Emotional Community Detection and Emotion Prediction in Social Network

Vikanksh Nath(MT19AI024)
nath.2@iitj.ac.in
Megha Kumari(MT19AI001)
kumari.9@iitj.ac.in
Neelu Verma(MP19AI002)
verma.14@iitj.ac.in
Suman Karan(MT19CS019)
karan.1@iitj.ac.in
Course Instructor: Dr. Suman Kundu

Department of Computer Science & Engineering,
IIT Jodhpur, India

Abstract

Community detection has been an important research problem in social network analysis. User's emotional behaviour also play a significant role in the detection of community in a social network, but emotional state of an user has been neglected in previous reported works by the researchers for a long time. Recently, researchers have started exploring this emotional dimension too, for the detection of a community in the network. So, The goal of this project will be reviewing the relevant reported literature on the topic "Emotional Behaviour for Community Detection in Social Network", which are understanding sentimental and emotional behaviours in social networks at more granular level using Ekman's psychometric scale.

1 Introduction

Humans are emotional being and emotions affect almost all aspect of human lives. Humans are able to express their emotions from their verbal and written communications explicitly and implicitly as well. Moreover, as an emotional being humans get affected from the emotions of others too. One of the best way of mass communication, these days is social media. Now a days most of the people are active on various social media, and people make posts on these social media platforms frequently. These posts are in the form of text as well as other media format such as images, videos etc. Most of the time, the textual data in a post carries the emotional state of the user at the moment user wrote the post. Moreover the media such as images, videos along with texts such as post caption, tags shared by an user also explains user's emotional dimension.

The availaiblity of such large data on social media which contains information of emotional dimension inherently, attracted researchers recently to study the emotional behaviours in the social networks. Detection of communities in the social network is a popular research domain, but recently researchers have started incorporating emotional dimensions too for detecting emotional communities in such networks. In this review, some of the innovative works on the topic emotional community detection are reviewed. [5],[11] are the works which analyse user's tweets (i.e. text data) and [10][9] are the works which do analysis over social image networks.

The next section of the review lights on the general overview of all of these mentioned works, one by one. The third section talks about the data sets used. The Fourth section explains the methodology and technical aspects involved in each work along with comparison. The comparison analysis also include independent and interrelated pros-cons of these reviewed works. The empirical analyses on the reported results are discussed in section five. Finally review is concluded with some future extension notes.

2 Overview of Reviewed Papers

In the field of community detection the path breaking work of [4] is highly appreciated where they have tried to detect the edges between communities and by removing then trying to get better communities.[3] have talked about modularity which emphasizes on the intra community edges against the inter community edges. The most popular approach which takes modularity matric into consideration is [1]. In [7] and [6] they have worked on the emotion detection on the tweets. [8] have worked on community detection based on sentiment where as [2] have used naive bayes for improving modularity. [11] also deals with emotional behaviour of users in social networks. They have used 7 emotions for this task such as happy, good, anger, sorrow, fear, hate and shock. But unlike [5], which has used Twitter for their experiment [11] has used Sina-microblog, one of the most popular social network of china for their experiments to measure user's emotional similarity. They have used RostEA to find the similarity of microblogs that gives emotion value as either positive or negative depending on the statistics of the emotional words. Further, PCA similarity and Distance similarity measures has been used in [11] to compute the users similarity values. Also, in order to verify the emotional community results of Guser and Gmicroblog they have set null Hypothesis. It is reported in the paper that homophily is the pre-requisite for this method. It is found that emotional network is more suitable for detection of communities(emotional). Basically, emotional community means that the users in the same community are more denser.

