Detection of Offline Social Communities Using Unique Identifier-Based Methods

Carlos M. Silva
Department of Software Engineering, University of Lisbon, Lisbon, Portugal

***Abstract* — with the extensive use of internet technologies, people today are likely to establish online interactive sessions from participating in social networking media like WhatsApp, Twitter, Facebook, LinkedIn etc. Based on the behavioral and structural aspects, communities can be distinguished and further be analyzed in a comprehensive fashion. However despite technological advancements in internet communications, some sections of people are neither lack of technical skill nor showing not interest to be participant. This is most happening thing in people who are involving in accomplishing mischievous activities. Our proposed system is to detect the fraudulent activities of members who are not explicitly or implicitly a part of online social networks. As on today Aadhaar identity playing a predominant role in linking the users with different agencies and delivering fruitful information to all major departments whenever is needed., it is taken as a criterion for investigating the ongoing crime activities in and around our country.**

Keywords—Offline Community; UIDA; community detection; Aadhaar; OSN; actors;

I. INTRODUCTION

A national wide unique identification project, Aadhar card launched by unique identification authority of India (UIDAI) [1], now widely recognized as a unique identity to residents based on demographic as well as on biometric statistics [2]. Having linked individual biometric finger prints and iris to their accounts [3], certainly resident may find duplicates of their own numbers [4]. In this context, customers with unique identification number eliminates frauds and duplicate identity, so that the individuals will no longer be able to represent themselves differently to different agencies, such as banks, post offices, NGO s, public distribution system, education, social welfare schemes etc. Nowadays, linking customer aadhar card to their accounts is made mandatory for acquiring government schemes, banking transactions, gas, water and other connections, sales and purchase [5]. The objective of this paper is to network the aadhar card users and build an offline community for identifying the relationship between the actors.

. In general terms, a community is a set of vertices that have more internal connections, then external connections [6]. In other words a set of people have densely connected when compared to other people [7]. Community detection is done in any of the three ways: using topological features or using addition information associated with nodes and edges or combination of both. Community detection can be applied to classify people viz., influencers, identifying correlated structures in protein-protein interaction networks, grouping web clients [8].

As on today every transaction is linked with aadhar number and hence all other details of individual can be inferred from the single unique identification number [9] i.e aadhar card. Hence each vertex of the graph is labeled with an aadhar identity and edge is weighed based on the count of transactions made by an individual with others. A critical study is made on different existing community network detection mechanisms till data and a survey report is presented to research community.

II. RELATED WORK

A. Online Social Network (OSN)

OSN is a dedicated website or other applications which enables users to communicate with each other by posting information,comments,messages,images etc. For example now a days most of the people are members of one or more online social networks like Facebook,Twitter,LinkedIn, Whatsup and Orkut etc. In their Social networks they have own network based on their interactions among the members. Some of the applications of OSN are

1) Gene network clustering
2) Web page clustering
3) Image segmentation
4) Detection of terrorist in SN
5) Link Prediction
6) To detect suspicious events in social networks.
7) Online Recommendation system etc.

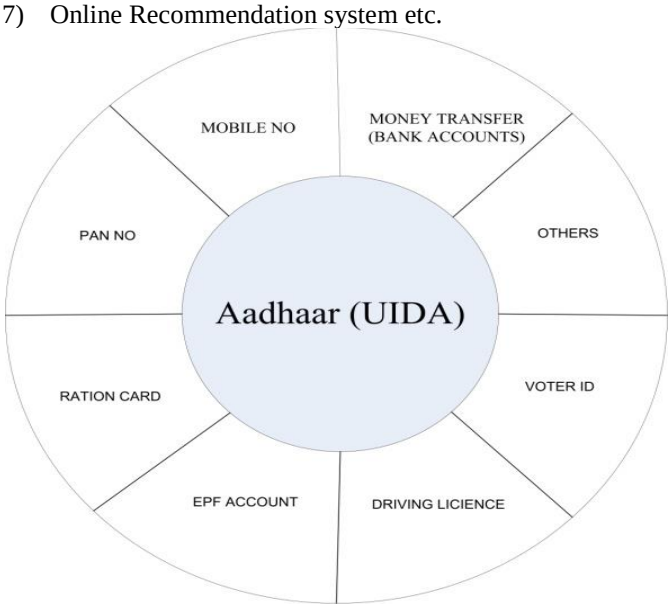


Fig.1

B. Off-Line Social Network (OffSN)

A network of social interactions and personal relationships in offline (ie these networks are derived from offline activities of the people in the society). Some of the applications of OffSN are

