

Characterizing Influence Factors Affecting Emotion Diffusion in Facebook

Kurt Junshean P. Espinosa, Alma Mae J. Bernales

Abstract— Previous studies have claimed that emotion can be transferred from one person to another via social media. As emotion is crucial to one's ability to adjust to the challenges of daily life and affect our relationship with others, this paper aims to characterize the factors influencing the diffusion of emotion in Facebook by using the Independent-Cascade diffusion model describing the diffusion of emotion in Facebook user's status messages and by using Multi-Regression analysis in analyzing the results of the diffusion model.

The final results show that there is a significant relationship between the similarity rate of users and the diffusion rate in an emotion diffusion process. Thus, a regression equation with a similarity variable is formulated. The resulting regression equation is $D_r = 0.061 + 0.518S$ where D_r is the diffusion rate and S is the similarity rate of a diffusion process. This equation may also be useful in investigating interpersonal communication in terms of emotion diffusion between lovers, parents and children, and other relationships we can find in Facebook. Moreover, results of this study can help viral marketers find clues on how to spread positive sentiments from their customers effectively on the internet.

Index Terms— Emotion Diffusion, Social Networks, Regression Analysis, Text-Mining, Facebook.

I. INTRODUCTION

EMOTION is a subjective experience of a strong feeling usually engaged to a specific object and usually complemented by physiological and behavioral changes. Emotional states can be transferred directly from one individual to another through mimicry and "emotional contagion". This claim is also supported by the study of Hill, Rand, Nowak and Christakis [6], in which they introduced a SIR disease model in order provide evidence of emotion diffusion. It was found out that emotions spread like an infectious disease in a large social network. Moreover, Fowler & Christakis [5] formulated a longitudinal statistical model, which shows the spread of happiness in a social network and it is not just a tendency for people to associate with similar individuals. These studies, however, focus on the real world social network, wherein emotions spread during face-to-face interaction.

Manuscript received July 23, 2014; revised August 22, 2014.

K. J. P. Espinosa is with University of the Philippines Cebu, Lahug, Cebu City, 6000 Philippines. (phone: +63 925 501 7354; e-mail: kpespinosa@up.edu.ph).

A. M. J. Bernales was a student of University of the Philippines Cebu, Lahug, Cebu City, 6000 Philippines. She is now with NPAX Cebu Corp. Banilad, Mandaue City, Cebu, 6014 Philippines (phone: +63 912 817 4998; e-mail: maebernales@gmail.com).

As the World Wide Web becomes a significant part of the modern society because of the internet, researchers are also becoming interested in the emotional contagion in the online world. Sites like Facebook, Twitter, MySpace and the like have been the subject of emotion diffusion studies in online social network. Diffusion models, sentiment analysis methods and statistical analysis help researchers find evidences of emotional contagion in online social network [1][10].

In this paper, the researcher will investigate the process of emotion diffusion in Facebook and will characterize the influencing factors that trigger such. If emotion diffusion in Facebook can be shown, this will add to the growing body of evidence suggesting that emotional contagion is present in online social network. Moreover, characterizing the influencing factors that affect emotion diffusion can help people become aware of the effects of these factors. In turn, knowing these factors will help people become aware of what to post in the social networking sites so that they would not negatively affect any of their audience. The factors that were investigated in this study are similarity, interactivity, and connectivity.

II. RELATED LITERATURE

Characterizing the influence factors that affect the diffusion of emotion in an online social network delves in three study areas: sentiment analysis, online text mining, and diffusion in social networks. Sentiment analysis literature provides techniques in classifying user-generated online contents in terms of emotion. Moreover, related studies on online text mining show processes of getting datasets for social network analysis, and the like. On the other hand, existing studies of information diffusion and diffusion in online social network, in general, gives some basic models, which can be used in determining how emotion spreads through the social network.

A. Sentiment Analysis

Sentiment analysis uses Natural Language Processing to analyse word use, order and combinations, in order to classify sentiments which are commonly categorized as positive, negative, or neutral[8].

It can be deduced from Table 1 that existing supervised and unsupervised sentiment analysis methods are more accurate for domain-specific datasets, with 90.48% accuracy rates (specific to movie reviews datasets). However, Table 2 shows that existing methods of both supervised and unsupervised sentiment analysis do not achieve satisfactory accuracy rates for domain independent datasets.

