A Comparative Analysis of Different Social Network Parameters Derived from Facebook Profiles

Paramita Dey and Sarbani Roy

Abstract In social network analysis (SNA), using online social media, it is possible to collect large open source information and to analyze those data for knowing the characteristics of these networks. The main objective of this work is to study online social network parameters commonly used to explain social structures. In this paper, we have extracted data from the three real-time facebook accounts using Netvizz application. Gephi, a open source free software, is used for analysis and evaluation of these network parameters. This analysis shows some well-known network parameters like calculating clustering coefficient (CC) of clusters, group formation, finding node degree distribution (NDD), identifying influential node etc., which can be used for further feature extraction.

Keywords Social network analysis · Facebook profiles · Social network parameters · Netvizz · Gephi

1 Introduction

Recently, with the advent of online social networks (OSNs), there is a boom in social information [1]. In December 2014, the monthly active users of facebook, a popular social network service reach the value over 1.39 billion (source: www.facebook.com). But the main drawback of using these data is the huge size of the network. These data are mainly stored in the form of graph, analysis of which required large time and computation overhead. The main objective of this paper is to study the dynamics of different parameters involved in social network analytics. We have considered three real-time Facebook profiles. Some selected network

P. Dey
Department of Information Technology, GCECT, Kolkata, India

S. Roy (✉)
Department of Computer Science & Engineering, Jadavpur University, Kolkata, India
e-mail: sarbani.roy@cse.jdvu.ac.in

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_13

parameters are derived from these profiles for analysis. A comparative analysis is also presented here. Filtering approaches are used to reduce the data set but without effecting basic properties like node degree distribution (NDD) and clustering coefficient (CC).

Different social network parameters relevant to this paper are briefly discussed in Sect. 2. Results derived from real-time Facebook accounts are presented in Sect. 3 and a comparative analysis is done in perspective of social networking. In Sect. 4, the analysis is concluded and future scope of this work is reviewed.

2 Social Network Parameters

Social network analysis (SNA) is the qualitative and quantitative measuring technique to find the relationships among different social community like people, groups, organizations, etc. Some useful network parameters are discussed in the Table 1.

3 Analysis and Comparative Study

For the analysis purposes, three Facebook profiles were used to extract the data using Netvizz [2]. These three profiles are three disjoint data sets as they are not in the friend list of each other. From the perspective of social network, it is informative to study their interactions within the network and find the implication of these results [3]. A partial view of data table of profile 1 is shown in Fig. 1. After importing the data tables from Netvizz (this application is inbuilt within facebook profile), an initial hairball like network can be formed using Gephi [4], where each node represents an individual user and each edge represents the communication between them. Though from these networks, no direct information can be derived but an initial idea about the profile can be made. From the following profiles, it can be seen that the profile 1 consists of small number of edges which signifies the users within these group shares less communication within them. On the other hand, profile 2 and profile 3 have almost similar initial network due to the size of network in terms of nodes (profile 2: 215 #nodes; profile 3: 234 #nodes) and edges (profile 2: 3114 #edges; profile 3: 3310 #edges). A comparative view of these three initial networks is given in Fig. 2. Different network layouts are available in Gephi software. Among them, for our simulation, we have used Force Atlas 2 model [5] and Fruchterman-Reingold model [6]. Force Atlas 2 use different techniques such as degree-dependent repulsive force, Barnes Hut simulation, and adaptive temperatures for their simulation purposes. The main idea of simulation is that the nodes repulse and the edges attract. It is a continuous force directed layout. Network layouts using Force Atlas 2 with dissuade hubs mode of three profiles is shown in Fig. 3a. Dissuade hubs prefer authorities (nodes with high in-degree) in

Table 1 Network parameters

Social network parameters	Definition
Radius	Minimum path between any two nodes of the network, represented as *rad*(*G*) [7]
Diameter	Maximum path between any two nodes of the network, represented as *diam*(*G*) [7]
Shortest path	The minimum distance between any two nodes is the shortest path between two nodes [7]
Average path length	Arithmetic mean distance among all possible shortest path between any two nodes of the network signifies the rank of the network [7]
Node degree distribution	For a directed graph, *in degree* denotes the number of edges ending at that node and *out-degree* denotes the number of edges beginning at that node [8]
Rank	Rank counts the the number and quality of a links connected to that node [9]
Node degree centrality	Node degree centrality depends on the node degree distribution (i.e., in degree and out degree) of individual node. The node with maximum node degree represents the maximum centrality [10]
Betweenness Centrality	It denotes the number of shortest path passes through a node. The nodes with high betweenness centrality implies maximum connectedness in the network and vulnerability of the network is dependent on that nodes [10]
Closeness centrality	It denotes the average shortest path of a node with other nodes [10]
Eigen vector centrality	This centrality measures is based on Eigen vector matrices [10]
Community detection	It is the parameters of a network to classify the nodes into separate groups according to some properties. In social network analysis, *mutuality*, *reachability*, *vertex degree* and internal versus external *cohesion* are the four properties, which are used for Community detection [11, 12]
Clustering coefficient	The measurements of average distance of connected clustering nodes in a graph is called the clustering coefficient (CC) [10]

