ASYMMETRIES OF MEN AND WOMEN IN SELECTING PARTNER

by

Ömer Başar

B.S., in Computer Engineering, Marmara University, 2006

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering Department
Boğaziçi University
2010

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Assist. Prof. Haluk Bingöl for his valuable criticism, extensive encouragement, continuous guidance and support. His enthusiasm, his great efforts to explain things clearly and simply enabled me to develop a total and comprehensive understanding of the subject throughout this study.

I would like to state my special thanks to Assoc. Prof. Yağmur Denizhan and Dr. Suzan Üsküdarlı for devoting their time, for their guidance, their helpful advice and criticism.

My deepest thanks goes to my fiance Nurgül Köroğlu who has stood by me and encouraged me throughout the entire work. Your endless love, understanding and persistent confidence in me has taken the load off my shoulder.

My family, especially my mother Nihal Başar, deserves very special mention for their inseparable support and love throughout my life. To know that they are always there for me makes me to live in comfort.

I would like to extend my special thanks to my employer Ersan Özer who has provided me with the whole data on which my research is based and who has entitled me to go home everytime I wish to study on my thesis.

It is a pleasure to express my gratitude to Assist. Prof. G. Tarcan Kumkale from Department of Psychology, Koç University, for his interest in our research and for positive attitude.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

ABSTRACT

ASYMMETRIES OF MEN AND WOMEN IN SELECTING PARTNER

People seek their love partner throughout their lives. As the internet grows rapidly, the traditional mate finding ways start to change. The new trend of finding a mate is to join the online dating services.

We work on the dataset of an online dating system in Turkey which has more than 4 million registered users. We investigate the asymmetries of men and women in selecting partner. We find out that some criteria are more crucial. The correlation between some properties are also studied. It is found that educational level and salaries are highly correlated.

Some men do not give up until he gets a message from the woman who impressed him up. We investigate whether the insisting is effective to get attention of women. We also investigate the effect of beauty for both men and women. If a man is handsome, not surprisingly, finding a mate is easier.

It is found that login times according to the gender, educational level and salary are asymmetric. Men with lower educational level or lower salary prefer to login to the system after 8 pm. The physical distance is also an important factor when selecting partner. We showed that men prefer women who lives nearby.

ÖZET

EŞ SEÇİMİNDE KADIN ERKEK FARKLILIKLARI

İnsanlar hayatları boyunca kendilerine uygun bir eş ararlar. İnternetin hayatımızda daha çok yer edinmesiyle birlikte, geleneksel eş bulma yöntemleri daha az tercih edilmeye başlandı. Artık internetten eş aramak daha popüler hale geldi.

Elimizde toplam üye sayısı 4 milyondan fazla olan ve Türkiye'de faaliyet gösteren bir arkadaşlık sitesinin verileri mevcuttur. Bu çalışmamızda eş bulma kriterleri açısından kadınlar ile erkekler arasındaki farklılıkları araştırdık. Ayrıca bazı özelliklerin daha baskın olup olmadığını inceledik. Hangi özelliklerin birbirleriyle alakalı olduğunu gösterdik. Yaptığımız çalışmalara göre eğitim düzeyi ve aylık gelir arasında pozitif bir ilişki bulunmaktadır.

Bazı erkekler, beğendikleri kadını etkileyebilmek için daha çok emek harcarlar. Bu ısrarların, kadınların ilgisini çekebilme üzerindeki etkisini inceledik. Güzelliğin hem kadınlar hem de erkekler için ne kadar önemli olduğunu gösterdik. Yaptığımız çalışmanın sonucuna göre, çekici bir erkek daha kolay eş bulabiliyor.

Kullanıcıların arkadaşlık sitesini kullandıkları saat dilimlerini cinsiyetlerine, eğitim seviyelerine ve aylık gelirlerine göre kıyasladık. Daha düşük eğitim seviyeli ve aylık kazancı daha az olan erkeklerin sisteme daha çok akşam 8'den sonra girdiklerini analiz ettik. Fiziksel mesafenin de eş ararken önemli bir kriter olduğunu gösterdik. Erkekler kendilerine daha yakın oturan kadınları tercih ediyorlar.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

σ	Standard Deviation
μ	Mean value
$\mathcal{F}$	Total number of women
$\mathcal{M}$	Total number of men
$\mathcal{N}$	Total number of users
a	Property age
b	Property body type
e	Property educational level
h	Property height
p	Properties defined in user profile
s	Property salary
z	Property job position
P_i	Value of property p of person i
$\overline{P}$	Average value of property p of some set of users.
API	Application programming interface
BMI	Body mass index
UUID	Universally unique identifier

1. INTRODUCTION

Online dating networks are becoming more popular, as the internet grows. Nowadays, people are seeking for their true love using matchmaking networks.

In online dating networks, almost every necessary data is collected for matchmaking. People descibe themselves, and the profile that they are looking for. So people search for others, who are suitable for themselves. When he finds a suitable candidate, he sends a message to her, and if the woman is also impressed, the conversation begins. Also an offline meeting is arranged by the individuals, in order to move the relationship to the real world.

In this thesis, we analyze the dataset of an online dating network. We investigated if some properties play an important role in partner selection. We also analyzed the asymmetries of male and female users according to the comparable properties such as age and height. As we mentioned before, the users describe themselves, so we studied whether some of these personal data are correlated with each other.

The beauty is an important factor for both male and female users, so the effect of beuty is also investigated. Some men think that women like men who are insistent to attract themselves. We analyzed if the insistence is an effective strategy when finding a mate in online dating networks.

The data used in this thesis belongs to an online network, so the login times are also investigated if there is any asymmetry between gender according to their salaries or edutional levels. Also the effect of neighbourhood is discussed.

Everyday more people join the network, so we observed the advantage of sending a message to a user as soon as she joins the network.

In the next section we will share the related studies and give information about the dataset we have. In Section 3 and 4, we will explain each study in more detail.

1.1. Related Work

Use of internet for social context is rapidly increasing. Online dating is one of the social use of internet where people search for potential partner via virtual dating services. Naturally, dating has strong cultural influence. Penetration of internet is also an important factor in investigation of online dating services.

Results given in the literature demonstrate significant statistical differences between behavioral patterns in different countries. In Germany, compared to the total population, online daters are rather male, younger, higher educated and live in households with a higher income [1]. For Dutch people, online dating was unrelated to income and educational level [2]. The results are based on an online questionnaire that is answered by $\mathcal{N}$=367 single Dutch Internet users between 18 and 60 years old. More than 29% of British people had used an online dating site [3]. $\mathcal{N}$=30,000 UK respondents are considered. 9% of the participants also find a marriage partner using online dating site according to an online survey [3].

Inormous data is collected as people make their moves in an online dating service. A statistical approach on data mining produces interesting relationships in various parameters.

The power law nature of social interactions is described [4]. Power law model is also applied other social networks such as e-commerce [5].

Online dating services use statistics, data mining, and activity monitoring to provide appropriate matches [6].

Some interesting statistical results reported in the literature. A study made in the USA showed that selections are not uniform among subgroups of the population. White males are more likely than non-white men to prefer to date thin and toned women, while African-American and Latino men are significantly more likely than white men to prefer female dates with thick or large bodies [7]. Some criteria of individuals are more important for a subset of users, such as, blue-eyed men prefer women with the same eye color [8].

According to a study about online communication preferences across age, gender, and duration of Internet use, there is no effect of gender for online communication [9]. The main effect for age was significant for online communication with friends and unknown individuals. Longer duration of internet use indicated higher scores on online communication and relationship building.

Time evolution of an online dating network is also analyzed [10, 11]. From time perspective, the dynamics of system use and network structures are discussed [10, 11].

When only the heterosexual relations are considered, the dating networks are bipartite scale-free networks [12, 13]. Using the rate equation approach, the relation between epidemic threshold and infection rates of female-to-male and male-to-female are analyzed [12].

The emergence of age in dating networks is discussed [14, 15, 16, 17]. The age and also the educational level difference play an important role on selecting partners for both male and female individuals [14]. As the age decreases, involvement in online dating may increase [17].

Attractiveness and activity effect on online dating networks is measured [18, 19]. The attractiveness and activity measures show broad power-law-like distributions [18]. Combining the different growth patterns of new entities and attraction patterns of old ones, different heavy-tailed distributions for popularity and activity which have been observed in real life, can be obtained [19].

2. THE DATA

We have the data of one of the largest commercial dating network in Turkey. The aim of the system is to help users to find a partner. People provide some information about themselves and the type of person that they are looking for. Then they start looking for potential partner using the system.

2.1. The Online Dating System

2.1.1. Some Numbers (Size)

As expected, users come and go. The turnover of the system is quite high. More than 3,000 users are registered daily. On the average, a user stays in the system for 3 months. It is a common pattern that user registers the system, uses for a while, finds a partner. Then there is no motivation to stay in the system. So they stop to renewal their membership. Later, the relation did not work. The user registers back to the system. So the exact numbers are continuously changing.

Total number of users of the system is more than 4,500,000. The system is organized according to cities. Istanbul has the largest user population of more than 2,000,000 users. This is unexpectedly high number. Considering the population of the city of Istanbul which is about 13,000,000, one out of five would be a user. This means that the users have more than one account in the system.

2.1.2. Membership

Since the system is a commercial one, there are three different memberships. Female users have no restriction to the usage of the system. On the other hand, there are two types of membership for male. Free of charge standard male membership allows some basic activity in the system. In order to have full access, male should be a gold member by paying a memhership fee. Almost 36% of the users are gold-members and

Table 2.1. Activities-membership matrix

Activity	Standard Male	Gold Male	Female
Text message		✓	✓
Winking		✓	✓
Vote	✓	✓	✓
Favorite list	✓	✓	✓
Virtual gift	1 TL for each virtual gift		

31% of the users are female. The remaining users are standard male users.

The activities that are available for each membership is shown in Table 2.1. The details of activities are given in Section 2.3. Female and gold male users can send text messages to each other, wink at each other. All users in the system may vote or add other users to their favorite lists. The virtual gifts costs 1 TL for each for all users. However, when a user buys a gold membership, he will get 3 free gifts to send to other users.

