Selected Papers of the 6th Polish Symposium of Physics in Economy and Social Sciences FENS2012, Gdańsk, Poland

Statistical Analysis of Emotions and Opinions at Digg Website

P. Pohorecki[a], J. Sienkiewicz[a,*], M. Mitrović[b], G. Paltoglou[c] and J.A. Hołyst[a]

[a]Faculty of Physics, Center of Excellence for Complex Systems Research, Warsaw University of Technology
Koszykowa 75, PL-00-662 Warsaw, Poland

[b]Department of Theoretical Physics, Jožef Stefan Institute, P.O. Box 3000, 1001 Ljubljana, Slovenia

[c]Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton
Wulfruna Str., WV1 1LY Wolverhampton, United Kingdom

We performed statistical analysis on data from the `Digg.com` website, which enables its users to express their opinion on news stories by taking part in forum-like discussions as well as to directly evaluate previous posts and stories by assigning so called "diggs". Owing to fact that the content of each post has been annotated with its emotional value, apart from the strictly structural properties, the study also includes an analysis of the average emotional response of the comments about the main story. While analysing correlations at the story level, an interesting relationship between the number of diggs and the number of comments that a story received was found. The correlation between the two quantities is high for data where small threads dominate and consistently decreases for longer threads. However, while the correlation of the number of diggs and the average emotional response tends to grow for longer threads, correlations between numbers of comments and the average emotional response are almost zero. We also suggest presence of two different mechanisms governing the evolution of the discussion and, consequently, its length.

DOI: 10.12693/APhysPolA.123.604

PACS: 02.50.−r, 07.05.Hd, 89.20.Hh, 89.70.Cf

1. Introduction

Although the concept of physical modelling of social processes is older than the idea of statistical modelling of physical phenomena (dating back to the times of Laplace, Comte and Stuart Mill) [1], it has not been until recent years that the methods and models of physics became widely (and successfully) employed to the description of social phenomena. Thanks to its very general conceptual framework, the field of statistical physics has proven a tool of exceptional use in this regard [2].

The lively increase observed in the past 15 years in the number of papers and works concerning the field in question was due to many factors. Firstly, with the advances of computational powers and storage capabilities, large digital databases have become available to scientists. Secondly, owing to the rapid development of the Internet, unprecedented social processes appeared and prompted more and more physicists to turn their attention to the rising domain of sociophysics.

Exploring the behaviour of complex systems comprising humans rather than particles, physicists have tackled so far a number of phenomena: from spontaneous formation of a common culture and its dissemination through evolution of opinions and crowd behaviour through isolation phenomena to language dynamics [2–4]. Very recently studying and modelling of collective emotions in Internet communities has also attracted considerable interest.

Collective emotions are relatively straightforward to notice in online communities. Highly emotional discussions are usually connected with very exciting, controversial or tragic events. When many users take part in such highly emotional communication we call this phenomenon collective emotion. The studies of collective emotions in online communities comprises two major areas. First is *sentiment analysis* — computational methods to extract emotional content of a written text and to classify this content according to a set of possible dimensions. Second — building mathematical models of the emergence of collective emotions based on the psychological and sociological body of knowledge on emotion. Computational simulations of the models and the comparison of their results to empirical data verifies the validity of these models.

During the past two years several papers were published presenting studies about the presence of collective emotions on the Internet — both empirical [5–13] and theoretical [14–17] let alone the ones that stress the applicability of the results [18–20]. The objective of this very paper was to analyse available `Digg.com` dataset trying to find regularities and relations that would give insight into the dependences between the emotional content of online discussions and the opinions issued by its users (understood as the number of so-called "diggs"). To be more precise we focus on the following aspects: (1) statistical description of emotionality in `Digg.com` data in division to discussions, users and websites, which is included in Sects. 3 and 4, (2) correlations among number of diggs, number of comments and emotional response, which are discussed in Sect. 5.1, (3) separation of opinions for highly emotional stories, described in Sect. 5.2

*corresponding author; e-mail: `julas@if.pw.edu.pl`

and finally (4) influence of the emotional level on the thread length shown in Sect. 4.

2. Data and methods

Internet discussion participants do not transmit their emotions directly; they communicate via text messages which, depending on their emotional content, may induce certain emotional responses in readers. Information on emotions of individuals can usually be inferred from the physiological signals sent by their body. However, in online communities this is not the case; we are left only with textual statements from discussion participants. The question arises how to infer from these statements information regarding emotions.

Before trying to detect emotions, one needs to decide on how to measure them. A well-established psychological theory of emotion, the circumplex model, is commonly used in sociophysics modelling. It takes into account two dimensions of emotion: *valence* and *arousal*. The former indicates how positive or negative the emotion is, the latter, the level of personal activity caused by that emotion (from lethargic to hyperactive) [6]. Depending on different methods, *sentiment analysis* can usually provide *valence* or a specific combination of the two values [21]. Different procedures of emotion detection are in use. The analysis in [21] and [22] engaged manual (human) annotation, but this method allows for only a limited number of textual statements to be assessed. Much larger amount of data can be processed with the use of automatic annotation developed within the already mentioned field of sentiment analysis.

Application of sentiment analysis (also known as *opinion mining*) to the detection and classification of emotions is a development of the field which initially concentrated on extracting opinions [21]. In the past ten years this area of research has seen a substantial growth [23], gaining a lot of attention from industry and academia alike. This is due to the phenomena of Web 2.0, which led to an unprecedented increase in the amount of online content generated by regular users, rather than website owners or publishers. The information contained in user-generated content (UGC) could be of pivotal importance to firms and institutions. Hence, the first efforts of sentiment analysis focused on analysing multiple movie reviews or comments regarding manufacturers' products with the purpose to determine which features receive most positive and negative feedback. The two fundamental tasks of sentiment analysis are: (i) identifying whether a text is objective or subjective (i.e. contains facts or opinions/emotions) and (ii) determining its subjective polarity (i.e. identifying how positive/negative it is).

In the case of this study we used the following approach: during the training phase, the program is fed with a set of documents classified by humans for emotional content (positive, negative or objective) from which it learns the characteristics of each type. Afterwards, during the second phase, the algorithm applies obtained sentiment classification knowledge to previously unseen documents. We trained a hierarchical language model [24, 7] on the Blogs06 collection [25] and applied the trained model to the Digg data, achieving 70.1% accuracy in distinguishing positive from negative comments and 69.7% accuracy in distinguishing objective from subjective (i.e., positive and negative) ones [8] (the description of the classifier and annotation methods is given in full detail elsewhere [7]). Each post is initially classified as objective or subjective and in the latter case, it is further classified in terms of its polarity, i.e., positive or negative. Each level of classification applies a binary language model [26, 7]. Eventually posts are annotated with a single value $e = -1, 0$ or 1 to indicate their valence (negative, neutral or positive, respectively). The dataset was obtained by a complete crawl of the Digg site [27] for months February, March and April 2009. Data concerning stories submitted during that period were collected. Information on users, diggs and comments relevant to the stories was also gathered. The most fundamental properties of the dataset are shown in Table I.

