International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Understanding the Users Personal Attributes Similarity Across Online Social Networks

Waseem Ahmad, Rashid Ali

Abstract*: In this modern era of technology, everyone accessing the Internet is obsessed with social media. A User accesses different social media services to fulfill his diverse needs. For instance, Instagram is mainly used for sharing personal visual content while Twitter is known for finding latest news and trends, similarly Facebook for personal posts. Such services lead to the distribution of personal information of an Internet user on these platforms. In this paper, we build a framework to discover the relationship among the attributes of a user across the social media.We use different fuzzy string matching algorithms to find the similarities between the attributes. We extract the 'name' and 'username' from a publicly shared dataset and apply two character based and token based algorithms on these features. The results are indicative of the fact that only a limited number of users share the same name and username across the sites. On further analysis, it is found that although name and username of most of the users do not exactly match, they tend to be similar with the infinitesimal difference like; underscore, period, one digit numbers, etc. This study provides an analysis of the typical variations in names and usernames, which can further be studied for the extension to other social networks This profile will help in behavior analysis of a user, which will further help us to improve recommendations and analyze for criminal behavior and similar applications.

***Keywords*: Personal Information, Social Account, User Identity, Cross link posts.**

I. INTRODUCTION

People use social media networks for different purposes, for example; a user chose Facebook for posting personal information, Twitter for consuming recent news, Instagram for posting own photograph, Pinterest for uploading and downloading preferred images. In a survey[1] it has been reported that 3.196 billion users created their profiles on social media services out of 4.021 billion Internet users. In the same study, it is estimated that the growth of social media users across the globe for the next two years will be about 13 percent per year [2]. The emergence of mobile phone minimizes the portability issue and encourages people to join different social media using mobile based social media apps.

As each social media service uses the different layout to represent the profile of a user, therefore a user shares his or her unique personal identity to differentiate him with the other users present on the same site. The main traits that a naive user shares across the sites are the name, username, email address,
location, profile pictures, home address, gender, description, etc. But each social network selects some features to represent a user profile as public attributes. For example, an email address is public on LinkedIn while protected on the other sites; similarly, gender and birth date are open on Facebook and private on the other social media. The main challenge with the different social media sites is that the same feature is represented by the different names across various social media. For example, the name attribute is represented on Twitter as 'full-name' while on Google Plus 'display-name,' and whereas on Facebook by 'name.' One more thing that social media sites make interesting is that, in "network relationship" (follower, following, friends) some of the social networks represent a user with limited attributes like 'name' and 'username' followed by 'profile pictures' while others use display name along with 'designation' or 'landmarks'. Recently, fake users started creating forged profiles on different social networks by selecting virtual attributes to imitate a particular user present on the site like 'Justin Bieber'. To trounce such challenges, social media service provider started verifying their users by sending OTP (One time password) on registered mobile number or email-Id. Therefore, an online user shares his updated contact details to these services by worrying his or her privacy. Such mechanism forces a user to share authentic personal information on social media service. Leakage of personal information from social media may be catastrophic. However, due to privacy reasons, many people do minor modification in their usernames, and they are careful in selecting their personal attributes like name, username, location, and description on each site. Finding identical users across different social networks may help to improve the business by ads, marketing using personal messages, crime detection, and its investigation by retrieving and matching user's posts, item recommendation, etc. To study the comprehensive behavior of a users, it is essential to amalgamate the fragmented features available across multiple social networks. Merging user profile data from different social media sites is not a straight forward task because of the security and privacy of the users. Before the profile attribute diffusion, it is essential to find a correlation among the common attributes available across the sites, which concludes the exact profile of a user across the different networks. As each social media service contains millions of users, many of them are involves in conversations. During an online interaction with friends, many users reveal their account relevant information like 'URLs' and 'username' on a particular site and some of them link their profile URLs (Uniform resource locator) in the profile field.