2.2. User

Users can define themselves, upload their photographs, express themselves through questions and define what they are looking for. In this section, we will explain them in more detail.

2.2.1. Define Yourself

Users define themselves in their profile. Profile information is visible to all users. User can change its profile at any time. Users must provide the following mandatory information about themselves:

- birthday
- marital status

- town where they live in

The optional part of the profile can be listed as follows:

- relation type that they are looking for
- height, weight, eye color and hair color
- body type
- education level
- occupation, job position and salary
- foreign languages that they are speaking
- with whom they are living with
- parenthood status and desire
- smoking or drinking status

For all questions except the “relation type” and “the foreign languages”, user is asked to select one out of predefined list of choices. For “relation type” and “the foreign languages”, user prefers to select more than one, since these questions are multiple choice questions. The details can be found in the Appendix A.

2.2.2. Express Yourself

Users can also express themselves by answering some questions. These questions can be listed as:

- How do you express yourself?
- What are the most important factors in a relationship?
- How long is your longest relationship?
- What are your hobbies?
- What are your favorite music genres?
- Which sports are you interested in?
- How often a person should bath?

The details can be found in the Appendix B.

2.2.3. Define What You are Looking for

As we expressed before, users can also give information about the properties that he or she is looking for their next relationship which can be seen by other members, so that a potential candidate can check if he fits to what is requested. These questions are similar to the questions about the basic profiles. The only difference is that questions for "what you are looking for" are multiple choice. For example, a user can only be single, divorced, separated or widow. However, a user may be looking for both single or divorced people. The details can be found in the Appendix A.

2.2.4. Photographs

In the system, all users can upload 4 profile and 10 album photographs. Uploading photographs is free for all users. The photographs can be deleted or added any time. Photographs can be seen by all members as soon as they are approved.

2.3. Interaction

The interaction is based on activities. Users have a wide variety of activities to communicate with each other. The list of activities are listed in Table 2.1. While the virtual gifts are public, the other activities such as sending text messages are private which are only visible to the sender and the receiver. Let us explain each activity in more detail.

2.3.1. Text Messages

The users can send messages to other users. Sending messages is free for female users, but if a male user wants to send messages, he should be gold member. There is no restriction on the number of messages that can be sent daily. More than 500,000 messages are sent daily.

2.3.2. Winking

One of the activities to communicate with other users is to wink at. When a user winks at other users, he or she may select a message from a predefined list. The seven available wink messages are listed in the Table 3.3. There is also an option to wink at a user without adding a message. More than 50,000 winks are sent daily in the system.

2.3.3. Vote

A user can provide his opinion about another user by means of voting. A user x can vote another user y by giving “yes”, “no” or “maybe”. User y may prefer to ignore user x's vote. User y choose to response by writing back to user x. When both parties vote for “yes”, a “match” happens. This is the begining of a potential partnership. More than 20,000 votes are given daily. Giving votes for other users is free for all users.

2.3.4. Favorite List

Users can make a list of favorite users. Favorite list application is free for all users. When member i has been added to the favorite list of member j, member i has been alerted by the system, so member i becomes aware of member j. Favorite lists provide an effective instrument to users, via which they can closely monitor the members they are interested in. More than 10,000 favorites are added daily.

2.3.5. Virtual Gift

One of the most important way of communication that is widely used in the system is sending virtual gifts. Each virtual gift costs 1 TL. Although it is $\frac{1}{25}$ of a monthly fee, the user community considers virtual gift valuable. Because of the fact that every single user has to pay money to send even only one gift, virtual gifts are more valuable for our research when compared with other types of communication. Moreover, as they are in sending messages, users are not inclined to send gifts to everyone. Therefore, this makes them to be more selective in this respect.

2.4. Software Platform

2.4.1. Live Platform

The live platform has hundreds of gigabytes of data in MySQL database. In order not to degrade the performance of the system, we backup the data and do our analysis on backup machines. Even if we use backup servers, due to large volume of data, our process was slow. We had to partition data in order to get acceptable response times for our queries.

2.4.2. Data Extraction Platform

In our situation it is not possible and feasible, to retrive all data to input data files, and analyze them with MATLAB. Instead of this way, we chose to retrieve data from MySQL with Java. For example, the messages and gifts are stored in 16 different servers and more than 400 database tables. So we should retrieve this data and merge them in order to be able to analyze in MATLAB. The data extraction platform representation is shown in Figure 2.1.

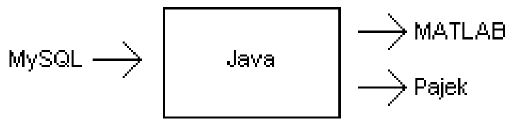


Figure 2.1. Data extraction platform representation.

Later, to increase the efficiency, we wrote a middleware which retrives data from MySQL or any other database, and creates MATLAB objects that are defined in the system. By this way, we automated the process of analysis.

Now, we have a nice, data mining tool that reads data from the database, and creates MATLAB data files and MATLAB code files. Network properties are calculated by means of Pajek which is a network analysis tool [20]. We also wrote a middleware for Pajek which again reads data from database, and creates files that are readible by Pajek.

By means of the middleware, we can speed up our analysis, especially repeated ones. We have plans to make the middleware an open source project.

2.4.3. Privacy Protection

We have access to all data, so we must pay attention to user privacy protection. The most and only critical information of users is the user identification number (userid). So when we are getting data from live server, we get the real userid's, however, when getting backup of these records, we create a universally unique identifier (UUID) for each user. We wrote a method that always converts userid to the same UUID. By the help of this method, the UUID's are all mapped to different userid's, so we know which data is belong to which user, however we do not know who the user is. In other words, there is no way to retrive userid from UUID. That's how we have fully protected the privacy of users.

3. ANALYSIS OF DATA

In this section, we will explain the studies in more detail. We use the convention that $\mathcal{M}$, $\mathcal{F}$, $\mathcal{N}$ represents the numbers of male, female and total users for each particular case, respectively.

3.1. Distribution of Declared Self Profile

Users give their profile information such as educational level, age and body type. In this section, we will group the users by some of their properties. The properties can be listed as below:

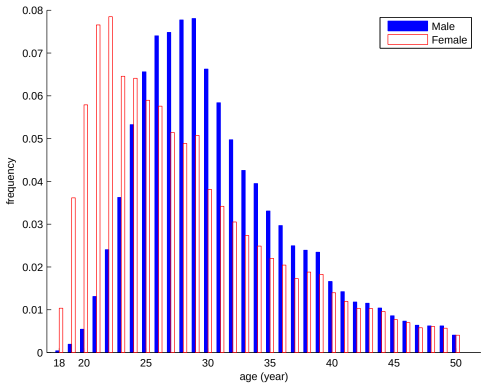
- Age (mandatory)
- Educational level
- Height
- Body type
- Salary

All these informations are based on the users' own declarations, so there is always a chance of lying or giving somewhat biased information.

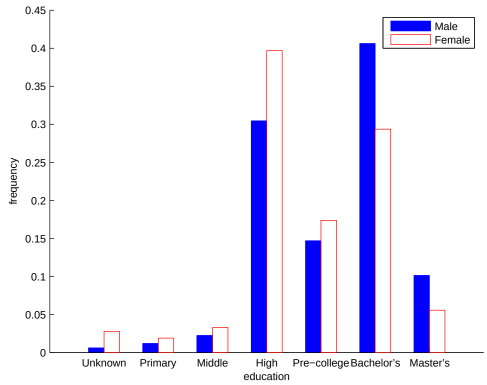
Except the age, the other properties are optional. Since the users who are aged under 18, can not join the network due to legal restrictions, the age is mandatory for all people. On the other hand, a user may not give the salary or the educational level. We only included male users who has gold membership, because the average lifetime of standard male users is very low and standard male users cannot get all the benefits of the network. Due to that, the number of male users are less than the the number of female users. $\mathcal{M}$=276,210 male and $\mathcal{F}$=483,963 female users are included in this study. Total number of users is $\mathcal{N}$=760,173.

3.1.1. Distribution of Age

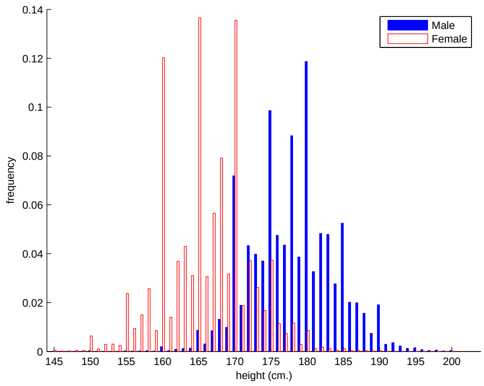
Age is a mandatory field. The average age of the men in the network is 30 years old. On the other hand, average of women is 28 years old. For both men and women the standard deviation is 6 years. Most of men are aged between 25 and 35 and most of women are aged between 20 and 30. It is observed that above 40 years old, the difference of men and women almost disappeared as seen in Figure 3.1(a).



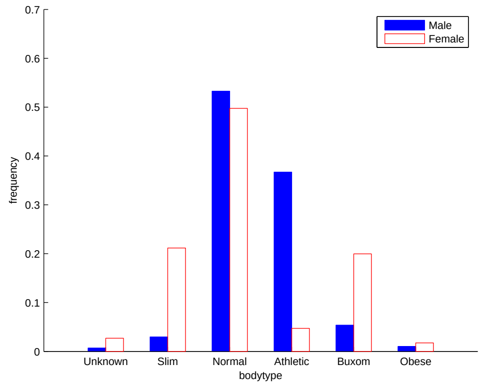
(a) Age



(b) Education



(c) Height



(d) Body type

Figure 3.1. Distribution of users according to the age, educational level, height and body type. ($\mathcal{M}$ = 276,210, $\mathcal{F}$ = 483,963 and $\mathcal{N}$ = 760,173.)

