Gender-specific preference in online dating

Xixian Su[1] and Haibo Hu[1*]

*Correspondence:
hbhu@ecust.edu.cn
[1] East China University of Science
and Technology, Shanghai, China

Abstract

In this paper, to reveal the differences of gender-specific preference and the factors affecting potential mate choice in online dating, we analyze the users' behavioral data of a large online dating site in China. We find that for women, network measures of popularity and activity of the men they contact are significantly positively associated with their messaging behaviors, while for men only the network measures of popularity of the women they contact are significantly positively associated with their messaging behaviors. Secondly, when women send messages to men, they pay attention to not only whether men's attributes meet their own requirements for mate choice, but also whether their own attributes meet men's requirements, while when men send messages to women, they only pay attention to whether women's attributes meet their own requirements. Thirdly, compared with men, women attach great importance to the socio-economic status of potential partners and their own socio-economic status will affect their enthusiasm for interaction with potential mates. Further, we use the ensemble learning classification methods to rank the importance of factors predicting messaging behaviors, and find that the centrality indices of users are the most important factors. Finally, by correlation analysis we find that men and women show different strategic behaviors when sending messages. Compared with men, for women sending messages, there is a stronger positive correlation between the centrality indices of women and men, and more women tend to send messages to people more popular than themselves. These results have implications for understanding gender-specific preference in online dating further and designing better recommendation engines for potential dates. The research also suggests new avenues for data-driven research on stable matching and strategic behavior combined with game theory.

Keywords: Social network; Online dating; Mate choice; Preference

1 Introduction

As a special type of social networking sites [1–3], online dating sites have emerged as popular platforms for single people to seek potential romance. According to a recent survey, nearly 40 million single people (out of 54 million) in the U.S. have been trying online dating, and about 20% of committed relationships began online [4]. Although some psychologists have questioned the reliability and effectiveness of online dating [5], recent empirical studies using the tracking data and survival analysis found that for heterosexual couples, meeting partners through online dating sites can speed up marriage [6]. Besides, one survey found that marriages initiated through online channels are slightly less likely to break than through traditional offline channels and have a slightly higher level of marital satisfaction for the respondents [7].

Mate choice and marital decisions, because of their importance to the formation and evolution of society, have drawn wide attention of scholars from different fields. Two hypotheses, potentials-attract and likes-attract, have been proposed to explain the preference and choice of long-term mates [8]. The potentials-attract means that people choose mates matched with their sex-specific traits indicating reproductive potentials: men pay more attention than women to youthfulness, health, and physical attractiveness of partners which are the characteristics of fertile mates, while women pay more attention than men to ambition, social status, financial wealth, and commitment of partners which are the characteristics of good providers. In other words, men tend to seek young and physically attractive women, while women pay more attention to men's socio-economic status [9, 10], which is consistent with the Chinese saying "lang cai nv mao" for the choice of long-term partners [8]. In fact, analyzing gender differences of online identity reconstruction in an online social network revealed that men value personal achievements more while women value physical attractiveness more [11]. The likes-attract means that people choose mates who are similar to themselves in a variety of attributes, which is consistent with the Chinese saying "men dang hu dui". From the perspective of evolutionary and social psychology [12], the difference in parental investment strategies determines the different mate selection strategies for both sexes [13]. Empirical studies on offline dating showed that mate choice is very much in line with the evolutionary predictions of parental investment theory on which potentials-attract hypothesis is founded [14, 15], while one research on a Chinese online dating site showed that mate choice is more consistent with the likes-attract hypothesis [8].

From a sociological perspective, compared with the offline environment, online dating largely expands the search scope of potential mates [16, 17]. The Internet allows users to form relationships with strangers whom they did not know before, whether through online or offline channels. For individuals who are difficult to find potential partners through offline channels, such as homosexuals and middle aged and elderly heterosexuals, the Internet provides an ideal platform for them to meet their partners. The preference of people for mate selection has been extensively studied [18–21], such as the preference on education level [22], age [23] and race [24, 25]. The matching pattern or the choice for potential mates, shows a homophily phenomenon [26, 27], that is, people prefer to choose mates who are similar to themselves. Three possible reasons lead to homophily. First, similar people are more likely to have the same hobbies and reach the same places, thus it is easier to see each other [17]. Second, there exists homophily for the relationship from the introduction of friends and relatives [28]. Finally, the similarity between partners can also be explained by individual preferences or cost/benefit calculation. By analyzing OkCupid data [21], Lewis found that although there is a similarity preference for partner selection, the preference is not always symmetrical for men and women. On some online dating platforms, users can browse the profiles of the other users anonymously, without leaving any trace of visit. A recent study on a major North American online dating site found that anonymous users viewed more profiles than nonanonymous ones, however nonanonymity can achieve better matching results [29].

Economists usually study mate choice and marriage problem from the perspective of game theory and strategic behavior [30–35]. Considering the difference of mate choice for both sexes in marriage market, Becker regarded the marriage matching problem of mate choice as a frictionless matching process, and by constructing a matching model, Becker

proved that the mate choice is not random, but a careful personal choice of attributes [30, 31], which is later extended to a barging matching by Pollak et al. [32]. Marriage market is the first stage of a multi-stage game and corresponds with the Pareto efficiency of equilibrium. In the Internet age, Lee and Niederle launched a two-stage experiment in online dating market using rose-for-proposal signals [36], and found that sending a preference signal can increase the acceptance rate. Some other scholars also studied the mate preference from the economic perspective [37, 38]. For example, Fisman et al. found that male selectivity is invariant to size of female group, while female selectivity is strongly increasing in size of male group [37].

Computer scientists usually study online dating from the perspective of user behaviors [39–41] and recommendation systems [4, 42–44]. By analyzing online dating data, Xia et al. found that there exists distinct difference between preferences of men and women [41], and there also exists difference between users' stated and actual preferences. Xia et al. also proposed a reciprocal recommendation system for online dating based on similarity measures [4]. For general social networks, gender differences lead to obvious differences in behaviors and preferences between men and women. Research on an online-game society showed that females perform better economically and are less risk-taking than males, and they are also significantly different from males in managing their social networks [45]. Another research found sex-related differences in communication patterns in a large dataset of mobile phone records and showed the existence of temporal homophily [46].

Although the research on mate choice, both offline and online, has been extended to many fields, the following problems still exist: (i) online dating sites are a special kind of social networking sites, but the most previous researches focus only on the users' demographic attributes, and have not considered users' network centrality in dating sites, which can be potential important factors associated with users' mate selection; (ii) most studies focus on male and female preferences in mate choice, but they do not properly examine the compatibility of the two parties' preferences; (iii) with the advent of big data era, the methods of machine learning, such as ensemble learning, have been widely applied to diverse fields to achieve good prediction performance. However, most of the existing literature still only uses the econometric methods to study users' mate choice.

To address the research gap, in this paper, using empirical data from a large online dating site in China, we explore the users' attribute preference compared with random selection, and use logistic regression to study how the users' demographic attributes, popularity and activity and compatibility scores are associated with messaging behaviors, which reveal the gender differences in potential mate selection. We also use ensemble learning classifiers to sort the importance of various potential factors predicting messaging behaviors. At last we use correlation analysis to study users' strategic behavior.

2 Dataset

This study is based on a complete anonymized dataset extracted in 2011 from a large online dating site in China for only heterosexual users. The dating site provides many features common to other popular online dating platforms: it allows users to set up a profile, browse the profiles of potential mates, be browsed by the potential mates, and send and receive messages. Specifically, when a registered member (user) *A* visits the dating site, at a specific position of his/her homepage, the site will recommend to him/her the members that he/she may be interested in according to certain rules. At this time, *A* can only see

the members' avatar (real photo), nickname, location and age. After A enters the members' homepage, he/she can browse their detailed personal information without leaving the trace of visit. After that, if A feels very interested in some member, he/she will contact the member through the internal letters of the site. There are three data tables in the dataset, including female profiles, male profiles and the user behavior data. There are total 548,395 users in the dataset including 344,552 male users and 203,843 female users. The users' profiles include 35 attributes, such as user ID, gender, birthday, education level, mate requirements and so on. The dating site requires the registered users to be at least 18 years old at the time of registration, thus on the platform the minimum user age is 18.

The behavior data about user recommendation and behavior information is in the form of triples: u_a, u_b, and *action*, where *action* has three possibilities, *rec*, *click*, and *msg*. *rec* means that the dating site recommended user u_b to user u_a, *click* means that u_a clicked u_b for further personal information, and *msg* means that u_a sent a message to u_b. There are totally 4,151,224 records in the user behavior data, and the numbers of *rec*, *click* and *msg* are 3,978,321, 138,502 and 34,401, respectively.

