactions between core-users, regular users, and newcomers in online health forums would provide a technical foundation of modern pro-health engagement research.

There is a gap in the literature of prescriptive models capable of monitoring, controlling, and improving user engagement in online health forums. One avenue towards accomplishing these goals is through targeted recommendations of threads to users. Thread recommendation systems apply knowledge discovery techniques to match users to threads. Given the diverse interests and needs of forum users, coupled with the large amount of information that they must sift through on a typical forum, recommender systems present an essential tool for improving end-user retention and facilitating meaningful user interactions. Thread recommender systems serve to simultaneously satisfy users’ information needs by directing them to appropriate content, and their social needs by connecting them to other users within the community. There are a number of domain-specific considerations, not emphasized or even present in conventional thread recommendation tasks, which are essential for the development of effective health forum recommender systems.

In contrast to conventional online forums, the participation of users in online health forums is primarily motivated by a desire to give and/or receive social support (White and Dorman 2001). Friendships between forum participants play an essential role in the provision of social support within such communities. Reading and participating in forum threads leads users to encounter other members like themselves with whom friendships can be built,

thus enabling *personalized* support. Therefore, threads serve not solely as platforms for the dissemination of static content, but also as conduits for meaningful user interactions, with thread value being generated by and representing its participants. Within this framework, each thread can be viewed as a resource for introducing new user ties and strengthening existing ones. The mechanisms by which friendships form between users, and the manner in which threads can be employed to facilitate the process, are essential components of the emerging methodology for health forum thread recommender systems.

The ensuing subsections discuss adjustments to current paradigms that can lead to models capable of informing and controlling online forum user engagement. These subsections offer more focused discussion about the domain-specific challenges confronting thread recommendation systems in online smoking cessation forums, the use of social network structure as a means to motivate thread recommendation, and present a new paradigm for modeling actor ties within a social network to better capture the manner in which friendships between users develop.

3.1 Thread recommendation within an online smoking cessation forum

Research on thread recommendation systems is just beginning to emerge (Tang et al. 2013). Traditional product recommendation systems have employed a combination of collaborative filtering and content-based approaches to match consumers with products, under the assumption that product appeal and consumer preference are static but initially unknown (Sarwar et al. 2001). Collaborative filtering methods identify and exploit consumer and product similarities to make predictions about user tastes or preferences. They may be reinforced via content-based approaches which function by comparing consumer preferences to product features, and thereby, provide even more suitable recommendations.

In addition to the traditional challenges confronting all recommender systems (e.g., cold starts and data sparsity) (Sarwar et al. 2001), user preferences/needs within online smoking cessation communities are dynamic, i.e., continually evolving, as users progress through health state changes (e.g., quitting stages). The process of changing smoking behavior has been subdivided into five distinct stages by smoking cessation researchers, including: pre-contemplation, contemplation, preparation, action, and maintenance (Prochaska and DiClemente 1984). Additionally, users may relapse, i.e., return to an earlier stage. An effective health forum thread recommendation system should therefore tailor its recommendations to reflect users' states of behavioral change (in this case, smoking cessation) in order to provide them with an appropriate level of support.

Forum threads are typically short-lived and quickly changing in their content/narrative, in contrast to the long lives and static characteristics of products being recommended in conventional settings. The content and narrative of a thread may evolve as contributions are made to it by users introducing uncertainty into the very defining characteristic of a thread as a product. An effective thread recommendation system should capture and account for such uncertainty. Note that thread evolution affects not only its future contributors: the benefit/utility that a user derived from participating in a thread is not immediately realized upon their initial posting, resulting instead from the responses made by the future contributors. Due to these dynamics, the benefit/utility of thread participation is unpredictable, being a function of time and depending on future, as yet unrealized, events.

3.2 Network structure considerations and complex behaviors

Threads in online smoking cessation forums facilitate user engagement, providing a platform for interactions between users and the provision of social support. A thread's value lies not solely in its narrative, but in the opportunity that it provides users to directly communicate with one another. A user's acceptance of a thread recommendation can be thought of as signifying that engagement is taking place. To reflect the importance of communication between users, and the role that threads serve to enable it, an effective thread recommendation system must consider both the social network structure of the overall community and within the thread itself.

Social network analysis takes an expanded view of a social environment, allowing for inferences about how network structure both enables and drives behavior change (Cobb et al. 2011). Smoking cessation is an example of a complex adoptable behavior, which is differentiated from simpler behaviors in social network literature (Centola and Macy 2007). The distinction between the simple and complex behaviors is an essential consideration for an effective thread recommender system due to fundamental differences in how behaviors are diffused through a network. Simple behaviors, such as the adoption of a new technology or product, are spread farther and more quickly by networks having many long-ties. Conversely, complex behaviors, such as smoking cessation, typically require a user to be in contact with multiple individuals capable of supporting them in their behavior change before it is adopted. Once adoption of a complex behavior has been realized, continued reinforcement is crucial to ensure that the newly adopted behavior persists and the user does not relapse back to their prior state. Research has shown that highly clustered networks are most effective in facilitating

the adoption of complex behaviors within a community (Centola 2010).