the center than hubs (nodes with higher out-degree). This system pushes the hubs in the periphery of the network. Force Atlas 2 network layout of these profiles with LinLog mode is shown in Fig. 3b. In this mode, LinLog energy model is used to make the cluster more dense but the convergence time with this model is much higher. According to Fruchterman-Reingold model [6], continuous network modeling was done depending on even distribution of the vertices in the frame, making edge lengths uniform and reflects inherent symmetry. Network layout of these profiles using Fruchterman-Reingold model are shown in Fig. 3c.

In profile 1, numbers of clusters are more but with less density which can be seen in two modes (dissuade hubs and LinLog) of Force Atlas 2. In profile 2, clusters are evenly distributed in nature. It is quite likely that there are better communications

1	Id	Label	sex	locale	agerank
2	10131316	Anthony Fernandes	male	en_US	69
3	5.4E+08	Atasi Deb Ray	female	en_US	68
4	5.6E+08	Raktim Ghosal	male	en_US	67
5	5.66E+08	Swaty Mitra	female	en_US	66
6	5.71E+08	Sohini Dasgupta	female	en_US	65
7	6.65E+08	Prabal Bagchi	male	en_US	64
8	6.71E+08	Urmi Choudhury	female	en_US	63
9	6.72E+08	Ram Rup Sarkar	male	en_US	62
10	6.96E+08	Madhumita Barua	female	en_US	61
11	7.01E+08	Alok Sharma	male	en_US	60
12	7.7E+08	Ranjan Som	male	en_US	59
13	8.22E+08	Sanjit Kumar Das		en_US	58
14	8.24E+08	Dipak Chatterjee	male	en_US	57
15	1.02E+09	Tapas RayMahapatra	male	en_US	56
16	1.06E+09	Gourab Ghosh	male	en_US	55
17	1.08E+09	Som Subhra Chakraborty	male	en_US	54
18	1.15E+09	Poulomi Chakraborty	female	en_US	53
19	1.18E+09	Samrat Laskar	male	en_US	52
20	1.27E+09	Sudip Chakraborty	male	en_US	51

Fig. 1 Partial data table derived from facebook using Netvizz

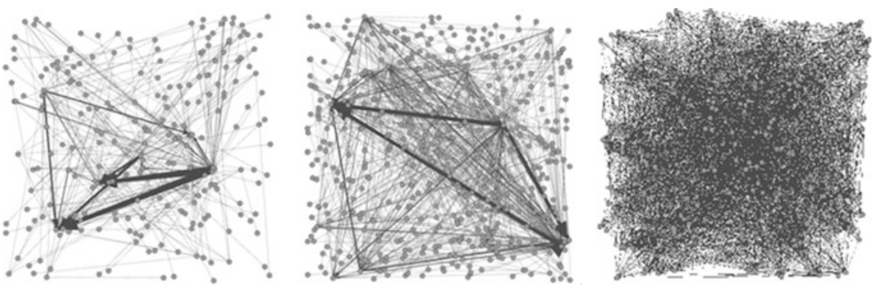

Fig. 2 Initial network formation from three facebook profiles

within the members of these profiles and the network is very stable. On the other hand, in profile 3, there are only two distinguishable clusters apart from very small clusters or isolated nodes. These clusters are very closely connected (as reflected from LinLog mode), which imply there is a regular communication between these dense clusters whereas it is less on the other nodes or clusters. In all three profiles, some of the links are very dense than the other which imply these nodes that share these links have a large betweenness, that means these nodes are very influential in that network. Node degree distribution (NDD) is an important networking property. NDD can be improved by reducing less important nodes and edges. This is the

