ISSN No. 0976-5697

Volume 5, No. 5, May-June 2014

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Framework for Knowledge Discovery from Facebook

Ashish Kumar
Computer Science and Engineering Department
Maulana Azad National Institute of Technology
Bhopal, India

Nikhil Kumar Singh
Computer Science and Engineering Department
Maulana Azad National Institute of Technology
Bhopal, India

Deepak Singh Tomar
Computer Science and Engineering Department
Maulana Azad National Institute of Technology
Bhopal, India

Abstract: In the 21st century Social Network became one of the most popular and easiest platforms for sharing information. Millions of people daily share billions of information over social networking site. Facebook is the most popular social media network around the globe today. It has a global audience of 1,230 million users. With the increasing number of posts, possibility of fake / illegal post are also being increase. Careless use of social media by user can also have a negative effect on their digital reputation. Effective analysis plays an important role in the social media crime investigation. Whenever an activity is conducted on social networks various controlling attributes are generated. So by analysing controlling attributes it is possible to investigate public activities and filter out the disruptive activities. In this paper structural information of Facebook, Facebook Query Language (FQL) and activity analysis using FQL are explored. A frame works for classify the end user by examine the flow of attributes for an activities is also presented.

.

Keywords: Social Network, Facebook, Graph Explorer, Activity, Facebook Query Language, Classification

I. INTRODUCTION

With the prevalence of social networks physical space between users is reduced. Social networks are the platform which allows user to user interaction. Beside interaction these platforms are also helpful to keep an eye over activities (photos, posts, location etc.) performed by a user. Social networks are the fastest way for spreading news so Nine out of ten Egyptians and Tunisians responded to a poll indicating that they used Facebook to organize protests and spread awareness [1]. As mentioned above the news can be spread fast over the social network, the main reason for this is the mouth publicity [2]. As it is the tendency of the human being to share any information with their beloved ones. Now-a-days social networks provide an easy platform to communicate or share any information. So these platforms are used for promote or defame the product [3].

.

With the introduction of social networks there are varieties of activities on internet. Like other media, social networks also become a source of data leakage, intrusions, and vulnerable entry point for malicious content. Government and Security agencies are using social network for finding activities performed by users. Tracking the activities of the user is not only necessary but there is need to focus on information generated by those activities. So there will be some mechanism to identify the truthfulness of any information and if it doesn't meet the requirements then that information will not be allowed to spread in social networks. The common social networks used today are Facebook, Twitter and LinkedIn. Now-a-days most used social network is the Facebook so it is a large pool of information about users.

Facebook is a social network platform which allows users to communicate and to look over the activities of users. It is different from other social networks such as Twitter and LinkedIn. Facebook provides facility of being friend to other users and to chat with them which is not provided by the Twitter. Twitter's only facility is to follow users for getting their updates. In LinkedIn the facility provided are somehow same to the Facebook but it is only for the professional purpose while in Facebook there are various applications by which user can be entertained and user can also develop its own application. This paper focuses on the identification of user's activities on Facebook with the help of association [4,5] and correlation [6] of different Facebook relation.

II. STRUCTURE OF FACEBOOK

In Facebook, graph (G(V,E)) is used to represent structure of Facebook. Where vertex (V) represents the objects such as users, applications, pages, groups and edge (E) is used to represent relationship between these objects.

Facebook objects are the discrete entities that are perceived by an individual. Facebook objects are categorized on basis of their common characteristics (login, share, like etc.) as shown in figure 1.

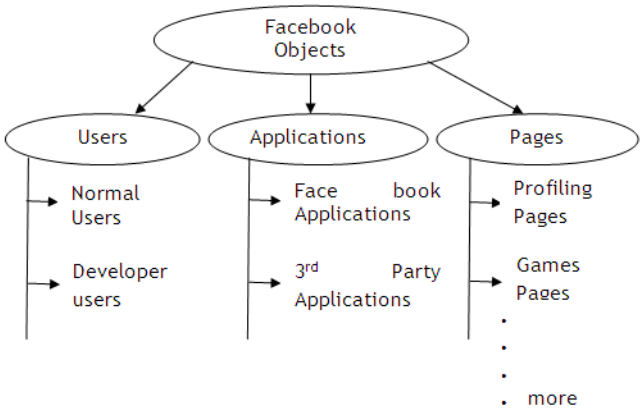
A. User Object

It is a registered and authorized person having authority to use Facebook services. On the basis of their services usage users are classified as normal User and developer User. Normal users refer as end user of Facebook whereas developer user is the admin user of its respective application..

B. Application Object

Applications are the utilities provided by Facebook for users to entertain and for accessing their performance. On the basis of their creator applications are classified as Facebook application and 3rd party application.Facebook application is developed by its own consortium and Facebook also

provides a facility to develop 3rd party application by developer user.



