Identification of Conflict Spreading Nodes in a Community Network

Hamid Ali, Muhammad Arshad Islam, Kifayat Alizai
Department of Computer Science
National University of Computer and Emerging Sciences,
Islamabad, Pakistan.
hamidriasat@gmail.com, arshad.islam@nu.edu.pk, kifayat.alizai@nu.edu.pk

Abstract—

Social media platforms are widely used for sharing public views.Those views can be used for finding different trends.In social media communities are formed by group of users and communities share views with one another. Understanding how communities fight and how to prevent conflicts is important to create better online environment.There is not much analysis on detections of online communities or individuals that leads to conflict. Because of lack of experimental data it is difficult to determine how conflict starts between different individuals or communities.In this research we studied how communities interect with other communities.Usually social media data is analyzed by performing sentiment analysis. These techniques do not perform well to transform available data facts collected by large number of system into useful knowledge.Also you cannot understand how users and communities are interacting with each other.Proposed approach is divided into five steps. Defining Communities, Sentiment Analysis,Bipartite Graph,TF-IDF and Page Rank.First step is to define communities.Second step is to calculate sentiment of tweet text.Third step is to generate bipartite graph to identify which communities are communicating or attacking to other communities.Fourth step is to find similarities between communities.In last step we use page rank to analyze attacker and defender interaction.We demonstrate the significance of our method on a Twitter dataset. We demonstrate that compared to previous methods, our proposed approach achieved significant results with more accuracy. h more accuracy.

Index Terms—Community, graph, twitter, bipartite, sentiment, conflict, network

1 Introduction

Rapid growth of internet has enabled public to share their opinions/interests with the rest of world. With invent of smart phones people are drastically sharing every event that happen with the rest of word.And news get viral within no time. [1] Web contains a lot of information about people's behavior. Web knowledge is used widely to understand user behavior or to predict different things based on the information collected from different social media platforms, blogs or websites. Now a days social medias are considered as an important platform where people from different parts of globe can communicate with each other and there is no check on who can post what? Because of ease access use of these social media platforms is increased drastically and a large amount of data is being generated rapidly. But [2] collecting data from different social media platforms require large amount of time and efforts. Mostly users on social media are linked to different communities (group of users with same thoughts) and communities interact with each other. There is a lot of research on the prediction of user behavior on the basis of social media platform data. But there is not much research how communities interact with each other. In this paper we are gonna study how communities interact with other communities and how conflict starts between different communities. It is very difficult to identify all the events and there is no proof of its authenticity.In social media communities are formed by group of users and communities share views with one another. Those views can be used for finding different trends. Some researches conclude that social media platforms like twitter, Facebook can be used for analyzing different trends.

Recently a lot of research is done to identify inter-community (user behavior with in one community) behaviors of users using social media platform data. [3] Usually social media data is analyzed by performing sentiment analysis. These techniques do not perform well to transform available data facts collected by large number of system into useful knowledge. Also you cannot understand how users and communities are interacting with each other [4] 74 percent of conflict between communities are started by less than one percent of communities. Understanding how communities fight and how to prevent conflicts is important to create better online environment. There is not much analysis on detection of online communities or individuals that leads to conflict.

According to [4] around 74 percent of conflict between communities are started by less than one percent of communities. Once conflict is started by the active members of communities they are carried out by less active members. Our aim is to identify those communities or individual who start conflict between communities.Some questions that we are gonna answers in our research. Do communities attack other random communities, or is there a relation between the source and target community? Attacker-Defender Interactions? Echo-chambers form during conflicts? Do communities attack other random communities, or is there a relation between the source and target community?

Proposed approach is divided into four steps. First step is the identification of communities. Second step is the Sentiment Analysis of the text in which three things will be determined polarity, subject, opinion holder. Third step is the creation of bipartite graph in which user and communities will be created to identify which communities attack whom. Then page rank will be used to identify how attacker and defender communicate with each other.

2 Related Work

A community is a place where people with same set of minds communicate and share their ideas with each other. There is not much analysis on detection's of online communities or individuals that leads to conflict.Mostly previous studies are based on small scale experiments that are carried out in labs or in small artificially created environment and because of lack of experimental data it is difficult to determine how conflict starts between different individuals or communities [5] [6] [7]. Previous work is mostly related to the detection of different communities [8] or interaction of individual within communities [9], or how users spread their time between multiple communities [10].