The consideration of network structure in thread recommendation tasks alters the manner in which a thread's value is determined and the purpose which it ultimately serves. The relationship between a thread's network structure and that of a user targeted by the recommender system, directly influences the ability of the thread to provide the user with social support. A thread containing contributions from friends (recognizable peers) may be assumed to provide a greater level of social support to an individual. However, this is not to say that threads containing relatively few of a user's friends are of no value to that user. Rather than providing high levels of immediate social support, such threads provide a user with the opportunity to change their local network structure through the introduction of new ties and/or the strengthening of existing ones. In summary, threads possess the capacity to provide both immediate and future benefits to users. In order to reflect a thread's capacity to influence a user's local network structure, a prescriptive modeling framework capable of capturing the influence of outside interventions (in the form of thread recommendations) on network structure is required.

Existing stochastic actor-based models lack the means to analyze and quantify influence imposed on social networks from the outside. Stochastic actor-based models are a popular methodology for modeling network evolution and predicting ties between actors. Within such networks, nodes represent social actors, e.g., forum users, while edges (ties) represent social relations between them such as friendship, trust, or cooperation. Ties between pairs of actors may be established, or existing ties dissolved, influenced by factors such as the actors' structural positions within the network, actor characteristics (actor covariates), and their relationships with other nearby actors (dyadic covariates). However, in stochastic actor-based models, network ties are actor-initiated, i.e., they can only be changed myopically by the actors themselves.

Formulation of an exogenous intervention strategy requires one to choose an aggregate, actor-based objective function, and decision variables to optimize this function. The modeling challenge lies in the identification and application of external interventions (in the form of thread recommendations) that serve to modify a user's local network in such a way to benefit that user and/or those around them. When recommending a thread to a user for the purpose of altering their local network structure, the likelihood of such changes is an essential consideration. The concept of link-prediction may be applied towards this task (Liben-Nowell and Kleinberg 2007). The link-prediction problem for a social network involves the identification of new links that are likely to appear in the future,

complementing the network's current structure and the characteristics of pairs of users (dyadic covariates).

In order to modify the existing actor-oriented modeling paradigm to accommodate exogenous interventions, the manner in which actor ties are modeled must be revisited. Ties in traditional stochastic actor-based models are assumed to be binary, with relationships between actors either existing or not. To capture the dynamics of user interactions within a smoking cessation community, actor ties should instead be weighted, reflecting varying levels of friendships between actors and the build-up from weak ties to strong ones.

3.3 Weak and strong tie dynamics

While the first co-posting in the same thread by two users may only constitute a weak tie, potential repeated communication between them can lead to tie strengthening over time, eventually resulting in the establishment of a strong friendship tie. In this way, altruistic behavior of core-users can be employed to "push" the network towards a state characterized by higher levels of user engagement by introducing users to one another by co-referencing threads, thereby facilitating meaningful user interactions.

A major deficiency of existing actor-oriented models lies in their inability to explicitly accommodate weak ties and their dynamics. One of the most significant premises, upon which the actor-oriented models are built, is that tie formation is a Boolean class of variables wherein a tie is either present or absent, and must be observable (Snijders et al. 2010). This paper posits that the problem of modeling exogenous interventions can be approached by considering two processes that together describe the formation of a social network. Process 1 expresses how an actor builds strong ties with other nearby actors, i.e., what drives their decisions about with whom to communicate more/less. However, such decisions are clearly made with respect to the actor's acquaintances, with most other actors treated as strangers. Strangers' attributes are unknown to an actor, and their influence on their decision-making mechanism, captured by Process 1, is minimal. This accentuates the importance of Process 2—building acquaintances, termed weak ties. This definition of a "weak tie" is different than that based on the structural holes theory (Walker et al. 1997).

Therefore, models incorporating varying levels of "affinity" between actors are required, enabling more detailed analyses of transitions between weak and strong ties: these transitions may serve as a key underlying facilitator for the growth of health behavior online social networks.

It is strong ties that people would report in a questionnaire, or that can be observed from time-stamped

interaction records. Meanwhile, weak tie patterns are hidden unless they trivially span a whole (small) network. Weak ties can be traced online in certain situations: they are “follow”-type ties as opposed to “friends”-type ties. Therefore, it is crucial to explore approaches to learning weak tie formation dynamics in large networks simultaneously with strong tie dynamics. This will allow (1) the accurate expression of actor decision-making logic, i.e., estimation of Process 1 parameters by removing the bias of the tie patterns that actors are unaware of, and (2) the quantitative evaluation of social influence effects inside networks as well as effects of interventions imposed from the outside.

