Friends' Recommendations in Social Networks: An Online Lifestyles Approach

KLIMIS NTALIANIS

Department of Marketing
Athens University of Applied Sciences (TEI of Athens)
Agiou Spyridonos, Egaleo, Athens
GREECE
kntal@teiath.gr

Abstract: - Several of the existing major social networking services such as Facebook and Twitter, recommend friends to their users based on social graphs analysis, or using simple friend recommendation algorithms such as similarity, popularity, or the "friend's friends are friends," concept. However these approaches, even though intuitive and quick, they consider few of the characteristics of the social networks, while they are typically not the most appropriate ways to reflect a user's preferences on friend selection in real life. To overcome these problems in this paper a novel scheme is proposed for recommending friends in social media, based on the analysis and vector mapping of online lifestyles. In particular for each user a vector is created that captures her/his online behavior. Then, in the simple case, vector matching is performed so that the top matches are selected as potential friends. In a more sophisticated case, the most similar profiles to the user under investigation are detected and a collaborative recommendations approach is proposed. Experimental results on real life data exhibit the promising performance of the proposed scheme.

Key-Words: - Friends' recommendations, social networks, social life style, social computing

1 Introduction

Social networks have experienced an explosive growth during the last decade. Currently billions of users share opinions, photos and videos every day, while many of them are looking for new online friends. Thirty years ago, this attitude was only described in science fiction stories, while people typically made friends with others who lived or worked close to themselves, such as neighbors or colleagues. However the Internet came to abolish geographical location-based limitations and open new horizons also to personal relations. For example there are several articles about falling in love through social media acquaintances [1]. Additionally most people act differently online than they do in real life [2], including making new friendships and/or relationships. This mainly has to do with the sense of safety: many online friendships are just remote and thus the dangers of abuse, bad treatment or even sexual assault are absent. It needs a real meeting in the real world so that the aforementioned and other problems and crimes can happen. This sense of safety makes people more and more interested in finding new friends from all over the world. However one very challenging issue of this research area is how to help users to efficiently detect new potential social friends.

Towards this direction, current schemes rely on pre-existing user relationships. For example, Facebook performs social links analysis among those who already share common friends and recommends users as potential friends. Unfortunately, this approach may not be the most appropriate based on sociology findings [3]-[5]. According to these studies, the rules to group people together include: 1) habits or life style; 2) attitudes; 3) tastes; 4) moral standards; 5) economic level; and 6) people they already know. Apparently, the third rule and the sixth rule are mainly considered by existing recommendation systems. The first rule, although probably the most intuitive, is not widely used because users' life styles are difficult, if not impossible, to capture through web actions.

For this reason in this paper we focus on online lifestyles, since online lifestyles are often different from real lifestyles and online friends may be different from real world friends. Thus we aim at automatically building a profile for each social media user. This profile should capture the online lifestyle of the user under investigation. Towards this direction for each social media user several factors are estimated such as the total number of friends the user has, the average number of likes per day the user makes, the average number of multimedia posts per day the user makes, the average time per day the user spends on social media etc. We claim that these factors characterize the online lifestyle of each user under investigation. Then a vector is created for each user that ideally represents her/his online behavior. Afterwards, in the simple case, vector matching is performed so that the top matches are selected as potential friends, while in a more sophisticated case the most similar profiles to the user under investigation are detected and a collaborative recommendations approach is proposed. Experimental results on real life data exhibit the advantages and disadvantages of the proposed scheme.

The rest of this paper is organized as follows: in Section 2 state-of-art approaches are presented. In Section 3 the details of the proposed friends' recommendations scheme are presented, while experimental results are provided in Section 4. Finally section 6 concludes this paper.

2 Previous Work

Recently, with the advance of social networking platforms, friends' recommendations have received great attention. Existing algorithms incorporated by Facebook, LinkedIn, Twitter, etc. recommend friends according to social relations (common friends).

Meanwhile, other recommendation mechanisms have also been proposed by researchers. For example in [6] MatchMaker is presented, which is a collaborative filtering friends' recommendations system based on personality matching. In [7] a friends' recommendations method is proposed that uses physical and social context. However, the authors do not explain what the physical and social context is and how to obtain the necessary information. In [8] a scheme is proposed that recommends geographically related friends in social networks by combining GPS information and social network structure.