.Figure 1: Facebook Objects

C. *Page Object*

A Facebook page is a public profile specifically created for businesses, brands, celebrities, causes, and other organizations. These pages share the pictures, videos, statuses update. So since these pages are public the person can use them to propagate disruptive stories.

Facebook objects are responsible for generating the activities. Some of the general activities performed by the Facebook objects are shown in table1 [7].

Table 1: General Activities

S.No.	Activities	Actions
1.	Login	Authorize user to use the service
2.	Stories	Stories are related to the post for the various attribute of Facebook.
3.	Sharing	Allows user to see the activities of other users and also share his activities.
4.	Likes	People share your pages and content back to their Facebook profile.
5.	Comments	The Comments lets people to give special attention on any piece of content by user.
6.	Chat	Used for allowing the friends to exchange messages.
7.	Build	Using Facebook platform for developing the application and games for user.
8.	Grow	With help of sharing and ads developer can increase the number of users using it.
9.	Monetize	Make application more interesting with help of new offers so people will buy them.

Action performed by an object is recognized as an activity. Object Activity play a vital role in detailed examination of action performed by object over the Facebook and help to verify the credibility of users and their posts.

III. ACTIVITY ANALYSIS

Activity analysis is a mind hunting task and become more challenging because of strong privacy settings like “public” and “custom” as follows:

- ***Public:*** Allows visibility to all users.
- ***Friends:*** Allows visibility to user’s friend only.
- ***Only me:*** Only visible to user.
- ***Custom:*** Allows visibility to some specified friends.

General data provided corresponds to any activity is not yet sufficient to analysis action over the network. For extracting more data corresponds to any activity numerous data extraction techniques had been proposed [8, 9].

Activity analysis also suffers from data inconsistency problem because different privacy settings will produce different types of data attributes. Feature selection plays a vital role to overcome this data inconsistency problem.

IV. DATA EXTRACTION

Data extraction is the process of crawling relevant information from data sources and plays a vital role in exploring the activities of Faccebook. Joseph Bonneau, Jonathan Anderson et. al. [8] proposed common methods to extract data from Facebook. Common methods for extracting data from Facebook are as follows:

A. *Public Listing*

With the help of crawlers public profiles of users are extracted. Public profiles consist of a user’s general information. Crawlers are used for extracting the profile information from Orkut [9] and from Facebook [10].

B. *False profile*

One can create a false profile to gather more data than it is available in public view. With help of false profile person can send friend request to highly connected persons in order to view more data than publically.

C. *Compromise and Phishing*

With help of operating system attacks such as malware and key-loggers Facebook profiles can be compromised. Phishing is effective against Facebook, as Facebook sent links for login to see the photos or messages.

D. *Malicious Applications*

Facebook provides platform for developing applications by 3rd party [11]. Mostly these applications have full access of the user’s profile data. Applications are created for extracting the user’s activities by integrating FQL with the application.

E. *Graph Explorer*

Facebook has provided the Graph Explorer tool [12] this tool is helpful to get the information corresponding to the nodes.

In this paper Graph Explorer Tool is used for extracting data. In Graph Explorer there are two methods defined for extracting information about any node. The two methods are:

1) Graph API

It is the primary way to extract data from the social network. Most of the Graph API calls required a valid access token along with some data extraction permission like user data permissions, friends' data permissions and extended permissions.

2) FQL Query

FQL is Facebook Query Language that supports to access the features available in Graph API in the SQL style interface. Basic query structure of FQL Query is:

SELECT [fields] FROM [table] WHERE [conditions].

V. FACEBOOK QUERY LANGUAGE

FQL is a query language that can be used to extract the relevant information from Facebook dataset same as SQL which is used to extract the relevant information form relational dataset on the basis of some command that can be executed over database portal.

In this paper FQL [13] is used to extract data rather than Graph API because FQL provides advanced features not available in the Graph API such as using the nested query. Same as SQL, FQL requires the knowledge of the structural data set of Facebook to extract data. Facebook provides the restricted data set with very limited access permissions to ensure the privacy and security of the Facebook database.

FQL can only extract the data on the basis of the indexed attribute of the respective table, as explored in Q1. Q1 is used to extract information about any post like image in post, message in post, date of post etc. on the basis of the post id. All this information is stored in stream table of Facebook data set with 3 indexed attributes post id, filter key source id.

VI. FEATURE SELECTION

Feature selection is the process of selecting key features from extracted data [14] to remove data inconsistency problem as defined in section 3. As extracted information have many features corresponding activity.

