User Behavior Under the Influence of Groups in Social Media

Lionel Martin
HCI, I&C, EPFL

***Abstract*—Social Media are at the heart of our communications and are among the most visited places on the web. Their user-generated content allows the gathering of immensely many information that cannot be processed entirely by the users. Thus it is of interest to understand how these users are deciding what pieces of information they are trusting and what are the reasons that influenced them.**
In this thesis proposal, we survey three papers presenting different methods to detect influence in social media. We show that the content produced is an important feature which hasn't been extensively studied yet and deserves a closer attention. Toward this goal, we present our first results of influence detection via emotion recognition and propose several extensions.

***Index Terms*—influence detection, emotion recognition, social media, group decision, thesis proposal, Human Computer Interactions**

I. Introduction

Nowadays on Internet, everyone has numerous possibilities to express his opinion, share content with others and access almost any information in a couple of seconds. Social Media defines all the web platforms that have been developed to allow the creation of user-generated content in any domain from product reviews to personal life sharing.

Proposal submitted to committee: August 16th, 2013; Candidacy exam date: August 23rd, 2013; Candidacy exam committee: Prof. Karl Aberer (exam president), Dr. Pearl Pu (thesis director), Prof. Boi Faltings (exam co-examiner).

This research plan has been approved:

Date: ____________________

Doctoral candidate: ____________________
(L. Martin) (signature)

Thesis director: ____________________
(P. Pu) (signature)

Doct. prog. director: ____________________
(B. Falsafi) (signature)

While the most successful ones, such as Facebook[1], Twitter[2] or Amazon[3] welcome millions of visitors daily, the content produced by these millions of users is obviously enormous. Theoretically this would permit to obtain a very high quantity of information in various fields if only we were capable to process all this information at the same speed that it is created. Unfortunately, internet users are only able to read a small portion of what they can access by lack of time and they have to make their mind based on this subset.
Moreover, it is a sure fact that all the comments left on social media are not equally informative and that each of the authors has his own way of expressing. For this reason, we believe that some content is more probable to lead a reader's opinion in a direction or another. Consider the case where you need to decide in which hotel to stay during your holidays next summer. Among all the customers giving their feedback about their own stay there, only a few of them will help you to decide where to go. The same thing could apply to people to elect, products to purchase, or actions to take in general.

On the one hand, the detection of influencers and propagation of thoughts has been widely studied as a structural problem. Ultimately, the structure of the network helped to determine which were the most influential users based on their number of acquaintances and their connectivity [1], [2]. On the other hand, the works focusing on the content exchanged between users are rare even though their results give good hope that it contains features that may improve the detection of influence. Knowing the content exchanged between any two people gives obviously more information than just the fact that they are connected with each other, and although more information means a finer analysis, we are facing challenges such as deciding what are the important data and how to aggregate these among the whole set of users.
In this proposal, we will first present three different models that aim at understanding the influence observed during decisions made in social media at different stages. First, we will be talking about the effect of the similarity between the users and its impact on votes and elections; then we will see how to detect helpful reviews, those with the overall most informative content about a particular topic; and finally we will dive into the details of the formation of comments

[1]www.facebook.com
[2]www.twitter.com
[3]www.amazon.com

and the detection of emotions in user-generated texts using text mining techniques. Then, in section V, we will present our current work about influence detection using emotion mining in reviews where we will show that influential reviews have the particularity to contain more than twice as many emotion signals than the other reviews in average. Finally, this proposal is concluded with future work that will hopefully help group decisions via a better understanding of user behavior in social media and the detection of influence.

II. Effects of User Similarity in Social Media

Recent studies tried to emphasize the important features of user's decision concerning the evaluation of another user or the content he produced in social media. Anderson et al. [3] presented two measures that combine to improve the prediction of this evaluation. These two measures, which we will introduce below, are descriptors of the influence induced by an author towards a reader.

A. Contributions

The basic idea of this paper comes from the fact that people usually have a higher degree of affinity with people that share common interests. Thus, one might either be kinder in the evaluation of someone else's content or simply agree more generally with the ideas that this other person presents in the comment he has written.

