Alike people, alike interests? A large-scale study on interest similarity in social networks

Xiao Han, Leye Wang, Soochang Park, Angel Cuevas Rumin, Noel Crespi

Alike People, Alike Interests? A Large-scale Study on Interest Similarity in Social Networks

Xiao Han*, Leye Wang*, Soochang Park*, Ángel Cuevas†* and Noël Crespi*
* Institut-Mines Télécom, Télécom SudParis,
{han.xiao, leye.wang, soochang.park noel.crespi}@telecom-sudparis.eu
† Universidad Carlos III de Madrid,
{acrumin}@it.uc3m.es

***Abstract*—In this paper, we present a comprehensive empirical study on the correlations between users' interest similarity and various social features across three interest domains (i.e., movie, music and TV). This study relies on a large dataset, containing $479,048$ users and $5,263,351$ user-generated interests, captured from Facebook. We identify the social features from three types of the users' information - demographic information (e.g., age, gender, location), social relations (i.e., friendship), and users' interests. The results reveal that the interest similarity follows the homophily principle, which could be further harnessed by various practical applications and services.**

I. Introduction

Online Social Networks (OSNs) have boomed and attracted a huge number of people to join them over the last decade. In OSNs, the participants publish their profiles, make friends, and produce various content (photos, answers/questions, videos, etc.). Unlike legacy web systems, OSNs are organized around both people and content, which provide us with unprecedented opportunities to understand human relationships, human communities, human behaviors and human preferences[1] [2] [3].

With the evolution of OSNs, understanding to what extent that two individuals are alike in their interests (i.e., interest similarity) has become a basic requirement for the organization and maintenance of vibrant OSNs. On the one hand, users' interest similarity could be leveraged to support friend recommendation and social circle maintenance. For instance, the decision to recommend the users who share many interests with each other to be friends could increase users' approval rate of recommendation, because people usually aggregate by their mutual interests [4]. On the other hand, knowing interest similarity between users also facilitates social applications and advertising. For example, instead of randomly hunting clients, exploring those users of a high interest similarity with the existing clients could efficiently enlarge client groups for application providers and businesses.

Although many previous studies have been widely conducted on various OSN platforms, most of them have only focused on discovering various structural properties including the small world effect, community structures, and clustering [2] [3] [4]. Such investigations could not be directly applied to the above-mentioned applications (e.g., personalized advertisements). Aiming to enhance specific social-based services and applications (e.g., friend prediction, recommendation), several existing researches have already examined how interest similarity changes with very limited social features: [5] [6] has studied that friends share more interests than strangers; [7] has verified that interest similarity strongly correlates to the trust between users. While none of them has extended this study to discuss how interest similarity varies with various social features.

Therefore, in this paper, we are motivated to carry out empirical studies on how users' interest similarity relates to various social features in a wide variety of cases. In particular, we quantify interest similarity over an aggregation of user pairs based on a cosine method to capture interest overlaps between two users. Besides, we extract the social features (e.g. profile similarity, geographic distance, friend similarity) from users' social information regarding three aspects: demographic information (e.g., age, gender, location), social relations (i.e., friendship), and users' interests. Specifically, we conduct the study in three interest domains, namely movie, music and TV, over a large dataset including $479,048$ users and $5,263,351$ user-generated interests crawled from Facebook.

To highlight our key findings, we reveal the homophily regarding interest similarityin Facebook based on the comprehensive analysis. Generally, homophily shows homogeneity in people's social networks regarding many sociodemographic, behavioral and intrapersonal characteristics [8]. Specifically, in this paper,

- homophily reveals that people are more likely to be interested in the same movie, music and TV series when they are more similar in their demographic information, such as age, gender and location;
- homophily also implies that friends have higher interest similarity than strangers do. Furthermore, the interest similarity becomes higher if two users share more common friends;
- in addition, homophily indicates that the users with a larger interest individuality are likely to share more interests with each other. Note that we define interest individuality to quantify the personalized characteristics of individual interests. A user's interest individuality is affected by two factors: the total number of a user's interests and the popularity of these interests. The more interests a user presents, and the more popular the interests are, the higher interest individuality the user gains.

