Original Research

Semantic networks of interests in online non-suicidal self-injury communities

D Zinoviev[1], D Stefanescu[1], G Fireman[2] and L Swenson[2]

Abstract

People who engage in non-suicidal self-injury (NSSI) often conceal their practices, which limits examination and understanding of their engagement. The goal of this research is to utilize data from public online social networks (namely, LiveJournal, a major blogging social networking site) to observe the NSSI population in a naturally occurring setting. Specifically, the focus of this paper is the interests publicly declared by LiveJournal users. In the course of study, we collected the self-declared interests of 25,000 users who are members of or participate in 139 NSSI-related communities. We constructed a family of semantic networks of interests based on their similarity. The semantic networks are structured and contain several dense clusters–semantic domains–that include NSSI-specific interests (such as *self-injury and razor*), references to music performers (such as *evanescence*), and general daily life and creativity related interests (such as *poetry and friendship*). Assuming users are genuine in their declarations, the clusters reveal distinct patterns of interest and may signal keys to NSSI.

Introduction

Non-suicidal self-injury (NSSI) is the direct, deliberate destruction of one's own body tissue in the absence of suicidal intent.[1] It is practiced primarily by adolescents and young adults[2,3] and is often concealed from others. Common NSSI activities include skin cutting, banging or hitting oneself, and burns.[4]

Recent prevalence estimates suggest that 14% to 21% of adolescents and 17% to 25% of young adults have engaged in NSSI at some point in their lives.[4–6] Furthermore, NSSI is repeatedly found to be associated with significant emotional and behavioral dysfunction, for example, eating disorders and suicide.[7–9] These findings highlight the need to enhance understanding and prevention of NSSI and its social and emotional sequalea.

The goal of this research is to undertake an exploratory study of self-declared interests that could identify NSSI people by automatically analyzing secondary data publicly available from massive online social networks (MOSN), without explicitly interacting with the subjects. Many popular MOSNs (e.g. Facebook and LiveJournal)[10,11] allow users to declare their interests, either explicitly or in the form of "likes." While these interests are often selected randomly and polluted with "status words," we found a very significant correlation between interest lists and membership in NSSI online communities in at least one major MOSN—LiveJournal, a blogging social network.

The association between interest lists and NSSI community membership suggests that "likes" or interest lists may be serving as identity signals "communicating aspects of individuals (e.g. group membership or other preferences) to others in the social world."[12] Such identity signals gain greater meaning (i.e. signal value) as their association with group membership strengthens. Identity signals with greater signal value can influence others, particularly others who aspire for group membership, to adopt behaviors characteristic of the larger group.

[1]Department of Mathematics and Computer Science, Suffolk University, Boston, USA
[2]Department of Psychology, Suffolk University, Boston, USA

Corresponding author:
D Zinoviev, Suffolk University, 73 Tremont St., Boston, MA 02114, USA.
Email: dzinoviev@suffolk.edu

To investigate the value of interest lists generated by members of NSSI online communities in LiveJournal, we use the self-declared interests as graph nodes and similarities between them as graph edges to build a semantic network.[13] We expect to see well-separated interest clusters—dense network communities representing semantic domains, ''area[s] of meaning and the words used to talk about [them].''[14] These domains reflect the mindscape of NSSI people: their conceptual view of the world.

The major advantage of our approach to studying NSSI-related behavior, compared to the traditional face-to-face interviews, is twofold: first, online data are plentiful and easy to obtain using automated tools; second, automated online data collection is non-invasive and unobtrusive, thus allowing us to observe public behavior of NSSI people without introducing unnecessary hesitation and bias and still preserving the subjects' privacy—to the extents defined explicitly by the subjects' online privacy settings.

We hypothesize that there exists a set of self-declared interests which are substantially more frequently used by the NSSI people than by the non-NSSI people (we call them NSSI-related interests). We also hypothesize that the semantic context (the perceived meaning) of certain self-declared interests would be different for the NSSI and non-NSSI people. Our findings support both hypotheses.

Our results have been partially presented at the Workshop on Words and Networks, Evanston, IL.[15] The new contribution of this paper includes:

- a much bigger data set (25% more users and 350% more communities) with a better coverage of the NSSI presence in LiveJournal;
- introduction of reference groups of LiveJournal users—those who are neither formal members nor participants of any known NSSI-related community—and use of these groups for calibrating the results;
- historical data validation through the Internet Wayback Machine;[16]
- study of the interests frequencies.

The paper is organized as follows: the second section describes the data acquisition and processing; the third section is the analysis of the semantic networks and discussion of the results; and the fourth section is conclusion and future directions.

Method

The results presented in this paper are based on data from LiveJournal.[10] The blogging social networking site was started in 1999 by American programmer Brad Fitzpatrick and sold to Russian media company SUP Media in 2007. At the time of data acquisition, LiveJournal hosted approximately 40 million individual blogging accounts and communities.[11,17]

A blogging social network allows individual bloggers to form contact lists, subscribe to their friends' blogs, comment on selected blog posts, declare interests, and participate in collective blogs known as communities. Thus, a blogging network is a bimodal venue where users engage in both publishing and social activities.[18,19]

The data acquisition and processing involved the following stages:

- download of original data (user demographics, interests, relationships, community membership, posts, and comments) from LiveJournal and other websites, including longitudinal data, where available;
- calculation of similarities among users and among interests;
- construction of semantic networks for each group and identification of clusters of interests.

Data acquisition

LiveJournal positions itself as an open blogging platform with a public application programming interface (API). In fact, LiveJournal provides several APIs, including RSS XML and Atom XML for the most recent posts and FOAF XML and plain text interface for contact lists. It is also possible to download profiles in HTML and parse them directly, although the variety of presentation styles makes parsing daunting. The LiveJournal administration encourages retrieval of public data for the purpose of research, provided that researchers follow certain guidelines.[20]

We used a combination of custom-written Python and Unix shell scripts to download and organize the data of interest into a MySQL database. In LiveJournal, a community of users can have its own interests declared by the community moderators, in the same spirit as users declare individual interests. We started our quest by identifying some thematic self-cutting related communities through the built-in LiveJournal search facility by searching for the community interests self-cutting, self-injury, NSSI, self-harm, and cutting. For each community, we downloaded the member list, as well as every publicly available post and comment, the author of each signed (not anonymous) post or comment, and the posting timestamp. Thus, we discovered the initial list of the NSSI-related people. Some of the NSSI-related people are not formal members of any NSSI-related community—but they are on

the list because they published at least one post or comment.

For all NSSI-related people, we downloaded their screen names, dates of birth (where available), titles and subtitles of the personal blogs, the contact lists ("friend-of" and "friended-by"), and the lists of self-declared interests. The contact lists also contain community membership information. They led us to more communities, some of which, after the visual inspection of their descriptions, were also identified as NSSI-related. We added the members of these communities, as well as the non-member authors of the posts and comments published in these communities, to the list of the NSSI-related people. We repeated the aforesaid procedure until no more clearly NSSI-related communities were found, and uncovered the majority of the NSSI-related people (30,562, further referred to as the "[NSSI] core members") and communities (139).

At the time of data collection, only 25,240 of the core members had active LiveJournal accounts. The remaining ca. 5,000 accounts had been permanently closed and purged. The list of the largest 100 NSSI-related communities (by membership size) is shown in Table 1.