1) Crime related issues
2) Detection of terrorist in Online SN
3) Relation Prediction
4) To detect suspicious events in Offline social networks.
5) Offline community Recommendation system
6) Information Diffusion etc.

Generally multiple communities' forms social networks. Community detection is a NP-Hard problem [9]. Community have many definitions, most of the definitions based on density of the nodes with more interactions between them. In social networks, community is group of nodes having more relations compared to other nodes. Analysis of social networks is two types: Link and Structure based and Content-based analysis. Network structure analysis describes how many nodes in the network and Link based analysis explains how they are interacted with each other. We mentioned various types of networks based on type and social network nature. They are static, dynamic, direct, heterogeneous, signed and positive.

Main static social community detection methods are Graph partitioning, Spectral bisection method, Max-flow min cut method, Level structuring partitioning, Geometric algorithms, multilevel algorithms, Girvan-Newman, Enf-Gibbs and Hierarchical clustering algorithm etc. In many social networks, interactions and activities among people change over time. Due to these reason dynamic networks comes into the picture. Various dynamic community detection algorithms are evolutionary clustering, Graph-Scope algorithm, Palla algorithm, FaceNet algorithm and Event-based Framework etc. Now a day's directed social networks such as web pages, Twitter and citation networks are very popular. We can analyze directed networks by using Leicht & Newman algorithm and satuluri & Parthasarathy algorithm etc. In heterogeneous social networks each entity represents a certain type of relationship, and performs a specific role in a specific task. SONAR, Mincut, Regression based algorithms are best for Heterogeneous networks. Signed networks having both positive and negative links (both positive relationships as well as hostile relations). FEC, CRA and Dorian & Mrvar algorithms get best results for Signed networks. In positive networks all edges represents positive relationships. There are three categories of methods for above mentioned networks: Graph theoretic methods such as spectral methods, physics-based methods, the spectral methods and Kernighan-Lin algorithm. Hierarchical methods such as division methods based on betweenness metrics, Tyler algorithm, Radicchi algorithm, agglomerative methods, MFC (max flow communities), HITS algorithm, SAE algorithm etc.

