Soc. Netw. Anal. Min. (2017) 7:46
https://doi.org/10.1007/s13278-017-0465-y

ORIGINAL ARTICLE

Effects of emotion and topic area on topic shifts in social media discussions

Kamil Topal[1] · **Mehmet Koyutürk**[1] · **Gültekin Özsoyoğlu**[1]

Received: 15 December 2016 / Revised: 20 September 2017 / Accepted: 20 September 2017 / Published online: 4 October 2017

Abstract Nowadays, discussing and commenting interactively on an article in Internet-based social media platforms is pervasive. The topic of a comment/reply in these discussions occasionally shifts, sometimes drastically and abruptly, other times slightly, away from the topic of the article. In this paper, we model and study the topic shift phenomena in article-originated social media comments, and identify quantitatively the effects on topic shifts of comments' (1) emotion levels (of various emotion dimensions), (2) topic areas, and (3) the structure of the discussion tree. We then propose and evaluate a new approach to measure and visualize named emotion scores of comment sets. We show that, with a better understanding of the topic shift phenomena in comments, personalized automated systems can be built to cater to comment-browsing and comment-viewing needs of different users.

Keywords Social media analysis · Topic shift · Emotion analysis · Emotion visualization

1 Introduction

Users of internet-based social media platforms commonly discuss/comment on a "topic" of their choice in an interactive manner. Usually, such social media discussions start by an article on the web that covers an event, a product, a situation, etc. Comments in these discussions have no size restrictions, allowing people to express their opinions more completely as compared to twitter tweets.

Comments in a social media discussion usually start around the topic of an article, and quite commonly shift, sometimes drastically and abruptly, other times slightly, away from the original topic. These topic shifts take away from the discussion at hand and create problems. In fact, due to large numbers of unrelated, inflammatory, or uncivilized comments, as well as the large numbers of unrelated and/or shifted comments on some popular articles, news Web sites and blogs have started to eliminate their comment sections (Gross 2014). This is an unfortunate action as readers of a discussion, not just the commenters, gather useful information by the simple act of reading these, sometimes informed, comments, and form more informed opinions themselves—a significant loss for both readers and those Web sites and blogs that eliminate comments from their software systems.

We hypothesize in this paper that there are three causes for topic shifts in comments: emotion levels, the specific area of the article (e.g., sports, politics, cancer, etc.), and the structure of the "discussion" (comment) "tree." These three factors collectively play a role on topic shifts in comments; and, understanding their roles in more depth can lead to building better automated comment-viewing software systems that help readers sift through large numbers of comments and gather information more effectively.

Motivated by our main hypothesis, we study the phenomena of topic shifts in article-originated social media comments. We attempt to identify quantitatively the effects on topic shifts of comments' (1) emotion levels (of various emotion dimensions), (2) topic areas, (3) the structure of the discussion tree. We show that, with a better understanding of the topic shift

✉ Kamil Topal
kamil.topal@case.edu

Mehmet Koyutürk
mehmet.koyuturk@case.edu

Gültekin Özsoyoğlu
tekin@case.edu

[1] Case Western Reserve University, Cleveland, USA

phenomena in comments, automated systems can easily be built to personalize and cater to the comment-browsing and comment-viewing needs of different users: users can be provided with options in real-time to (a) selectively view and reply to comments or discussion threads that are "on the topic," or within a range of either the original article or a specific, possibly shifted, comment of interest within the discussion tree, (b) link and view discussions of interest in temporal order even when they belong to different discussion threads within the discussion tree, (c) prune the discussion tree in real-time by specifically eliminating those discussion threads that are of no interest to them, (d) view comments from all over the discussion tree that may have shifted from the original topic in a certain way, such as shifted to a certain "drifted topic," or (e) visualize on-the-topic/shifted comment sets. This paper only discusses items (b), (d), and (e), as items (a) and (c) are left as future research topics.

In addition to topic shift analysis, this paper presents and evaluates a new comment set visualization and analysis technique (based on dimension reduction Fodor 2002 via singular value decomposition Wall et al. 2003) to visualize aggregated emotion levels of on-the-topic and shifted comment sets that can perhaps be used by comment readers to pass a judgment on the properties of the comment set at hand. This technique also allows for mass separation of on-the-topic and off-the topic comments, allowing for the possibility of building automated systems that, at reader's discretion eliminate shifted comments from their view of the full discussion tree. Finally, we propose using more complex named emotion scores both for topic shift analysis and for future research.

This paper is an extended version of work published in Topal et al. (2016), with both extensions to the analysis of shifted topic analysis and the newly proposed comment set visualization and analysis technique.

For our experimental studies, we have collected about 580,000 news article comments on ten topics in different areas (though, due to space restrictions, we only discuss results of six topics) and analyzed the effects of three factors on topic shift: (1) the comment's location within the discussion tree—in terms of both the level and path of the comment within the tree, (2) comments' emotion dimensions (i.e., sensitivity, aptitude, attention and pleasantness) and the associated emotion levels (e.g., for the sensitivity dimension, the six levels are rage, anger, annoyance, apprehension, fear, and terror), and (3) the topic area (e.g., sports, politics, or health). We have found that:

- In terms of a comment's location in the discussion tree, the first comment of the discussion tree sets the tone for all of its descendants: if it is on the topic, usually, the descendant comments in its discussion subtree also stay on the topic. In rare occasions where a descendant comment, say c, is off-topic (i.e., has a topic shift), regardless of the location of c in the discussion tree, most ($\sim 85\%$) of the descendant comments of c also end up having topic shifts of varying degrees.
- The role of emotion on topic shifts, as one would expect, is very significant: different emotion levels in different emotion dimensions cause differing degrees of topic shifts: highly emotional comments (such as those with high sensitivity dimension scores, e.g., *rage* and *terror* emotion levels) shift away from their original topics with very high frequency (around 90% of the time). And, comments with high emotion levels in emotion dimensions *sensitivity* and *aptitude* are associated with higher topic shift frequencies, as compared to comments with high emotion levels of *attention* and *pleasantness* dimensions.
- The role of the topic area on topic shifts is also quite significant: topic areas such as sports or politics are more prone to higher levels topic shifts in comments (perhaps because they evoke higher levels of emotions on commentators) than other topic areas such as health (also perhaps because they evoke lower levels of emotions on commentators). This leads us to believe that all topic areas can easily be pre-classified as *high, medium,* or *low* emotion level provoking topic areas. Automated tools can then be built to help users identify (and take actions such as perhaps not view) comments with certain types of topic shifts, taking into account this classification and other factors.
- Topic shifts can easily be predicted via unsupervised or supervised learning techniques with around 80% accuracy based on the comments' emotion levels.
- One can easily partition and visually observe on-the-topic and off-the topic comment sets via the SVD-based dimension reduction technique introduced here.
- In addition to extracting emotional dimension scores of comments and identifying the corresponding emotion scores on four different dimensions, one can also obtain pairwise combinations of emotion dimension scores, and observe more complex named emotions within comments such as *love*, *optimism*, *rivalry*, etc. We believe these scores can also provide additional hints to readers' in choosing what to read and what to skip within a comment discussion tree, though this research direction is not pursued further in this paper.

2 Related work

2.1 Topic Shifts

The notion of topic shift has been studied in the field of web community discovery (Kleinberg et al. 1999) via

focused/topical crawlers (Manning et al. 2008), to identify those web pages (i.e., documents) that "stay on the topic at hand" (He et al. 2002), using an information retrieval model, usually a vector space model (Manning et al. 2008), that characterizes the topic of each web page and the distance between two pages that specifies the amount of topic shift. This approach is used in many other environments, e.g., O'Hare et al. (2009) apply sentiment analysis to financial blog corpus and identify topic shifts among documents in that corpus. Liu et al. (2013) study topic drift on micro blog posts by using Latent Dirichlet Allocation model. Knights et al. (2009) detect topic drift with Compound Topic Models, to see how a topic evolves and changes to a different topic over a specific time. Our study borrows from these studies in that we also use the vector space model. However, we are mostly interested in the causes of topic shift, and add emotion to our model. Vector space model has numerous advantages over its alternatives; it can extract the knowledge from text itself without using any lexicon, and performs very well measuring similarity between texts (Turney and Pantel 2010).

In identifying topic shifts between an article/comment and another comment, we take an approach similar to topic shift detection in web community discovery, with a number of provisions, namely discussion tree structure, revised comment similarity score functions in discussion trees, and emotion dimensions.

2.2 Emotion modeling

In social media discussions, commenters' emotions influence their comments, which in turn cause abrupt or slowly-changing topic shifts from one comment to another. To this end, there is a need to identify/classify emotions of commenters and investigate the causal effects of different emotions. In a recent study, Hasan et al. (2014) build and use a system, called EMOTEX, to extract emotions from Twitter data. To label data for training, EMOTEX uses Twitter hashtags, without an effort to annotate data for any form of learning. In comparison, article-based comments do not contain annotated data. For this reason, we use manually labeled data to extract comment emotions.

Sentiment Analysis, or Opinion Mining, aims to find the polarity (Wilson et al. 2005) of sentiments and detects their subjectivity (Wiebe and Ellen 2005) via Natural Language Processing techniques. These techniques usually produce two or three labels for documents, e.g., positive, negative, or neutral sentiments, together with a score ranging between two polarities of, say, -1 and 1. There are free, academic, or commercial tools available for sentiment analysis (NLP 2013; Alchemy 2015). There are two main models for representing emotions. The Circumplex model (Russell 1980) characterizes emotions in two dimensions: activation and pleasure with General Inquirer (Smith et al. 1967) database containing lexicon of emotions, and more than 100 categories and 11,000 words.