The overall number of posts and comments extracted from the NSSI-related communities (both signed and anonymous) is 76,849 and 218,134, respectively. These posts and comments were published between 6 January 2000 and 1 March 2012, and thus cover twelve years of operations. The total size of the posts and comments, including HTML markup and punctuation, but excluding images and multimedia, is 80 and 48 megabytes, respectively.

Reference groups

The goal of the project is to identify the characteristics of the core members' self-declared interests. This implies comparing the interests of the core members to the interests of LiveJournal users who are neither formal members nor de facto participants of any known NSSI-related community.

To do so, we compared the characteristics of NSSI core members with those of their friends or their friends of friends in the hope that direct structural relationships of the core members (that is, their friendships) shall lead us to the non-core users most similar to them and, thus, facilitate a better characterization of NSSI-related people by separating the interests and semantic associations that are NSSI-specific from those that are either general or specific to the age group and subculture.

We downloaded profiles and interests lists of ca. 20,000 randomly selected friends of the core members and another ca. 20,000 randomly selected friend of the

Table 1. Top 100 NSSI-related communities in LiveJournal, reversed sorted by membership size. The names of the communities with any activity in the past two years are shown in bold.

Name	Created	Updated	Size
cuttingimage	11 Mar 2003	3 Apr 2012	5591
cuttersselfharm	20 Oct 2002	19 Mar 2012	3136
_knowyoulive	27 Jan 2004	3 Apr 2012	3094
dont_cut	7 Dec 2001	13 Feb 2012	2196
pro_si	8 Mar 2003	13 Mar 2012	1872
_trigger_happy_	26 Jun 2003	1 Dec 2010	1496
bleed_me_dry	7 Nov 2002	1 Dec 2010	1469
self_mutilation	22 Jun 2001	29 Jan 2012	1317
recoveryourlife	12 Dec 2003	23 Feb 2012	1118
blades_edge	13 Oct 2001	15 Nov 2011	1093
inthecut_	9 Jul 2004	4 Nov 2010	1084
the_cutters	14 Nov 2001	17 Feb 2011	1048
scar_myself	8 Oct 2001	20 Mar 2012	997
asi	25 Dec 2000	14 Jan 2012	980
razor_dreams	12 Apr 2004	2 Nov 2011	974
self_injury	20 Dec 2000	19 Aug 2011	936
bleed_me_skinny	21 Mar 2005	5 Sep 2011	806
csp_support	13 Jun 2002	4 Mar 2012	783
bleeding_beauty	21 Jul 2001	11 Mar 2011	756
cuts_of_pain	7 Feb 2002	26 Nov 2009	669
lovexmyxcrimson	6 Mar 2004	31 Dec 2010	621
blunt_razors	6 Dec 2002	1 Dec 2010	587
_secret_scars_	26 Mar 2005	25 Mar 2011	572
beautifulsi	24 Apr 2002	7 Jul 2012	509
cut_bleed	17 Feb 2005	4 Apr 2013	461
21_gashes	16 Nov 2003	4 Dec 2011	442
icut	13 Oct 2006	29 Mar 2012	375
scarlet_designs	16 Aug 2005	8 Dec 2011	371
razorotik	8 Nov 2001	7 Jul 2010	346

Table 1. Continued.

Name	Created	Updated	Size
__cut	30 Jul 2004	25 Feb 2008	345
siawareness	21 Feb 2004	26 Dec 2011	343
bloody_wrist	19 Dec 2001	2 Feb 2011	326
christiansiers	17 Jun 2001	1 Dec 2010	312
painfulsecrets	28 Jun 2001	8 Feb 2011	310
razors_portrait	12 May 2004	20 Nov 2010	289
cutting_freely	30 Jul 2004	3 Jan 2012	287
xcut_me_nowx	11 Nov 2003	29 Nov 2009	285
bio_etching	12 Mar 2003	3 Feb 2011	276
lovelyscars	23 Jul 2002	1 Dec 2010	270
cuttingsins	30 Jun 2001	25 Jul 2010	270
german_cutter	16 Sep 2002	16 Sep 2009	263
recoverthepain	27 Apr 2009	28 Jan 2011	262
____pain	23 Nov 2004	11 Oct 2009	262
cutters_unite	4 Jul 2004	25 Jun 2009	257
selfinjuryanon	22 Jul 2006	6 Jul 2010	250
anticutting	26 Feb 2004	2 Nov 2012	223
cuttersrepublic	11 Jul 2003	19 Aug 2011	199
0oabdo0	1 Jul 2004	12 Apr 2009	199
cutters	15 May 2001	9 May 2010	188
original_s_i	15 Apr 2005	11 Sep 2009	177
__selfinflicted	21 Apr 2004	30 Dec 2007	174
endingselfharm	31 Aug 2009	22 Oct 2011	163
_self_afflicted	8 May 2004	16 Nov 2010	154
cut_yourself	10 Jan 2002	1 Dec 2010	148
cantcut	16 Mar 2003	21 Sep 2010	148
selfharmhome	18 May 2004	1 Dec 2010	142
self_injure	24 Feb 2002	8 Jul 2007	133
16_gashes	8 Jun 2005	26 Oct 2009	127
injurious	14 Apr 2004	5 Feb 2011	114
bleed_it_out	12 May 2003	2 Feb 2011	109

(continued)

Table 1. Continued.

Name	Created	Updated	Size
selfinjury_ru	10 Dec 2009	8 Apr 2013	109
_bleedingcolors	28 Jun 2004	21 Apr 2012	99
scars_run_deep	12 Nov 2003	13 Dec 2010	98
bloodbracelet	5 Mar 2004	20 Nov 2010	87
sharpobjectsinc	14 Apr 2005	17 Feb 2011	82
poetic_cuts	15 Feb 2004	21 Nov 2006	75
xrazorxkissesx	28 Jan 2004	8 May 2007	74
cutingawaypain	18 Jul 2004	20 Apr 2006	74
why_cut	10 Jul 2004	4 Jan 2008	71
____return	2005-06-29	2012-08-21	71
bleeding_parts	6 Jan 2003	23 Aug 2009	70
selfmutilation	31 Jul 2001	17 Jan 2011	68
cuttingart_	9 Oct 2004	29 Apr 2007	66
love_is_blood	17 Feb 2004	31 Oct 2006	66
painaddicts	27 Apr 2002	23 Aug 2009	63
cutmyvein	5 Mar 2004	25 Jul 2010	60
get_safe	27 Jan 2003	31 Mar 2008	55
letme_bleed	22 Oct 2004	30 Nov 2008	54
a_crimson_love	31 May 2004	16 Nov 2010	54
sharp_objects	12 May 2002	24 Aug 2012	50
adult_siers	20 Aug 2004	16 Nov 2010	50
plural_selfharm	25 Oct 2010	11 Jul 2012	48
bleeding_fae	23 Sep 2001	2 Feb 2011	45
cuttingmeup	24 Jul 2004	16 Nov 2008	43
its_only_flesh	19 Dec 2005	12 Mar 2008	43
pain_is_pain	27 Dec 2004	21 Aug 2008	42
razorbladecuts	2 Sep 2004	28 Oct 2010	41
cutters_club	22 Jul 2004	23 Jun 2006	36
_bleed_like_me_	30 Apr 2005	1 Mar 2011	34
selfburning	15 Jun 2002	17 Feb 2011	33
love_your_blood	15 Feb 2006	16 Nov 2010	32

(continued)

Table 1. Continued.