On the other hand there are some fundamental works which try to perform activity recognition and thus build a lifestyle profile of each user, by incorporating smartphones. In [9] and [10] the authors tried to discover daily location-driven routines from large-scale location data and also combined location and physical proximity sensed by the mobile phone. The work in [11] has been inspired from these approaches. In particular Friendbook is presented, which recommends friends to users based on their life styles. The scheme takes advantage of sensor-rich smartphones and discovers life styles of users from user-centric sensor data. Then, by measuring the similarity of life styles between users, it recommends friends. In [12] the structure of social networks is investigated and an algorithm for network correlation-based social friends' recommendations is developed. Different "social role" networks, are also considered and their relationships are examined. In [13] users' daily lives are modeled as life documents, from which their life styles are extracted using the Latent Dirichlet Allocation model. Furthermore a similarity metric is proposed to measure the similarity of life styles between users. In [14] life documents of each user are collected from the client with the help of a browser. The life styles of users are extracted by using either Hadoop technology or SQL depending on the type of file. Recommendations are based on different properties like similar interests, similar blood group, nearby locations etc. In [15] a big data analytics solution that uses the MapReduce model is proposed. The work aims at mining these big social networks for discovering groups of frequently connected users, so that friends' recommendations are made. In [16] a study is presented of the community structure of ego-networks. Toward this goal, a technique to efficiently build and cluster all the ego-nets of a graph in parallel is designed. In [17] an approach that recommends friends with similar location preference is proposed, in which both the online friendship information and the offline user behavior are considered. This approach uses Markov chains, cosine similarity and threshold evaluation, while the effectiveness of the algorithm is verified on a real dataset. Finally in [18], a friends' recommendations method with two stages is proposed. In the first stage the information of the relationship between texts and users, as well as the friendship information between users is utilized, and some "possible friends" are chosen. In the second stage, a topic model is built to further refine the recommendation results.

3 Description of Social Lifestyles and the Proposed Algorithm

Let $U = \{1, 2, \ldots, N_U\}$ be the index set of all users of a social network, which has N_U users in total. Let also u_i be the i^{th} user of this social network. Users usually have specific characteristics and perform several activities. In order to describe in a typical way the online lifestyle of each user, in this paper several factors are considered: (a) the total number of friends TF_i that u_i has, (b) the average number of likes per day AL_i that u_i makes, (c) the average number of loves per day ALv_i that u_i makes, (d) the average number of hahas per day AH_i that u_i makes, (e) the average number of wows per day AW_i that u_i makes, (f) the average number of sads per day Asd_i that u_i makes, (g) the average number of angrys per day Aas_i that u_i makes, (h) the average number of comments per day AC_i that u_i makes, (i) the average number of shares per day AS_i that u_i makes, (j) the average number of text posts per day Atp_i that u_i makes, (k) the average number of multimedia posts per day AM_i that u_i makes, (l) the average number of friends per day AF_i that u_i makes, (m) the average number of deleted friends per day AdF_i for u_i, (n) the average number of total actions per day AA_i that u_i makes, (o) the total number of groups GP_i that u_i has joined, (p) the average number of groups per day AG_i that u_i joins, (q) the average number of deleted groups per day Agd_i for u_i and (r) the average time per day AT_i that u_i spends on the social network under investigation. By gathering all this online lifestyle information, a feature vector $\mathbf{f}_i$, $i= 1, 2, \ldots, N_U$ is formulated for each social media user:

$$\begin{aligned}\mathbf{f}_i = [&n(TF_i), n(AL_i), n(ALv_i), n(AH_i), n(AW_i), n(ASd_i),\\ &n(AAs_i), n(AC_i), n(AS_i), n(ATp_i), n(AM_i), n(AF_i),\\ &n(AdF_i), n(AA_i), n(GP_i), n(AG_i), n(AGd_i), n(AT_i)]\end{aligned} \tag{1}$$