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
Waseem Ahmad*, Department of Computer Engineering, ZHCET, AMU, Aligarh, India. Email: waseemahmad2@zhcet.ac.in
Rashid Ali, Department of Computer Engineering, ZHCET, AMU, Aligarh, India. Email: rashidaliamu@rediffmail.com

Therefore, many users disclose their personal information across the social media networks and respective services makes their profile attributes public by the APIs. Esfandyari et al.[4], describes that thousands of links are shared by the users on different social networks but only two percent links are useful. In this work, we use such links to find the relationship among users using common attributes available across the two popular social networks, namely: Twitter and Facebook. To find the relationship between the common traits, we used four different fuzzy string matching algorithms (Stringdist package available on CRAN) and applied them on the two common attributes 'name' and the 'username' field open on both thesocial networks. After that, we compare the performances of each algorithm for respective characteristics and their effectiveness of a particular algorithm for the specific attributes.

In section II, we describe the related works published in user account matching across the different online social network. Section III presents the problem formulation and modeling, while section IV gives an overview of the the statistical description of data along with data gathering and cleaning procedure. Section V demonstrates the experimental results and interpretation. Finally, section VI describes the conclusion and some future direction.

II. RELATED WORKS

Social networking services persuade people's to describe more personally identifiable information (PII) in different formats. There are two major cause of personal information revelation in online social media networks; (a) popularity (b) Trust. PII generally classified into three main categories, namely; Profile based information (personal treats), content-based information (user-generated text and content) and structure-based information (interaction networks, friendship, etc.).

A. *Profile based Information*

In online social media platforms, users regularly divulge their identities like username, name, gender, birth-date, etc. In literature, authors exploited publicly accessible attributes on social media services and matched them to find identical users across the networks. Vosecky et al. [3], utilized profile attribute to detect identical profile of users across MySpace and Facebook. Then, Esfandyari et al.[4], exploits profile features available on the three popular social media networks; Twitter, Facebook, and Google+ and applied machine learning algorithms(supervised) on the common features present on these media to find user's profile on different networks. Zafarani et al. [5], suggested that a user generally selects a maximum up to four user-names and use them across the sites. It happens due to human memory limitation, exogenous, and endogenous factors. They further described that around 59% of the users selected the same username across the sites. Malhotra et al. [6] described that 'user-id' and 'name' attributes are the discriminating features for finding the user's identity across LinkedIn and Twitter. Motoyama and Verghese [7] compare the user's profile attribute to find the same or similar user's identity across Myspace and Facebook. Further, Raad et al. [8], studied social networks and compared the profile features using various similarity measures. Profile-based information retrieval from various social networks in the current scenario is a very tedious job for researchers due to users' privacy and security roped by law and enforcement agencies. In literature, the authors used vector based attribute similarity, but it requires lots of effort during data retrieval and pair-wise matching across the different social networks using different similarity measure algorithms

B. *Content-based Information*

Social media sites support the construction, curation, and exchange of online user-generated contents. Zhang et al. [9] and Goga et al. [10] study user-generated online content for user account linking across the different social networks by exploiting 'time of post,' 'user's writing style,' 'social tagging', 'geo-location', etc. In literature, people assume that a user has a behavioral property that the same user may post the same content on different social networks at different intervals of time. Li et al. [11], discusses the roles of user-generated content in different online social networks like tagging, question answering, and micro-blogging services. They further, studied the various types of information available on social networks and suggested that single point information may be efective for finding users accounts across the networks. Zhang et al. [9] and Goga et al. [10], used user-generated content like the spatio--temporal location to discover the unique user's identity shared on different social networks.

