SocialMatching++: A Novel Approach for Interlinking User Profiles on Social Networks

Hussein Hazimeh[1], Elena Mugellini[1], Omar Abou Khaled[1], Philippe Cudré-Mauroux[2]

[1] University of Applied Sciences of Western Switzerland, Fribourg, Switzerland
{fisrtname.lastname}@hefr.ch
[2] eXascale Infolab, University of Fribourg, Fribourg Switzerland
phil@exascale.info

Abstract With the large number of users connected to social networks, screenname duplication is a rising problem, which leads to interference when trying to recognize users. A number of algorithms have been proposed to distinguish user profiles on one or multiple social networks. The main task in this context is to have robust features. According to the state-of-the-art approaches, features can be: content and behavioural based features, that compare content similarity between posts or behaviour similarity (timestamps between posts (behavioural), or overlapping between content (content) for example). Attribute-based features that compare profiles attributes, such as gender, age, location or image. In this paper, we tackle this problem and propose SocialMatching++ a novel approach that leverages: (1) user life events such as graduation, marriage or new job, which used to enhance the behavioural approaches (2) profile biographies, which consist in small paragraphs that users write to comprise arbitrary information about themselves. These are used to enhance the attribute approaches. To evaluate our approach, we conducted experiments on 2,263 different profiles from Facebook matched with 5,694 Twitter users, and compared them with two baseline approaches. Our results show that SocialMatching++ achieves better results compared to the baselines approaches, showing that our system successfully bridges the gap between behavioural and attribute based approaches.

Keywords: Identity Linkage, Profile Matching, Social Networks

1 Introduction

The process of registration on any social networking website is accomplished in an easy fashion, launched by the creation of a new user account. Typically, it is necessary to provide an email and a password to obtain a valid user profile. A user can later access his own profile and modify a diversity of settings concerning several aspects, such as profile image, location, relationship status, social network, interests and other information. People can register themselves on several sites, moreover, the same user can create different accounts on a single platform by providing the same information for each different profiles while using different email accounts. So far, social networking sites did not setup any mechanisms yet to detect discover if two accounts are similar and merge them. This problem is mainly originating from databases known as Record Linkage (RL). RL is the task of identifying records corresponding to the same entity from one or more data sources accurately, [22, 27]. The process of RL normally starts by resolving entities in database (Entity resolution), matching them using convenient data matching techniques and finally merging similar records.

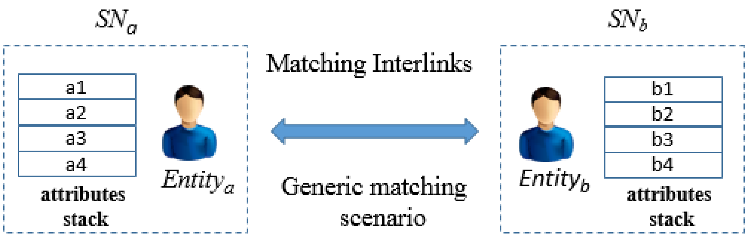


Fig. 1. Abstraction of Profile Matching Scenario

The scenario is similar in social networks; Figure 1 shows a generic model that explains how it works. Suppose that we have two social networks, social network SN_a and social network SN_b, and we have a user entity $Entity_a$ on SN_a and user entity $Entity_b$ on SN_b. The objective is to discover whether these two profile accounts are linkable. Approaches to user profile matching usually start by defining a set of features (matching interlinks) to link user profiles. Interlinks are categorized into attribute information or context and semantic information [38]. Attribute based matchers employ profile attributes, such as screennames, profile images, birthdates, etc. Context or semantic matchers compare the behavioural likeness of user profiles. Both categories of matching algorithms poses a set of challenges: in attribute ones, information can be private or not updated; in semantic (behavioural) ones user's activity can be completely different between two social networks.

The research community works forward to investigate new matching mechanisms. Timestamp variations between user posts, is a widely used behavioral feature. However, this feature could be weak if the user is active in only one social network rather than the others. Similarly, profile attributes like image, location, and others, might be not updated over time.

In this paper, we tackle this problem by leveraging two novel profile-matching features: (1) user life events and (2) profile biographies. Our approach enhances the behavioral approaches and attribute ones, arguing that even if a user is not active in one social network, the potential of sharing his life events such as marriage, graduation, new job, etc, could be high. Furthermore, given life events cannot be common for two different users. Profile biographies contain key information about the user, with the advantage that it is always public and easy to be fetched. The proposed approach is tested on profiles extracted from Facebook, and linked them to their correspondent Twitter accounts.