Comparing to previous methods, which are mainly focused on text data, [10],[9] tried to fuse both the textual and image information to detect emotions. There have been studies on exploring emotional state of the user solely based on content of images posted by users. But, predicting user's emotions based on only images can hardly be precise enough. So, [10] try to detect emotions from images from a new angle. Their formulation is based on extracting emotions from an image as well as the comments of other users on that image. Moreover out of these comments on an image post some comments from close friends are given more significance. The rationale behind that can be understood with an example, suppose if a user posts a sad dark image, then general comments on the post might be related to photography skills, while the comments from close friends may reflect care which will lead to true emotion revelation of the user who posted an image. The authors of [10] propose a network formulation having users, images and comments as node. Moreover they have proposed a probability theory based framework, to combine textual and image (visual) features in the latent space, using LDA and Gaussian mixture models. The details of this proposed frame work is given in methodology section.

In [10], the friends of a user are supposed to reveal emotions in an image network more precisely, But in [9] authors, proposed a new framework where the position of a user is also taken into account to affect or reveal other connected users' emotional state. The rationale behind this formulation comes from the influential strengths of opinion leaders in a social media network. So, in [9] authors studied the emotion contagion phenomenon in social network considering the social role of a contagious user in the social media network. Their framework is briefly described in methodology section.

3 Datasets

The given table summaries about the datasets used in reviewed works.

Datasets Details			
Paper	Name	Type	Description/Statistic
[5]	Twitter dataset	Text	Dataset consists of tweets based on incidents like Malaysia Airlines Flight 370 disappearance etc.
[11]	Sina-microblog dataset	Text	The dataset collected from an online social network of China named Sina-microblog of more than 500 million users with unique IDs along with comments and tag.
[10], [9]	Flickr dataset	Text Images	Random selection of 3,50000 images posted by 4800 users, along with associated comments and tags.

Table 1: Dataset description of each reviewed work

4 Methodology

This section describes overall methodology of reviewed works ([5],[11],[10],[9]). Starting from describing the framework of [5] which is a multi stage pipeline. In this pipeline, the first step is to get the emotional status of the tweet on Ekman's scale using POS tagger, parser and knowledge base. then calculate the user impact based on post impact. The post impact is calculated by the following formula-

$$Post\_Impact = \frac{(Clicks+1)*(Favorites+1)*(Mentions+1)}{Tweets} * \frac{(Replies+1)*(Retweets+1)}{Tweets}$$

Next step is about how to categorized users based on user is emotional or neutral. tweets now collected from each user's timeline of the last 3 weeks and Then categorized them based on Ekman's scale. So whether user is emotional or neutral is depend on the percentage of their tweets marked as emotional. So 10 percent is marked tweets treated as emotional user. A modularity optimization algorithm a community detection algorithm is used to detect the community. The weight formula is given as-

$$Influence\_Metric = Post\_Impact * Frequency * \log(FtF + 1)$$

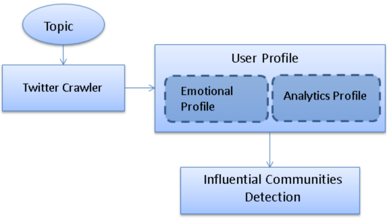

Figure 1: System Architecture.

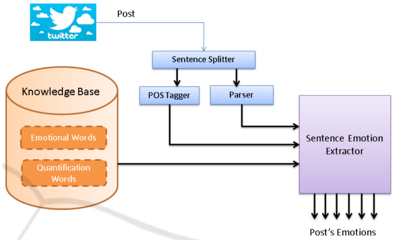

Figure 2: Architecture of Emotion Recognition Tool.

Figure 1: Working and Methodology of community detection ref[5]

In community detection step the graph is formed from nodes and edges where users are represented by nodes as the vector of their tweets' emotional scale and edges represent the following relationship between them.

Like previous paper, this paper, [11] also deals with emotional behaviour of users in social networks such as Sina-microblog in order to measure user's emotional similarity. It is found that emotional network is more suitable for detection of communities(emotional). Basically, emotional community means that the users in the same community are more denser. These communities have mainly two features: one is that the users in the same community should poses similar posting pattern and second is that they should have interconnection between them that means they should form edges. They have used modularity method with the help of CNM and BGLL algorithm for community detection task. The CNM follows hierarchical agglomeration algorithm to find community in a very large network. Being faster than many other algorithms, they take running time of the order of O(mdlogn), where n and m are vertices and edges respectively and d is depth of dendrogram. On the other hand, BGLL is a heuristic method based on modularity optimization that also takes less computation time and gives better insight of the communities so that we can discover sub-communities. However, homophily is the pre-requisite of the methodology used in this paper.