Features provide information related to post and user, on the basis of that features are classified as post features and user features. Post features helps to identify what type of post it is and how it is propagating in the Facebook network while user features helps to know about the user who posted the story, to whom story is posted.

Table 2: Features and Their Significance

Feature Selection			
Post Features		**User Features**	
Features	**Information**	**Features**	**Information**
Privacy	Find out to whom post is shared.	Can Like	Users who are viewing the post have permission to like it.
Like Info	How many likes on a post.	User Like	User who have posted story have liked it or not.
Share Info	How wide the post is spread	Target Id	To whom story is published.
Message Body	Content of post.	Parent Post Id	Original source of the story.
Source Id	From where the story is published.	Actor Id	ID of the user, page, group, or event that published the post.
Type	Story is a status, group post, link posted etc.		
App Data	Information about attachments of a post		
Attribution	Tells us from which app the post is published.		

Q1: Query for finding the information about the post:

FQL Statement

```
SELECT actor_id, app_data, attribution, comment_info, is_hidden, like_info, message, parent_post_id, permalink, privacy, share_info, timeline_visibility, type, created_time, updated_time FROM stream WHERE post_id="344128252278047_792267827464085"
```

Result of Q1


```
{
 "data": [
  {
   "actor_id": "344128252278047",
   "app_data": {
    "attachment_data": "[]",
    "images": "[792267740797427]",
    "photo_ids": [
     "792267740797427"
    ]
   },
   "attribution": "Facebook for Android",
   "comment_info": {
    "can_comment": true,
    "comment_count": 1459,
    "comment_order": "ranked"
   },
   "is_hidden": false,
   "like_info": {
    "can_like": true,
    "like_count": 111878,
    "user_likes": false
   },
   "message": "Great to meet a former 'India' player and a legend.",
   "parent_post_id": null,
   "permalink":
"https://www.facebook.com/SachinTendulkar/photos/a.792251527465
715.1073741827.344128252278047/792267740797427/?type=1",
   "privacy": {
    "value": ""
   },
   "share_info": {
    "can_share": true,
    "share_count": 1635
   },
   "timeline_visibility": "normal",
   "type": 247,
   "created_time": 1401298323,
   "updated_time": 1401773092
  }
 ]
```

VII. FRAMEWORK

Whenever user post a story over the social networking site, user and server machine encapsulates numerous of controlling attributes such as (privacy setting, story id, type of post) with original story in order to provide security and aggregation of that story.

Proposed frame work is deployed to discover control flow of selected features by gathering the controlling attributes added by user and server side, which is extracted by using relational query language (like FQL in Facebook) and generate effective association rule that can be use to classify end user.

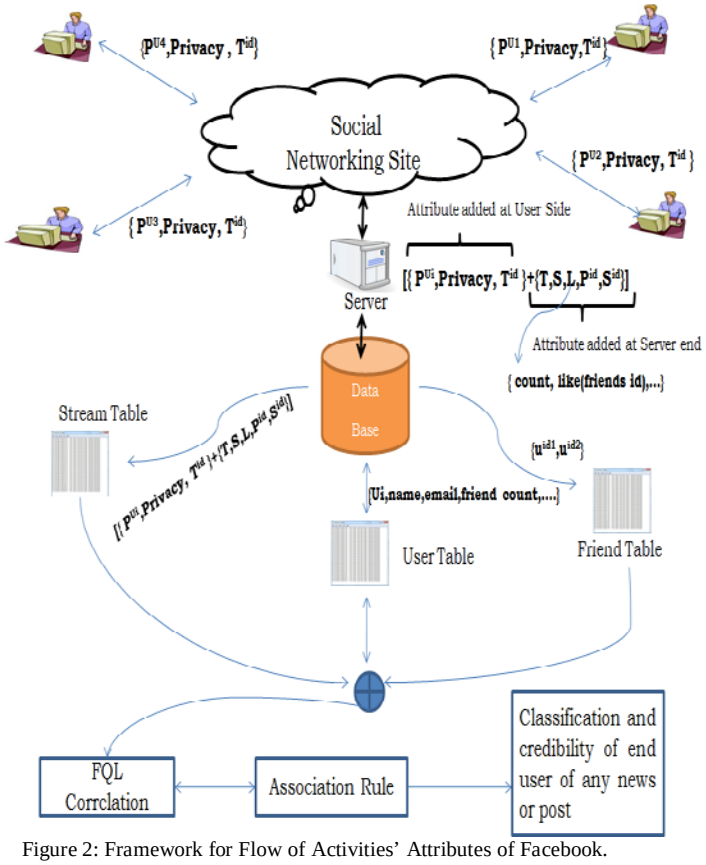


Figure 2: Framework for Flow of Activities' Attributes of Facebook.