Toward this end, they decided to associate to every user of a social media, binary vectors tracking all the actions that one has done. They focused on different types of actions and constructed for example a vector for the topics of interest (topics vector), one for the users evaluated so far (evaluation vector) and also one for the threads in which they posted (edit vector). Then they defined the similarity between any two users as follows.

They used two binary vectors at a time with as many components as there are possible actions for the two different users (respectively **e** for the evaluator and **t** for the target). They set the entries to 1 if and only if the user has done the action associated with the entry before the decision that they are studying was made. Finally, they took the cosine of the two vectors and called it the *similarity*:

$$s(\mathbf{e}, \mathbf{t}) = \frac{\mathbf{e} \cdot \mathbf{t}}{\|\mathbf{e}\| \, \|\mathbf{t}\|}$$

Their evaluation exploits three different datasets. The first one concerns elections to the admin status on Wikipedia[4] with a total of almost 120k votes distributed among 3'422 elections. Here, users are asked to decide whether the candidates, which are frequent contributors of Wikipedia, should obtain the administrator status on the community website or not, giving them rights to manage the encyclopedia. The remaining two datasets are composed of 1.1M questions and 3.2M answers from Stack Overflow[5], a Q&A website about programming, possessing 7.5M votes whose 93.4% are positive and 1.5M reviews of Epinions[6], a product review website, posted by 132k users which were rated 13.6M times (with 78% of 5-stars, the positive ones).

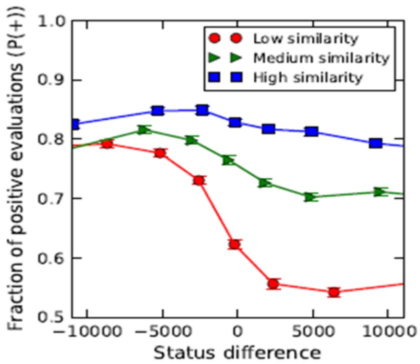

Fig. 1. Combined effect of similarity and status on the positivity of evaluators for the Wikipedia dataset.

The effect of similarity on the positivity of the voters has been proved for all the three datasets with a particular impact on Wikipedia. Indeed, a relative gain of 60% has been established between the users with the smallest edit similarity (similarity between the edit vectors) and those with the highest. Stack Overflow and Epinions both showed a positive result as well but with less impact. A relative gain of 20% has been observed for the study of tag similarity between the 25th percentile and the 80th percentile in these datasets.

Moreover, a second basic characteristic influences user evaluations in many situations: the hierarchical status [4], [5]. In social media, this can be modeled by the seniority and the number of actions that each user has been doing. In their three datasets, the authors decided to count the number of articles, questions or reviews written. Rather than the status of the evaluator σ_e or that of the target σ_t, this is the difference between them which has the most impact in the decision: $\Delta = \sigma_e - \sigma_t$. Indeed, the authors have observed that for a given Δ, the probability of voting positively is almost constant with respect to the target status (and thus to the evaluator status as well) but for two different values of Δ and a fixed target status, the probability changes (reaching higher values for lower Δ). Then, a combined impact of both similarity and status was studied. They discovered that with high similarity between the evaluator and the target, the status difference had less impact but that in general positive status difference (where evaluator has a higher status that the target) was coming along with a smaller probability of positive evaluation. In their Wikipedia dataset, they showed that the decrease of positivity between the smallest and the largest status difference is 10 times smaller in the highly similar setting with an observed loss of 2%, than in the least similar ones where the decrease reaches 20% as depicted in fig. 1.