TABLE I. SENTIMENT CLASSIFIER ACCURACY RATE FOR DOMAIN DEPENDENT DATASETS

Algorithm	Domain	Accuracy Rate
Turney	Movie	66%
	Automobile	80%
	Bank	84%
SVM (Na et. al)	Director Review	75.54%
	Cast Review	78.74%
	Movie Review	90.48%

TABLE II. SENTIMENT CLASSIFIER ACCURACY RATE FOR DOMAIN INDEPENDENT DATASETS

Algorithm	Accuracy Rate
Average Perceptron (Llaguno)	46%
Senti-Strength	66%
Rocchio(Llaguno)	73%
Naïve Bayes (Pang and Lee)	78.7%
SVM(Yang, Lin, Chen)	82.59%
SVM (Pang and Lee)	82.9%
CRF (Yang, Lin, Chen)	84.07%

Thus, existing literature shows that automated methods in emotion classification are not that accurate. Moreover, considering the locale of this study which is Philippines, varied dialects would pose a very big challenge in analysing sentiments automatically, as it implies about having to deal with words that may have the same spelling but are entirely different in meaning.

Since the purpose of this study is to characterize the influence factors of emotion diffusion, it is a must that classifying contents according to the right emotion category must be accurate. To the best of the researcher’s knowledge, manual labeling is still the most accurate approach of sentiment analysis, though this method requires huge amounts of time and effort in determining the sentiments.

B. Online Text Mining

With the increasing number of social networking sites, there has been a surge of user generated content online. A study revealed that online social networks contain some sort of subjectivity, which proves that these online social networks are emotionally rich environments [15].

Because of this, online social networking sites have become the subject of many research, pointing out the scientific trials and significance of sentiment analysis, information diffusion and emotion diffusion in different fields of study [4][10][11]. The most common focus of these studies are Twitter and Facebook, since they are used by diverse types of people in showing opinions about different topics, and they also contain a huge number of posts which grow significantly everyday [11].

Yassine & Hajj [15] provided a framework for emotion mining from text in online social networks. Since the subject of this study is on online social network, the

following based on available literature, are various challenges that must be overcome.

- **Opinion mining.** Contents such as spam and marketing content have no emotional bent, which can obscure emotion diffusion. Moreover, the dialects used by Filipino social media users, given the fact that some of their words have the same spelling but have entirely different meaning, are bound to create some complications [12].
- **Data sources.** Since datasets of this study are user-generated, getting data from these users poses some privacy issues, which means that the researcher should get permission from the users to get the desired data for the study [1].
- **Emotion Classification.** In most online social networks, there are no direct, consistent and standard indicators of emotion. However, if there are any, these indicators are not constantly applied by social media users when they generate contents. Thus, appropriate techniques must be applied to classify contents according to their right emotion label.
- **Lack of ground truth.** Emotions do not only spread in an online social network. The truth of whether a user’s emotion is influenced by either an online or an offline social interaction is a limitation of the current study.

C. Models of Diffusion in Online social network

Studies have been presented to illustrate the spread of emotion in online social networks [1][4][9][10].

A study conducted by Cole[1] about emotional contagion in an online social network employed an information diffusion approach in detecting emotions of blogs in LiveJournal. Cole [1] modified the General Threshold model to better represent the propagation of emotion. He trained the model on a network of bloggers and emotion-labeled blogs in order to learn the influence probabilities of the network. The idea is, if a network of bloggers is proven to have influenced each other in the past, there is a higher probability that a blogger’s emotion will be affected in the future by the same network. The model is then used to run a diffusion simulation that predicts the emotions of a set of blogs. Through this, conclusions about the existence of propagation in the network based on the accuracy of the model’s predictions can be drawn, given that they are based on a propagation model. The results show that there is indeed an existing emotional contagion in an online social network, such as LiveJournal. Kramer [10] investigated on the diffusion of emotion via Facebook by conducting a sentiment analysis to examine the emotion of Facebook users according to their Facebook status messages. He analysed the words used in each Facebook status message, with the help of the Linguistic Inquiry Word Count (LIWC) software available on the web. It rates words according to whether they have positive or negative meanings. However, this software is only limited to Arabic, Chinese, Dutch, English, French, German, Italian, Portuguese, Russian, Serbian, Spanish, and Turkish languages.