[11] created communities on a Twitter data. in his method user was considered as a node and his friends was considered as a edges. After identifying user his friends were mapped on his user node forming a small community. then two different approaches were used to create communities. First approach uses a well known Girvan Newman algorithm [12]is used whose first step is to calculates edge betweenness. Second step is to remove edges with high betweenness. then first step is repeated iteratively to remove all edges which has high betweenness . Second approach used Strongly Connected Components(SCC) concept. Scc consider communities in which every node is linked with other node. And there is no disconnected node in a community mean every member of community can communicate with other member.

[2] creates communities using common interest on twitter data. To find users with common interest celebrities were used.User with more than 10000 followers were considered as a celebrity and to find the category of each celebrity Wikipedia data was used. Data was converted as a directed graph where user was a node and a his followers were acting as a links.

Sentiment analysis is the process of finding opinion of given text using Natural Language Processing(NLP) techniques. Different techniques exist to find sentiment of a given text.

[1] takes the sentiment of tweet into account. For calculating the sentiment of tweet two methods was used .First method used lexicons from OpinionFinder which has manually annotated words into four categories (negative, neutral, positive or both) and overall sentiment of tweet is calculated by summing the polarities of each word in a tweet. Also one tweet was selected for representing each user. After applying it on a MAsen10 data set the accuracy of results was 41.41%.Second method [3] used lexicons from SentiWordNet 3.0.This contain many senses for each word/around 207000 words. For calculating sense of each word from multiple senses it checks if word has many negative senses then its negative and vice versa. After applying it on a MAsen10 data set the accuracy of results was 47.19%.

[13] uses page rank to identify the most important nodes in a graph. It incorporates both number and quality of links to determine the most important node in a graph. A well known search engine GOOGLE uses page rank to rank the websites.

3 Proposed Approach

This section contain detailed explanation of our proposed approach.Proposed approach is divided into 5 steps.

3.1 Defining Communities

First step is to define communities.For this we used twitter celebrities profiles. Each user with more than 1000 followers is considered as a celebrity.A directed graph G = (C,F) where C is set of celebrities and F is set of followers of celebrities. A follower-celebrity link (i, j) $\in$ F means that user i $\in$ C is a follower of user j $\in$ C . while on th e other hand a friend ship relation is where i follow j and j follows i. Next step is to define a point of interaction where different members of communities can communicate with each other. We consider twitter hashtag as point of interaction of twitter users. Through hashtag we can identify different community members who are taking about the same topic. Pseudocode for creating communities is given in ALgorithm 1. steps for assigning users to communities is defined in Algorithm 2. Steps for creating interaction point is defined in Algorithm 3.

Algorithm 1: Pseudocode for separating users with hashtags

Input: tweets file
Output: userTweetsList;
communityUsersHashtagsList;
hashTagUsersList;

userTweetsList $\Leftarrow \phi$;
communityUsersHashtagsList $\Leftarrow \phi$;
hashTagUsersList $\Leftarrow \phi$;
while *end of file* **do**
    read tweet from file;
    if *tweet user not in userTweetsList* **then**
        add new user to userTweetsList;
    else
    end
    read list of hashtags from user tweet text;
    if *hashtag not in hashTagUsersList* **then**
        add new hashtag to hashTagUsersList;
    else
    end
    calculate sentiment of tweet text;
    add sentiment to hashTagUsersList;
end

Algorithm 2: Pseudocode for reading user celeberties

```
Input: tweets followers;
userTweetsList;
communityUsersHashtagsList;
hashTagUsersList;
Output: updated userTweetsList;
updated communityUsersHashtagsList;
updated hashTagUsersList;
while end of file do
    read celeberty and follower from file;
    if follower in userTweetsList then
        set celeberty to userTweetsList ;
    else
    end
    if celeberty not in communityUsersHashtagsList
      then
        add new celeberty to
          communityUsersHashtagsList;
    else
    end
    set celeberty from communityUsersHashtagsList
      to new follower;
end
```