This study is distinct from the existing work on interest similarity by three aspects. Firstly, we carry out a more comprehensive analysis on the correlations between users' interest similarity and diverse social features. We attempt to dig out

more relative factors which can be harnessed to enhance social recommendations and advertisement services. Secondly, the majority of existing studies on interest have not distinguished the different types of interests - they usually relied merely on users' favorite music or movies [4]. Additionally, they typically measure interests in terms of genre. In this paper, we consider interest similarity with respect to three interest domains - movie, music and TV - respectively. And we measure interest similarity founded on every single interest item - a finer grain.

In summary, the main contributions of this paper include:

- Relying on a large dataset crawled from Facebook, the analytical results can advance the collective knowledge of OSNs.
- The findings about homophily regarding interest similarity could practically benefit numerous applications and services, such as recommendation system and advertisement service.

II. Data description and statistics

A. Data description

Facebook is the largest online social network in the world, and leaves open-ended spaces to explicitly present their interests in several domains as movies, music, TV, books and so on. For studies about interest similarity among users, we crawled Facebook from March to June in 2012 and collected data from $479,048$ users. To our knowledge, these data represent one of the largest and most comprehensive social information databases up to date, involving 9 interest domains and $5,263,351$ user-generated interest items (including $626,294$ distinct items). The analyzed data can be split into three parts: *User Interest*, *Demographic Information*, and *Social Relationship*:

- ***User Interest***: We conduct the analysis across three representative interest domains - music, movie and television (TV) - since more users report interests in these three domains than the others.
- ***Demographic Information***: It contains 7 attributes of users' profiles including age, gender, current city, hometown, high school, college and employer.
- ***Social Relationship***: This is represented by users' friend list. The friendship relation in Facebook is bidirectional, i.e., A is B's friend when B is a friend of A.

Note that we construct our dataset merely with users' public information and anonymize all the users during the analysis.

B. Characteristics statistics

In this section, we first examine some high-level characteristics and patterns of demographics that emerge from the collective users.

1) Demographic characteristics of individuals: Gender, location, and age are the three specific demographic attributes being considered. $256,163$ (53.5%) users in our dataset report their gender, while $173,027$ (36.1%) users publish their current city which is used to represent users' location. Compared with reporting gender and current city, users are more reluctant to uncover their age and only $14,055$ (2.9%) users have their age in the crawled profiles.

Among the $256,163$ gender reporters, $124,677$ of them are self-reported as females while $134,486$ are males. Although males account for a slightly greater proportion than females in our dataset, females dominate over males of reporting interests. Table I presents the numbers and percentages of females and males that report their interests in terms of music, movie and TV respectively. The results infer that females are more likely to report their interests than males.

	Music	Movie	TV
Male	35516	50692	40620
Female	42648	58850	47225
Male (%)	26.4	37.7	30.2
Female (%)	34.2	47.2	37.9

TABLE I. Distributions of interests by gender

Figure **??** displays the geographical location distribution of $173,027$ current city reporters over the globe. We decode the geographical coordinate of users' current city with latitude and longitude via Facebook Graph API. We can see that the red dots are mainly located in the east coast of North America as well as Europe, thus we infer that people from North America and Europe are the dominant users on Facebook. We also observe that people in coastal regions are relatively more active than people situate inland. In addition, a few blue dots are noticed in the oceans, which might indicate some users report fake locations. We ignore them as the number is very small.

Moreover, we study the distribution of users by age. Figure 2 displays the distributions of age reporters with respect to female, male, unknown gender and all. Among all the age reporters, 4196 are male and 4096 are female. We notice that the age distributions of males and females are similar to each other. We also observe that the user distributions are skewed by age following with a long tail. The users in the 20-30 span of years are the most representative users in our dataset; while the proportion of the users older than 40 years or younger than 20 in our dataset is rather small (less than 10% in total). Besides, we choose 3 years as an age interval and cluster age reporters in the age range of 20-40 into seven age groups. Figure 3 examines the average number of interests that each user exhibits according to different age groups. It reveals that the young users report more interests than the users in middle-age.

2) Demographic characteristics of friends: In this section, we further reveal the demographic characteristics between friends in terms of gender, location, and age respectively.