3.1.2. Distribution of Educational Level

The education level is an optional profile information. However, most of the users are giving their educational level. Almost %97 of women and %99 of men are stating

Table 3.1. Percentages of

Property	% of population	
	Male	Female
age	100	100
height	98	88
education	99	97
bodytype	99	97
salary	65	35

their educational level as seen in Table 3.1. High school is the most frequent educational level for women. On the other hand, almost 1 of every 2 men have bachelor's degree or master's degree in the system. The number of users whose educational level are below the high school are very rare for both men and women. Figure 3.1(b) shows the distribution of educational levels in the network.

3.1.3. Distribution of Height

The average height of male users is higher than the average height of the female users. In Figure 3.1(c), the height distribution of men and women are displayed separately. One interesting observation is that when entering height, user prefers to round their height to muliplies of 5. For example, there are few 168 cm, but many 170 cm. The average height of women is $\sigma = 167\ cm.$ with standard deviation of $\mu = 6.46\ cm.$ The average height of men is $\sigma = 178\ cm.$ with standard deviation of $\mu = 5.87\ cm.$

3.1.4. Distribution of Body Type

Also the body type question is optional. Again, the users prefer to share their body type with other users. Almost 3% of women and 1% of men do not give their body type. Normal body type is the most frequent one, for both men and women. Second most frequent body type for male users is athletic, whereas, buxom and slim body types are the second most frequent ones for women. Almost 1 of every 1000 users are obese. In Figure 3.1(d), the distribution of body types are graphed.

3.1.5. Distribution of Salary

The last property for this study is salary. There are three types of answer for salary: "no answer", "answer" and "do not want to share". Most people, 1 out of 2 women and 1 out of 3 men, do not want to share their salaries. The salary of 27% of the male users is between 1,000 TL and 2,000 TL. In Figure 3.2 the distribution of salary is displayed separately for men and women.

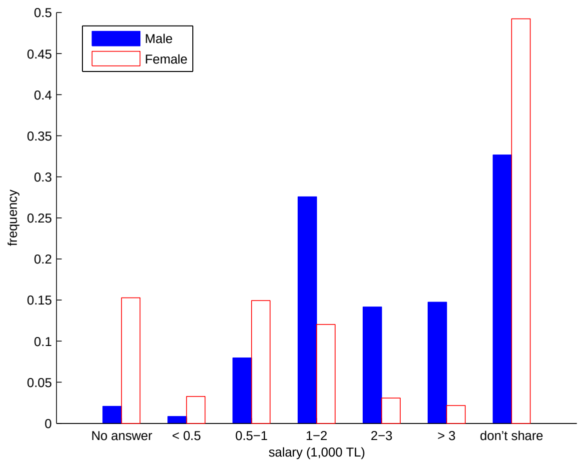


Figure 3.2. User distribution according to salary. ($\mathcal{M}$ = 276,210, $\mathcal{F}$ = 483,963 and $\mathcal{N}$ = 760,173.)

3.2. Property Difference

It is generally believed that people interact with people who has similar characteristics. Consider member i and member j. i does some action to many members in the system. If i does some action to j, it is said that i touches j. It is interesting to investigate if i touches members of the similar properties or not.

3.2.1. Distribution of Property Difference

A member touches many members during his stay in the system. All the members that are touched by member i is denoted by set A_i. Clearly two members j, k in A_i have different values for any given property P. Let P_i denotes the value of the property P given in the profile of i. Then the distance in property P of two members i and j can be represented as $\Delta_{P_i} = P_i - P_j$. Since i touches all the members in A_i, the average value of the property P of A_i is defined as

$$P_{A_i} = \frac{\sum_{\gamma \epsilon A_i} P_\gamma}{|A_i|} \tag{3.1}$$

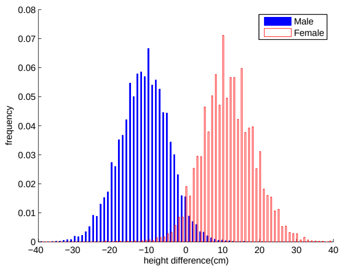
Then the distance in property P between i and all the members that are touched by i, that is A_i, is given as

$$\Delta_{P_i} = P_{A_i} - P_i \tag{3.2}$$

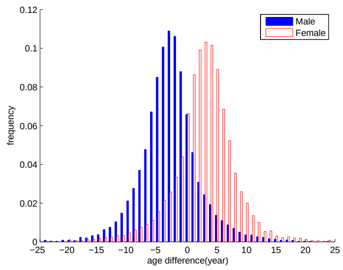
The distribution of Δ_{P_i} for the properties height, age, educational level, salary, BMI and body type properties are analyzed for men and women explicitly. Height, age and BMI properties have numeric values. On the other hand, educational level, salary and body type properties do not have numeric values. So in order to calculate Δ_{P_i} for these properties, they should be converted to numeric values. Since the answers for these properties are all discrete, conversion is easy. The details can be found in Appendix A.

We included the male users who sent at least one gift and female users who replied at least one gift. So $\mathcal{M}$=29,274 men and $\mathcal{F}$ =14,981 women are only considered in this study. Figure 3.3 shows the distribution for male users and female users explicitly. One general observation is that at all properties, male prefer one level low female, and women prefer one level up male.

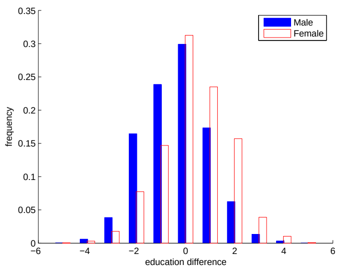
According to Figure 3.3, it can be easily seen that, most of the male users prefer younger and shorter female users. Also women prefer men who are taller and older than themselves. Men prefer women who has a lower BMI. Men also prefer women who has lower educational level and salary. On the other hand, women prefer men with higher education level and salary.



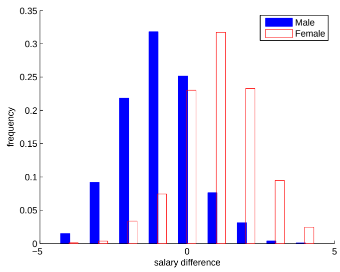
(a) Height



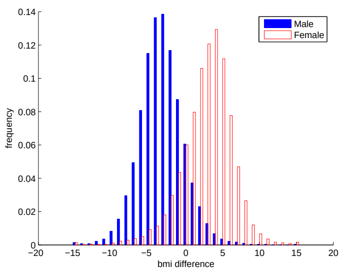
(b) Age



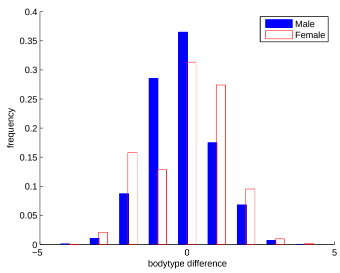
(c) Education



(d) Salary



(e) BMI



(f) Body Type

Figure 3.3. Distribution of Δ_{P_i} according to the height, age, educational level, salary, BMI and body type. ($\mathcal{M}$ = 29,274, $\mathcal{F}$= 14,981 and $\mathcal{N}$ = 44,255.)

3.2.2. Distribution of Normalized Property Difference

The mean values of women for height, age, educational level, salary, BMI and body type are lower than the mean values of men for the same properties. In Table

Table 3.2. μ and σ values for age, height, weight, educational level, bodytype, salary and BMI. ($\mathcal{M}$ = 173,698, $\mathcal{F}$= 214,284 and $\mathcal{N}$ = 387,982.)

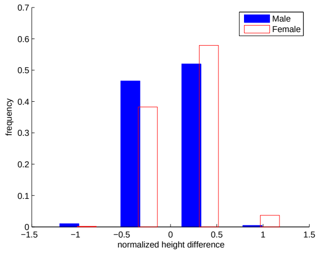
Property	Male		Female		Asymmetry	
	μ_m	σ_m	μ_f	σ_f	$\mu_m - \mu_f$	$\sigma_m - \sigma_f$
age	30.63	6.17	28.23	6.82	2.39	-0.64
height	177.95	5.87	167.20	6.46	10.75	-0.59
weight	75.40	8.91	57.76	8.54	17.64	0.37
education	4.23	1.11	3.99	1.13	0.24	-0.02
bodytype	2.47	0.68	2.24	1.05	0.23	-0.37
salary	3.52	1.00	2.64	1.00	0.88	0
BMI	23.29	2.47	20.19	2.99	3.10	-0.52

3.2, the average μ and standard deviation σ values are listed for each property. So we investigated whether men prefer women over the average for these properties. We also analyzed for women. For this study, we normalized the values. Let us explain how we normalize the values.

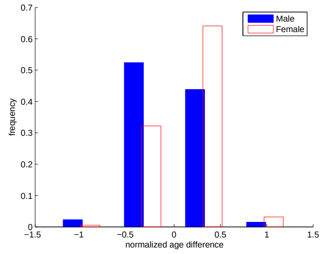
We know the maximum, minimum and mean values for each property such as age. We normalize all users' ages to values from 0 to 1. Here if the age of the user is equal to the mean value of age property, its normalized value is set to 0.5. The minimum age value which is 18 is normalized to 0 and the maximum age value which is 50 is normalized to 1. Then, we analyzed the difference distributions for normalized values as we did in the previous section. Figure 3.4 shows the difference of normalized distributions for each property.

According to the age, there is an asymmetry between male and female users. Female users prefer men who are aged over the average. In contrast, men prefer women who are younger than the average. The histogram for age is shown in Figure 3.4(b) for both men and women.