When a user publishes any post on Facebook various attribute are added at user & server side.

User Side: User side add attribute like post_id (P^{Ui}), target_id (T^{id}), where post_id (P^{id}) is used to identify every post uniquely and target_id is use to define destination address.

Along with that user side also responsible for privacy setting of post in order to maintain the proper visibility of post such as public or private.

Server Side: Other attributes related to post are added by server side. Attributes (T, L, S, P^{id}, S^{id}) contains the more detailed information about post.

T: Defines type of post such as status update, photo update, post in a group etc.

L: Details corresponds to likes of the post such as total like count, users who likes it etc.

S: Details corresponds to share information such as total shares, user liable to share etc.

P^{id}: Parent post Id i.e the originator of the post.

***S^{id}*:** Source Id i.e publisher of the post.

These attributes are globally distributed over the different face book database tables. The attribute set [{PUi,Privacy,Tid}+{T,S,L,Pid,Sid}] shown in figure 2 stored in the stream table of face book database.

After extracting important feature through FQL, correlation will apply to find the association rules which will be used classify end user on the basis of feature of stories posted by him.

Correlation technique will be helpful for knowledge engineer to find the statistical relationship between selected features. Features such as like count, share count, comment count will be used to find correlation. On the basis of degree of correlation between different feature association rule will be generated and used to classify end user into different classes based on their browsing behaviour on Facebook.

Currently the implementation of data extraction and feature identification from Facebook are all most completed and we are at the stage of formulised association rule from extracted information.

VIII. CONCLUSION

When user navigate on to social site such as Facebook or twitter the huge amount of data is generated which may be used for web based knowledge discovery. In this work a framework is proposed for effective knowledge discovery from Facebook. Due to the design complexity of Facebook it is difficult for researcher to understand the internal design and structure of facebook.

Towards a better understanding for how Facebook works, data extraction and feature selection has been explored which is further used by knowledge engineer for enhancing Facebook services and privacy preservation. The future work includes the formulation of association rule.

IX. REFERENCES

[1] F. Jin, E. Dougherty, P. Saraf, Y. Cao, N and Ramakrishnan, "Epidemiological Modelling of News and Rumors on Twitter", Proceedings of the 7th Workshop on Social Network Mining and Analysis, 2013, Article No. 8.

[2] Yuko Tanaka, Yasuaki Sakamoto and Toshihiko Matsuka "Toward a Social-Technological System that Inactivates False Rumors Through the Critical Thinking of Crowds", Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS-46), pp. 649-658.

[3] Mukherjee, A., B. Liu, and N. Glance, "Spotting Fake reviewer groups in consumer reviews", Proceedings of the 21st international conference on World Wide Web 2102, pp. 191-200.

[4] Rakesh Agrawal, Tomasz Imieliński and Arun Swami, "Mining association rules between sets of items in large databases". Proceedings of the 1993 ACM out of a social

network. Proc. of the First SIGMOD Int. Conference on Management of data - SIGMOD '93. p. 207.

[5] Yang, J.,"Classification by association rules: The importance of minimal rule sets", In ICML-2003.

[6] Nikhil Kumar Singh, Deepak Singh Tomar, Bhola Nath Roy, "An approach tounderstand the end user behavior through log analysis" International Journal of Computer Applications (0975 – 8887), Vol. 5, no-11, August 2010.

[7] Facebook Developers, "Social Plugins" Internet: https://developers.facebook.com/docs/plugins/ [April 14, 2014].

[8] J. Bonneau, J. Anderson, and G. Danezis, "Prying data out of a social network", International Conference on Advances in Social Networks Analysis and Mining, 2009

[9] Rajni Ranjan Singh and Deepak Singh Tomar (2009). Approaches for user profile Investigation in Orkut Social Network". In: International Journal of Computer Science and Information Security 6.2, pp. 259-268.

[10] Zhefeng Xiao, Bo Liu and Huaping Hu, "A Facebook crawler based on interaction simulation and MHRW-DA," Computer Science and Network Technology 2012 2nd International Conference pp. 2041-2044, Dec. 2012

[11] Makridakis, E. Athanasopoulos, S. Antonatos, D. Antoniades, S. Ioannidis, and EP Markatos. Designing malicious applications in social networks. In IEEE Network Special Issue on Online Social Networks, 2010.

[12] Facebook Developers, "Graph API Explorer" Internet: https://developers.facebook.com/tools [April 20, 2014].

[13] Facebook Developers, "Facebook Query Language" Internet: https://developers.facebook.com/docs/reference /fql/ [April 26, 2014].

[14] Jiliang Tang, Xia Hu, Huiji Gao and Huan Liu, "Unsupervised Feature Selection for Multi-View Data in Social Media" In KDD 2012.