[4]www.wikipedia.org
[5]www.stackoverflow.com
[6]www.epinions.com

B. Evaluation

With these observations they tried to predict the outcome of group voting. They selected a dataset of more than 3'000 elections from Wikipedia with an average of 44 votes per election. They only used the identity of the first five voters to predict whether the candidate obtains or not his administrator status. First they constructed the edit vectors (the binary vectors of articles that the users edited previously) and then computed similarity and status difference with the user to be elected for each voter. This setting is called ballot-blind prediction since they didn't use the result of the votes for those five voters but only their identity and past actions in order to predict.
They derived a simple model to predict the probability that the user voted positively in this election based on the average positivity of the evaluator (the percentage of times he voted positively in the past) and the average deviation that is perceived over all the users in the same quadrant than him. Quadrants are defined by the status difference on the x-axis (negative on the left, positive on the right) and the similarity on the y-axis (below the 50^{th} percentile at bottom, above on top).
For instance, assume that the candidate and the evaluator have the following history of actions: $v_e = (1, 1, 1, 0, 0), v_t = (1, 0, 1, 0, 0)$. Then their similarity is sim($\mathbf{e}, \mathbf{t}$) $= \frac{2}{6} = 0.33$ (65^{th} percentile) and their status difference is $\Delta = 3 - 2 = 1$. Moreover, this evaluator has an average positivity of 75% since he voted 4 times in the past and only once negatively. Finally, the other voters in the quadrant of positive Δ and high similarity are in average voting 8% more positively than the average of all votes so far thus Pr($\mathbf{e}$ votes for $\mathbf{t}$) = 0.83.
They chose to aggregate these five results and to predict based on a threshold function. In other terms, the outcome of the vote was decided positive if the average of the five results is higher than τ (whose value is decided by machine learning) and otherwise it was decided negative. With this simple statistical analysis, they achieved an accuracy of 75%, which is only 10% above what can be achieved by the gold standard where one could look at the decision of the first five voters. This also corresponds to a relative gain of 69% compared to an a priori decision.

C. Discussion

In this paper, the authors showed that the similarity between two users has an impact on the way one rates or votes for the other. This representation of influence has the particularity to be relative to the person voting and the one being evaluated. This means that for two different evaluators, the influence might be completely different for the same content to vote on. However, it also means that we expect the same evaluator to cast the same vote for two different reviews of the same author. Indeed, this characterization never takes into account the content which is published and needs to be evaluated. Nonetheless, even if two users are really similar on a particular topic, they might not have similar interest for the topic about which one just wrote something. For this reason, we believe that it is necessary to take also the content into account for most datasets in order to obtain better predictions.
Finally, in their prediction process they decided to treat the group decision as an aggregation of single users' decisions, neglecting the fact that the users may influence each other during the decision process both via the comments they leave and the fact that previous votes are available. We believe that this is an interesting choice that to simplify the group decision problem to the aggregation of personal opinions as a starting approach. However, it might introduce a bias since part of the influence might be due to the decisions of the previous voters that is visible.
In our work, we will later try to mix relation-based characteristics (such as similarity and status) with the features of the content exchanged. However, our first approach considers solely the content exchanged.

III. Helpful or Unhelpful: a Linear Approach for Ranking Product Reviews

With their work on the prediction of helpfulness, Zhang and Tran [6] proposed a solution to the estimation of influence which is focusing on the content created by the users.
Compared to the previous works based on the structure of the network, they established a methodology that aggregates the opinion of all the users of the social media and focus on the impact of a document (for example, a review).

A. Contributions

Their willingness to evaluate the content of the reviews posted by the users made them consider a Bag-of-Words approach to identify the words that are the most relevant to the characterization of influence. They considered each word independently of the others and constructed a dictionary with every word that appears at least once in any of the reviews of their dataset. Then they associated with these words the measure of influence observed over all the reviews. This influence is defined by the interactions of the different readers of the comment. Two buttons allow the users to say whether the review seems helpful or unhelpful to them. This vote is not mandatory and only a small proportion of the readers take the time to give some feedback. To measure it, the authors simply counted the proportion of helpful votes among the total number of votes for each document and compare the difference between the documents containing the word and those not containing it.

1) Definition of Gain: They defined the gain of a term using the concepts of Entropy and Mutual Information of Shannon which are basic measures of Information Theory. They extended their definition to obtain the gain of a word t as

$$G(t) = H(S) - H(S|t) - H(S|\bar{t}),$$

where $H(S)$ is the entropy of the set in which the documents belong. $S = \{s_1, s_2\}$ is a random variable whose universe contains two outcomes "being unhelpful" or "being helpful" respectively. A document is said to be truly helpful if at least 90% of the votes are helpful and truly unhelpful if less than 10% of the votes are. Thus the threshold to decide if a document is helpful or not is set at 60% which represents the mean of the helpfulness proportion for their dataset. Then, the gain can be rewritten as:

$$G(t) = -\sum_{i=1}^{m} \Pr(s_i)\log(\Pr(s_i)) + \Pr(t)\sum_{i=1}^{m} \Pr(s_i|t)\log(\Pr(s_i|t)) + \Pr(\bar{t})\sum_{i=1}^{m} \Pr(s_i|\bar{t})\log(\Pr(s_i|\bar{t}))$$