2 Related Work

The following two sections present the most relevant approaches closest to ours. In Table 1, we mention the feature(s) used by each approach.

2.1 Attribute-based approaches

Attribute-based approaches use basic profile attributes such as: gender, age, location or profile images. Goga et al. 2015 [9] define four characteristics: availability, consistency, non-impersonability, and discriminability. By using each one of the following attributes, they can reach a reliable matching. The second contribution of their work is how to select the training and testing sets properly.
Raad et al. [2] goal is to create a framework that finds the similarity among user profiles across different social networks. They exploit to solve this problem all the user's profile attributes. The matching mechanism starts by assigning weights to each one of the profile attributes based on its importance. To decide the matching occurrence, they create a decision-making algorithm and assign it this task.
Jain et al. 2015 [25] Link user profiles' accounts by detecting the historical modifications of user's profile attribute information. Historical values of attributes can definitely link two same users. However, they are not always available and hard to fetch them.
Bennacer et al. [18] match social accounts across Flickr, Live-Journal, Twitter and YouTube using names, emails and links to other webpages. In addition, they define a set of rules on the aforementioned attributes to ensure the matching accuracy.
Zafarani et al. [1] provide an approach for mapping user across communities. They use data collected form twelve different communities. The main goal of this research is to connect these platforms using community mappings. To achieve this task the authors rely on usernames and URL with an accuracy of 66%.

2.2 Content and behavioural based approaches

Behavioural-based approaches use comparisons between user behaviours leveraging features such as posting rates, timestamps, or comparing content of posts.
Van Le et al. [10] propose a system for user profile modelling which exploits Latent Dirichlet Allocation (LDA) to discover the hidden topics that lies inside user-generated contents.
Liu et al. [12] introduce HYDRA for linking different accounts of the same user using a large-scale approach, which models the behavior across different social networks during a long period, in order to raise the level of consistency. This approach allows overcoming shortcomings of basic behavioral comparisons, due to the heterogeneity of behavior modelled across different social networks. Another crucial matching key proposed is the social network structure of a user.
Zafarani et al. [3] propose an attribute-independent user profile mapping approach, by exploiting redundant information that exists from user's behavioral patterns in social media sites. They argue that behavioral information is unique, due to a variety of

factors, such as user personality and others. Since user personality cannot be changed, this will lead to effective mapping approach. The second contribution is the use of machine learning techniques to increase the efficiency or accuracy of user identification.

Roedler et al. [20] exploite timestamps between user posts and geo-tags. They hypothesize that users use their social networks simultaneously; hence, if a user for example update his status on Facebook, he will do the same on Twitter. They use goe-tags to infer user's geographical area by calculating distances between these tags.

Table 1. Matching feature used by other approaches

Reference	Matching features (Interlinks)
Goga et al. [7]	Geo-location, timestamp of posts, writing style
Sha et al. [8]	User message (posts, tweets, retweets)
Zafarani et al. [1]	Usernames
Goga at al. [9]	Profile public-attributes
Bennacer et al. [18]	Network topology, public information
Van Le et al. [10]	Topics exists in user posts
Raad et al. [2]	Profile public-attributes
Jain et al. [11]	Public-attributes, social network, self-mentions (URIs)
Liu et al. [12]	Long-term behavioral analysis
Zafarani et al. [3]	Information redundancies in behavioral patterns
Nunes et al. [13]	Profile public-attributes
Motoyama et al. [4]	Profile public-attributes, email
Bartunov et al. [5]	Profile public-attributes, friendship links
Vosecky et al. [6]	Profile public-attributes
Shen et al. [14]	Public attributes, neighborhood features, quasi (inferred) features
Liang et al. [19]	Profile attributes, friendship links
Roedler et al. [20]	Timestamp of posts, device generated geo-tags
Panchenko et al. [17]	Usernames, friend lists
Nguyen et al. [16]	User public information
Peled et al. [15]	Profile public-attributes, network features
Jain et al. 2015 [27]	Historical values of attributes
Perito et al. [28]	Usernames
Szomszor et al. [30]	Tag-clouds
Iofciu et al. [31]	Usernames, tags
Malhotra et al. [34]	username, display name, location, profile image, and number of connections
Zhang et al. [37]	Local features: Usernames, language, URL, popularity. External features: location, avatar
Vosoughi et al. [35]	Language models, temporal activity

Based on the aforementioned state-of-the-art analysis, we propose two features life events and profiles biographies. In addition, we show that they can enhance each other.