We first examine the distribution of friends by gender combinations: cross-gender friends and same-gender friends. This analysis is conducted on the $256,163$ gender reporters. Particularly, for each gender reporter, we rely on his/her friends that are also gender reporters and calculate the percentage of friends in the same-/cross- gender respectively. Figure 4 displays the CDF of the percentage of friends by gender combinations. We observe that only around 40% of users exhibit the same gender with less than half of their friends, while more than 60% of gender reporters make fewer friends (i.e., less than half) with opposite gender. It indicates that

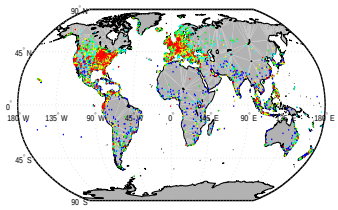


Fig. 1. Location distribution

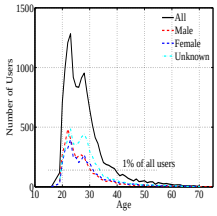


Fig. 2. User distribution by age

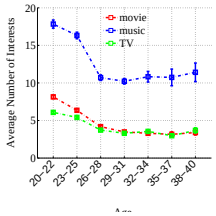


Fig. 3. Interest distribution by age

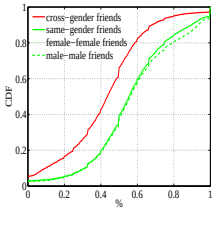


Fig. 4. CDF of friends distribution by gender combinations

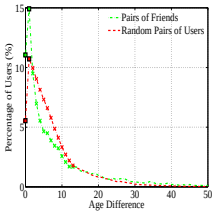


Fig. 5. Pairs distribution by age difference

people prefer to make friends with others of the same gender, especially for men.

In addition, we track how age affects the friendship between people. Figure 6 displays the distribution of pairs at various age differences. It reveals that people are more likely to make friends with others at the same age or at an age gap of 1-2 years. The percentage of friend pairs decreases rapidly as age difference increases when it is larger than 1 year. Besides, we also notice that the percentages of friend pairs are less than the numbers of random pairs at the age differences in the range of 3 - 13 years. When age difference is larger than 13 years, people make friends following the random probabilities. We infer that people are more likely to make friends with others who are in the similar ages.

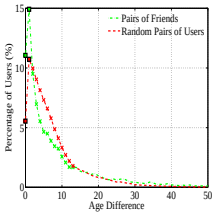


Fig. 6. Pairs distribution by age difference

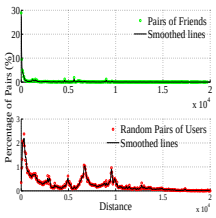


Fig. 7. Pairs distribution by distance

We calculate geographical distances between pairs and illustrate the pairs distribution with distances in Figure 7. From the upper subfigure, we see that the distance distribution of friend pairs is strongly skewed to the left. It falls dramatically from the start, bottoms out at the distance of 400 kilometers, and then stays at a very low value as the distance increases. Among all the friend pairs in the experiments, 28.9% of them come from the same city and 43.43% of the friends live less than 100 kilometers apart. Whereas, the lower subfigure shows that the percentages of random pairs fluctuate by distances with a gradual downward trend. The peaks and drops at some specific distances may reveal geographical characteristics. For instance, the peaks at distances of 5000 km and 6500 km may respectively indicate the width of America and the width of Atlantic. The different distributions of friend pairs and random pairs, in other words, mean that people tend to make friends within a short distance.

III. Effects on Interest Similarity

In this section, we first define the metric of interest similarity, followed by the studies on how interest similarity correlates to demographic information, social relationships and interest individuality sequentially.

A. Definition of interest similarity

We formalize a notion of *Interest Similarity* that measures how much two users' interests overlap. We denote user u's interests by an interest set I_u instead of a binary interest vector, in order to avoid the very sparse interest vector. Drawing on the calculation of cosine similarity, interest similarity between users u and v is then defined as the cosine distance between their respective interest sets: $s_I(u,v) = \frac{\|I_u \bigcap I_v\|_1}{\|I_u\|_2 \cdot \|I_v\|_2}$ where $\|I_u\|_2 = \sqrt{l_u}$ (l_u is the number of interests of u) and $\|I_u \bigcap I_v\|_1$ is the number of the same interests of u and v. If either $l_u = 0$ or $l_v = 0$, $s_I(u,v)$ is undefined.

In the dataset, each user might report various items in various interest domains. We think of users' interest similarity separately in different domains, i.e., interest similarity in terms of movie, music, TV. For the analysis of each particular interest domain, we only consider the users who have more than three items in the domain.