C. *Structure-based Information*

The structure of social media networks is formed by the collection of a user's identity present on the particular network (user's profile). Friends, followers, following, connection, etc. are assumed as structural (network) information. In literature, several authors used friend relationship available on the user's profile and exploited such relationships to find identical users accounts across social media sites. Structural information is categorized two classes, depending on the nature of a particular site like display name by 'username' or 'real name'. Facebook, LinkedIn, Pinterest use 'display-name' whereas Twitter, Instagram uses both 'username' and 'real name' but the obligatory field is the username which is used to get user profile on social media. In this context, Narayan and Somaticov [12] studied Flickr and Twitter network structure and given a re-identification algorithm to match the anonymized identity of users across both the networks. The algorithm works fine with around 12% error rate. Further, Bartunov et al. [13] used users profile attributed along with social links and further they applied conditional random field to find the user identity across Facebook and Twitter, the proposed method demonstrate better results than single attributes used in the literature. Korula et al. [14] proposed a reconciliation algorithm which improves the efficiency of user identification by minimizing the error. Li. et al. [15], exploit two dissimilar social networks Twitter and Foursquare to get identical user's profile by using the follower /following relationship. Tan et al. [16] investigated that username based matching works well when a user puts the same username across the networks. It is difficult to find the user's identity across the networks when the user chooses different user-name and same username held by the different individuals.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Zhou et al. [17] proposed an efficient framework for user identification using Friendship networks (QQ, Sina Weibo) , the given framework requires priory user matched pairs(seed users) as input. Li and Su [18], analyzed the FRUI algorithm proposed in and suggested that when there is no proper 'seed user' is not selected in the experiment then several contentious nodes may be generated and the algorithm may stop its working without finding any new node . They suggested that the use of closeness centrality may work as supplement and termed a method known as FRUI with the proposed suggestion p-FRUI. Later, Zhou et al. [19] modified the friend relationship-based method in which prior seed users were required and it requires lots of effort. They proposed a new framework without using priory seed users; it uses real network relation in the absence priory user matched pairs.

In this paper, we selected only two most appropriate profile attributes publicly available across the network, and they are also available in network relationships. Vosecky et al. [3] used all the attributes available across the social networks and matches all common traits and required lots of effort, and the accuracy is less due to the use of only one algorithm. In this paper, we used four string matching algorithms and compared the results to realize the effectiveness of each algorithm for particular

III. PROBLEM FORMULATION

During online conversation, social network users often share their social account link as URLs or username. For example; to strengthen their friend relationships, many naïve users or social media marketing people usually leave some footprints on public media to increase the number of followers. Such links are very intuitive to know the user's personal information filled during online account creation i.e., profile information.

Example1: Suppose a user U has an account on Twitter and have a certain number of followers there. After some time, he has creates a new profile on any other social network like Pinterest and desires that his friends should follow him on Pinterest. Therefore, the user is compelled to post his or her profile URLs on Twitter. The cross link post provides a link between the user's account on Twitter and Pinterest (fig.1).

Fig. 1: Single user broadcasts post (http://twitter.com)

Example2: In another scenario, to increase the number of customers many business organizations post interactive content on social media accounts (fig. 2) and broadcast their social account (URLs) on public social media. Therefore, many users obliged to reveal their social account information by following such accounts. By collecting such users' account, such an organization can learn the user's behavior and recommend the desired item to a particular user.

Fig. 2: Single user broadcasts post (http://twitter.com)

Assume that a user (U), who reveals his or her identity on a particular social networks S by sharing links (profile URLs) or username as depicted in fig. 3. Further, this figure reveals how personal information like 'name,' 'username,' and 'location' can be extracted from these links using social network APIs (Application program Interface). The matching of these common attributes is learned using different fuzzy string matching algorithms. To find the similarity among the profiles of a user across various social networks, we matched two standard features 'username' and 'name' using distance measure algorithms (two tokens based and two characters based) namely; cosine, Jaccard, Levenshtein and Jaro Wrinkler distance (Cohen et al.[20]). Social network display-name often includes a 'username,' 'real name' and 'profile pictures'.

Let a user U whose profile on source network is P_s and P_t on the target network. Further, consider that P_s contain n attributes P_s (a_1, a_2, a_3,. a_n) and P_t comprises m attribute P_t (b_1, b_2, b_3,.. b_m). Suppose F represents the matching function across the networks with degree d. Then the profile of a user on both the social networks is represented by (1) and (2).

$$U(P_s(a_i)) \quad (1)$$

Where P_s is the profile of a user on source social network (Twitter), a_i denotes the number of personal attributes available on the network and $1 \leq i \leq n$.

$$U(P_t(a_j)) \quad (2)$$

Where, P_t is the profile of a user on the target social network (Fcaebook), a_j denotes the number of personal attributes available on the network and $1 \leq j \leq m$.