3 SocialMatching++: A Novel Approach for Interlinking User Profiles in Social Networks

The approach we propose, called SocialMatching++, aims at linking user profiles from Facebook to their exact profiles on Twitter by using life events and biographies as two novel matching links.

3.1 SocialMatching++ Conceptual Model

SocialMatching++ is divided into: (1) LEBL (Life Event Based Linking) that link profiles using life events, and (2) DEBL (DEscription Based Linking) that link profiles based on profile biographies. Figure 2 shows a case study of SocialMatching++ (LEBL). It presents the user interaction on two different platforms (Facebook and Twitter) over time. We observe that user could have same life event (graduation) mentioned on both social networks, even if the content and behavior does not exists on Twitter.

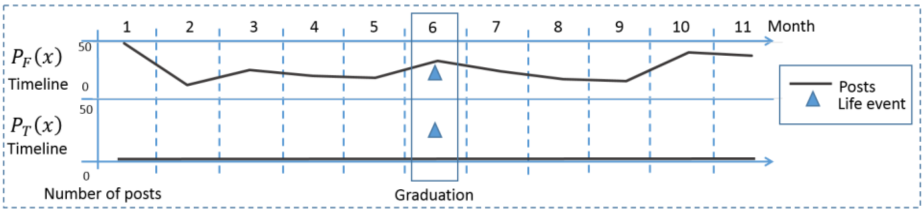


Fig. 2. Case study of user timelines show the existence of life events in the absence of profile content on Twitter, **which constitutes the key motivation of this research**

Life events

A life event post is not frequent. This means, it occurs in a specific circumstances, hence users who are not active on all their social networks and want to keep the audience updated, usually share these events, by writing a post that describe them or by adding them to their timeline. In our work, life events were extracted from the users' timelines on Facebook. Unlike Facebook, no formal representation of life events on Twitter is available. Consequently, we have to perform alternative mechanisms to detect them; named entity recognition is used to recognize entities inside event posts.

Profile descriptions

Profile description and biography are attributes that exist both on LinkedIn and Twitter. On LinkedIn, users can write a detailed description about themselves. However, Twitter descriptions are shorter. Recently, Facebook developed a new feature that allows users to define anything about themselves using characters. People can mention any thing inside it (hobbies, life events, biographies, etc). Users can mention many things inside like new job, hobbies, favorite food, etc.

3.1 SocialMatching++ Problem Formulation

Let P_F be a user profile on Facebook and P_T a user profile on Twitter. P_F is a known entity and it consists of a username (U) which is composed of a first name, last name, a list of life events L where ($P_F \leftarrow L$, U) and a description D. For each P_F we have to match the exact user names on Twitter P_T . F_m (1) is the matching function and F_s (2) the similarity function.

$$F_m(P_F, P_T) = \forall P_T \ cal \ F_s \ \left(L\left(P_T^i\right), L(P_F)\right) and \ F_s \ \left(D\left(P_T^i\right), D(P_F)\right) \tag{1}$$

F_s is the similarity function, and i is the total number of events on Facebook. We use the vector space model (cosine similarity) because we have a sequence of tokens to compare in both life events and biographies.

$$F_s = \cos(\theta) = \sum_{i=1}^{n} \frac{L,D\left(P_T^i\right)L\left(P_F^i\right)}{\sqrt{\sum_{i=1}^{n} L,D\left(P_T^i\right)^2} \sqrt{\sum_{i=1}^{n} L,D\left(P_F^i\right)^2}} \tag{2}$$

Two profiles are considered to be matched, if they have exact screenname matches and if the value of similarity score is higher than a predefined threshold *t*.

$$F_m(P_F, P_T) = \begin{cases} 1, \ if \ F_s > t \ and \ P_F.\text{screenname} = P_T.\text{screename} \\ \\ 0, \qquad\qquad\qquad\qquad\qquad\qquad otherwise \end{cases} \tag{3}$$

SocialMatching++ Implementation

The complete architecture of SocialMatching++ is detailed in Figure 3. It starts by retrieving screennames from Facebook, in parallel the set of life events and biographies corresponding to each user. After, we find the exact matching screennames from twitter. For each Twitter account, we search for a similar life events on their timeline and compare them to those on Facebook, as well as we do the same for biographies. Finally, we decide if two user profiles are linkable or not.