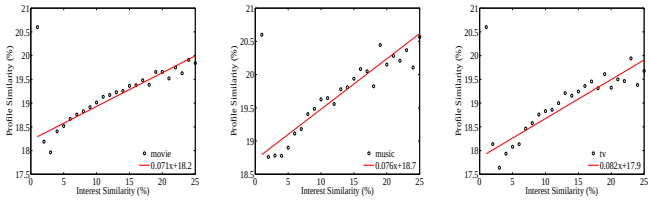
B. Homophily of interest similarity by demographics

In this section, we study how demographic information affects interest similarity between users. We separately conduct several experiments by using different user samples. For instance, to test the relation between gender and interest similarity on movie, we select users who present gender and more than three interested movies and construct a gender/movie set of pairs.

1) Profile similarity with interest similarity: We first look into how the interest similarity between users changes with their profile similarity. Similar to the interest similarity evaluation, we perform cosine to profile vectors of two users and formulate profile similarity as $s_p(u,v) = \frac{\|P_u \bigcap P_v\|_1}{\|P_u\|_2 \cdot \|P_v\|_2}$. In particular, 7 demographic attributes of age, gender, current city, hometown, high school, college and employer are considered. We assume if the corresponding attributes a_i of two users are completely the same, then $\|P_u(a_i) \bigcap P_v(a_i)\|_1$ is set to 1, otherwise it equals 0. As long as one user of a pair misses information about the attributes a_i, $\|P_u(a_i) \bigcap P_v(a_i)\|_1$ is also equal to 0. Hence, the profile similarity can be simplified as $\frac{\sum_{P_u = P_v} 1}{7}$.

We generate $500,000$ user pairs for each interest domain and show the collective relation between interest similarity and profile similarity in figure 8. Regarding all the three interest domains of movie, music and TV, we observe that the profile similarity gets higher if the users share more common interests. We can fit their relations with linear functions as $y = ax + b$. In other words, the observations reveal the positive correlation

between interest similarity and profile similarity regardless of interest domains, whereas the coefficient are different in these domains.



(a) movie (b) music (c) TV

Fig. 8. Profile similarity with interest similarity

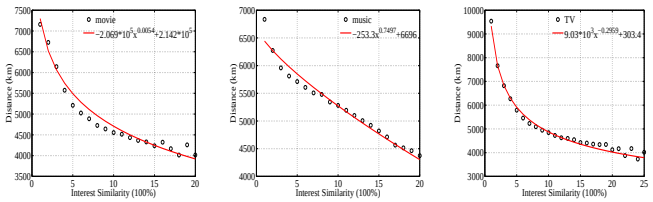
2) Interest similarity by gender: We group user pairs according to the categories of their gender combinations. Table II shows the average interest similarities of male-male pairs, female-female pairs, as well as male-female pairs. We observe that people have a higher interest similarity with the others when they are in the same sex. For instance, the interest similarity regarding movies between two males is close to 0.02 while the value between female and male only exhibits 0.014. This demonstrates that the homophily of interest similarity holds for gender.

	Movie	Music	TV
Male & Male	0.0202	0.0190	0.0347
Female & Female	0.0188	0.0154	0.0430
Female & Male	0.0136	0.0145	0.0276

TABLE II. INTEREST SIMILARITY BY GENDER

In addition, we also notice that user pairs share much higher interest similarity in terms of TV than the other two interest domains. For example, the male-female user pairs generate an average interest similarity of 0.028 regarding TV, compared with 0.014 and 0.015 for movie and music respectively. It might be due to the fewer selections for TV shows (there are $66,396$, $93,846$ and $370,456$ distinct items of TV, movie and music respectively in our dataset). Moreover, we find that males are more alike to each other on the interests of movie and music whereas females have higher similarity in the domain of TV.

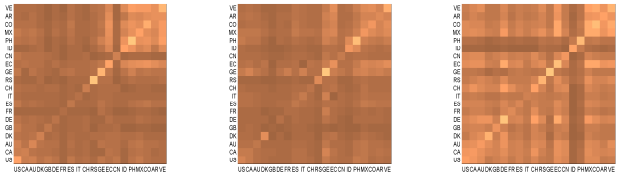
3) Interest similarity by location: Location has been proved as a key factor in various social network scenarios and applications. We investigate whether and how the location of a user pair affects their interest similarity from two perspectives: the geographic distance and users' countries.