While strong tie formation driven by actors themselves cannot be influenced from the outside, weak tie formation can. People cannot be expected to become friends just because a model-based tool says they should. However, they can be introduced to each other, informed of congruent interests, and invited to vote on or contribute to “hot” forum threads, etc. Such actions help build acquaintances, as they unobtrusively increase the probability that people will more quickly expand friendship circles, begin communicating with newly found acquaintances, and eventually build stronger ties. A model incorporating weak ties can quantify weak influence effects, and suggest feasible interventions for actor outcomes.

It is only with time that a network actor (e.g., a smoking cessation forum user) expands their local neighborhood, on which they will make decisions about building long-lasting relationships, getting engaged, or staying inactive and leaving for good. Thus, strong tie formation depends on weak ties. On the other hand, it is through communication with friends (i.e., people already trusted) that an actor will learn about other trustworthy actors, begin to distinguish those actors from strangers and explore communication pathways to them. Thus, weak tie formation is facilitated by strong ties.

A potential pathway to incorporating both strong ties and weak ties into a mathematical model lies in studying the behavior of any actor based on the actor’s local network structure, i.e., the actor’s acquaintances, under the assumption that a part of the network is hidden, which is a critical omission in all the existing actor-oriented models the investigators are aware of. The exploration of a network, i.e., the discovery of its hidden parts that may contain useful information, then becomes an important task for an actor, where they may benefit from “outside” assistance.

4 Concluding remarks

Calls for the design and implementation of prescriptive social network analysis techniques for the growth and

maintenance of online health communities continue to emerge. The National Institutes of Health have called for research addressing “the emergence of collective behaviors that arise from individual elements or parts of a system working together” through an exploration of “complex and dynamic relationships among the parts of a system and between the system and its environment” (Marcus 2013). Recent papers such as “The Spread of Behavior in an Online Social Network” (Centola 2010), have improved our understanding of how network structure influences the diffusion of complex behaviors. The present study contributes to this research direction by paving the way for the prescriptive modeling of behavior dynamics. This section touches upon some additional aspects of prescriptive social network modeling for reach enhancement of online pro-health communities, in particular, the treatment of lurkers and recent trend towards using gamification for therapeutic purposes.

A challenge facing the present and prior analyses of online health communities is that passive users (lurkers) are difficult to account for, although they have been found to make up a significant proportion of users in online health forums (Selby et al. 2010). Other research has indicated that lurkers enjoy many of the same benefits as active posters, with more than half of lurkers reporting that “just reading/browsing is enough” (Preece et al. 2004). User anonymity has been observed to play a significant role in WATIs and other online health communities. Although known contacts are potentially more influential than anonymous ones, typically having more detailed knowledge of a particular user’s needs and emotional state (Newman et al. 2011), many users are disinclined to discuss sensitive issues pertaining to habits and behavior on non-anonymous social networks such as Facebook® (Ploderer et al. 2013; Morris et al. 2010).

While the empirical work presented in this paper relies on the data from an online health forum with limited capabilities beyond posting, the more recent introduction of user-controlled features (profile creation, friendship assignment, thread tracking, etc.), the use of gamification as a modern treatment delivery mechanism (Primack et al. 2012), and mobile-based treatment delivery mechanisms (Whittaker et al. 2008; Stanton et al. 1999; Lawrance 2001), may require further effort for analyzing how modern health portals, such as “medhelp.com”, deliver treatment to their participants.

Acknowledgments This work was supported in part by the Academy of Finland Grant #268078 “Mining social media sites” (MineSocMed) and the National Cancer Institute (R01CA152093-01 to S.M.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Academy of Finland, the National Cancer Institute or the National Institutes of Health.

