2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Emotion -and Area- Driven Topic Shift Analysis in Social Media Discussions

Kamil Topal
Electrical Enginnering and Computer Science
Case Western Reserve University
Cleveland, OH
Email: kxt147@case.edu

Mehmet Koyuturk
Electrical Enginnering and Computer Science
Case Western Reserve University
Cleveland, OH
Email: mxk331@case.edu

Gultekin Ozsoyoglu
Electrical Enginnering and Computer Science
Case Western Reserve University
Cleveland, OH
Email:tekin@case.edu

***Abstract*—Internet-based social media platforms allow individuals to discuss/comment on the ”topic” of an article in an interactive manner. 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 study the phenomena of topic shifts in article-originated social media comments, and identify quantitatively the effects on topic shifts of comments (i) emotion levels (of various emotion dimensions), (ii) topic areas, and (iii) the structure of the discussion tree. We show that, with a better understanding of the topic shift phenomena in comments, automated systems can easily be built to personalize and cater to the comment-browsing and comment-viewing needs of different users.**

***Index Terms*—Social Media Analysis, Topic Shift, Emotion Analysis**

I. Introduction

Internet-based social media platforms allow individuals to discuss/comment on a ”topic” of their choice in an interactive manner. Usually, social media discussions are started 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.

While comments in a social media discussion start, usually, around the topic of an article, it is not uncommon for the discussion topic to shift during comments and replies of the ”discussion”, sometimes drastically and abruptly, other times slightly, away from the original topic. This has been a problem, and, in fact, due to large numbers of unrelated, inflammatory, or uncivilized comments, as well as the large numbers of comments on some popular articles, news websites and blogs have started to eliminate their comment sections [1]. This is an unfortunate action as readers of a discussion, not just the commenters, gather highly useful information by the simple act of reading these, sometimes informed, comments, and form more informed opinions themselvesa significant loss for both readers and those websites 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 (i) emotion levels (of various emotion dimensions), (ii) topic areas, and (iii) the structure of the discussion tree. We show that, with a better understanding of the topic shift 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, or (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”. Due to space restrictions, this paper only discusses items (b) and (d).

For our experimental study, 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: (i) the comments location within the discussion tree–in terms of both the level and path of the comment within the tree, (ii) 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 (iii) the topic area (e.g., *sports, politics,* or *health*). We have found that:

- In terms of a comments 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,

IEEE/ACM ASONAM 2016, August 18-21, 2016, San Francisco, CA, USA
978-1-5090-2846-7/16/$31.00 ©2016 IEEE

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 of 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 frequency, 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.

II. Related Work

A. Topic Shifts

The notion of topic shift has been studied in the field of web community discovery [2] via focused/topical crawlers [3], to identify those web pages (i.e., documents) that ”stay on the topic at hand” [4], using an information retrieval model, usually a vector space model [3], 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., OHare et al [5] apply sentiment analysis to financial blog corpus and identify topic shifts among documents in that corpus. Liu et al study topic drift on micro blog posts by using Latent Dirichlet Allocation model [6]. Knights et al detect topic drift with Compound Topic Models [7], 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 [8].

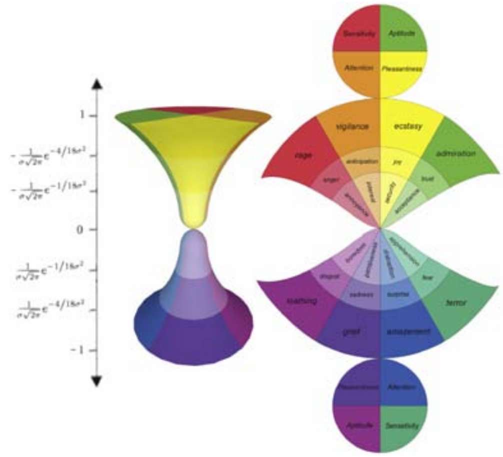

Fig. 1. HourGlass of Emotion Model [17]

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. 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 [9] 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 [10] of sentiments, and detects their subjectivity [11] 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 [12] [13]. There are two main models for representing emotions. The Circumplex model [14] characterizes emotions in two dimensions: activation and pleasure with General Inquirer [15][16] database containing lexicon of emotions, and more than 100 categories and 11,000 words.