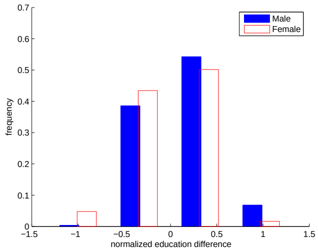
Female and male users both prefer users whose salaries are more than average. In



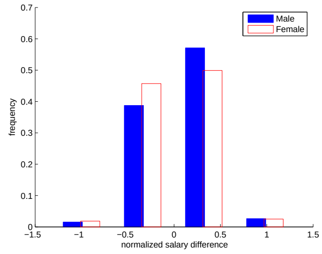
(a) Height



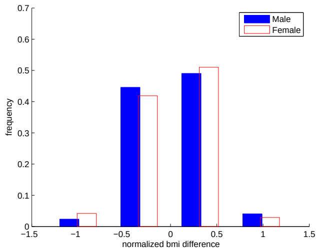
(b) Age



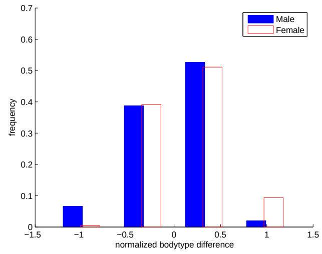
(c) Education



(d) Salary



(e) BMI



(f) Body Type

Figure 3.4. Distribution of the normalized property differences according to height, age, educational level, salary, BMI and body type. ($\mathcal{M}$ = 29,274, $\mathcal{F}$= 14,981 and $\mathcal{N}$ = 44,255.)

Figure 3.4(d), the normalized salary difference histogram is shown. For BMI property, men and women prefer users who have bigger BMI value than average. In Figure 3.4(e) shows the normalized BMI difference histogram.

Lastly, we investigated the body type difference for both men and women. Women prefer men who have more than average body type value. Also men prefer women whose body type is over the average. The Figure 3.4(f) shows the normalized body type difference histogram for men and women separately.

3.3. Gender Classification

First of all, we thought about whether or not we can correctly guess, if a user may vote yes for another user. However when we reliazed that the yes vote frequency of man is very high, this study will not give us a good result. Male users are giving 500 yes votes on average. After that, we wondered if we can guess the gender using classification algorithms [21]. The inputs for the classification are as follows:

- the height of the user
- the average height of users who s/he voted yes
- age of the user
- the average age of people who s/he voted yes
- education level of the user
- education level average of the users s/he voted yes
- salary level of the user
- average salary level of users s/he votes yes.

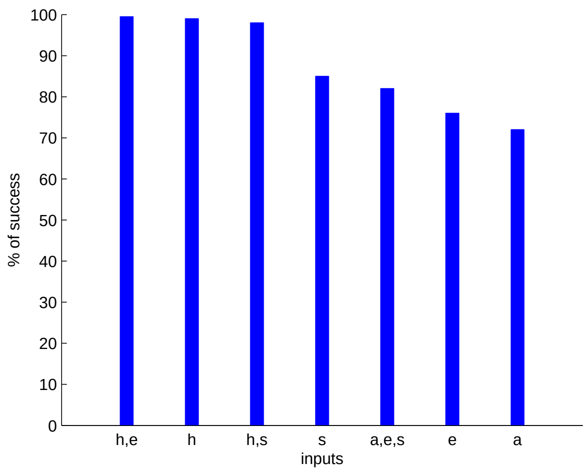


Figure 3.5. Percentage graph of valid gender guesses according to different input types. ($\mathcal{M}$ = 169,845, $\mathcal{F}$= 162,376 and $\mathcal{N}$ = 332,221.)

We selected some of these properties and run the 2-class classification algorithm. We used the arbitrary covariance matrices for classification [21]. The most accurate result is gathered when the height and the salary are the only input types with a

Table 3.3. Wink messages

Value	Language	Message
1	TR	Fotoğrafına bayıldım. Süpersin!
	EN	I loved your photo. You are cool!
2	TR	Fotoğrafını beğendim. Çok güzel gülüyorsun!
	EN	I liked your photo. You smile nice!
3	TR	Profilini beğendim. Ama niye fotoğrafın yok?
	EN	I liked your profile. Why don't you have a photo?
4	TR	Profilini beğendim. Bana kendinden bahsetsene.
	EN	I liked your profile. Tell me about yourself.
5	TR	Birbirimize çok uygunuz. Profilime göz atsana.
	EN	We are suitable for each other. Look at my profile.
6	TR	Bir merhaba ile başlar her şey... Merhaba!
	EN	Everything starts with a hello. Hello!
7	TR	İstediğin eğlenceli bir arkadaş ise ben buradayım.
	EN	If you're looking for a funny friend, I am here.
8	TR	Mesaj eklemek istemiyorum.
	EN	I don't want to add a message.

percentage of 99.1%. When the only input parameter is height, the accuracy is 98.2%, too. If only age is used as input parameter, the accuracy decreases to 70%. It is the worst accuracy among the input parameters. In Figure 3.5 the valid percentage results are shown. In the figure, h stands for height, s stands for salary, e stands for educational level and a stands for age.

3.4. Distribution of Wink Messages

One of the communication ways is to wink at other users with an optional message. The message must be selected from a predefined list. The available messages are shown in the Table 3.3.

In this study, we analyzed if there exists a group of users which prefers a particular message. It is also investigated if there is any asymmetry between male and female users according to the frequency of selected messages while winking at. $\mathcal{M}$ = 216,891 male and $\mathcal{F}$=241,748 female users are included in this study. The number of male users are almost 10 times bigger than the number of female users in this study, because female users do not prefer winking as much as the male users do.

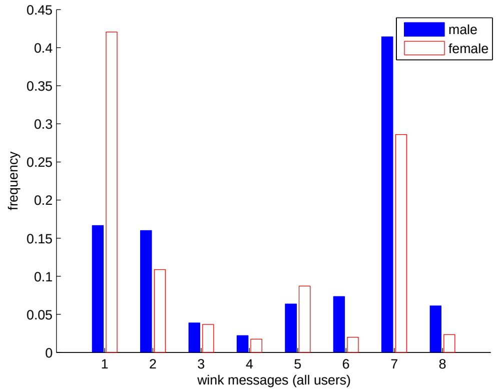
In Figure 3.6, we can see the frequencies of each messages according to the gender of the sender. There are 4 graphs. In Figure 3.6(a), we show the frequencies of selected messages of all users. The message "If you're looking for a funny friend, I am here." is the most prefered message by male users. In almost one of every two winks, this message is selected. For female users this message is the second most prefered message. The most frequent message of female users is "I loved your photo. You are cool!".

When we compared the graphs that contains age intervals, we can clearly state that, the frequency order of the messages is not changing as the age interval changes for male users. However, as the age increases, the frequency of the message "If you're looking for a funny friend, I am here." increases.

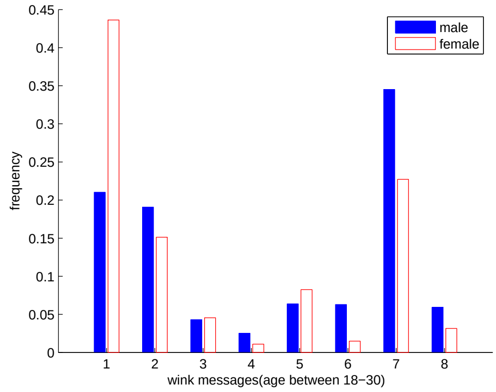
On the other hand, when we compared the graphs for female users, we can also say that the frequency of the message "If you're looking for a funny friend, I am here." increases, as the age increases. Although the frequencies of messages are different for male and female users, there is a common pattern of all users. The message "If you're looking for a funny friend, I am here." is prefered more as the age increases.

3.5. Login Time Analysis

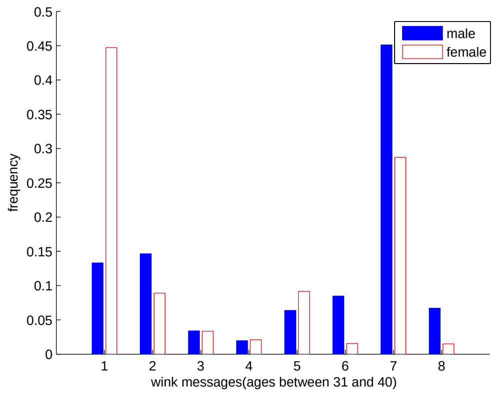
People use the system at different hours. We have login times of each individual. However, once a user logs in, we do not know how long he stays in the system. In this study, we investigated which hours are more prefered for login. We also analyzed whether or not there is asymmetry for male and female users according to the login hours. Finally we grouped users by their educational levels and salaries in order to



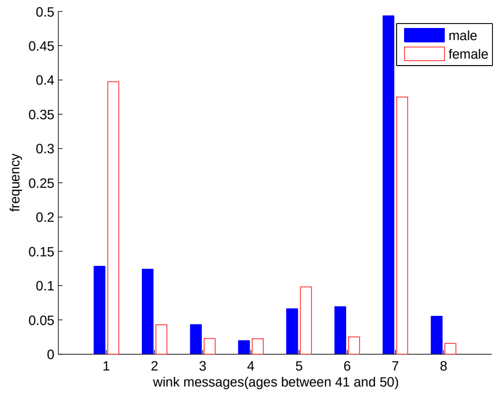
(a) All users



(b) Users between ages 18-30



(c) Users between ages 31-40



(d) Users between ages 41-50

Figure 3.6. Distribution of wink messages according to gender and age. ($\mathcal{M}$ = 216,891, $\mathcal{F}$= 241,748 and $\mathcal{N}$ = 458,639.)

investigate any difference between different groups.

This study contains users who are logged in the year of 2007. Some people register to the system but not use much. We consider people who use the system more than 20 times, so the users considered in this study are low when compared with other studies. $\mathcal{M}$ = 44,504 male and $\mathcal{F}$= 59,456 female users are considered.

As expected, the users mostly prefer to login after 21:00. The crowd increases as the time goes to 1:00. After 1 am the number of logins starts to decrease. Figure 3.7 shows the graphs for male and female users separately. For male users there is a small decrease in the number of logins at 14:00. It is because many users who are working are logged into the system at their lunch time and after the end of the break, some of them are logged out from the system. Female users log on system between 17:00 - 19:00 rarely. A possible explanation could be that they are busy preparing dinner. The peek time for male users is 22:00 and for women is 23:00.