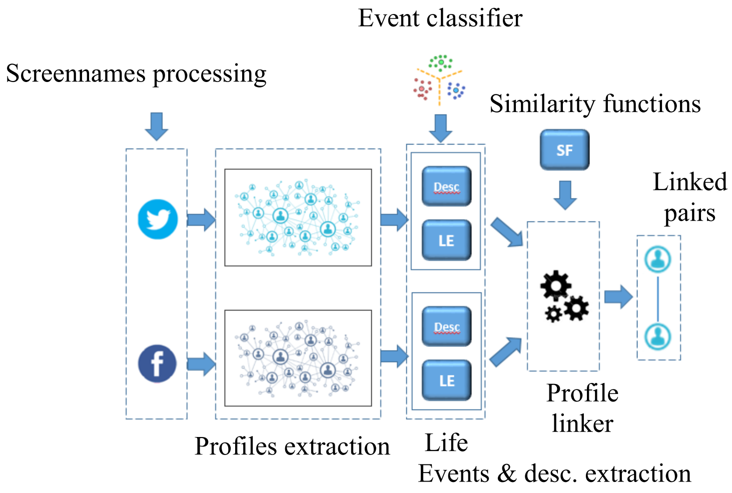


Fig. 3. SocialMatching++ Complete Architecture

Social Network Selection

In SocialMatching++, we decide to use Facebook and Twitter websites. Facebook is the largest social networking platform across the world, followed by Twitter. In our research, we observed that only 48.9% from the user profiles extracted from Facebook own a Twitter account. Facebook permits users to create structured life events that describe a certain circumstance, these events are commonly posted on the user's wall (timeline). Users also can post a new update (status) describing their life events. Contrariwise, Twitter does not provide any official feature through which users can update their life events. Twitter users can nevertheless post their own life events as Tweets.

Processing Screennames and Profile Extraction

Each user registered on a social network have to define a valid first name and last name, which are called screen names. People normally distinguish their names to prevent ambiguity with other users, through the modification of their usernames (a username is an id that can be accessed via the URL, e.g. facebook.com/userid). For instance, the exact screen name "Hussein Hazimeh" is available in 50 different profiles on Facebook, and each one has a different username. In plus, users differentiate themselves by adding a nickname to the original screen names (e.g. Hussein Hazimeh (PhD Student)), or to write it in two different languages. Screennames retrieval process starts by acquiring a set of screennames from the Facebook directory[1], this directory contains people, pages and place names sorted alphabetically, with the URL for each entity. All screennames extracted were composed of Latin characters for both Facebook and Twitter datasets. In Figure 2, we

[1] https://www.facebook.com/directory

started by querying Facebook with each screenname, after this we obtained a list of exact screennames from Twitter. The maximum number of matches were 18 for a

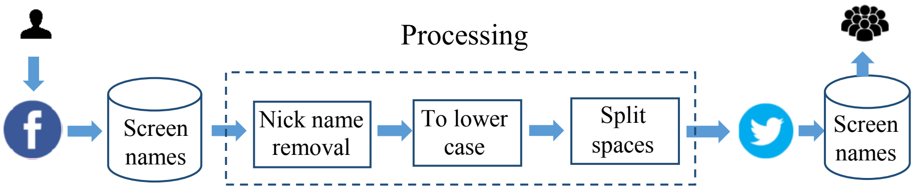


Fig. 4. Screennames Processing

single Facebook screenname. All the nicknames that are different from Latin are removed in the screennames processing phase.

Life Events Extraction

After all screennames are manipulated carefully, and matched them to the correspondent screennames on Twitter, we started in retrieving the life events for each username with the date for each one. The maximum number of events extracted per entity was seven.

Biography Extraction

A set of operations must be applied on biographies before storing them. Due to the messy content exists in many free texts, many special characters might exist, and that could diminish the matching performance. For this, we clean all biographies by removing stop words, stemming the text, remove special characters and defining entities inside the text.

Named Entity Extraction

Named entities are extracted from both life events and biographies. A named entity can be a person, organization, location or a university. In life events, user can mention their work place or organization name for example, as well as in biographies, same entities can be shared, in addition to persons. We used the state-of-the-art techniques to annotate entities. Conditional Random Fields (CRFs) are widely used models. Given a specific biography or life event text, applying CRF we can obtain and annotate all the entities inside this text.