The Hourglass model (Cambria et al. 2012), the most recent emotion categorization proposed by Cambria et al. (2012), uses a more advanced model, and has four independent, but concomitant, dimensions, namely *pleasantness, attention, sensitivity,* and *aptitude* (Fig. 1). Each dimension captures a different type of emotion: (a) *Pleasantness* captures the user's "amusement level" with interaction modalities, (b) *Attention* captures interaction contents, (c) *Sensitivity* captures the comfort level of the user with interaction dynamics, and (d) *Aptitude* captures the user's confidence in interaction benefits. As also seen in Fig. 1, each of the four dimensions of the Hourglass model has six levels of activation, which collectively characterize the emotional state of an individual. As an example, *pleasantness* has six different activation levels, namely, *ecstasy* (the most pleasant), *joy, security, pensiveness, sadness,* and *grief* (The least pleasant). To form a dataset, Cambria et al. (2012) first create the AffectNet dataset by blending ConceptNet (Havasi et al. 2007) and WordNet-Affect (Strapparava and Valitutti 2004) datasets. Then, they apply truncated singular value decomposition on AffectNet, and use dimension reduction on AffectNet by finding the best approximation. Finally, they use the k-means approach to cluster Sentic space to the Hourglass model. SenticNet 3.0 (http://sentic.net/downloads/) database, which is publicly available, has more than 30,000 words and phrases that are already scored (in the range of $[-1, 1]$) for all dimensions. A snapshot of the SenticNet database is in Table 2. SenticNet database also has polarity scores for each word. In this paper, we use the Hourglass Model to classify the emotion dimensions and levels of commenters.

3 Modeling article and comment similarities, and topic shifts

3.1 Discussion trees

We present an abstract representation of the structure of a social media discussion via its discussion tree. We identify the main characteristics of (article-based and size-free) social media discussions as follows:

a. Each comment is either about the original article, or a reply to another comment/reply.
b. Each comment has a timestamp, indicating the posting time of the comment.

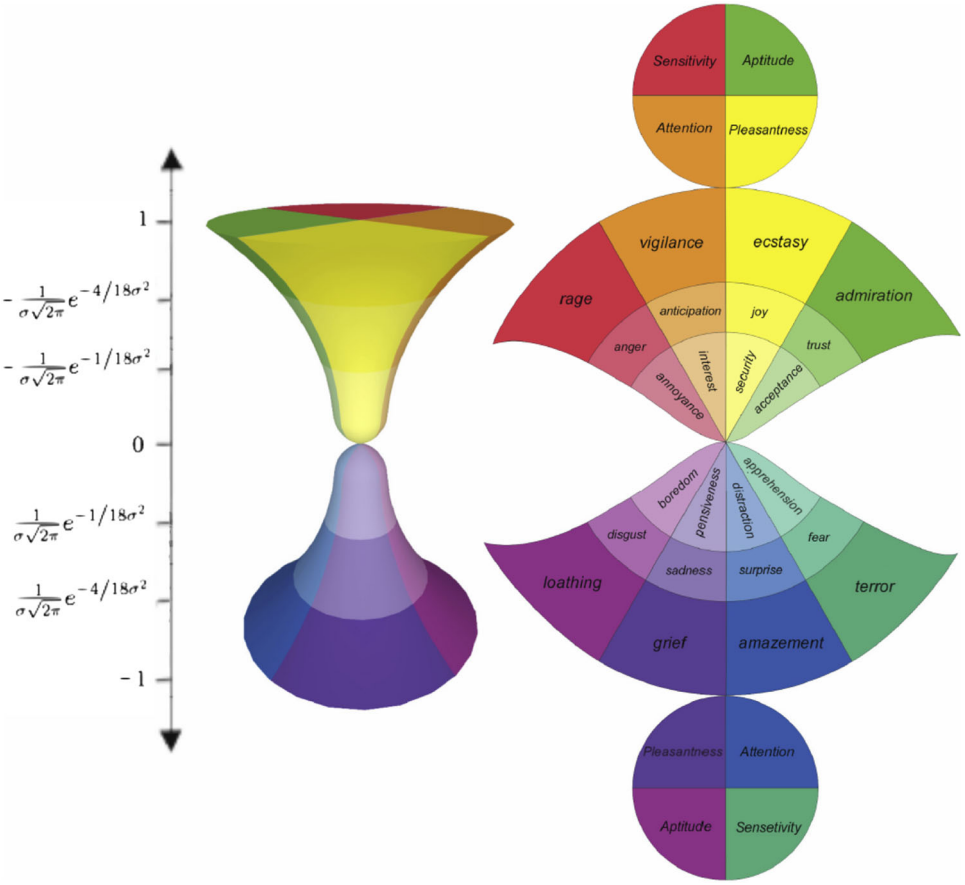


Fig. 1 Hourglass of emotion model (Cambria et al. 2012). Please see all the dimension and emotion names in Table 1

Table 1 Hourglass of emotion model

	Pleasantness	Attention	Sensitivity	Aptitude
+ 3	Ecstasy	Vigilance	Rage	Admiration
+ 2	Joy	Anticipation	Anger	Trust
+ 1	Serenity	Interest	Annoyance	Acceptance
0				
− 1	Pensiveness	Distraction	Apprehension	Boredom
− 2	Sadness	Surprise	Fear	Disgust
− 3	Grief	Amazement	Terror	Loathing

Table 2 A snapshot of SenticNet database

	Pleasantness	Attention	Sensitivity	Aptitude
Accept	0.972	0	0	0.894
Acceptability	− 0.997	0	0.878	− 0.957
Acceptable	0.258	0.357	− 0.088	0.351
Acceptableness	− 0.997	0	0.878	− 0.957
Acceptance	0	0	0	0.3
Acceptation	0.577	− 0.77	0	0.517
Accepted	0.683	− 0.553	0	0.552

c. Each sequence of comments is either about the original article (represented by the topic t) or another "parent" comment (represented by its own topic).
d. Each comment sequence has a nesting level, which is equal to the number of comments preceding that comment in the reply-chain that contains the comment.

3.2 Vector space model

We use the vector space representation of the article and comments. The article and each comment are tokenized and stemmed using the Porter stemming algorithm (Porter 1980). Word tokenization involves removing characters from words (such as punctuations), and attaching a unique id for each word. Stemming and stop-word removal are applied. Related words are mapped to the same stem by removing their inflections. Stop-words are common in sentences, and add grammatical, but no context, value, and thus, they are not useful to determine keywords for a topic. Some researches do use stop-words, however. There is no universal stop-word list in the literature. In our study, for stop-word removal, we employ a comprehensive list from the web (XPO6 2009). Then, we use a modified version of the vector space model (Salton and Buckley 1988) as

follows. Term frequency for each document (article story and comments) is calculated using the Cornell SMART system's smoothened version (Salton and Buckley 1988). Let t be a term in document d, where d is a comment or the article. Then, the term frequency $TF(t, d)$ is computed as follows. $n(t, d)$ = frequency of term t in document d

$$TF(t,d) = \begin{cases} 0, & \text{if } n(t,d) = 0 \\ 1 + log(1 + n(t,d)), & \text{otherwise} \end{cases}$$

We compute the inverse domain frequency of each term t across all documents, $IDF(t)$, to scale up the effects of terms that occur in many comments or the article.

$$IDF(t) = log\left(\frac{1 + |D|}{|D_t|}\right) \quad (1)$$

Here, D denotes the document collection (in our case, the set of comments and the article), D_t denotes the set of documents containing t, and $log()$ is a dampening function. Note that this analysis is performed independently for each article and its associated comments; i.e., the set D is unique for each article. We then compute the relative frequency $x_d(t)$ of term t in document d as

$$x_d(t) = \frac{TF(t,d)}{IDF(t)} \quad (2)$$

Clearly, Eq. 2 is the opposite of the standard approach of $TF() * IDF()$ where rare terms are considered important, and "rewarded" by the $IDF()$ factor. In our case, however, the "universe" of the documents for each article is composed of the article and its associated comments. Therefore, a term that is frequent in this collection of documents indicates relevance to the article, whereas rare terms signal decrease in importance. In other words, our premise is that important words are not usually rare. For example, if the article is about Ebola virus dissemination, then the relevant comments are more likely to have the terms "Ebola," "hospital," "health," etc. If we multiply $TF()$ with $IDF()$ scores, then these frequent terms will get smaller weights, which is an undesirable effect. On the other hand, dividing $TF()$ by $IDF()$ assigns more weight to these frequent terms. Furthermore, if a comment includes a term that is rarely used in other comments as well as the article, dividing $TF()$ by $IDF()$ lowers that term's weight. Note that uninformative words are already removed in the stop-word removal stage of analysis.

3.3 Similarity scores

Since we have a vector representation of each comment and the article, we use the cosine similarity (Manning et al. 2008) (Eq. 3) to calculate the topical similarity between the article and each comment. Namely, for a given article a and comment c, we compute the similarity between a and c as

$$C(\mathbf{x_a}, \mathbf{x_c}) = \frac{\mathbf{x_a}\mathbf{x_c}}{|\mathbf{x_a}||\mathbf{x_c}|} = \frac{\sum_{t=1}^{n} x_{a(t)} x_{c(t)}}{\sqrt{\sum_{t=1}^{n} x_{a(t)}{}^2}\sqrt{\sum_{t=1}^{n} x_{c(t)}{}^2}} \quad (3)$$

where x_a and x_c, respectively, denote the vector space representation of a and c, and n denotes the total number of terms. In choosing cosine similarity, we experimented with Jaccard index (Hamers et al. 1989), dice similarity (Murguia and Villasenor 2003), and cosine similarity in a small set of comments, manually judged their performance, and chose cosine similarity since it performed slightly better than the others in our environment.

Once we quantify the similarity between each article and comment, we use a threshold to distinguish between off-the-topic and on-the-topic comments. In order to set the threshold, we use the k-means algorithm (Hartigan and Wong 1979) to create two different clusters for off-the-topic and on-the topic comment sets, and consider small centroid clusters as having shifted comments.