By employing a multi-day lagged regression design, Kramer [10] found out that a user's status update that has more positive than negative words resulted into a 7% increase of positive sentiment strength and a 1% decrease of negative sentiments to friends' status updates. His study proved that there is indeed a diffusion of emotion in Facebook even three (3) days after the posting of a status update.

However, one of the study's limitations is that its methodology could not ensure that the Facebook users analyzed really saw the Facebook status updates which has been thought to have influenced them. A user's Facebook status update may be similar to that of his/her friend even though he/she had not really seen that said status update. Thus, Kramer [10] suggested to conduct more research about whether a Facebook status update can really influence other Facebook user's emotions.

Another study conducted [9] focused on discovering the emotion influence patterns in Twitter. They investigated the transition of tweets to their reply tweets and examined the influence of a user's tweet (on a certain topic) to the sentiment of the conversation partner's sentiment. In order to analyse the transition and influence of emotions, they used the aspect and sentiment unification model (ASUM) to automatically discover topics and sentiments. ASUM discovers topics that are closely coupled with sentiment in an unsupervised way. Moreover, they proposed a way of finding interesting conversations by looking at the overall sentiment patterns of the conversers.

Furthermore, they concluded that Twitter users in general tend to have similar sentiments with their conversational partners. However, their results also showed that other users tend to feel good even when the conversational partners do not. This model is only appropriate for a pairwise interaction.

Fan, et. al. [4] studied Weibo, a Twitter-like service which has attracted more than 500 million users in less than four years, and discovered that anger could spread more quickly and broadly in an online social network. They used Naïve Baye's method in classifying emotions of tweets. In case of diffusion of emotion, they used an emotion correlation (Pearson correlation) metric to quantify the strength of sentiment influence between connected users in an undirected graph $G(V, E, T)$, in which V is the set of users, E represents the set of interactive links among V , and T is the minimum number of interactions on each link. E is determined by the sum of retweets and mentions between two ends in a specified time period.

Conversely, research in information diffusion is also a prevalent subject in literature. Several models of diffusion are proposed by different authors to illustrate the process of diffusion in social networks.

Three of the most common models used in representing diffusion are Linear Threshold Model (LTM), Independent Cascade Model (ICM) and Epidemic Model.

According to [1], LTM uses a network represented by a weighted graph where each node v chooses a threshold function which will be compared with the activation function. Moreover, the activation function is evaluated at t discrete time steps and v becomes active when the activation function exceeds the threshold function.

Similarly, ICM uses a weighted graph in determining diffusion in discrete time steps. It suggests that a node v has the probability to influence his neighbor w given a link from v to w as the basis of probability function [1], regardless of the neighbors of w . Moreover, node v does not have the chance to influence others anymore after the contagious stage[1][13].

Dargatz [2] described Epidemic Model as based on the cycle of infectious disease in a host. It proposes that under a population of interest, individuals are classified as being susceptible, infected or recovered (S-I-R), depending on their contact with infected nodes. The transition between these classes is defined as:

$$S + I \xrightarrow{\alpha} 2I, I \xrightarrow{\beta} R \quad (1)$$

where α is the rate of infection when an individual comes in contact with an infectious individual, whereby making him/her susceptible. β is the rate of recovery between two interacting infected and susceptible individual.

Lastly, Lahiri & Cebrian [14] used the concept of Independent Cascade model formulating the genetic algorithm as a general diffusion model for social networks. They proposed that each contact in a dynamic network can elicit an activation probability in a particular time step defined as the set of users $V = \{v_1, v_2, \dots, v_n\}$ interacting over a period of T discrete time steps.

D. Factors of Diffusion in a Social Network

There are several factors that can affect diffusion in social network. However, studies delve more on the information diffusion because of its significance in viral marketing, and the like. Moreover, diffusion of news is also investigated to help government and non-government organizations (NGOs) in spreading or suppressing the diffusion of relevant news, such as emergencies, incoming natural disasters and others.

According to Katona Zubcsek & Sarvary [7], network characteristics (i.e., network density and network size) and personal influences affect diffusion in online social networks, more specifically in information diffusion. They cited that two related individuals connected to the same third parties transmit information better because of the stronger relationships that influence them. They were able to "model the adoption decision of individuals as a binary choice affected by three factors: 1) local network structure formed by already adopted neighbors; 2) average characteristics of adopted neighbors (influencers); and, 3) the characteristics of the potential adopters".