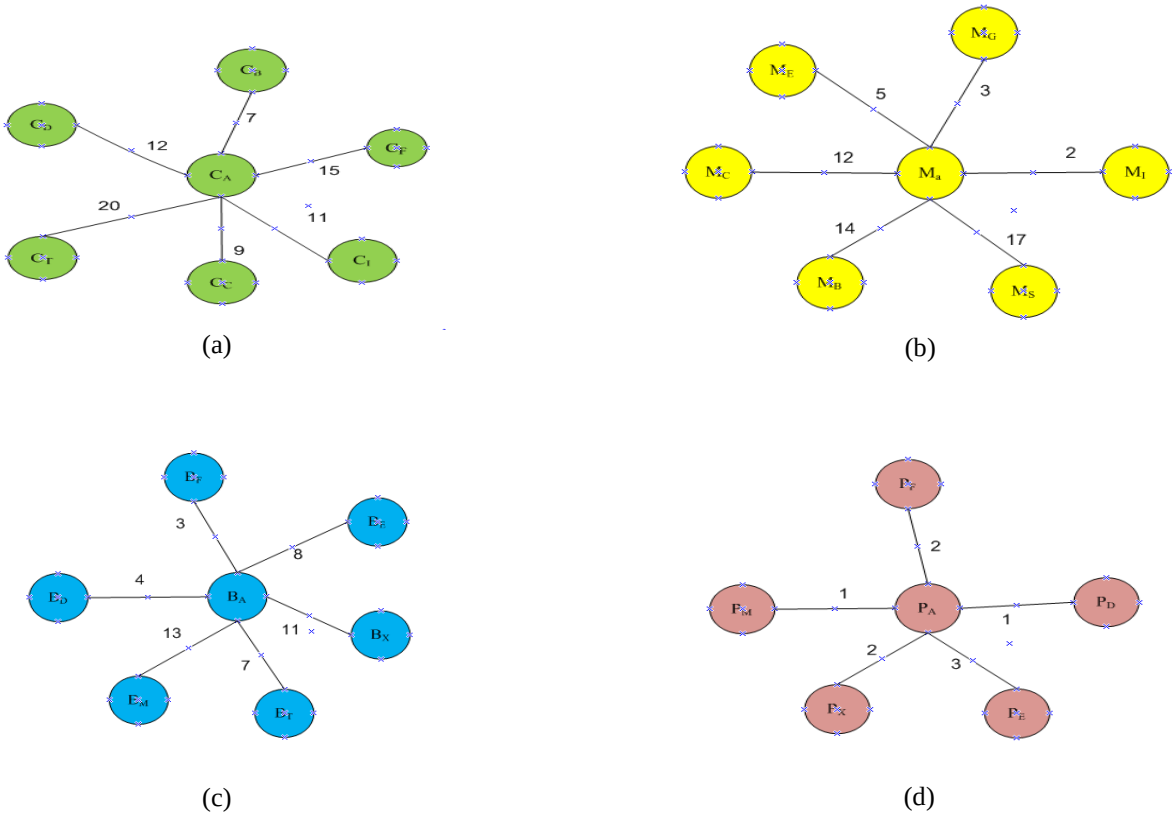
Nowadays people have adopted the latest GPS-enabled services in order to enable non-direct interaction among the communities. This leads to increase the active members of the online social networks (Facebook, Twitter, LinkedIn etc.). Many of them common for more than one community, this reflects overlapping communities (OC) comes into the picture. Some properties of Overlapping Communities as follows

- Multi-communities can share common set of actors.
- Actors must obey the properties of those communities.
- The behaviour of actors in the network depends on community

As shown in the Fig.2 (a) depicts the voice communication in centralized environment .(b) depicts message passing in centralized system.(c) links various neighboring nodes via customer accounts.(d)links various neighboring nodes via customer pan identities. In all the above four cases, edge weights are evolved based on frequency of communication at a particular instant with immediate neighbors.

The voice communication in centralized environment as shown in Fig.2 (a) is depicted using an undirected graph comprising one central vertex C_A and six immediate neighboring vertices C_B C_C C_D C_F C_I C_T with their corresponding edges whose weights numbered as 7,9,12,15,11,20. As actor in social network may have more than one subscription, in such case more than one central node exists.

The message passing in centralized environment as shown in Fig.2 (b) is depicted in a similar fashion using an undirected graph comprising one central vertex M_A and six immediate neighboring vertices M_B,M_C,M_E, M_G,M_I,M_S with their corresponding edges whose weights numbered as



(a) (b)

(c) (d)

Fig.2

14, 12, 5, 3,2,17. The message that is communicated in the network may be text/audio/video message.

In Fig.2(c), each individual is identified in the network based on their account number the network comprises with vertices B_D B_E B_F B_M B_T B_X and edges labeled as 4, 8,3,13,7,11. There are also exceptions that each individual may have more than one account number.

A similar graph is obtained with the pan identities representing each individual as vertex vertices P_D P_E P_F P_M P_X with and the corresponding edges 1, 3,2,1,2

ALGORITHM

Today every transaction is linked to unique identify number, so we can start with one actor/person in the society.