Life Events Querying Mechanism: a Time Window Approach

Each Facebook life event consists of a set of entities and a date. We model the timeline of a Twitter user as a series of time windows, Figure 5. A single time window is a fixed interval of time when the data stream is processed for querying. Suppose that we have a Facebook life event *e* that has a specific date *d*. Our objective is to query the time window with date *d* and try to find if we have existing similar life events to *e*. The similarity of two life events represented by the function $S(e_1, e_2)$. *S* is positive if one of the two following scenarios is occurred: (1) if the value of F_S

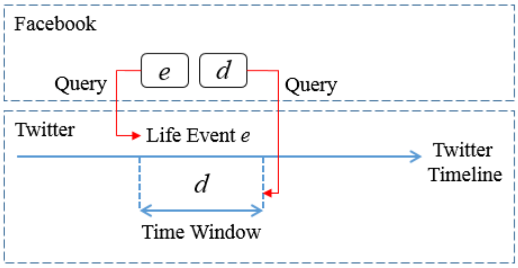


Fig. 5. Querying Twitter Time Window Using Specific Life Event and Date

Function (2) is greater than threshold, or if the similarity between two life events entities' is positive.

Profile linker Workflow

The first two inputs of the profile linker are: the Facebook screen name and its corresponding Twitter ids. We have two matching problems, the first one is a (*1 to 1*) matching problem. This exists when we have only one matched screen name on Twitter. The second one is a (*1 to n*) matching problem. This exists when we have more than 2 existing Twitter profiles. The matching procedure starts by comparing two biographies considering both named entities and matching score to take the matching decision. For e.g.

*Married one year ago. My life **Sarah***

*Love you **Sarah***

In the biographies (1) and (2), the similarity score is very low. However, if we consider common entities between them, we can observe that the user writes about his wife Sarah. In this case, we decide to link the two profiles rather than ignoring them.

If the description comparison returns a null result, we query each user's Twitter timeline with the named entities extracted from Facebook in a specific interval of time, Figure 5.

4 Experiments and Evaluation

In all our experiments, we used Facebook and Twitter websites. Alternative social networks like Instagram have more number of users compared to Twitter. However, Instagram allows people to share only images, which is not enough to conduct matching studies. LinkedIn contains a lot of information about users. But the platform is dedicated to and we do not have access to the user's timeline. Hence, even if the portion of connected users between Facebook and Twitter is not big, they remain two better choices due to the richness of information tweeted and posted on the user's timelines.

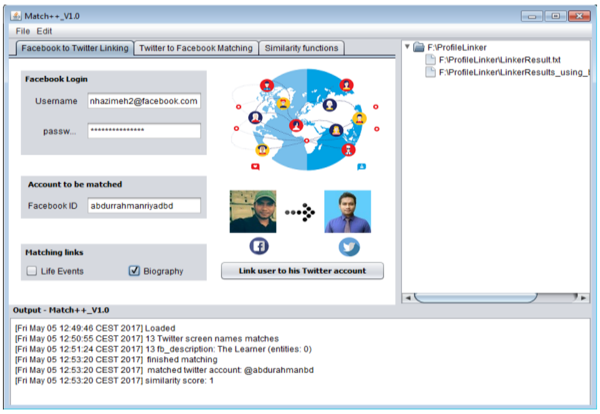


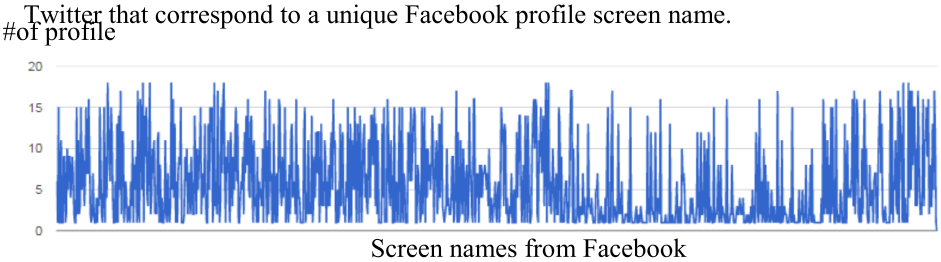
Fig. 6. Screenshot of Our System

We conducted tests on 2,263 different profiles having 6,630 life events and 1948 profile descriptions from Facebook matched with 5,694 Twitter users. The constructed Facebook dataset is open to the public and available for download (upon request). To implement our system we used Selenium web driver coded in Java using eclipse, and used a PC with 8GB of RAM and eight cores. The complete system code can be downloaded[2].

Dataset analysis and profiles selection

The 2,263 profiles from Facebook were selected using the following mechanism: we select two random profiles from Facebook that have a public friend list, each one of these two profiles has more than thousand friends. We have crawled these friend lists and for each user in this list we have extracted from his profile the public life events published on his timeline and his profile description. For each event, we have extracted its content and the exact date of publishing.