TABLE I

Fundamental statistics of the dataset.

Stories/posts	Comments	Users	Stories per user	Comments per user
1195808	1646153	484985	2.47	3.39

An example of an online user-generated content rating network, Digg.com relies on users to submit and moderate news stories. Each newly submitted story goes to the *Upcoming* section, which is the place where users browse and vote for (or using website's nomenclature: *digg*) the stories they like most. Once the story fulfils special criteria it gets promoted and moves to the *Popular* section, displayed as the website's *front page*. The exact promotion algorithm is not known to the public (and changes on a regular basis), but the number of votes (diggs) and the rate at which a story receives them are the most important factors [28–34]. The order in which promoted stories are featured on the *front page* is also subject to the algorithm. The most interesting and relevant stories (according to the promotion mechanism) are placed at the top of the page.

Apart from receiving diggs, each story can be commented on. Comments, in turn, can obtain diggs (approvals), but in addition they can also be disapproved of by getting *thumbs down/diggs down* (an action called *buring*). In the context of analysing the emotional content of websites, Digg's voting system seems of pivotal importance. In addition to emotional classification, the dataset allows for analysis of the number of *diggs up/diggs down* which to some extent reflects how interesting and/or emotionally engaging the assessed story/comment was.

A schematic plot of the Digg structure is shown in Fig. 1. Nonetheless, this very study focuses only on dependences between the emotionally classified content and *certain* structural properties or opinions. Throughout

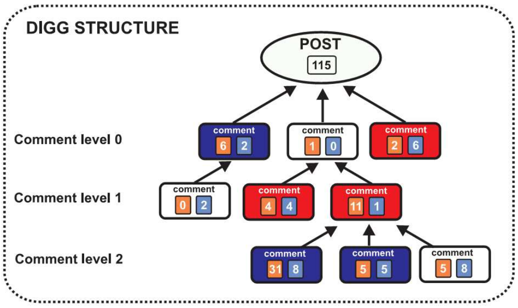

Fig. 1. Digg structure as present in the gathered data. Each story starts with a post (an empty ellipse) that can obtain diggs (a box inside the ellipse). Both the post and the comments can be commented on, however no deeper than to the second level. The comments (rectangular boxes) themselves also obtain diggs, but contrary to the post they can be both positive (diggs up) or negative (diggs down). Each comment is subject to emotional classification, shown with different colours of the comment box.

the paper the term "opinion" will be used with respect to the number (or sort) of diggs issued to a story or to a comment.

3. Structural properties

3.1. *Threads' lifetime*

We define a thread's lifetime as the time that elapsed between the first and the last comment in a thread. Using comments' timestamps (exact date and time) threads' lifetimes were calculated. Their histograms (using different ranges and time scales) are plotted in Fig. 2a. They reveal clear increase of threads counts for lifetimes of 24 h and multiples of this period (inset in Fig. 2a). Also, using different time scale, the character of the graph seems to change around the value of 30 days (Fig. 2b).

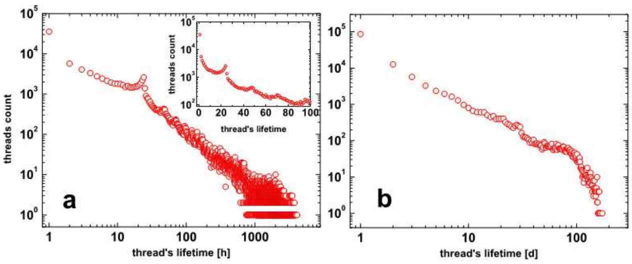

Fig. 2. (a) Log–log plot of the histogram of threads' lifetime given in hours. The inset shows the same data in a semi-logarithmic scale. (b) Log–log plot of the histogram of threads' lifetime given in days. Both hours and days are rounded up to integer values.

Most certainly this behaviour is due (at least in part) to the promotion mechanism of the site and the presentation of top-ranked material. Next to thematic categories, Digg.com allows for viewing the most popular stories within a specific period of time. On the front page an Internet user may choose to browse through most recent stories or top in 24 h, 7 days, 30 days or 365 days.

3.2. *Comment distributions*

A histogram of the number of comments for all data (Fig. 3a) shows two distinct distributions: for lower values a power law can be observed, and then starting around the 20th comment a significantly different distribution takes over. One can hypothesise that the two distributions might be generated by two distinct classes of users (e.g. regular ones and spammers or advertising/marketing professionals). However, this assumption could be tested only by a careful analysis of the actual content of postings.

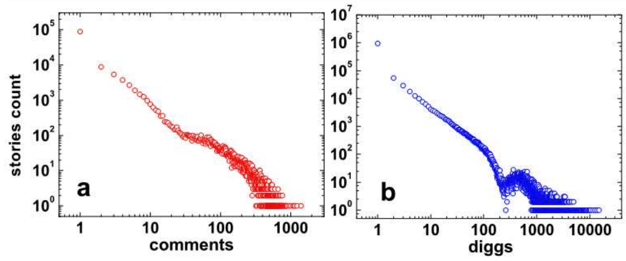

Fig. 3. (a) Log–log histogram of the number of comments for all data. (b) Log–log plot of number of stories that obtained a certain amount of diggs.

3.3. *Diggs distributions*

Similar to comments, diggs histogram — presented in Fig. 3b — displays two distinct distributions. Plots first start with power law and then evolve into Gaussian peak. Power law relation may be explained by a preferential process [35] — a phenomenon quite common in complex systems, responsible for fat-tailed distributions including power-laws [36, 37]. Obtained histograms represent real data, and they are not an outcome of a simulation; however the same mechanism of preferential attachment was at work here. The more diggs the main post (story) obtained in a specific period of time, the better promoted it was through the Digg.com algorithm (front-page placement, higher ranking position, etc.). The unimodal part of the plot has been reported to follow the log-normal distribution [33, 34]. Its origin can be explained either by the law of proportionates affect [33] or a counting process that takes into account a non-constant proportionality factor [34].

4. Average emotional response

Different posted materials trigger different emotional responses. In order to measure overall reaction we introduce a new quantity. We define average emotional response to a main post (story) as a mean value of the emotional content of the comments (all levels) submitted

TABLE II

Number of users engaged in different activities.

Activity	Number of users
gave at least 1 digg	460584
posted at least 1 comment	137183
submitted at least 1 story	218686
posted at least 2 comments	77804
submitted at least 2 stories	108068
submitted at least 1 story, each of the stories had at least 1 comment	44962
submitted at least 1 story, each of the stories had at least 4 comments	7128
submitted at least 1 story, each of the stories had at least 8 comments	3515
submitted at least 1 story, each of the stories had at least 15 comments	2224

to this story

$$\langle e \rangle = \frac{1}{N} \sum_{i=1}^{N} e_i , \qquad (1)$$

where $e_i \in \{-1, 0, 1\}$ is the emotional content of the i-th comment and N is the number of comments (all levels) submitted to a given story.