3.4 Emotion modeling

To represent the emotional landscape of each comment, we use the SenticNet 3.0 database (http://sentic.net/downloads/) containing a large collection of phrases. For each phrase in the database, there are five different scores (one for each of the four emotion dimensions, and a polarity score) in the range $[-1, 1]$. We introduce two different emotion score computation techniques per dimension, one that uses (aggregated) absolute emotion scores (and thus merges negative and positive emotion activation levels to only observe the level of topic shifts), and another one that obtains exact named emotion scores using dimension reduction via SVD to both visualize topic shifts via heatmaps, and to investigate the properties of comments in more depth.

3.4.1 Computing aggregated scores across emotion dimensions for topic shift analysis

We map each comment to the SenticNet database by identifying all phrases that match the comment in SenticNet. Since each comment may map to multiple phrases in the database, we then aggregate the scores of each phrase to compute an emotion representation for the comment. For this purpose, for each emotion dimension, we compute the average of the absolute values of the respective dimension score across all matching phrases. This gives us a five-dimensional representation of the emotional landscape of the comment. We average the absolute values of the scores, since we are mainly interested in quantifying the "level" of emotionality of the

comment, as opposed to quantifying the polarity of the emotion.

In the Hourglass of emotion model (Cambria et al. 2012), there are four dimensions with scores in $[-1, 1]$, namely *pleasantness*, *attention*, *sensitivity*, and *aptitude*, each with six levels of activation that represent six different emotion levels. As an example, there are *grief*, *sadness*, *pensiveness*, *security*, *joy*, and *ecstasy* in the *Pleasantness* dimension. We eliminate polarity by taking absolute values, and, thus, the computed dimension scores are in the range of 0–1, resulting in 3 distinct levels of activations in each dimension (instead of 6), with the emotion levels in each dimension symmetrically combined. E.g., if the *pleasantness* dimension score is less than -0.66 than it is *grief*; if it is between -0.66 and -0.33, then it is *sadness* and so on. After taking absolute values of all scores, we end up with three combined emotion levels for *Pleasantness*, namely, {*grief*, *ecstasy*} ("high" emotion level), {*sadness, joy*} ("medium" emotion level), and {*pensiveness, security*} ("low" emotion level).

3.4.2 Computing named emotion scores for topic shift analysis

We use dimension reduction for each comment to obtain a single comment score per emotion dimension. The scores of all comments for a news article m per emotion dimension d_i forms the "emotion vector of news m for emotion dimension d_i."

Let a news article m have n comments, and the maximum number of SenticNet word instances in any comment of news article m, duplicates included, be k. We create a k by n matrix to store all emotion scores of news article m for emotion dimension, say, d_i. If a comment has less than k scores, we use bootstrapping (Freedman 1981) to k dimensions for that comment. Next, in order to obtain an emotion dimension-specific comment score for a comment of the news article m, we reduce the matrix dimension from k by n to 1 by n by using singular value decomposition (SVD), as shown in Eqs. 4 and 5.

$$M_{d_i} = U_{d_i} \Sigma_{d_i} V^*_{d_i} \tag{4}$$

M_{d_i} is a $k \times n$ original matrix for each dimension, U_{d_i} is a $k \times k$ unitary matrix, Σ_{d_i} is a $k \times n$ diagonal matrix with nonnegative real number on the diagonal where the diagonal entries are singular values such as $\sigma_{11} > \sigma_{22} > \sigma_{33} > \cdots$, and $V^*_{d_i}$ is $n \times n$ unitary matrix.

$$E_{d_i} = \left(U^{(1)}_{d_i}\right)^T M_{d_i} \tag{5}$$

Since σ_{11} is the largest singular value, we take $U^{(1)}_{d_i}$ as first column of U_{d_i} and transpose it from $k \times 1$ to $1 \times k$ and multiply by M_{d_i}, then we get E_{d_i} a $1 \times n$ vector, which contains a single comment score for each comment, which we call the "emotion vector of news article m for dimension d_i," where d_i is one of dimensions *pleasantness, attention, sensitivity*, or *aptitude*.

To visualize the aggregated per-dimension emotion score of a comment, we use the news article's emotion map which is really a special heatmap (Skuta et al. 2014), a two-dimensional graphical representation of data with colors instead of numbers. Cluster heatmaps are suited for row/column hierarchical clustering in order to observe similar data within the same cluster. In our case, since each comment has four different scores, we cluster over comments (columns) and emotion dimensions (rows). Also, within any cluster, subclusters where emotional activations are slightly different can be observed. Figures 9 and 13 are examples of heatmaps consisting of randomly chosen 200 comments. On-the-topic and shifted comment sets can then be instantly identified via color differences.

Finally, the Hourglass model has a more complex level of emotions obtained by combining two emotional dimensions as seen in Fig. 2 which represent more complex named emotions such as "love," "optimism," "rivalry," etc. We analyze these emotions in terms of topic shifts in Sect. 4.2.5.

4 Experimental evaluation and results

4.1 Dataset

Our datasets come from news article comments on various topics (Table 3). We collected 581,952 comments from 130 news articles in the period of June 2015–September 2015 from blog comment hosting service Disqus API (Disqus 2015). We chose Disqus for the following reasons:

- It provides a service for major Web sites such as Politico, CNBC, ABC News, and The Washington Times.
- It allows replies to all comments and replies, which allows for deeper discussion trees shift behavior as a function of the structure tree.
- Discussion trees can be created as JSON-formatted data.

Our data preprocessing removes all spam comments in each dataset. Each discussion tree and the level of each comment are extracted based on the reply relationship. Figure 3 lists six different topic similarity distributions as histograms, each with 40 buckets, a title, and a threshold value (as determined by the k-means algorithm) in log scale. Because of space limitations, we here focus on the results related to these six topics and their discussion trees.

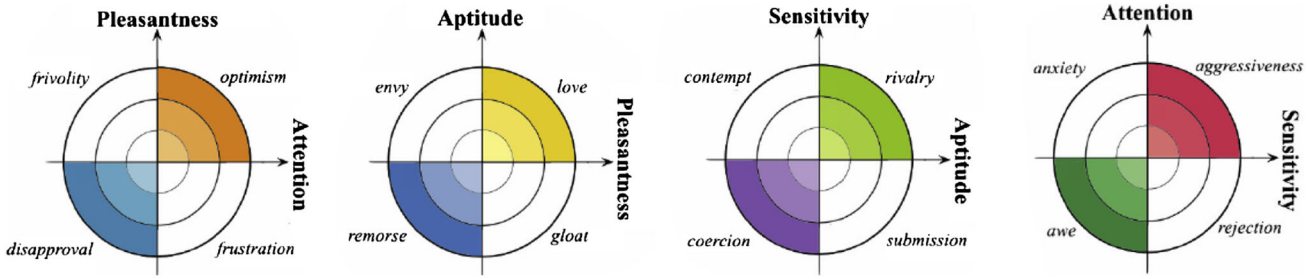


Fig. 2 Compound emotions of second levels: combination of two dimensions (Cambria et al. 2012)

Table 3 Topics and number of comments of our dataset

News article topic	# of Articles	# of Comments
2016 US presidential election debates	11	83,604
Gun laws discussion	5	18,480
Immigration in USA	11	34,481
Supreme Court decision on LGBT marriage	16	83,487
Media Brawl with politicians and journalist	9	89,243
Hillary Clinton email controversy	9	93,580
Iran nuclear deal with USA	13	58,388
Stocks	20	16,788
Planned parenthood and abortion	19	87,946
Economy	17	15,955
Total	**130**	**581,952**

4.2 Topic shift analysis

4.2.1 Level-based analysis of shifted comments

Each comment has its level information: level 1 (root) nodes are comments directly on the article, and level 2 nodes are replies to a level 1 comment, and so on. We gather all comments about an article, and group the comments logarithmically according to their levels into buckets. Figure 4a–f contains *box plots* (FlowingData 2008), one for each bucket, summarizing the similarity distribution of comments in each bucket to the article. In each box plot of Fig. 4a–f, the *X*-axis represents the levels of comments represented by a bucket, and the *Y*-axis displays five different statistics for similarity scores between comments within that bucket and the article, namely minimum, first quartile, median, third quartile, and maximum (FlowingData 2008). Note that the X dimension scale grows exponentially since the number of comments at higher level buckets decreases exponentially. Thus, logarithmic bucketing provides a more balanced distribution of comments into buckets (but the number comments in a bucket still goes down with increasing number of levels). We have the following observations:

Observation 1 Among discussion tree levels for a topic, the level with the highest average similarity score to the article is level 1, i.e., the root.

That is, root-level comments (per discussion tree) are on the average the most similar comments to the article. Observation 1 is to be expected since topic shifts are likely to occur more frequently as the discussion thread continues.

There are exceptions to Observation 1 within our topics, namely *Gun Laws* and *Hillary Clinton Email Controversy*. In these discussions, if the comment's level goes deep enough in the discussion tree, i.e., after level 32, surprisingly, the average of on-the topic comments exceeds their root averages. This shows us that there are still on-the-topic (and, perhaps, useful) comments at deeper levels among the descendants of shifted comments.

Observation 2 Topical similarity to the article is highly variable for comments that are at different levels. Even in buckets that have low similarity to the article on average, there are still significant numbers of useful comments, as represented by their higher similarity scores. At least 25% of all comments in each bucket have similarity scores higher than 0.3, which are classified as on-the-topic comments by the k-means algorithm. This observation suggests that the level of a comment in the discussion tree may not be sufficient to predict the relevance or usefulness of a comment.