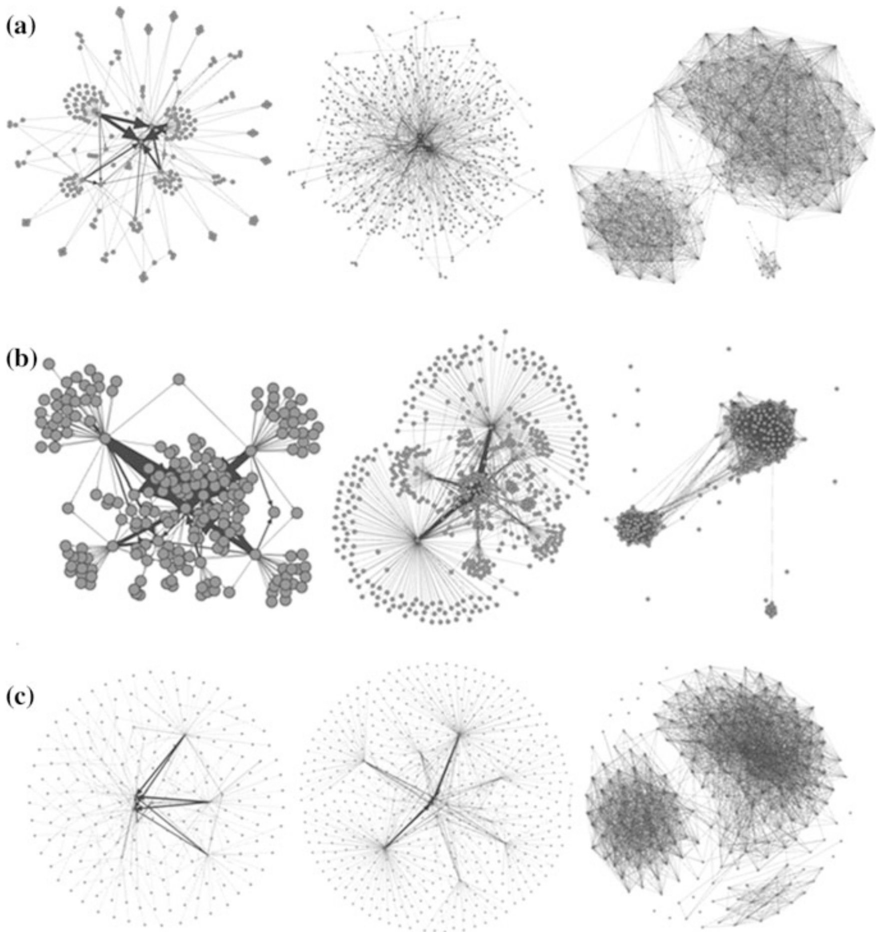

Fig. 3 Network formation of three profiles using different techniques. **a** Dissuade hubs mode of force Atlas 2. **b** LinLog mode of force Atlas 2. **c** Fruchterman-Reingold mode

basic process of filtering. Rank denotes the number of connections of each node. The nodes with maximum ranks can form the most significant network. In Fruchterman-Reingold distribution, the nodes with maximum nodes are connected with darker edges. Maximum eccentricity is the diameter and minimum eccentricity is the radius of a network. From the simulation result of eccentricity distribution and centrality distribution of the facebook profiles different parameters like radius, diameter, average path length, number of shortest paths can be calculated.

In social networking terms, centrality defines how fast the information can be spread. The distribution for closeness centrality distribution is shown in Fig. 4.

Clustering coefficient (CC) defines the centers of different communities. Community distribution based on the sex of these three profiles is shown in Fig. 5.