Algorithm 3: Pseudocode for creating interection point

```
Input: communityUsersHashtagsList;
Output: interectionPointList
foreach community key in
  communityUsersHashtagsList do
    foreach hashtag key in community users hashtags
      dict.hashtags list do
        if hashtag key not in interectionPointList then
            create empty temp list;
            add community key to temp list;
        else
            add hashkey of interectionPointList to
              temp list;
            add community key to temp list;
        end
        assign temp list to hashkey of
          interectionPointList;
    end
end
```

3.2 Sentiment Analysis

Second step is to calculate sentiment of tweet text. For calculating tweet text sentiment we are using TextBlob library of python. Textblob returns sentiment of tweet in a range of -1 to +1. if tweet sentiment is less than zero than it means its negative tweet. if its greater than zero then its mean its positive tweet. if sentiment is equal to zero then its neutral tweet. For our research purpose we classify tweets into two categories positive and negative.if tweet sentiment is less than 0 then its negative tweet other wise its positive. Pseudocode for calculating sentiment of tweet is given in Algorithm 1.

3.3 Bipartite Graph

Third step is to generate bipartite graph to identify which communities are communicating or attacking to other communities. Through bipartite graph we created who-posts-where network.We used NetworkX [14] library to create bipartite graph. Pseudocode for creating bipartite graph is given in Algorithm 4.

Algorithm 4: Pseudocode for creating bipartite graph

```
Input: interectionPointList;
communityUsersHashtagsList;
Output: graph
graph ⇐ φ;
tempListNodes ⇐ φ;
tempListEdges⇐ φ;
foreach hashtag in interectionPointList do
    add hashtag to tempListNodes ;
    foreach communityId in hashtag of
      interectionPointList do
        add edges to
          tempListEdges(hashtag,communityId)
    end
end
add tempListNodes and tempListEdges to graph
```

3.4 TF-IDF similarity

Fourth step is to find similarities between communities. We used sklearn library to calculate TF-IDF similarity between communities tweets text.

3.5 Page Rank

In last step we use page rank to analyze attacker and defender interaction.This step is divided into two phases. In pahase one run PageRank but restrict the teleport set to just attackers. In phase two run PageRank but restrict the teleport set to just defenders. In order to calculate page rank we used NetworkX [14].

4 Experiments Result

Proposed approach is evaluated on publicly available data set of twitter followers [14].This was complete followers data set of twitter in 2009. Two files celebrities profiles and twitter followers is used in our research.Both are tab separated files. An example of celebrity profile entry is shown in Table 1. An example of twitter followers entry is shown in Table 2.

User tweets dataset was removed from website because it was violating Twitter Terms and Conditions. So we created our own Twitter scraper using Twitter Streaming Api. we scraped 200 tweets of each users. it is total of 1857270 tweets of 8971 users. Our scraped tweet data set is available here [14].

After applying threshold of minimum users of 10 users and minimum 50 hashtags in a community we detected 18 communities. Description of communities with users and hashtags is shown in Table 3.

15 hashtags that were used by mostly users are shown in figure 1.

We created bipartite graph to identify which communities are communicating or attacking to other communities. A sample bipartite graph is shown in figure 4. We found that 5 communities (14, 18, 21, 34, 53) are creating conflict because they are using more negative sentiment for other communities.

From attacker and defender page rank score (shown in figure 2 and 3) we can see that attackers are close to other attackers and defenders are close to others defenders in their respective community. This also means that they are forming echo chambers.

TABLE 1
A detailed description of celebrity profile entry.

Format	Example
numeric id	12
verified	False
profile sidebar fill color	EADEAA
profile text color	333333
followers count	895829
protected	False
location	San Francisco
profile background color	8B542B
utc offset	-28800
statuses count	4209
description	Chairman and co-founder of Twitter
friends count	574
profile link color	9D582E
screen name	jack
profile background tile	False
favourites count	614
name	Jack Dorsey
url	None
created at	Tue Mar 21 20:50:14 +0000 2006
time zone	Pacific Time (US and Canada)
profile sidebar border color	D9B17E
following	False
gender infered by name	m

TABLE 2
A detailed description of celebrity-follower profile entry.