This gain symbolizes the uncertainty of classification of term t into a class or the other. Thus, the higher the gain, the more certainly one can affirm that it belongs to a class.
Moreover, with the information gain, they were able to define the term gain as the participation of a term to the helpfulness sentiment the readers have about the review. This second measure uses the proportion of positive votes in average among all the reviews containing the term t, denoted $\bar{h}(D|t)$ and the conditional probability $\Pr(S|t)$ as follows:

$$\text{gain}(t_j) = \begin{cases} G(t_j)\bar{h}(D|t_j) & \text{if } \Pr(s_1|t_j) < \Pr(s_2|t_j) \\ -G(t_j)(1-\bar{h}(D|t_j)) & \text{otherwise} \end{cases}$$

2) Helpfulness Score: The authors defined helpfulness score next to measure the predicted helpfulness of a document (a review) based on the different words that it contains. The idea is that each word has a positive or negative term gain as presented above and thus by addition of the gains of each term in the document, we can obtain the document helpfulness estimator:

$$\text{score}(d_i) = \sum_{j=1}^{W} \text{gain}(t_j)\mathbb{1}_{\{t_j \in d_i\}}$$

Finally, they decided to normalize the results on a scale [0, 1] to compare more easily with the actual helpful proportion.

B. Evaluation

The authors proposed an experimental evaluation of their method for the classification of reviews about digital cameras on Amazon. The dataset contains 1'486 reviews evaluated by at least five consumers. The evaluation can be either positive (helpful) or negative (unhelpful). They perform 10-fold cross validation on a dataset of 600 randomly chosen reviews. They decided that the threshold to define a review as Helpful during the training phase was 60% of positive votes and that the learning algorithm defines the threshold for the classification during the testing phase as the helpful score of the $|V^+|^{\text{th}}$ (the cardinality of the set of helpful reviews in the training set) sorted review.
Their evaluation presents good precision, recall and F-measure results (all close to 75%). The authors presented their new method in comparison to Naive Bayes, Support Vector Machine (i.e, SVM) and Decision Tree Classification, and declared that it outperforms two of the three and works as well as the third one. Moreover, they computed the log-likelihood of the predicted and original helpfulness probabilities and showed qualitative results about the goodness-of-fit of their model.

C. Discussion

Intuitively, the different measures presented are not selected in the way that seems to match our perception of the datasets. Consider a simple example with 1'010 reviews out of which 10 only are helpful. Moreover, consider that a term t is present in those 10 reviews plus 11 more of the unhelpful category. Proportionally, this term is in all the helpful reviews and in 1.1% of the unhelpful ones thus it is probably a term which conveys the sentiment of helpfulness. However, since the majority of the documents containing this term are unhelpful, it will be a descriptor of unhelpfulness and his term gain will be negative with respect to the measures defined in section III-A. Moreover, they assessed that two of the three methods compared in the evaluation are outperformed by their new technique but if we look closely to the table in their paper, it seems that Naive Bayes is performing as well as their technique and that SVM is outperforming it clearly. Finally, the original and predicted probabilities computed to study the goodness-of-fit of the model are not correct distributions. For example, you can consider a document voted by 100 users with $h(d_i) = 0.86$. Their original probability for a voter to decide the documents d_i as helpful becomes $2.6e^{-18}$ which is not reflecting the actual behavior of the vote (with 86% of helpful votes). Given this inconsistency, the goodness-of-fit of their model cannot be studied in details.
Conclusively, this work is very inspiring because it proposes an approach taking the content of the reviews into account and focusing on the influence of reviews more than reviewers. Moreover, such technique also outputs a good lexicon of influential terms that can be reused outside the context of the study, either with Bag-of-Words or Part of Speech approach. Nonetheless, it would be interesting to reproduce this method with a few modifications such as the consideration of n-grams instead of unigrams only and by changing the measures that are taking as features of prediction (term gain, helpfulness score, etc.). I believe that we could obtain a better accuracy by considering the border cases, as the one highlighted above, more rigorously.

