

A Study of User Behavior on an Online Dating Site

Peng Xia [†] Bruno Ribeiro ^{*} Cindy Chen [†] Benyuan Liu [†] Don Towsley ^{*}

[†] Department of Computer Science
University of Massachusetts Lowell
Lowell, MA 01854, USA
{pxia, cchen, bliu}@cs.uml.edu

^{*} Department of Computer Science
University of Massachusetts at Amherst
Amherst, MA 01002, USA
{ribeiro, towsley}@cs.umass.edu

Abstract—Online dating sites have become popular platforms for people to look for potential romantic partners. It is important to understand users’ dating preferences in order to make better recommendations on potential dates. The message sending and replying actions of a user are strong indicators for what he/she is looking for in a potential date and reflect the user’s actual dating preferences. We study how users’ online dating behaviors correlate with various user attributes using a real-world dataset from a major online dating site in China. Our study provides a firsthand account of the user online dating behaviors in China, a country with a large population and unique culture. The results can provide valuable guidelines to the design of recommendation engine for potential dates.

I. INTRODUCTION

Online dating sites have become popular platforms for people to search for and communicate with potential dates, offering an unprecedented level of access to potential romantic partners that is otherwise not available through traditional means. According to a recent survey¹, out of 54 million single people in US, 40 million have signed up with various online dating sites such as Match.com, eHarmony, etc, and around 20% of currently committed romantic relationships have begun online, which is more than through any means other than meeting through friends.

On an online dating site, a user can create a profile that typically includes the user’s photos, basic demographic information, behavior and interests (e.g., smoking, drinking, hobbies), self-description, and desired characteristics of an ideal partner and sometimes a personality questionnaire. After creating a profile, users can search for partners based on a variety of user attributes and exchange messages with them. Many sites provide suggestions on compatible partners based on proprietary matching algorithms.

There is often considerable discrepancy, or dissonance (to use a social psychology concept), between a user’s stated preference and his or her actual dating behavior [1]. Therefore, it is important to understand users’ true dating preferences in order to make better dating recommendations. The message sending and replying actions of a user are strong indicators for what he/she is looking for in a potential partner and reflect the user’s actual dating preferences.

In this paper we study the behavior of online daters, correlating various user attributes with user actions using observed real-world dataset obtained through a collaboration with a major online dating site in China. In particular, we will address the following research questions: (i) *Temporal characteristics*: How often does a user send and receive messages and how does this change over time? (ii) *Sending behaviors*: What is the relationship between the attributes of initiators and recipients of the initial messages? (iii) *Reply behaviors*: How does the reply probability of a message correlate with various attributes of the sender and receiver?

Many of our results on user messaging behavior align with notions in social and evolutionary psychology [2]–[4]. Males tend to look for younger females while females put more emphasis on the socioeconomic status (e.g., income and education level) of a potential mate. While geographic distance between two users plays an important role in dating considerations – the volume of messages quickly decreases as users live farther apart – we observe that females are more likely than males to send and reply to messages between distant big cities. Profile photos affect male and female’s messaging behaviors differently. Females with a larger number of photos are more likely to receive more messages and secure more replies from males, while male photo count plays a less prominent role in attracting and securing replies from females. These results on users’ dating preference can provide valuable guidelines to the design of recommendation engine for potential dates.

The rest of the paper is structured as follows. Section II presents an overview of previous studies on the data analysis of online dating sites. Section III describes the dataset that we obtained from a major online dating site in China. Section IV shows the temporal characteristics of users’ online dating behavior. Users’ message sending and replying behaviors are studied in Section V. We conclude the paper in Section VI.

II. RELATED WORK

Hitsch et al. [5] shows that in online dating there is no evidence for user strategic behavior shading their true preference. Both male and female users have a strong preference for similarity along many (but not all) attributes. U.S. users display a strong same-race correlations. There are gender

¹<http://statisticbrain.com/online-dating-statistics>

differences in mate preferences; in particular, women have a stronger preference than men for income over physical attributes. In their follow-up work [6] they show that stable matches obtained through the Gale-Shapley algorithm are similar to the actual matches achieved by the dating site, which are also approximately efficient.