3.1 Simple Case: User to User Correlation

In the simple case of friends' recommendations, the correlation between the feature vector of user u_i and the feature vector of user u_j is estimated for any $i \neq j$. Users with the highest correlation values are recommended as friends to user u_i. In particular the correlation coefficient between two feature vectors (representing two users) is defined as [19]:

$$\rho(\mathbf{f}_i, \mathbf{f}_j) = \frac{C(\mathbf{f}_i, \mathbf{f}_j)}{\sqrt{C(\mathbf{f}_i, \mathbf{f}_i)} \cdot \sqrt{C(\mathbf{f}_j, \mathbf{f}_j)}} \tag{2}$$

with

$$C(\mathbf{f}_i, \mathbf{f}_j) = (\mathbf{f}_i - \mathbf{m})^T(\mathbf{f}_j - \mathbf{m}) \tag{3}$$

and

$$\mathbf{m} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{f}_i \,, \tag{4}$$

where $C(\mathbf{f}_i, \mathbf{f}_j)$ is the covariance between $\mathbf{f}_i$ and $\mathbf{f}_j$, and $\mathbf{m}$ is the average feature vector of the group of social media users under investigation.

3.2 Collaborative Recommendations Case

This case is significantly different from the simple case of subsection 3.1, since friends' recommendations are based on the opinions of other users. More specifically, initially users with similar profiles to the user under investigation are detected. Detection is accomplished by the method proposed in subsection 3.1.

Let us now assume that users with similar profiles to the user under investigation have already evaluated several profiles of recommended users in previous recommendation sessions (e.g. by sending friends request to the recommended users, by expressing an interest to become friends with the recommended users, by poking the recommended users etc.). In this case each recommended user (which has already been recommended to other users in previous recommendation sessions but not to the user under investigation) receives a score. The score is calculated for each recommended user by aggregating the ratings provided by users with similar – to the user under investigation – profiles.

$$Ar_{ru_i} = aggr_{u_j \in SU} rt_{u_j, ru_i} \tag{5}$$

where: (a) ru_i is the ith recommended user, (b) Ar_{ru_i} denotes the aggregated rating of the ith recommended user, (c) SU is the set of top "N" users that are most similar to the user u under investigation and they have also rated the ith recommended user and (d) rt_{u_j, ru_i} is the rating score of the ith recommended user provided by user u_j.

In this paper the aggregation function is expressed as:

$$Ar_{ru_i} = \overline{rt_u} + k\sum_{u_j \in SU} sim(u, u_j)\left(rt_{u_j, ru_i} - \overline{rt_{u_j}}\right) \tag{6}$$

where k is a normalizing factor defined as:

$$k = \frac{1}{\sum_{u_j \in SU} \left| sim(u, u_j) \right|} \tag{7}$$

$\overline{rt_u}$ is the average ratings value provided by user u under investigation and for all recommended users (it refers to ratings that have been given in previous recommendation sessions). Similarly $\overline{rt_{u_j}}$ is the average ratings value provided by user u_j in previous recommendation sessions. Furthermore $sim(\cdot)$ is a function that estimates the similarity of the two arguments of the function (i.e. users u and u_j). This neighborhood-based algorithm calculates the similarity between two users (the user u under investigation and user u_j who belongs to the top N most similar users compared to u) and produces a prediction for the ith recommended user, by taking the weighted average of all the ratings. Similarity computation between users is an important part of this approach. Again similarity can be easily computed by Eq. (2). However other sophisticated measures can also be incorporated such as Pearson's correlation, which takes into consideration the previous ratings and not each user's profile information:

$$sim(u,u_j)=\frac{\sum_{ru_i\in I_{u,u_j}}(rt_{u,ru_i}-\overline{rt_u})((rt_{u_j,ru_i}-\overline{rt_{u_j}}))}{\sqrt{\sum_{ru_i\in I_{u,u_j}}(rt_{u,ru_i}-\overline{rt_u})^2\sum_{ru_i\in I_{u,u_j}}(rt_{u_j,ru_i}-\overline{rt_{u_j}})^2}} \quad (8)$$

where I_{u,u_j} is the set of recommended users rated by both user u under investigation and user u_j during previous recommendation sessions.