Name	Created	Updated	Size
_razor_tears_	9 Apr 2005	8 May 2007	31
x_cut	6 Mar 2005	7 Oct 2005	30
breakingviolets	17 Feb 2009	21 Sep 2009	29
__cut__here_	10 May 2005	29 Nov 2010	29
sicons	24 Sep 2005	13 Jun 2006	28
silent_nectar	14 Jun 2002	10 Jun 2005	26
shades_of_blood	13 Dec 2003	25 Aug 2007	25
bleed_pain_away	8 Apr 2004	14 Jan 2013	22
poetsoftheblood	17 Apr 2008	1 Dec 2010	21

friends who are not direct friends of any known core member. The new groups are comparable in size with the NSSI core. Note that some members in each of the three groups did not declare any interests.

We looked at the age distribution and the mean age of the members of all three groups (core members, their friends, and the friends of friends) who declared their dates of birth. An average core member is (27.2 ± 4.6) years old, an average friend of core member is (29.4 ± 6.0) years old, and an average friend of a friend is (29.4 ± 5.9) years old (this is the age at the time of data acquisition). The average ages are similar, and the shapes of the age distributions in all three groups, denoted respectively by C[ore], F[riends], and B[ase], are similar as well. Additionally, they are similar to and consistent with the age of young adult populations most prone to NSSI.

We assume that the members of the groups F and especially B, who are removed or twice removed from the NSSI-related communities, do not exhibit an intense NSSI behavior and do not declare abundant NSSI-specific interests. We use these groups for "calibrating" our approach.

Demographics and posting behavior

The only demographics readily available to LiveJournal researchers is the self-reported age of some (but not all) LiveJournal users. In our data set, only 20% of users specified what seems to be a valid date of birth, yielding the current age in the range from 12 to 80 years.

Based on the posting behavior, the NSSI core population consists of three factions: one-time (35%) and returning (39%) core members, and non-contributing members (26%). The one-time visitors engage in the core communities for no more than four consecutive days and publish no more than two public posts or comments. The engagement time of the returning core members has a log-normal distribution with the mean of 98 days.

The average ages of the returning core members at the time of their first and last observed NSSI-related contribution, as well as of the one-time visitors, are 19.4, 20.4, and 20.3 years, respectively. The closeness of the mean ages gives us a hope that there is no strong demographic difference between one-time visitors and returning core members.

Several NSSI-related communities on our list (e.g. cuttingimage) are private and not available for automated data collection. We do not have posts, comments, and their dates and authors for these communities, but we know who their members are, and the total numbers of posts (but not of comments). In addition, there are some private posts in generally public communities. Only 65% of all posts in the whole data set are public and attributable. The privacy restrictions prevent us from seeing the actual (rather than visible) posting patterns.

Historical validation

The discrepancy between the current average age of a core member and the age at the time of posting to an NSSI community is eight to nine years. However, the lists of the declared interests that we have in our possession are recent. These lists do not necessarily reflect the interests that seemed important to the core members in the past (8–9 years ago). If we plan to use self-declared interests as indicators of an NSSI identity, we must separate them into "sticky" (not frequently added to the lists or removed from the lists) and "volatile" (added to the lists or removed from the lists frequently and unpredictably). The current analyses focus on those interests determined to be "sticky."

LiveJournal does not provide a facility for obtaining historical data. Fortunately, partial snapshots of some LiveJournal accounts have been preserved by the Wayback Machine—a digital time capsule created by the Internet Archive.[16] The Wayback Machine had collected LiveJournal data in several major waves: in 2007, 2008, 2009, 2011, and 2012. The oldest data date back to Fall 2006. No earlier LiveJournal records are available. There is a major gap between Fall 2009 and Fall 2011 with very few observations.

We used custom Python software to download longitudinal interests data (one or more snapshots per user) for 3174 core users, 3304 friends, and 6609 friends of friends. By combining the historical and current interests lists, we calculated interests add/drop patterns and rates. We defined change rate of an interest X as

Table 2. Interests change rates over the Wayback Machine observation period (mean μ and standard deviation σ) for the samples from each group: C (the NSSI core), F (their friends), and B (the reference group).

Group	Users	μ	σ	Interests per user	Volatile interests
B'⊂B	6,609	−3.8%	6.7%	49	144
F'⊂F	3,301	−5.2%	6.9%	60	173
C'⊂C	3,172	−4.9%	5.6%	71	156
All					393

the number of users who added or dropped X over the observation period, divided by the number of users who had X declared at any time during the observation period. The change rate is positive for the interests that are mainly added, negative for those mainly dropped, and close to zero (±10%) for the sticky interests. Additionally, we require that at least 10 users in each group declare X before considering it volatile.

We discovered that the NSSI-related people have the highest average number of declared interests (71 versus 49 for the reference group). The change rate means and averages do not differ substantially between the groups. The mean change rates in all three groups are slightly negative (between −4% and −5%), suggesting that it is more typical for the LiveJournal users to drop interests than to add interests. The most frequently added interests are dexter, doctor, heart, oscar wilde, science, surf, text, heavy metal, murder, ruin, satan, star trek, weight loss, adventure, bubble bath, christmas, horse, insomnia, kink, laughter, lie, lolita, pool, rave, rock music, sexual, sunshine, and viva la bam. The list of the most frequently dropped interests is long and available electronically from the corresponding author. The summary of our findings is shown in Table 2.

Highly positive change rates of the volatile interests makes it hard to argue if the users who currently declare them, also had them in the past. Conversely, highly negative change rates make it impossible to speculate if the users had certain interests in the past but dropped them later. That is why for the rest of the study we restrict ourselves only to the interests that are sticky in all three groups, that is: declared by at least 10 users and having the change rate within ±10% in each group. This restriction will contract the list of the most frequently declared interests (to be defined in the ''Vocabulary construction'' section) from ∼1000 interests to ∼660. However, we have greater confidence in the results that are derived from the more historically stable data.

Vocabulary construction

Only 46,172 (70%) users in the data set have at least one self-declared interest, which results in 369,582 lexicographically distinct interests. Many of the interests are misspelled (cemetery/cemetary), contain punctuation (usa/u.s.a.) and abbreviations (nyc/new york city), represent cognates (france/french), or are otherwise not semantically distinct enough. We normalized all interests by removing all punctuation and stemming each word, thus yielding ∼300 thousand semantically distinct interests.

Further processing of the complete collection of the interests is computationally infeasible due to time constraints and arithmetic loss of precision. We restricted our study to the interests declared by at least 150 core members each—which happened to be the top ≈1000 most frequently declared interests. The complete list of the most frequently declared interests is available electronically from the corresponding author. From this list we eliminated 393 volatile interests, since their prevalence in LiveJournal could not be confirmed at the time when the NSSI people were most active (8—9 years before the data collection). For the similarity calculation, we select the users who declared at least two of the remaining top 660 interests. There are 40,207 users satisfying the requirement, somewhat equally distributed among the groups (C—43%, F—32%, B—25%).