In Figure 7, a chart displays the total number of exact screen names matchings from Twitter that correspond to a unique Facebook profile screen name.



[2] https://github.com/HusseinHESSO/ProfileLinking_v1.0

Fig. 7. Number of matched screennames from Twitter

The maximum number of screen names matches is 18, and only 1,022 screen names were found on Twitter.

The maximum number of life events extracted from Facebook for each user is eight events. The total number of events extracted for each class is shown in table 2. For each life event, we define the named entity.

Table 2. Life events dataset analysis

Life event	Total number extracted	Named entities
Travelled	190	City name
Started a new job	1,481	Company name
Left his job	175	Company name
Graduated from a school/university	809	School/university name
Started school	2,261	School name
Moved to a new city	99	City name
Engaged	169	Person name
Published a new paper	1	Paper/conference name
Get married	278	Person name
In a new relationship	510	Person name
Left studying at university	363	University name
Get a new award	1	Award name

Evaluation Metrics

To evaluate our system, we use precision as a metric. In addition, we compare our system to two baseline systems.

With regard to life events, we chose one state-of-the-art system that compares the behavioural similarity between two user profiles. The system is HYDRA [12] published in 2014. HYDRA compares also long-term behavioural activity on large-scale datasets. We decide to use HYDRA because it is one of the most relevant and important contributions in this field.

Concerning biographies, we compare our system with systems that rely on profile attribute information, and prove that even if users do not share public attribute information, it is possible to link these profiles using biographies. @I seek 'fb.me'. [11] was the baseline compared with us because it uses a variety of profile attributes, however, it missing the biography.

Precision

We compare the precision of our system before and after defining entities. Figures 8 and 9 show how that the precision can be enhanced after defining entities using both the LEBL and DEBL approaches. We compare also the precision of our system with the baseline approaches, in Figures 10 and 11 and we show that both of our approaches are highly precise compared to the baselines. All the results shown in the figures take into account between 10 and 50 user profiles

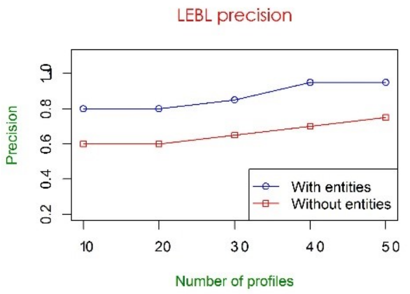


Fig. 8. LEBL precision with/without entities

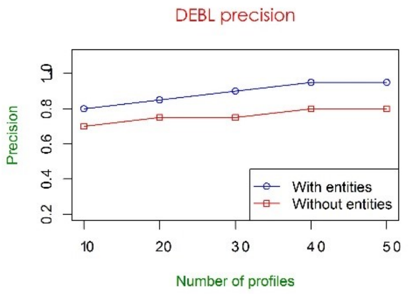


Fig. 9. DEBL precision with/without entities

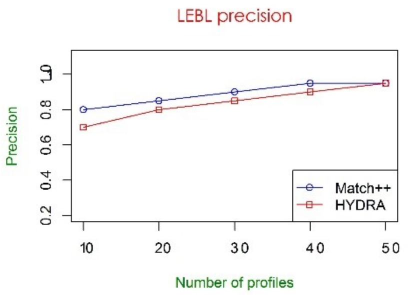


Fig. 10. LEBL precision compared to the baseline

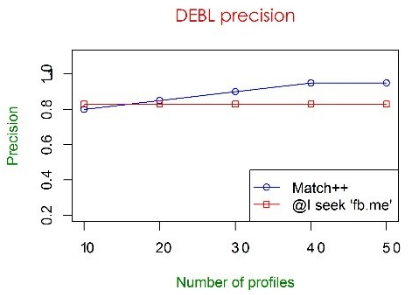


Fig. 11. DEBL precision compared to the baseline

Baseline systems comparison

DEBL Baselines

To compare our system with the two baseline systems, we select four random profiles that can fail to be linked using public attribute-based approaches, and show that they can be linked using profile descriptions. Table 2 show a set of four random profiles linked using biographies compared with @I seek 'fb.me' [11], providing the Twitter id and Facebook id of the user (we do not mention the screen names to respect the privacy of users).