Finally cosine-based similarity can also be computed between users u and u_j:

$$sim(u,u_j)=\cos(\vec{u},\vec{u}_j)=\frac{\vec{u}\cdot\vec{u}_j}{\|\vec{u}\|\times\|\vec{u}_j\|}=$$

$$=\frac{\sum_{ru_i\in I_{u,u_j}}rt_{u,ru_i}\cdot rt_{u_j,ru_i}}{\sqrt{\sum_{ru_i\in I_{u,u_j}}rt^2_{u,ru_i}}\cdot\sqrt{\sum_{ru_i\in I_{u,u_j}}rt^2_{u_j,ru_i}}} \quad (9)$$

4 Experimental Results

In order to evaluate the performance of the proposed social media friends' recommendation scheme, we have setup an experimentation phase on real world Facebook profiles. In particular we have recorded the activities and profiles' information of the 219 members of the Online Computing Group (Onlog) (www.facebook.com/klimis.ntalianis.7) for a period of two months (1 February 2017 – 31 March 2017). Onlog is part of the Department of Marketing of the Athens University of Applied Sciences (TEI of Athens) and its members mainly include students in the age group of 18 – 26 years. Furthermore six of the profile pictures (for privacy reasons pictures have been processed so that they do not reveal any clues about the real user) are presented in Figure 1.

Here it should be mentioned that the parameters *TF* and *GP* have been measured for each user on the 20th of March 2017. Furthermore a feature vector $\mathbf{f}_i$, i= 1, 2, …, 219 has been created for each member. In particular the average *TF* was 465.46, the average *AL* was 5.23, the average *Alv* was 2.11, the average *AH* was 1.15, the average *AW* was 1.29, the average *Asd* was 0.61, the average *Aas* was 0.73, the average *AC* was 2.25, the average *AS* was 0.08, the average *Atp* was 1.12, the average *AM* was 3.31, the average *AF* was 2.15, the average *AdF* was 0.03, the average *AA* was 15.67, the average *GP* was 14.15, the average *AG* was 0.01, the average *Agd* was 0.01 and the average *AT* was 333.16 (minutes). *AT* was high since several of the participating members are "always" connected to Facebook, through their smartphones, even though there are sufficiently long intervals of no activity. Furthermore the average number of friends (*TF* = 465) is also higher than the global average, since younger ages tend to use Facebook more than the older age groups.

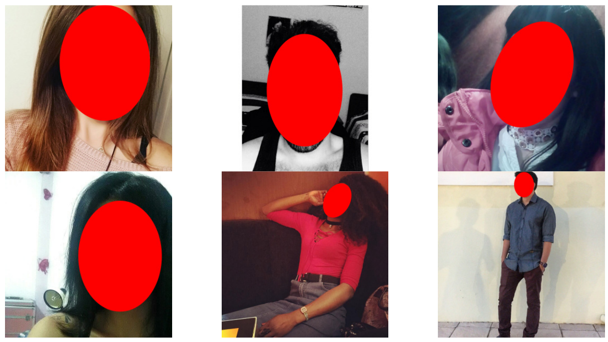

Figure 1: Six of the profile pictures of the users under investigation, regarding friends' recommendations.

Now in order to evaluate the proposed scheme, we use the precision and recall metrics. In particular:

$$precision=\frac{\text{correctly recommended friends}}{\text{all recommended friends}} \quad (10)$$

$$recall=\frac{\text{correctly recommended friends}}{\text{all existing friends}} \quad (11)$$

In case of precision, the correctly recommended friends are those profiles which really attract the interest of the user under investigation. In a real world case the user under investigation could send a friends request to the recommended friend, could express her/his interest to become friends by sending e.g. a message or interacting with a post created by the recommended friends, by poking the recommended friends etc. Since in this phase we are interested in examining the behaviors of users against the new friends' recommendations scheme, in the current experiments the users under investigation expressed their interest to become friends with a recommended friend by answering a questionnaire. More specifically 3,000 open Facebook profiles have been crawled and a vector $\mathbf{f}_i$ has also been formed for each profile. In Figure 2 six processed profile pictures are presented. Additionally all existing friends in the recall measure represent the total number of friends (*TF* parameter) a user under investigation has.

Figure 2: Six of the profile pictures of the 3,000 possibly recommended friends.

After creating a vector for each profile, we have estimated the correlation coefficient (Eq. 2) between each profile under investigation and the 3,000 open profiles. In Figure 3 the normalized correlation coefficient for user u_{64} is presented for the top 200 matches among the 3,000 profiles.