1. Consider one actor/person aadhaar no and find out respective bank account number, cell no, PAN no and email ids etc.
2. Draw a weighted graph of each Bank transactions, Cell communications (Text/Voice), PAN related (assert transaction) and email communication etc based on their communication with in a particular time of interval, here weight of an edge is depended on how many times communication happened in between persons in respective communication networks.
3. Finally we got four types of networks (graphs) Bank transaction, Cell communication, PAN based transaction and message/Email transactions, merge these networks into a single network, which consists of all unique nodes in four networks and sum of existed edge weights between pair of persons.
4. For neighbors of initial people, repeat step 2 until required no of levels.

In the next phase, all the four communities as defined in Fig. 2 are merged in to single network having all their unique vertices representing the aadhar ID's and edges whose value is taken by summing up the edge weights corresponding to similar neighbors in all the four groups.

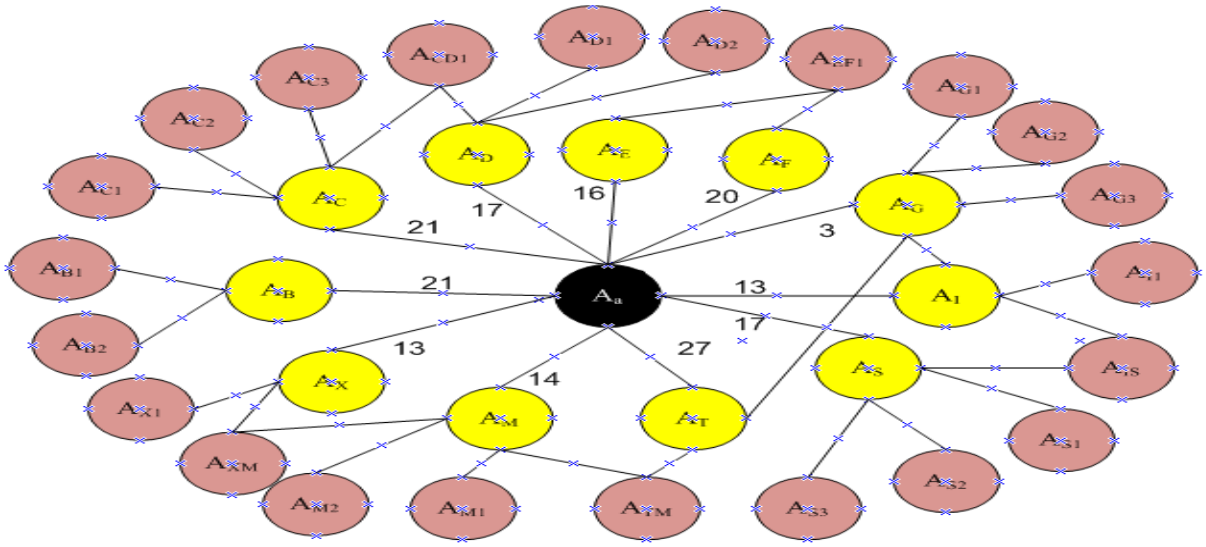


Fig.3

As observed in Fig.3, the central position of the node in the network determines the most influential contact . The degree of its influence in the social network is assessed using centrality measures as mentioned in ….. Degree centrality simply gives number of immediate contacts an actor has in the network. Some of the formulae are mentioned below depicts the actor's level of involvement in activity of the network.

Degree centrality

Formula for degree centrality, for actor i:

$$C_D(i) = \sum_{j=1}^{n} X_{IJ}$$

Where , X_{ij}= the value of the tie from actor i to actor j. Thus, it is the sum of all ties.

n= the number of actors in the network.

Betweenness centrality

Formula for betweenness centrality:

$$C_B(k) = \sum \frac{\partial_{ikj}}{\partial_{ij}} \, , i \neq j \neq k$$

$\partial_{ikj=}$ the number of geodesics linking actors I and j that pass through actor k.

$\partial_{ij=}$ the number of geodesics linking actors i and j

Eigenvector centrality

Rather than a formula for eigenvector centrality, one is making use of an algorithm to search for the largest eigenvalue of an adjacency matrix .