In this paper, the first step of the method used, is the construction of emotional network for which they have remodeled the network to a weighted and undirected emotional network to avoid the link structure. As shown in the figure. 2, the upper half of this figure shows original Sina-microblog network and the next diagram shows emotional network. In the lower half of figure.2, Guser and Gmicroblog represents user and microblogs emotional networks respectively. The second step of community detection has been done using CNM and BGLL algorithm. The CNM detects 5 emotional communities in which 2 are large, 2 is small and 1 is medium. Similarly, BGLL detects 14 communities out of which 8 were found small containing 1.61% of users whereas 6 were large with 98.39% of users.These communities represents different emotions like happy, anger, fear, hate, good etc. Further, the third step was experiment with contrastive network so that they could verify the emotional network as most suitable way for the purpose of community detection. In this case, they have taken 1 unweighted along with 3 weighted undirected networks. Finally, they have applied both CNM and BGLL algorithm on this contrastive network to find the communities that uses the 'following relations of each users.

The frameworks of methods fusing textual and visual features propsed in [10] is shown

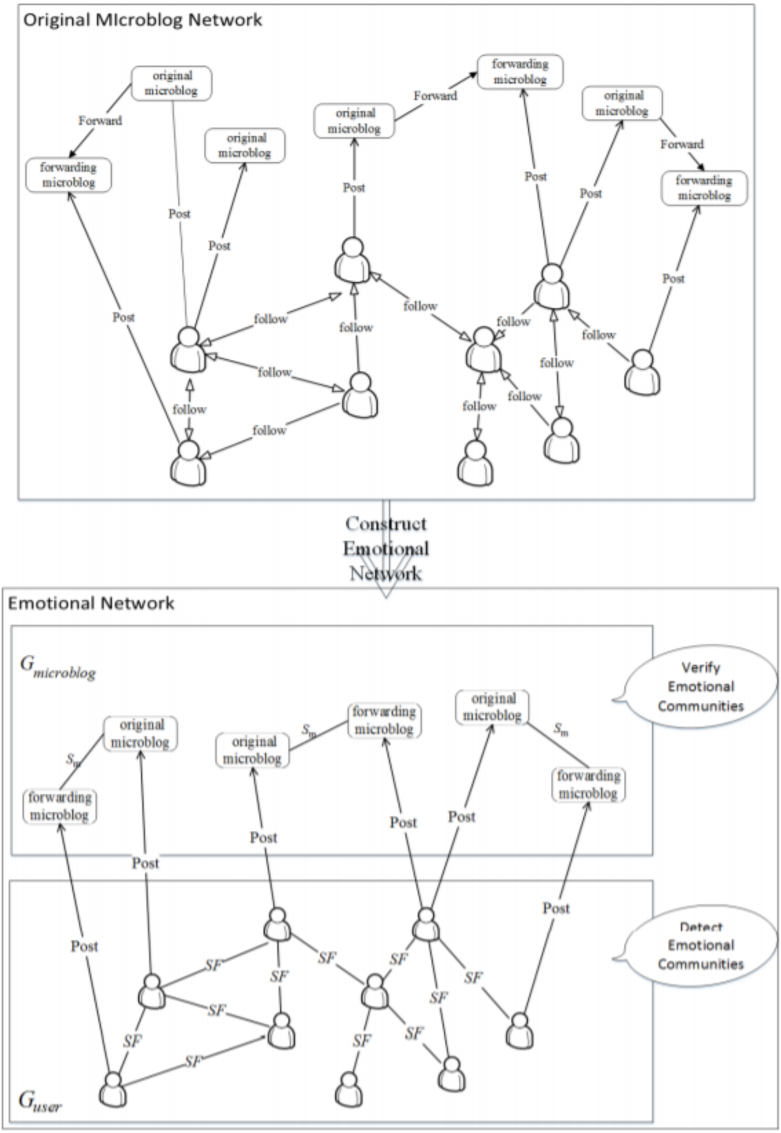

Figure 2: The Construction of emotional network from original Sina-microblog[1]