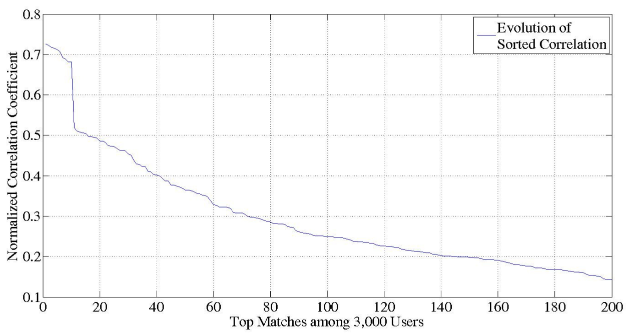

Figure 3: Evolution of the sorted normalized correlation coefficient for the user u_{64} (top 200 matches among 3,000 possibly recommended friends).

Next for each user – among the 219 available – under investigation, 30 recommendation sessions have been carried out. In each session one friend was recommended to one user under investigation. The 30 recommended friends among the 3,000 available were those that presented maximum correlation to the user under investigation.

The recommendation was visualized to the user under investigation by the profile picture of the recommended friend as well as the 10 most recent items that the recommended friend has posted to her/his wall. Then the user under investigation was asked whether she/he was/was not interested to become friends with the recommended friend. Figure 4 presents the precision diagram for all 219 users under investigation while Figure 5 provides the recall diagram. The average precision was 0.18 (or 5.55 recommended friends per user under investigation among the presented 30) while the average recall was 0.013.

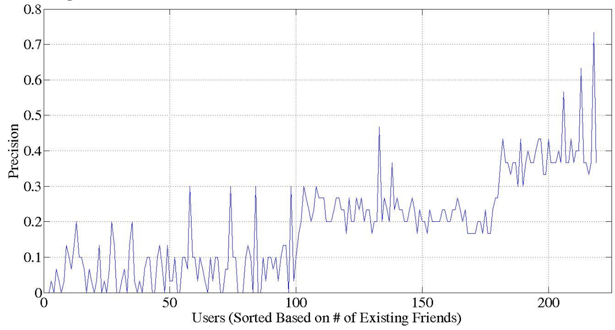

Figure 4: The precision diagram for all 219 users under investigation. Users have been sorted according to their numbers of existing friends.

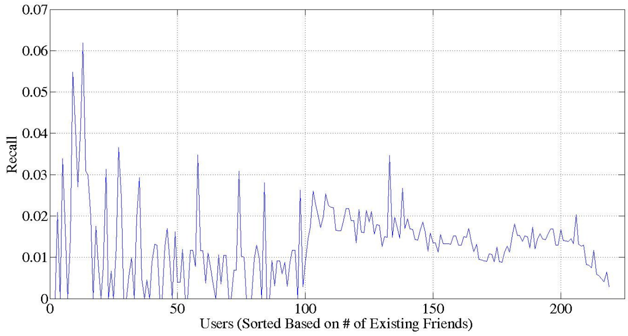

Figure 5: The recall diagram for all 219 users under investigation. Users have been sorted according to their numbers of existing friends.

As it can be observed it seems that people are not very willing to become Facebook friends with completely unknown persons, possibly from other countries, and probably with different socioethnical characteristics, even if those people follow similar online lifestyles. Of course in other types of social networks where Professional "Friends" are made (e.g. LinkedIn), maybe this observation is not so valid. However some stereotypes may still be valid. In any case research on each social network should be carried out. Additionally users with more online friends seem to be less reluctant to make new online friends, for example the users with more than 1,000 online friends (11 users in total), were willing to make 13 new friends on average, from the set of their 30 recommended friends. This is a significant difference compared to the overall average of 5.5 new friends among the recommended.

5 Conclusion

In this paper an online lifestyles-based friends' recommendation scheme is proposed. Different from the friends' recommendations mechanisms relying on social graphs, the proposed scheme maps the online lifestyles of each user into a vector. Then correlation among vectors is performed and the top matches are recommended as possible friends. In a more sophisticated case, users with similar profiles are considered and their choices of recommended friends in previous recommendation sessions are analyzed. Experimental results on real data exhibit the promising performance of our scheme.