B. Emotion Modeling

The Hourglass model [17], the most recent emotion categorization proposed by Cambria et al [17], uses a more advanced model, and has four independent, but concomitant, dimensions, namely, *pleasantness, attention, sensitivity,* and *aptitude* (Figure 1). Each dimension captures a different type of emotion: (a) pleasantness captures the users ”amusement level” with interaction modalities, (b) Attention captures inter-

TABLE I
A SNAPSHOT OF SENTICNET DATABASE

	Pleasantness	Attention	Sensitivity	Aptitude	Polarity
accept	0.972	0	0	0.894	0.622
acceptability	-0.997	0	0.878	-0.957	-0.944
acceptable	0.258	0.357	-0.088	0.351	0.293
acceptableness	-0.997	0	0.878	-0.957	-0.944
acceptance	0	0	0	0.3	0.1
acceptation	0.577	-0.77	0	0.517	0.621
accepted	0.683	-0.553	0	0.552	0.596

action contents, (c) Sensitivity captures the comfort level of the user with interaction dynamics, and (d) Aptitude captures the users confidence in interaction benefits. As also seen in Figure 1, each of the four dimensions of the Hourglass model have 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. first create the AffectNet dataset [17] by blending ConceptNet [18] and WordNet-Affect [19] 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 [20] 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 I. 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.

III. MODELING ARTICLE AND COMMENT SIMILARITIES, AND TOPIC SHIFTS

A. 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.
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.

B. 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 [21]. 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 [22]. Then, we use a modified version of the vector space model [23] as follows. Term Frequency for each document (article story and comments) is calculated using the Cornell SMART systems smoothened version [23]. 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.

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

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

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

Clearly, Eqn 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, than 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 terms weight. Note that uninformative words are already removed in the stop-word removal stage of analysis.

C. Similarity Scores

Since we have a vector representation of each comment and the article, we use the cosine similarity [3] (Eqn 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

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

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 [24], dice similarity [25], 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 [26] to create two different clusters for off-the-topic and on-the topic comment sets, and consider small centroid clusters as having shifted comments.

D. Emotion Modeling

To represent the emotional landscape of each comment, we use the SenticNet 3.0 database [20] 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 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 [17], 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 to 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

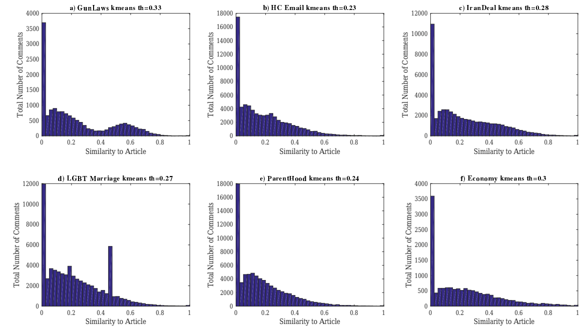

Fig. 2. Article-comment similarity distribution for six different topics

TABLE II
TOPICS AND NUMBER OF COMMENTS OF OUR DATASET

News Article topic	# of Articles	# of Comments
2016 U.S. Presidential Election Debates	11	83,604
Gun Laws Discussion	5	18,480
Immigration in U.S.	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 U.S.	13	58,388
Stocks	20	16,788
Planned Parenthood and Abortion	19	87,946
Economy	17	15,955
Total	**130**	**581,952**

combined emotion levels for Pleasantness, namely, {grief, ecstasy} (which forms the "high" emotion level), {sadness, joy} ("medium" emotion level), and {pensiveness, security} ("low" emotion level).

IV. EXPERIMENTAL EVALUATION AND RESULTS

A. Dataset

Our data sets come from news article comments on various topics (Table II). We collected $581,952$ comments from 130 news articles in the period of June 2015 to September 2015 from blog comment hosting service Disqus API [27]. We chose Disqus for the following reasons:

- It provides a service for major websites 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 data set. Each discussion tree and the level of each comment is extracted based on the reply relationship. Figure 2 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). Because of space limitations, we here focus on the results related to these six topics and their discussion trees.