The constructed list of the 660 sticky interests most frequently declared by the NSSI core members is the vocabulary that we use to establish the semantic mindscape of each of the three user groups: the NSSI core, their friends, and the reference group. The vocabulary has been generated based on the self-declared interests of the core members and imposed onto the members of the other two groups. It is plausible that some of the interests that are important to the non-NSSI-related people, but not to the core members, have been overlooked. Conversely, it is plausible that some interests in the vocabulary are of little or no significance to the other users. Since our goal is to study the core group in the first place, we do not consider this possible mismatch a deficiency.

Similarity calculation and semantic network construction

We used the collected data to build three bipartite networks (C, F, and B). Each network connects the set of users U and the set of interests I. A node $u \in U$ is connected to a node $i \in I$ if the user u declares the interest i. Each network is described by its adjacency matrix M, where $M_{ui} = 1$ if there is a link from the user u to the interest i, and $M_{ui} = 0$, otherwise. Each row of the matrix is a vector of declared interests for a particular user, and each column is a vector of users who declared

a particular interest. For each bipartite network, we calculated two projection networks with the adjacency matrices Φ and Θ. The matrix Θ is the adjacency matrix of a semantic network[21] of interests. The nodes of the network represent interests i and the edges represent the similarities between the interests θ_{ij}. The matrix Φ is the adjacency matrix of the users' similarity network. The nodes of the network represent users u and the edges represent the similarities between the users ϕ_{ij} with respect to the self-declared interests. Both projection networks are signed, weighted, and undirected. Both similarity measures, ϕ and θ, are in the range from -1 for totally structurally dissimilar actors (both users and interests) to $+1$ for structurally identical actors.

The measure of similarity between the actors can be as simple as the Chebyshev distance, Pearson correlation, or cosine distance. These three measures treat all dimensions equally, without emphasizing or de-emphasizing any particular interest or user, and assume that all dimensions are independent. While these assumptions may be valid for a random subset of unrelated LiveJournal users, they do not hold in our case: homophily, as well as membership and participation in the NSSI communities, influence the choice of at least some self-declared interests.[12] This intrinsic similarity between the actors should not be ignored when calculating the distances.

Kovacs[22] proposed generalized similarity measures that take the population structure into account. The measures are defined in a mutually recursive way:

- two interests are similar (with generalized similarity θ_{ij}) if they are declared by similar people;
- two people are similar (with generalized similarity ϕ_{ij}) if they declare similar interests.

Let $M_i = M_{i,\star} - \overline{M_{i,\star}}$ and $M^i = M_{\star,i} - \overline{M_{\star,i}}$ be the ith row or the ith column of the matrix M, respectively, normalized by subtracting the corresponding row or column averages. Then the matrices Θ and Φ can be calculated iteratively by starting with two appropriately sized identity matrices I

$$\Phi^0, \Theta^0 = I \tag{1}$$

$$\theta_{ij}^{k+1} = \frac{M_i \Phi^k M_j^{\mathrm{T}}}{\sqrt{\left(M_i \Phi^k M_i^{\mathrm{T}}\right)\left(M_j \Phi^k M_j^{\mathrm{T}}\right)}} \tag{2}$$

$$\phi_{ij}^{k+1} = \frac{M_i^{\mathrm{T}} \Theta^k M_j}{\sqrt{\left(M_i^{\mathrm{T}} \Theta^k M_i\right)\left(M_j^{\mathrm{T}} \Theta^k M_j\right)}} \tag{3}$$

For each consecutive iteration, the previously computed matrices Θ and Φ serve as metric tensors. After a number of iterations, the algorithm approaches the "true" $\Phi \approx \Phi^\infty$ and $\Theta \approx \Theta^\infty$. Note that the first iteration of the algorithm (equation (2)) precisely calculates the Pearson correlation between the actors.

We calculated the matrices Θ and Φ for each of the three user groups on a 64-bit AMD desktop computer with 12GB of RAM, using our custom, highly optimized, sparse array-based generalized similarity calculator written in C++. Only the matrix Θ is important for the rest of the paper.

The matrix Θ is a dense symmetric signed square matrix with few or no true zero terms. The distribution of similarity measures in the matrix is close to uniform. The similarities in the matrix are sustained by the whole body of interests used for its construction and are robust against random variations of individually declared interests.

To explore the structure of the semantic network of interests, we extracted some of the strongest generalized similarities from Θ and created another adjacency matrix, Ψ

$$\Psi_{ij} = \begin{cases} \Theta_{ij} & \text{if } \Theta_{ij} \geq 0.9 \\ 0 & \text{else} \end{cases} \tag{4}$$

The choice of the threshold value of 0.9 is a compromise between the computational complexity of a clustering algorithm and ease of visually observing the hierarchy (both calling for a higher threshold), on the one hand, and connectedness of the network and the completeness of its giant component (both calling for a lower threshold), on the other hand. To study the effect of the threshold, we ran a small size experiment with the threshold of 0 (which eliminated the negative ties and kept all the positive ties) and obtained comparable results, although in a much longer time.

We applied the simple sparsification procedure (equation 4) to each of the three semantic networks: Θ_C, Θ_F, and Θ_B, thus yielding three new matrices Ψ_C, Ψ_F, and Ψ_B. These new matrices are still square and symmetric, but sparse (their densities are $6...9\%$). Each matrix induces a weighted, unsigned, undirected semantic network with weights for interests in the range $[0.9...1]$.

Analysis and discussion

In this section, we analyze the structure of the individual semantic networks of the self-declared interests in the NSSI-related blogging communities and the relationship between the semantic networks.

Table 3. Semantic networks of interests: basic descriptive statistics of C (the NSSI core), F (their friends), and B (the reference group).

Group	Users	Nodes	Edges	Density	Modularity	Clusters
B	10,333	645	16,715	8.0%	0.65	7
F	12,749	649	13,099	6.2%	0.67	8
C	17,125	656	11,009	5.1%	0.69	8

Semantic network community structure

All three groups of users—the NSSI groups members and non-member contributors C, their friends F, and the reference group B—share the same ~660 most frequently declared interests. However, the semantic similarity θ_{ij} between any two interests i and j is not the same for the members of different groups. For example, the core group members perceive the words razor and candle as similar because both can be used to cause self-harm ($\Theta_{ij}^C = 0.50$), but the reference group members perceive the same two words as quite dissimilar ($\Theta_{ij}^B = -0.19$). Consequently, the semantic networks for the three groups have very similar or identical node sets, but different edges and, therefore, different community structure.

Strong perceived similarity between the top interests comes in the variety of forms, such as:

- different graphical representations of the same word (♡–hearts);
- abbreviations and cognates that were overlooked or intentionally ignored at the preprocessing stage (anorexia–ana, nine inch nails–nin, japan–japanese);
- specializations and generalizations (painting–art);
- other associations (soccer–basketball, kissing–hugging, writing–books, etc).

This variety of associations leads us to expect that the networks have a clear community structure: most interests belonging to the same node cluster are connected in many ways to other interests in the same cluster, thus producing a dense web of significant ties. On the contrary, two interests pertaining to very different cognitive, aesthetic or emotional aspects of life, consistently fall into different clusters together with their network neighbors. Thus, a cluster is essentially a network representation of a semantic domain.