Again, assume that k is the number of common attributes available across the site and ranges ($0 \leq k \leq l$) then the matched degree of a particular set of the attribute is represented by the equation (3)

$$F(U(P_s(a_k), P_t(b_k)))=d \quad (3)$$

Where, d denotes the matched degree of common attributes across the social networks. Where, the value of d lies $0 \leq d \leq 1$ for distance measuring algorithm, except Levennshtein ($0 \leq d \leq N$). N represents a positive integer. If the value of d is zero, then the corresponding attributes match accurately, and if the matched degree range is either 1 or (.01-0.05) then the traits are similar otherwise distinct.

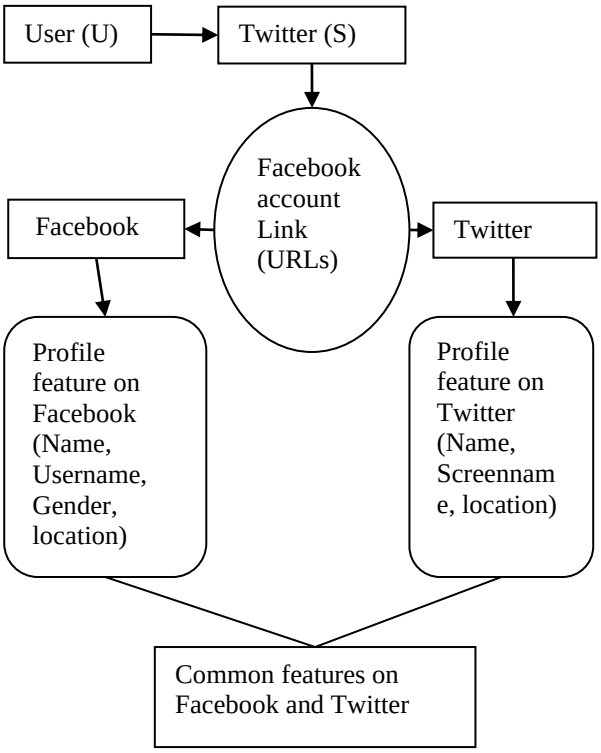

Fig. 3. A framework for identity matching

To match the user's identity across social networks, we used following string distance measuring algorithms.

A. *Character based Distance Measuring Algorithm*

In this form of algorithms, two given strings are matched character by character if characters on respective positions are same then the distance is zero; otherwise weight of the distance parameter is increased. The well known character based algorithms are Levenshtein distance (minimum edit operation required to transform one string into another), Jaro Wrinkler algorithm.

B. *Token based Distance Measuring Algorithm*

In this type of algorithms, the whole word of the given string is matched with the corresponding word in the other series. Such string strings do not check the words in lexicographic order. The well known distance measure algorithms are cosine, Jaccard, etc.

IV. DATASET

We have downloaded a dataset from the URLs mentioned in the paper (Esfanyari et al. 2018). The downloaded data is available in ".json" format. Therefore, we converted the given file in '.csv.' format using R studio platform. Further, we extracted two most essential features of the Twitter and Facebook users, i.e. 'screenname' and 'name'. Screenname attribute of Twitter is equivalent to the username on Facebook while full-name attribute of Twitter is tantamount to username present on Facebook. The purpose of extraction and matching of these attributes can play a vital role in user identification using network relation because these two attributes are most often present in the network relationship of a user across different social networks. The Strength wise description of the username based dataset is presented in table I. While the strengths of the name attribute are depicted in table 2. We extracted only English names available in the given dataset; therefore, the total number of name counts in the extracted dataset is less than the username. We carefully aligned all the data available in the dataset and found that 9520 'screenname' available for Twitter accounts while 13790 usernames available for Facebook. We extracted the name available on Twitter and Facebook. The total number of full-name obtained on Twitter and Facebook is 8570 and 13118, respectively.

Table-I:Dataset: Twitter-Facebook (Username/Screenname)

S. No.	Name of Social Network	Total number of Usernames/Screenname
1.	Facebook	13790
2.	Twitter	9520

Table- II: Dataset: Twitter-Facebook (Full-name/name)

S. No.	Name of Social Network	Total number of Name/Fullname
1.	Facebook	13118
2.	Twitter	8570

V. RESULT AND DISCUSSION

To find the similarity between the profiles feature of a user across the two social networks, in this paper, we use two characters based distance measure algorithm, namely; Levenshtein and Jaro Wrinkler. The other two are token-based (Cosine and Jaccard). We implemented both types of algorithms in R language. The experiment was performed on a core i5 processor with 8GB RAM.

The result of our experiment corresponding to username and name matching is demonstrated in table III and IV by using different distance measuring algorithms.

A. *Username similarity across Facebook and Twitter*

From the observation of table III, we found that both the character based algorithms give the same results while token based algorithms give different results. From the inspection of table III, it is also clear that users generally put the same or a similar username across the sites.

In online social networks, due to limited human recall capability, people often select same or similar username across different social media services. To test this hypothesis in this paper, we match the usernames of Twitter and Facebook using four distances measuring algorithms and the corresponding result is shown in table III. From the observation of Fig. 4, it is clear that in online social networks 'user identification' using only 'username' contributes maximum 16.36 % across Twitter and Facebook.To understand the status of each account, we selected a hundred exactly matched username pairs and redirected URLs on the corresponding sites. The results obtained are shown in the fig. 5.

From the observation of the fig. 5, it is observed that only eighty six percent account pairs matched on both the networks and remaining account pairs are not found due to the dynamic nature of social media sites. Many people deleted their profiles from a particular social media due to the personal issue, and many accounts are suspended from the social media service because of suspected activities. From the evaluation of the results, we found that seven percent link missing on Facebook while five percent of Twitter URLs are missing. It happens because some people alter their username after a certain interval of time. Only two percent user match pairs are missing on both the social networks. Such accounts may be either spam or reluctant users.

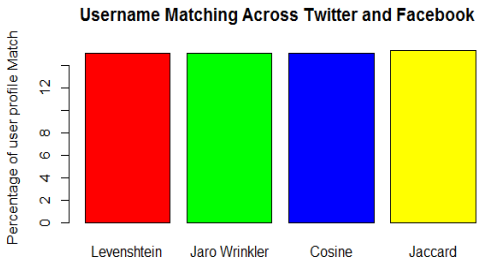

Fig.4. Username match across Facebook and Twitter

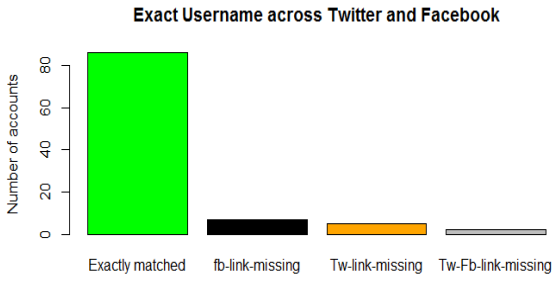

Fig. 5. Status of users matched accounts

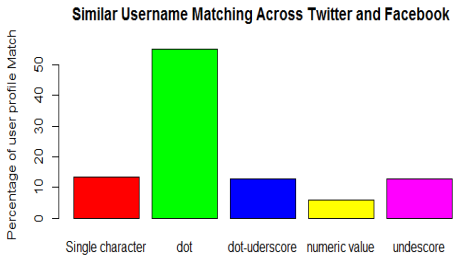

Fig.6. Similar username across Facebook and Twitter

Due to privacy reasons, many users alter their 'username' across the sites using specific symbols that support by the particular sites; like a dot on Facebook while 'underscore' on Twitter and other characters may be ‘numbers’ or ‘alphabets’. To verify similar ‘username’ across the different social site, we selected the results of Levenshtein distance with score one. Only 4.86 % of users selected such username across Facebook and Twitter. Fifty seven percent of Facebook users used 'dot' symbol and their corresponding matched users on Twitter using the same username without 'dot'. Twitter users generally select 'underscore symbol in the username, about fourteen percent of users selected the underscore on Twitter, whereas Facebook users chose the same 'screenname' without 'underscore' symbol. Similarly, Fourteen percent of users used the 'dot' symbol on Facebook while 'underscore' on Twitter. The remaining users chose the numeric or alphabet on the respective social networks. Fig. 6 demonstrates the users' username selection behavior across Twitter and Facebook.

B. *Name similarity across Facebook and Twitter*

To find the user identity across these social networks, we selected the ‘full-name’ attribute, which is the most consistent feature available across the sites. To understand the name consistency across these social networks, we used four different distance measuring algorithms, namely; Levenshtein distance, Jaro Wrinkler distance, cosine distance, and Jaccard distance measure. The name matches using the different algorithms is depicted in table IV.

Table- III: Dataset: Twitter-Facebook (Screen-name/username)

Attributes	Total number of users on Facebook	Total number of users on Twitter	Number of matched pairs using Levenshtein Distance	Number of matched pairs using Jaro Wrinkler Distance	Number of matched pairs using Cosine	Number of matched pairs using Jaccard Distance
Screenname or Username	13790	9520	1437	1437	1443	1558

Table- IV: Dataset: Twitter-Facebook (Full-name/name)

Attributes	Total number of users on Facebook	Total number of users on Twitter	Number of matched pairs using Levenshtein Distance	Number of matched pairs using Jaro Wrinkler Distance	Number of matched pairs using Cosine	Number of matched pairs using Jaccard Distance

Name or Fullname	13118	8570	1864	1864	1869	2068

From the observation of fig. 7, we found that exact name similarity across Twitter and Facebook ranges from 20.42% to 22.65% using different algorithms. From the observation of the results, we found that Jaccard distance consistently outperforms than the other algorithms.

Levenshtein distance matches only those strings which are exactly the same on the source and the target networks. The performance of Jaro Wrinkler distance is exactly the same as Levenshtein distance for exact name matching. Cosine similarity shows better performance than the character based distance (Levenshtein and Jaro Wrinkler) while poorer than the Jaccard distance. The reason is that Jaccard distance considers overlapping of characters in the present strings. As a result, this algorithm improves the performance by adding, reverse strings, prefixes, suffixes, and middle words available in the name strings available across the sites.

From the study of two publicly available attributes on both the social networks: we find that eighty four percent of users selected the different username to make their profiles anonymous. Around seventy eight percent of users selected their 'names' different across these two popular social networks because many users are extra conscious regarding their personal information. Another reason behind the selection of such a large amount of different names by the same user across Facebook and Twitter is the nature of these services (Facebook is personal while Twitter is public). Therefore, we

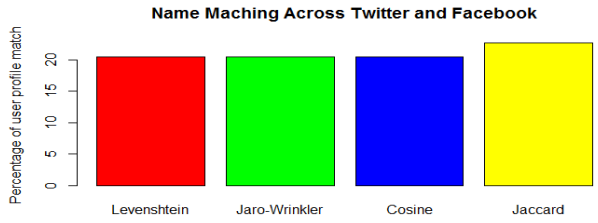

Fig. 7. Name similarity across Twitter and Facebook.

conclude that different sets of social media sites contain a different degree of correlation with these public attributes.

VI. CONCLUSION AND FUTURE WORK

Social networking services are becoming more popular among the masses because of the decreasing cost of the Internet and smart handheld devices (Mobile Phone and Tablet). Consequently, the strength of social media users is increasing day by day. Finding the account of a user on different social media is a challenging task because of the non-uniform nature of social media. This paper examines different fuzzy string matching algorithms to find the similarity among names and usernames of users across two popular social media networks: Twitter and Facebook. These findings can be exploited in users' behavior analysis, which will further help in improving social recommendation system for the creation of customized and novel online services.

REFERENCES

1. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/. (Last accessed on 20/01/2019).
2. http://www.pewinternet.org/2016/11/11/social-media-update-2016. (Last accessed on 20/01/2019).
3. Vosecky J., Hong D., and Shen V. Y., (2009) User identification across multiple social networks. In First International Conference on Networked Digital Technologies, 360-365.
4. Esfandyari A., Zignani M., Gaito S., Rossi G. P., (2018) User identification across online social networks in practice: Pitfalls and solution," Journal of Information Science, 44 (3), 377-391.
5. Zafarani R., Tang L., Liu H., (2015) User identification across social media, ACM Transactions on Knowledge Discovery from Data, 10 (2),16.
6. Malhotra A., Totti L., Meira W. J., Kumaraguru P., and Almeida V., (2012) Studying user footprints in different online social networks. In International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 1065-1070.
7. Motoyama Marti, and George Varghese. (2009) I seek you: searching and matching individuals in social networks. Proceedings of the eleventh international workshop on Web information and data management, ACM.67-75.
8. Raad, E., Chbeir, R., & Dipanda, A. (2010). User profile matching in social networks. In International Conference on Network-Based Information Systems (NBiS), 297-304.
9. Zheng R., Li H. C., and Huang Z., (2006) A framework for authorship identification of online messages: writing-style features and classification techniques. J. of the American Society for Information Science and Technology, 57(3), 378-393.
10. Goga O., Lei H., Parthasarathi S.H.K., Friedland G., Sommer R., and Teixeira R., (2013) Exploiting innocuous activity for correlating users across sites.In Proceedings of the 22nd international conference on World Wide Web, 447-458,
11. Li J., Wang G. A., and Chen H., (2011) Identity matching using personal and social identity features. Information Systems Frontiers,13 (1), 101-113,.
12. Narayanan A. and Shmatikov V., (2009) De-anonymizing social networks. In Proceedings Of the 30th IEEE Symposium on Security and Privacy, 173-187.
13. Bartunov S., Korshunov A., Park S., Ryu W., and Lee H., Joint link-attribute user identity resolution in online social networks. The 6th SNA-KDD Workshop(2012).
14. Korula N., and Lattanzi S., (2014) An efficient reconciliation algorithm for social networks. In proceedings of the VLDB Endowment, 7 (5), 377-388.
15. Li Y., Zhang Z., Peng Y., Hongzhi Y., and Xu Q. (2018) Matching user accounts based on user generated content across social networks. Future Generation Computer Systems 83: 104-115
16. Tan S., Guan Z., Cai D., Qin X., Bu J., and Chen C., (2014) Mapping users across networks by manifold alignment on hypergraph, In AAAI, 14, 159-165.
17. Zhou X., Liang X., Zhang H., and Ma Y., (2016) Cross-platform identification of anonymous identical users in multiple social media networks. IEEE transactions on knowledge and data engineering,28 (2), 411-424.
18. Yongjun L.I., and Su, Z.(2018) A Comment on Cross-Platform Identification of Anonymous Identical Users in Multiple Social Media Networks. IEEE Transactions on Knowledge and Data Engineering, 30(7),1409-1410.
19. Zhou, Xiaoping, Xun Liang, Xiaoyong Du, and Jichao Zhao (2018) Structure based user identification across social networks. IEEE Transactions on Knowledge and Data Engineering, 30(6),1178-1191.
20. Cohen W., Ravikumar P., and Fienberg S., (2003) A comparison of string metrics for matching names and records. In Kdd workshop on data cleaning and object consolidation 3, 73-78.

AUTHORS PROFILE

Waseem Ahmad received the B. Tech from VBS Purvanchal University, Jaunpur, UP, India in 2007 and M. Tech degree from MD University, Rohtak,Haryana,India in 2011. He is currently working toward the PhD degree in the Department of Computer Engineering, Aligarh Muslim University, India, and is an assistant professor at the Al-Falah University, Faridabad Haryana. His research interests include web mining and information retrieval.

Rashid Ali obtained his B.Tech. and M.Tech. from A.M.U. Aligarh, India in 1999 and 2001 respectively. He obtained his PhD in Computer Engineering in February 2010 from A.M.U. Aligarh. His PhD work was on performance evaluation of Web Search Engines. He has authored about 125 papers in various International Journals and International conference proceedings. He has presented papers in many International conferences and has also chaired sessions in few International conferences. He has reviewed articles for some of the reputed International Journals and International conference proceedings. He has supervised 19 M.Tech Dissertation and three PhD Thesis. Currently, he is supervising five PhD candidates. His research interests include Web-Searching, Web-Mining, soft computing Techniques (Rough-Set, Artificial Neural Networks, fuzzy logic etc), Recommender Systems and Online Social Network Analysis