Zhu, et. al. [16], analysed users' retweeting behavior by studying the factors that may affect their decision, including "context influences, network influences, and time decaying factors," and by using logistic regression to formulate the problem into a retweeting probability conditioned on the incoming tweet and targeted users. Through this model, they were able to investigate the spread of messages, since disaster messages do not surpass other communication in the Twitter medium.

On emotion diffusion specifically, Doherty, et. al. [3], in their study on primitive emotional contagion, proved that an individual's gender can also affect the diffusion rate of an emotion. In their recent study conducted in Weibo, Fan,

Zhao, Chen & Xu[4] found out that anger spreads faster than joy. Thus, these suggest that both content expressed and gender of the one who posted the content can affect the diffusion rate of the emotion expressed.

III. METHODOLOGY

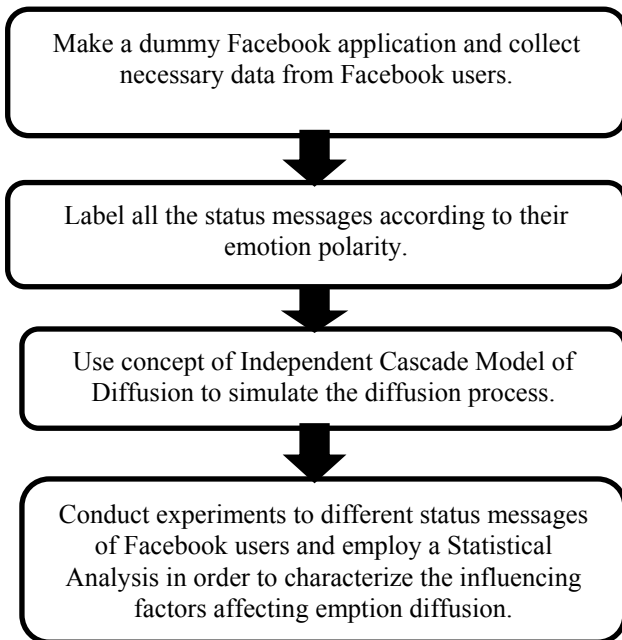


Fig. 1. Operational Framework of the Study

Since the objective of this study is to characterize the factors that affect emotion diffusion in Facebook, the researcher gathered data from Facebook by asking users to login the Facebook application which collects necessary data for the study.

There are two types of users involved in this study. The first one is the **direct user**. It is the user who logged in the Facebook application. The other one is the active user. It's the user who interacts with the direct user from January, 2013 – December, 2013.

The following data were gathered from each user who logged in the application:

- Basic information (work, education, language, religion, gender, current location, hometown, relationship status)
- Status Message
- Posts (photos, links, videos)
- Interaction on posts and status messages (people who interacted on the post, including the date of interaction)
- Liked pages
- Groups List

After gathering all data from users, status messages of users posted from October, 2013 – December, 2013 were manually classified according to their emotion polarity. Moreover, the researcher made a java program to aid the classification with a more user-friendly graphical user interface.

Status messages of direct users which are posted from October, 2013 – December, 2013 were used as sources of emotion diffusion for the experiments.

The researcher used Independent Cascade Model of diffusion to illustrate the process of emotion diffusion from the direct users to the active users.

The following are the rules considered by the researcher to determine if emotion is diffused from A to B:

- B should have commented or have liked the status message of A.
- B should post a status message similar to the emotion of A's status message.
- B's status message should be at most 3 days after liking or commenting on A's status message.

The process of emotion diffusion was evaluated according to the diffusion rate:

$$D_r = \frac{d_d}{d_i} \quad (2)$$

where d_d is the number of friends who were diffused while d_i is the number of unique friends who liked or commented on the status messages.

Moreover, the similarity, interactivity and connectivity between the users in the diffusion process will be measured according to the following equation:

Let A be the source user, B be the user who interacted with A's status message and n be the number of people who commented on the status message.

$$\text{a. similarity: } S = \frac{A \cap B}{A \cup B} \quad (3)$$

where $A \cap B$ is the number of similar basic information (friends, interests, likes, location, language, work, school) of A and B while $A \cup B$ is the number of unique basic information (friends, interests, likes, language, religion, relationship status, location, work and education) of A and B.

$$\text{b. interactivity: } I = \frac{\sum \frac{B_a}{A_b}}{n} \quad (4)$$

where B_a is the total number of interaction (i.e., number of likes and comments) of A to B and A_b is the total number of interaction of B to A.

$$\text{c. connectivity: } C = \frac{\sum \frac{B_a}{\sum A_i}}{n} \quad (5)$$

where $\sum A_i$ is the total number of interactions of A to all his/her friends.