IV. Crowdsourcing a Word-Emotion Association Lexicon

In this section, we present a project of annotation of emotions to create a large lexicon of term-emotion associations [7]. The creation of such lexicon has numerous applications such as creating emotion-aware systems that manage customer relations or respond appropriately; identifying and characterizing the emotions used by people for different purposes including influencing others using a particular vocabulary; and many more.

A. Contributions

Mohammad and Turney [7] propose to associate to a very large quantity of common words the emotion that they represent or more generally that people feel when they encounter these words in texts. Similar lexicons associating words simply with positive or negative emotions exist for some years and have been used in different works, e.g. to interpolate the rating based on the review [8]. The main difference in the work that we are presenting is that they achieve a more precise characterization of the author's opinion which goes beyond the positivity and is also concerned with basic emotions. As of now, there also are a few limited-coverage lexicons such as WordNet Affect Lexicon [9] or General Inquirer [10] and those will be used to compare with the lexicon the authors are constructing. The reason why there is no perfect lexicon can be attributed to several things: first, it is costly and time consuming to ask expert annotators to classify each word. Moreover, several well-known models of emotion classification are used in research and thus each of them would need its own lexicon.

Regarding this last point, the authors chose to focus on 8 emotions which form one of the basis commonly used and with which any complex emotion can be associated. This model is named after its inventor Plutchik [11]. The principal concept is that emotions are associated by pairs of opposed meaning: Joy and Sadness, Trust and Disgust, Fear and Anger, and Surprise and Anticipation. He also introduced the idea that these families could be expressed at different intensity defining three layers (e.g., Annoyance, Anger and Rage) and that they could be mixed together (e.g, Joy and Trust makes Love). Nonetheless, even though Plutchik pairs his emotions, some of the pairs are composed with two negative emotions (e.g., Fear and Anger) and thus his model is not balanced regarding positivity.

In their paper, Mohammad and Turney proposed an alternative to the costly method usually applied to construct a lexicon. They decided to use online crowdsourcing and ask lots of non-expert users about the emotions they felt after reading words. Ultimately, since this lexicon's goal is to interpret the feelings of writers and readers or such words, these people are a good population to ask to. The only issue is that non-experts might have different opinions about the same word or think about different meanings of a word. Quality control and fine annotation have been added in the process to circumvent the previous issues without reducing the population that can answer. In the small questionnaire that they ask to the users, the first question checks that the annotator knows the meaning of the word by proposing 4 different possibilities among which one is a synonym of the meaning of interest of the word. Moreover, they rejected the answers of users that were mistaken too often (above a third of failures).

B. Evaluation

During the test, they asked the users to tell how much the word was positive or negative to them, and how much it conveyed each of the 8 basic emotions of Plutchik's model. Then, they compared the results of their experiment with the lexicons already existing. The General Inquirer lexicon associates words with their positivity only while the WordNet Affect Lexicon associates words to Ekman's emotion model which is a subset of Plutchik containing 6 basic emotions: Anger, Disgust, Fear, Joy, Sadness and Surprise. The comparison between the EmoLex that the authors are presenting and the two others are really encouraging. They showed that negative terms of the General Inquirer are marked negative at 83% and positive terms are marked positive at 82%. Words of the WordNet Affect Lexicon are marked correctly in the majority of cases as well: disgust and sadness terms are classified correctly at 94%, anger and joy around 80%.

This evaluation highlighted difficulties to classify surprise terms of the WordNet Affect Lexicon. The users decided that only 66% of the terms associated with Surprise in the WordNet Affect Lexicon are terms that actually inspire surprise, half of them being positive and a quarter negative (the remaining being neutral). Given the results for the other categories, such a difference probably means one of the following things: Surprise is an emotion which is hard to classify or the surprise terms of the WordNet Affect Lexicon themselves were wrongly classified.

C. Discussion

Since a consequent part of content analysis will require to understand the vocabulary used in the reviews on social media, we need to be comfortable with techniques of lexicon formation. This, for example, is useful if we decide to detect more than what is available at the moment by expanding available resources or creating our own.