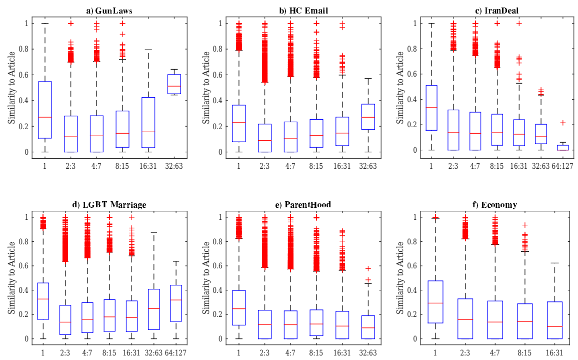

Fig. 3. The relationship between the level of an article in a conversation tree and its topical similarity to the article for six different topics. the x-axis shows the level of the comment in its respective conversation tree, where comments are binned logarithmically.

B. 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 3 (a)-(f) contain box plots [28], one for each bucket, summarizing the similarity distribution of comments in each bucket to the article. In each box plot of Figure 3 (a)-(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 [28]. Note that the X dimension scale grows exponentially since the number of comments at higher level buckets decrease 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 comments level goes deep enough in the discussion tree, i.e., after level 32, surprisingly, the average of on-the topic comments exceed 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.

C. Dissimilar versus Similar Trees

We distinguish between ”low similarity” and ”high similarity” comment sets by employing a similarity threshold defined as the maximum similarity score of the low similarity comment set.

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 Figure 4, we dissect the analysis of Figure 3 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 Figure 3 (labels omitted for readability), and box plot similarity scores of similar/dissimilar trees per bucket, as error bars [29] 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% to 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.

D. Shift Behavior Within Paths and Subtrees of Discussion Trees

Motivated by the observation that the root comment appears to set the tone, we further investigate whether there could be deeper comments that initiate a topic shift. For this purpose, we extract each path, say, path p, in a discussion tree, and locate the first shifted comment, say, comment c, in p. We then divide the number of shifted comments after c in p by the total number of comments in the subpath from c to the leaf node of p, and obtain the fraction of shifted comments in a sub-path. Figure 5-a and 5-b show, for each topic, the mean and the standard deviation of the fraction of shifted comment over respectively all subpaths and subtrees that are rooted at a shifted comment.

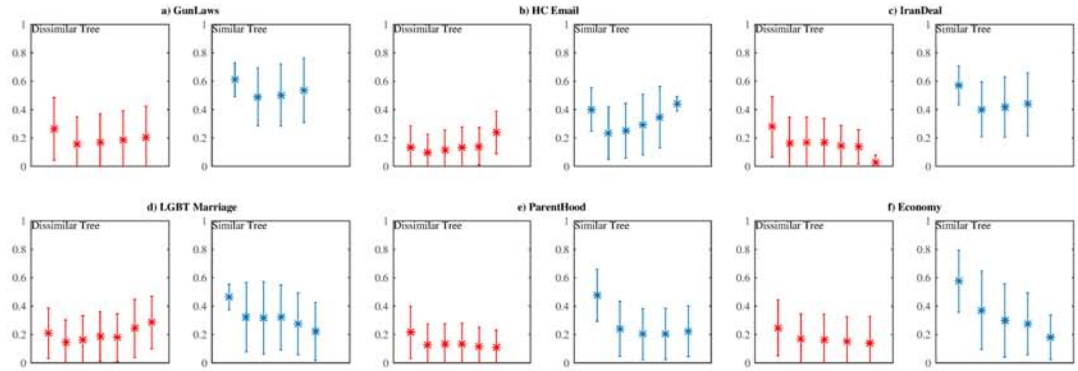

Fig. 4. The relationship between comment level and topical similarity to the article (y-axis) for Dissimilar versus Similar Trees. The x-axis shows the level of the comment in its respective decision tree, where comments are binned logarithmically.

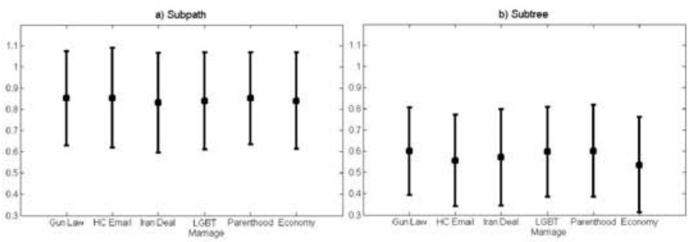

Fig. 5. Fraction of shifted comments (y-axis) that come after the first shifted comment in **a) path b) subtree**

Observation 5: In each path of the discussion tree, after the first shifted comment c, on average, 85% of the comments from c to the leaf comment shift away from the articles topic.