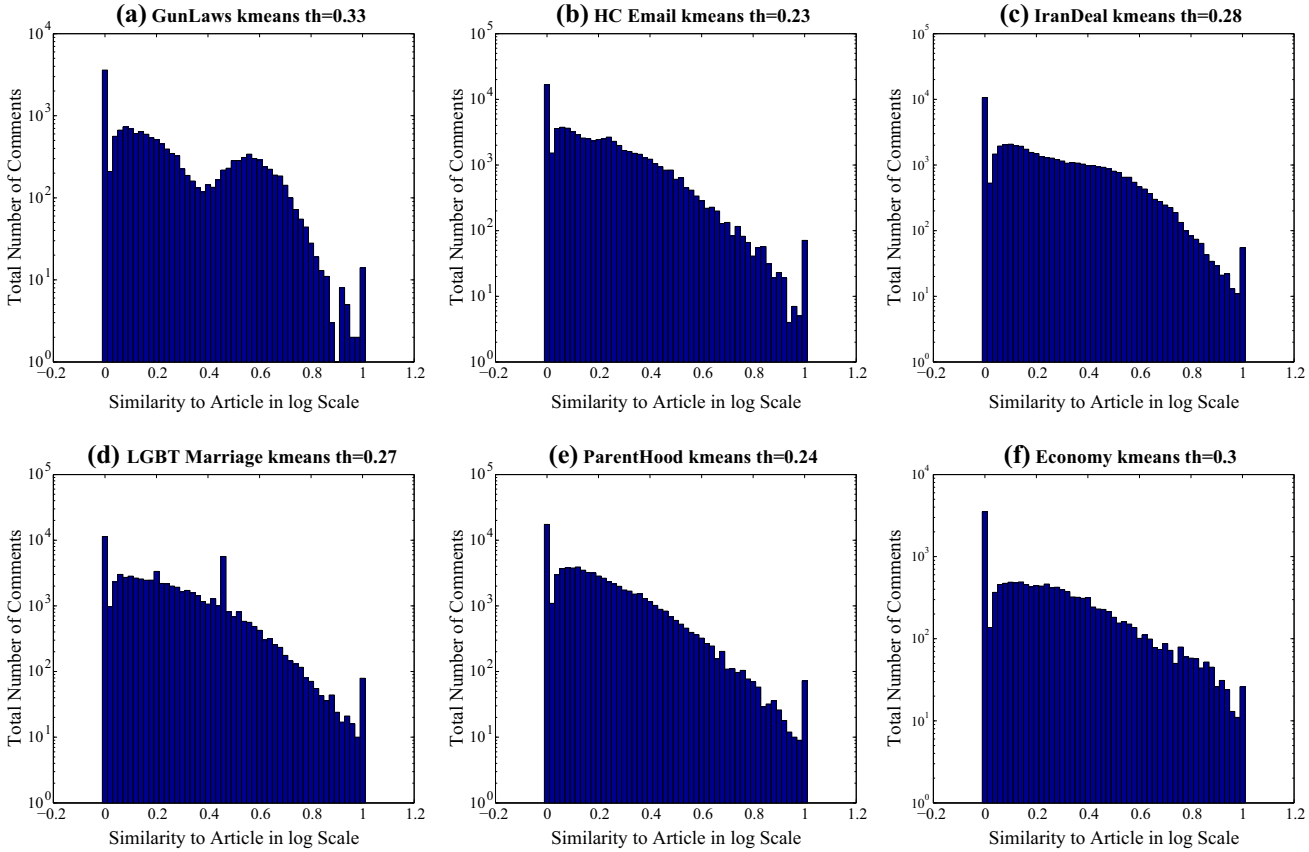


Fig. 3 Article-comment similarity distribution for six different topics in log scale. Each subfigure has a title that includes their topic name and threshold score for separation of on-topic and off-topic comment set

4.2.2 Topic shifts in similar and dissimilar trees

We distinguish between "low similarity" and "high similarity" comment sets by employing a similarity threshold α defined as the maximum similarity score of the low similarity comment set. We run the k-means algorithm ($k = 2$), and the cluster having a higher centroid is the on-the-topic comment set. So, the minimum value of the on-the-topic comment set is α.

Defn-dissimilar/similar discussion tree: Let β_i be the average similarity of all the comments in the discussion tree $T_i, 1 \leq i \leq n$. Then, T_i is a similar tree (to the article) if $\beta_i > \alpha$; otherwise, it is a dissimilar tree.

In Fig. 5, we dissect the analysis of Fig. 4 into dissimilar and similar trees, by plotting the distribution of similarity to the article separately for comments in dissimilar trees and those in similar trees. For this purpose, we use the logarithmic bucketing of Fig. 4 (labels omitted for readability), and box plot similarity scores of similar/dissimilar trees per bucket, as error bars (Motulsky 2002) for dissimilar and similar discussion trees. The purpose of this analysis is to understand whether trees are uniform in terms of the "topic shift" behavior of the comments they contain.

Observation 3 Similar trees have more on-the-topic comments even at deeper levels, compared to dissimilar trees. For all topics, similar trees have on-the-topic root comments 65–85% of the time. In comparison, these values decrease at least 20% for dissimilar trees.

Observation 4 In each discussion tree, the root comment for each article sets the tone and mostly decides as to how the following discussions evolve. If a root stays on the topic, then the following comments usually also stay on the topic (i.e., result in a similar tree), at least 20% more frequently than those for a dissimilar tree. In comparison, for a dissimilar tree, there is up to 50% more shifted root comments, and this leads to, on average, a 40% higher topic shift.

4.2.3 Relationship between topic shifts and the topic area

Some of the comments we extracted from Disqus do not match any phrases in SenticNet dataset. For example "Nope" or "So does Crump and Truz!" are comments for

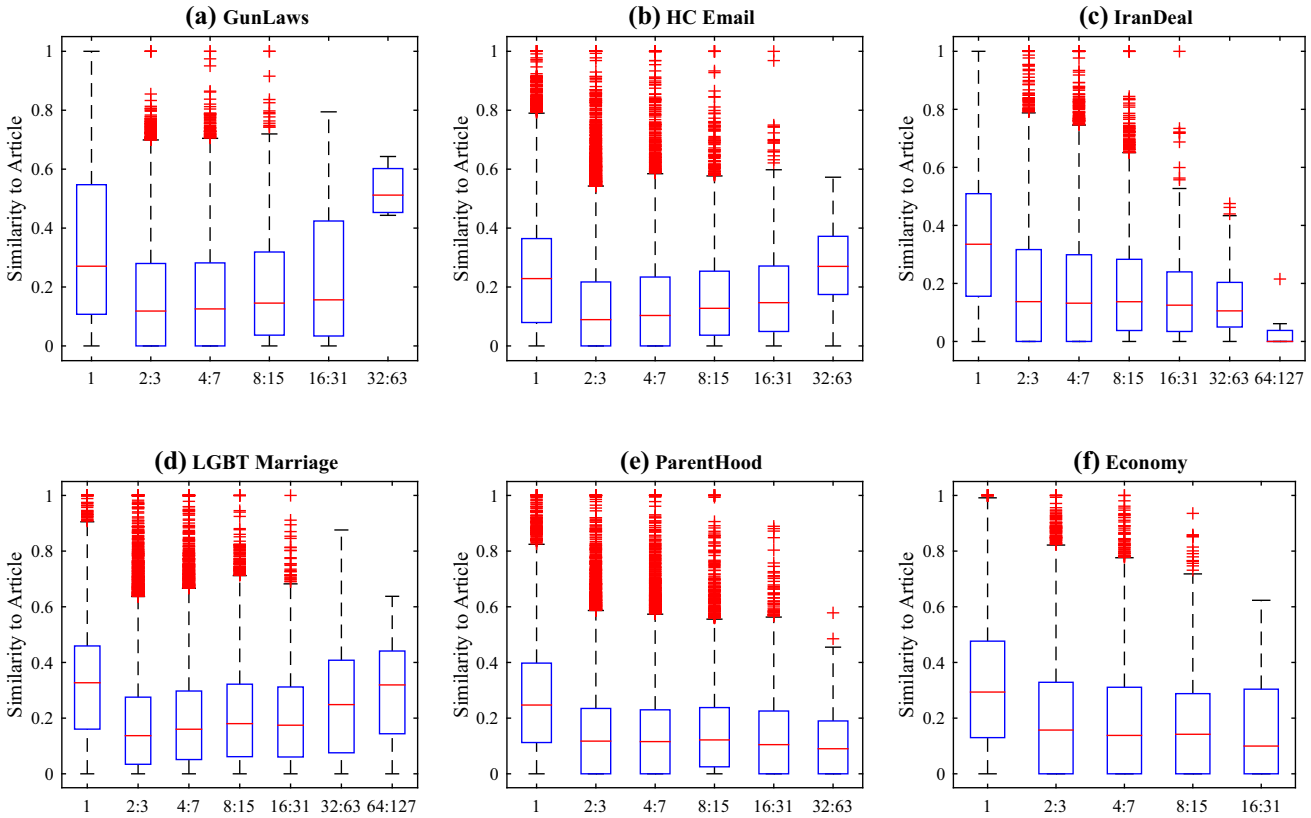


Fig. 4 The relationship between the level of an article in a conversation tree and its topical similarity to the article for six different topics. The *x*-axis shows the level of the comment in its respective conversation tree, where comments are binned logarithmically