Fiore et al. [7] analyzes messaging behavior and find them consistent with predictions from evolutionary psychology, women state more restrictive preferences than men and contact and reply to others more selectively. Lin et al. [8] studied how race, gender, and education jointly shape interaction among heterosexual Internet daters. They find that racial homophily dominates mate searching behavior for both men and women. This is not the case of Chinese online daters where the overwhelming majority of users are of the same race. Finkel et al. [9] states that online dating has fundamentally altered the dating landscape by offering an unprecedented level of access to potential partners and allowing users to communicate before deciding whether to meet them face-to-face. On the other hand, the authors also argue that there is no strong evidence that matching algorithms promote better romantic outcomes than conventional offline dating. Part of the problem is that main principles underneath the algorithms (typically similarity but also complementarity) are much less important to relationship well-being than online sites are willing to assume. Interesting on-the-fly statistics of OKcupid users is found at the OkTrends blog [10].

III. DATASET DESCRIPTION

Our dataset includes the profile information of 200,000 users uniformly sampled from users of a major Chinese online dating site registered in November 2011. The data contains all messages exchanged to and from these users until the end of January 2012.

A user's profile provides a variety of information including user's gender, age, current location (city and province), home town location, height, weight, body type, blood type, occupation, income range, education level, religion, astrological sign, marriage and children status, number of photos uploaded, home ownership, car ownership, interests, smoking and drinking behavior, self introduction essay, among others.

Each user also provides his/her preferences for potential romantic partners in terms of age, location, height, education level, income range, marriage and children status, etc. Of the sampled users, 69.7% are males and 30.3% females. Figure 1 shows the reported age distribution of the sampled users. We observe that the site is used by a young population. Moreover, in our dataset we observe that female income levels trail that of males. The median income ranges of male and female users are 3,000–4,000 and 2,000–3,000 Chinese Yuan, respectively. With respect to users' education level, females stated education levels tend to be higher than males. About 66.5% of females state to have at least a community college degree in contrast with only 53.2% of the males. The fraction of users with stated doctoral and post-doctoral degrees is 0.61%.

Unlike the U.S. and European online daters, where race

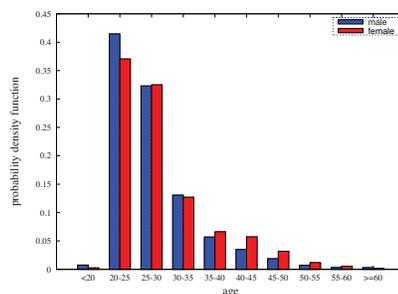


Fig. 1. Age distribution of the male and female users. plays an important role when it comes to finding potential romantic partners [8], [10], most of the users (98.9%) are of the same ethnicity (Han). Moreover, 97.0% of the users declare to be non-religious. Thus, race and religious traits of our dataset are significantly different than those of U.S. and european online dating sites [5], [8].

IV. TEMPORAL CHARACTERISTICS

We are interested in how a user's online dating activity level changes over time after he or she registered an account. We have eight full weeks of messaging behavior, allowing us to analyze the activities of each user during the first eight week of his/her membership. During these eight weeks 76,654 males (55.0% of males in the dataset) in our sample initiated 2,089,029 message exchanges with 508,118 distinct female users, which in turn generated 156,774 replies (a reply rate of 7.5%). During the same time period, 29,535 females (48.8% of females in the dataset) in our sample initiated 1,217,672 message exchanges with 440,714 unique male users, which in turn generated 112,696 replies (a reply rate of 9.3%). During the same period 94,179 females initiated 332,665 message exchanges with distinct 44,509 males in our sample, which in turn generated 62,801 replies (reply rate of 18.9%). Similarly, 288,602 males initiated 1,604,354 message exchanges with 45,623 females in our sample, which in turn generated 167,497 replies (a reply rate of 10.4%). Note that male users are more likely to initiate contact than female users while messages from female users are more likely to generate replies than those from male users.