Observation 6: The fraction of shifted comments within subtrees with the first shifted comment as the root are 20% to 25% lower than paths on an average. The shifted comment averages in these subtrees are more than 55%.

Observation 6 is expected because if a comment is a descendant of a shifted comment, it cannot place another subtree. On the other hand, a comment can place multiple paths.

E. 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 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

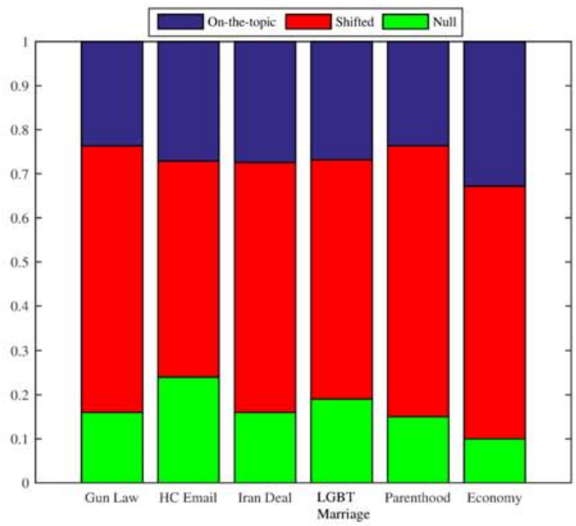

Fig. 6. On-the topic, Shifted, Null Comment percentages among different topics

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

Observation 7: After null score elimination, 64% to 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-topic comments and the lowest fraction of null comments.

F. Effect of Emotion on Topic Shifts

This section analyzes the effects of emotion on comments. For each comment, we compute the four emotional dimension

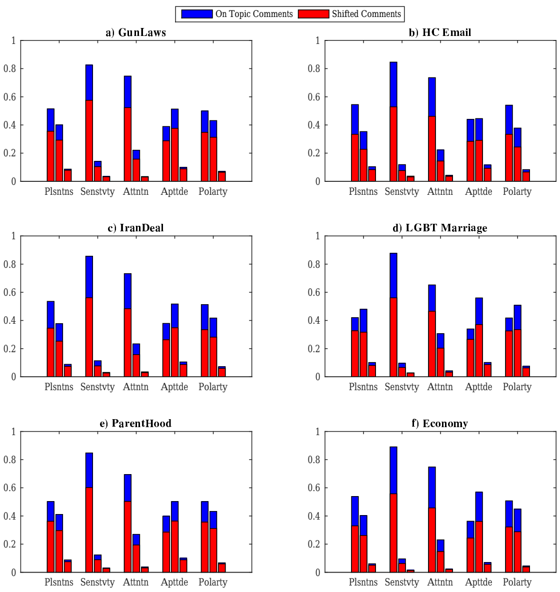

Fig. 7. Emotion bucket distribution and their shifted comment percentage

scores, namely, *pleasantness, sensitivity, attention*, and *aptitude*, as well as the similarity of the comment to the article, and compute the relevant statistics.

1) 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 8: 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 Figure 7-b, 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 Figure 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 (Figure 7-d), 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 Figure 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% to 95% of all comments in this bar have shifted. For example, among *Gun Laws* (Figure 7-a) comments, the third bar of the *sensitivity* dimension (*rage* and *terror* emotions) have only 3% of all comments; but, 93% of these comments have shifted.

Observation 9: 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 10: *Pleasantness* and *aptitude* dimensions have, for all topics, more highly emotional comments (9% to 11%) than *sensitivity* and *attention* dimensions (3% to 4%).

Observation 11: 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.

2) 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 12: 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 13: 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 Figure 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 Figure 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 Figure 7.