in figure 3. In this framework, the authors define a heterogeneous social media network as a directed graph, in which images posted by a user, all users, comments posted by users on images are considered as nodes. The four types of edges are defined as user-image edges indicating user posts that image, user-comment edges indicating user posts that comment, image-comment edges indicating that comment is posted about that image and user-user edges indicating that a user follows another user. They have used the image network from flicker website, dataset description is given in relevant section. To annotate the data, the authors define a set of wordlists from the synonyms, adjectives of the Ekman's emotional scale words (happy, surprise, anger, disgust, fear, sad) using SysNet and WordNet.

In this framework visual features of the images are extracted using Gaussian mixture models (Purple part of figure 3) and corpus of comments is modeled as bag of word models for baseline, for their enhanced approach the comments are modeled as mixture of topic modeling using LDA (Blei, Ng, and Jordan 2003) (Yellow part of figure 3). The tie between textual and visual information is learned in latent space using Bernoulli distribution generative process. This tie tells about the importance of a textual sample for an image.

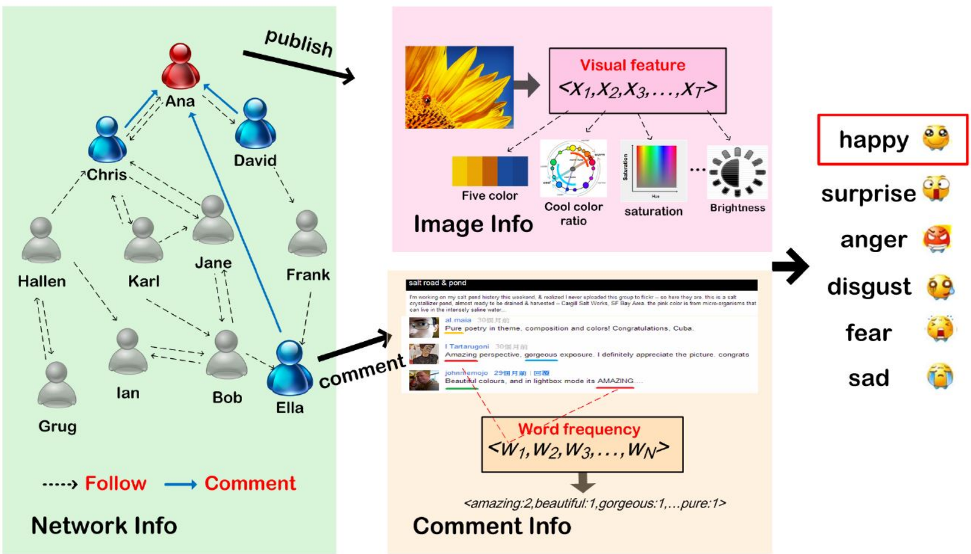

Figure 3: The proposed emotion learning framework presented in [10]

On the sameline similar to [10], the work presented in [9] has image network but it also incorporates the social role information of a user. Similar to the work in (Yang et al. 2015), Authors of [9] categorize users into three roles, named as ordinary users, opinion leaders and structural hole spanners depending on their network properties. In their framework, The emotion prediction of an image posted by user at a time t, is mapped as a probabilstic function. This function depends on the emotional status of user's friends at time t-1, user's own emotional state at t-1. Briefly saying this framework learns three things, 1. learn influential strength between friends by considering the social roles, 2. learn the dependency between emotions of the users at each next time step 3. learn the user dependent visual features i.e. learning the how an image posted by a user reflects user's emotion.

Now, after describing the approaches and technical aspects involved in these reviewed works. The comparison of the methodologies used in these works are done in subsequent subsection.