The aforementioned equations are based on the assumptions of a co-researcher whose research involves the use of connectivity, similarity and interactivity in modeling the emotion diffusion among Facebook users.

In order to characterize the factors that affect emotion diffusion in Facebook, the researcher used the multiple regression analysis in analyzing the result of the experiments. However, the assumed factors must undergo

first with the correlational tests in order to establish that there is a significant relationship between the three factors and the diffusion rate.

The following mathematical formula to express the relationship between similarity/interactivity/connectivity and emotion diffusion:

$$\sigma_r = b_0 + b_1S + b_2I + b_3C \quad (6)$$

where:

σ_r = average diffusion rate (D_r)

S = similarity score

I = interactivity score

C = connectivity score

b_0, b_1, b_2, b_3 = numerical constants which must be determined from observed data.

Positive values of b_1, b_2 , and b_3 will show that increasing score of similarity, interactivity and connectivity will result to higher diffusion rate.

The researcher conducted the Multiple Regression Analysis with the aid of SPSS Statistics application.

IV. FINAL RESULTS AND DISCUSSIONS

This chapter presents the data gathered by the researcher from Facebook which would help characterize the influencing factors that affect emotion diffusion in Facebook. Moreover, analysis from the results of calculations of diffusion rate, similarity, interactivity, and connectivity will be presented in this chapter to come up with a regression equation which characterizes the influencing factors affecting emotion diffusion.

A. Emotion Polarity

The Facebook application made by the researcher gathered about 375 source status messages from 10 Facebook users. 186 of these source status messages are positive while 189 source messages are negative. These source status messages are the status messages of users from October, 2013 – January, 2013.

B. Emotion Diffusion

For each the source status messages, similarity, interactivity and connectivity between source user (user who posted on each source status message) and active friends (user who liked or commented on the source status message) were calculated. The data showed that there are 210 or 56% of all the source status messages have 0% diffusion rate. On the other hand, 8 or only 2.13% of all the source status messages have 100% percent diffusion. Considering the aforementioned data, characterizing the influencing factors of emotion diffusion in Facebook is still reasonable since almost half of the source status messages have evidences of emotion diffusion.

C. Characterizing the Factors of Emotion Diffusion

In characterizing the influencing factors affecting emotion diffusion in Facebook, it is important to see first if the assumed factors have significant correlation with diffusion rate. The factors which will have significant

correlation with the diffusion rate will be considered to be a predictor of diffusion rate, thus, these factors will be included in the regression analysis.

TABLE III. CORRELATION BETWEEN DIFFUSION RATE AND SIMILARITY

Pearson Correlation (r)	0.213
Sig. (2-tailed)	0.000
N	375

TABLE IV. CORRELATION BETWEEN DIFFUSION RATE AND INTERACTIVITY

Pearson Correlation (r)	-0.100
Sig. (2-tailed)	0.052
N	375

TABLE V. CORRELATION BETWEEN DIFFUSION RATE AND CONNECTIVITY

Pearson Correlation (r)	-0.026
Sig. (2-tailed)	0.616
N	375

TABLE VI. COEFFICIENTS OF REGRESSION EQUATION

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	0.061	0.016		3.897	0.000
Similarity	0.518	0.123	0.213	4.205	0.000

Table III shows the correlation between diffusion rate and similarity. The r-value 0.213 says that there is a weak positive correlation between the two variables. Moreover, the sig-value 0 tells us that the relationship is significant for 375 number of diffusion processes. This implies that there is a statistical evidence that the higher the similarity rate of a source status message (i.e., the average similarity between the source user and all users who interacted on the source status message), the diffusion rate is also higher.

Table IV shows the correlation between diffusion rate and interactivity. The r-value -0.100 signifies that there is no relationship between diffusion rate and interactivity. However, the sig-value 0.052 shows that the relationship is insignificant. This suggests that the finding could have happened by chance. Therefore, for this study, interactivity cannot be considered as factor in emotion diffusion.

Conversely, Table V shows the correlation between diffusion rate and connectivity. The r-value -0.026 means that there is no relationship between the two variables. However, the sig-value 0.616 implies that the finding is insignificant since the former is greater than 0.05. Therefore, connectivity cannot be used as one of the factor of diffusion rate in our regression equation.