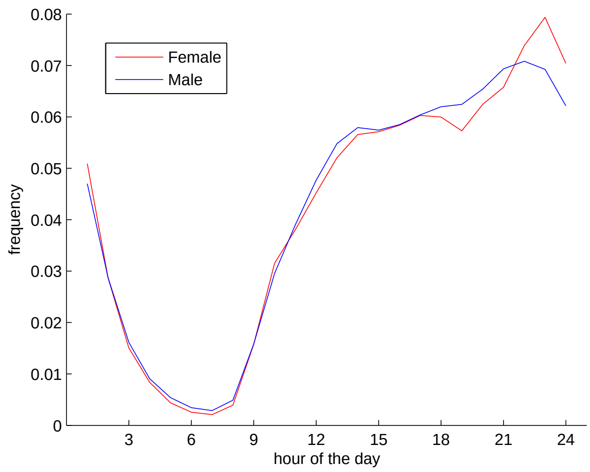
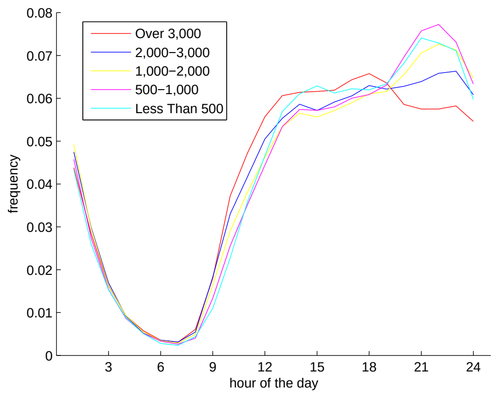


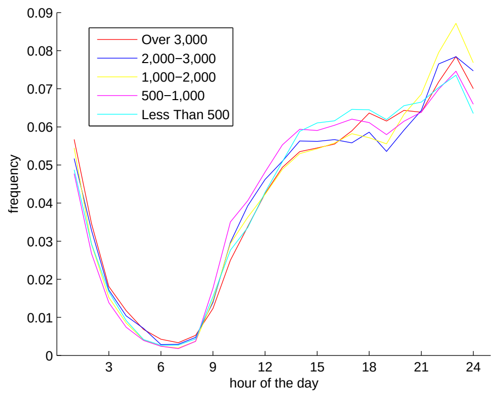
Figure 3.7. Login hour histogram according to the gender. ($\mathcal{M}$ = 44,504, $\mathcal{F}$= 59,456 and $\mathcal{N}$ = 103,960.)

After the analysis of all logins made by all users, we investigated if login times differ according to the salary or education level of individuals. First of all, we separated the users according to their salaries as given in Table A.7. In the Figure 3.8(a), we can clearly see that, the men who has lower salary prefer logging into the system after 19:00. However, the male users with higher salaries log into the system at daytime more when compared with users who has lower salaries. It will be explained as, the users with low salaries can not find much chance to log into the network at daytime. So they can only login at night. On the other hand, for female users, the login times does not have a pattern when the users are grouped by salary. Contrary to male users, the female users whose salary are under 500 TL do not prefer night time to log in to the system.

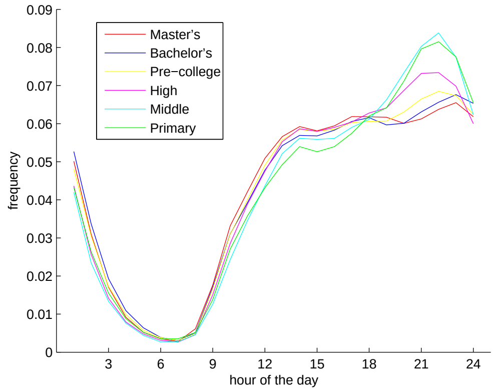
Finally, we separated users according to their educational level for males and females separately as given in Table A.5. The Figure 3.8(c) shows the histogram of logins of male users who are grouped by their educational levels. The male users with



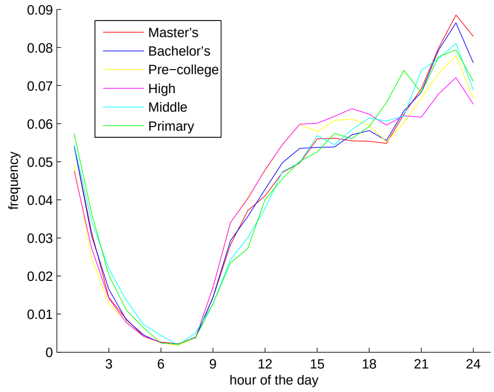
(a) Male users grouped by salary



(b) Female users grouped by salary



(c) Male users grouped by educational level



(d) Female users grouped by educational level

Figure 3.8. Login hour histogram according to the educational level and salary.($\mathcal{M}$ = 44,504, $\mathcal{F}$= 59,456 and $\mathcal{N}$ = 103,960.)

higher level of education prefer logging into the system at daytime. On the other hand, the users with lower level of education are logging into the system more than the users with higher level after 18:00. There is an asymmetry in login histograms between male and female users when grouped by educational levels. The female users with higher educational levels do prefer to login more after 22:00, however, male users with higher educational levels do not.

When we look deeper to the Figure 3.8(a) and 3.8(c) graphs for education level and salary, we can easily see that they are very similar to each other. At the graphs, the highest level of education level and salary have the similar characteristics. Similarly, the lowest level of education level and salaries have the similar characteristics. At the section 3.6, we will see that there is a positive correlation between salary and education level of male users. So we will say that, the results are consistent with each other.

3.6. Correlations

In this part of the study we will investigate if there is any correlated properties. We have chosen one-tailed statistical significance level of $\alpha = 0.05\%$. We randomly selected 500 users for each gender. We studied separately for males and females. We also searched for correlation with the properties of the actual choices. Pearson's linear correlation algorithm is used while computing the correlation between two variables [22].

For various properties such as height and education level, we create $n \times 1$ vectors where i^{th} entry is the value of the i^{th} person and n is the number of people. For example, h_i^p is the height of the i^{th} person where p represents the profile. Similarly for actual values there is a corresponding h_i^a where h_i^a is the average of the values of people who i^{th} person is voted "yes".

Here are the vectors that are used:

- h^p: $n \times 1$ height vector of users.
- e^p: $n \times 1$ education vector of users.
- s^p: $n \times 1$ salary vector of users.
- a^p: $n \times 1$ age vector of users.
- z^p: $n \times 1$ job position vector of users.

- h^a: $n \times 1$ age average vector of users who are voted "yes".
- e^a: $n \times 1$ education average vector of users who are voted "yes".
- s^a: $n \times 1$ salary average vector of users who are voted "yes".
- a^a: $n \times 1$ age average vector of users who are voted "yes".
- z^a: $n \times 1$ job position average vector of users who are voted "yes".

Table 3.4. Correlation table for women

	h^p	e^p	s^p	a^p	z^p	h^a	e^a	s^a	a^a	z^a
h^p	1.00									
e^p	0.17*	1.00								
s^p	0.18*	0.42*	1.00							
a^p	-0.17*	0.12	0.07	1.00						
z^p	0.15*	0.16*	0.26*	-0.07	1.00					
h^a	0.01	-0.03	-0.01	-0.11	-0.05	1.00				
e^a	0.04	0.18*	0.17*	0.19*	0.00	0.07	1.00			
s^a	-0.08	0.06	0.09	0.33*	-0.06	0.10	0.42*	1.00		
a^a	-0.14	0.08	0.06	0.84*	-0.07	-0.18*	0.17*	0.36*	1.00	
z^a	-0.04	-0.03	-0.05	0.18*	0.06	-0.01	0.15*	0.29*	0.19*	1.00

*Significant at 0.05% level.

Correlations of properties are given in Table 3.4 and 3.5. Since no correlation with body type is found, the body type is not shown in the tables. The highest correlated properties are age (a^p) and actual choice age (a^a) properties for both male and female users. The correlation coefficient for female users is 0.84, twice of the correlation for male users. The education level (e^p) and salary (s^p) properties are also highly correlated for both men and women. Not surprisingly, the job position level (z^p) and the salary (s^p) are highly correlated for all users. As expected the education level (e^p) and age (a^p) are not correlated. Surprisingly, for women, height (h^p) is correlated with educational level (e^p), salary (s^p), age (a^p) and job position (z^p) which is hard to explain. However, for male users, there is no correlation for any property with height (h^p).

It is found that job position (z^p) and age (a^p) are correlated for male users. On the other hand, the correlation between job position (z^p) and age (a^p) is not statistically significant for female users. One possible explanation is that the female users are not working at older ages, the correlation will be weaker.

We calculated the unemployment rates for different ages both for males and

Table 3.5. Correlation table for men

	h^p	e^p	s^p	a^p	z^p	h^a	e^a	s^a	a^a	z^a
h^p	1.00									
e^p	0.08	1.00								
s^p	0.10	0.35*	1.00							
a^p	-0.08	0.09	0.31*	1.00						
z^p	0.05	0.19*	0.41*	0.21*	1.00					
h^a	0.08	-0.02	0.00	-0.12	-0.04	1.00				
e^a	0.02	0.16*	0.10	0.09	0.05	0.08	1.00			
s^a	0.03	0.10	0.06	-0.01	-0.02	0.23*	0.41*	1.00		
a^a	-0.09	0.05	0.19*	0.41*	0.11	-0.23*	0.13	0.01	1.00	
z^a	-0.04	0.01	-0.04	-0.05	0.00	0.16*	0.14	0.25*	-0.04	1.00

*Significant at 0.05% level.

females. The graph is shown in the Figure 3.9. The X-axis shows the ages from 18 to 60 grouped by 4 years. The Y-axis shows the percentage of the unemployment. The red line shows the female percentages. The percentages of male are shown by the blue line. It is easily seen that the percentages for males are very stable when compared with the percentages of females.

The unemployment percentage of male users is always lower than 2%. On the other hand, the percentage of female users starts with 5% for the users whose ages are between 18 and 22, reaches its highest percentage on the age group from 38 to 42. After the age 42, the percentage decreases.