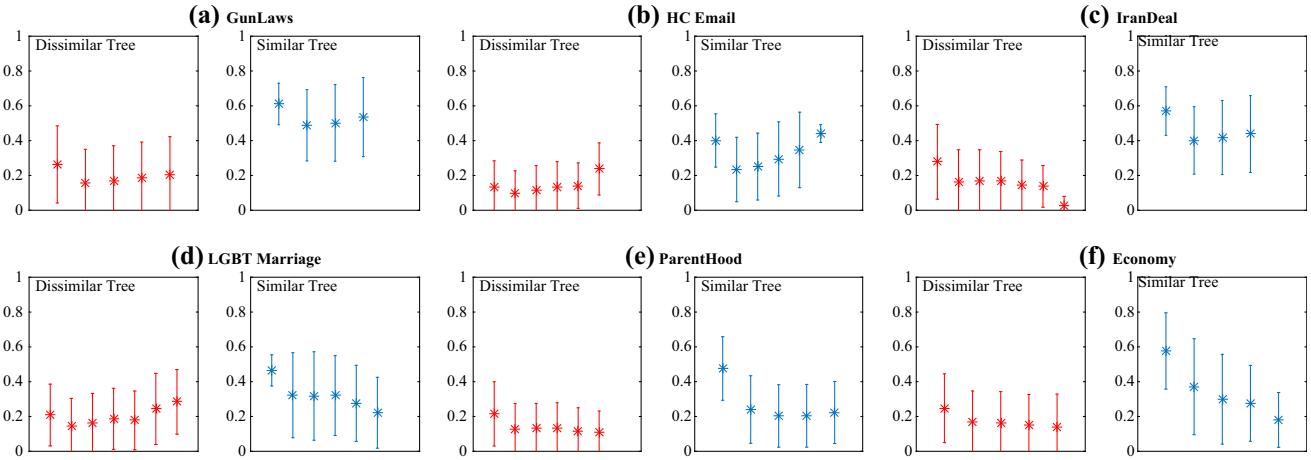


Fig. 5 The relationship between comment level and topical similarity to the article (*y*-axis) for dissimilar versus similar trees. The *x*-axis shows the level of the comment in its respective decision tree, where comments are binned logarithmically

which there is no emotional score in the SenticNet 2.0 database. (We have refrained from adding our emotion scores for such comments, even though one can, after a diligent due-analysis, add a content-based emotion score for such phrases). Also, some of these comments are only stop-words and sarcastic words like "Crump" (as opposed to (Donald) "Trump") (that would indeed evoke emotions on the replies and would shift them); we have similarly refrained from adding scores to such comments, and used a similarity score of 0 (i.e., "null emotion" comment) for such comments. Moreover, we removed these null emotion comments from our analysis. Figure 6 shows the

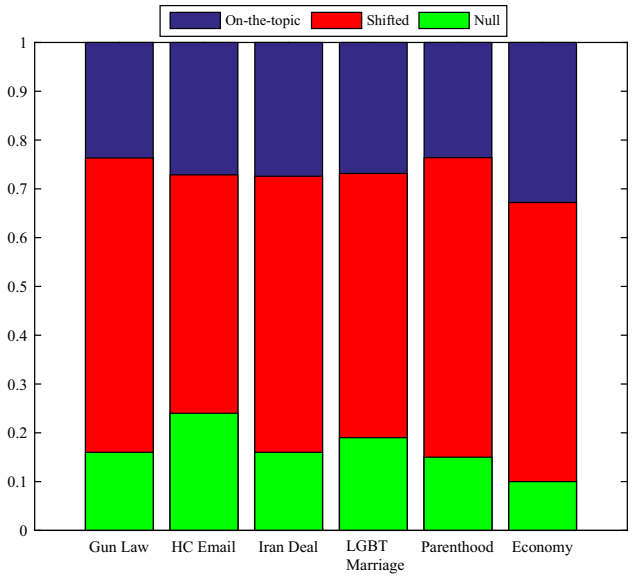


Fig. 6 On-the-topic, shifted, null comment percentages among different topics

percentages of on-the-topic, shifted, and null emotion comments for each topic.

Observation 5 After null score elimination, 64–72% of all comments are shifted from their original topics for all dataset.

The fraction of shifted comments varies topic by topic. For example, in *Hillary Clinton Email Controversy* 73% of comments are shifted away from the article topic. On the other hand, in *Economy* news article this number drops to 67%.

Hillary Clinton Email Controversy and *Supreme Court decision on LGBT marriage* news comments have more null comments than others.

Economy news comments have highest fraction of on-the-topic comments and the lowest fraction of null comments.

This section analyzes the effects of emotion on comments. For each comment, we compute the four emotional dimension scores, namely *pleasantness, sensitivity, attention*, and *aptitude*, as well as the similarity of the comment to the article, and compute the relevant statistics.

4.2.4 Effect of emotion on topic shifts

a) Emotion statistics

Figure 7 lists emotion scores of comments (per topic and per emotion dimension) as histograms, where, for each dimension, there are three bars. (Note that, by taking absolute values, we reduce the number of emotion activation levels from 6 to 3). Within each bar of a given emotion dimension, we capture two emotion scores for that dimension, i.e., blue and red, representing the percentages of on-the-topic and off-the-topic comments, respectively, in that emotion dimension.

Observation 6 For most topics, more than 50% of comments have low levels of emotion, i.e., located in the leftmost bar for each histogram.

E.g., *Hillary Clinton Email Controversy* comments in Fig. 7b, for the sensitivity dimension, 85% of all comments is in the lowest level (i.e., *apprehension* or *annoyance* emotions). However, even though most comments have low emotion levels, they nevertheless evoke replies with high emotion levels, causing topic shifts in their replies.

In the histograms of each emotion dimension in Fig. 7, the second bars (i.e., the medium-level emotion category with scores in the range [0.33, 0.66]) have the second largest numbers of comments, with percentages ranging from 14% and higher. There are some exceptions where most of the comments are located in this bar; e.g., for *LGBT marriage* comments (Fig. 7d), 56% of all comments are in the *Aptitude* dimension (i.e., *trust* and *disgust* emotions). The percentage of shifted comments in different dimensions of the medium-level emotion category ranges from 63 to 76%.

And, finally, in Fig. 7, the third bars (i.e., the high emotion level category) usually have the smallest number of comments, ranging from 2 to 12% of all comments. In comparison, 79–95% of all comments in this bar have shifted. For example, among *Gun Laws* (Fig. 7a) comments, the third bar of the *sensitivity* dimension (*rage* and *terror* emotions) has only 3% of all comments; but, 93% of these comments have shifted.

Observation 7 Almost in every dimension of all topics, the largest percentages of shifted comments are at the third (i.e., the highest emotion) bars.

As an exception, for *LGBT marriage* comments, in the *Pleasantness* dimension, 78% of the comments in the first (i.e., the lowest emotion) level (*serenity* and *pensiveness* emotions) have shifted; but in the third (i.e., the highest) level (*ecstasy* and *grief* emotions), this number is 79%; so, for *LGBT marriage* comments, high emotions do not affect topic shift itself.

Observation 8 *Pleasantness* and *aptitude* dimensions have, for all topics, more highly emotional comments (9–11%) than *sensitivity* and *attention* dimensions (3–4%).

Observation 9 The numbers of shifted comments increase for all topics when commenters choose words with higher emotion levels to express their opinion on a specific topic.

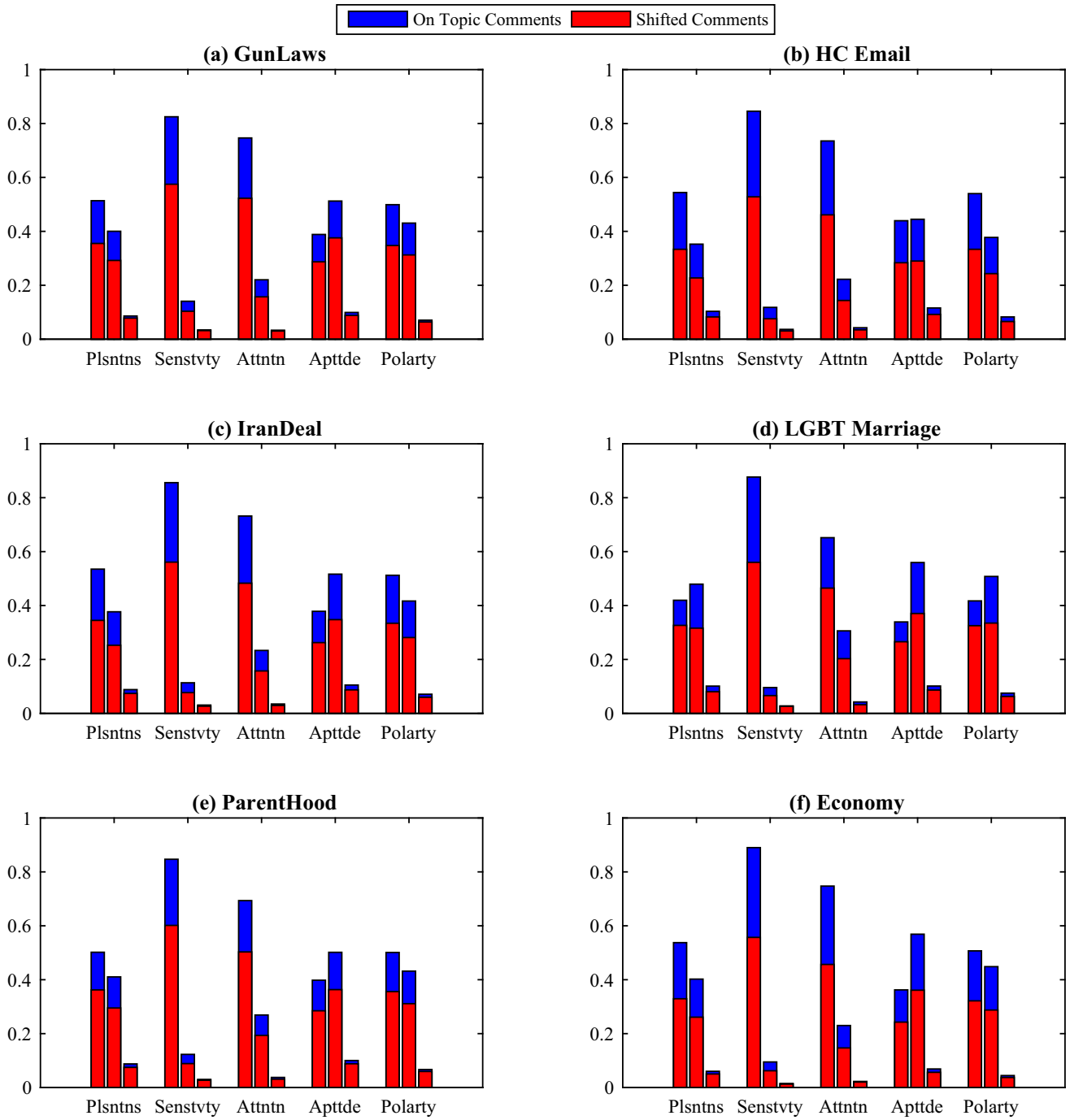


Fig. 7 Emotion bucket distribution and their shifted comment percentage

b) Discussions driven by the first-shifted comment

We take, in each root-to-leaf path, the first-shifted comment and the following comments (subtrees) in discussion trees, and measure the effect of these emotions on the topic shift and the following comments' emotions. First, we take all "first-shift" comments in a discussion tree just like we do in section IV.D, and have the following comments (without the accompanying figure, due to space restrictions).

Observation 10 90% of the time, emotion levels of the first-shifted comments fall into the first and second (i.e., low and medium) emotion levels.

That is, the first-shifted comments do not contain high-levels of emotions per dimension.

Observation 11 Within the subtrees of all first-shifted comments, on-the-topic comments decrease almost more

than 50% of the time for all topics (as compared to distribution shown in Fig. 7).

Figure 8 summarizes, per topic, the topic shift percentage changes (increase or decrease) (Y dimension) as histograms where the X dimension represents, similar to Fig. 7, the low-medium-high emotion bars of comments as a three-bar histogram (per emotion dimension). We see that, after a first-shift comment with pleasantness scores higher than .5 (i.e., *ecstasy* and *grief* emotions), all subsequent comments end up with higher percentages of shifted comments as compared to Fig. 7.

For example, in *Hillary Clinton Email Controversy* (Fig. 8b), after a first-shift comment with high *pleasantness* score, the topic shift in the comments of the following subtree increases more than 10%. This behavior of increased topic shift percentages is also observed, with very few exceptions (that occur in the *Economy* topic for our datasets), in the remaining emotion dimensions, leading us to conclude:

Observation 12 After a first-shift comment with emotion dimension scores higher than 0.5 for all dimensions, all subsequent subtrees end up with higher percentages of shifted comments as compared to Fig. 7.

That is, first-shifted comments have a very large effect on decreasing the number of on-the-topic comments that follow, regardless of the emotion dimension.

c) Predicting a comment's shift by observing its emotion

Can we predict the topic shift in a comment by just looking its emotional dimension scores? To answer this question, we randomly take 80% of the data and train a decision tree model (Rokach and Maimon 2005) and then predict the rest of the 20% of the data. We create a decision tree using *pleasantness, sensitivity, attention, aptitude,* and *polarity* scores to decide whether that comment is shifted or not. First, we create an $n \times 5$ matrix M such that each row represents a comment, and each column represents one

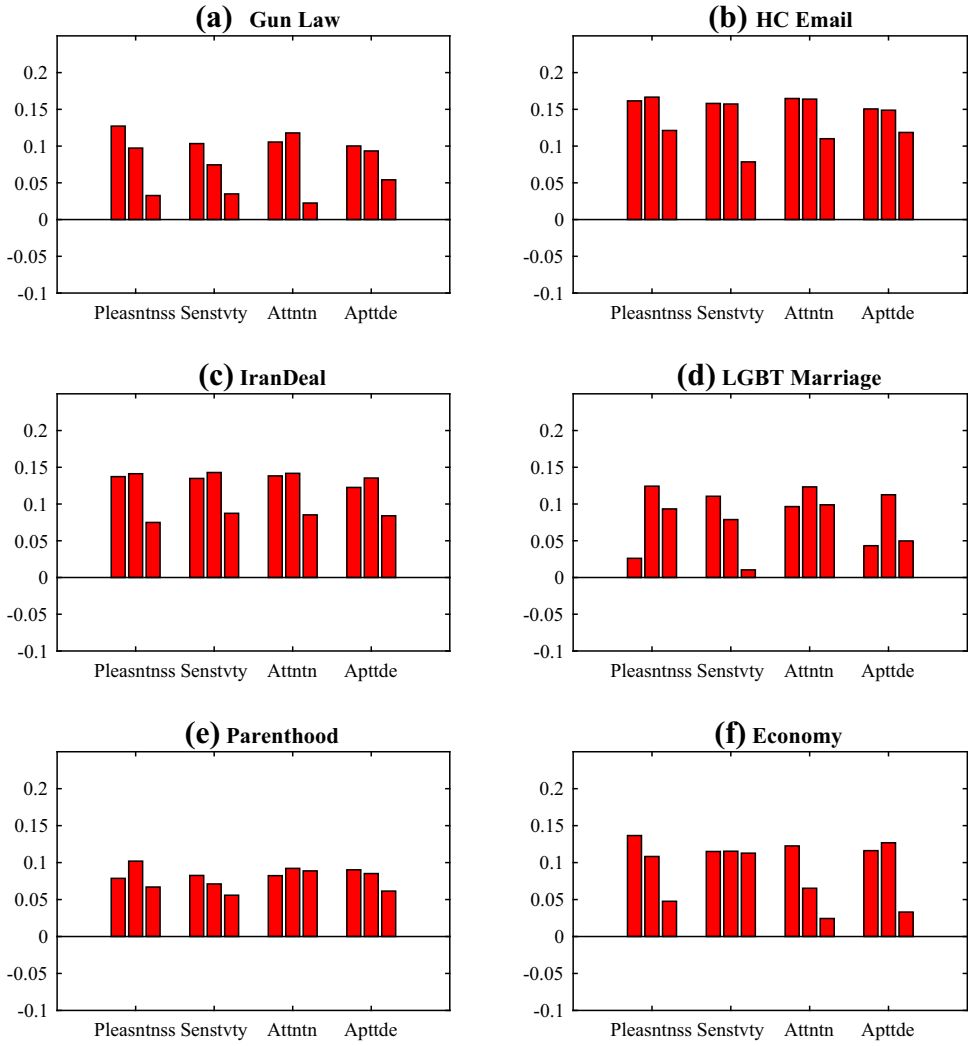


Fig. 8 Shifted comments change (*y*-axis) in response to comments that have larger pleasantness scores

of *pleasantness, sensitivity, attention, aptitude,* and *polarity* scores. Also, we know if a comment has shifted or not by looking its similarity scores to the article. We (a) set shifted comments to "0" and on-the-topic comments to "1" as "labels," (b) randomly select 80% of the data and create the decision trees by giving emotion scores and "labels," and (c) test the rest of the 20% of the data to see whether the comments are shifted or not. We use fivefold cross validation 100 times.

To improve the results, we add more fields to the decision tree. In addition to emotion scores, we add level information and create an $n \times 6$ matrix M^{+} and repeat all the steps. Then we add another row "number of words in a comment" and an $n \times 7$ matrix M^{++}, and repeat all the steps. We take the average of precisions and recalls for 6 different topics, and display them in Table 4.

Observation 13 Decision trees can predict shifted comments with precision and recall higher than 75% for most of the topics by just looking at five emotion dimensions.

In other words, emotions of comments can tell whether or not a comment has shifted with at least 75% accuracy.

Observation 14 Adding level information to emotional dimensions increases the precision and recall 1–2% for all topics.

As discussed in Sect. 4.2.1, shifted comments are located at some specific levels; so this feature increases the predictions.

Observation 15 Adding the number of words in a comment as a new feature increases the precision and recall 1–2% more. This is because short comments may have wrong emotion scores at times; so, knowing the level and the number of words of comments adds 3–5% to the accuracy.

We have also used other learning techniques with similar results. In summary, learning techniques (SVM and *k*-nearest neighbor classifiers) identify on-the-topic and shifted comments, even within the same (low, medium, or high) emotion levels.

4.2.5 Topic shifts and named and complex emotions

The emotions of each dimension is calculated by the model introduced in Sect. 3.4.2 We use this model separately within on-the-topic and shifted comment sets, per news article as to observe the differences between the two comment sets on emotional levels. We select two news articles to show how emotional levels are different within on-the-topic and shifted comments. To construct a news article emotion map, we use randomly chosen 100 on-the-topic and another 100 shifted comments from all article comments.

The first news article is about 2016 U.S. Presidential Election Debates and Primaries and article's title is *"Ramos: Trump can't handle 'uncomfortable' questions"*. Journalist Jorge Ramos was removed from one of the Republican Party candidate Donald Trump's news conferences, and the news article contains an interview with Jorge Ramos. Figures 9, 11 and 12 contain representations of the related comments' emotions. The following observation applies to all the comment sets of all articles in our dataset.

Observation 16 Heatmaps perfectly separate on-the-topic and shifted comment sets based on comments' named emotion scores.