This work provides a good starting point for the recognition of emotions in texts. However, they chose to use Plutchik's model of emotion categories with 4 pairs of families. We believe that this model doesn't have the best repartition of emotions and is driven towards negative feelings. In what follows, and especially in our current work, we focus on finer-grained evaluation of emotion in text with 20 emotion categories because it has better chances to represent the correlation with influence.

V. RESEARCH PROPOSAL

A. Current Work

The three papers that were presented above are approaching the characterization of influence in various manners. We believe that focusing on content-based techniques is the best way to improve our understanding of influence. For this reason, we developed a method using emotion retrieval in comments partially based on the works described above. We chose to apply a different emotion classification model than Mohammad and Turney [7] that we considered to be inappropriate for two reasons: it contains more negative than positive emotions and these 8 families are rather large and were not matching our attempt to characterize entirely the emotions in the reviews in details. Thus, we selected Scherer's model instead [12], proposing a solution with 20 different categories, 10 of which being positive emotions and 10 being negative ones. These emotions are arranged on a wheel, named Geneva Emotion Wheel, separated in 4 quadrants describing the valence (positive or negative emotion) and control (low-control or high-control on the emotion) of each of them. This classification globally feats best our objectives of characterization of the influence via the recognition of emotions.

We focused our approach on the content and the sense of

the words encountered to characterize the emotions that the author tried to convey in his writing. To detect the emotions contained in the reviews, we designed an algorithm in several steps inspired from the standard Bag-of-Words approach. Given a review as input, the text is first split into a list of words. These words are then matched one by one to a dictionary of emotional words, named GALC [12]. This lexicon associates words to emotions of the Geneva Emotion Wheel. Words that are associated with an emotion are indexed and we are searching for negation as well as intensity markers in the neighborhood of these words. For example the sentence "I was absolutely not happy that I pay for a suite that wasn't available." contains the word "happy" that is associated with an emotion (Happiness in this case) but also a word that triggers negation ("not") and one for intensity ("absolutely"). Once all the emotion-related words have been retrieved from the comment and the negation and intensity checks have been completed, the output is simply a measure of the expression of each of the 20 emotions of the Geneva Emotion Wheel. This measure is summing the number of occurrences of each emotion individually and applies modifications to the counter for the occurrences linked to a negation word and/or an intensity word. Next to characterize the influence of reviewers based on the emotions present in the comments we had to determine how we chose the influent comments. We decided to consider helpfulness score as a measure of influence in our work because it is a good representation of what consumers declare of importance even though it is time dependent. Such votes reflect the influence that a text can have for the user voting. Moreover, even if a lot of different comments might express the same thing, only a small subset is selected by the crowd as the most relevant ones. We thus labeled the comments with the highest helpful score as influential and put a threshold for the selection.
We hypothesized that the comments which are influential will contain fewer emotions than those which are not influential because people prefer to rely on objective facts and would trust more easily comments that are informative than users telling the personal problems they had with the product and for which they might be angry.

However, the results are quite surprising since we showed that on average, influential reviewers are using more emotion-associated words than the others. In our dataset, we discovered an increase of +114% (more than twice as many markers). We also presented the differences that some parameters could imply such as the choice of the lexicon (for this part, we used a lexicon close to what Zhang obtained in [6] but constructed for positivity analysis) or the dataset. We showed for these cases that the difference was really negligible compared to our main dataset. Moreover, we split the reviews in two categories: the positive ones (with a good rating) and the negative ones (with a bad rating) and attested that negative emotions are more often present in negative comments and conversely. Regarding influence we noticed that positive comments are even more prone to the use of emotions than the negative ones.
Our approach is the first fine-grained emotion retrieval that we know of and also a novel approach of influence detection solely based on the content exchanged by the users. As presented before, this method seems to be effective and predict partially at least the source of influence. For this reason, we will propose future work going in this direction in the next section as thesis proposal.

B. Future Work

We believe that content-based techniques haven't been studied extensively so far and that they might allow a better understanding of user behavior as presented with our current work. For this reason, we present a list of possibilities for our future works that use what have been surveyed in this thesis proposal as well as our latest results to develop a few ideas. Moreover, since it is often assumed that group decision is mostly based on the aggregation of each user's decision, we hope that these results might also help the study of group decisions in a second time.