To determine responses to materials submitted by individual users (or the ones published at individual websites), we group together threads started by the same user (or originating from the same website) and calculate averages of $\langle e \rangle$ in those groups. In the following three subsections, average response to individual posts (that is $\langle e \rangle_{\mathrm{thread}}$ in each thread) as well as to individual users $\langle e \rangle_{\mathrm{user}}$ and websites $\langle e \rangle_{\mathrm{website}}$ will be presented.

4.1. Threads

Analysis concerned only commented stories, i.e. those which initiated a thread. There were 129 998 such main posts (stories). As can be seen from Fig. 4a, due to a large number of very short threads in the whole data there is a prevalence of $\langle e \rangle$ having exact (it employs 1000 bins) value of either -1, 0 or 1. Similarly, smaller peaks correspond to values of $\pm\frac{1}{2}$, $\pm\frac{1}{3}$, $\pm\frac{1}{4}$, $\pm\frac{3}{4}$, etc. which are the only possible results for short (and quite numerous) threads. When we take into account a subset comprising threads of the length of eight or more comments (Fig. 4b), we eliminate peaks for the three values and get a distribution resembling normal distribution shifted slightly to negative values. However performing the Kolmogorov–Smirnov test for normal and log-normal distributions resulted in hypothesis rejection ($\mu = -0.17$, $\sigma = 0.36$, for normal distribution $D_n = 0.068$, for log-normal $D_n = 0.074$ while the critical value equals $\mathrm{CV} = 0.01$ for $\alpha = 0.05$).

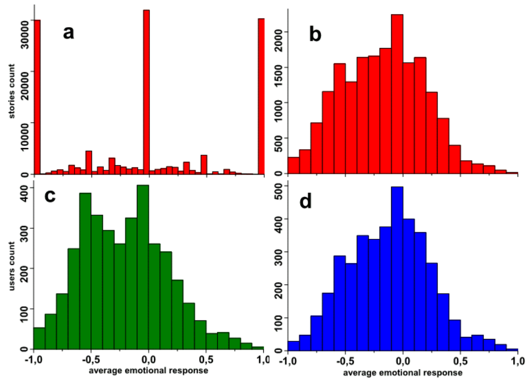

Fig. 4. Histograms of the average emotional response. (a) Average emotional response of threads $\langle e \rangle_{\mathrm{thread}}$, number of bins $N_{\mathrm{bins}} = 50$, all data. (b) Average emotional response of threads $\langle e \rangle_{\mathrm{thread}}$, number of bins $N_{\mathrm{bins}} = 20$, threads with a threshold $N_{\mathrm{comments}} \geq 8$. (c) Average emotional response to material submitted by specific users $\langle e \rangle_{\mathrm{user}}$, number of bins $N_{\mathrm{bins}} = 20$, threads with a threshold $N_{\mathrm{comments}} \geq 8$. (d) Average emotional response to a content from specific websites $\langle e \rangle_{\mathrm{website}}$, number of bins $N_{\mathrm{bins}} = 20$, threads with a threshold $N_{\mathrm{comments}} \geq 8$.

4.2. Users

Collected data enabled tracking of material submitted by specific users and to measure average overall emotional response to their posts. During the crawl there were 484985 registered and active users who either: posted a story, wrote a comment or dug some material. Table II shows the number of users who engaged in different activities.

It can be seen from the table that more users were interested in submitting a story than posting a comment. This is probably due to fact that the primary appeal of Digg.com to the vast majority (so-called "light users") is to share content of interest with others, the desire to exchange comments on the website being of slightly lower importance. We believe that this fact could be a starting point for a further separate analysis.

In order to cut off peaks for -1, 0 and 1, thresholds in the number of comments for a given story were used. The number of stories posted by a user did not seem a good threshold as there were users who posted many stories on which few were commented. For example: user with ID 59919, who posted links to a website on golf in the UK submitted 1402 stories (the greatest number of all users) of which only 5 were commented with the total of 6 comments, i.e. 4 stories received 1 comment and 1 story received 2 comments. He posted no comments of his own.

Figure 4c shows that histograms for users are similar in character: without a threshold in the number of comments there are 3 distinct peaks for -1, 0 and 1. Once the threshold is introduced, we can distinguish two peaks for values near 0 and -0.5.

TABLE III

Top 50 most popular websites (in terms of the number of appearances).

Rank	Website	Counts	Comments	$\langle e \rangle_{\rm website}$
1	www.youtube.com	6808	76433	0.067
2	www.examiner.com	2512	14280	−0.131
3	www.nytimes.com	1465	34719	−0.280
4	www.huffingtonpost.com	1451	35478	−0.338
5	www.news.bbc.co.uk	1222	15092	−0.301
6	www.news.yahoo.com	1066	21225	−0.394
7	www.cnn.com	965	17773	−0.373
8	www.telegraph.co.uk	884	32447	−0.232
9	www.reuters.com	838	12928	−0.342
10	www.rawstory.com	729	14136	−0.563
11	www.washingtonpost.com	688	12435	−0.410
12	www.flickr.com	666	20378	0.091
13	www.foxnews.com	661	6675	−0.451
14	www.arstechnica.com	658	16697	0.064
15	www.squidoo.com	632	825	0.310
16	www.online.wsj.com	558	13298	−0.300
17	www.i.gizmodo.com	549	15481	0.132
18	www.time.com	532	22557	−0.243
19	www.dailymail.co.uk	505	16299	−0.245
20	www.alternet.org	497	6446	−0.389
21	www.msnbc.msn.com	479	16409	−0.239
22	www.guardian.co.uk	475	12341	−0.311
23	www.news.cnet.com	468	13633	0.064
24	www.worldnetdaily.com	459	4214	−0.483
25	www.blog.wired.com	415	11508	0.009
26	www.chicagotribune.com	402	9077	−0.121
27	www.latimes.com	379	8885	−0.313
28	www.collegehumor.com	371	9034	0.138
29	www.opednews.com	348	644	−0.249
30	www.breitbart.com	338	6899	−0.421
31	www.politico.com	325	8571	−0.499
32	www.news.com.au	317	10085	−0.127
33	www.bnp.org.uk	314	1753	−0.556
34	www.hubpages.com	301	439	0.189
35	www.break.com	295	4068	0.026
36	www.firedoglake.com	291	696	−0.012
37	www.npr.org	283	5501	−0.373
38	www.cracked.org	277	17976	0.074
39	www.pcworld.com	273	10124	0.078
40	www.newsciencist.com	267	10473	−0.052
41	www.engadget.com	257	10362	0.186
42	www.usatoday.com	247	4658	−0.262
43	www.dailykos.com	243	4043	−0.384
44	www.slate.com	240	3006	−0.157
45	www.google.com	240	2920	−0.246
46	www.ebaumsworld.com	239	578	−0.003
47	www.salon.com	237	6367	−0.459
48	www.abcnews.go.com	234	6170	−0.257
49	www.blog.propertynice.com	226	248	0.338

4.3. Websites

As in the case of individual users, collected data enabled also the measure of emotional response to content from specific websites (e.g. www.youtube.com, www.nytimes.com, news.bbc.co.uk). The top 50 most popular websites (in terms of the number of times their content appeared at Digg.com) are presented in Table III. The outright winner is video-sharing website www.youtube.com, followed by www.examiner.com (citizen journalism) and a number of professional online newspapers/news outlets (www.nytimes.com, news.bbc.co.uk, www.cnn.com, www.telegraph.co.uk) and news aggregators (www.huffingtonpost.com, news.yahoo.com).