Since only similarity has significant relationship with diffusion rate, we make a regression analysis using the former to predict the latter.

Table VI shows the resulting coefficients for the regression equation. It can be construed from Table XII that the value of b_0 and b_1 are 0.061 and 0.518 respectively. Thus, this forms the regression equation:

$$D_r = 0.061 + 0.518S \quad (7)$$

V. CONCLUSIONS AND RECOMMENDATIONS

Base on the final results of the study, it can be concluded that similarity among Facebook users significantly affect the diffusion rate of emotion in Facebook. Thus, the prediction equation $D_r = 0.061 + 0.518S$ can be used to predict the rate of emotion diffusion in Facebook.

Moreover, the result of this study can be used to investigate the interpersonal communication between lovers, parents and children, and other relationships we can find in Facebook.

In addition, the researcher recommends further investigation about the effects of similarity, interactivity, and connectivity on emotion diffusion in an online and real world. With a large dataset, the study does not prove that interactivity and connectivity can affect the emotion diffusion rate in Facebook. However, the number of source users may affect the result of this study since there are only 10 source users with different number of source status messages. Thus, the researcher suggests more number of varying source users to see if the result of this study holds true.

References

- [1] Cole, W. D. (2011). An information diffusion approach for detecting emotional contagion in online social networks. Retrieved from ASU Libraries.
- [2] Dargatz, C. (2007). A diffusion approximation for an epidemic model. (No. 517). Discussion paper. Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München.
- [3] Doherty, W., Orimoto, L., Singelis, T., Hatfield, E. & Hebb, J. (1995). Emotional contagion: Gender and occupational differences. *Psychology of Women Quarterly*, 19, pp. 355-371.
- [4] Fan, R., Zhao, J., Chen, Y., Xu, K. (2013). Anger is more influential than joy: Sentiment Correlation in Weibo. Retrieved from <http://arxiv.org/pdf/1309.2402>
- [5] Fowler, J. H. & Christakis, N. A. 2008 Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *Br. Med. J.* 337, a2338. Doi:10.1136/bmj.a2338
- [6] Hill, A. L., Rand, D. G., Nowak, M. A., Christakis, N. A. (2010). Emotions as infectious diseases in a large social network: the SISa model. Doi: 10.1098/rspb.2010.1217
- [7] Katona, Z., Zubcsek, P. P. & Sarvary, M. (2010). Network effects and personal influences: The diffusion of an online social network. Retrieved from <http://www.cs.bme.hu/~zskatona/pdf/diff.pdf>
- [8] Kennedy, H. (2012). Perspectives on sentiment analysis. *Journal Of Broadcasting & Electronic Media*, 56(4), 435-450. Doi:10.1080/08838151.2012.732141
- [9] Kim, S., Bak, J., Oh, A. (2012). Discovering emotion influence patterns in online social network conversations. *SIGWEB Newsl.* DOI=10.1145/2362631.2362634
- [10] Kramer, A. (2012). The spread of emotion via facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 767-770.
- [11] Kumar, A. & Sebastian, T. M. (2012). Sentiment analysis on twitter. *IJCSI International Journal of Computer Science Issue 9(3)*, 372-378. Retrieved from <http://ijcsi.org/articles/Sentiment-analysis-on-twitter.php>
- [12] Llaguno, K. (2012). *Facebook status sentiment analysis*. University of the Philippines Cebu, Gorordo Ave., Lahug, Cebu City.
- [13] Mertsalov, K., Magdon-Ismail, M., Goldberg, M. (2009). Models of communication dynamics for simulation of information diffusion. In *2009 Advances in Social Network Analysis and Mining*. Retrieved from <http://ieeexplore.ieee.org/>
- [14] Lahiri, M. & Cebrian, M. (2010) The genetic algorithm as a general diffusion model for social networks. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI 2010)*. Retrieved from <http://ieeexplore.ieee.org/>
- [15] Yassine, M. & Hajj, H. (2010). A framework for emotion mining from text in online social networks. *IEEE International Conference on Data Mining Workshops*. 13(13), pp. 1136-1142.
- [16] Zhu, J., Xiong F., Piao, D., Liu, Y., Zhang, Y. Statistically modeling the effectiveness of disaster information in social media. In *2011 IEEE Global Humanitarian Technology Conference*. Retrieved from <http://ieeexplore.ieee.org/>