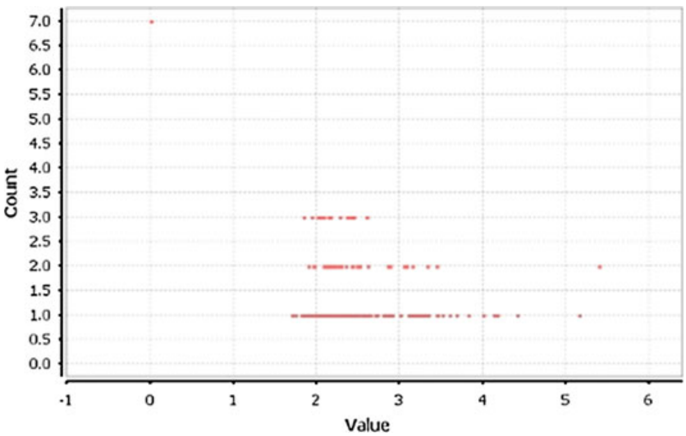

Fig. 4 Closeness centrality distribution of profile 2

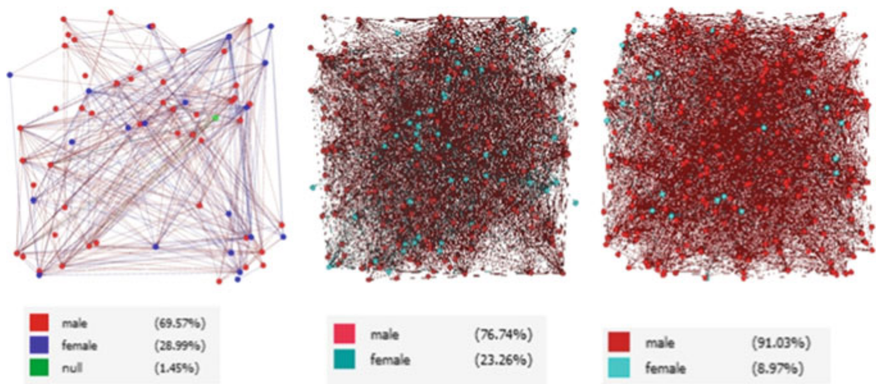

Fig. 5 Male–female percentage distribution of three profiles

It can also be done for other types of community distribution. Summarize all the properties derived from above distributions, it is shown in the Table 2. Though a number of nodes in all three profiles are almost same but in the first profile number of edges are significantly less. It signifies the cohesion between all the nodes are relatively less. It also reflects in density parameters which is only 1 % compare to 13 and 12 % of other two profiles. Diameter of profile 3 is maximum, refers that any two nodes of that profile attached through a distant communication. So, profile 3 plays a powerful role in the network. Average clustering coefficient in profile 2 and profile 3 is around 0.5 whereas for profile 1, it is less than 0.01, which signifies clusters of this profile are less connected compare to other two profiles. In this work, we have considered three profiles of male users. Here one observation is that female community in these profiles in respect to male community is very less.

Table 2 Comparative study of three facebook profiles

Matrices	Profile 1	Profile 2	Profile 3
Number of nodes	259	215	234
Number of edges	357	3114	3310
Diameter	6	7	9
Radius	1	0	0
Number of communities	21	17	13
Density	0.012	0.135	0.121
Average weighted degree	3.202	28.967	28.291
Average clustering coefficient	0	0.561	0.592
Average path length (after clustering)	3.492	2.538	2.794
Number of shortest path (after clustering)	47,326	43,056	49,062

4 Conclusion and Future Scope

The main motivation of this work is to extract the well-known parameters of the social network, make an analysis and make a comparative study between all the profiles. As we are using open-source tools like GEPHI software, which can use external modules like JAVA net-beans as add-on, these data can be classified further for more advanced feature extraction. Extracted information can be used for designing graph sampling algorithms and game theory-based social network designing. In future, we will try to design an interactive model for finding a relation between an individual with population to measure the influence of that person in the society.

Acknowledgments We would like to thank Subrata Dey, Aniket Sinha and Dr. Krishnendu Dutta for sharing their Facebook profiles with us for data collection purposes.

References

1. Benevenuto, F., Rodrigues, F., Cha, M., Almeida.: Characterizing user behavior in online social networks. In: Proceedings of ACM IMC (2009)
2. Rieder, B.: Studying facebook via data extraction: the Netvizz application. In: WebSci, Paris, France, ACM 978-1-4503-1889-1, May 24 2013
3. Wilson, R.E., Gosling, S.D., Graham, L.T.: A review of facebook research in the social sciences perspectives. Psychol. Sci. **7**(3), 203–220 (2012)
4. Gephi Official Website: https://gephi.github.io
5. Jacomy, M., Heymann, S., Venturini, T., Bastian, M.: ForceAtlas2, a continuous graph layout algorithm for handy network visualization. PloS One **9**(6), (2010)
6. Fruchterman, T.M.J., Reingold, E.M.: Graph drawing by force-directed placement. Softw: Pract. Expert **21**(11), 1129–1164 (1991)
7. Wayne, G., Oellermann, R.O.: Distance in Graphs, Structural Analysis of Complex Networks, pp. 49–72. Springer (2011)

8. Wang, T., Chen, Y., Zhang, Xu, T., Jin, L., Hui, P., Deng, B., Li, X.: Understanding graph sampling algorithms for social network analysis. In: Simplex, IEEE ICDCS, Minneapolis, USA, pp. 123–128 June 20–24 2011
9. Kleinberg, J.: Authoritative sources in a hyperlinked environment. In: Proceedings of 9th ACM-SIAM Symposium on Discrete Algorithms (1998)
10. Freeman, L.C.: Centrality in social networks: conceptual clarification. Soc. Netw. **1**, 215–239 (1978/1979)
11. Ronald, S.B.: Social contagion and innovation: Cohesion versus structural equivalence. Am. J. Sociol. **92**(6), 1287–1335 (1987)
12. Fortunato, S., Castellano, C.: Community structure in graphs. Phys. Rep. (2007)