Out of the top 50 most popular websites those with the lowest and highest value of average emotional response were listed in Table IV and Table V, respectively. As can be seen, websites dealing with politics generate most negative emotions (blogs and news websites with political bias, website of the British National Party) while those concentrated on gadgets and technology/giving advice/with humorous content receive the most positive responses. Similar conclusions were reached at in a paper by P. Sobkowicz and A. Sobkowicz [22]. The authors analysed data from the *Politics* section of discussion fora at one of the most popular Internet portals in Poland, www.gazeta.pl. Using human assessment of comments for a sample of discussion threads, they noted that aggressive comments (disagreeing, provocative or invectives) accounted for 75% of communication between *Politics* forum users. This was not the case in the *Sports* or *Science* forum. Moods and opinions of *Sports* section participants turned out to be usually similar while discussions on science, if they happened to be longer, tended to be more factual in character.

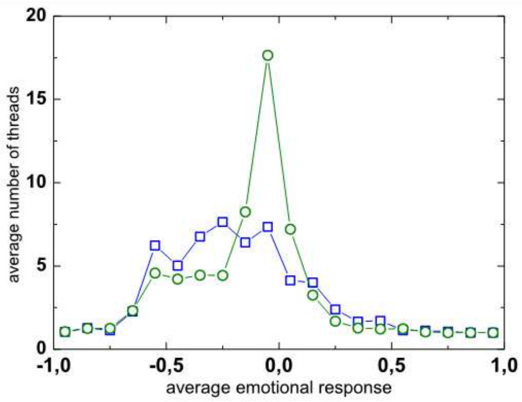

Fig. 5. Average number of discussions (threads) initiated by websites (squares) and users (circles) that have received a specific average emotional response. All threads with a threshold $N_{\rm comments} \geq 8$.

In addition, a graph presenting average number of discussions (threads) initiated by websites grouped according to the value of average emotional response to materials coming from these websites was plotted (squares in Fig. 5). It shows that most successful in terms of the number of appearances and discussions' initiation were websites that received $\langle e \rangle_{\rm website}$ of slightly and mildly negative values. Similarly to distributions for threads and users, $\langle e \rangle_{\rm website}$ histograms without a threshold show three large peaks for values −1, 0 and 1, though their proportional heights differ among the groups of threads, users and websites. After introducing a threshold, a Gaussian-like distribution appears, slightly shifted to the left.

4.4. Background removed

In order to make more visible the differences among distributions for threads, users and websites, a procedure of background removal was carried out. In this context, background means the distribution of threads.

TABLE IV

Top 15 websites (out of top 50 most popular) with the lowest $\langle e \rangle_{\mathrm{website}}$.

Rank	Website	Counts rank	$\langle e \rangle_{\mathrm{website}}$	Description
1	rawstory.com	10	−0.563	liberal news, politics and blogs
2	bnp.org.uk	33	−0.556	British National Party
3	www.politico.com	31	−0.499	American political journalism, conservative and Republican bias
4	www.worldnetdaily.com	24	−0.483	news and associated content, American conservative perspective
5	www.salon.com	47	−0.459	online magazine, focuses on U.S. politics, criticized for its left-leaning content
6	www.foxnews.com	13	−0.451	news channel perceived as promoting conservative political positions
7	www.breitbart.com	30	−0.421	news site, its Blog & “Network” links tend to run to the right within the U.S. political spectrum
8	www.washingtonpost.com	11	−0.41	The Washington Post
9	news.yahoo.com	6	−0.394	news site provided by Yahoo!
10	www.alternet.org	20	−0.389	progressive/liberal activist news service
11	www.dailykos.com	43	−0.384	American political blog, liberal or progressive point of view
12	www.npr.org	37	−0.373	news site of the National Public Radio, a media organization that serves as a national syndicator to most public radio stations in the United States
13	www.cnn.com	7	−0.373	CNN
14	www.reuters.com	9	−0.342	Reuters
15	www.huffingtonpost.com	4	−0.338	liberal/progressive American news website and aggregated blog

TABLE V

Top 15 websites (out of top 50 most popular) with the highest $\langle e \rangle_{\mathrm{website}}$.

Rank	Website	Counts rank	$\langle e \rangle_{\mathrm{website}}$	Description/category
1	blog.propertynice.com	49	0.338	properties
2	www.squidoo.com	15	0.310	publishing platform for posting overview material on the topic of interest, e.g. “50 Things you can Reuse”
3	hubpages.com	34	0.189	publishing tool, all sorts of topics
4	www.engadget.com	41	0.186	gadgets, technology
5	www.collegehumor.com	28	0.138	humour
6	i.gizmodo.com	17	0.132	gadgets, technology
7	www.theonion.com	50	0.118	website of a parody newspaper
8	www.flickr.com	12	0.091	image hosting and video hosting website
9	www.pcworld.com	39	0.078	website of a computer magazine
10	www.cracked.com	38	0.074	American comedy website
11	www.youtube.com	1	0.067	video sharing website
12	arstechnica.com	14	0.064	news and reviews, analysis of technology trends and expert advice
13	news.cnet.com	23	0.064	top technology news headlines
14	www.break.com	35	0.026	humor website, formerly Big-boys.com
15	blog.wired.com	25	0.009	aggregated blogs of Wired.com, an online technology news website

We divide counts from $\langle e\rangle$ histogram for users/website (dark-colored bars in Fig. 6a, b) by corresponding counts from $\langle e\rangle$ histogram for threads (dashed bars). The proportion is plotted for data with (Fig. 7a) and without a threshold in the number of comments (Fig. 7b).

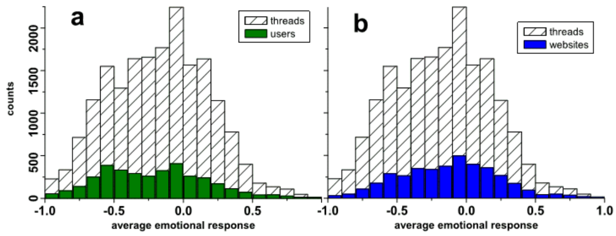

Fig. 6. Comparison of average emotional histograms: (a) threads and users, (b) threads and websites. Number of bins $N_{\rm bins} = 20$, all threads with a threshold $N_{\rm comments} \geq 8$.