We extracted the clusters using the modularity-based louvain algorithm[23] built into the Gephi network visualization package.[24] Each of the three semantic networks has six to seven easily distinguishable and interconnected clusters of comparable size. Note that the density of connections of a more NSSI-specific network is slightly lower and its modularity is slightly higher than of a less NSSI-specific network (Table 3). In other words, the most NSSI-specific network is the most structured one: fewer connections allow clearer separation of the clusters, the clusters have sharper boundaries, and most interests are unambiguously anchored in their respective clusters.

Twenty most frequently used interests from each cluster were posted to Amazon Mechanical Turk[25] and presented to 25–50 randomly selected workers, each. The workers were asked to describe the words with one word or a two-word or three-word phrase. The most frequent responses were used to produce a consistent cluster name. Based on the analysis of the top twenty interests in each cluster, we assigned each cluster to one of the following types (Table 4):

- "Music" (MUS), found in all three groups;
- "Activities, entertainment" (ACT), found in all three groups;
- "Hobbies, life" (LIFE), found in all three groups;
- "Hobbies, creativity" (CRE), found in all three groups;
- "Entertainment, goth, horror" (ENT), found in all three groups;
- "Sex, dark" (SEX), found in all three groups.

The F and B semantic networks also included a "Games, fantasy" (GAM) cluster. In addition, the C and F semantic networks included a "Health, body, disorders" (DIS) cluster that was not present in the B network. Finally, the C semantic network included a "Rock, pop" (ROCK) cluster that was not found in the F or B semantic networks.

Figure 1 shows the semantic metanetwork—the structure of each of the networks B, F, and C, and the connections between them. Each network occupies a horizontal layer. A node is a semantic domain; the node size represents the number of interests in the domain (we defer the discussion of node and edge colors until the next section). Two nodes P and Q in one layer, representing two different domains in the same network, are connected if some interests in P are similar to some interests in Q, and the thickness of the connection reflects the number of similarities.

Table 4. Top 20 sticky interests in each semantic domain in each group of users (C—the NSSI core, F—their friends, and B—the reference group) in the order of decreasing frequency.

Cluster	B	F	C
Music (MUS)	beatles, pierce, rock, punk, green day, cure, pink floyd, use, brand new, bright eyes, dashboard confessional, linkin park, weezer, thursday, nin, incubus, queen, emo, killer, yellow card	guitar, beatles, punk, green day, cure, nin, use, bright eyes, brand new, dashboard confessional, tool, pink floyd, linkin park, thursday, emo, incubus, weezer, drum, yellow card, sublime	green day, use, linkin park, emo, dashboard confessional, brand new, bright eyes, incubus, thursday, yellow card, sublime, fall out boy, alkaline trio, metallica, something corporate, weezer, story of a year, sugarcult, finch, atreyu
Rock, pop (ROCK)			punk, eyeliner, nin, rock, beatles, cure, kitty, tool, korn, jack & jill, pink floyd, hellokitty, 80s, garbage, kurt cobain, disturbed, mindless self indulge, door, fiona apple, pixi
Activities, entertainment (ACT)	music, movies, friend, dance, sleep, love, sing, shop, kiss, food, guitar, rain, swim, beach, laugh, star, boy, hug, cheese, family	music, movies, friend, love, sleep, dance, sing, kiss, rain, shop, star, food, beach, swim, laugh, hug, boy, concert, cuddle, drive	music, love, movies, friend, sleep, dance, kiss, sing, star, guitar, hug, shop, concert, boy, swim, cuddle, laugh, picture, beach, food
Hobbies, life (LIFE)	photography, computer, coffee, chocolate, dream, dog, internet, fashion, dvds, act[ing], camp, bike, candle, jazz, piano, romance, comedy, ocean, thunderstorm, road trip	photography, computer, dream, internet, dog, act[ing], candle, life, dvds, romance, piano, thunderstorm, bike, ocean, road trip, comedy, beauty, journal, peace, kitten	photography, draw, computer, paint, chocolate, internet, water, dog, sarcasm, life, fashion, act[ing], thunderstorm, romance, journal, ocean, family, god, happy, cartoon
Health, body, disorders (DIS)		anorexia, diet, thin, perfect, skinny, fast, bulimia, lose weight, pro-ana, starve, thinspire, restrict, calorie, purge	anorexia, diet, bone, bulimia, perfect, thin, exercise, anxiety, bpd, fast, skinny, mental health, lose weight, pro-ana, recovery, starve, therapy, control, ocd, restrict
Hobbies, creativity (CRE)	read, write, art, book, travel, poetry, cat, cook, history, paint, theater, philosophy, literature, psycho log, tea, film, hike, sushi, polite, nature	write, read, art, book, poetry, cat, draw, travel, coffee, paint, cook, chocolate, psycho log, theater, history, philosophy, literature, tea, sarcasm, nature	write, read, art, anime, book, cat, psycho log, harry potter, coffee, travel, philosophy, video game, fantasy, cook, fanfict, theater, nature, manga, literature, tea
Games, fantasy (GAM)	anime, harry potter, draw, fanfict, sci-fi, video game, manga, fantasy, lord of the ring, japan, game, vampire, comic, mythology, roleplay, slash, rpgs, cartoon, firefly, final fantasy	anime, harry potter, sci-fi, video game, fanfict, fantasy, lord of the ring, manga, mythology, japan, game, cartoon, comic, pirat, roleplay, rpgs, angel, dragon, final fantasy, firefly	
Sex, dark (SEX)	sex, beer, fire, vodka, mac, depress, tear, diet, anorexia, control, perfect, addict, sad, anger, thin, pill, fast, self-expression, hurt, lose weight	sex, pierce, fire, eyeliner, alcohol, 80s, dark, hair dye, depress, vodka, marilyn monroe, fishnet, ♪, hate, fuck, scream, handcuff, polkadot, cut, bunny	poetry, tattoo, pierce, cut, sex, blood, rain, dream, depress, death, candle, suicide, fire, self-injury, dark, girl, alcohol, metal, cry, night
Entertainment, goth, horror (ENT)	tattoo, david bowie, metal, depech mode, bondage, blood, bdsm, kevin smith, death, aquateen	tattoo, vampire, goth, david bowie, metal, death, halloween, beer, tori amos, blood,	vampire, goth, bondage, lesbian, halloween, corset, horror movies, edgar allan poe,