For example, in *Hillary Clinton Email Controversy* (Figure 8-b), after a first-shift comment with high 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 data sets), in the remaining emotion dimensions, leading us to conclude:

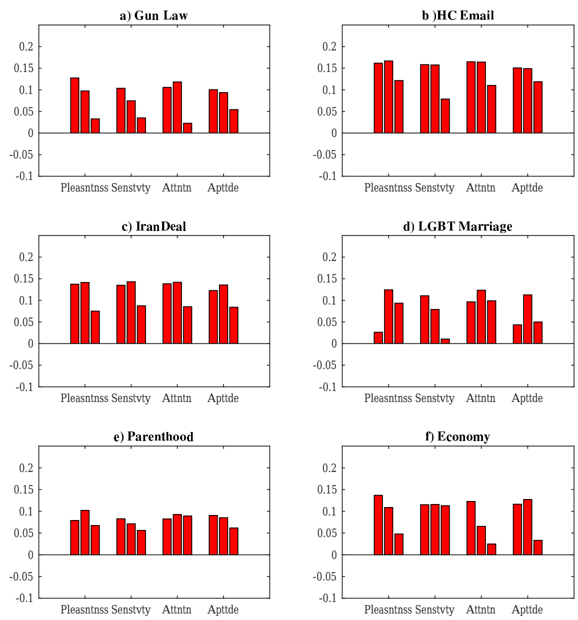

Fig. 8. Shifted comments change (y-axis) in response to comments that have larger Pleasantness scores

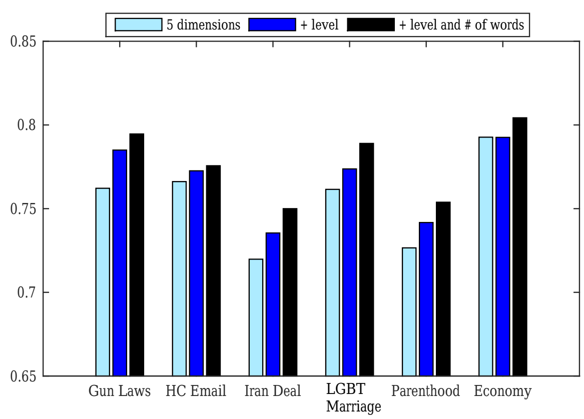

Fig. 9. Five-fold cross validation precision for decision trees

Observation 14: 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 Figure 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.

G. Predicting a Comments 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 [30], 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 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 five-fold 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 display the histograms in Figure 9.

Observation 15: Decision trees can predict shifted comments with precision 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 16: Adding level information to emotional dimensions increases the precision 1% to 2% for all topics.

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

Observation 17: Adding the number of words in a comment as a new feature increases the precision 1% to 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% to 5% to the accuracy.

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

ACKNOWLEDGMENT

This research is supported by a MEB scholarship for the first author from the goverment of Turkey and the US NIH grant R01HS020919-01. The conference travel funds for the presentation of this research by the first author are provided by the government of Turkey and the School of Graduate Studies, Case Western Reserve University.

REFERENCES

[1] D. Gross, “Online comments are being phased out,” 2014. [Online]. Available: http://www.cnn.com/2014/11/21/tech/web/online-comment-sections/

[2] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins, “The Web as a graph : measurements , models , and methods,” *Computer*, vol. 1627, no. 2, pp. 1–17, 1999. [Online]. Available: http://portal.acm.org/citation.cfm?id=1765751.1765753

[3] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, 2008, vol. 1, no. c. [Online]. Available: http://nlp.stanford.edu/IR-book/

[4] D. He, A. Goker, and D. J. Harper, “Combining evidence for automatic Web session identification,” *Information Processing and Management*, vol. 38, no. 5, pp. 727–742, 2002.

[5] N. O’Hare, M. Davy, A. Bermingham, P. Ferguson, P. P. Sheridan, C. Gurrin, A. F. Smeaton, and N. OHare, “Topic-Dependent Sentiment Analysis of Financial Blogs,” *International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement*, pp. 9–16, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1651461.1651464

[6] Q. Liu, H. Huang, and C. Feng, “Micro-blog Post Topic Drift Detection Based on LDA Model,” pp. 106–118, 2013.

[7] D. Knights, M. C. Mozer, and N. J. Nicolov Power, “Detecting Topic Drift with Compound Topic Models.”

[8] P. D. Turney and P. Pantel, “From frequency to meaning: Vector space models of semantics,” *Journal of Artificial Intelligence Research*, vol. 37, pp. 141–188, 2010.