Format	Example
USER	12
FOLLOWER	13

5 Conclusions

This paper proposed a solution for identification of conflict spreading nodes in a community network. First we formed communities on twitter data and then we identified which communities are creating conflict. we used hashtag as a point of interaction to check how users of one community communicate with users of other communities. We have use page rank algorithm. Where teleport set was restrict to only one type of node to find the importance of that node in a community.Through this we proved that mostly users communicate to users of same community.

TABLE 3
Name of communities with users and hashtags

Name	Users	Hashtags
12	1419	298
13	2294	298
14	431	290
15	537	282
16	53	92
17	421	289
18	322	270
20	1172	298
21	342	276
22	410	273
23	213	217
31	210	171
34	127	182
47	332	265
52	105	104
53	217	245
56	12	11
57	167	246

[p]

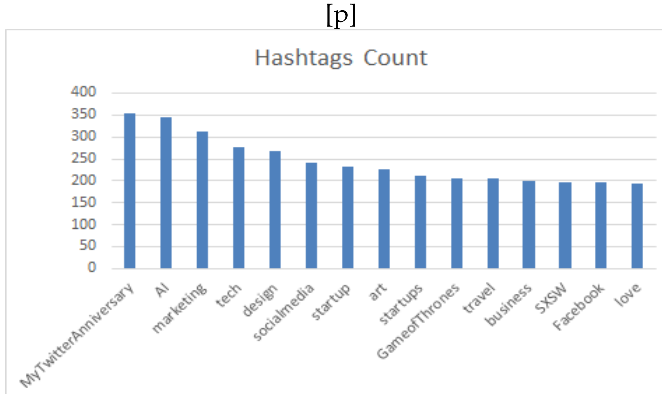

Fig. 1. 15 highest used hashtags with number of tweets in which they were used.

[p]

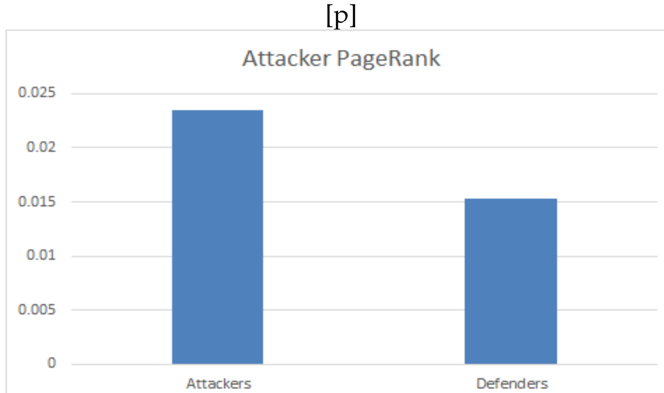

Fig. 2. Page Rank where teleport set is on for attackers. Attackers have higher page rank score than defenders which means that attackers are more close to other attackers.

6 Future Work

For future work, we will explore machine learning to make prediction on identification of conflict spreading nodes.

[p]

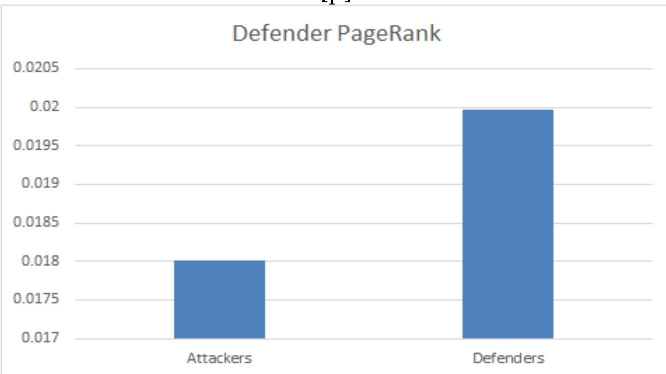

Fig. 3. Page Rank where teleport set is on for defenders. Defenders have higher page rank score than attackers which means that defenders are more close to other defenders. With respect to attackers defenders are more close to each other because their average page rank score is higher compared to attackers average score.