Figure 9 depicts the emotional heatmap of the *"Ramos: Trump can't handle 'uncomfortable' questions"* news article. On-the-topic and shifted comment sets are distinctly identifiable in the heatmap. For this news article, named emotions of on-the-topic and shifted comment sets are different in every dimension. For example, for the *sensitivity* dimension, on-the-topic comments have the *annoyance* emotion (dark red colors refer to scores from 0

Table 4 Average precision and recall values for fivefold cross-validation decision trees for six different topics

	M 5 dimension scores		M^{+} 5 dimension scores, tree level		M^{++} 5 dimension scores, tree level,# of words	
	Precision	Recall	Precision	Recall	Precision	Recall
Gun laws	0.76	0.79	0.79	0.8	0.8	0.81
LGBT	0.76	0.78	0.77	0.78	0.78	0.79
HC email	0.72	0.74	0.74	0.75	0.75	0.76
Parenthood	0.76	0.77	0.77	0.78	0.79	0.8
Iran deal	0.73	0.74	0.74	0.75	0.75	0.77
Economy	0.79	0.8	0.79	0.8	0.81	0.81

We run decision trees with three different inputs: (1) M: "5 emotion dimension scores," (2) M^{+}: "5 emotion dimension scores" and "comments" levels in discussion trees" (3) M^{++}: "5 emotion dimension scores," "comments" levels in discussion trees," and "number of words in comments"

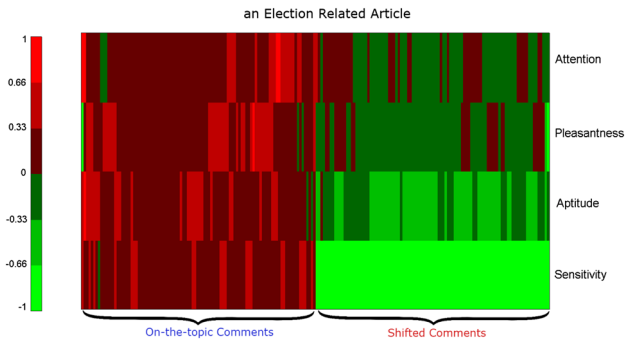


Fig. 9 Heatmap of an election related article

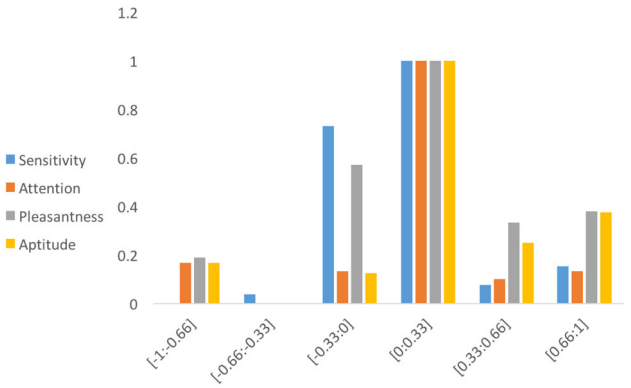


Fig. 10 Normalized emotional distribution of an election related news article

to 0.33), whereas shifted comments have the *terror* emotion (lightest green colors refer to scores of -0.66 to -1).

We also run Hourglass of emotion model on news article and count the number of emotional words on the article for each dimension. Then, we normalize these numbers to see the emotional distribution of the main article and compare with on-the-topic and off-the-topic comment sets.

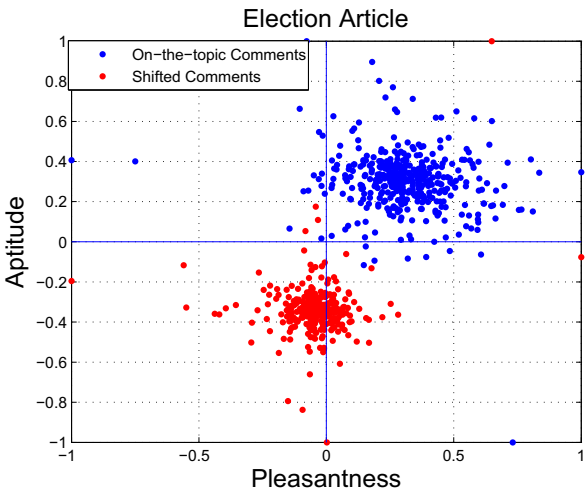
Observation 17 Emotions of news article shows more similarity to on-the-topic comment set.

Emotional distribution of the news article *"Ramos: Trump can't handle 'uncomfortable' questions"* can be seen in Fig. 10. [0 : 0.33] interval is the highest number for all dimension, which means the news article shows these emotions (*Serenity, Interest, Annoyance* and *Acceptance*) the most. Observe that on-the-topic comment set of Fig. 9 has the same intervals, which means on-the-topic comment set and news article show emotionally similar behaviors.

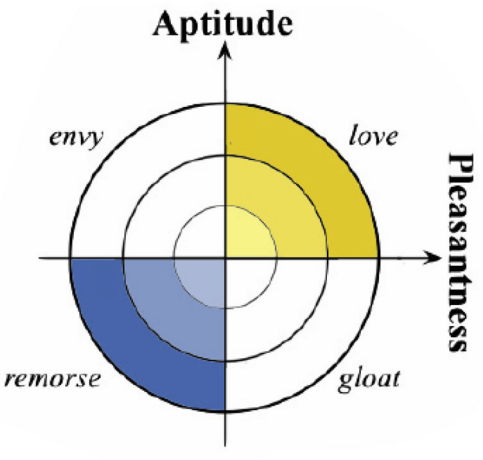
The Hourglass of emotion model (Cambria et al. 2012) defines compound emotions by combining dimensions. These pairwise combinations can be seen in part b's of Figs. 11, 12, 14 and 15.

Observation 18 There is at least one pairwise dimension combination such that on-the-topic and shifted comments cluster at different places.

Figure 11 shows the distinct clusters on the compound emotions within on-the-topic and shifted comments. We use all comment scores on two dimensions (*Pleasantness* vs. *Aptitude*) instead of the 200 comments in heatmaps. On-the-topic comments have low and moderate levels of *love* emotion, whereas shifted comments show low and moderate *remorse* and *gloat* emotions. In Fig. 12, shifted comments indicate stronger emotions (*coercion*) compared to on-the-topic comments (low or moderate *rivalry*).



(a) Comment scores, blue dots repre- sent on topic, red dots represent off topic comments



(b)

Fig. 11 Compound emotions on election related article: *Pleasantness* versus *Aptitude*

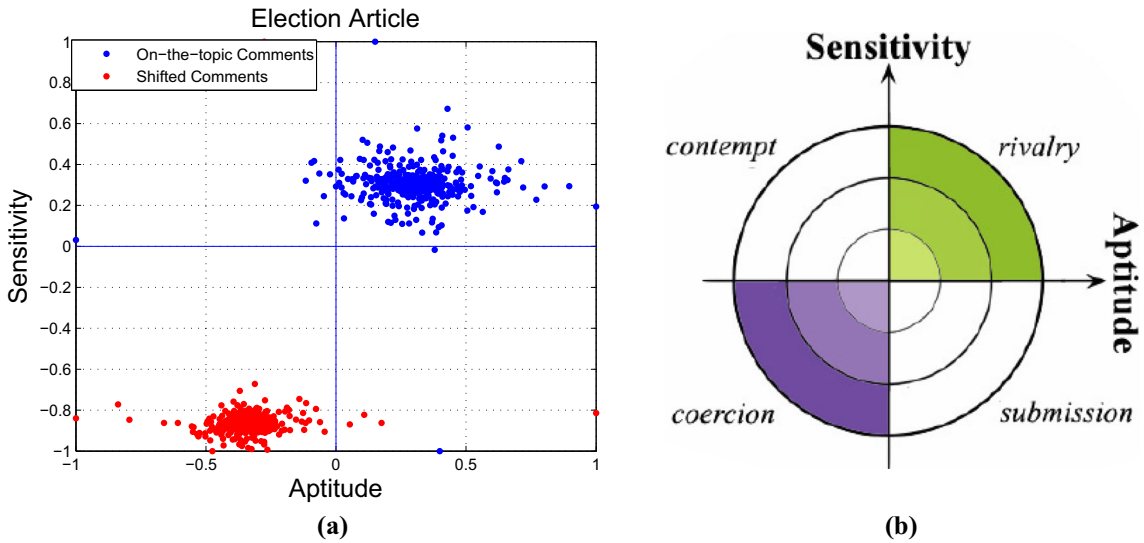


Fig. 12 Compound emotions on election related article: *Sensitivity* versus *Aptitude*

Table 5 Most frequent phrases for election news

On-the-topic phrases	Off-the-topic phrases
Deport 11 million ramos acted like like illegal alien trump going president man want see	11 million people father modern conservatism 1982 eligible amnesty entered country 1982 like Mexico central 1986 ronald reagan reagan signed sweeping

Next, we look into the most frequent phrases and words occurring in on-the-topic and shifted comments sets. Table 5 lists the most frequent phrases for each set. There are more distinct phrases in the shifted comment set because when people digress, then it can be about anything. On the other hand, for on-the-topic comment set, one would expect smaller number of unique terms. In this news article, people talked about Jorge Ramos's origin and entrance to the country, Ronald Reagan's 1986 amnesty, and immigration issues. On-the-topic comments talk more about Trump's presidency and the behavior of Ramos. And, the most frequent Sentic database words for shifted comments are "out," "off," "white," "tax," "illegal," "wrong," "border," "amnesty," whereas the most frequent Sentic database words for the on-the-topic comment set are "right," "win," "back," "ask," "country," "put," "rude," "wrong."

Observation 19 Heatmap clusters comments by their emotional scores. However, the topic shifts can be seen in one of the clusters by looking at their most frequent phrases.

The second news article we choose to view emotions of comments is related to Hillary Clinton email controversy,

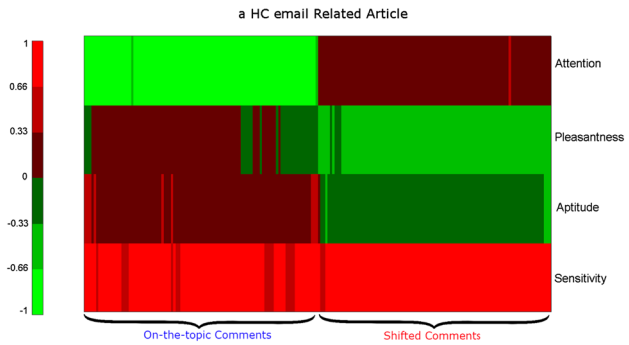


Fig. 13 Heatmap of HC email controversy news article comments

and its title is "Allies fault Hillary's response on emails." As one can see, Figs. 13, 14, and 15 are consistent with Observations 18–20.

Observation 20 Occasionally, for an emotional dimension for a news article, on-the-topic and shifted comment sets may have a named emotion with similar scores.

As an example, in Fig. 13, even though on-the-topic and shifted comment sets are perfectly separated by heatmap clustering, the *sensitivity* dimension scores remain the same for both comment sets, which is the lightest red color with a score between 0.66 and 1, corresponding to the named emotion *rage*.

Observation 21 Even though on-the-topic comment set has mostly neutral comments, for some news articles, on-the-topic comments exhibit some "strong" named emotion scores, i.e., those close to -1 or 1.