Table 3. DEBL comparison with baselines

Facebook id	**Twitter id**	**@I seek 'fb.me'**	**DEBL**	**#of Twitter profiles**
stevenjong	@StevenJong	No	Yes	5

fernanda.vasconcelo.9822	@qbooomm	No	Yes	14
studioandrew	@andrewandraos	Yes (same image)	Yes	5
rodwell.mupungu	@minyango	Yes (same location)	Yes	1

As we observe in Table 3, our approach DEBL performs better than baselines. We can see that the baseline approaches succeed in linking users via image comparison. However, image matching can be more challenging than text matching. Some of them can link users via location too. Location can be same for different users. Hence, biographies can play a vital role when linking users.

LEBL Baselines

We also compare LEBL to the behavioural approaches. We select four random profiles, and show that information existing between users' timelines can vary in terms of content of posts, timestamps between posts and others. However, if we query the timeline with a specific type of information (Life events) we can detect that the two profiles belong to the same user. This can be enhanced in terms of accuracy (life events cannot be the same for two different users) and lack of activity in one of the two profiles. In Table 4, we show in detail the analysis of the four different profiles.

Table 4. LEBL comparison with baselines

Facebook id	Twitter id	HYDRA	LEBL	#of TW profiles
itspeterwish	@peter_wish	TW(textweets),FB(p hotos & videos)	Yes	2
hassan.mans our.528	@mansourhassanhm	TW(latin letters),FB(Arabic latters)	Yes	1
1406189719 414584	@IhabMortada	TW(Last public post 2015) FB(last public post 1 hour ago)	Yes	4
1015402229 4674058	@venessabassil	Similar hashtags	Yes	6

5 Conclusions and Future Works

In this work, we described SocialMatching++ a novel system that links user profiles on Facebook to their Twitter accounts, using life events (LEBL) and descriptions (DEBL). After conducting a comprehensive state-of-the-art on this research problem, we found that none of the related works has used these two features. We proved that our approach can achieve promising results. We concluded that even users do not share similar profile attribute information, such as images, locations; they can be matched using profile descriptions. Furthermore, users with different timeline behaviours and content can be matched using life events.

This work is strongly intended to match the topic of dataset profiling. The two built datasets can be later used for profiling reasons as a future work, motivating that dataset profiling in the context of social media is quite novel.

However, our work still has some limitations. Users are matched only from Facebook to Twitter, and not vice versa. Life events classification needs a more sophisticated, to detect more possible number of events on both social networks. As future work, we are working on enhancing the flexibility of our system. System users can define later the similarity functions based on their expertise, in addition to the flexibility of selecting the social network of interest. In order to test the scalability of our approach, we are working on building large-scale datasets and extend the number of social network channels (Google+ is under study).

References

1. R. Zafarani and H. Liu. Connecting corresponding identities across communities. ICWSM. 2009.
2. E. Raad, R. Chbeir, and A. Dipanda. User profile matching in social networks. In Network-Based Information Systems (NBiS). 2010.
3. R. Zafarani and H. Liu. Connecting users across social media sites: A behavioral-modeling approach. SIGKDD, ACM 2013.
4. M. Motoyama and G. Varghese. I seek you: searching and matching individuals in social networks. In WIDM, 2009.
5. S. Bartunov, A. Korshunov, S-Taek Park, W. Ryu, and H. Lee. Joint link-attribute user identity resolution in online social networks. SNAKDD Workshop, 2012.
6. J. Vosecky, D. Hong, and V.Y. Shen. User identification across multiple social networks. NDT '09. 2009.
7. O. Goga, H. Lei, S. Hari Krishnan, G. Friedland, R. Sommer and R. Teixeir. Exploiting Innocuous Activity for Correlating Users Across Sites. WWW Conference 2013.
8. Y. Sha, Q. Liang and K. Zheng. Matching User Accounts across Social Networks based on User Message. ICCS 2016.
9. O. Goga, P. Loiseau, R. Sommer, R. Teixeira and K. P. Gummadi. On the Reliability of Profile Matching Across Large Online Social Networks. KDD 2015.
10. T. Van Le, T. Nghia Truong and T. Vu Pham. A Content-Based Approach for User Profile Modeling and Matching on Social Networks. MIWAI 2014.
11. P. Jain, P. Kumaraguru, Anupam Joshi. @i seek 'fb.me': identifying users across multiple online social networks. WWW (Companion Volume) 2013.
12. S. Liu, S. Wang, F. Zhu, J. Zhang, R. Krishnan. HYDRA: large-scale social identity linkage via heterogeneous behavior modeling. SIGMOD 2014.
13. A. Nunes, P. Calado, B. Martins. Resolving user identities over social networks through supervised learning and rich similarity features. SAC 2012.
14. Y. Shen, H. Jin. Controllable Information Sharing for User Accounts Linkage across Multiple Online Social Networks. CIKM 2014.
15. O. Peled, M. Fire, L. Rokach, Y. Elovici. Matching entities across online social networks. Neurocomputing 210. 2016