3 Results

3.1 Attribute preference analysis

3.1.1 Attribute difference distribution

In online dating, there are significant gender differences in terms of attribute preference, self-presentation and interaction [47]. Users usually have a certain preference for mates' age or height. For both men and women, when they send messages to their potential partners, we compute the age difference as age(receiver) – age(sender), and the height difference as height(receiver) – height(sender). Figures 1 and 2 show the age difference and height difference distributions, respectively. As a comparison, we also show the randomized results by assuming that female(male) users randomly send messages to male(female) users.
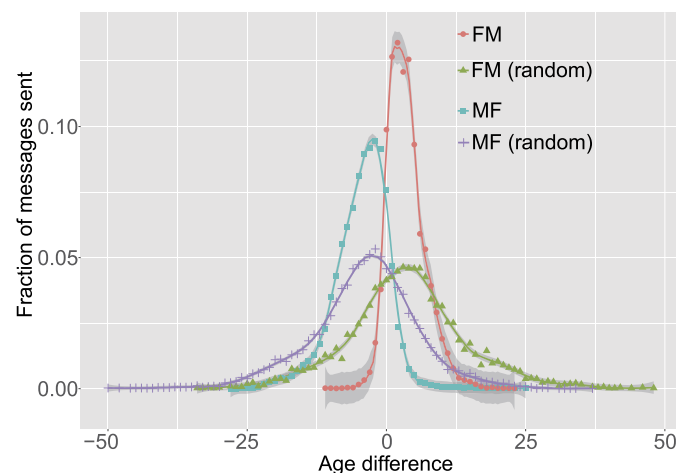

Figure 1 Age difference distribution. **FM** represents that female users send messages to male users and **MF** represents that male users send messages to female users. Solid lines represent the locally weighted polynomial regression fitting of their corresponding data points, and the gray interval represents a 95% confidence region

Figure 2 Height difference distribution. **FM** represents that female users send messages to male users and **MF** represents that male users send messages to female users. Solid lines represent the locally weighted polynomial regression fitting of their corresponding data points, and the gray interval represents a 95% confidence region
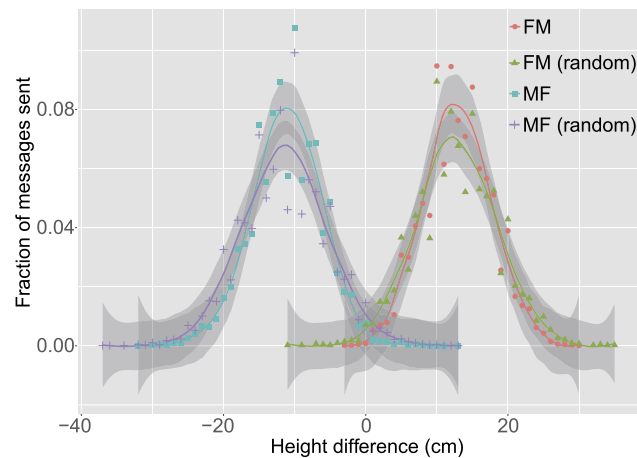
In most times and places, women usually marry older men [48, 49]. Figure 1 shows that in modern Chinese society, on average, men prefer women two years younger than them and women prefer men two years older than them. However, the range of age difference that women accept is smaller than that of men: the minimum age women accept is that men are 11 years younger than them and the maximum age they accept is that men are 23 years older than them, while the minimum age men accept is that women are 25 years younger than them and the maximum age they accept is that women are 28 years older than them. If only the age difference distributions are considered, in line with previous findings from a range of cultures and religions [50], we find that the range of ages that women are willing to message is narrower than the range of ages that men are willing to message. Male and female preferences are not random; they seek potential dates with a smaller age difference than predicted by random selection, which shows the characteristic of likes-attract.

Figure 2 shows that generally the height difference for women sending messages to men (most are 12 cm) are larger than that for men sending messages to women (most are 10 cm) when choosing potential mates. In China, for men, the ideal height difference is that they are 10 cm taller than the person they message, while for women, the ideal height difference is that they are 12 cm shorter than the person they message. According to the data from Yahoo! dating personal advertisements, for users in the U.S., height also matters for dating, especially for females [51]. In Fig. 2, the height difference range for women is smaller than that for men: the minimum height women accept is that men are 3 cm shorter than them and the maximum height they accept is that men are 30 cm taller than them, while the minimum height men accept is that women are 13 cm shorter than them and the maximum height they accept is that women are 32 cm taller than them. Females show the characteristic of likes-attract in terms of preference for height. As is same with age, users seek potential mates with a smaller height difference than predicted by random selection, although the difference is not as obvious as age difference.
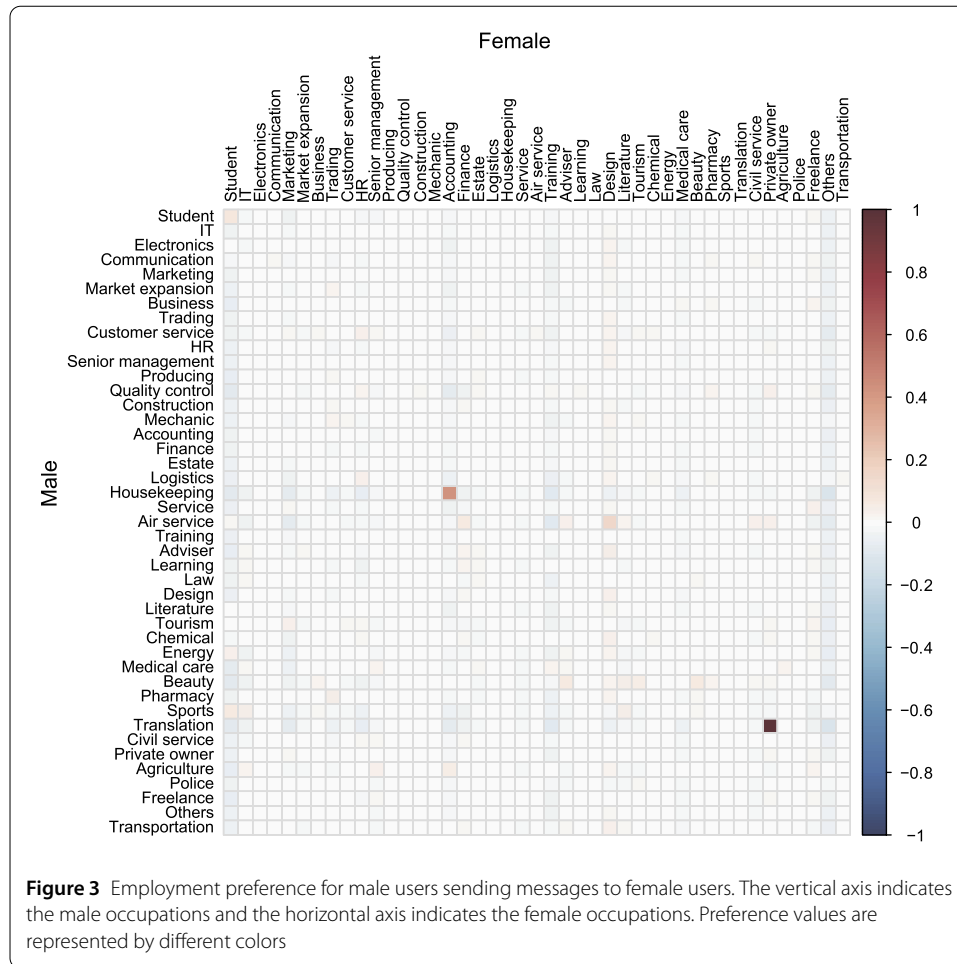
It is noteworthy that in the dating site, users' characteristics are all self-reported. For impression management considerations [52], users can exaggerate their personal characteristics [53]. For example, a recent research on online self-reported height against objectively measured data in young Australian adults revealed that self-reported height is significantly overestimated by a mean of 1.79 cm for males and 1.29 cm for females [54]. Men lie more than women about their height, which is also found in the online daters of New York City [55]. We note that users seem to have not accurately reported their physical height in the dating site. In the dataset, the average heights of female and male users are 161.99 cm (*SD* = 4.18) and 173.08 cm (*SD* = 4.68), respectively. However, in real world the average heights of adult females and males in China are 160.88 cm and 169.00 cm, respectively, which means that female and male users can exaggerate their height by an average of 1.11 cm and 4.08 cm, respectively. After correcting these, we find that real height differences 10 − (4.08 − 1.11) = 7.03 cm for men, and 12 − (4.08 − 1.11) = 9.03 cm for women would be significant. However we also notice that in the dating site, the average ages of male and female users are 28.73 and 28.58 years old, respectively, while in the overall adult population in China, the average ages of men and women are 40.56 and 41.01 years old respectively according to the population census data. The dating population is younger than the overall adult population, thus is likely taller, and users may not exaggerate their height by quite as much as calculated.