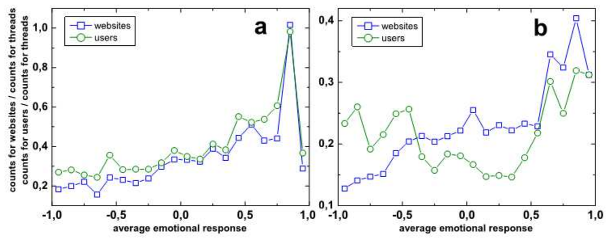

Fig. 7. Proportion of the average emotional response histogram of users to the histogram of threads (circles) and websites to threads (squares). (a) All data, $N_{\rm bins} = 20$. (b) Threads with a threshold $N_{\rm comments} \geq 8$, $N_{\rm bins} = 20$.

For data without any cut-off we get similar plots both for websites and users (squares and circles in Fig. 7, respectively). However, there is not much point in inspecting them in much detail on their own, because as we have already seen, numerous short threads tend to blur the picture. More informative are plots for data with a cut-off. We can see that when a threshold is introduced ($N_{\rm comments} \geq 8$), the character of plot for websites does not change (squares in Fig. 7b), while plot for users transforms into one resembling a U-shape (circles in Fig. 7b).

Based on the figures, a number of hypotheses could be formulated. However, in order to avoid false conclusions, additional graphs have been plotted. They shed more light on the matter under investigation. Namely, Fig. 5 presents average number of discussions (threads) initiated by websites (squares) and users (circles) grouped according to the value of average emotional response to materials coming from these websites/submitted by these users.

Now, let us turn back to the analysis of the results obtained in Fig. 7. Most interesting is the difference in graphs for websites and users, with introduced threshold. It is shown that the group of websites receiving exclusively very positive responses and groups of users receiving exclusively either very positive or very negative responses are relatively more numerous. On the other hand, the group of websites receiving exclusively very negative responses is relatively less numerous.

We could hypothesise that largely negative threads group into a relatively small number of websites which in turn originate many discussions. However, this logic fails when confronted with Fig. 5. Here, for values between -1 and -0.7, the average number of threads is equal or very close to 1, meaning that almost all websites which receive $\langle e\rangle$ of values from this region initiated only one thread. The same is true for the region of positive $\langle e\rangle$ (from 0.5 upwards) and for users (negative and positive regions). Also, in the case of users, those whose submitted material received $\langle e\rangle$ of values between -0.4 and 0.5 are slightly less numerous.

It is worth emphasising that observed relations are not a special case of the threshold used ($N_{\rm comments} \geq 8$). Starting with threads of the length of 5 or more comments, the U-shape for users clearly appears and becomes more and more evident with the increase of the threshold. Plots for websites express the same behaviour irrespective of the threshold used.

5. Correlations

The objective of the following part of the analysis was to determine whether emotional content obtained by computer program (both continuous and binary classifications used) correlates with the number of users' approvals (diggs up), disapprovals (diggs down) or some function thereof. Data was divided into a number of subsets for which correlation coefficients were calculated. In addition, some graphs presenting behaviour for selected ranges were plotted.

5.1. *Correlations of diggs and emotional response*

For main posts (stories), only the number of approvals (diggs up) is registered by `Digg.com` mechanism, hence it was the only metric that could be used for determining correlation with average emotional response $\langle e\rangle_{\rm thread}$.

Correlation coefficient (Pearson's product-moment correlation coefficient) between number of diggs $N_{\rm diggs}$ and $\langle e\rangle_{\rm thread}$ for all commented stories (threads) equals -0.027, implying no correlation on a global level. In order to establish whether or not any correlation can be observed for limited ranges of data, correlation coefficients were calculated for various data subsets obtained by introduction of different thresholds in the number of comments ("high-pass" mechanism, i.e. threads with a specific and higher number of comments were taken into account).

In Fig. 8a obtained coefficients were plotted versus threshold values. For example, threshold point equal to 200 means that the coefficient was calculated for the subset of threads of the length equal to 200 comments and more.

As Fig. 8a reveals, the correlation level between $\langle e\rangle_{\rm thread}$ and $N_{\rm diggs}$ (squares) is positive and, except for

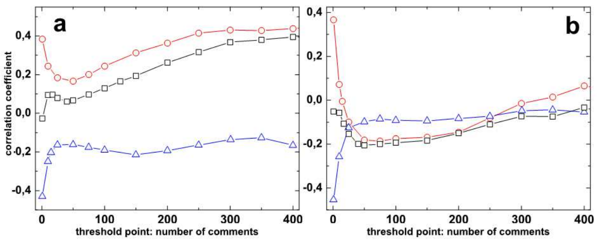

Fig. 8. (a) Correlation coefficients between number of diggs N_{diggs} and the average emotional response $\langle e\rangle_{\text{thread}}$ versus the threshold on the number of comments N_{comments}. Squares — correlation coefficients for all threads, triangles — correlation coefficient for threads with $\langle e\rangle_{\text{thread}} > 0$, circles — correlation coefficient for threads with $\langle e\rangle_{\text{thread}} < 0$. (b) Correlation coefficients between number of comments N_{comments} and the average emotional response $\langle e\rangle_{\text{thread}}$ versus the threshold on the number of comments N_{comments}. Squares — correlation coefficients for all threads, triangles — correlation coefficient for threads with $\langle e\rangle_{\text{thread}} > 0$, circles — correlation coefficient for threads with $\langle e\rangle_{\text{thread}} < 0$.

initial bump, increases with the length of the threads (linearly for a long range of values). This behaviour is accounted for mainly by threads with negative $\langle e\rangle_{\text{thread}}$ (circles in Fig. 8a) owing to their prevailing number. Unintuitive is the fact that correlation for threads with positive $\langle e\rangle_{\text{thread}}$ is negative for all thresholds (triangles in Fig. 8a). Initially it sharply increases and from around the point equal to 25 stays at more less the same level. One would expect that more popular stories (i.e. those with a larger number of diggs) will trigger more positive responses. A contrary behaviour is implied by data — popular stories tend to have lower (though still positive) $\langle e\rangle_{\text{thread}}$. This is most significant in the region where small threads dominate — the most negative value of correlation coefficient can be observed there.

TABLE VI

Correlation coefficients c between average emotional response $\langle e\rangle_{\text{thread}}$ and the number of diggs N_{diggs} versus the number of comments with an imposed threshold $N_{\text{comments}} \leq x$; M stands for number of threads.

x	c	M
100	-0.017	125399
150	-0.019	127038

Similarly, Fig. 8b also implies that it is the stories with the value of $\langle e\rangle_{\text{thread}}$ nearing zero (either positive or negative) that attract larger numbers of diggs. Positive correlation for negative $\langle e\rangle_{\text{thread}}$ means that the higher (closer to zero) the negative values, the more diggs they obtain. In addition, correlation for two "low-pass" thresholds were also calculated and listed in Table VI, proving explicitly that it is the impact of shorter threads that hides in the overall data the correlations for longer threads.