4.1 Methodology Comparison

The comparison analysis of the methodologies of these reviewed works is summarised in the form of tables. Table

Methodology Comparison			
Paper	Targeted Problem	Network Structure	Main Techniques
[5]	Emotional Community detection	Weighted and Undirected	Weighted community detection algorithm
[11]	Emotional community detection	Weighted and Undirected	Modularity method - CNM and BGLL algorithm. Both aggregation and split algorithms.
[10]	Emotion prediction	Heterogeneous (Images, Users, Texts/Comments) directed social network	Gaussian Mixture Model, LDA, SVM as classifier
[9]	Emotion contagion	Dynamic heterogeneous directed social network	Conditional probabilistic supervised learning

Table 2: Methodology comparison each reviewed works

Novelty Comparison	
Paper	Key Contribution and Novelty
[5]	Performed emotional profiling of user and calculated Post Impact and impact fact, then based on that community detection.
[11]	While both [5] and [11] deals with same problem i.e., emotional community detection but in different social networks, [11] shows that emotional network is better for this task.
[10]	Fusion of textual and visual features in latent space while other works [5],[11] based on textual features only, Major improvement in Precision score for emotion prediction
[9]	Extension of [10] using social role awareness based contagion study

Table 3: Contribution and Novelty comparison of reviewed works

5 Empirical Analyses

5.1 Comments on Results Reported

In this subsection, the improvements in metrics reported are analyzed and key findings are mentioned in the point wise manner for each reviewed work one by one.

- [5] gives better result than blondel method upto 5th rank community in terms of percentage of tweet. For percentage of retweets and follower it gives mixed result with respect to blondel method.

- The verification result with respect to density of the emotional networks in [11] showed that the metric factor of CNM (0.7567) and BGLL (1.0270) were larger as compared to K-means (3.4109).

- Also, from the results obtained in [11] from the average users' emotional similarity, it was clear that the value of f K-means (0.7241 for PCA) was larger than those of CNM (0.6205 for PCA) and BGLL(0.6210 for PCA).

- The experiments done on Flicker dataset in [10] shows the +37.4 % improvement in F1 score for inferring emotions with fused feature information comparing to baseline utilizing only visual features.

- Authors also compared with other approach which uses both features but without fusion in latent space, their finding suggests that using both features without proper fusion in latent space only slightly improves the performance.

- The results reported in [10] also indicates that SVM is a superior classifier for similar emotion prediction task comparing to Logistic Classifier.

- While the work in [10] reported that the emotional state of a user linearly influenced from the user's friends emotions. The work in [9] depicts more granularity indicating that a user's emotion state is superlinearly influenced by those friends who are opinion leaders in the network while sublinearly influenced by those friends who are structural holes (less important) in the network.

5.2 Pros and Cons of Reviewed Works

To summaries the pros and cons of the reviewed works in more organised way,Tables are formulated which compares the implementation pipeline requirement (IPR), assumptions involved, applications of each method and the robustness of each method.

Implementation Pipeline Requirements	
Paper	Involved Requirements
[5]	Need to access tweet from past three weeks for each user and find out emotion of the tweets and then calculate post impact and then calculate impact factor.
[11]	It is required to filter the grabled data. This has been done using regular expression that detects chinese characters.
[10]	Word-list based on synonyms of Ekman's scale words is required to annotate the training data
[9]	Similar to [10], the annotation of true emotion of an image based on synonyms extracted from SysNet and HowNet dictionary is required

Table 4: The implementation pipeline requirements of each reviewed paper

More the implementation preprocessing requirements involved, lesser the robustness of the framework, because a fault in initial preprocesing steps may incur higher errors in the results of the main methodology, e.g. If annotation of ground truth of emotions of the images in [10] is not reliable then the value of reported results is nullified. The social media networks are also affected by the propagation of the false information and every framework works under some predefined assumptions. So, a brief analyses on the robustness of these frameworks is made in next table, considering these two factors. Moreover one common yet well accepted assumption in all of these methods is the granularity of the emotions classified among only six classes (happy, surprise, anger, fear, disgust, sad) based on Ekman's Scale.