3.1.2 Attribute preference

When a user sends a message to another user, his/her choice of recipient may not be random, but rather has some preference for certain attributes, such as preference for employment, education, income, and so on. To characterize the preference of sender with attribute i for receiver with attribute j, let m_{ij} be the number of messages sent from users with attribute i to users with attribute j, m_i be the total number of messages sent from users with attribute i, n_j be the number of receivers with attribute j, and n be the total number of receivers, then the attribute preference is $p_{ij} = m_{ij}/m_i − n_j/n$. $p_{ij} > 0$ indicates that compared with random selection, senders with attribute i have a preference for receivers with attribute j, $p_{ij} = 0$ indicates that there is no preference and $p_{ij} < 0$ indicates negative preference, i.e. preferring not to select the receivers with attribute j.
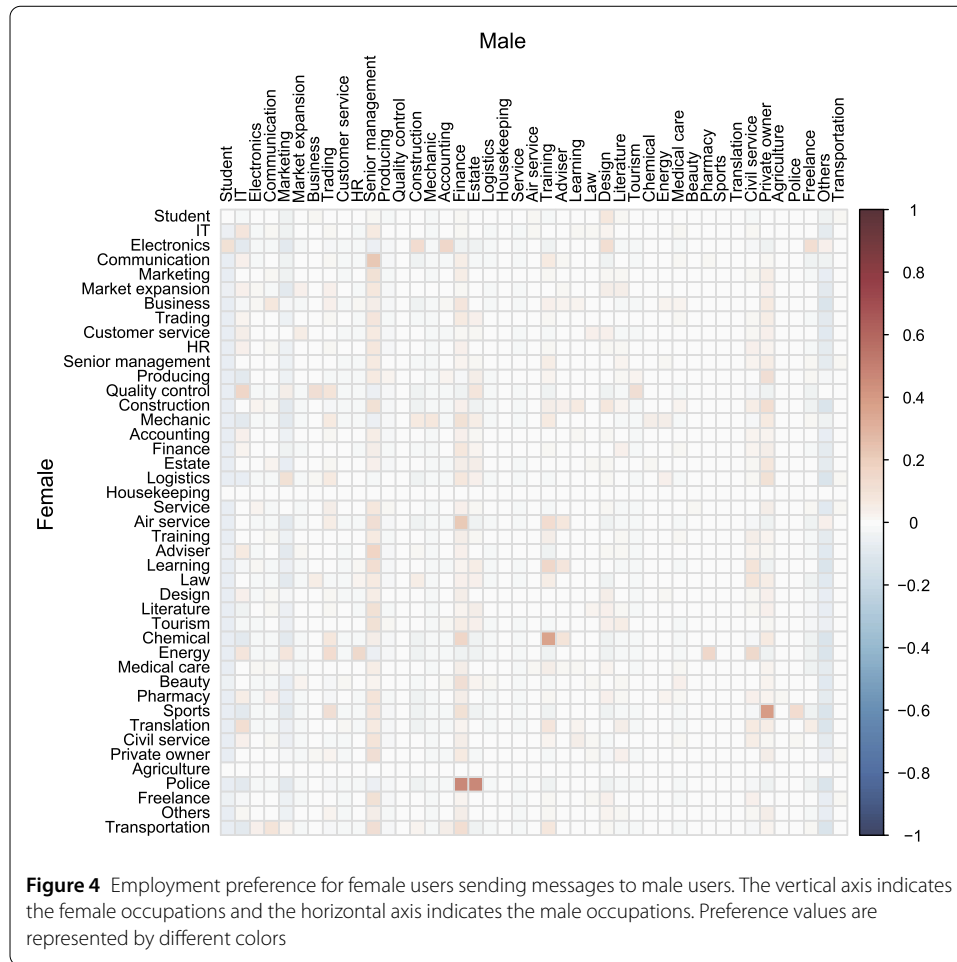
Employment preferences are shown in Figs. 3 and 4 (see Tables 1 and 2 in Additional file 1 for the meanings of attributes and the number and proportion of men/women for each employment). We find that compared with males sending messages to females, when female users send messages to male users, there is a stronger preference for the employments of their potential mates. In Fig. 3, we find that women who are students, accountants, educators or in other uncategorized occupations are not preferred by men, while women engaged in design are slightly popular in terms of the relative amount of messages received, especially for men in aviation service industry. At the same time, we also find that in these data, men engaged in housekeeping only send messages to women in accounting and men engaged in translation industry only send messages to women who are private owners, which may be due to the small sample size of user behavior with respect to these attributes.

From Fig. 4, we find that the most popular professions for men are senior management, finance, education and private owners. Most people in these four occupations have high income or are well-educated. Unpopular male users are school students, salesmen and

Figure 3 Employment preference for male users sending messages to female users. The vertical axis indicates the male occupations and the horizontal axis indicates the female occupations. Preference values are represented by different colors

those engaged in other uncategorized occupations. At the same time, women engaged in chemical industry tend to seek men engaged in education and training, women engaged in sports tend to seek men who are private owners, and women engaged in police only send messages to men engaged in finance and real estate in these data, which may also be attributed to the small sample size of user behavior with respect to these attributes.

Education levels have a significant impact on mating and marriage [22]. Education level preferences are shown in Figs. 5 and 6 (see Tables 3 and 4 in Additional file 1 for the meanings of attributes and the number and proportion of men/women for each education level). In China, like in the other countries, postdoctor also refers to a position rather than an educational achievement. However, in many Chinese websites, when a user registers, postdoctor is also considered an education level beyond obtaining a PhD. Similarly we find that compared with males sending messages to females, when female users send messages to male users, there is a stronger preference for the education level of their potential mates. Figure 5 shows that men whose education level is below the undergraduate degree tend to look for women the same academic qualifications as them or lower than their qualifications, men with education level higher than bachelor degree but lower than doctoral degree tend to look for women with bachelor degree, and men with a PhD degree or postdoctoral training tend to look for women with graduate degree. In terms of preference for education levels, generally men show likes-attract

Figure 4 Employment preference for female users sending messages to male users. The vertical axis indicates the female occupations and the horizontal axis indicates the male occupations. Preference values are represented by different colors

characteristic. For female users sending messages to male users, Fig. 6 shows that men with undergraduate and graduate degrees are popular and, for most women, undergraduate males are more popular, but graduate females are more likely to seek potential mates with graduate degree. In terms of preference for education levels, generally women show potentials-attract characteristic. Research on a German online dating site revealed that preference for similar educational background increases with educational level. Females are reluctant to communicate with males with lower educational levels, however there are no barriers for males to contact females with lower educational qualifications [22].

