

A Review on Resemblance of User Profiles in Social Networks using Similarity Measures

Nidhi Goyal^{a*}, Jaswinder Singh^b

^aM.Tech Student CSE, GJUS&T, Hisar (Haryana) -125001,India

^bAssistant Prof. CSE, GJUS&T, Hisar (Haryana) -125001,India

^aEmail: nidhigoyal1993@gmail.com

^bEmail: jaswinder_singh_2k@rediffmail.com

Abstract

Online Social Networking is increasing at a fast rate. There are lots of profiles of the users and there is too much resemblance between the user profiles which can help recruiter's to select the best candidates for the Job Profile. Now, each similarity measure has its own applicability and best suited to a particular type of attribute values and if these measures are collectively combined then it can help us to find the best resemblance among the user profile, the result of which matches to the actual result. In this paper, the discussion of the past studies is done and how our research is proposing a framework for finding the resemblance is being discussed.

Keywords: Best Resemblance; Heterogeneous Similarity measures; user profiles.

1. Introduction

As the users of Social networking sites are increasing at the alarming rate. Online Social Networking has become popular among all the age groups, especially among the School and the College students that it has knitted the people all over the world. It does not matter to which class a person belongs to or in which field the person is working whether Researchers, Industrialists, Celebrities, Government officials, Politicians, Entrepreneurs, Sportsperson etc. Social media and networking has connected all of them. It seems that Social Networking Sites have hypnotized the people. Social Networking could be a boon for an individual leading him to grow faster than others or could be a bane if being addicted to it. The need of the Social Sites is also being concluded by Keith Ferrazi. As the author has rightly said in [17].

* Corresponding author.

"Poverty, I realized, wasn't only a lack of financial resources; it was isolation from the kind of people that could help you make more of yourself."

A platform is being served by Social Media and with the advent of technology, the number of Social websites users is increasing tremendously. Social Networking websites gave the user a base for sharing, expressing, interacting and for cultivating relationships. Figure 1 shows the connectivity of users on the Social Network. The Users from various countries interact with each other using social networking sites. The top 15 Most Popular Social Networking Sites are described in [22]. A person can have multiple profiles over multiple networks or on the same network. There are lot of user profiles which are resembling with each other that can help the recruiter's to select the most promising and convincing candidate. Social Networking websites like Facebook has most of the same profiles. Duplicate user profiles or fake user profiles [5] can be detected using the various similarity measures like Overlap, Dice, Cosine ,Simpson [6,16]. User can have many accounts with the same name and other details or two or more accounts with different names or other information could be related to each other as being accessed and handled by the same user or organization. Users share their information like Date of birth, favorites like music, movies, dances, videos etc.

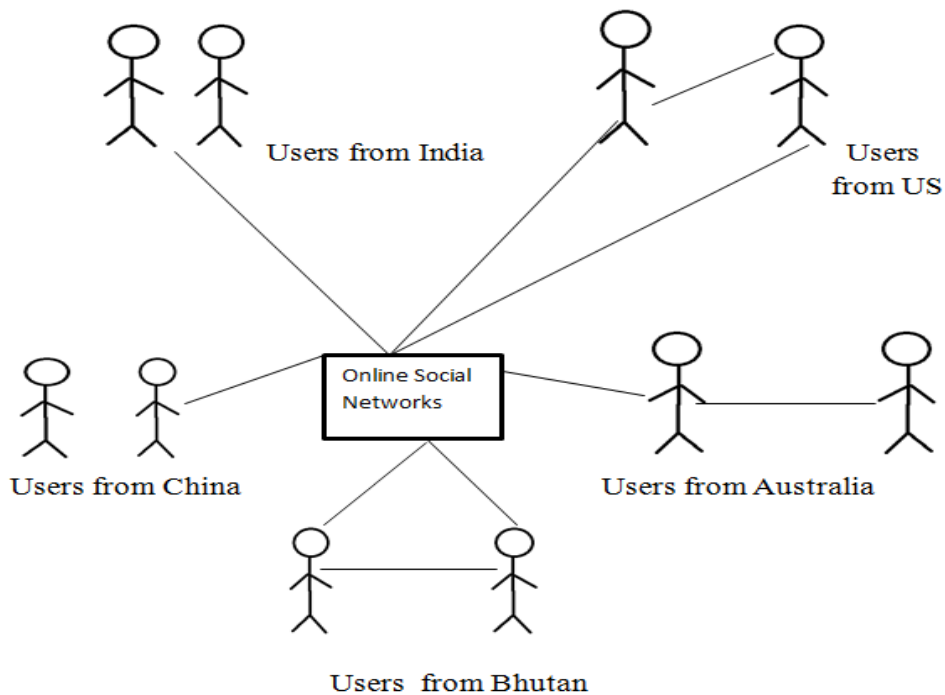


Figure 1: Interaction among people all over the world using social networking website.