Robustness Analysis		
Paper	Susceptibility to False Data	Assumptions Involved
[5]	It is less susceptible to false data as it depends on knowledge base for classification of tweet on Ekman Scale.	If 10% of total tweet of a user is emotional them, they have marked the user as emotional user.
[11]	The closed emotion value shows most similar microblogs and users in the same community will have more significance	It is assumed that the distance data follows Gaussian distribution.
[10]	The method involved give higher significance to close friends' comments to infer true emotions, so method is very less susceptible to false comments by an adversory	The visual features of Images belonging to different emotion class would be separable, Close friends comment may reveal true emotional state, The annotations based on the word list overlap pipeline are actual ground truth
[9]	The method gives more significance to opinion leaders to infer a user's emotion, so in a network if a prominent opinion leader may propagate false emotions if somehow corrupted by an adversory	Same assumptions as of [10]

Table 5: Robustness Analysis of each reviewed work

6 Conclusion and Future work

The proposed framework in [5] is able to detect the emotional communities in the social network (Twitter) effectively. There are many fields for improvement in [5], such as how this framework be scaled up for bigger network.It can be used for advertising or influencing people. Also what features influences the community detection process and how feature-tuning can make it a better model can be analysed.

From the results of the experiments used in this paper[11], it can be concluded that emotional network works better for emotional communities detection task. Further, the emotional communities detection in other social network such as in [5] can also be done using the method proposed in [11]. An extension of the method used in [11] can be identification of more diverse communities such as combination of emotional similarity with other things like topics of interest etc.

Furthermore works like [10] are able to introduce the bridge between two different piece of information written and visual media on social network to identify emotions at more granular level. [9] incorporates influential strength of a user in the network based on user's social role. Moreover, the works [10],[9] targets only prediction of emotion of a user, these work can be extended into emotional community detection based on fused textual and visual features in a social role aware network structure.

References

[1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[2] William Deitrick, Benjamin Valyou, Wes Jones, Joshua Timian, and Wei Hu. Enhancing sentiment analysis on twitter using community detection. 2013.

[3] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[4] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[5] Andreas Kanavos., Isidoros Perikos., Ioannis Hatzilygeroudis., and Athanasios Tsakalidis. Integrating user's emotional behavior for community detection in social networks. In *Proceedings of the 12th International Conference on Web Information Systems and Technologies - Volume 1: SRIS, (WEBIST 2016)*, pages 355–362. INSTICC, SciTePress, 2016. ISBN 978-989-758-186-1. doi: 10.5220/0005862703550362.

[6] Jasy Suet Yan Liew and Howard R Turtle. Exploring fine-grained emotion detection in tweets. In *Proceedings of the NAACL Student Research Workshop*, pages 73–80, 2016.

[7] Valentina Sintsova, Claudiu Musat, and Pearl Pu. Semi-supervised method for multi-category emotion recognition in tweets. In *2014 IEEE International Conference on Data Mining Workshop*, pages 393–402. IEEE, 2014.

[8] Kaiquan Xu, Jiexun Li, and Stephen Shaoyi Liao. Sentiment community detection in social networks. In *Proceedings of the 2011 iConference*, pages 804–805. 2011.

[9] Y. Yang, Jia Jia, Boya Wu, and J. Tang. Social role-aware emotion contagion in image social networks. In *AAAI*, 2016.

[10] Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Qicong Chen, Juanzi Li, Chunxiao Xing, and Jie Tang. How do your friends on social media disclose your emotions? Proceedings of the National Conference on Artificial Intelligence, pages 306–312. AI Access Foundation, 2014.

[11] Jiang ZHU, Bai WANG, Bin WU, and Weiyu ZHANG. Emotional community detection in social network. *IEICE Transactions on Information and Systems*, E100.D(10): 2515–2525, 2017. doi: 10.1587/transinf.2016EDP7473.