We also calculated the correlation between job position and age for female users without the unemployed users. The correlation is increased from -0.005 to 0.061. It is also not statistically significant, but the correlation gets stronger.

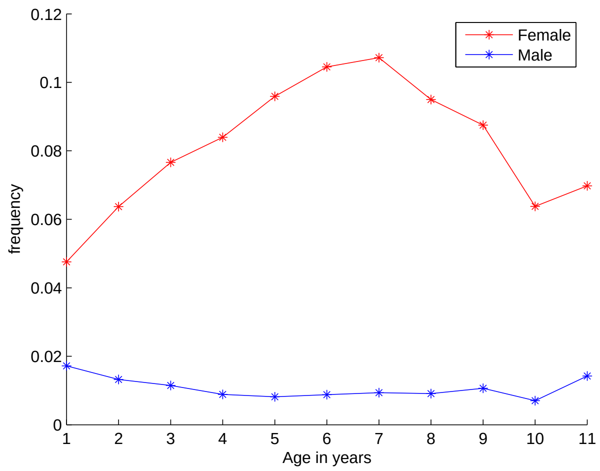


Figure 3.9. Unemployed percentages by age.

3.7. Profile, Declared Preference and Actual Choice Distribution

For each property p, there are three cases. A person specifies his value in his profile, called "profile". He also specifies the same property for the partner that he is looking for which is called "declared preference". Finally, he actually votes for some people. The profile value of voted people is called "actual choice". Profile value is unique. On the other hand declared preference and actual choice values could be multiple choices which must be handled. If there are n choices, then each value would have $\frac{1}{n}$, so that the total would sum up to 1 as we have for the profile.

Each person has unit contribution to profile, declared preference and actual choice. Profile value is easy since it is a single value. Declared preference and actual choice values are not single valued. If he selects, n options, then his conribution to each option is $\frac{1}{n}$ so that overall contribution is unit.

A person votes for many people. Each vote means one point for the profile value of the voted person. If one votes k times, his contribution should be normalized by $\frac{1}{k}$ in order to obtain the unit contribution.

In Table 3.6, an example for person i is shown. c_1, c_2, c_3, c_4 are choices for property

Table 3.6. Profile, declared preference and actual choice example for person i

P_i	c_1	c_2	c_3	c_4	Σ
profile		1			1
declared preference	$\frac{1}{3}$		$\frac{1}{3}$	$\frac{1}{3}$	1
$vote_1$	1				1
$vote_2$		1			1
$vote_3$	1				1
$vote_4$			1		1
$vote_5$	1				1
actual choice	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{0}{5}$	1

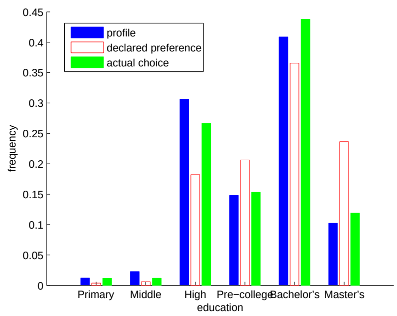
p. Person i has $P_i = c_2$. His first vote, v_1 goes to a person with c_1. His second vote, v_2 to c_2, so on. Note that if he makes a new vote, actual choice value vector changes. He can change his profile value and declared preference vector at any time. We take the current value of the profile value and the declared preference vector.

In this study we analyzed three properties:

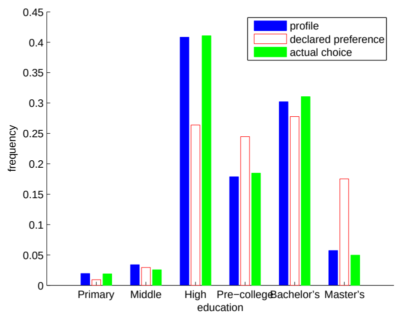
- Education level
- Salary
- Body Type

These properties are all comparable. For this purpose we sorted all of the available values ordered by their degree.

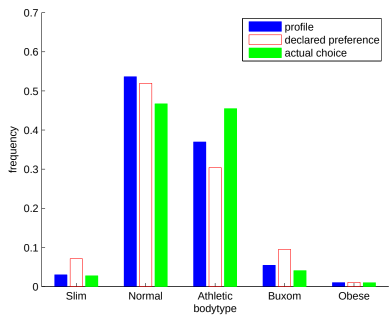
In Figure 3.7, the frequencies of profile, declared Preference and actual choice are shown. For women, the frequencies of profile and actual choice are very close to each other. So, although men states some declared preferences, they do not strict to them, even if a women do not satisfy their declared preferences. They tend to say “yes”. On the other hand, women strict to their preferences. If a man does not fit to their declared preferences, they tent to say “no”. So this can be interpreted as women is the one who makes the decision.



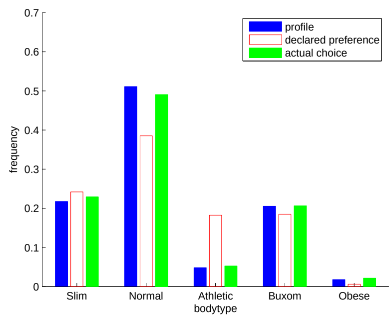
(a) Education for men



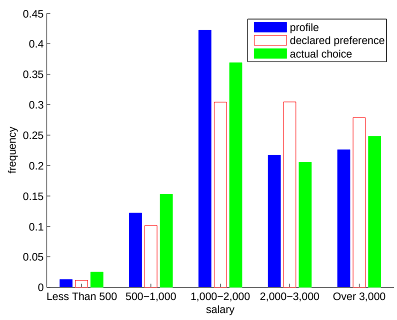
(b) Education for women



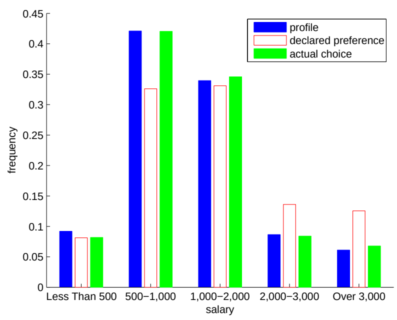
(c) Body type for men



(d) Body type for women



(e) Salary for men



(f) Salary for women

Figure 3.10. Profile, declared preference and actual choice distribution of educational level, body type and salary.

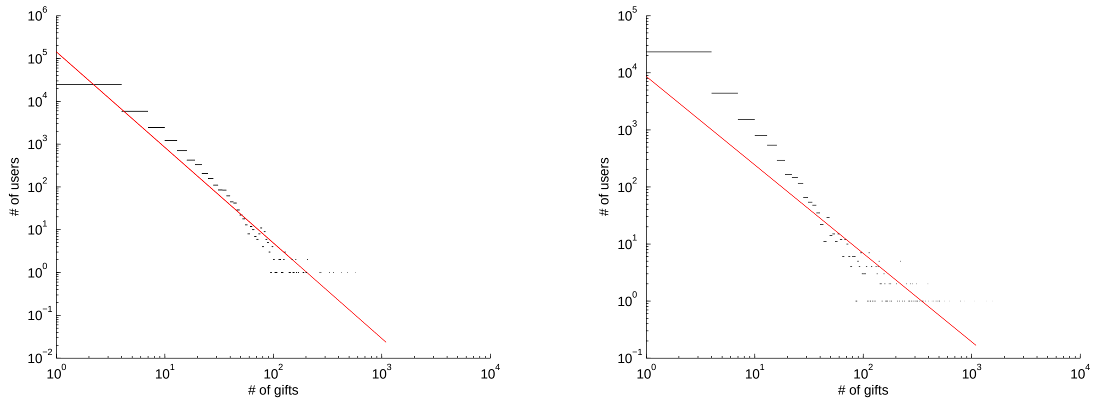
4. ANALYSIS OF VIRTUAL GIFTS

Any member can send gift to any other member. The number of gifts received by a member can be seen by other members. Therefore receiving more gifts is a kind of value in the system. It is observed that some members send gift to themself in order to increase their perceived value.

In this section, we will explain the studies that we made about virtual gifts. There is a very sharp asymmetry of gender in sending gifts. Although all users can send virtual gifts, female usage is almost none. 99.9% of the gifts are sent by male users, so in this section we only analyze gifts that are sent by male users.

4.1. Gift Distribution

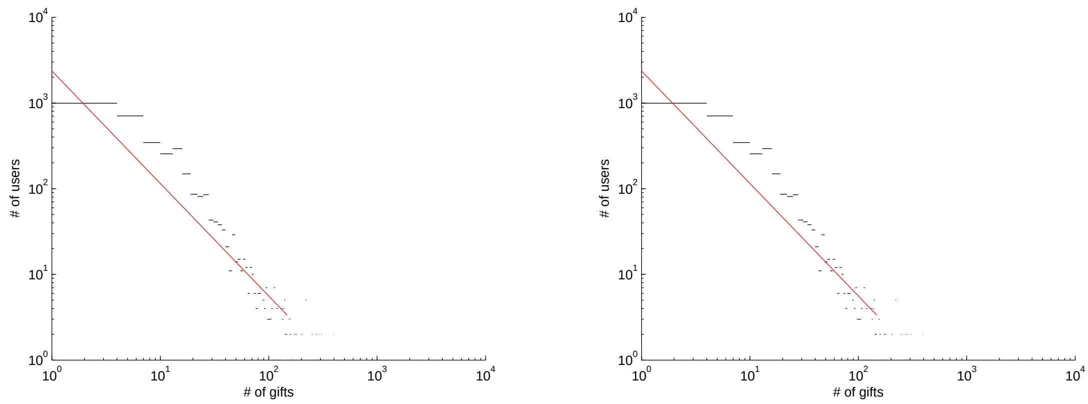
One person can sent as many gifts as he wants. On the other hand, a person can receive many gifts. In order to understand the gift process, two different distributions are investigated. In this study, we analyzed 166,213 gifts sent by $\mathcal{M}$=31,938 male users to $\mathcal{F}$=36,934 female users. Total number of users in this study is $\mathcal{N}$=68,872.