Thus, $C_E(i)$= the eigenvector centrality for actor I, which is the i^{th} entry of the unit eigenvector e. Hence, e refers to the largest eigevalue of the adjacency matrix.

Closeness centrality

Formula for closeness centrality, for actor i:

$$C_C(i) = \sum_{j=1}^{n} d_{ij}$$

Where d_{ij} = the distance connecting actor i to actor j.

Conclusion

In this paper, the concepts of social networks, offline community structure which is considered as one of the most important features of social networks and necessity of detecting offline communities were introduced. Then some of the main criteria for evaluating identified offline communities in the social networks were noted followed by the introduction of main challenges in community detection and mining. Beside introduction of social networking, a classification for offline community detection approaches based on the nature and type of the social networks have been proposed. This classification includes offline community detection approaches in Aadhaar based communication networks, GPS based mobile networks and GOOGLE Time Line based networks. Based on the proposed classification, the status, the advantages and disadvantages of the methods were addressed based on their application. In conclusion, providing a classification for the sustainability of the proposed approaches in terms of the nature and use of social networks as an achievement of this study is noteworthy.

References

[1]. Aadhaar Authentication API 1.2, <http://uidai.gov.in/images/ FrontPageUpdates/aadhaar_authentication_api_1.2_4dec.pdf> 2010. [online] accessed on Aug 28, 2011.

[2]. Aadhaar Handbook for Registrar, <http://uidai.gov.in/images/FrontPageUpdates/aadhaarhandbookver1.2.pdf >, 2010. [online] accessed on Aug 27,2011.

[3]. Biometrics Design Standards For UID Applications, <http://uidai.gov.in/UID_PDF/Committees/Biometrics_Standards_Committee_report.pdf> 2009.[online] accessed on Aug 26, 2011.

[4]. Envisioning A role for Aadhaar in Public Distribution System, <http://uidai.gov.in/UID_PDF/Working_Papers/Circulated_Aadhaar_PDS_Note.pdf>2010. online] accessed on Aug 28, 2011.178

[5]. Kanakia , Hemant, Srikanth Nadhamuni and Sanjay Sarma, “A UID Numbering Scheme”. 2010. [online] accessed on Aug 26, 2011.

[6]. S. Fortunato, "Community detection in graphs," Phys. Rep, Elsevier.Italy, vol. 486, pp. 75-174, December 2009.

[7] M.E.J Newman,”The structure & function of complex network”. Vol 2, pp 167-256, 2003.

[8]Maryam pourkazemi, Mohammadreza k “ A survey on community detection methods based on the nature of social networks”,ICCKE 2013.

[9] Bo Shen*, Ningwei Wang, Huihuai Qiu “A New Genetic Algorithm for Overlapping Community Detection”,2014

[10] J. Qiu, ZH. Lin, CH. Tang and SH. Qiao, "Discovering Organizational Structure in Dynamic Social Network," IEEE. USA, pp.932-937, December 2009 [the 9th IEEE International Conference on Data Mining]

[11] M. Takaffoli,F . sangi, J. Fagnan and O. Zaiane,” A Framework for Analyzing Dynamic Social Networks,” ASNA, PP.1-15,2010

[12] S. Asur, S. Parthasarathy and D. Ucar, "An event-based framework for characterizing the evolutionary behavior of interaction graphs," Trans. Knowl. Discov. Data, ACM. USA, vol. 3, pp. 16, November 2009.

[13] D. Chakrabarti, R. Kumar and A. Tomkins, "Evolutionary Clustering," ACM, pp. 554-560, [Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining]

[14] D. Cai, ZH. Shao, X. He, X. Yan and 1. Han, "Mining Hidden Community in Heterogeneous Social Networks," ACM. USA, PP: 58-978-1-4799-2093-/13/$31.00 ©201 3 IEEE 65, 2005 [Proceedings of the 3rd international workshop on Link discovery]

[15]] B. Yang, K. Cheung and J. Liu, "Community Mining from Signed Social Networks," IEEE Trans. Knowl. Data engineering. USA, vol. 19,pp. 1333-1348, October 2007.