5.2. Correlations of comments and average emotional response

Similar calculations to those for diggs were carried out for correlation between $\langle e\rangle_{\text{thread}}$ and the number of comments N_{comments} (Fig. 8b). Correlation coefficient between the average emotional response and the number of comments is slightly negative with a minimum of 0.2 for the threshold point equal to around 50. The cut through 0 around point 500 should be treated with care as fluctuations are more probable around this region due to lower statistics. The plot implies that longer threads tend to be (slightly) more negative. In order to further investigate the relation between $\langle e\rangle_{\text{thread}}$ and N_{comments}, coefficients for threads with exclusively negative and positive $\langle e\rangle_{\text{thread}}$ were calculated and plotted in Fig. 8b.

Relatively high (in the case of $\langle e\rangle_{\text{thread}} < 0$) and low (in the case of $\langle e\rangle_{\text{thread}} > 0$) initial values stem from the fact that in the region where short (but solely negative/positive) comments dominate their large (in absolute values) $\langle e\rangle_{\text{thread}}$ have a huge impact, resulting in, respectively, very high (the longer the negative thread the less negative it gets) and very low (the longer the positive thread the less positive it gets) values of coefficients. As was the case with diggs, surprisingly the correlation for $\langle e\rangle_{\text{thread}} > 0$ is (slightly) negative for larger values of comment thresholds as well. For $\langle e\rangle_{\text{thread}} < 0$ an expected behaviour for the threshold range between 10 and 300 occurs — the negative value of correlation indicates that (for this range) longer threads indeed tend to be more negative. However, for exclusively long threads this tendency is not the case.

In the case of comments, the data consisted also of the information about the number of disapprovals (diggs down, negative diggs) submitted by users. This fact allows consideration of another quantity — digg difference, defined as

$$\Delta d = d_{\text{up}} - d_{\text{down}}, \tag{2}$$

where d_{up} (d_{down}) is, respectively, the number of diggs up (down) submitted to the comment. The histogram $H(\Delta d)$ is shown in Fig. 9, suggesting a similar law behind the process of issuing both positive and negative diggs. However, as the largest $|\Delta d|$ for the positive branch is about 10 times the value for the negative one, and taking into account that $H(|\Delta d|)$ for $\Delta d < 0$ drops much faster than for $\Delta d > 0$, it seems that the users are much more reluctant to submit negative diggs.

On this level of analysis it is also possible to check the relation between the average emotional value of comments $\langle e_{\text{c}}\rangle$ and the digg difference Δd. All comments that acquired a specific value of digg difference Δd were grouped together and their average emotional value was calculated (see grey points in Fig. 10a). Figure 10a suggests that sentiment of the comments characterised with a negative value of Δd tends to be more negative than in the case of the comments with $\Delta d > 0$. Moreover, the value of $\langle e_{\text{c}}\rangle$ saturates for higher Δd, being close to average emotional value of the whole dataset ($\langle e_{\text{c}}\rangle_{\text{digg}} = -0.16$, marked with a dotted line in Fig. 10a).

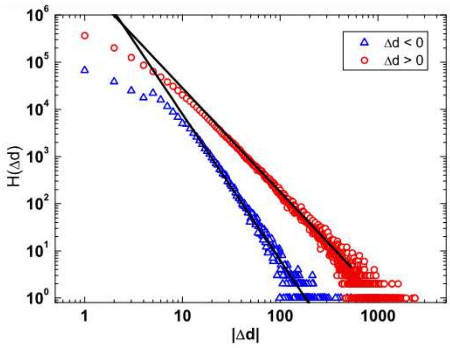

Fig. 9. Log–log histogram of the absolute value of digg difference $H(|\Delta d|)$. Triangles represent the branch $\Delta d < 0$ while circles $\Delta d > 0$. The solid lines represent power-law fitting to the data: in case of positive Δd the line follows $H(|\Delta d|) \sim |\Delta d|^{-2.2}$, as for negative Δd it is $H(|\Delta d|) \sim |\Delta d|^{-3.1}$.

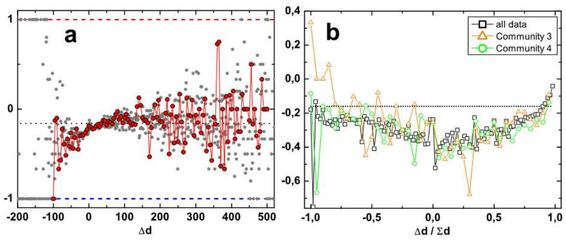

Fig. 10. (a) Average emotional value of comments $\langle e_{\mathrm{c}} \rangle$ and the digg difference Δd. Grey points are real data while circles are obtained using a 5-point binning. Dotted line marks average emotional value of the whole dataset $\langle e \rangle_{\mathrm{digg}} = -0.16$ and solid lines indicate levels $\langle e \rangle = 1$ and $\langle e \rangle = -1$. (b) Average emotional value of comments $\langle e_{\mathrm{c}} \rangle$ versus normalised digg difference $\frac{\Delta d}{\sum d}$ for all data (squares) and selected user communities (circles and triangles). Dotted line marks average emotional value of the whole dataset $\langle e \rangle_{\mathrm{digg}} = -0.16$.

One has to take into account the fact that the values shown in Fig. 10a are not normalised. In order to present a more accurate quantity one should divide Δd by $\sum d = d_{\mathrm{up}} + d_{\mathrm{down}}$, thus obtaining the relative value of digg difference. The results, marked with squares, are plotted in Fig. 10b, demonstrating a clear minimum around $\frac{\Delta d}{\sum d} = 0$ and increasing values for both positive and negative $\frac{\Delta d}{\sum d}$ once again stopping in the vicinity of $\langle e \rangle_{\mathrm{digg}} = -0.16$. It leads to a rather stunning conclusion that no matter if all diggs submitted to a comment are positive or negative, its content, on average, will have a similar emotional value. A possible explanation might be put in the following way: a comment with the emotional content close to average value does not provoke a separation of opinions — it is either commonly liked or disliked. On the other hand, those comments that seem to divide the users into almost equal fractions seem to have very low $\langle e_{\mathrm{c}} \rangle$. Bearing in mind that the above conclusion might be an artefact we checked the relation $\langle e_{\mathrm{c}} \rangle \left(\frac{\Delta d}{\sum d} \right)$ for four different user communities obtained via eigenvalue spectral analysis of the weighted bipartite (i.e., users and comments) network of the most popular comments (for method details see [5, 7, 8]). The results, shown for the two largest communities with number of comments $N_{\mathrm{c}} = 10214$ (circles in 10b) and $N_{\mathrm{c}} = 51166$ (triangles in 10b) confirm the previous observations. Although there are small discrepancies, the tendency stays the same, suggesting that the described behaviour is common regardless of the dataset partition scheme.