(a) Received gifts (b) Sent gifts

Figure 4.1. Gift distribution power-law graphs. ($\mathcal{M}$ = 31,938, $\mathcal{F}$ = 36,934 and $\mathcal{N}$ = 68,872.)

If the frequency of having k gifts is proportional to $k^{-\gamma}$ and γ is larger than 1, then we can state that, the distribution shows power law characteristics. We discovered that both received and sent gift distributions shows power-law characteristics as seen in Figure 4.1, since for the receiving case $\gamma_{received} = 2.23$ and for the sending case $\gamma_{sent} = 1.55$.



(a) Received purchased gifts (b) Sent purchased gifts

Figure 4.2. Purchased gift distribution power-law graphs. ($\mathcal{M}$ = 3,591, $\mathcal{F}$ = 23,310 and $\mathcal{N}$ = 26,901.)

As a part of gold membership packages, user gets 3 gift tokens. This has its effect on the distribution of sent gifts. Another way to get gift tokens is to purchase them. The data for purchased gifts only is obtained by discarding the gifts that are sent free of charge. This reduces the numbers to $\mathcal{M}$=3,591 male users to $\mathcal{F}$=23,310 female users. Total number of purchased gifts is 71,740. The distributions of purchased gifts are also found to be power law where $\gamma_{sent} = \gamma_{received} = 1.31$. The power law graphs of purchased gifts can be seen in Figure 4.2.

4.2. Effect of Insistence

One of the activity that is widely used to communicate with other users is sending virtual gifts. The usual scenario is that male user sends a message to a female user. Women decides whether to return the message or not. Sending gift means, the sender pays real money though it is a small amount. The cost of a gift depends on which service person is using but it is in the range of 1/20 of a monthly membership fee. A person i can send virtual gift to person j. One hopes to increase his chances of getting

Table 4.1. Number of gifts sent

gifts	Messages	Returns	Success Ratio
0	9493	1711	18.03
1	1633	478	29.28
2	548	172	31.39
3	462	141	30.52
4	264	84	31.82
5	162	63	38.89
6	78	33	42.31
7	63	33	52.39

a message from person j if he sends more gifts to person j. Validity of this hypothesis is investigated in this study.

It is possible that person i and j are already in touch. If two people are in touch, it means that person i and j have exchanged at least one message before. This case is not what we are looking for. We consider only people that are not previously touched. So this reduces our data set to $\mathcal{M}$ = 9,603 male users, $\mathcal{F}$= 21,511 female users. Total number of users considered in this study is $\mathcal{N}$ = 31,114.

The data shows that sending more gifts to the same woman has big impact. Sending two or three gifts increases the success rate. Sending more gifts also increases the rate of getting a message. The visual presentation is shown in the Figure 4.3 by the label "all users".

We also investigate effect of some properties in this study. We take the subset of men whose some properties are set. For example when we analyze the men who are 5 cm taller than the women, we obtain better success rates. Height of the male is declared by h_m, while h_f is the height of the female user. In the Figure 4.3, the graph is labelled by $h_m > h_f + 5$. As expected the return rate increased.

Finally in this study, contribution of the educational level is included. We take the subset of men who have a higher educational level than the women that are being sent gifts. Education level is denoted by e_m for male users and e_f for the female users. In the Figure 4.3, the graph is labelled by $h_m > h_f + 5, e_m > e_f$. It is also expected that if the educational level of male user is higher than that of women, the success rate still increases. When a person i sends 8 gifts to the person j, the percentage of getting a message from person j is more than 60%.

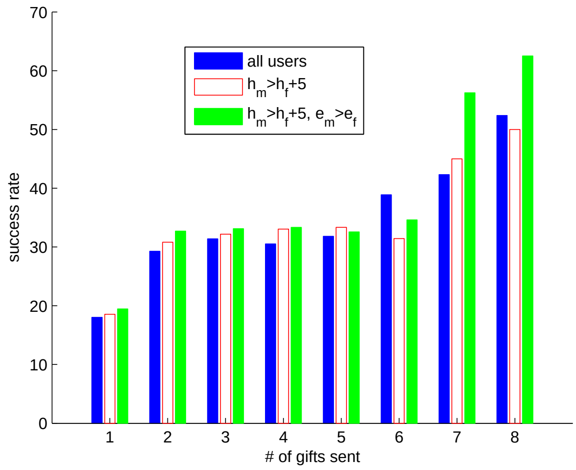


Figure 4.3. Effect of insistance graph. ($\mathcal{M}$ = 9,603, $\mathcal{F}$= 21,511 and $\mathcal{N}$ = 31,114.)

4.3. Effect of Beauty

In the Section 4.2, we already showed that sending more gifts to the same women increases the rate of success. Let's analyze the beauty effect of getting a reply from a woman.

As a company procedure, all photographs are checked by operations department. During this process, they mark users if they found them beautiful or handsome. This marking is used to select front page person that is randomly selected to be put on the front page as one enters the system. These marked ones are called "featured" users. Having marked as featured is a metric of physical beauty or attractiveness decided by human operators. Also users do not know which users are marked as featured.

Table 4.2. Number of featured and not featured users.

	To Featured	To Not Featured
From Featured	21,456	11,649
From Not Featured	30,473	22,097

We already showed that a man has a higher rate of getting a reply from a female when he sends more than one gift. So in order to be able to analyze the effect of beauty, we take the subset of men users who send a single gift to a woman. We also did not include the users who are already in touch, as we do in the Section 4.2. This reduction on data causes $\mathcal{M}$ = 29,056, $\mathcal{F}$= 32,865 and $\mathcal{N}$ = 61,921.

We separated users into four groups. These group are are defined as:

- FN: <u>F</u>eatured male sending a gift to <u>N</u>ot featured female.
- FF: <u>F</u>eatured male sending a gift to <u>F</u>eatured female.
- NN: <u>N</u>ot featured male sending a gift to <u>N</u>ot featured female.
- NF: <u>N</u>ot featured male sending a gift to <u>F</u>eatured female.

where F, N coding means featured and not featured, respectively. The number of users for each group are given in Table 4.2.

The male users that are selected as featured, are presumably handsome so it is expected that they will get higher success rates. So, not surprisingly, the FN group has the highest return rate. If a featured male user sends a gift to a woman that is not selected as featured, he will get a message with a success rate of 17%. NF case, as expected, has the lowest success rate, since the featured women are less likely to send messages to not featured male users. Interestingly NN case is the second highest. Unexpectedly FF case has the third return rate. Figure 4.4 shows the results of beauty effect in gifts.

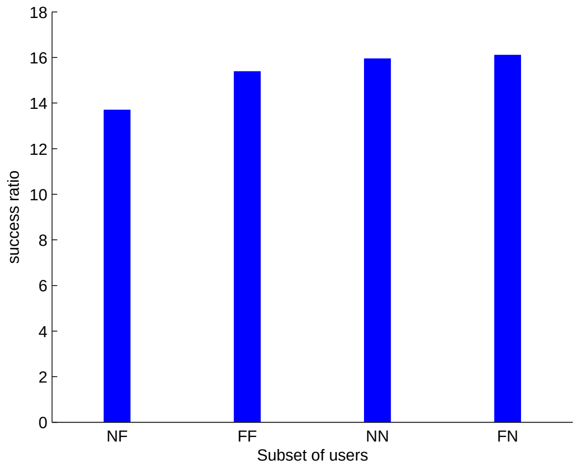


Figure 4.4. Effect of beauty graph. ($\mathcal{M}$ = 29,056, $\mathcal{F}$= 32,865 and $\mathcal{N}$ = 61,921)

4.4. Effect of First Gifts

The next study about the gifts is about the effect of first gifts. Male users are sending the 99.9% of the gifts that are sent in the system. So a woman potentially gets many gifts. Those gifts could be from different men as well as from the same man. We investigate the effect of getting the first gifts. It is expected that the very first gift is more important than the receiving the 20^{th} gift. Since we are investigating the success of the gifts, in this study we do not include the users who are already in touch. As we already said before, some men send more than one gift to the same woman. So if it is the case, we do not include the gifts except the very first one. $\mathcal{M}$=27,251 male and $\mathcal{F}$=32,871 female users satisfy these constraints are considered. The success rate for the very first gift is 18%. If the sender is selected as featured, the success rate is more than 20%. This is close to 1 out of 5 gifts gets successful return. But if the gift sent is not the first or the second gift of the woman, success rate drops below to 13%. Therefore being the very first gift sender is a very good strategy once we have this global information about the system.

We showed that sending a female's first gifts is an important factor in order to get a reply. However, some men already discover this strategy, by their experience on the system. These men usually watch for females as soon as they join to the system.

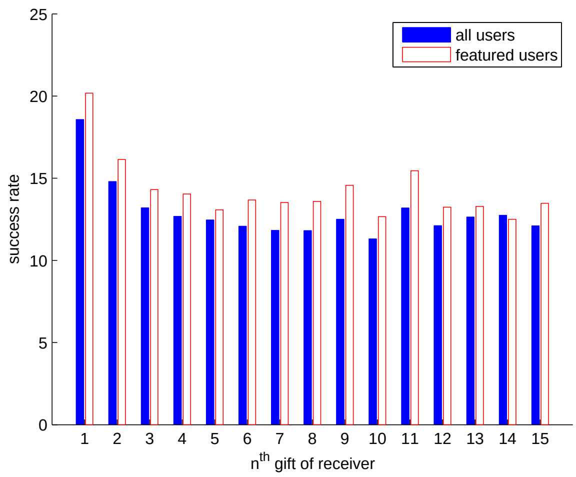
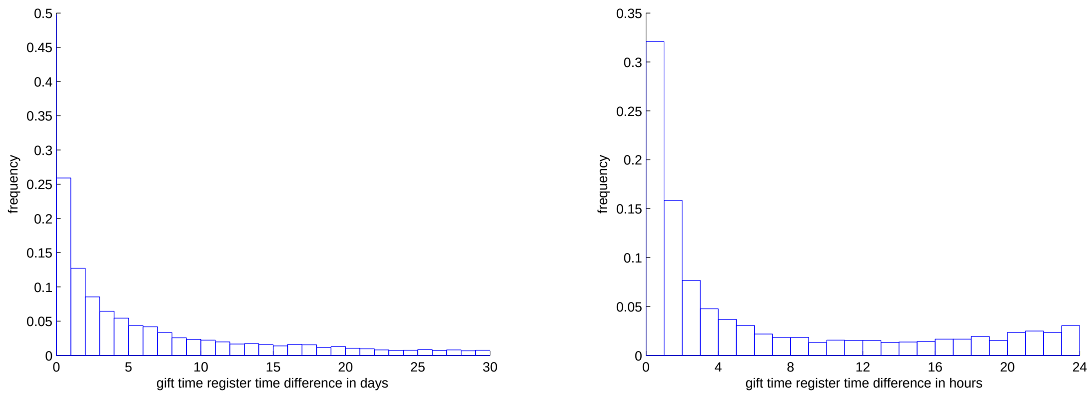


Figure 4.5. Effect of first gifts graph. ($\mathcal{M}$ = 27,251, $\mathcal{F}$= 32,871 and $\mathcal{N}$ = 60,122.)