Education level and income are two important indicators of a person's social and economic status. From Figs. 7 and 8 (see Tables 5 and 6 in Additional file 1 for the meanings of attributes and the number and proportion of men/women for each income level) we find that, in terms of income levels, there is less obvious preference on potential mate selection for male users compared with female ones. On the one hand, as shown in Fig. 7, all men obviously prefer women whose monthly income is between RMB 5000 and RMB 10,000 (the RMB is the Chinese currency, and RMB 1 = 0.145 US Dollars = 0.128 Euros), while women whose income is below RMB 2000 are obviously excluded. However, men show no obvious preference or exclusion for women whose income is above RMB 10,000. On the other hand, as shown in Fig. 8, all women dislike men who earn less than RMB 5000, and
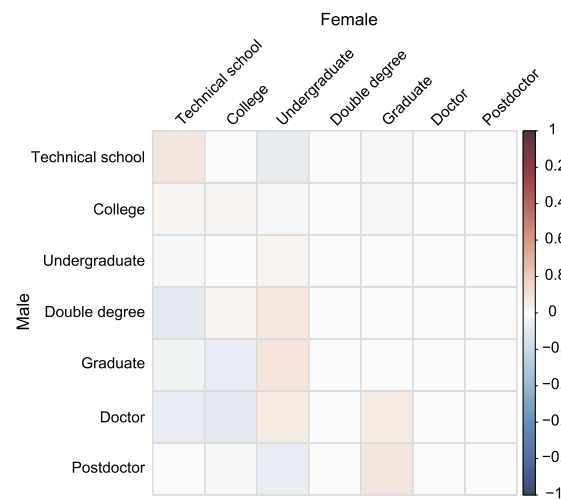
Figure 5 Education level preference for male users sending messages to female users. The vertical axis indicates the male education levels and the horizontal axis indicates the female education levels. Preference values are represented by different colors

Figure 6 Education level preference for female users sending messages to male users. The vertical axis indicates the female education levels and the horizontal axis indicates the male education levels. Preference values are represented by different colors. Postdoctoral females did not send any message to men in the dataset, and we set the elements in the corresponding row to 0

men who earn RMB 10,000 to RMB 20,000 are the most popular. In terms of preference for income levels, generally women also show potentials-attract characteristic. A field experiment on a Chinese online dating site found that men visited the profiles of women of different incomes with roughly the same rates, while for women, the higher the male incomes are, the greater the rates of visiting their profiles will be [38], which is different from our findings.
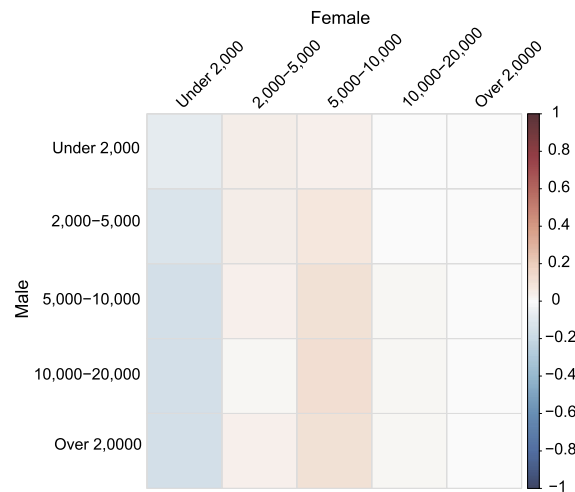
Figure 7 Preference for monthly income levels for male users sending messages to female users. The vertical axis indicates the male income levels and the horizontal axis indicates the female income levels. Preference values are represented by different colors
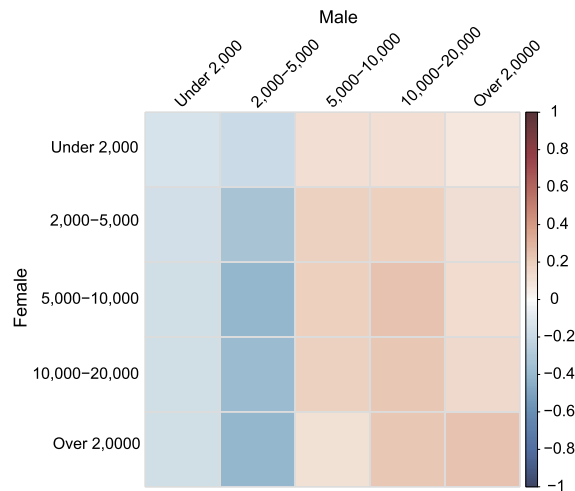

Figure 8 Preference for monthly income levels for female users sending messages to male users. The vertical axis indicates the female income levels and the horizontal axis indicates the male income levels. Preference values are represented by different colors

3.2 Logistic regression classification

3.2.1 Compatibility scores

On users' personal homepages, each user has shown the demands to the potential mates, including requirements for 7 attributes, i.e. age, avatar, education level, height, credit rating, place of residence and marital status (see Figs. 1–4 in Additional file 1 for the selection requirements of several attributes). As for credit rating, on the dating site, after a user passes the quick identity authentication, or uploads one of three documents (the ID card, the passport or the Hong Kong and Macau Pass) and passes the review, he/she will obtain the first star, i.e. credit rating equals 1. On the basis of the first star, each time a new document is uploaded and approved, an additional star or rating can

be added (up to five stars, i.e. five-star member). Besides although on the platform the minimum age of users is 18, there are still very few users who set their requirement for minimum or maximum age below 18 (see Fig. 3 in Additional file 1 for details). We apply the concept of compatibility score to describe the match between users based on whether or not a user meets another user's selection requirement. When women send messages to men, for each message and for each attribute, we can obtain the proportion of women who match the mate preferences of men and the proportion of men who meet the preferences of women, i.e. we can get two vectors including 7 proportions. According to the data we obtain $\mathbf{w}_{\text{FMm}} = (0.701, 0.886, 0.462, 0.826, 0.919, 0.786, 0.920)$, and $\mathbf{w}_{\text{FMf}} = (0.912, 0.976, 0.681, 0.962, 0.994, 0.864, 0.912)$, where $\mathbf{w}_{\text{FMm}}$ is the proportions of female attributes meeting male preferences and $\mathbf{w}_{\text{FMf}}$ is the proportions of male attributes consistent with female preferences. Similarly when men send messages to women, we obtain $\mathbf{w}_{\text{MFm}} = (0.877, 0.977, 0.402, 0.980, 0.992, 0.831, 0.960)$ and $\mathbf{w}_{\text{MFf}} = (0.671, 0.867, 0.572, 0.678, 0.758, 0.771, 0.892)$. Thus the compatibility scores of women sending messages to men are

$$c_{\text{FMm}} = \frac{\mathbf{w}_{\text{FMm}} \cdot (\text{female attr. in male pref.})}{\text{sum}(\mathbf{w}_{\text{FMm}})}, \tag{1}$$

$$c_{\text{FMf}} = \frac{\mathbf{w}_{\text{FMf}} \cdot (\text{male attr. in female pref.})}{\text{sum}(\mathbf{w}_{\text{FMf}})}, \tag{2}$$

and the compatibility scores of men sending messages to women are

$$c_{\text{MFm}} = \frac{\mathbf{w}_{\text{MFm}} \cdot (\text{female attr. in male pref.})}{\text{sum}(\mathbf{w}_{\text{MFm}})}, \tag{3}$$

$$c_{\text{MFf}} = \frac{\mathbf{w}_{\text{MFf}} \cdot (\text{male attr. in female pref.})}{\text{sum}(\mathbf{w}_{\text{MFf}})}, \tag{4}$$

where (female attr. in male pref.) is a vector characterizing whether female attributes meet male preferences for a pair of users (1 for yes and 0 for no), and similarly (male attr. in female pref.) is a vector characterizing whether male attributes meet female preferences for a pair of users. Equations 1 and 3 are the compatibility scores between a male preference and the profile of his chosen mate, and Eqs. 2 and 4 are the compatibility scores between a female preference and the profile of her chosen mate. For a pair of users, u_a and u_b, we use a score, i.e. reciprocal score, to quantify how much the attributes of u_b match the preferences of u_a and how much the attributes of u_a match the preferences of u_b. The reciprocal score between u_a and u_b is the mean of the compatibility scores of these two users, that is, for women sending messages to men the reciprocal score is $rs = (c_{\text{FMm}} + c_{\text{FMf}})/2$, and for men sending messages to women $rs = (c_{\text{MFm}} + c_{\text{MFf}})/2$.

3.2.2 Logistic regression

Let *click* be the number of times a user is clicked, *msg* be the number of messages received by a user, and *rec* be the number of times a user is recommended and shown on the other users' homepages, we define $pop_1 = click/rec$ and $pop_2 = msg/rec$ which can characterize the popularity of a user based on actions. We also use PageRank centrality (pop_3) to quantify how focal or popular a user is in a network by considering all connections in the network. Attractive people, such as the people with advantageous demographic attributes and

Table 1 Variables and their corresponding meanings

Variables	Meanings
MobileF	Whether a female mobile phone is verified
HouseF	Whether a female has a flat
AutoF	Whether a female has a car
LevelF	Female credit rating
Pop1F	Female pop_1
Pop2F	Female pop_2
Pop3F	Female PageRank (pop_3, the damping factor is 0.85) in the messaging network
IndegreeF	Female indegree in the click network
OutdegreeF	Female outdegree in the click network
CompatFM	The compatibility score between a female preference and the profile of the corresponding other side
MsgFM	Whether females send messages to males
MobileM	Whether a male mobile phone is verified
HouseM	Whether a male has a flat
AutoM	Whether a male has a car
LevelM	Male credit rating
Pop1M	Male pop_1
Pop2M	Male pop_2
Pop3M	Male PageRank (pop_3, the damping factor is 0.85) in the messaging network
IndegreeM	Male indegree in the click network
OutdegreeM	Male outdegree in the click network
CompatMF	The compatibility score between a male preference and the profile of the corresponding other side
MsgMF	Whether males send messages to females
RS	Mean of the compatibility scores of a sender and the corresponding receiver

higher socio-economic status, tend to be more demanding than average people in terms of potential mate choice, which can be revealed in the preference analysis of income and education level in Sect. 3.1.2. Those who are perceived as attractive by attractive people can be even more popular/attractive. The variables used in the paper and their meanings are shown in Table 1.