As an example, in Fig. 13, on-the-topic comments have polar named emotions of *amazement*, whereas shifted

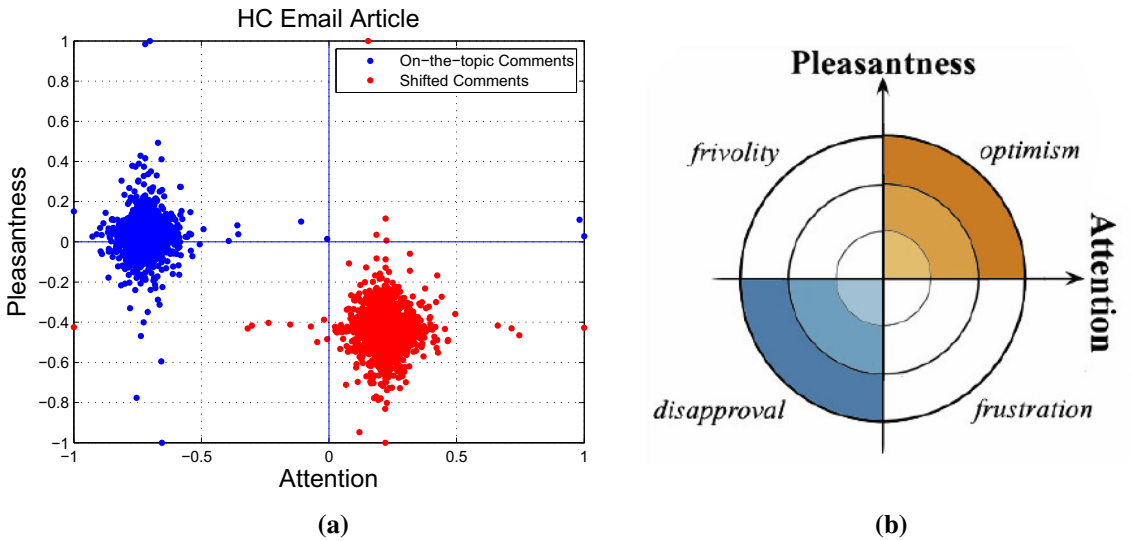


Fig. 14 Compound emotions on HC email controversy news article: *Pleasantness* versus *Attention*

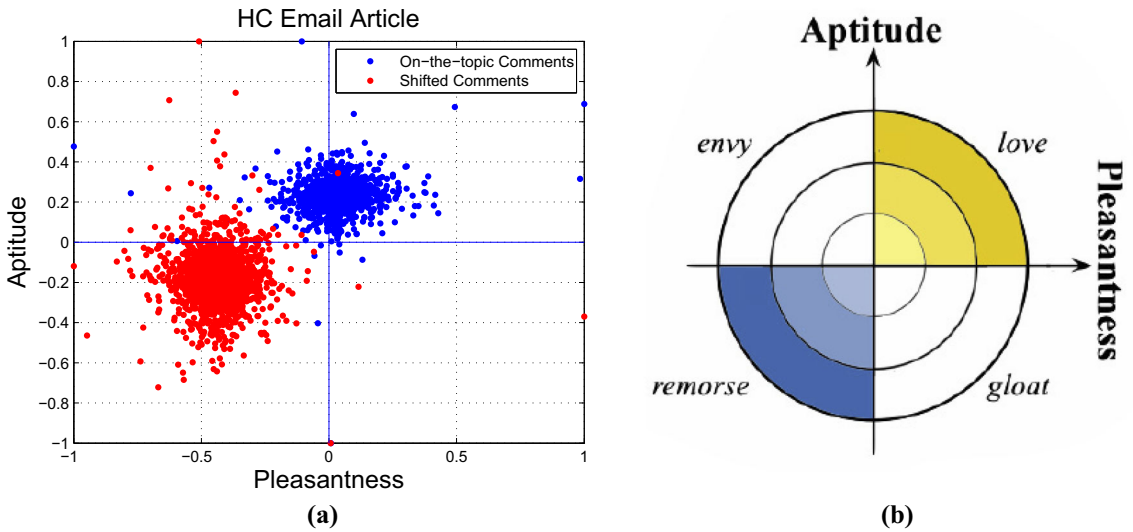


Fig. 15 Compound emotions on HC email controversy news article: *Pleasantness* versus *Aptitude*

comments have more neutral named emotions of *interest* on the dimension *attention*.

5 Conclusions

In this paper, we have studied the effects on topic shifts of comments' (1) emotion levels (of various emotion dimensions), (2) topic areas, and (3) the structure of the discussion tree, and summarized our findings that can lead to effective automated tools to identify and take actions on shifted comments, such as hiding highly emotional comments because they will be most probably shifted, showing some specific emotional levels of comments, e.g., show the neutral comments only, showing on-the-topic comments with high emotional levels, showing the statics of emotional levels of on-the-topic and shifted comment sets.

In addition, we presented and evaluated a new comment set visualization and analysis technique (based on dimension reduction via singular value decomposition) to visualize aggregated emotion levels of on-the-topic and shifted comment sets that can be used by readers to (a) pass a judgment on the properties of a comment set that hand, and/or to (b) separate all on-the-topic and shifted comments in an automated manner.

Acknowledgements This research is supported by a MEB scholarship for the first author from the government of Turkey and the US NIH Grant R01HS020919-01.

References

Alchemy (2015) Sentiment analysis API. http://www.alchemyapi.com/api/sentiment-analysis

Cambria E, Livingstone A, Hussain A (2012) The hourglass of emotions. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol 7403 LNCS, pp 144–157

Disqus: Disqus API (2015). https://disqus.com/api/docs/

FlowingData (2008) How to read and use a box-and-whisker plot. https://flowingdata.com/2008/02/15/how-to-read-and-use-a-box-and-whisker-plot/

Fodor IK (2002) A survey of dimension reduction techniques. Library 18(1):1–18

Freedman DA (1981) Bootstrapping regression models. Ann Stat 9(6):1218–1228

Gross D (2014) Online comments are being phased out. http://www.cnn.com/2014/11/21/tech/web/online-comment-sections/

Hamers L, Hemeryck Y, Herweyers G, Janssen M, Keters H, Rousseau R, Vanhoutte A (1989) Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula. Inf Process Manag 25(3):315–318

Hartigan JA, Wong MA (1979) Algorithm AS 136: a *K*-means clustering algorithm. J R Stat Soc C 28(1):100–108

Hasan M, Rundensteiner E, Agu, E (2014) EMOTEX: detecting emotions in Twitter messages. In: ASE BIGDATA/SOCIALCOM/CYBERSECURITY conference, pp 27–31

Havasi C, Speer R, Alonso JB (2007) ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. In: Proceedings of recent advances in natural languges processing 2007, pp 1–7. http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.115.2844

He D, Göker A, Harper DJ (2002) Combining evidence for automatic Web session identification. Inf Process Manag 38(5):727–742

Kleinberg JM, Kumar R, Raghavan P, Rajagopalan S, Tomkins AS (1999) The Web as a graph: measurements, models, and methods. Computer 1627(2):1–17. http://portal.acm.org/citation.cfm?id=1765751.1765753

Knights D, Mozer MC, Nicolas N (2009) Detecting topic drift with compound topic models. In: ICWSM

Liu Q, Huang H, Feng C (2013) Micro-blog post topic drift detection based on LDA model. pp 106–118

Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval, vol 1

Motulsky H (2002) The link between error bars and statistical significance. https://egret.psychol.cam.ac.uk/statistics/local_copies_of_sources_Cardinal_and_Aitken_ANOVA/errorbars.htm. Accessed 13 Nov 2015

Murguia M, Villasenor JL (2003) Estimating the effect of the similarity coefficient and the cluster algorithm on biogeographic classifications. Annales Botanici Fennici 40:415–421. https://www.jstor.org/stable/23726799?seq=1#page_scan_tab_contents

NLP S (2013) Sentiment analysis. http://nlp.stanford.edu/sentiment/

O'Hare N, Davy M, Bermingham A, Ferguson P, Sheridan PP, Gurrin C, Smeaton AF, OHare N (2009) Topic-dependent sentiment analysis of financial blogs. In: International CIKM workshop on topic-sentiment analysis for mass opinion measurement, pp 9–16

Porter M (1980) An algorithm for suffix stripping. Program 14:130–137

Rokach L, Maimon O (2005) Decision tree. Data mining and knowledge discovery handbook, pp pp 165–192. http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf

Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39(6):1161–1178. http://psycnet.apa.org/journals/psp/39/6/1161

Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Inf Process Manag 24(5):513–523

SenticNet SenticNet. http://sentic.net/downloads/

Skuta C, Bartunek P, Svozil D (2014) InCHlib—interactive cluster heatmap for web applications. J Cheminform 6(1):44–52

Smith MS, Ogilvia DM, Stone PJ, Dunphy DC, Hartman JJ (1967) The general inquirer: a computer approach to content analysis. Am Sociol Rev 32. doi:10.2307/2092070

Strapparava C, Valitutti A (2004) WordNet-affect: an affective extension of WordNet. In: Proceedings of the 4th international conference on language resources and evaluation, pp 1083–1086

Topal K, Koyuturk M, Ozsoyoglu G (2016) Emotion and area-driven topic shift analysis in social media discussions. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 510–518

Turney PD, Pantel P (2010) From frequency to meaning: vector space models of semantics. J Artif Intell Res 37:141–188

Wall M, Rechtsteiner A, Rocha L (2003) Singular value decomposition and principal component analysis. In: A practical approach to microarray data analysis, pp 91–109

Wiebe JM, Ellen R (2005) Creating subjective and objective sentence classifiers from unannotated texts. Comput Linguist Intell Text Process 3406:486–497

Wilson T, Wiebe J, Hoffman P (2005) Recognizing contextual polarity in phrase level sentiment analysis. Acl 7(5):12–21

XPO6 (209) List of English stop words. http://xpo6.com/list-of-english-stop-words/