(a) movie (b) music (c) TV

Fig. 9. Geographical distance with interest similarity

We intuitively hypothesized that the pairs would exhibit a higher interest similarity if they are geographically closer to each other. Figure 9 plots the aggregate relation between distance and interest similarity based on $500,000$ pairs in each interest domains. We observe that interest similarity changes with the distance between users - the average distance of pairs decreases with the increase of interest similarity. And their relationships follow power functions ($y = ax^b + c$).



(a) movie (b) music (c) TV

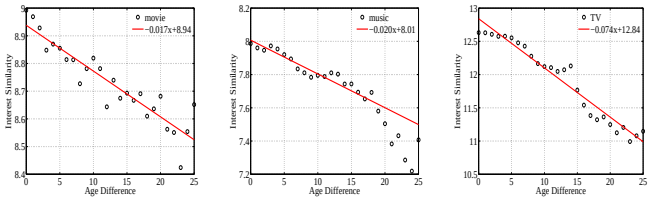
Fig. 10. Interest similarity by country

We further investigate the interest similarity of user pairs according to their current country-country combination. We select users from 20 countries over six continents. We produce $200,000$ user pairs for each country-country combination and then compute its average interest similarity. Figure 10 displays the heatmaps of the average interest similarity in terms of movie, music and TV respectively. This figure maps the average interest similarity of country-country combination into corresponding small cell in the figure. The larger the interest similarity is, the more brightly the corresponding cell presents. The countries on the same continent are put together in the figures following the order of North America, Australasia, Europe, Africa, Asia, and South America. The cells on the secondary diagonal represent the interest similarity of pairs from the same countries.

The results reveal that users from the same country share more items of interests regardless of the interest domains. However, with respect to the different interest domains, the interest similarities between users from two countries probably exhibit different characteristics. For instance, we observe that the users from Philippines have high similarity on movie with the users from countries in South America like Mexico, while the interest similarity in terms of TV is low among them. In addition, we observe that the cells in right upper areas are relatively brighter in all the four subfigures, which might imply that the countries in South America share more culture with each other. The users from the U.S.A., Canada and Australia also report more similar interests. We also notice that interest similarity does not correlate to distance very strictly concerning countries. The interest similarity between pairs from nearby countries is perhaps low, and vice versa. For instance, in terms of movie, users in China show low interest similarity with users from Philippines and Indonesia, but report relatively high similarity of movie with users in North America and Europe (shown in figure 10(a)). Besides, users in different European countries do not share many interests even though they are close to each other.

4) Interest similarity by age: Intuitively, people in various generations appreciate diverse styles of music or have different tastes of movies in specific eras. For instance, young generation of 1990s probably likes *Justin Bieber*; while middle-age people who were born in 1970s might listen to the music from *The Beatles* more. Therefore, in this section, we are interested

in how the age difference influences on the interest similarity of pairs.



(a) movie (b) music (c) TV

Fig. 11. Age difference with interest similarity

According to the distribution of users by age (shown in figure **??**), the experiments in this section only depend on the users whose age falls between 20 and 45 years. Therefore, the age difference ranges from 0 to 25 years. In addition, although the number of age reporters is relatively small $(14,055)$, the amount of user pairs generated by randomly coupling users is huge enough. We also produce $500,000$ user pairs for each interest domain.

Figure 11 displays how the interest similarity of user pairs changes with their age difference. We observe that the interest similarity declines as the age difference goes up with respect to all the three interest domains. This observation demonstrates that the homophily of interest similarity holds for age - the users share more interests if they are more similar at age. We employ linear models to depict the trends of the correlations.

C. Friendliness of interest similarity

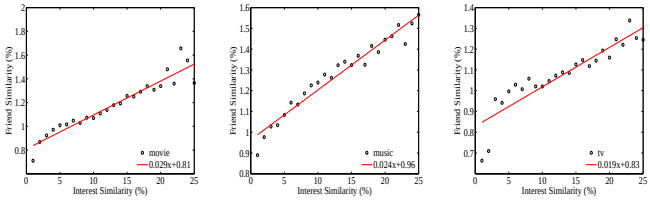
Relationship is considered as a special element, which distinguishes social network from general web sites and blogs. With user-generated relationship, OSNs are constructed by connecting people. They generally involve many real social relations. For example, the friends on OSNs perhaps have known each other in their real life or have engaged in a same event or in a same interest group. We hypothesize that the friendship among users in Facebook would strongly correlates to their interest similarity. And the examinations are carried out in two parts: the effects of friendship relations of pairs and quantified friend similarity.