We introduce several centrality indices, such as pop_1, pop_2, pop_3, and indegree, to evaluate their correlation with messaging behaviors. It is noteworthy that the centrality indices are aggregated indicators describing users' desirability or popularity, and users do not know their indices, nor do they know the indices of others. We use outdegree to characterize users' activity level, and in the dating site, users also do not know the outdegree of other users. In reality, instead of using the indices to identify or select attractive partners, users will message another based on more specific clues, such as higher income, better education background, attractive photos or good demographic and socio-economic compatibility. In the paper, we will evaluate whether the indices are significantly associated with messaging behaviors.

Suppose p_i is the probability of sending messages for a female user i, $1 - p_i$ is the probability of not sending messages, then $L_{f_i} = \ln(\frac{p_i}{1-p_i})$, i.e., for all women, $L_f = \ln(\frac{p}{1-p})$. Similarly, suppose q_j is the probability of sending messages for a male user i, $1 - q_j$ is the probability of not sending messages, then $L_{m_j} = \ln(\frac{q_j}{1-q_j})$, i.e., for all males, $L_m = \ln(\frac{q}{1-q})$. We obtain logistic regression models as follows:

$$L_f = \alpha_1 + \boldsymbol{\beta}_1 \cdot \textbf{attribute} + \varepsilon_1, \tag{5}$$

$$L_m = \alpha_2 + \boldsymbol{\beta}_2 \cdot \textbf{attribute} + \varepsilon_2. \tag{6}$$

Table 2 Logistic regression results for female users sending messages to male users

Variables	Model 1		Model 2		Model 3	
	b	SE	b	SE	b	SE
Intercept	−5.322***	0.014	−5.548***	0.016	−5.640***	0.017
MobileF	−0.092***	0.014			−0.090***	0.014
HouseF	0.061***	0.014			0.038**	0.014
AutoF	−0.118***	0.016			−0.116***	0.016
LevelF	−0.059***	0.014			−0.072***	0.014
Pop1F	−0.162***	0.018			−0.167***	0.018
Pop2F	0.016	0.014			0.012	0.014
Pop3F	−0.110***	0.021			−0.121***	0.021
OutdegreeF	0.209***	0.005			0.211***	0.006
MobileM			0.015	0.014	0.025	0.014
HouseM			0.238***	0.015	0.243***	0.015
AutoM			0.153***	0.013	0.157***	0.013
LevelM			0.016	0.013	0.029*	0.013
Pop1M			0.393***	0.007	0.390***	0.007
Pop2M			0.053***	0.004	0.051***	0.004
Pop3M			0.142***	0.006	0.146***	0.006
OutdegreeM			0.028**	0.011	0.029**	0.011
CompatFM					0.057***	0.014
CompatMF					0.061***	0.014
AIC	72,160		67,462		65,958	
N	1,115,363		1,115,363		1,115,363	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In this study, multicollinearity tests are conducted to find out independent variables among which the correlation coefficients are less than 0.5 (see Tables 7 and 8 in Additional file 1 for details). The logistic regression results for women sending messages to men are shown in Table 2. We find that almost all the variables are significant when only considering the attributes of women (model 1), i.e., the attributes of senders, but only housing and outdegree of women are positively associated with the probability of women sending messages to men. When only considering the male attributes (model 2), except male mobile phone verification and credit rating, all the others are significant and are positively associated with the probability of women's sending messages. When considering the two parties' attributes and compatibility scores (model 3), among the significant variables, female mobile phone verification, car ownership, credit rating and popularity levels (pop_1 and pop_3) are negatively associated with the probability of women's sending messages, while the other variables are positively associated. We find that, when women send messages to men, they are concerned about not only whether they meet the requirements of men but also whether men meet their own requirements.

The logistic regression results for men sending messages to women are shown in Table 3. We find that when only the female attributes are considered (model 1), except female mobile phone verification, credit rating and outdegree, all the other variables are significant, but only female house ownership affects probability of male messaging in a negative way. When only male attributes are considered (model 2), all the variables are significant but only male outdegree is positively correlated with messaging behaviors, others negatively correlated. With all variables considered (model 3), except for female credit rating, outdegree, and the compatibility score between a female preference and the profile of the corresponding other side, all other variables are significant. Among the significant variables, female mobile phone verification, car ownership, popularity (pop_1, pop_2 and pop_3),

Table 3 Logistic regression results for male users sending messages to female users

Variables	Model 1		Model 2		Model 3	
	b	SE	b	SE	b	SE
Intercept	−4.800***	0.008	−4.732***	0.008	−4.973***	0.009
MobileF	0.007	0.007			0.022**	0.007
HouseF	−0.021**	0.007			−0.015*	0.007
AutoF	0.037***	0.006			0.038***	0.006
LevelF	−0.013	0.007			0.012	0.007
Pop1F	0.468***	0.004			0.462***	0.004
Pop2F	0.023***	0.003			0.022***	0.003
Pop3F	0.126***	0.003			0.141***	0.003
OutdegreeF	−0.004	0.008			0.000	0.008
MobileM			−0.146***	0.007	−0.147***	0.007
HouseM			−0.062***	0.008	−0.071***	0.008
AutoM			−0.100***	0.008	−0.107***	0.008
LevelM			−0.195***	0.009	−0.197***	0.009
Pop1M			−0.036***	0.009	−0.044***	0.009
Pop2M			−0.026*	0.012	−0.026*	0.012
Pop3M			−0.196***	0.013	−0.215***	0.014
OutdegreeM			0.338***	0.003	0.335***	0.004
CompatFM					−0.013	0.007
CompatMF					0.062***	0.008
AIC	219,384		227,215		210,184	
N	2,066,668		2,066,668		2,066,668	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

male outdegree and the compatibility score between a male preference and the profile of the corresponding other side are positively correlated with messaging behaviors, while all the other variables are negatively correlated. In addition, by analyzing the significance of the two compatibility scores, we find that men only pay attention to whether women meet their own requirements when sending messages to women.

As can be seen from Tables 2 and 3, for males or females sending messages, popularity of the other side is significantly positively associated with messaging behaviors. On the one hand, pop_1 and pop_2 values, according to their calculation method, represent a user's local popularity. On the other hand, pop_3 value, i.e. PageRank, represents the popularity of a user from a global perspective.

For females sending messages to males, $\exp(0.390) = 1.477$ for male pop_1 is larger than $\exp(0.146) = 1.157$ for male pop_3, and for males sending messages to females, $\exp(0.462) = 1.587$ for female pop_1 is also larger than $\exp(0.141) = 1.151$ for female pop_3. Thus, for both males and females, the other party's pop_1 is more important than pop_3. Besides we also find that, when females send messages to males, $\exp(0.390) = 1.477$ for male pop_1 is less than $\exp(0.462) = 1.587$ for female pop_1 when males send messages to females, which indicates that compared with females, for males the other side's pop_1 is more associated with their messaging behaviors. However, when females send messages to males, $\exp(0.146) = 1.157$ for male pop_3 is larger than $\exp(0.141) = 1.151$ for female pop_3 when males send messages to females, which indicates that compared with males, for females the other side's pop_3 is more associated with their messaging behaviors.

In China, having an apartment and a car is a symbol of a person's wealth and social status, and in some regions, they have become necessities for getting married. When women send messages to men, it is important for men to have a house and a car. When men send messages to women, it is not important for women to have a house but it's some-

what important for women to have a car. We find that exp(0.038) = 1.039 for whether the other side has a car when men send messages to women is smaller than exp(0.157) = 1.170 for whether the other side has a car when women send messages to men, indicating that women pay more attention than men to whether the other side has a car.