1) Predicting Influence: First, with the results that we have obtained so far, we would like to apply prediction techniques to the emotion retrieval work we have been doing recently. Since more emotional comments are associated with influential ones, we hope to be able to improve significantly the prediction of influence in the near future. We will need to test several algorithms and compare with different well-known method to get the most of our findings.
We planned also for some time to match words directly to influence without passing by emotions before computing predictions; in the direction of what was done in [6]. The main differences will appear in the selection of expressions to evaluate because we plan to do it for n-grams instead of unigrams and we think that we will also apply different measures and different thresholds.

2) Different Emotion Mining Approaches: Bag-of-Words is the simplest way to consider the content of a text because it treats each word separately. However, the relations between words in a sentence can improve our understanding of the meaning of a text. For example, we already tried to interpret negation and amplification of emotions with Bag-of-Words but we faced a couple of limitations like the size of the neighborhood in which the negation applies or the double negation that might appear in some sentences. Alternatively, we could consider working on Part of Speech where words in a sentence are grouped into categories such as nouns, verbs, adjectives, etc. It would allow differentiating "like" as a verb from the preposition or the conjunction (the former being an emotion whereas the others are not) for example. Another possibility is to consider adjective-noun pairs which have been studied for sentiment analysis and recommendation [13].

3) Constructing different lexicons: In our current work, we saw that lexicons have a large impact on the results and we would like to obtain a lexicon which is more specific to our dataset to be able to compare the results. Namely, we claim that having a lexicon which is domain-specific to reviews and perhaps even to establishment reviews (including restaurants for example) or hotel reviews could improve the emotionality perceived in the comments and the characterization of the influent comments for our current dataset (reviews on Trip Advisor) even though GALC is already a good lexicon for our purpose. One way to do so is to extend our current lexicon (GALC) with a technique close to Zhang's work [6] once again to find out new words that often appear in reviews conveying a particular emotion. The idea would be to associate a set of emotions to each review and to look for words that often appear in review conveying an emotion but rarely in the other reviews.

4) Characteristics of Influence: Besides emotions, there are potentially others characteristics of influence in social media's comments. Without any evidence, it might make sense that humorous people are more susceptible to be followed because they present the fact in a subtler manner for example. Thus we could decide to work on other characteristics of the comments and then mix the different characteristics found so far to improve further our estimate. This would also rejoin the discussion about relation-based techniques like the work of Anderson et al. [3] presented in section II.

5) Group Decision: Later, we hope to be able to help group decisions with our understanding of influence in social media. For now, researchers are principally assuming that group decisions consist of the aggregation of each individual decision but the interactions during the decision between the voters might have an impact on the final decision. Understanding who the influencers in a group are and what influenced the mind of the voters (potentially outside the group of voters [2]) might improve group decision and reduce the contestations about the outcome. Toward this goal, we need to evaluate different datasets with a group structure and correlate their decisions with their content they read in social media.

REFERENCES

[1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference*, 2003.

[2] S. Myers, C. Zhu, and J. Leskovec, "Information diffusion and external influence in networks," in *Proceedings of the 18th ACM SIGKDD international conference*, 2012.

[3] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Effects of user similarity in social media," in *Fifth ACM International Conference on Web Search and Data Mining*, 2012.

[4] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010.

[5] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *Proceedings of the 13th international conference on World Wide Web*, 2004.

[6] R. Zhang and T. Tran, "Helpful or unhelpful: A linear approach for ranking product," *Journal of Electronic Commerce Research*, 2010.

[7] S. M. Mohammad and P. D. Turney, "Crowdsourcing a wordemotion association lexicon," in *Computational Intelligence*, 2012.

[8] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, 2002.

[9] C. Strapparava and A. Valitutti, "WordNet-Affect: an Affective Extension of WordNet," in *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 2004.

[10] P. J. Stone, D. C. Dunphy, and M. S. Smith, *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press, 1966.

[11] R. Plutchik, "A general psychoevolutionary theory of emotion," *Emotion: Theory, research, and experience*, 1980.

[12] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, 2005.

[13] K. Yatani, M. Novati, A. Trusty, and K. Truong, "Review spotlight: a user interface for summarizing user-generated reviews using adjective-noun word pairs," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.