2. Related work

There is lots of research works carried out in context of matching the user profiles. Some of the findings are worth mentioned here below:

Anshu Malhotra and his colleagues [1] proposed use of automated classifiers for classifying the user profiles. Bhumiratana [3] demonstrated a model for Automating Persistent Identity Clone in Online Social Networks for

exploiting availability weak trust in social networks. Akcora and his colleagues [4] proposed a network similarity measure that considers only the graph structure and takes into consideration also how two users are indirectly connected. Secondly, a similarity measure based on user profile information, such to find semantic similarities between users.

Raad and his colleagues [8] worked upon the FOAF attributes in User Profile Matching in the Social Networks. There can be many profiles that refer to the same person. In order to find those profiles various similarity measures have been used. The four Components used in the work are Profile Generator, Profile Retriever, Weight Assignment, Profile Matcher. Spertus and his colleagues [9] studied similarity measures relative to each other in a large real-world environment. This paper presented an extensive empirical comparison of six distinct measures of similarity for recommending online communities to members of the Orkut Social Network. Jaswinder and his colleagues [10] discussed the different similarity measures that are used in information retrieval. Reference [11] elaborated a new and precise method to detect profile cloning in online social networks.

In the method, the social network is shown, then the matching profiles to the real profile are gathered from the same network, then strength of relationship (among all selected profiles and the real profile) is calculated, and those which have the less strength of relationship will be verified by mutual friend system. Kontaxis and his colleagues [14] contributed for designing of the architecture (Information Distiller, Profile Hunter and Profile Verifier) and implemented a tool to detect cloned profiles in a Linked-In network. The limitation of this system is that it basically used the LinkedIn social network. In this implementation exact string matches are taken in view by the Profile Verifier. Here, Fuzzy String matching could be used. Jin and his colleagues [15] demonstrated an active detection framework for detection of cloned profiles. Profile similarity and multiple-faked identities profile similarity. Three step process is being proposed in which first step is to search and separate identities as a set of profiles. Second step is detection of suspicious profiles using profile similarity measures. Third is detection of looking alike or cloned profiles using the list of the friends. This whole of the detection model detects existing faked identities but cannot defend against ICAs in future. Q.Xu and his colleagues [20] described the various semantic similarity models for measurement of similarity. S.Soundarajan and his colleagues [21] demonstrated through empirical study about which network similarity measure to choose.

Reference [23] measured the similarity on the basis of professional, social, geographical, educational, shared interests, pages liked in a social network. The identification of the connection between two user profiles and their closeness level is explained. Profile similarity is being discussed step by step and finally computed similarity score on the basis of the String Similarity metrics.

The assignment of the weight of the attributes by using binary weight assignment algorithm but is not done according to their rank. Since some attributes are more important than the other attributes hence some algorithm must be used to find best resemblance. Reference [24] discussed about the various text based similarity approaches. The discussion of the existing works on text similarity through partitioning them into three approaches; String-based, Corpus-based and Knowledge-based similarities is being done.

3. Organised review

Literature review has been organized in form of a table for making the study more and more simpler.

Table 1: Comprehensive and Systematic Review Table

Work Title	Author	Year	Methods used and work done
Prevention of Fake Profile Proliferation in Online Social Networks[18]	S.Priyanga et al.	2015	Identified original users for creating another profile with same data existing in the server in case of them losing their password for their access.
Similarity Measure for Social Networks – A Brief Survey[2]	Ahmad Rawashdeh and Anca L.Ralescu	2015	SimRank, PageSim, SimFusion, P-Rank, Vertex Similarity, FriendTNS, E-Rank, Improved SimRank
User Profile relationships using string similarity metrics in Social networks[23]	V.A.Dabeeru	2014	1.Jaro Metric Similarity 2.Cosine Similarity
An IAC Approach for Detecting Profile Cloning in Online Social Networks [11]	Kharaji and Rizzi	2014	1.Discovering community the social network graph 2.Computing strength of relationship
Detecting Cloning Attack in Social Networks Using Classification and Clustering Techniques[13]	Kiruthiga.S et al.	2014	1.Jaccard Similarity 2.Cosine Similarity 3.Naive Bayes Classifier 4.K-Means Clustering
An approach for detecting profile cloning in online social networks[12]	M.R.Khayyambashi and F. S. Rizzi	2013	1.Profile Similarity 2.Friend's Network Similarity
@I seek 'fb.me': Identifying Users across Multiple Online Social Networks[19]	Paridi Jain et al.	2013	1.Content Search 2.Self –Mention Search 3.Network Search 4.Application of Identity search algorithms by evaluating Accuracy and