[p]

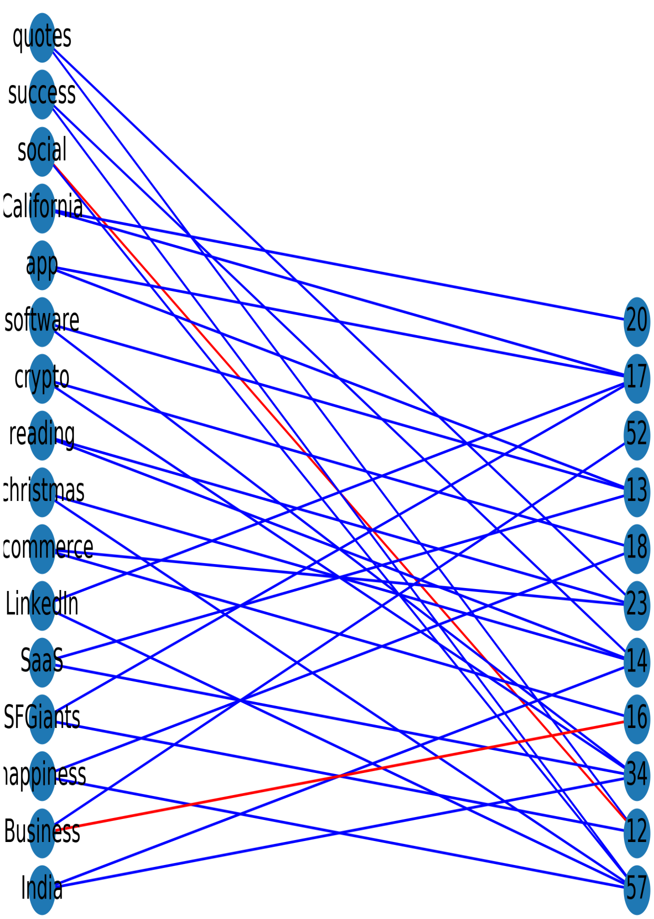

Fig. 4. Bipartite Graph of communities with hashtags. Each numerical node represent a community. Nodes with string represent a hashtags. Edges represent which community talked about which hashtag. Here blue edges represent that an average of positive sentiment of tweet is associated with this hashtag. And red edges represent that an average of negative sentiment is used for this hashtag.

7 Acknowledgment

We like to acknowledge National University of Computer and Emerging Sciences, Islamabad to give us opportunity and their resources to research in this field.

References

[1] J. E. Chung and E. Mustafaraj, "Can collective sentiment expressed on twitter predict political elections?" in *AAAI*, vol. 11, 2011, pp. 1770–1771.

[2] G. Suryateja and P. Saravanan, "Optimization of community detection in twitter social network," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 20, pp. 745–758, 2018.

[3] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, "Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to italy and france," *New Media & Society*, vol. 16, no. 2, pp. 340–358, 2014.

[4] S. Kumar, W. L. Hamilton, J. Leskovec, and D. Jurafsky, "Community interaction and conflict on the web," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web.* International World Wide Web Conferences Steering Committee, 2018, pp. 933–943.

[5] M. E. Hewstone and R. E. Brown, *Contact and conflict in intergroup encounters.* Basil Blackwell, 1986.

[6] M. Sherif, *Group conflict and co-operation: Their social psychology*. Psychology Press, 2015.

[7] H. Tajfel, J. C. Turner, W. G. Austin, and S. Worchel, "An integrative theory of intergroup conflict," *Organizational identity: A reader*, pp. 56–65, 1979.

[8] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study," *Acm computing surveys (csur)*, vol. 45, no. 4, p. 43, 2013.

[9] R. E. Kraut and P. Resnick, *Building successful online communities: Evidence-based social design*. Mit Press, 2012.

[10] W. L. Hamilton, J. Zhang, C. Danescu-Niculescu-Mizil, D. Jurafsky, and J. Leskovec, "Loyalty in online communities," in *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[11] K. H. Lim and A. Datta, "Finding twitter communities with common interests using following links of celebrities," in *Proceedings of the 3rd international workshop on Modeling social media*. ACM, 2012, pp. 25–32.

[12] M. E. Newman, "Fast algorithm for detecting community structure in networks," *Physical review E*, vol. 69, no. 6, p. 066133, 2004.

[13] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[14] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*. AcM, 2010, pp. 591–600.