A user's outdegree quantifies the user's activity. Seemingly high activity means contacting many other users, however, essentially it may imply that users invest more time and resources in attempting to find potential partners. Outdegree is an attribute different for men and women. When a woman sends a message to a man, the other side's outdegree is significantly positively associated with the messaging behavior, while not when a man sends a message to a woman. When women send messages to men, network measures of popularity and activity of the men they contact are significantly positively associated with their messaging behaviors, but when men send messages to women, only the network measures of popularity of the women they contact are significantly positively associated with their messaging behaviors.

3.3 Ensemble learning classification

With the advent of the big data era, ensemble learning classification methods have gradually been introduced into the field of social network research. As early as 1996, Breiman proposed the method of bagging [56], and five years later, he further proposed the method of Random Forest [57]. Freund proposed the AdaBoost method in 1997 [58], and with the continuous improvement of machine learning classifiers, in 2016, Chen et al. proposed a classifier—XGBoost [59], which can greatly improve the efficiency and accuracy of algorithm in some cases. As an application, recently Reece et al. have already applied machine learning tools to identify depression from Instagram photos [60].

Regression analysis often has certain requirements on the independent variables, such as the absence of multicollinearity, however ensemble learning classification methods relax the constraints on independent variables. In this section, ensemble learning classification methods including bagging, Random Forest, AdaBoost and XGBoost are used to evaluate the importance of each attribute in Table 1. We use package 'adabag' in R software to perform AdaBoost and bagging methods, package 'randomForest' to perform Random Forest method and package 'xgboost' to perform XGBoost method. For the dataset, 5-fold cross validation is used to assess the classifiers' performance, and the algorithm parameters are chosen to obtain the stable error rate. The numbers of sending and not sending messages are unbalanced in the dataset, and the larger set is subsampled randomly to obtain a set the same size as the smaller one.

The error rates of four ensemble learning classification methods are shown in Table 4. We find that the error rates of Random Forest and AdaBoost are the lowest for females sending messages to males while XGBoost is the lowest for males sending messages to females. Attribute importance ranking is shown in Figs. 9 and 10. Figure 9 shows that when women send messages to men, the three most important attributes are the pop_3 and pop_1

Table 4 Error rates using ensemble learning classification methods

	Bagging	Random Forest	AdaBoost	XGBoost
FM	0.183	0.157	0.157	0.158
MF	0.240	0.162	0.195	0.158

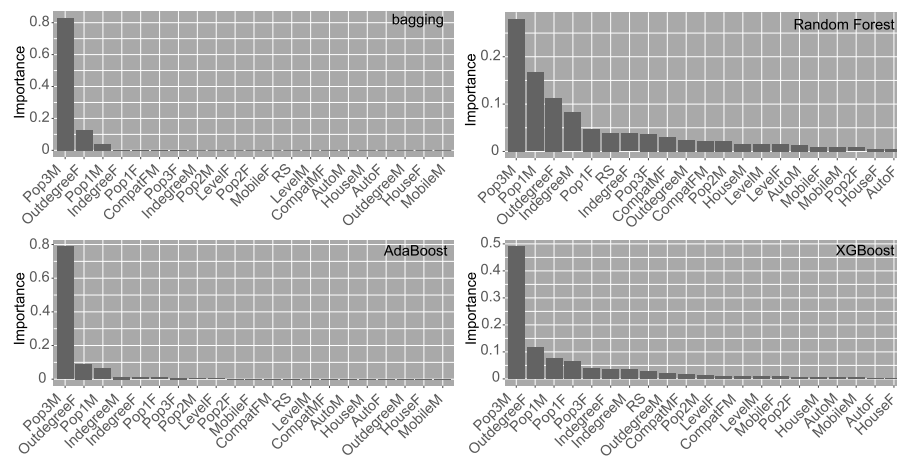FM: females send messages to males; MF: males send messages to females.

Figure 9 Attribute relative importance rankings when women send messages to men for different classification methods. The horizontal axis indicates the attributes and the vertical axis indicates their corresponding importance. For bagging, Random Forest, and AdaBoost, the relative importance of each variable in the classification task is measured by the Gini index, and for XGBoost the relative importance is measured by the Gain parameter
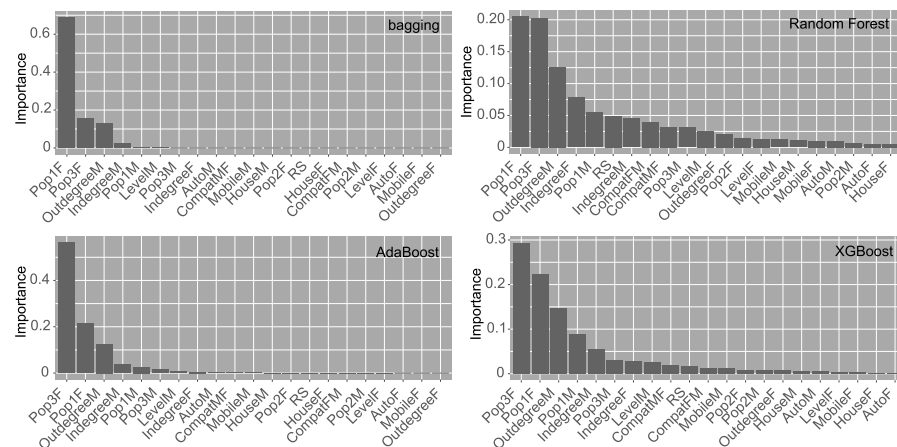

Figure 10 Attribute relative importance rankings when men send messages to women for different classification methods. The horizontal axis indicates the attributes and the vertical axis indicates their corresponding importance. For bagging, Random Forest, and AdaBoost, the relative importance of each variable in the classification task is measured by the Gini index, and for XGBoost the relative importance is measured by the Gain parameter

values for men, and the outdegree for women. Similarly, Fig. 10 shows that when men send messages to women, the three most important attributes are the pop_3 and pop_1 values for women, and the outdegree for men. The most important factors predicting the decision of sending messages of both men and women are the pop_3 and pop_1 values representing the popularity of potential mates, which are also significantly positively associated with messaging behaviors in the logistic regression.

The purpose of ensemble learning classification is different from logistic regression analysis. According to Figs. 9 and 10, the centrality indices indeed show the overwhelming importance, and the other variables show the relative lack of predictive power. However

this does not mean that the other variables are useless, and they can still be significantly associated with users' messaging behaviors in logistic regression.

3.4 Strategic behavior analysis

The concept of strategic behavior [61] derives from economics, where the original implication is that firms take action that affects the market environment to increase profits (referring to the message response rate in this study), which is then extended to matching problems [35], such as mate matching.

In our research, strategic behavior refers to whether a user will send a message to another user depends on whether his/her decision may increase the reply probability of the message. Since without user response data, we would like to use centrality indices characterizing user popularity to analyze whether users tend to send messages to people who are more popular than themselves or to those who are less popular. We study the users' strategic behavior by analyzing the correlation between centrality indices. Smoothing fitting curves for the correlation with generalized additive model show that there is a nonlinear or approximate linear relationship between users' centrality indices (see Figs. 5 and 6 in Additional file 1 for details), thus we use the Spearman correlation coefficient to characterize the correlation. As shown in Tables 5 and 6, We find that in the dating site men and women show different behavior patterns in messaging despite the reduced cost of rejection in the network environment. For males sending messages to females, there exist weak positive correlations between centrality indices, which can be characterized by small positive and significant correlation coefficients, while for females sending messages to males, there exist weak or modest positive correlations between centrality indices characterized by small or slightly larger positive and significant correlation coefficients. Men do not show strategic behavior to a large extent when sending messages, while for women, as their centrality indices increase, the corresponding indices of men who received their messages could also increase.

By studying the correlations between the same centrality index pairs for users, we further analyze whether users tend to send messages to people who are more popular than

Table 5 Spearman correlation coefficients among centrality indices when females send messages to males

	Pop1M	Pop2M	Pop3M	IndegreeM
Pop1F	0.105***	−0.023	0.145***	0.092***
Pop2F	0.050***	−0.016	0.046***	0.023
Pop3F	0.064***	−0.011	0.144***	0.062***
IndegreeF	0.083***	−0.013	0.163***	0.104***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6 Spearman correlation coefficients among centrality indices when males send messages to females

	Pop1F	Pop2F	Pop3F	IndegreeF
Pop1M	0.055***	0.027***	0.005	0.001
Pop2M	0.039***	−0.003	0.037***	0.034***
Pop3M	0.083***	0.025***	0.052***	0.028***
IndegreeM	0.087***	0.052***	0.061***	0.045***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 7 The proportions of the receivers' centrality indices that are larger than those of the senders' when sending messages

	pop_1	pop_2	pop_3	Indegree
FM	0.779	0.232	0.902	0.777
FM (random)	0.729	0.234	0.827	0.656
MF	0.828	0.116	0.833	0.840
MF (random)	0.768	0.125	0.775	0.727

FM: females send messages to males; MF: males send messages to females.

themselves or to those who are less popular. For each centrality index of senders, we give the mean and standard deviation of the corresponding receivers' indices, and the proportion of the receivers' centrality indices that are larger than those of the senders' in Figs. 7 and 8 in Additional file 1. For each centrality index, Table 7 presents the proportion of the receivers' centrality indices that are larger than those of the senders' when sending messages. As a comparison, we also give the randomized results. Compared with men, more women tend to send messages to people who are more popular than themselves.