1) Interest similarity by relation of pairs: We take into account users' friendships by two hops and categorize users' relations into three groups: pairs of friends, pairs of indirect friends and random pairs. We define two users u and v as indirect friends if u is a friend of v's friend. We report interest similarity by friendship in table III. We observe that the interest similarity between friends is the highest, and indirect friends also share more interests than random pairs. For various interest domains, the average interest similarity of friend pairs could be 1 to 4 times larger than the one of random pairs. Therefore, we conclude that friends are more likely to have same tastes on any interest domains.

Interest Similarity (%)	Music	Movie	TV
Friends	3.58	4.98	7.45
Indirect Friends	1.73	1.71	3.67
Random Pairs	1.54	1.41	3.04

TABLE III. INTEREST SIMILARITY BY FRIENDSHIP

2) Interest similarity with friend similarity: Much previous work differentiates the relationship between two users by its strength [9] [10]: strong connections (e.g., intimate friends, or close friends) and weak connections (e.g., acquaintances, or strangers but knowing several same acquaintances). In this section, we further measure the effect of relationship on interest similarity by its strength. We quantify the strength of the connection between two users by friend similarity and assume that the pairs with a stronger relationship have a higher friend similarity.



(a) movie (b) music (c) TV

Fig. 12. Friend similarity with interest similarity

Similar to the calculation of interest similarity in section III-A, friend similarity is measured by the overlap of friends between two users, based on the method of cosine similarity. Denote the friend sets of user u and user v as F_u and F_v respectively, and then the friend similarity between users u and v is computed by $s_F(u,v) = \frac{\|F_u \bigcap F_v\|_1}{\|F_u\|_2 \cdot \|F_v\|_2}$. Where $\|F_u \bigcap F_v\|$ stands for the number of the same friends that u and v own, and $\|F_u\|_2$ is equal to square root of the number of u's friends.

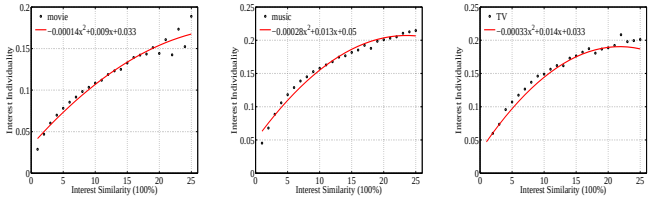
Figure 12 plots the aggregated relation of interest similarity versus friend similarity among $1,000,000$ pairs for each interest domain. We observe that interest similarity is positively correlated to friend similarity in all the three interest domains. In other words, the observations demonstrate that the user pairs generally share more interests if they obtain a higher friend similarity.

D. Effects of interest individuality

In this section, we are interested in whether the personalized characteristics of individual interests, namely *interest individuality*, would affect the interest similarities of pairs.

To define interest individuality, we consider two factors of individual interests: 1) the number of interests and 2) the popularity of his/her interests. In particular, the popularity of a specific interest item associates with the number of its fans. We normalize the popularity of interests (p_i) between 0 to 1 with a power function of $p_i = \frac{e^{(x-1)/k}-1}{e^{(k-1)/k}-1}$. In this function, x represents the fan number of a particular interest i, and $k = \max(x)$. Eventually, we define a interest individuality by the production of the average interest popularity and accumulative interest popularity, denoted as $E_u = (\frac{1}{N}\sum p_i) \cdot \frac{1-e^{-\sum p_i}}{1+e^{-\sum p_i}}$. Here we apply the function of $\frac{1-e^{-x}}{1+e^{-x}}$ to normalize the accumulative quality of interests. Note that the user would gain a higher interest individuality if he/she has more interests and his/her interests are more popular. In the following analysis, we average the interest individuality of two users to represent the interest individuality of the pair.

Figure 13 displays the correlation between interest similarity and user's interest individuality. It reveals that interest similarity grows with the increase of user's interest individuality regardless of interest domains. To wit, two users tend to share many interests if both of them exhibit many highly popular interests (i.e., larger individuality). In addition, we also notice that the sensitivity (i.e., slope) of the relations decreases when the values of interest similarity and the user individuality becomes larger. This indicates that the correlation is stronger when the values are smaller.