5.3. Correlation of comments and diggs

The correlation between the number of diggs N_{diggs} and the number of comments N_{comments} (Fig. 11) shows a very interesting behaviour. The coefficient is very high (more than 0.8) for the data where small threads dominate and consistently decreases for longer threads, though having positive value all the time (slightly over 0.2 at the lowest). The behaviour has a very convincing heuristic explanation. Longer threads are usually developed thanks to multiple comment exchanges between a limited number of users (usually a few, but binary exchanges — just between two people — are also frequent). Those users post additional comments, but do not digg the story again — it is not allowed by the system, even if they had such an (unlikely) wish. Hence the discrepancy in the number of diggs and comments for long discussions. For short and medium-sized threads such a duality does not occur and the numbers are roughly the same.

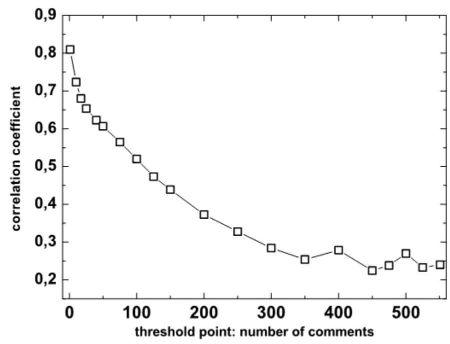

Fig. 11. Correlation coefficients between number of diggs N_{diggs} and the number of comments N_{comments} versus the threshold on the number of comments N_{comments}.

6. Average response to a story

As a development of the analysis of diggs correlations from the previous section, we examine here the dependence of average emotional response to a story $\langle e \rangle_{\mathrm{thread}}$ on the number of diggs N_{diggs} and comments N_{comments} the story receives. The graphs presented in Fig. 12 exhibit a very interesting behaviour. They imply that there is a specific value at which average emotional response assumes a minimum. The point in question is equal to approximately $N_{\mathrm{diggs}} = 50$ in the case of diggs (Fig. 12a) and approximately $N_{\mathrm{comments}} = 20$ in the case of comments (Fig. 12c).

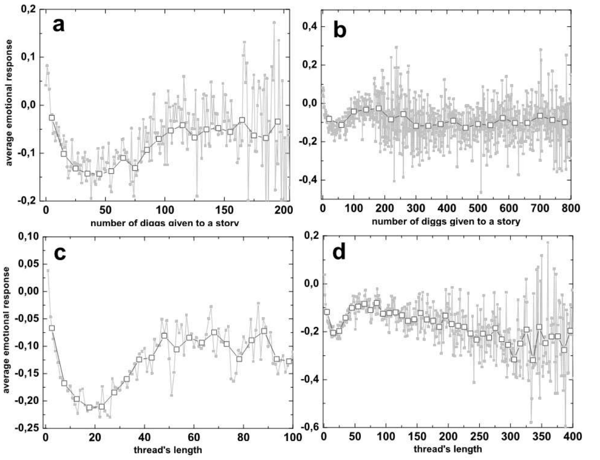

Fig. 12. Average emotional response to a story $\langle e \rangle_{\text{thread}}$ (a)–(b) versus the number of diggs given to a story N_{diggs} (c)–(d) versus the thread's length. Small squares are original data, big squares represent averaging with 5 bins. Plots (a, c) are close-ups of plots (b, d), respectively.

In the further part, the graphs differ: $\langle e \rangle_{\text{thread}}$ (N_{diggs}) fluctuates around a fixed value (Fig. 12b) whereas $\langle e \rangle_{\text{thread}}$ (N_{comments}) visibly decreases toward the end (Fig. 12d). The behaviour for comments is of great importance as it indicates that (beyond certain length) longer threads tend to be more negatively charged. Hence we could assume that it is the negative comments that fuel the communication. Similar conclusions were reached in [22] where it was observed that confrontational or abusive discussions lasted much longer than neutral ones. It is worth mentioning that the overall behaviour of the two graphs comply with the previous results concerning correlation between the number of diggs and the number of comments. However, one should not draw any conclusions about the causality between the activities of reading the comments and digging. As was presented in Fig. 11, for short threads, the correlation is very high and substantially decreases for longer discussions. That is also the case here: the assumption of minimal value for short threads is observed in plots for diggs and comments alike while the behaviour of the graphs towards longer discussions does not match.

In order to explain the phenomena of minimal value seen for shorter threads, average response histograms for the initial groups of threads were plotted in Fig. 13. It can be assumed that the ratio of probability that all (or almost all) comments in a thread are negative to an analogous likelihood for a positive dominance is responsible for the behaviour in question. For very short threads, probabilities for extreme values of $\langle e \rangle_{\text{thread}}$ are roughly the same (cf. equal bars in Fig. 13a); for threads between 10–20 a major dominance for very low negative values is observed culminating for threads of the length of around 20 comments. For still longer discussions the dominance tends to diminish as the probability of obtaining solely negative comments in longer threads is very low. Around

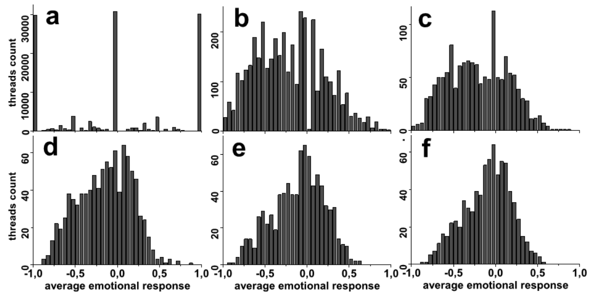

Fig. 13. Average emotional response histograms for threads of specified length N: (a) $N \in [0, 10]$, (b) $N \in [11, 20]$, (c) $N \in [21, 20]$, (d) $N \in [31, 40]$, (e) $N \in [41, 50]$, (f) $N \in [51, 60]$.

the value of 50 comments both probabilities for extreme $\langle e \rangle_{\text{thread}}$ values are again similar (this time both are very close to zero), though the graph as such is slightly shifted to the left.

7. Conclusions

During the period of analysis a number of possible study directions were penetrated. Some provided interesting results. Below there are listed conclusions deemed by the authors as bearing the greatest significance.

It was established that once a certain length of the thread is reached, the regularity that longer threads acquire more negative charge is valid. We can assume that it is the negative emotions that (starting from some point) propel the discussion in longer threads. However, for the thread to develop, for certain lengths it should not be launched with highly negative emotions. If the first comments are largely negative the thread dies quickly. A possible explanation of this mechanism is that when participants give vent to their emotions early on in the thread, they later do not have enough emotional reserves to carry on discussion. Similarly, mildly and highly positive launches tend to lead to shorter discussions. This suggests presence of two different mechanisms governing the evolution of the discussion and, consequently, its length.

With the use of averages it was ascertained that the most negative emotional responses were prompted by websites dealing with politics. On the other hand, those concerned with technology, giving advice or humour generated the most positive reactions. The most successful in terms of the number of appearances and discussions' initiation were websites that received average response of slightly and mildly negative values.