References

- An L, Schillo BA, Saul JE, Wendling AH, Klatt CM, Berg CJ, Ahulwalia JS, Kavanaugh AM, Christenson M, Luxenberg MG (2008) Utilization of smoking cessation informational, interactive, and online community resources as predictors of abstinence: cohort study. *J Med Internet Res* 10(5):e55
- Bishop J (2007) Increasing participating in online communities: a framework for human-computer interaction. *Comput Hum Behav* 23(4):1881–1893
- Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
- Centola D, Macy M (2007) Complex contagions and the weakness of long ties. *Am J Sociol* 113(3):702–734
- Cobb NK, Graham AL, Byron Abrams DB (2011) Online social networks and smoking cessation: a scientific research agenda. *J Med Internet Res* 13:e119
- Crutzen R, De Nooijer J, Candel MJ, de Vries NK (2008) Adolescents who intend to change multiple health behaviours choose greater exposure to an internet-delivered intervention. *J Health Psychol* 13:906–911
- Glasgow RE, Klesges LM, Dziewaltowski DA, Estabrooks PA, Vogt TM (2006) Evaluating the impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues. *Health Educ Res* 21(5):688–694
- Iriberry A, Leroy G (2009) A life-cycle perspective on online community success. *ACM Comput Surv* 41(2):1–29
- Lawrance KG (2001) Adolescent smokers' preferred smoking cessation methods. *Can J Public Health* 92(6):423–426
- Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. *J Am Soc Inform Sci Technol* 58(7):1019–1031
- Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani, et al (2012) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010. *Lancet* 380(9859):2224–2260
- Marcus S (2013, Nov 4) Modeling social behavior funding opportunity [Web log comment]. Retrieved from <https://loop.nigms.nih.gov/2013/11/modeling-social-behavior-funding-opportunity>
- Morris MR, Teevan J, Panovich K (2010) What do people ask their social networks, and why? A survey study of status messages Q&A behavior. In: *Proceedings CHI* pp 1739–1748
- Myung SK, McDonnell DD, Kazinets G, Seo HG, Moskowitz JM (2009) Effects of web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. *Arch Intern Med* 109:929–937
- Narayan KM, Ali MK, Koplan JP (2010) Global noncommunicable diseases—where worlds meet. *N Engl J Med* 363(13):1196–1198
- Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME (2011) It's not that I don't have problems, I'm just not putting them on Facebook: challenges and opportunities in using online social networks for health. In: *Proceedings CSCW*, pp 341–350
- Norman CD, McIntosh S, Selby P, Eysenbach G (2008) Web-assisted tobacco interventions: empowering change in the global fight for the public's (e)health. *J Med Internet Res* 10:e28
- O'Neill B, Ziebland S, Valderas J, Lupiáñez-Villanueva F (2014) User-generated online health content: a survey of internet users in the United Kingdom. *J Med Internet Res* 16(4)
- Ploderer B, Smith W, Howard S, Pearce J, Borland R (2013) Patterns of support in an online community for smoking cessation. In: *Proceedings of the 6th international conference on communities and technologies*, pp 26–35
- Preece J (2001) Sociability and usability in online communities: determining and measuring success. *Behav Inform Technol* 20(5):347–356
- Preece J, Nonnecke B, Andrews D (2004) The top 5 reasons for lurking: improving community experiences for everyone. *Comput Hum Behav* 20(2):201–223
- Primack BA, Carroll MV, McNamara M et al (2012) Role of video games in improving health-related outcomes: a systematic review. *Am J Preventative Med* 42:630–638
- Prochaska JO, DiClemente CC (1984) *The transtheoretical approach: towards a systematic eclectic framework*. Dow Jones Irwin, Homewood
- Rheingold H (2000) *The virtual community: homesteading on the electronic frontier*. MIT Press, MA
- Sarwar B, Karypis G, Konstan J, Riedl J (2001) Item-based collaborative filtering recommendation algorithms. In: *Proceedings of the 10th international conference on World Wide Web*
- Scarborough P, Bhatnagar P, Wickramasignhe KK, Allender S, Foster C, Rayner M (2011) The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006–07 NHS costs. *J Public Health* 33(4):527–535
- Selby P, van Mierlo T, Cunningham J (2010) Online social and professional support for smokers trying to quit: an exploration of first time posts from 2,562 members. *J Med Internet Res* 12(3):e34
- Shahab L, McEwen A (2009) Online support for smoking cessation: a systematic review of the literature. *Addiction* 104(11):1792–1804
- Snijders T, van de Bunt G, Steglich C (2010) Introduction to stochastic actor-based models for network dynamics. *Soc Netw* 32:44–60
- Stanton WR, Lowe JB, Fisher KJ, Gillespie AM, Rose JM (1999) Beliefs about smoking cessation among out-of-school youth. *Drug Alcohol Depend* 54(3):251–258
- Tang X, Zhang M, Yang CC (2013) Leveraging user interest to improve thread recommendation in online forum. In: *2013 international conference on social intelligence and technology*, pp 11–19
- van Mierlo T (2014) The 1% rule in four digital health social networks: an observational study. *J Med Internet Res* 16(2)
- 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 StopSmoking-Center.net. *J Med Internet Res* 14(3):e66
- Walker G, Kogut B, Shan W (1997) Social capital, structural holes and the formation of an industry network. *Organ Sci* 8(2):109–125
- White M, Dorman SM (2001) Receiving social support online: implications for health education. *Health Educ Res* 16(6):693–707
- Whittaker R, Maddison R, Rodgers A (2008) A multimedia mobile phone-based youth smoking cessation intervention: findings from content development and piloting studies. *J Med Internet Res* 10(5):e49
- Young C (2013) Community management that works: how to build and sustain a thriving online health community. *J Med Internet Res* 15(6):e119