(a) movie (b) music (c) TV

Fig. 13. Interest individuality with interest similarity

IV. Related work

A lot of work has focused on studying the characteristics of social graph in large-scale online social networks. Among the early initial studies, [2] conducted a comprehensive analysis on the MSN message network, while [1] examined and compared four social networks (Flickr, YouTube, LiveJournal, Orkut) simultaneously. These early studies mainly shed light on the high-level characteristics and verified many properties of online social networks, such as power law and small world. Afterwards, as Facebook became the largest online social network in the world, much work turned to use Facebook as testbed. [3] carried out an experiment on the complete Facebook social network, including 721 million users at that time. The analysis results verified most common structural network characteristics that had been found by earlier work on smaller social networks. Complementary to the basic friendship social graph, some work began to aim at users' interactions, such as posts, comments and mentions, and tried to analyze features on the user interaction graph [11], [12]. Different from the existing work, during the analysis of the dataset crawled from Facebook, we focus on a more specific question - how the interest similarity between two users relates to various social features.

Although some existing work has already examined that friends share more interests than strangers and has confirmed that interest similarity decreases with the increase of friend distance[5] [6]. It also has been verified that the interest similarity strongly correlates to the trust between users [7]. A close work to our paper attempted to deduce a user's interests by considering this user's social neighbors' interests [13]. This just turns out that understanding the interest similarity between users is practically essential for building social applications. In this paper, we carry out a more comprehensive analysis on the correlations between users' interest similarity and diverse social features, which associate with users' demographic information, social relationships and their interests. And we look into interest similarity with respect to movie, music and TV respectively, while the majority of existing studies on interests have not distinguished the different types of interests.

V. Conclusion

In this paper, we conduct a comprehensive empirical study on how users' interest similarity relates to various social features in a large Facebook dataset including 479,048 users and 5,263,351 user-generated interests. We conduct the study in three interest domains (i.e. movie, music, and TV). The result reveals that interest similarity follows the homophily principle and correlates with many social features: people tend to exhibit more similar tastes if they have similar demographic information (e.g., age, location) or share more common friends; besides, the individuals with a higher interest individuality would generally share more interests with the others. We believe the observations could be harnessed to improve various social applications and services.

VI. Acknowledgments

The research leading to these results was funded by the European Union under the project eCOUSIN (EU-FP7-318398) and the project SITAC (ITEA2-11020).

References

[1] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, ser. IMC '07, New York, NY, USA, 2007, pp. 29–42.

[2] J. Leskovec and E. Horvitz, "Planetary-scale views on a large instant-messaging network," in *Proceedings of the 17th international conference on World Wide Web*, ser. WWW '08, New York, NY, USA, 2008, pp. 915–924.

[3] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," *CoRR*, vol. abs/1111.4503, 2011.

[4] K. Lewis, M. Gonzalez, and J. Kaufman, "Social selection and peer influence in an online social network," *Proceedings of the National Academy of Sciences*, vol. 109, no. 1, pp. 68–72, 2012.

[5] L. Adamic and E. Adar, "Friends and neighbors on the Web," *Social Networks*, vol. 25, no. 3, pp. 211–230, Jul. 2003.

[6] D. H. Lee and P. Brusilovsky, "Social networks and interest similarity: The case of citeulike," in *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, ser. HT '10, New York, NY, USA, 2010, pp. 151–156.

[7] C.-N. Ziegler and J. Golbeck, "Investigating interactions of trust and interest similarity," *Decision Support System*, vol. 43, no. 2, pp. 460–475, Mar. 2007.

[8] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[9] R. Xiang, J. Neville, and M. Rogati, "Modeling relationship strength in online social networks," in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10, New York, NY, USA, 2010, pp. 981–990.

[10] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: inferring user profiles in online social networks," in *Proceedings of the third ACM international conference on Web search and data mining*, ser. WSDM '10, New York, NY, USA, 2010, pp. 251–260.

[11] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in *Proceedings of the 2nd ACM workshop on Online social networks*, ser. WOSN '09, New York, NY, USA, 2009, pp. 37–42.

[12] C. Wilson, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, "Beyond social graphs: User interactions in online social networks and their implications," *ACM Trans. Web*, vol. 6, no. 4, pp. 17:1–17:31, Nov. 2012.

[13] Z. Wen and C.-Y. Lin, "On the quality of inferring interests from social neighbors," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '10, 2010, pp. 373–382.