16. N. Hoang Long, J. J. Jung. Privacy-Aware Framework for Matching Online Social Identities in Multiple Social Networking Services. Cybernetics and Systems 46(1-2) 2015.
17. A. Panchenko, D. Babaev, S. Obiedkov. Large-Scale Parallel Matching of Social Network Profiles. AIST 2015.
18. N. Bennacer, C. Nana Jipmo, A. Penta, G. Quercini. Matching User Profiles Across Social Networks. CAiSE 2014.
19. W. Liang, B. Meng, X. He, Xianchao. GCM: A Greedy-Based Cross-Matching Algorithm for Identifying Users Across Multiple Online Social Networks. PAISI 2015.
20. R. Roedler, D. Kergl, G. Dreo Rodosek. Profile Matching Across Online Social Networks Based on Geo-Tags. NaBIC 2015.
21. O. Goga. Matching user accounts across online social networks : methods and applications. PhD thesis, Pierre and Marie Curie University, Paris, France 2014
22. L. Gu, R. Baxter, D. Vickers, and C. Rainsford. Record Linkage: Current Practice and Future Directions. CSIRO 2003
23. H. Noureddine, I. Jarkass, H. Hazimeh, A. Abou Khaled, E. Mugellini. CARP: Correlation Based Approach for Researcher Profiling. SEKE 2015.
24. W. W. Cohen, P. Ravikumar, S. E. Fienberg. A Comparison of String Distance Metrics for Name-Matching Tasks. IIWeb 2003.
25. P. Jain, P. Kumaraguru. Other Times, Other Values: Leveraging Attribute History to Link User Profiles across Online Social Networks. ACM (HT) 2015.
26. D. Perito, C. Castelluccia, M. Ali Kâafar, P. Manils. How Unique and Traceable Are Usernames? Privacy Enhancing Technologies . PETS 2011.
27. Elmagarmid, A. K., Ipeirotis, P. G., and Verykios, V. S. Duplicate record detection: A survey. IEEE Transactions on Knowledge and Data Engineering 19, 2007.
28. M. Szomszor, I. Cantador, H. Alani Correlating user profiles from multiple folksonomies. Hypertext 2008.
29. T. Iofciu, P. Fankhauser, F. Abel, K. Bischoff. Identifying Users Across Social Tagging Systems. AAAI 2011.
30. R. Nayak, M. Zhang, L. Chen. A Social Matching System for an Online Dating Network: A Preliminary Study. ICDM Workshops 2010
31. A. Krzywicki, W. Wobcke, Y. Sok Kim, Xiongcai Cai, M. Bain, A. Mahidadia, P. Compton. Collaborative Filtering for people-to-people recommendation in online dating. Int. J. Hum.-Comput. Stud. (2015)
32. X. Cai, M. Bain, A. Krzywicki, W. Wobcke, Y. Sok Kim, P. Compton, A. Mahidadia. Learning Collaborative Filtering and Its Application to People to People Recommendation in Social Networks. ICDM 2010.
33. J. S. Breese, D. Heckerman, C. Myers Kadie. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. UAI 1998.
34. A. Malhotra, L. C. Totti, W. Meira Jr., Ponnurangam Kumaraguru, Virgílio A. F. Almeida. Studying User Footprints in Different Online Social Networks. ASONAM 2012.
35. S. Vosoughi, H. Zhou, D. Roy. Digital Stylometry: Linking Profiles Across Social Networks. Social Informatics - 8th International Conference. SocInfo 2015.
36. G. Sujatha Das, X. Li, A. Sun, H. Kardes, X. Wang. Person-Name Parsing for Linking User Web Profiles. WebDB 2015
37. H. Zhang, M-Yen Kan, Y. Liu, S. Ma. Online Social Network Profile Linkage. AIRS 2014
38. H. Kopcke, E. Rahm. Frameworks for entity matching: A comparison. Data and Knowledge Engineering, 2009.
39. Thomas Dickinson, Miriam Fernández, Lisa A. Thomas, Paul Mulholland, Pam Briggs, Harith Alani: Automatic Identification of Personal Life Events in Twitter. WebSci 2015