The fraction of users from the dataset who sent out at least one message and the average number of messages sent by each user are shown in Figures 2 and 3, respectively. A considerable fraction of users (51.2 % male users and 43.0% female users) sent out at least one message during the first week of their memberships. This fraction decreases sharply in the second week (down to 11.3% for males and 12.8% for females) and in subsequent weeks. Except for the first week, females are slightly more likely to send out a message than males.

The average number of messages a male sends given that he sends at least one message stays between 15 and 20 messages per week, while the same number for females is more than twice as that in the first week, decreasing sharply after the second week at a level significantly below the males. Most users sent few messages with 94.6% of males and 96.5% of females sent less than 100 messages during their first eight weeks. On the other hand, a small fraction of the users send a

large number of messages. The online dating site identifies these users as likely spammers and quickly disable their accounts.

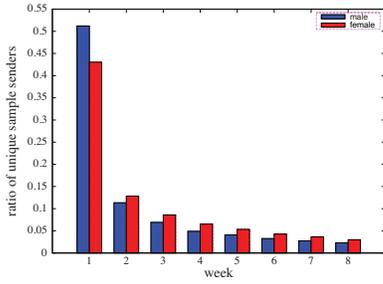


Fig. 2. Fraction of users who sent out at least one message during a week.

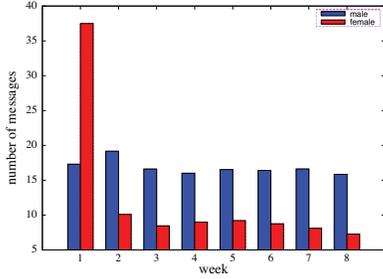


Fig. 3. Average number of messages a user sent out each week given that a user sends at least one message.

Figure 4 shows the average number of messages received at each week of membership by sampled users that received at least one message during the first eight weeks of their memberships. We observe that the fraction of both males and females who receive at least one message each week gradually decreases over time, and females are much more likely to receive messages than males. Also, for each week, the average number of messages a user received generally decreases over time for both genders, and the number of messages received by a female each week is much larger than that for a male user.

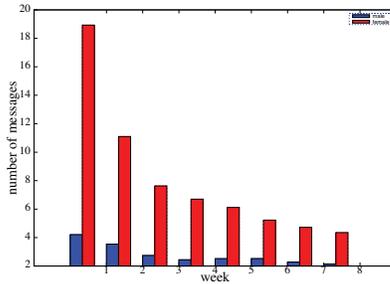


Fig. 4. Average number of messages a user received each week given that a user received at least one message.

For those users who received at least one message during the first eight weeks of their membership, we show the complementary cumulative distribution function (CCDF) of the number of messages received by each user in Figure 5. We observe that the distributions for both male and female users exhibit a log-normal-like behavior, and female users tend to receive more messages than male users.

V. MESSAGE SENDING AND REPLYING BEHAVIORS

After a user creates an account on the online dating site, he/she can search for potential dates based on information

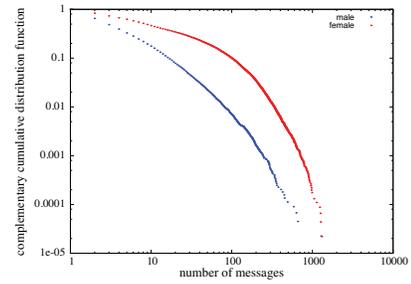


Fig. 5. CCDF of the number of messages a user received during the first eight weeks of his/her membership, within the profiles provided by other users including user location, age, etc. Once a potential date has been discovered, the user then sends a message to him/her. The message sending and replying behaviors of a user are strong indicators for what he/she is looking for in a potential partner and reflect the user's actual dating preferences. In this section, we present how user's message sending and reply behaviors correlate with various user attributes including age, income, education level, distance, and photo count.