In the future other types of SNs should be investigated, such as LinkedIn, where professional "friendships" are met. Furthermore more behavioral features should be taken into consideration. Finally comparisons to other relevant or different philosophy schemes should be carried out.

References:

[1] "Why it's possible to fall in love online," http://edition.cnn.com/2013/02/14/tech/social-media/online-love/, by Brenna Ehrlich, Feb. 2013, last accessed: 3 April 2017.

[2] "Facebook Psychology: 7 Reasons Why We Act Differently Online," http://www.huffingtonpost.com/2012/10/11/facebook-psychology-7-reasons_n_1951856.html, Nov. 2012, by Britney Fitzgerald, last accessed: 3 April 2017.

[3] I. Røpke, "The Dynamics of Willingness to Consume," *Ecological Economics*, Vol. 28, No. 3, p.p. 399–420, 1999.

[4] G. Spaargaren and B. Van Vliet, "Lifestyles, Consumption and the Environment: The Ecological Modernization of Domestic Consumption," *Environmental Politics*, Vol. 9, No. 1, p.p. 50–76, 2000.

[5] M. Tomlinson, "Lifestyle and Social Class," *European Sociological Review*, Vol. 19, No. 1, p.p. 97–111, 2003.

[6] L. Bian and H. Holtzman, "Online friend recommendation through personality matching and collaborative filtering," In Proc. Of *UBICOMM*, p.p. 230–235, 2011.

[7] J. Kwon and S. Kim, "Friend recommendation method using physical and social context," *International Journal of Computer Science and Network Security*, Vol. 10, No. 11, p.p. 116–120, 2010.

[8] X. Yu, A. Pan, L.-A. Tang, Z. Li, and J. Han, "Geo-friends recommendation in gps-based cyber-physical social network," In Proc. of *ASONAM*, p.p. 361–368, 2011.

[9] K. Farrahi and D. Gatica-Perez, "Discovering Routines from Largescale Human Locations using Probabilistic Topic Models," *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 1, 2011.

[10] K. Farrahi and D. Gatica-Perez, "Probabilistic mining of sociogeographic routines from mobile phone data," *IEEE Jour. Select. Topics Sign. Proc.*, Vol. 4, No. 4, p.p. 746–755, 2010.

[11] Z. Wang, J. Liao, Q. Cao, H. Qi, "Friendbook: a semantic-based friend recommendation system for social networks," *IEEE Trans. Mobile Comput.*, Vol. 14, p.p. 538–551, 2015.

[12] S. Huang, J. Zhang, L. Wang, and X.-S. Hua, "Social Friend Recommendation Based on Multiple Network Correlation," *IEEE Trans. Multimedia*, Vol. 18, No. 2, Feb. 2016.

[13] N. M. Eklaspur, A. S. Pashupatimath, "A Friend Recommender System for Social Networks by Life Style Extraction Using Probabilistic Method – Friendtome," *International Journal of Computer Science Trends and Technology*, Vol. 3, No. 3, May-June 2015.

[14] T. R. Kacchi, and A. V. Deorankar, "Friend recommendation system based on lifestyles of users", *AEICB*, 2016.

[15] F. Jiang, C. K. Leung, and A. G. M. Pazdor "Big data mining of social networks for friend recommendation," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 324 – 335, 2016.

[16] A. Epasto, S. Lattanzi, V. S. Mirrokni, I. Sebe, A. Taei, and S. Verma, "Ego-net community mining applied to friend suggestion," *PVLDB*, Vol. 9, No. 4, p.p. 324 – 335, Dec. 2015.

[17] M. Wum, Z. Wang, H.Sun, and H. Hu, "Friend Recommendation Algorithm for Online Social Networks Based on Location Preference", *ICSCE*, 2016.

[18] S. Huang, J. Zhang, D. Schonfeld, L. Wang, and X.-S. Hua, "Two-Stage Friend Recommendation Based on Network Alignment and Series-Expansion of Probabilistic Topic Model," *IEEE Transactions on Multimedia*, Vol. PP, No. 99, Jan. 2017.

[19] A. Papoulis, Probability, Random Variables, and Stochastic Processes. New York: McGraw Hill, 1984.