[9] M. Hasan, E. Rundensteiner, and E. Agu, “EMOTEX: Detecting Emotions in Twitter Messages,” *ASE BIG-DATA/SOCIALCOM/CYBERSECURITY Conference*, pp. 27–31, 2014.

[10] T. Wilson, J. Wiebe, and P. Hoffman, “Recognizing contextual polarity in phrase level sentiment analysis,” *Acl*, vol. 7, no. 5, pp. 12–21, 2005.

[11] J. M. Wiebe and R. Ellen, “Creating Subjective and Objective Sentence Classifiers from Unannotated Texts,” *Computational Linguistics and Intelligent Text Processing*, vol. 3406, no. FEBRUARY 2005, pp. 486–497, 2005. [Online]. Available: http://www.springerlink.com/index/10.1007/b105772

[12] S. NLP, “Sentiment Analysis,” 2013. [Online]. Available: http://nlp.stanford.edu/sentiment/

[13] Alchemy, “Sentiment Analysis API,” 2015. [Online]. Available: http://www.alchemyapi.com/api/sentiment-analysis

[14] J. A. Russell, “A circumplex model of affect.” *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980. [Online]. Available: http://psycnet.apa.org/journals/psp/39/6/1161

[15] “General Inquirer,” 2002. [Online]. Available: http://www.wjh.harvard.edu/ inquirer/spreadsheet_guide.htm

[16] M. S. Smith, D. M. Ogilvia, P. J. Stone, D. C. Dunphy, and J. J. Hartman, “The General Inquirer: A Computer Approach to Content Analysis.” p. 859, 1967.

[17] E. Cambria, A. Livingstone, and A. Hussain, “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, 2012, pp. 144–157.

[18] C. Havasi, R. Speer, and J. B. Alonso, “ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge,” *In Proceedings of Recent Advances in Natural Languges Processing 2007*, pp. 1–7, 2007. [Online]. Available: papers3://publication/uuid/154B5FA8-A3AB-4993-9833-B1C7340A9707

[19] C. Strapparava and A. Valitutti, “WordNet-Affect: an affective extension of WordNet,” *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pp. 1083–1086, 2004. [Online]. Available: http://wndomains.fbk.eu/publications/lrec2004.pdf

[20] SenticNet, “SenticNet.” [Online]. Available: http://sentic.net/downloads/

[21] M. Porter, “An algorithm for suffix stripping,” pp. 130–137, 1980.

[22] XPO6, “List of English Stop Words,” 2009. [Online]. Available: http://xpo6.com/list-of-english-stop-words/

[23] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, 1988.

[24] L. Hamers, Y. Hemeryck, G. Herweyers, M. Janssen, H. Keters, R. Rousseau, and A. Vanhoutte, “Similarity measures in scientometric research: The Jaccard index versus Salton’s cosine formula,” *Information Processing and Management*, vol. 25, no. 3, pp. 315–318, 1989.

[25] M. Murguia and J. L. Villasenor, “Estimating the effect of the similarity coefficient and the cluster algorithm on biogeographic classifications,” *Annales Botanici Fennici*, vol. 40, no. December, pp. 415–421, 2003. [Online]. Available: ISI:000188193000004

[26] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A K-Means Clustering Algorithm,” *Journal of the Royal Statistical Society C*, vol. 28, no. 1, pp. 100–108, 1979. [Online]. Available: http://www.jstor.org/stable/10.2307/2346830?origin=crossref

[27] Disqus, “Disqus API,” 2015. [Online]. Available: https://disqus.com/api/docs/

[28] FlowingData, “How to Read and Use a Box-and-Whisker Plot,” 2008. [Online]. Available: https://flowingdata.com/2008/02/15/how-to-read-and-use-a-box-and-whisker-plot/

[29] H. Motulsky, “The link between error bars and statistical significance.” [Online]. Available: https://egret.psychol.cam.ac.uk/statistics/local_copies_of_sources_ Cardinal_and_Aitken_ANOVA/errorbars.htm

[30] L. Rokach and O. Maimon, “Decision Tree,” *Data Mining and Knowledge Discovery Handbook*, pp. pp 165–192, 2005. [Online]. Available: http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf