International Journal of Database Theory and Application
Vol.9, No.12 (2016), pp.101-110
http://dx.doi.org/10.14257/ijdta.2016.9.12.10

Trust Evaluation on Social Media based on Different Similarity Metrics

Anubha Maurya and M. P. Singh

National Institute of Technology Patna
Bihar, India
National Institute of Technology Patna
Bihar, India
anubhamaurya@nitp.ac.in, mps@nitp.ac.in

Abstract

With advancement in internet era, the importance of social media is increasing day by day. It enables users to share their profile data, ideas, videos and any content they have with them. With benefits, it also has several issues related to it. One of the issue is "how to protect users from after effect of friendship over social media?". This paper proposes a trust model to overcome it. The proposed model calculates trust to assist end users to take decision about accepting friend-request on social media. Trust evaluation is based upon profile similarity analysis. Trust computation uses preferred attribute among profile attributes to evaluate trust of users. The paper analyzes different trust evaluation methods based on the proposed model.

Keywords*: Social Media, trustor, trustee, profile similarity analysis, similarity metrics, Ordered Weight Average (OWA)*

1. Introduction

Nowadays social media applications are being used by people as a part of their day-to-day online activities. The diverse nature task (comment writing, tagging friends', uploading of contents or expanding friend circle) performed on social media application and the availability of large number of users and content empowers people to develop and enrich their common interests with each other. It also provides a new platform for deriving users' relationship but at same time, increases complexity to store and extract data and information.

In general way, the relationship among users on social media sites can be established in two ways: *i)* Familiarity indication and *ii)* Similarity indication.

Familiarity indication can be collected from users' data when they know each other in prior. For instance, explicit message conservation on public sites, connection on some social network sites or bonding together on a webpage. Whereas similarity indication between two people (may be strangers) can be defined as when users' share some common attribute between each other. It may be common interest, tagged in same photograph or connects with same person. Mining of similarity indices between two persons may be useful for several scenarios such as review of similar interest person can be utilized when domain experts are at geographical far place to give suggestions, feedbacks or opinion [10]. Recommendation system also utilizes the similarity analysis (collaborative filtering [1]), to understand and gain from different similarity resources. In [8], authors shown the "kindness of strangers" that people are ready to provide aid to others who are similar in interest.

In today's social media environment trustworthy relationship is in demand. Trust is defined as an estimated parameter which helps a user to judge it neighbors behavior in future. It helps one user to trust its neighbor user. This paper uses similarity indices to

ISSN: 2005-4270 IJDTA

achieve trust between users from their profile attributes. Based on each profile attribute, model calculates similarity indication and then a similarity score (trust score) is calculated from these similarity indices. The similarity score assists in either accepting or rejecting the friend request. Experiment of this paper evaluates the trust between trustee (friendship requestor) and trustor (friendship request receiver) on basis of similarity between them.

The paper organization is as follows. Section 2 presents the review about the metrics used for similarity evaluation and attributes used for profile similarity analysis. Section 3 describes the working of proposed model. It may assist a user whether he/she can accept the friend request or not based on trust evaluation between them. Section 4 explains about the data collection method and its generalized representation in social media. Section 5 analyzes the experimental results. Finally, Section 6 concludes the paper and presents future scope of this work.

2. Related Works

This paper presents review based on two aspects, namely: *i) Measuring Similarity between Users* and *ii) Attributes used for Profile Similarity.*

2.1. Measuring Similarity between Users

Measuring similarity invokes two factors, namely the approach used in measuring and attributes used in measuring (*i.e.*, the features used to measure similarity among users). Therefore, in this section, first introduction is about the widely used similarity measuring methods and then the social attributes that are engrossed to measure user similarity. For assessment of similarity metrics, two objects (*e.g.*, social users) denoted by u and v, and their associated attributes of profiles (*e.g.*, friends, location, interest, education *etc.*) represented as u_i and v_i is considered where $i \in$ the profile attributes of the user and $s(u, v)$ represents the similarity between u and v. And hence, Table 1 represents description of similarity metrics.

Table 1. Similarity Metrics Description

Similarity Metrics	Description
Cosine Similarity Metrics (S_{cos}) [2, 11, 12, 13]	It uses two vectors (*e.g.*, users u and v) attributes (u_i, v_i) to calculate the cosine angle (θ) between the two vectors that is cos θ.
Levenshtein Similarity Metrics ($Lev_{u,v}(u_i, v_i)$) [16]	It measures the distance between two attributes (u_i, v_i). It calculates the minimum number of single-character edits (insertions, deletions or substitutions) needed to change one attribute value into another.
Jaro-Winkler Similarity Metrics (S_{JW}) [4]	It measures attributes (u_i, v_i), characters in common, such that no more than half length of the lengthier attribute value in distance, with transpositions. It gives more significance to differences near the start than differences at the end.
Jaccard Similarity Metrics (S_{Jac}) [14, 17]	It gives association between two attributes (u_i, v_i). It is the result of division between numbers of features (attributes) that are common between them divided by total no. of features selected for similarity measure.
Dice Similarity Metrics (S_{Dice})	It is similar as Jaccard Metrics. Only difference is that it omits true negatives from both common value as well as total feature set.
Monge-Elkan Similarity Metrics S_{ME} [3]	It gives average similarity value between pairs of more similar attributes within u_i *and* v_i
Letter Similarity Score Metrics ($S_{LetterPair}$)	It is a string similarity metric that uses both *common substrings* and *common ordering* of substrings for matching purpose. In addition to this, it considers other common substrings along with longest common substrings

2.2. Attributes used for Profile Similarity

The different social media platform uses different attributes to represent a user profile. However, this paper uses generalized profile attributes to calculate similarity between the social users. These attributes may be classified under two types:

i) **User's Social Attributes:** User's social attribute are those features that can be derived from their basic information. It includes the gender, age, education background, location of user, their interests, employment background *etc.* Different social media platforms usually have different user basic information. These attribute features are used to identify the similar peers into peer-to-peer (P2P) networks [5]. These social attribute features analysis may be sufficient to understand the character of a social user. These attributes may provide a foundation for modeling and applying them at time of evaluating trust of a user in Online Social Networks (OSNs). Nevertheless, we cannot fully rely on these attributes. But still they are the essential part of user information and may aid in evaluating a user trustworthiness based upon similarity among users.

ii) **User's Structural Attributes:** These attributes are extracted from links present in user's social graph. The relationship that a user share with others in a group [15] and mutual friends [6] are natural link that a social media exhibits. These attributes give information that there exist an indirect link between two users and recommendation from mutual friends can be used to form a friendship link between them.

Since the attributes selected for analysis of profile are from different sources, hence an associated weight is assigned to each attribute to evaluate the final trust of the user from trustor's perspective. Therefore profile similarity analysis will include the dynamic ordered weighted averaging (OWA) operator for calculation of weight to each attribute of profile.

In general OWA operator of elements *m* is a mapping *F*: $\mathbb{R}^m \rightarrow \mathbb{R}$ that has a correlated weighting vector $W = (w_{1,} \dots \dots \dots, w_m)^T$ of having the following properties [7]:

$$w_1 + \dots + w_m = 1, \;\; 0 \leq w \leq 1, i = 1, \dots \dots .., m \tag{10}$$

and such that

$$F(a_1, \dots, a_m) = \sum_{i=1}^{m} w_i \, a_i \tag{11}$$

where a_i is the ith largest element of the aggregated objects

$\{a_1, \dots \dots \dots a_m\}$. OWA operator has two properties namely is *i) measure of orness of aggregation* and *ii) measure of dispersion of aggregation.*

The profile attributes common among trustee and trustor are taken into account. This paper uses OWA operator with slight modification. The trustor's preferred attribute may be given importance but other attributes also need to be considered to give final outcome as only single attribute cannot define all characteristics of a user on social media. Hence for analysis, choice of trustor will be given more preference and placed at a_1 followed by similarity score of remaining attributes arranged in decreasing order.

3. Proposed Model

A friendship link can be established between two users if they exhibit some common attributes among themselves. For instance, if a new user *u* is working at XYZ, he might know other person *v* who was/is employee of XYZ. Then there is a probability that they may become friends on social media. Therefore, this paper proposes a model to assist in either accepting or rejecting the friendship between the users on social media as shown in Figure 1. This model uses different similarity metrics on profile attributes to evaluate trust based on similarity factor. The similarity analysis on basis of these attributes will help one to judge how much the requestor is similar to the receiver. This model uses OWA to assign dynamic weights for all attributes. If any attribute is missing the weight assignment to them is made 0 (zero).

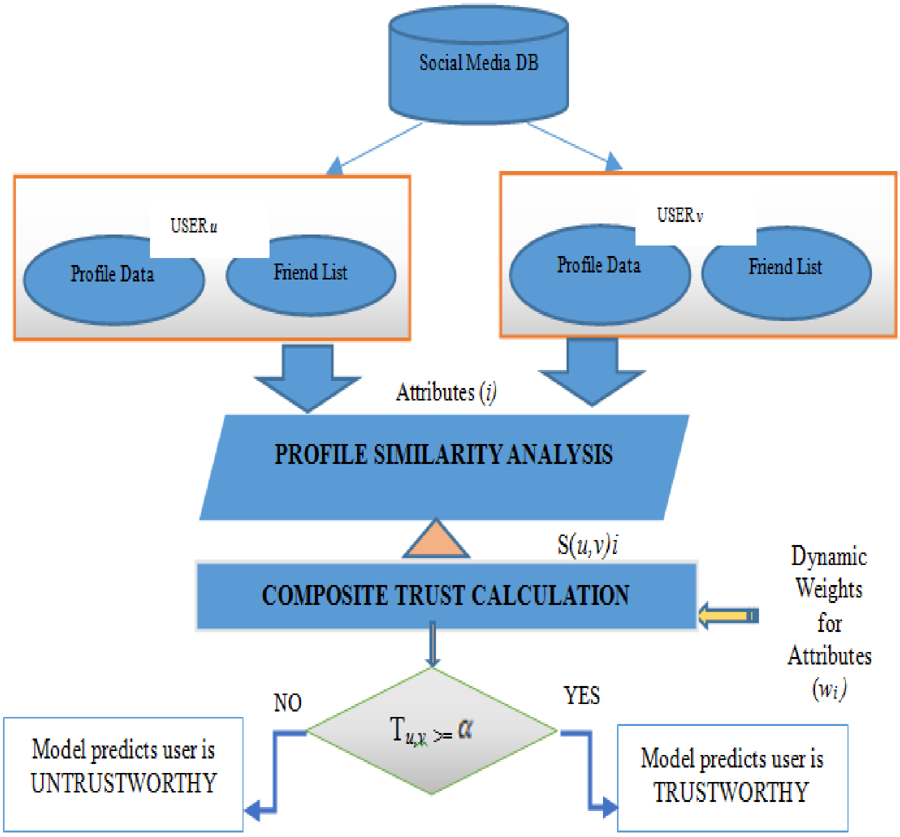


Figure 1. Workflow of Proposed Model

The corresponding trust value $T_{(u,\,v)}$ is calculated as follows:

$$T_{u,v} = \sum_{i=1}^{12} S_{(u,v)i} * w_i * 100 \tag{12}$$

where $S_{(u,v)i}$ is similarity score for attribute i, w_i is dynamic weight assigned to attribute i and α is a threshold value for trust evaluation. If $T_{u,v}$ is greater than α then user may be considered as trustworthy for friendship.

4. Data Collection and Description

Profile attributes may be considered as fundamental unit used to analyze the user's behavior and structure on social media. An issue arises when the attributes from different OSN platforms are collected. They are represented in different format on different social media. Unfortunately, the data which is available is also limited or requires access token, which may not fulfill the researchers' requirements for accomplishing deep studies and analysis. Therefore to generalize, attributes format an API is built and named as "TRUSTBOOK". This paper demonstrates the representativeness of data set by comparing the characters (*e.g.*, location, education, interest, friends *etc.*) which are revealed by users in "TRUSTBOOK". Therefore, it is believed that this data set and the research based on the data set are representative.

(a) Login Page of Trustbook (b) User Profile Information (Attributes)

Figure 2. (a) Shows the First Page of the API where Registered User Login or New User Can Sign Up

4.1. User Profile Information

Each user profile is composed of three parts of information including Identity Data, Social Graph Data and Content [9]. For profile analysis in this paper identity data and social graph information is used.

- *Identity Information*: It refers to specific profile attributes: age, gender, current city, hometown, high school, college and employer (*i.e.*, work place). Current city and hometown are two location attributes which are linked to corresponding latitude/longitude position. High school, college and employer, as location relevant attributes, are associated with a city name and latitude/longitude values.
- *Social Relationships*: It includes users' friend lists, thus this paper defines social relationship as user-claimed friendship. This paper considers friendship is bidirectional that is if u is v*'s* friend if and only if v is u*'s* friend.

Figure 2b represents Profile Information Page of a registered user on the "TRUSTBOOK" API. The attributes used for similarity evaluation are: current place of living, hometown of user, education information (10th School, 12th School, Graduation and Post-graduation), Employment details (current employment and previous working data), Relationship status, languages known by a user and Interest details and mutual friends between trustor and trustee.

5. Analysis Results

For trust evaluation based on profile similarity in social media, this paper analyzes the results of trust between users.

5.1. Analysis based on Test Cases

Five different cases have been used to test the proposed model. Following are the cases that may exist when user u sends friend request to user v and user v evaluates trust on user u based upon preferred attribute.

Case 1: When only the preferred Attribute between user u and user v is similar: There may a case when users' have similar interest only on the attribute used for similarity evaluation.

Case 2: When only the preferred attribute is dissimilar between user u and user v: It may happen that the attribute selected for profile similarity evaluation is dissimilar between users'.

Case 3: When all profile attributes are similar between user u and user v: The case may exist that two users' have similar interest to characterize themselves.

Case 4: When all profile attributes are dissimilar between user u and user v: It may also happen that two users' are entirely different in all aspects.

Case 5: When profile is partially similar but preferred attribute is similar: There may be a case when both partial attributes and attribute selected for analysis are similar between the users'.

<u>Case 6</u>: *When profile is partially similar preferred attribute is dissimilar*: It may be a case when partial attributes are similar between the users' but the attribute preferred is dissimilar.

Table 2. Trust Evaluation of User in Different Test Case Environment

	Cosine	Jaccard	Levenshtein	Jaro Winkler	Letter Pair	Dice	Monge Elkan
Case 1	18.36	5.69	35.08	55.8	25.09	13.83	10
Case 2	71.6	30.87	97.11	82.24	67.21	50.58	71.6
Case 3	89.26	43.6	89.26	89.26	89.26	89.26	89.26
Case 4	4.99	2.53	25.45	53.32	18.72	3.87	0
Case 5	77.44	33.66	77.26	78.7	76.79	75.38	70.74
Case 6	35.95	25..27	40.83	58.74	35.24	34.32	32.23

Table 2 shows the trust value of a user under different circumstances evaluated on different similarity metrics. Figure 3 shows a comparison statistics drawn from Table 2.

From analysis it has been observed that Levenshtein similarity metrics and Jaro Winkler similarity metrics behaves almost same in all cases. In *case 4*, they shows that users are similar and give high trust value as compared to others because it calculates similarity on basis of number of modification within the string and convert it to nearby similar string.

Whereas Jaccard similarity metrics evaluates trust very low in almost every case. In *case 3*, the users are similar in all respects but the trust evaluated by it is very low in comparison to others because it calculates the similarity between users by considering the completely common attributes and discards the partial matching features.

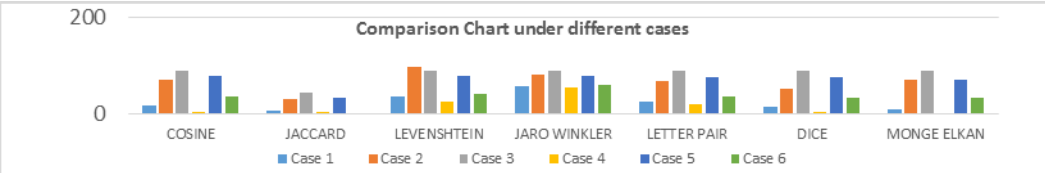


Figure 3. Trust Evaluation in Different Cases by Different Similarity Metrics

5.2. Analysis on Dataset

For deeper analysis, proposed model is implemented with different similarity metrics that are applied on data collected from API "TRUSTBOOK". Figure 4 shows the profile dataset of user *u* and user *v*. In real world environment of social media, the analysis result is shown in Table 3. Here the trust evaluation is done on dataset of user *u* and user *v* by using different similarity metrics.

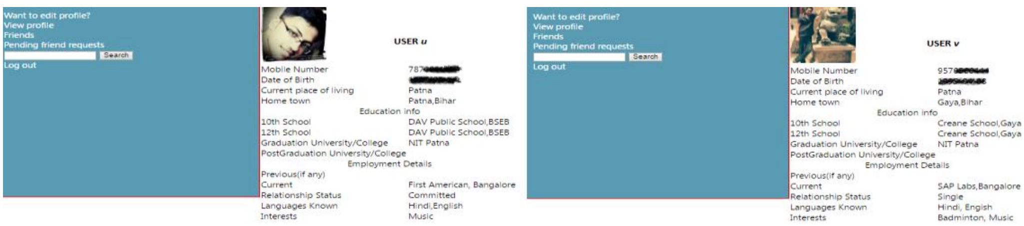


Figure 4. Dataset of User *u* and User *v*

From Table 3, three aspects of profile similarity analysis for trust evaluation is derived. 12 profile attributes are used out of those 3 attributes are similar, 4 are partially similar and 5 attributes are dissimilar between user *u* and user *v*. Experimental analysis is described below.

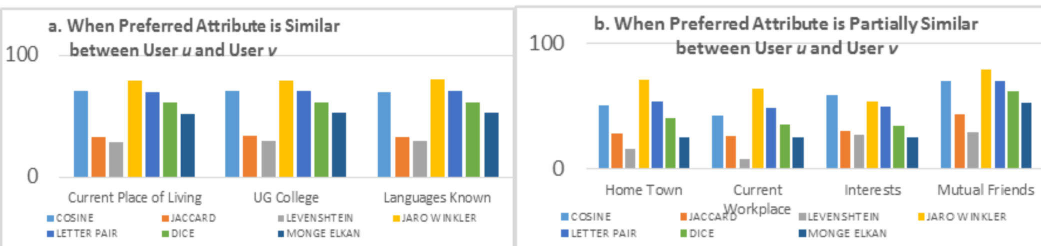
1. Figure 5a represents the comparison among similarity metrics when the attribute selected for trust evaluation is similar between the users' *u* and *v*. It is observed that Cosine metric and Jaro winkler metric shows best similarity between them. Levenshtein and Jaccard results are very low in comparison to other metrics.

2. Figure 5b depicts the comparison among metrics when the preferred attribute is partially similar. In this scenario cosine metric shows better result as compared to Jaccard metric and Levenshtein metric. Cosine metric is considered better is this case because preferred attribute is partially similar between the user and remaining attributes those are either similar or dissimilar plays little role but they are considered at time of evaluating trust.

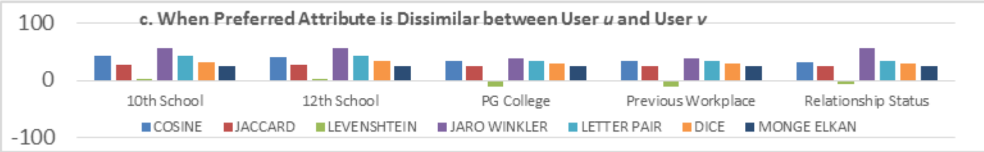
3. In Figure 5c, dissimilar attribute is the preferred attribute for evaluating the similarity between the users'. For this Jaccard, Levenshtein and Cosine gives comparatively better output because preferred attribute is dissimilar but still rest attributes constitute the trust factor for a user.

Table 3. Trust Evaluation of User by Different Similarity Metrics

	USER u vs USER v							
		Cosine	Jaccard	Levenshtein	Jaro Winkler	Letter Pair	Dice	Monge Elkan
Preferred attribute is similar	Current Place of Living	70.32	33	29.1	79.09	69.99	61.46	51.8
	UG College	70.7	33.5	29.22	79	70.44	61.1	52.4
	Languages Known	70	32.9	29.55	79.9	70.66	61.46	52.8
Preferred attribute is partially similar	Home Town	50.38	28.5	16.31	71.09	53.93	40.39	25.12
	Current Workplace	42.68	26.5	7.97	64.02	48.49	35.25	25
	Interests	58.4	30.1	27.2	53.69	49.46	34.36	24.8
	Mutual Friends	70.32	43.9	29.22	79.09	70.44	61.46	52.4
Preferred Attribute is Dissimilar	10th School	42.68	26.5	0.54	56.58	43.4	32.14	25.12
	12th School	42	27.2	1.6	57	43.77	33.1	24.99
	PG College	33.3	24.2	-11.4	38.5	33.74	29.4	25.01
	Previous Workplace	33.46	24.1	-10.5	37.87	33.74	28.78	24.78
	Relationship Status	32.5	24.2	-7.35	57.59	33.74	29.09	25.23



(a) (b)



(c)

Figure 5. Trust Evaluation by different Similarity Metrics when Preferred Attribute between User *u* and User *v* is (a) Similar (b) Partially Similar (c) Dissimilar

6. Conclusion and Future Work

This paper proposes a model to evaluate trust of a user before forming a friendship link on social media. For trust evaluation, profile attributes are matched between the user *u* and user *v*. The similarity ratio between the users is calculated and a composite trust value is formed. The composite trust score based on profile matching, may also assist the receiver (trustor) to take an appropriate decision about accepting a friendship request of requestor (trustee). By analyzing the results of all similarity metrics in different scenarios, it is concluded that cosine metrics behaves better in all situation since it considers all attributes importance and give appropriate weightage to the preferred attribute.

In this paper one-to-one matching using similarity metrics is applied. It can be further move to next level where an ontological structure of the profile may be created and then similarity techniques are applied to analyze the results.

References

[1] G. David, D. Nichols, B. M. Oki and D. Terry, “Using collaborative filtering to weave an information tapestry”, Communications of the ACM, vol. 35, no. 12, **(1992)**, pp. 61-70.

[2] C. Charalampos and V. K. Prasanna, “Social link prediction in online social tagging systems”, ACM Transactions on Information Systems (TOIS), vol. 31, no. 4, **(2013)**, pp. 20.

[3] M. Alvaro and C. Elkan, “An efficient domain-independent algorithm for detecting approximately duplicate database records”, **(1997)**.

[4] E. W. William, “String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage”, **(1990)**.

[5] S. Haiying, Z. Li and K. Chen, “Social-P2P: An Online Social Network Based P2P File Sharing System”, IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 10, **(2015)**, pp. 2874-2889.

[6] E. J. N. Mark, “Clustering and preferential attachment in growing network”, Physical review E, vol. 64, no. 2, **(2001)**, pp. 025102.

[7] R. Y. Ronald, “On ordered weighted averaging aggregation operators in multi-criteria decision making”, IEEE Transactions on systems, Man, and Cybernetics, vol. 18, no. 1, **(1988)**, pp. 183-190.

[8] C. David, L. Sproull and S. Kiesler, “The kindness of strangers: The usefulness of electronic weak ties for technical advice”, Organization science, vol. 7, no. 2, **(1996)**, pp. 119-135.

[9] M. Anubha and M. P. Singh, “A Survey on Social Networks: Issues and Attacks”, In Recent Advances in Mathematics, Statistics and Computer Science, **(2016)**, pp. 634-642.

[10] B. Krisztian and M. de Rijke, “Finding similar experts”, In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, **(2007)**, pp. 821-822.

[11] Y. Mao, P. Yin and W. C. Lee, “Location recommendation for location-based social networks”, In Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems, **(2010)**, pp. 458-461.

[12] A. Ashton, D. Huttenlocher, J. Kleinberg and J. Leskovec, “Effects of user similarity in social media”, In Proceedings of the fifth ACM international conference on Web search and data mining, **(2012)**, pp. 703-712.

[13] A. Xavier, N. Lathia, J. M. Pujol, H. Kwak and N. Oliver, “The wisdom of the few: a collaborative filtering approach based on expert opinions from the web”, In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, **(2009)**, pp. 532-539.

[14] N. Suphakit, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of Jaccard coefficient for keywords similarity", In Proceedings of the International Multi-Conference of Engineers and Computer Scientists, vol. 1, **(2013)**, pp. 13-15.
[15] L. Vincent, B. B. Cambazoglu and F. Bonchi, "Cold start link prediction", In Proceedings of the 16th ACM SIGKDD International conference on Knowledge discovery and data mining, **(2010)**, pp. 393-402.
[16] C. William, P. Ravikumar and S. Fienberg, "A comparison of string metrics for matching names and records", In Kdd workshop on data cleaning and object consolidation, vol. 3, **(2003)**, pp. 73-78.
[17] J. Yu, G. Li, J. Feng and W. S. Li "String similarity joins: An experimental evaluation", Proceedings of the VLDB Endowment, vol. 7, no. 8, **(2014)**, pp. 625-636.