Towards Active Detection of Identity Clone Attacks on Online Social Networks[15]	Jin et al.	2011	precision 1. To search and separate identities as a set of profiles. 2. Detection of suspicious profiles by using profile similarity schemas. 3. Deleting cloned profiles from friend list
'A Model for Automating Persistent Identity Clone in Online Social Network.'[3]	Bhume Bhumiratana	2011	Exploited Weak trust in Social networks
Detecting Social Network Profile Cloning[14]	Kontaxis et al.	2011	Detected Cloned profiles in Linked-In Network
Network and Profile Based Measures for User Similarities on Social Networks [4]	Cuneyt Gurcan Akcora et al.	2011	Network Based Profile Based
User profile matching in social networks[8]	Raad et al.	2010	Generation of the user profiles Retrieval of the profiles Weight assignment of those profiles attributes Matching of the profiles

4. Proposed framework

This Framework effectively merges the various similarity measures which are useful in designing of the Hybrid Binary Weight Assignment algorithm. First of all, the user profiles are being extracted from the social media platform i.e. Facebook using the Graph API Explorer. From these user profiles, various profile attributes are being extracted. The profile attributes which are being extracted are of different category. Based on their category type whether numeric or string or character type, different type of similarity measures are being applied. Different type of similarity measures best suited to specific type of the value likewise Jaro Similarity measure is best suited to the short strings and hamming distance well works on dichotomous values and TF-IDF [7] cosine similarity for weighted similarity measurement. After application there is formation of the incidence matrix from the result of the Hybrid Binary Weight Assignment algorithm that consists of the rows of the target user profiles which are to be compared with the source user profile and columns which has the attributes in the user profile. Rank Order Clustering algorithm concept utilization is to be used in the calculation of the weighted factor. This algorithm assigns the weighted factor according to the rank of the attributes. Similarity Scores are being calculated using cosine Similarity and with the usage of the weighted factor adjusted Similarity Score can

be calculated. The result of this hybrid binary weight assignment algorithm can be compared with traditional binary weight assignment algorithm [23]. This algorithm produce correct and better results matching to the real result. Figure2 shows the Framework for Similarity measurement of the user profiles. This framework can be helpful for job recruiter's to select the best candidate for the particular job profile.

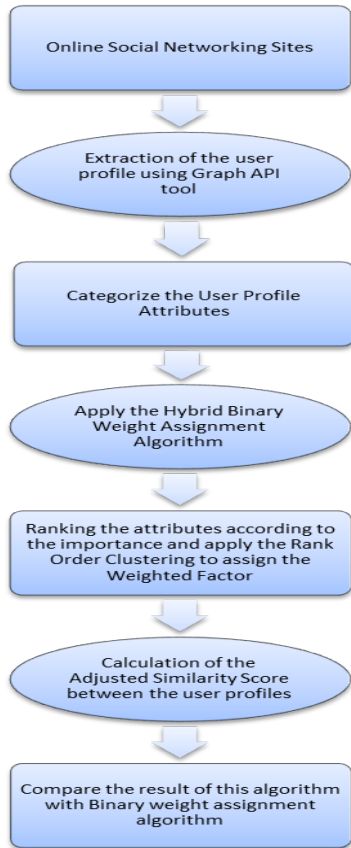


Figure 2: Framework for Similarity measurement between user profiles

5. Conclusion

After studying the various researches done for finding resemblance between the various user profiles using similarity measures for Social Networks, it has been seen that the different measures are not combined collectively to find the best resemblance of the various profiles across social networks and analytical perspective of those profiles attributes effectively need to be worked considering all the parameters of similarity measurement like rank of the attributes of the user profile. This research can be helpful in duplicate profile detection and researches being done in the field of Social Networks. Our concern is to find the best resemblance of the User profiles by using Binary Hybrid Weight assignment algorithm which will generate a incidence matrix using hybrid similarity measures for weight assignment and then assigning the weights according to the importance to the attributes by Rank-Order clustering algorithm .