There have been several studies on users' strategic behavior in online dating. Some studies have found a significant positive correlation between the popularity of male and female users. For example, the research by Taylor et al. on the users from the U.S. showed that, they tend to select and be selected by other users whose relative popularity is similar to their own, although it does not necessarily mean a higher success rate, i.e. receiving more responses [62]. A recent empirical analysis of users in four U.S. cities from an online dating site used PageRank to characterize their desirability, and found that, both men and women sent messages to partners who are on average about 25% more desirable than themselves [63]. However, there are also some studies that have not found correlation between users' popularity. For example, the research on users in Boston and San Diego did not find evidence of strategic behavior [33, 34]. Another research on online dating data from a mid-sized southwestern city in the U.S. revealed that, regardless of their own desirability levels which characterize users' physical attractiveness, popularity, personableness, and material resources, both men and women tend to send messages to the most socially desirable users [20]. We find that users on different platforms or in different cultural contexts have different strategic behaviors, and the underlying mechanisms still need to be explored further.

4 Conclusion

In summary, we analyze online dating data to reveal the differences of choice preference between men and women and the important factors affecting potential mate choice. We find that, with compatibility scores considered, when women send messages to men, they pay attention to not only whether men's attributes meet their own requirements for mate selection, but also whether their own attributes meet the requirements of men, while when men send messages to women, they only pay attention to whether women's attributes meet their own requirements. When considering centrality indices, we find that for women, the popularity and activity of the men they contact are significantly positively associated with their messaging behaviors, while for men only the popularity of the women they contact are significantly positively associated with their messaging behaviors. At the same time, we also find that compared with men, women attach greater importance to the socio-economic status of potential partners and their own socio-economic status will affect

their enthusiasm for interaction with potential mates. The machine learning classification methods are used to find the important factors predicting messaging behaviors. At last strategic behavior is analyzed and we find that there are different strategic behaviors for men and women. Although users do not know the centrality indices of themselves and their potential partners, compared with men, for women sending messages there is a stronger positive correlation between the centrality indices of women and men, and more women are inclined to send messages to people more popular than themselves.

This paper provides a foundation for gender-specific preference of potential mate choice in online dating. On the one hand, this study can provide references for the online dating sites to design better recommendation systems. On the other hand, an in-depth understanding of mate preference, such as the compatibility scores, can help users to select the most appropriate and reliable mates. There are still some limitations for the paper. Firstly, we lack the avatar or photo information and the body type data, and thus cannot evaluate the influence of users' physical attraction and body mass index (BMI) on messaging behaviors [33, 34, 64, 65]. In fact, BMI can compensate for the disadvantages of wages or education [65]. Secondly, we only have the message sending data and lack the reply data, which makes it impossible for us to study the interaction between users. Thirdly, the lists of potential partners presented to users are generated by the recommendation algorithm of the website, not the result of users' own search, and therefore could not reflect users' preference well. Ranking effects caused by recommendation algorithms in online environments have been shown to influence the music people select [66] and the politicians people favor [67]. Fourthly we study the users' attribute preference without considering the potential impact of other attributes. In real life, sending a message to another user is usually not affected by a single attribute. The additional attributes included in users' profiles—their avatar, place of residence, and marital status—could also influence whether a message was sent or not, which means that the users' preference for an attribute can be an illusion and may be based on other considerations. Fifthly, there are significant differences between Chinese and western cultures, and the website is only for heterosexual users, thus the conclusions of this paper may not be applicable to western society or homosexual people [68, 69]. Finally, people's preferences for certain attributes in potential partners can change over time [70], while we only study users' preferences in mate choice at a particular time. There are several avenues for future research. We can examine the influence of recommendation algorithms on potential mate choice in online dating. We can also use the results obtained in the paper to further study the problem of stable matching for potential mate choice. And by combining game theory with the real online dating data, we can further understand the users' behaviors.

Additional material

Additional file 1: Supplementary information (DOC 6.5 MB)

Abbreviations

MobileF, Whether a female mobile phone is verified; HouseF, Whether a female has a flat; AutoF, Whether a female has a car; LevelF, Female credit rating; Pop1F, Female pop_1; Pop2F, Female pop_2; Pop3F, Female PageRank (pop_3) in the messaging network; IndegreeF, Female indegree in the click network; OutdegreeF, Female outdegree in the click network; CompatFM, The compatibility score between a female preference and the profile of the corresponding other side; MsgFM, Whether females send messages to males; FM, Females send messages to males; MobileM, Whether a male mobile phone is verified; HouseM, Whether a male has a flat; AutoM, Whether a male has a car; LevelM, Male credit rating; Pop1M, Male pop_1; Pop2M, Male pop_2; Pop3M, Male PageRank (pop_3) in the messaging network; IndegreeM, Male indegree in the click network; OutdegreeM, Male outdegree in the click network; CompatMF, The compatibility score between a male preference and the profile of the corresponding other side; MsgMF, Whether males send messages to females; MF, Males send messages to females; RS, Mean of the compatibility scores of a sender and the corresponding receiver; BMI, Body mass index.

Availability of data and materials

The datasets supporting the conclusions of this paper are available in the figshare repository, https://doi.org/10.6084/m9.figshare.6429443.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HH and XS designed the research and wrote the paper. XS and HH preprocessed the data and performed the data analysis. All authors reviewed the manuscript, read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Hu H, Wang X (2009) Evolution of a large online social network. Phys Lett A 373:1105–1110
2. Hu HB, Wang XF (2009) Disassortative mixing in online social networks. Europhys Lett 86, 18003
3. Hu H, Wang X (2012) How people make friends in social networking sites—a microscopic perspective. Physica A 391:1877–1886
4. Xia P, Zhai S, Liu B, Sun Y, Chen C (2016) Design of reciprocal recommendation systems for online dating. Soc Netw Anal Min 6:32
5. Finkel EJ, Eastwick PW, Karney BR, Reis HT, Sprecher S (2012) Online dating: a critical analysis from the perspective of psychological science. Psychol Sci Public Interest 13:3–66
6. Rosenfeld MJ (2017) Marriage, choice, and couplehood in the age of the Internet. Sociol Sci 4:490–510
7. Cacioppo JT, Cacioppo S, Gonzaga GC, Ogburn EL, VanderWeele TJ (2013) Marital satisfaction and break-ups differ across on-line and off-line meeting venues. Proc Natl Acad Sci 110:10135–10140
8. He QQ, Zhang Z, Zhang JX, Wang ZG, Tu Y, Ji T, Tao Y (2013) Potentials-attract or likes-attract in human mate choice in China. PLoS ONE 8:e59457
9. Schwarz S, Hassebrauck M (2012) Sex and age differences in mate-selection preferences. Hum Nat 23:447–466
10. Li NP, Yong JC, Tov W, Sng O, Fletcher GJO, Valentine KA, Jiang YF, Balliet D (2013) Mate preferences do predict attraction and choices in the early stages of mate selection. J Pers Soc Psychol 105:757–776
11. Huang J, Kumar S, Hu C (2019) Physical attractiveness or personal achievements? Examining gender differences of online identity reconstruction in terms of vanity. In: Mohamad Noor M, Ahmad B, Ismail M, Hashim H, Abdullah Baharum M (eds) Proceedings of the regional conference on science, technology and social sciences (RCSTSS 2016). Springer, Singapore, pp 91–99
12. Buss DM (1989) Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. Behav Brain Sci 12:1–14
13. Trivers R (1972) Parental investment and sexual selection. Biological Laboratories, Harvard University, Cambridge
14. Todd PM, Penke L, Fasolo B, Lenton AP (2007) Different cognitive processes underlie human mate choices and mate preferences. Proc Natl Acad Sci 104:15011–15016
15. Castro FN, Hattori WT, de Araújo Lopes F (2012) Relationship maintenance or preference satisfaction? Male and female strategies in romantic partner choice. J Soc Evol Cult Psychol 6:217–226
16. Rosenfeld MJ, Thomas RJ (2012) Searching for a mate: the rise of the Internet as a social intermediary. Am Sociol Rev 77:523–547
17. Stauder J (2014) Friendship networks and the social structure of opportunities for contact and interaction. Soc Sci Res 48:234–250
18. Lin KH, Lundquist J (2013) Mate selection in cyberspace: the intersection of race, gender, and education. Am J Sociol 119:183–215
19. Tsunokai GT, McGrath AR, Kavanagh JK (2014) Online dating preferences of Asian Americans. J Soc Pers Relatsh 31:796–814
20. Kreager DA, Cavanagh SE, Yen J, Yu M (2014) "Where have all the good men gone?" Gendered interactions in online dating. J Marriage Fam 76:387–410
21. Lewis K (2016) Preferences in the early stages of mate choice. Soc Forces 95:283–320
22. Skopek J, Schulz F, Blossfeld HP (2011) Who contacts whom? Educational homophily in online mate selection. Eur Sociol Rev 27:180–195