Contrary to expectations, no correlation was found between the number of diggs received by a comment and its emotional charge. This leads to a conclusion that *digging* and *burying* are driven by the interest of the comment, rather than by its emotional content. People vote for comments that are interesting or witty, or that say something familiar and disapprove of those which are boring, irrespective of the emotions they convey.

While analysing correlations on a story level, an interesting behaviour of the relation between the number of diggs and the number of comments received by a story was found. The correlation between the two quantities is high for data in which small threads dominate and consistently decreases for longer threads (though staying positive all the time). This behaviour has a convincing explanation. Namely, longer threads form as a result of exchanges between a limited number of discussion participants. No matter how many comments these users write in a specific thread, they digg the story only once. Hence the longer the thread, the wider the discrepancy between the number of diggs and the number of comments.

Results indicate that threads with a small number of diggs (corresponding to a small number of comments) are relatively more objective. With the increase in the length and the number of diggs the threads' subjectivity increases.

It is worth highlighting that a considerable number of distributions plotted for `Digg.com` followed (at least for some ranges of values) power laws. This included comments and digg distributions as well as users' productivity. Thus it was confirmed also in this work that scale-free relations are often found in data concerning online behaviour.

Acknowledgments

The work was supported by EU FP7 ICT Project Collective Emotions in Cyberspace CYBEREMOTIONS (contract 231323) and Polish Ministry of Science Grant 1029/7.PR UE/2009/7. J.S. and J.A.H. acknowledge the support from the European COST Action MP0801 Physics of Competition and Conflicts and from the Polish Ministry of Science grant No. 578/NCOST/2009/0. J.S. acknowledges support from a special grant of the Dean of the Faculty of Physics, WUT.

References

[1] P. Ball, *Critical Mass: How One Thing Leads to Another*, Farrar, Straus and Giroux, London 2004.

[2] C. Castellano, S. Fortunato, V. Loreto, *Rev. Mod. Phys.* **81**, 591 (2009).

[3] R. Axelrod, *J. Conflict Resolut.* **41**, 203 (1997).

[4] K. Sznajd-Weron, J. Sznajd, *Int. J. Mod. Phys. C* **11**, 1157 (2000).

[5] M. Mitrović, B. Tadić, *Eur. Phys. J. B* **73**, 293 (2010).

[6] F. Schweitzer, D. Garcia, *Eur. Phys. J. B* **77**, 533 (2010).

[7] M. Mitrović, G. Paltoglou, B. Tadić, *Eur. Phys. J. B* **77**, 597 (2010).

[8] M. Mitrović, G. Paltoglou, B. Tadić, *J. Stat. Mech.* **2011**, P02005 (2011).

[9] A. Chmiel, P. Sobkowicz, J. Sienkiewicz, G. Paltoglou, K. Buckley, M. Thelwall, J.A. Hołyst, *Physica A* **390**, 2936 (2011).

[10] A. Chmiel, J. Sienkiewicz, G. Paltoglou, K. Buckley, M. Thelwall, A. Kappas, J.A. Hołyst, *PLoS ONE* **6**, e22207 (2011).

[11] P. Weroński, J. Sienkiewicz, G. Paltoglou, K. Buckley, M. Thelwall, J.A. Hołyst, *Acta Phys. Pol. A* **121**, B-128 (2012).

[12] D. Garcia, A. Garas, F. Schweitzer, *EPJ Data Sci.* **1**, 3 (2012).

[13] A. Garas, D. Garcia, F. Schweitzer, *Sci. Rep.* **2**, 402 (2012).

[14] A. Czaplicka, A. Chmiel, J.A. Hołyst, *Acta Phys. Pol. A* **117**, 688 (2010).

[15] A. Czaplicka, J.A. Hołyst, *Int. J. Mod. Phys. C* **23**, 1250020 (2012).

[16] A. Chmiel, J.A. Hołyst, *Int. J. Mod. Phys. C* **21**, 593 (2010).

[17] S. Rank, in: *Proc. Agent-Directed Simulation Symp. 2010 Spring Simulation Conf.*, Eds. T. Oeren, L. Yilmaz, Society for Modeling & Simulation International (SCS), Orlando 2010, p. 82.

[18] S. Gobron, J. Ahn, G. Paltoglou, M. Thelwall, D. Thalmann, *Vis. Comput.* **26**, 505 (2010).

[19] M. Skowron, H. Pirker, S. Rank, G. Paltoglou, S. Gobron, in: *Proc. 24th Int. FLAIRS Conf.*, Eds. R.Ch. Murray, P.M. McCarthy, AIII Press, Palm Beach 2011, p. 117.

[20] M. Skowron, S. Rank, M. Theunis, J. Sienkiewicz, *Lecture Notes in Computer Science* **6974**, 337 (2011).

[21] M. Thelwall, D. Wilkinson, S. Uppal, *J. Am. Soc. Inform. Sci. Technol.* **61**, 190 (2010).

[22] P. Sobkowicz, A. Sobkowicz, *Eur. Phys. J. B* **73**, 633 (2010).

[23] B. Pang, L. Lee, *Foundat. Trends Inform. Retriev.* **2**, 1 (2008).

[24] F. Sebastiani, *ACM Comput. Surv.* **34**, 1 (2002).

[25] I. Ounis, C. Macdonald, I. Soboroff, in: *Proc. Second Int. Conf. on Weblogs and Social Media*, Eds. E. Adar, M. Hurst, T. Finin, N. Glance, N. Nicolov, B. Steng, AAAI Press, Menlo Park 2008, p. 93.

[26] F. Peng, D. Schuurmans, S. Wang, in: *Proc. HLT-NAACL '03*, Eds. E. Hovy, M. Hearst, M. Ostendorf, Association for Computational Linguistics, Edmonton 2003, p. 110.

[27] `www.digg.com` .

[28] K. Lerman, A. Galstyan, in: *Proc. WOSP*, Eds. Ch. Faloutsos, T. Karagiannis, P. Rodriguez, ACM, New York 2008, p. 7.

[29] Y. Zhu, in: *Proc. 19th Int. Conf. on WWW*, Eds. M. Rappa, P. Jones, ACM, New York 2010, p. 1039.

[30] G. Szabo, B. Huberman, *Commun. ACM* **53**, 80 (2010).

[31] K. Lerman, R. Ghosh, in: *Proc. Fourth Int. AAAI Conf. on Weblogs and Social Media*, Eds. M. Hearst, W. Cohen, S. Gosling, AAAI Press, Menlo Park 2010, p. 90.

[32] H. Rangwala, S. Jamali, *IEEE Intellig. Syst.* **25**, 36 (2010).

[33] F. Wu, B. Huberman, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17599 (2007).

[34] P. Van Mieghem, N. Blenn, C. Doerr, *Eur. Phys. J. B* **83**, 251 (2011).

[35] P. Krapivsky, S. Redner, F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).

[36] A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999).

[37] P. Krapivsky, S. Redner, *Phys. Rev. E* **63**, 066123 (2001).