References

- [1] A. Malhotra, L. Totti, W. Meira Jr., P. Kumaraguru, and V. Almeida, "Studying User Footprints in Different Online Social Networks," Jan. 2013.
- [2] A. R. Ahmad Rawashdeh, "Similarity Measure for Social Networks – A Brief Survey," in CEUR Workshop Proceedings, At Greensboro,NC,USA, 2015, vol. 1353.
- [3] Bhume Bhumiratana. "A Model for Automating Persistent Identity Clone in Online Social Network," 681–86. IEEE, 2011. doi:10.1109/TrustCom.2011.87.
- [4] C. G. Akcora, B. Carminati, and E. Ferrari, "Network and profile based measures for user similarities on social networks," in 2011 IEEE International Conference on Information Reuse and Integration (IRI), 2011, pp. 292–298.
- [5] Conti, Mauro, Radha Poovendran, and Marco Secchiero. "FakeBook: Detecting Fake Profiles in On-Line Social Networks." in Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), 1071–1078. ASONAM '12. Washington, DC, USA: IEEE Computer Society, 2012. doi:10.1109/ASONAM.2012.185.
- [6] D. Bollegala, Y. Matsuo, and M. Ishizuka, "A Web Search Engine-Based Approach to Measure Semantic Similarity between Words," IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 7, pp. 977–990, Jul. 2011.
- [7] E. Cohen, D. Delling, F. Fuchs, A. V. Goldberg, M. Goldszmidt, and R. F. Werneck, "Scalable Similarity Estimation in Social Networks: Closeness, Node Labels, and Random Edge Lengths," in Proceedings of the First ACM Conference on Online Social Networks, New York, NY, USA, 2013, pp. 131–142.
- [8] E. Raad, R. Chbeir, and A. Dipanda, "User profile matching in social networks," in Network-Based Information Systems (NBIS), Japan, 2010, pp. 297–304.
- [9] E. Spertus, M. Sahami, and O. Buyukkokten, "Evaluating Similarity Measures: A Large-scale Study in the Orkut Social Network," in Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, New York, NY, USA, 2005, pp. 678–684.
- [10] Jaswinder Singh, Parvinder Singh, Yogesh Chaba, "Performance Modeling of Information Retrieval Techniques Using Similarity Functions in Wide Area Networks," in International Journal of Advanced Research in Computer Science and Software Engineering, vol.4, pp.786-793, 2014.
- [11] Kharaji, Morteza Yousefi, and Fatemeh Salehi Rizi. "An IAC Approach for Detecting Profile Cloning in Online Social Networks." International Journal of Network Security & Its Applications 6, no. 1 (January 31, 2014): 75–90. doi:10.5121/ijnsa.2014.6107.
- [12] Khayyambashi, Mohammad Reza, and Fatemeh Salehi Rizi. "An Approach for Detecting Profile Cloning in Online Social Networks." In E-Commerce in Developing Countries: With Focus on E-Security (ECDC), 2013 7th International Conference on, 1–12. IEEE, 2013.
- [13] Kiruthiga S., Kola Sujatha P., and Kannan A., "Detecting cloning attack in Social Networks using classification and clustering techniques," in 2014 International Conference on Recent Trends in Information Technology (ICRTIT), 2014, pp. 1–6.
- [14] Kontaxis, Georgios, Iasonas Polakis, Sotiris Ioannidis, and Evangelos P. Markatos. "Detecting Social

- Network Profile Cloning.” In *Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2011 IEEE International Conference on, 295–300. IEEE, 2011.
- [15] L. Jin, H. Takabi and J. Joshi, ” Towards Active Detection of Identity Clone Attacks on Online Social Networks”, In *Proceedings of the first ACM Conference on Data and application security and privacy*, pp. 27-38, 2011.
- [16] M. S. Hamani and R. Maamri, “Word Semantic Similarity Based on Document’s Title,” presented at the *Database and Expert Systems Applications (DEXA), 2013 24th International Workshop on*, 2013, pp. 43–47.
- [17] “Never Eat Alone Quotes by Keith Ferrazi.”<https://www.goodreads.com/work/quotes/81738-never-eat-alone-and-other-secrets-to-success-one-relationship-at-a-tim>.
- [18] N. Hariharan, S. Priyanga and V.M. Priyadharshini, “Prevention of Fake Profile Proliferation in Online Social Networks”, in *IJIRSET*, May 2015, vol. 4, pp. 25-32.
- [19] P. Jain, P. Kumaraguru, and A. Joshi, “@I Seek ‘Fb.Me’: Identifying Users Across Multiple Online Social Networks,” in *Proceedings of the 22Nd International Conference on World Wide Web, Republic and Canton of Geneva, Switzerland*, 2013, pp. 1259–1268.
- [20] Q. Xu and W. Shi, “A comparison of semantic similarity models in evaluating concept similarity,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS’2012)*, vol. 39, p. B2, 2012.
- [21] S. Soundarajan, T. Eliassi-Rad, and B. Gallagher, “Which network similarity measure should you choose: an empirical study,” in *Workshop on Information in Networks*, New York, USA, 2013.
- [22] “Top 15 Most Popular Social Networking Sites January 2016.” [Online]. Available: <http://www.ebizmba.com/articles/social-networking-websites>.
- [23] V.A. Dabeeru, “User Profile Relationships using String Similarity Metrics in Social Networks,” Aug. 2014.
- [24] W. H. Gomaa and A. A. Fahmy, “A survey of text similarity approaches,” *International Journal of Computer Applications*, vol. 68, no. 13, pp. 13–18, 2013.