A. Age

Figure 6 shows the distribution of the age difference between the sender and receiver of all messages sent by the sample users in the dataset. The age difference is computed as the sender's age less the receiver's age. While the age difference between senders and receives covers a wide range, the preferences of males and females are opposite of each other. Males tend to look for younger females and the distribution is skewed towards much younger females. On the other hand, females tend to look for older males and the distribution is skewed toward older males. The median age difference is 2 for messages sent from males to females and -4 vice versa.

A further analysis of the data shows that as a male gets older, the skewness towards younger females becomes even more pronounced. A female in her 20's is more likely to look for older males, but older females become more open towards younger males. This is the cause for the reply probability increase in the age difference range from -3 to -10. These results are consistent with observation made in [7].

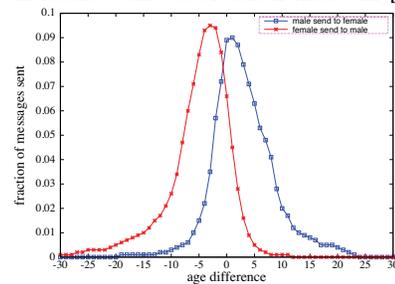


Fig. 6. Distribution of age difference between senders and receivers.

Figure 7 plots the reply probability as a function of the age difference between the sender and receiver of a message. For both males and females, the reply probability exhibits a bell shape mode at age difference of ten years older and eight years younger, respectively.

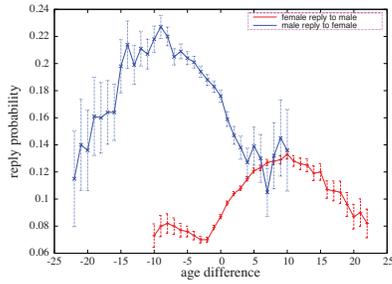


Fig. 7. Reply probability for users with different age difference.
B. Income

Figure 8 shows the distribution of income difference between senders and receivers. A user’s monthly income is reported as different ranges such as below 2,000, 2,000-3,000 (all in Chinese Yuan), etc. We take the median value of the reported income range as a user’s income and the income difference between the sender and receiver of a message is computed as the sender’s income less the receiver’s income. As shown in Figure 8, males tend to send messages to females with lower income, while females tend to send messages to males with higher income.

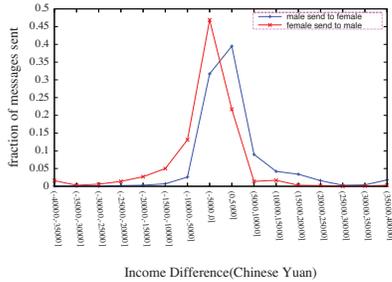


Fig. 8. Distribution of income difference between senders and receivers.

Figure 9 shows how reply probability varies with the sender’s income. The reply probability of female recipients increases steadily with male senders’ income. There is a very strong correlation coefficient of 0.91 between the reply probability and male sender’s income. On the other hand, the income of a female does not have as significant an effect on the likelihood of her messages being replied to. The correlation between the reply probability and female sender’s income is much weaker with a correlation coefficient of 0.47.

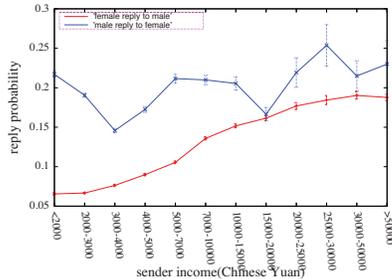


Fig. 9. Reply probability for senders with different incomes.

C. Education level

Figure 10 shows the fractions of messages sent to users of different education levels. Compared to a male, a female is more likely to contact a person with higher education levels. As shown in Figure 11, the higher the education levels of a

male sender, the more likely his messages will be replied to. On the other hand, the education level of a female does not have as significant an effect on the likelihood of her messages being replied to.

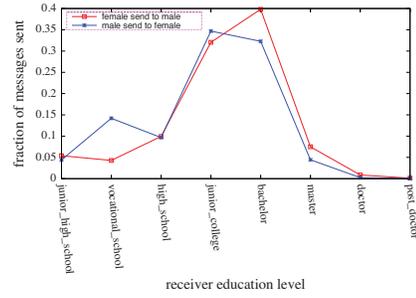


Fig. 10. Fraction of messages sent to users of different education levels.

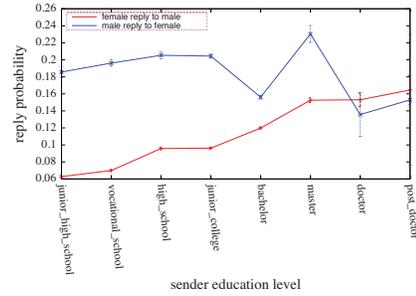


Fig. 11. Reply probability for messages from users of different education levels.

D. Geographic distance

The geographic distance between two users plays an important role in their online dating behavior. As mentioned in Section III, a considerable portion (41.5%) of the communications occurred between users within the same city. For communications between users in different cities, we further study how the message sending behavior and reply probability vary with the distance between the users (computed as the straight line distance between the two cities).

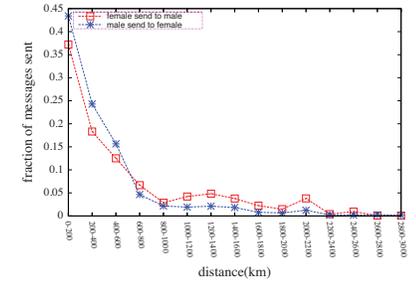


Fig. 12. Distribution of messages of different send-receiver distances.

As shown in Figure 12, in general the fraction of messages decreases as the distance between users increases. The messages between users of at least 1,000 km apart constitutes only a small fraction (11.7%) of the total number of messages. This bias towards shorter geographic distance is common in socio-economic complex networks (see [11] and references therein).

Figure 13 depicts how reply probability varies with distance. When a male receives a message from a female, the reply probability generally decreases with distance between them. For females, the reply probability first decreases with distance but increases in the range from 800 to 1,400km.

Note that there is a small increase in the fraction of messages between distance 800 and 1,400km for female senders. This is caused by the following factor. There is an increasing number of big cities (Shanghai, Beijing, Hong Kong, Chongqing, Guangzhou, Xi'an, etc) between many of which the distance falls into this range, and unlike males, females are more likely to send and reply to messages between these cities.

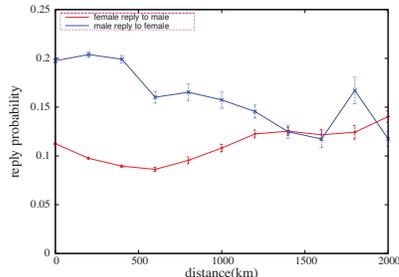


Fig. 13. Reply probability for users with different distances.

E. Photo count

On the dating site, a user can post photos on his/her profile page. Figure 14 plots the distribution of the number of photos posted by a user. A large fraction of users did not post or posted only a small number of photos. In our dataset, about 69.0% of male users and 59.0% of female users did not post any photos. Female users tend to post more photos than male users.

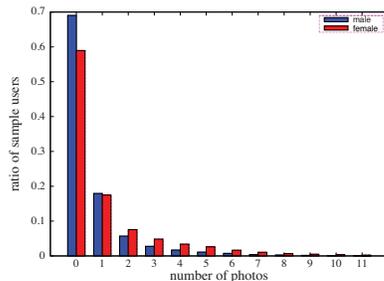


Fig. 14. Distribution of users' photo count.

As shown in Figure 15, a user tends to receive more messages if he/she has posted more photos online, with the trend being more pronounced for females than for males. The number of received messages by a male user starts to level off after some point.