Cluster	B	F	C
	hunger force, slayer, lesbian, boot, vegan, fetish, joy division, fuck, chuck palahniuk, queer, fishnet	techno, bondage, corset, horror, invade zim, horror movies, rocky horror picture show, zombi, edgar allan poe, lesbian	horror, invade zim, alice & wonderland, rocky horror picture show, david bowie, bdsm, insane, ghost, crow, boot, industry, fetish
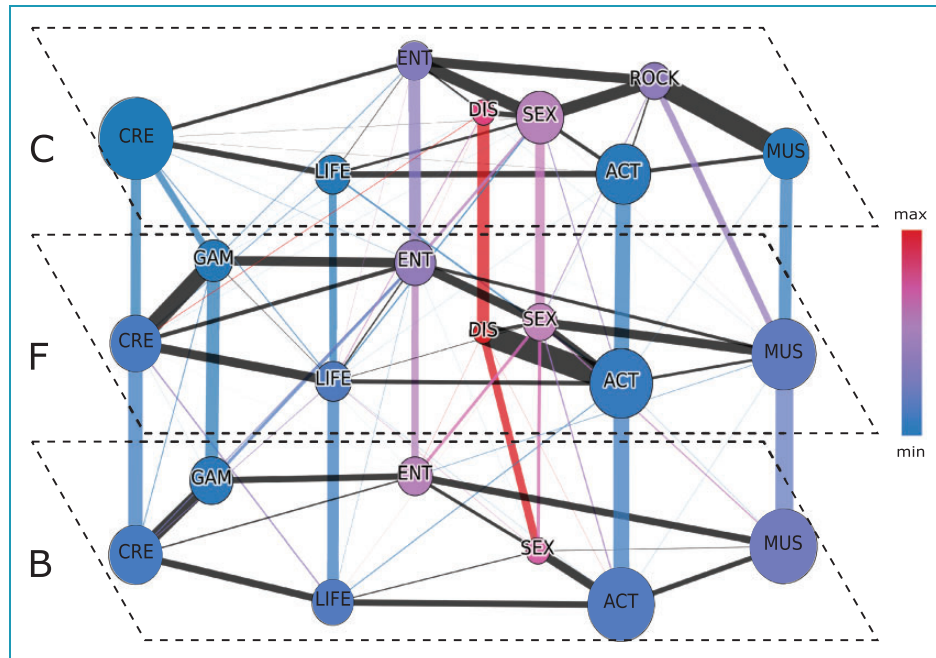
Figure 1. Clusters of interests in the three groups of users: C (the NSSI core), F (their friends), and B (the reference group). Node size corresponds to the number of interests. Line width corresponds to the number of shared interests (vertical connections) and similarities (horizontal connections). Line and node color represents the fraction of the NSSI-related interests in the node or among the shared interests.

Two nodes P and Q in different layers, representing the same domain in two different networks, are connected if some interests in P are also found in Q, and the thickness of the connection reflects the number of shared interests.

The figure portrays the similarity and the difference in perception of interests by the members of the different groups. We would like to focus on four structural aspects.

- For a "base" LiveJournal user, "music" is associated with "entertainment, goth, horror" and "activities." The F group members see another strong connection between "music" and the SEX cluster. The members of the NSSI core go one step further and recognize two types of music: "just music" (still linked to activities) and music dubbed "rock, pop" by the AMT workers, which, in turn, has a strong association

with "sex" and "entertainment, goth, horror." The NSSI-related people are more specific about music.
- "Games and fantasy," an important separate cluster for the reference group and for the groups of friends, is merged into "creativity" by the NSSI-related people. The NSSI-related people are less specific about creativity types.
- The SEX cluster at the reference group level contains some health-related interests (diet, anorexia, thin, fast[ing], lose weight). From the F and C members' perspective, these interests form a separate DIS cluster ("health, body, and disorders"), still strongly connected with the SEX cluster.
- The SEX cluster contains more interests at the friends level than at the reference group level, and even more interests at the core level. The additional interests are drawn from the network neighborhood (mostly from ENT: black rose, boob, burn, cut, die,

fishnet, fuck, handcuff, marijuana, rape, razor blade, self-injury, sharp object—but also from ACT: alcohol, 80s, hair dye—and MUS). These interests are viewed as more SEX-related by the NSSI-related people than by the NSSI-unbiased LiveJournal users. As a side effect, the bond between the SEX and ACT clusters at the core level is much weaker than at the reference group level.

To summarize, the major difference between the NSSI core and the reference group so far is that the NSSI core users have a more structured view on rock and pop music, sex, dark entertainment, and health-related interests—at the expense of a less structured perception of general creativity and gaming activities.

NSSI-related interests

Another observable and potentially important aspect of a self-declared interest (in addition to the interest's position in a semantic network) is its frequency of use by a specific cohort (B, F, or C). We expect that certain interests are declared more often or less often by NSSI-related people and, therefore, can be considered as more or less NSSI-related.

The use frequency of the same interest changes consistently between the three groups. In other words, if the NSSI core members declare X more frequently that their friends, then the reference group members

in general declare it even less frequently (Figure 2). The correlation between $\log(\text{freq}_C/\text{freq}_F)$ and $\log(\text{freq}_C/\text{freq}_B)$ is 0.93. The frequency log ratios are distributed almost normally along the first principal component dimension. The higher log ratios correspond to the interests that are more commonly declared by the NSSI core members than by the reference group. We focus on the interests whose log ratios are larger than $\mu + \sigma$, where μ and σ are the mean and the standard deviation of the log ratio along the first principal component, respectively.

There are 75 NSSI-related interests in the selection (in the order of decreasing frequency ratio $\text{freq}_C/\text{freq}_F$, which is shown in the parentheses):

self-harm (381), *si* (*self-injury*; 293), *self-injury* (124), *razor* (66.5), *bulimia* (51.5), *cut* (50.5), *bpd* (48.5), *razor blade* (39), *purge* (36), *starve* (32), *suicide* (28.5), *bipolar disorder* (26), *abuse* (25), *ocd* (23.5), *self-destruct* (23), *burn* (21.5), *mental health* (19.5), *anorexia* (19.5), *bipolar* (18.5), *hurt* (18), *pro-ana* (*pro-anorexia*; 17.5), *pill* (17), *recovery* (17), *thin* (16), *sad* (15), *depress* (14.5), *anxiety* (14), *needle* (13.5), *restrict* (13.5), *thin-spire* (13.5), *calorie* (13.5), *fast* (13), *perfect* (12.5), *scarification* (12.5), *anger* (12), *rape* (12), *skinny* (11.5), *lose weight* (11.5), *sharp object* (11), *addict* (10), *scarling* (10), *self-expression* (8.5), *alone* (8.5), *blood* (8), *tear* (8), *jack & jill* (7.5), *cry* (7.5), *black rose* (7.5), *blade* (7), *death* (7), *diet* (7), *therapy* (7), *hate* (6.5), *die* (6.5),
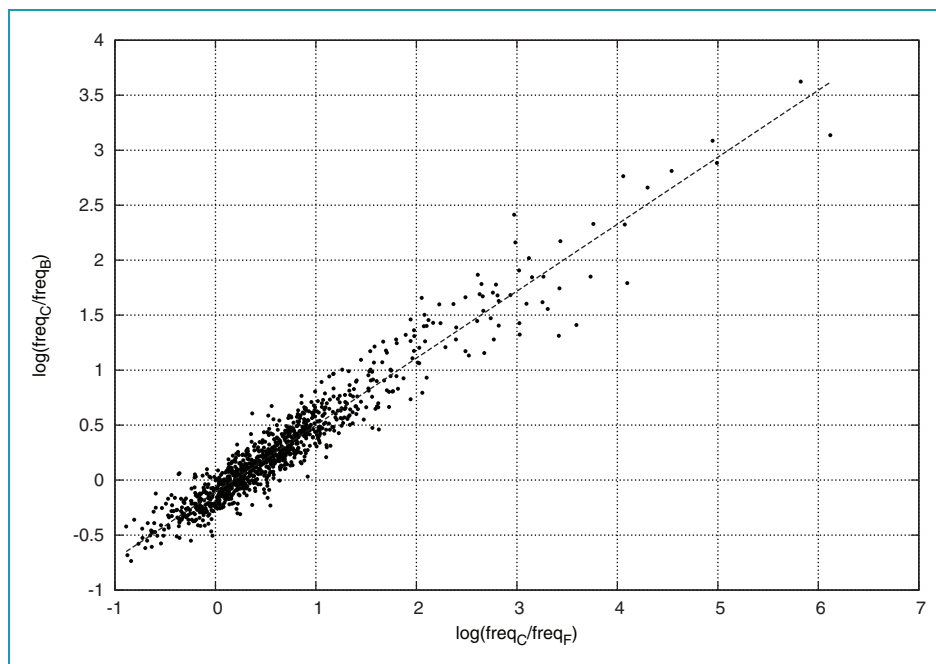
Figure 2. Ratios of interest use frequencies: the NSSI core to the friends vs the NSSI core to the reference group. The dashed line shows a linear regression.

