Emotional isolation in BBC Forum

J Sienkiewicz¹ and A Chmiel²
¹ Faculty of Physics, Center of Excellence for Complex Systems Research, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland
² Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, Vienna, A-1090, Austria
E-mail: julas@if.pw.edu.pl, anna.chmiel@meduniwien.ac.at

Abstract. We analyze emotionally annotated massive data from BBC Forum and examine properties of the isolation phenomenon of negative and positive users. Our results show the existence of a percolation threshold dependent on the average emotional value in the network of negatively charged nodes.

1. Introduction
Owing to instant access to tons of data on the online customs of people it is now possible to use physics approaches in order to model human behavior in this media (e.g., [1, 2, 3, 4]). A very special case of this analysis is devoted to emotional component of this activity. Recent research has already revealed certain interesting phenomena, such as collective emotional behavior [5], emotional persistence [6], emotional avalanches [7] or emotional equilibration of discussions [8]. Here we touch the problem of emotional isolation among the users of BBC Forum.

2. Data description
As a source of data we chose the BBC Forum data [5, 9], whose basic properties were gathered in Table 1. The analysis is limited only to seven categories of the forum: UK News, Christian Topics, World News, Ethics and Freethought, Jewish Topics, Eastern Religions and TV and Radio in the period July 2005 - June 2009 (see Table 2). All the downloaded comments were processed using the emotional classifier program that is based on a machine-learning (ML) Language Model [10] approach, which functions in two phases: during the first one, it acquires a set of documents classified by humans for their emotional content (positive, negative or objective) from which it learns the characteristics of each category. During the application phase, the algorithm applies the acquired sentiment classification knowledge to new documents. Posts are annotated with a single value \( e = -1, 0 \) or 1 to express their emotional content (i.e., valence [11]) as negative, neutral or positive, respectively. The accuracy of the classifier checked for 594 humanly annotated comments is 73.73 % for subjectivity detection and 80.92 % for polarity detection (see [7, 5] for details).
**Figure 1.** Networks of connections between negative (upper row) and positive (bottom row) nodes for specific values of the threshold $p$.

**Figure 2.** Size of the giant component $S$ as a function of the threshold parameter $p$ (see Eq. (2) for negative (blue circles) and positive (red triangles) user networks.

**Table 1.** Properties of the BBC dataset: number of comments $C$, number of threads $N$, number of unique users $U$, average and valence $\langle e \rangle$. See also [5, 9]

|   | $C$   | $N$   | $U$     | $\langle e \rangle$ |
|---|-------|-------|---------|---------------------|
|   | 2,474,781 | 97,946 | 18,249  | -0.44               |
3. Results

In order to address the problem of isolation in the network (e.g., [12, 13]) we have used the data to create a network of unique users. It was achieved by setting a link between users \( i \) and \( j \) if one of them has written a comment in response to the other’s comment. We preserved the direction of the relation, thus creating a directed network, where the value of the link is equal to emotional valence of the comment. In the case there were several comments between the same pair of users, the value \( e_{ij} \) of the link was calculated as the arithmetic average of them. Finally, for each node we calculated the mean weighted emotional value on all the in-going links

\[
s_i = \frac{1}{\sum_j w_{ij}} \sum_j w_{ij} e_{ij};
\]

where \( w_{ij} \) is the weight of the link, i.e., the number of of messages sent by \( i \) to \( j \).

In the case of the whole network, the majority of nodes tend to have negative value of \( s_i \) (the average value \( \langle s_i \rangle = -0.42 \)). However, it is of use to examine the nature of connections among negatively and positively charged nodes. To do that we perform percolation analysis [14, 15]: we keep only the nodes that fulfill the following condition

\[
|s_i \pm 1| < p,
\]

Table 2. Properties of the BBC dataset: number of comments \( C \), number of threads \( N \), number of unique users \( U \), average and valence \( \langle e \rangle \). See also [5, 9]

| category               | \( N \)  | \( \langle e \rangle \) | category               | \( N \)  | \( \langle e \rangle \) |
|------------------------|---------|-------------------------|------------------------|---------|-------------------------|
| UK News                | 1,068,803 | -0.37                   | Christian Topics       | 719,617 | -0.53                   |
| World News             | 491,528  | -0.54                   | Ethics and Freethought | 86,579  | -0.47                   |
| Jewish Topics          | 65,367   | -0.42                   | Eastern Religions      | 22,314  | -0.37                   |
| TV and Radio           | 20,573   | -0.50                   |                        |         |                         |

Figure 3. Size of the giant component \( S \) as a function of the node’s emotional value \( e \) for different categories (red points, see Table 2). The blue line is a linear fit \( p_c = A \langle e \rangle + B \) with \( A = 0.78 \pm 0.13 \), \( B = 0.764 \pm 0.059 \) and quality of fit \( R^2 = 0.86 \). The shaded area reflects standard deviation of \( A \).
where the plus sign reflects the procedure for negative nodes and minus sign for the positive ones and \( p \) is the threshold. In effect we obtain a directed weighted network for each value of the threshold \( p \) (see examples in Fig. 1). The obtained plots suggest that (i) there is a specific value of \( p = p_c \) (a percolation threshold) for the network of negative nodes, (ii) positive nodes are isolated in this system. These observations are confirmed in Fig 2, where we present the size of the giant component \( S \) (i.e., the largest connected part of the network) versus the threshold \( p \). In the case of the network of negative nodes, the percolation threshold \( p_c \) is a linear function of the overall emotional value of all post (see Fig. 3).

4. Conclusions
We have shown the existence of a percolation threshold in the emotionally annotated BBC user network. Moreover, our analysis suggest that the positively charged node (users) are isolated in the network.

Acknowledgments
The work was supported from Polish Ministry of Science and Higher Education (grant no 0490/IP3/2011/71). All the data were obtained from EU FP7 CyberEMOTIONS (Collective Emotions in Cyberspace) project (contract 231323).

References
[1] Eckmann J-P, Moses E and Sergi D 2004 Proc. Nat. Acad. Sci. U.S.A. 101 14333.
[2] Kujawski B, Holyst J A and Rodgers G J 2007 Phys. Rev. E 76 036103.
[3] Chmiel A, Kowalska K and Holyst J A 2009 Phys. Rev. E 80 066122.
[4] Radicchi F 2009 Phys. Rev. E 80 026118.
[5] Chmiel A, Sienkiewicz J, Thelwall M, Paltoglou G, Buckley K, Kappas A and Holyst J A 2011 PLoS ONE 6 e22207.
[6] Garas A, Garcia D, Skowron M and Schweitzer F 2012 Sci. Rep. 2 402.
[7] Mitrović M, Paltoglou G and Tadić B 2010 Eur. Phys. J. B 77, 597.
[8] Sienkiewicz J, Skowron M, Paltoglou G and Holyst J A 2013 Adv. Comp. Sys. 16 1350026.
[9] Chmiel A, Sobkowicz P, Sienkiewicz J, Paltoglou G, Buckley K, Thelwall M and Holyst J A 2011 Physica A 390 2936.
[10] Sebastiañi F 2002 ACM Computing Surveys 34.
[11] Feldman L A 1995 Journal of Personality and Social Psychology 69 153.
[12] Sienkiewicz J and Holyst J A 2009 Phys. Rev. E 80 036103.
[13] Sienkiewicz J, Siudem G and Holyst J A 2010 Phys. Rev. E 82 057101.
[14] Stauffer D and Aharony A 1994 Introduction to Percolation Theory, 2nd Ed. (London: Talor & Francis).
[15] Achlioptas D, D’Souza R M and Spencer J 2009 Science 323 1453.