23. Skopek J, Schmitz A, Blossfeld HP (2011) The gendered dynamics of age preferences—empirical evidence from online dating. J Fam Res 23:267–290
24. Potârcă G, Mills M (2015) Racial preferences in online dating across European countries. Eur Sociol Rev 31:326–341
25. Curington CV, Lin KH, Lundquist JH (2015) Positioning multiraciality in cyberspace: treatment of multiracial daters in an online dating website. Am Sociol Rev 80:764–788
26. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415–444
27. Laniado D, Volkovich Y, Kappler K, Kaltenbrunner A (2016) Gender homophily in online dyadic and triadic relationships. EPJ Data Sci 5:19
28. Brooks JE, Neville HA (2017) Interracial attraction among college men: the influence of ideologies, familiarity, and similarity. J Soc Pers Relatsh 34:166–183
29. Bapna R, Ramaprasad J, Shmueli G, Umyarov A (2016) One-way mirrors in online dating: a randomized field experiment. Manag Sci 62:3100–3122
30. Becker GS (1973) A theory of marriage: part I. J Polit Econ 81:813–846
31. Becker GS (1974) A theory of marriage: part II. J Polit Econ 82:S11–S26
32. Pollak RA (2017) How bargaining in marriage drives marriage market equilibrium. http://www.nber.org/papers/w24000. Accessed 20 Dec 2017
33. Hitsch GJ, Hortaçsu A, Ariely D (2010) Matching and sorting in online dating. Am Econ Rev 100:130–163
34. Hitsch GJ, Hortaçsu A, Ariely D (2010) What makes you click?—mate preferences in online dating. Quant Mark Econ 8:393–427
35. Jiao Z, Tian G (2017) The Blocking Lemma and strategy-proofness in many-to-many matchings. Games Econ Behav 102:44–55
36. Lee S, Niederle M (2015) Propose with a rose? Signaling in Internet dating markets. Exp Econ 18:731–755
37. Fisman R, Iyengar SS, Kamenica E, Simonson I (2006) Gender differences in mate selection: evidence from a speed dating experiment. Q J Econ 121:673–697
38. Ong D, Wang J (2015) Income attraction: an online dating field experiment. J Econ Behav Organ 111:13–22
39. Fiore AT, Donath JS (2005) Homophily in online dating: when do you like someone like yourself? In: CHI'05 extended abstracts on human factors in computing systems. ACM, New York, pp 1371–1374
40. Wang T, Liu H, He J, Jiang X, Du X (2011) Predicting new user's behavior in online dating systems. In: Tang J, King I, Chen L, Wang J (eds) ADMA 2011: advanced data mining and applications. Lecture notes in computer science, vol 7121. Springer, Berlin, pp 266–277
41. Xia P, Tu K, Ribeiro B, Jiang H, Wang X, Chen C, Liu B, Towsley D (2014) Characterization of user online dating behavior and preference on a large online dating site. In: Missaoui R, Sarr I (eds) Social network analysis—community detection and evolution. Lecture notes in social networks. Springer, Cham, pp 193–217
42. Pizzato L, Rej T, Chung T, Koprinska I, Kay J (2010) RECON: a reciprocal recommender for online dating. In: Proceedings of the fourth ACM conference on recommender systems. ACM, New York, pp 207–214
43. Pizzato L, Rej T, Akehurst J, Koprinska I, Yacef K, Kay J (2013) Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. User Model User-Adapt Interact 23:447–488
44. Tu K, Ribeiro B, Jensen D, Towsley D, Liu B, Jiang H, Wang X (2014) Online dating recommendations: matching markets and learning preferences. In: Proceedings of the 23rd international conference on world wide web. ACM, New York, pp 787–792
45. Szell M, Thurner S (2013) How women organize social networks different from men. Sci Rep 3:1214
46. Kovanen L, Kaski K, Kertész J, Saramäki J (2013) Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. Proc Natl Acad Sci 110:18070–18075
47. Abramova O, Baumann A, Krasnova H, Buxmann P (2016) Gender differences in online dating: what do we know so far? A systematic literature review. In: The 49th Hawaii international conference on system sciences. IEEE Press, New York, pp 3858–3867
48. Bergstrom TC, Bagnoli M (1993) Courtship as a waiting game. J Polit Econ 101:185–202
49. Choo E, Siow A (2006) Who marries whom and why. J Polit Econ 114:175–201
50. Dunn MJ, Brinton S, Clark L (2010) Universal sex differences in online advertisers age preferences: comparing data from 14 cultures and 2 religious groups. Evol Hum Behav 31:383–393
51. Yancey G, Emerson MO (2016) Does height matter? An examination of height preferences in romantic coupling. J Fam Issues 37:53–73
52. Ward J (2017) What are you doing on Tinder? Impression management on a matchmaking mobile app. Inf Commun Soc 20:1644–1659
53. Ellison N, Heino R, Gibbs J (2006) Managing impressions online: self-presentation processes in the online dating environment. J Comput-Mediat Commun 11:415–441
54. Pursey K, Burrows TL, Stanwell P, Collins CE (2014) How accurate is web-based self-reported height, weight, and body mass index in young adults? J Med Internet Res 16:e4
55. Toma CL, Hancock JT, Ellison NB (2008) Separating fact from fiction: an examination of deceptive self-presentation in online dating profiles. Pers Soc Psychol Bull 34:1023–1036
56. Breiman L (1996) Bagging predictors. Mach Learn 24:123–140
57. Breiman L (2001) Random forests. Mach Learn 45:5–32
58. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55:119–139
59. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 785–794
60. Reece AG, Danforth CM (2017) Instagram photos reveal predictive markers of depression. EPJ Data Sci 6:15
61. Besanko D, Dranove D, Shanley M, Schaefer S (2012) Economics of strategy, 6th edn. Wiley, New York
62. Taylor LS, Fiore AT, Mendelsohn GA, Cheshire C (2011) "Out of my league": a real-world test of the matching hypothesis. Pers Soc Psychol Bull 37:942–954
63. Bruch EE, Newman MEJ (2018) Aspirational pursuit of mates in online dating markets. Sci Adv 4:eaap9815

64. McGloin R, Denes A (2018) Too hot to trust: examining the relationship between attractiveness, trustworthiness, and desire to date in online dating. New Media Soc 20:919–936
65. Chiappori PA, Oreffice S, Quintana-Domeque C (2012) Fatter attraction: anthropometric and socioeconomic matching on the marriage market. J Polit Econ 120:659–695
66. Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. Science 311:854–856
67. Epstein R, Robertson RE (2015) The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Proc Natl Acad Sci 112:E4512–E4521
68. Ha T, van den Berg JEM, Engels RCME, Lichtwarck-Aschoff A (2012) Effects of attractiveness and status in dating desire in homosexual and heterosexual men and women. Arch Sex Behav 41:673–682
69. Potârcă G, Mills M, Neberich W (2015) Relationship preferences among gay and lesbian online daters: individual and contextual influences. J Marriage Fam 77:523–541
70. Dinh R, Gildersleve P, Yasseri T (2018) Computational courtship: understanding the evolution of online dating through large-scale data analysis. https://arxiv.org/abs/1809.10032. Accessed 21 Feb 2019