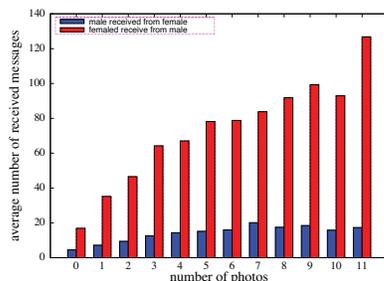


Fig. 15. Average number of received messages during first eight weeks of their memberships for users with different photo count.

Figure 16 shows how message reply probability varies with the number of photos posted by the sender. We observe that the male reply probability tend to increase with the number of photos posted by the female sender. Interestingly, when

a female receives a message, the reply probability remains relatively stable as the number of photos of the male sender increases.

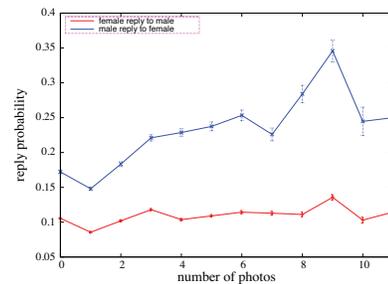


Fig. 16. Reply probability for users with different photo counts.

VI. CONCLUSION

We study how people's online dating behaviors correlate with various user attributes using a real-world dataset from a major online dating site in China. Many of our results align with notions in social and evolutionary psychology. In particular, males tend to look for younger females while females put more emphasis on the socioeconomic status (e.g., income, education level) of a potential date. Moreover, geographic distance between two users and the photo count of users play an important role in dating considerations. We are currently studying how users' actual online dating behaviors deviate from random selection and their stated preferences. Our study provides a firsthand account of the user online dating behaviors in China. These results on users' dating preference can provide valuable guidelines to the design of recommendation engine for potential dates.

REFERENCES

- [1] P. W. Eastwick and E. J. Finkel, "Sex difference in mate preferences revisited: Do people know what they initially desire in a romantic partner?" *Journal of Personality and Social Psychology*, vol. 94, no. 2, pp. 245–264, 2008.
- [2] D. M. Buss, "Sex difference in human mate preferences: Evolutionary hypotheses tested in 37 cultures," *Behavior and Brain Sciences*, vol. 12, pp. 1–49, 1989.
- [3] A. H. Eagly and W. Wood, "The origin of sex differences in human behavior: Evolved dispositions versus social roles," *American Psychologist*, vol. 54, no. 6, pp. 408–423, 1999.
- [4] S. Luo and E. C. Klohnen, "Assortative mating and marital quality in newlyweds: A couple-centered approach," *Journal of Personality and Social Psychology*, vol. 88, no. 2, pp. 304–326, 2005.
- [5] G. Hitsch, A. Hortasu, and D. Ariely, "What makes you click? mate preferences in online dating," *Quantitative Marketing and Economics*, vol. 8, no. 4, pp. 393–427, December 2010.
- [6] G. J. Hitsch, A. Hortacsu, and D. Ariely, "Matching and sorting in online dating," *American Economic Review*, vol. 100, no. 1, pp. 130–163, 2010.
- [7] A. T. Fiore, L. S. Taylor, X. Zhong, G. Mendelsohn, and C. Cheshire, "Who's right and who writes: People, profiles, contacts, and replies in online dating," in *Proc. of Hawaii International Conference on System Sciences*, 2010.
- [8] K.-H. Lin and J. Lundquist, "Mate selection in cyberspace: The intersection of race, gender, and education," *American Journal of Sociology*, forthcoming.
- [9] E. J. Finkel, P. W. Eastwick, B. R. Karney, H. T. Reis, and S. Sprecher, "Online dating: A critical analysis from the perspective of psychological science," *Psychological Science in the Public Interest*, vol. 3, pp. 3–66, 2012.
- [10] "Oktrends," <http://www.okcupid.com>.
- [11] D. Braha, B. Stacey, and Y. Bar-Yam, "Corporate competition: A self-organizing network," *Social Networks*, vol. 22, no. 3, pp. 219–230, 2011.