lighter (6), *torture* (6), *black eyeliner* (5.5), *moshpit* (5.5), *kurt cobain* (5.5), *otep* (5.5), *smile empty soul* (5.5), *help* (5.5), *white oleander* (5.5), *murder dolls* (5), *graveyard* (5), *hole* (4.5), *gore* (4.5), *coal chamber* (4.5), *switchblade symphony* (4.5), *scream* (4), *kill* (4), *marijuana* (3.5), *rainbow* (3.5), *type-o negative* (3), and *dyke* (3).

Most of them identify negative emotional, cognitive, and social terms associated with self-harm and mental illness (e.g. eating disorder, bipolar, ocd). To validate this observation, we used color in Figure 1 to visualize the fraction of NSSI-related interests both ''in'' (nodes) and ''between'' (edges) the clusters.

The NSSI-related interests are consistently located in the SEX, ENT, DIS, and ROCK clusters in the networks where these clusters are available. Additionally, the SEX, ENT, and DIS clusters in the friends and especially core level absorb NSSI-related interests from the non-NSSI-related clusters, thus depleting them even further from the NSSI-related content.

Finally, we compared the mean similarities of interests in the NSSI core and in the reference group. The comparison was performed separately for the NSSI-related and all other interests. The mean difference in similarity for the NSSI-related interests is ≈ 0.29 (they are more similar in the NSSI core than in the reference group), but is close to zero for all other interests. The dissimilarity between the two groups of interests is statistically significant ($p < .00001$), which further supports our proposition that the NSSI-related interests indeed form a coherent group—the vocabulary that can be used to track NSSI-related activities.

The role of friends

According to Figure 1, the semantic domains of the F group are somewhat similar both to the semantic domains of the NSSI-related people and to the domains of the reference group. The similarity is seen both in terms of structure of the interests and their frequency of use.

On the one hand, unlike the NSSI-related people, the friends differentiate the GAM and CRE clusters—the gaming-related creativity versus general creativity—but do not emphasize the difference between the general pop music MUS and the niche rock music ROCK. On the other hand, unlike the reference group, the friends discriminate between proper sex-related activities SEX and various health and mental disorders DIS, which makes them closer to the NSSI core.

The friends are clearly a transitional group between the reference group and the NSSI core. One possible explanation for this is that it is composed of two factions: people interested in NSSI but not formally affiliated with any NSSI communities (at least not on LiveJournal) and the reference group members not formally twice removed from the NSSI core. By blending the semantic domains of the two sub-populations, we obtain the semantic domains of the friends. Alternatively, all or the majority of the friends may be indeed interested in NSSI, but not to the same extent as the NSSI core members, which would again result in a blended set of the semantic domains, but for a different reason. The only feasible key to understanding the actual nature of the friends group is to examine their personal blogs in a hope to find traces of NSSI-related vocabulary (say, the NSSI-related interests calculated in the ''NSSI-related interests'' section).

Conclusion and future work

Modern MOSNs combine many features that make them attractive and desirable sources of open, public, and easily retrievable information for research in social sciences. They provide data about individual user accounts (often including basic demographics, such as age, gender, occupation, and location), self-declared users' interests or preferences, brief users' autobiographies or statements, data about relationships between users (one- and two-way friendships, romantic, and family ties), and various epistolary data (wall posts, blog entries, status updates, and comments). This information can be organized to form complex transient social, semantic, geographic, and epistemic networks, and even more complex co-evolving geo-social, socio-semantic, socio-epistemic, etc., overlays.[18,31–34]

Of the plethora of information available from LiveJournal, we used only the structural information (friendship and community membership), self-reported age, and interests lists, to construct a bipartite network of users and self-declared interests. The subjects of our study were identified with or reported interest in NSSI. In related research, individuals with a history of NSSI are found to view themselves as having lower social capital—for example, less attractive, weak social skills.[35] The extent of NSSI-related communities on LiveJournal could evidence the limited opportunities for face-to-face social networking among self-harmers who find themselves excluded from their local communities/local peer networks. In addition, LiveJournal and similar online social networking sites like Safe Haven,[36] purvey excellent grounds for observing thriving NSSI behavior in an undisturbed way.

We extracted and analyzed semantic networks of interests of the members and non-member contributors of 139 NSSI-related LiveJournal communities (the NSSI core), as well as their friends and the friends of friends (the reference group). The networks illustrate how the top ~660 self-declared interests form

clusters—semantic domains—based on their similarity, as perceived by the members of the different groups. The networks have a similar structure: each of them consists of 7–8 major semantic domains. The domains correspond to the deeply emotional sex- and NSSI-related interests (self-injury, razor), creativity (writing, poetry), everyday adolescent and young adult interests (movie, friend), and alternative music (evanescence, nirvana). In other words, the structure of the networks reflects the global middle-class youth culture revolving around leisure activities reflecting adolescent development in MOSNs.[37]

The NSSI-related interests (e.g. self-injury, sharp object) differ from the other self-declared interests in two aspects.

1. They are more frequently declared by the NSSI-related people than by the other users (the quantitative aspect).
2. They are aggregated in the SEX and DIS (eating disorder) semantic domains of the NSSI core network, but more spread across all domains in the other two networks (the structural aspect).

The quantitative aspect implies that the NSSI-related interests are more important to the NSSI-related people than to the other users. The NSSI-related interests, therefore, appear to be valuable identity signals[12] serving as linkages between NSSI group membership and larger youth culture. Future research targeting individuals use of these interests as a means to identify NSSI-oriented social groups would further support this interpretation. The identified NSSI-related interests could also serve as the foundation for online search strategies aiming to identify at-risk persons in need of secondary prevention or intervention efforts (for a related discussion see[38]).

The structural aspect suggests that the perceived interpretation of the NSSI-related interests differs for the NSSI core and the base reference group members with the friends appearing more similar to the NSSI core than to the base reference group. This is illustrated by the semantic cluster DIS being common to both the friends and the NSSI core populations, but not the base reference group, and the similarity of the friends and the NSSI core SEX cluster. This pattern further supports our interpretation of self-declared interests as "identity signals"[12] potentially influencing others (e.g. friends) who aspire for group membership. Greater awareness of NSSI-related identity signals within MOSNs may help to explain how NSSI onset and maintenance are influenced via online social contagion.[26–28]

The amount of data that we have at our disposition or that we deem feasible to obtain, warrants future research in several directions. Our short-term goals include:

- retrieval of comprehensive longitudinal lists of interest and friends from the Internet Wayback Machine[16] and exploration and possibly validation of the contagion hypothesis:[39] do users become LiveJournal friends because they share similar interests or do they share similar interests because they are LiveJournal friends?
- analysis of the corpus of blog posts and comments in the search for influence, persuasion, and pro- and anti-NSSI behaviors;
- exploratory analysis of the "sister" communities dedicated to eating disorders, BPD, OCD, and depression;
- extension of the techniques developed for this project to retrieve and process data from alternative NSSI-related online venues, such as SafeHaven[36] and Tumblr;[40]
- extension of the techniques developed for this project to other domains of dysfunction (e.g. Major Depression) to evaluate the generalizability and possible modifications of the methodology.

NSSI is a dangerous, addictive, and potentially epidemic pathology. We believe that the detailed and in-depth study of the semantic networks of interests of NSSI-related people will substantially inform the prevention and early detection of NSSI.

Declarations

References

1. Nock M and Favazza A. Nonsuicidal self-injury: Definition and classification. In: Nock MK (ed.) *Understanding nonsuicidal self-injury: Origins, assessment, and treatment*. Washington, DC: American Psychological Association, 2009, pp. 9–18.

2. Muehlenkamp JJ. Self-injurious behavior as a separate clinical syndrome. *Am J Orthopsychiatry* 2005; 75(2): 324–333.

3. Jacobson C and Gould M. The epidemiology and phenomenology of non-suicidal self-injurious behavior among adolescents: a critical review of the literature. *Arch Suicide Res* 2007; 11(2): 129–147.

4. Klonsky E and Olino T. Identifying clinically distinct subgroups of self-injurers among young adults: a latent class analysis. *J Consult Clin Psychol* 2008; 76(1): 22–27.

5. Ross S and Heath N. A study of the frequency of self-mutilation in a community sample of adolescents. *J Youth Adolesc* 2002; 31(1): 67–88.

6. Whitlock J and Heath N. Self-injurious behaviors in a college population. *Pediatrics* 2006; 117(6): 1939–1948.

7. Favazza AR, DeRosear L and Conterio K. Self-mutilation and eating disorders. *Suicide Life-Threat* 1989; 19(4): 352–361.

8. Klonsky ED and Muehlenkamp JJ. Self-injury: a research review for the practitioner. *J Clin Psychol* 2007; 63(11): 1045–1056.

9. Nock M. Why do people hurt themselves? New insights into the nature and functions of self-injury. *Curr Dir Psychol Sci* 2009; 18(2): 78–83.

10. LiveJournal.com. http://www.livejournal.com (2012, accessed 2 March 2012).

11. Zakharov P. Structure of LiveJournal social network. In: Kertész J, Bornholdt S and Mantegna RN (eds) *SPIE conference proceedings*. Florence, Italy, 20 May 2007.

12. Berger J. Identity signaling, social influence, and social contagion. In: Prinstein M and Dodge KA (eds) *Peer influence processes among youth*. New York: Guilford Press, 2008, pp. 181–199.

13. Steyvers M and Tenenbaum J. The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cogn Sci* 2005; 29: 41–78.

14. Moe R. What is a semantic domain? http://www.semdom.org/description (accessed 20 August 2015).

15. Zinoviev D, Stefanescu D, Swenson L, et al. Semantic networks of interests in online NSSI communities. In: *Proceedings of the workshop "Words and Networks"* Evanston, IL, 2012.

16. Kahle B. Preserving the Internet. *Sci Am* http://www.scientificamerican.com/article/preserving-the-internet/ (1997).

17. Lafayette L. Social networking and weblog sites for researchers. In: *eResearch Australia, Gold Coast, Australia*, November, 2009.

18. Zinoviev D, Llewelyn S and Walrond-McClean K. Co-evolution of friendship and publishing in online blogging social networks. In: *NetSci-2012*, Evanson, IL, 2012.

19. Merry SK and Simon A. Living and lurking on LiveJournal: the benefits of active and non-active membership. *Aslib Proc* 2012; 64(3): 241–261.

20. LiveJournal Bot Policy. http://www.livejournal.com/bots/ (2012, accessed 6 March 2012).

21. Shapiro S. *Encyclopedia of artificial intelligence*. New York, NY: Wiley.

22. Kovács B. A generalized model of relational similarity. *Soc Networks* 2010; 32(3): 197–211.

23. Bastian M, Heymann S and Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: *Proceedings of the international AAAI conference on weblogs and social media*, San Jose, CA, 2009.

24. Blondel VB, Guillaume JL, Lambiotte R, et al. Fast unfolding of communities in large networks. *J Stat Mech-Theory E* 2008; 2008: P10008.

25. Buhrmester M, Kwang T and Gosling S. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspect Psychol Sci* 2011; 6(1): 3–5.

26. Lewis SP, Heath N, St Denis JM, et al. The scope of non-suicidal self-injury on YouTube. *Pediatrics* 2011; 127(3): 552–557.

27. Jarvi S, Jackson B, Swenson L, et al. The impact of social contagion on non-suicidal self-injury: a review of the literature. *Arch Suicide Res* 2013; 17(1): 1–19.

28. Zinoviev D and Duong V. Toward understanding friendship in online social networks. *Int J Technol Knowl Soc* 2009; 5(2): 1–8.

29. Roth C and Cointet JP. Social and semantic coevolution in knowledge networks. *Soc Networks* 2010; 32: 16–29.

30. Scellato S, Mascolo C, Musolesi M, et al. Distance matters: geo-social metrics for online social networks. In: *3rd Workshop on Online social networks*, Boston, 2010, p.8. Boston, MA: Usenix.

31. Ibarra A. Epistemic networks. New subjects for new forms of (scientific) knowledge production. *Sci Technol Innov St* 2012; 8(1): 61–74.

32. Buffa M, Delaforge N, Erétéo G, et al. ISICIL: semantics and social networks for business intelligence. In: *39th International Conference on Current Trends in Theory and Practice of Computer Science*, Špindlerův Ml^ýn, Czech Republic, 26–31 January 2013, pp.67–85.

33. Claes L, Houben A, Vandereycken W, et al. Brief report: the association between non-suicidal self-injury, self-concept and acquaintance with self-injurious peers in a sample of adolescents. *J Adolesc* 2010; 33(5): 775–778.

34. Safe Haven. Self-Injury.net, http://gabrielle.self-injury.net (2013, accessed 29 July 2013).

35. Larson R, Wilson S, Brown B, et al. Changes in adolescents' interpersonal experiences: Are they being prepared for adult relationships in the twenty-first century? *J Res Adoles* 2002; 12: 31–68.

36. Bragazzi NL. A Google trends-based approach for monitoring NSSI. *Psychol Res Behav Manag* 2015; 2014(7): 1–8. DOI: 10.2147/PRBM.S44084.

37. D'Agostino G, D'Antonio F, De Nicola A, et al. Interests diffusion in social networks. *Physica A* 2015; 436(1): 443–461.

38. Tumblr. Tumblr.com, http://tumblr.com/tagged/cutting (2012, 29 July 2012).