In order words, some men know that sending the very first gift of a female, increases their chance of receiving a reply from her.

There are 55,928 distinct users who sent one or more gifts. Out of that, 3,764 users sent nothing else but the very first gifts. This 6.73% of the gift senders maximize their return of investment to 18%.



(a) First 30 days of the registration

(b) First 24 hours of the registration

Figure 4.6. Number of women who received their very first gifts in the first 24 hours or 30 days of the registration.

In the Figure 4.6, we can see the histograms of how early the female users get their first gift. There are two histograms in the figure. Figure 4.6(a) shows the number of female users who received their first gifts in the first 30 days of the registration.

Figure 4.6(b) shows the histogram of female users who received their first gifts in the first 24 hours of registration. We can state that the most of the female users received their first gifts in their first day and more than that most of them receive their first gifts in the first hour of registration to the system.

4.5. Effect of Neighbourhood

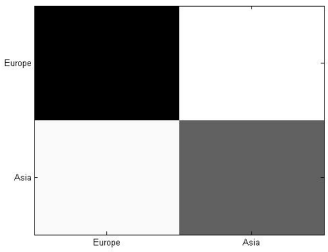
When the users join to the system, they should give the name of the town where they live in. The most important concept of the system is that the users can search for friends by the towns or even by the districts. So in this study, we analyzed if the users really care about the towns of the users, when they are selecting partner. We made this study in Istanbul. Istanbul is the most populous city in Turkey with its more than 10 million habitants. Istanbul resides in two continents, Asia and Europe, and these two sides are separated by the Bosphrous. Transportation in Istanbul is a serious problem. Due to that problem, it is expected that side where your partner lives in Istanbul should be important. It is also expected that users should prefer users who live in their towns.

First we investigated, if the Asia part and Europe part of Istanbul effects the users while selecting partner. Then we expanded this study by analyzing the town effect.

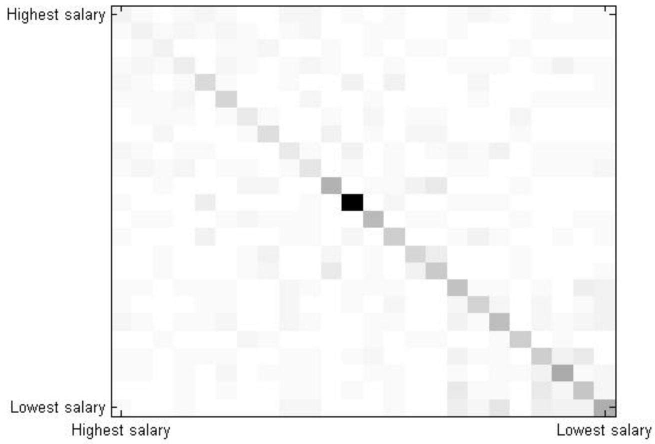
We made this study for male users. We used the gift data to analyze the town effect.

Let t_i, t_j be two towns. Consider those people in t_i sending gifts to people in t_j. The percentage of gifts sent from t_i to t_j can be given as

$$c_{ij} = \frac{\#\, of\ gifts\ sent\ from\ a\ person\ in\ t_i\ to\ a\ person\ in\ t_j}{\#\, of\ gifts\ sent\ from\ a\ person\ in\ t_i}$$



(a) Gray Scale Graph for Europe and Asia sides of Istanbul



(b) Gray Scale Graph of Towns of Istanbul

Figure 4.7. Gray scale graphs for towns.

Figure 4.7(a) gives the two sided view when t_i is Europe, t_j in Asian sides of Istanbul. There is a very strong tendency to being in the same side.

One expects that people have tendency to go out with people of the similar economical status. In order to check that districts are sorted with respect to property tax rates and gift flow from the members living in district t_i to the members living in t_j calculated. No meaningful results obtained.

Instead of using the property tax rate, average incomes of the members living in the district t_i is calculated. The districts are ordered with respect to the average income of the members. Then gift flow from a district t_i to a district t_j becomes meaningful as seen in Figure 4.7(b).

The tendency of staying in the same neighbourhood is clearly observed since the main diagonal has darker gray values. Another interesting observation is that lowest-lowest corner has some accumulation compared to highest-highest corner. Mid-range neighbourhoods has the highest values.

5. CONCLUSION

In this study, we work on the dataset of an online dating network in Turkey. First, we showed the distributions of age, educational level, height, salary and body type for male and female users. We investigated the difference distributions for age, height, educational level, monthly salary, BMI and body types values. Women prefer older, taller, thinner, more educated male users. On the contrary, men prefer younger, shorter and less educated women. We also normalized these differences and find out that men prefers younger women than the average.

Using classification algorithm, we guessed the gender of users with a success rate of 99.1% by using the input values salary and height. According to the wink distribution study, we analyzed that as the age increases, the frequency of the message "If you are looking for a funny friend, I am here." increases.

We find out the asymmetries of login hours according to the gender, educational level and monthly salaries. One of the findings of this study is that neighbourhood is an important factor for users.

We also showed the correlations between height, educational level, salary, age, job position, body type and their actual choice values. There are some asymmetries for men and women. For instance, position and age are correlated for male users, however they are not correlated for women.

We also showed the distributions for profile, declared preference and actual choices. It is found out that the expectations for men are not as high as women. So if a relation starts, the woman decides.

Received and sent gift distributions are found to be power law. Also the purchased received and sent gift distributions suits power law. We showed that if a man send more gifts to a single woman, he increases his chance of getting a reply. We also find out that beauty and attractiveness is an important factor in partner selection. Finally we showed that as the order of the received gift increases, the percentage of sending a reply increases.

Here are some of the topics of our future work.

- We will investigate if there is a group of men who are using the system more efficiently than other users. If there is such a group, we will also try to find out the similarities of these men.
- We will make users tell us the reason why they are living the system. Then we will analyze this data.
- We analyzed the neighbourhood effect on selecting partner. As the next step we will try to create a social network according to the average of salaries of their habitants. Than we will study on this network.

APPENDIX A: BASIC PROFILE QUESTIONS

User defines himself by answering questions in the profile. There is no fill in the blank type of questions. All the answers are selected from a list of possible answers. Therefore answers are discrete.

The aim of the system is to make people to find a suitable partner for their next relationship. So users can select what kind of relationship they are expecting. There are five options for this question and more than one option can be selected. These options can be found in the Table A.1.

The next profile questions are about the eye and hair color of the users. There are eight options for each question. The options are listed in the Table A.2 and Table A.3.

Height can be between 140 cm and 200 cm. Users can also set their weight from 40 kg. to 200 kg. Although users can enter their height and weight, they can also state their body types.

According to the National Heart Lung and Blood Institute of U.S (NHLBI), the humans are divided into four groups [23]. These groups are underweight, normal weight, overweight and obesity. They are ordered by their range of Body Mass Index (BMI) values. BMI is a measure of body fat based on height and weight that applies to both men and women. In our dataset there are five groups. They are slim, normal, athletic, buxom and obese. The only difference is that "normal weight" group according to NHLBI coverts both the "normal" and "athletic" of our dataset. So we assigned values

Table A.1. Options for relationtype questions.

cyber friendship	friendship	marriage
short-term relationship	long-term relationship	

Table A.2. Options for eye color

brown	light brown	hazel	green
blue	black	gray	other

Table A.3. Options for hair color

black	brown	yellow	red
auburn	gray	white	other

1 to 5 according to the BMI order. By this way, we are able to graph the body type property and also to compare two body type values. In Table A.4, the body types are ordered according to their BMI ranges.

Users can also declare which foreign languages they can speak. A user may select more than one language. The languages that are available in the system are German, Arabic, Bulgarian, Chinese, Persian, French, Dutch, English, Spanish, Italian, Japanese, Russian, Greek. For the remaining languages there is also "other language" option.

One of the basic profile questions is "Who are you living with?". The anwers for this question are "alone", "with my children", "with housemate", "with roommate" and "with family". For any other answers there is also an option "others".

Table A.4. Options for body type

Value	Option	NHLBI Name	BMI Range
1	slim	underweight	< 18.5
2	normal	normal weight	$18.5 - 24.9$
3	athletic	normal weight	$18.5 - 24.9$
4	buxom	over weight	$25 - 29.9$
5	obese	obesity	> 30

Table A.5. Options for educational level

Value	Option	Year in Education
1	Primary School	5
2	Middle School	8
3	High School	11
4	Pre-college	13
5	Bachelor’s Degree	15
6	Master’s Degree	17

There is also a question about their status of parenthood. There are four options for this question. They can be listed as:

- I have no child
- I have children and living with them
- I have children but now living with them
- I have children and sometimes living with them

Another question is about the child desire. A user may declare whether or not he wants a child or not in the future.

If a user does not drink alcohol, he can select the option “I am not drinking alcohol”. If he drinks alcohol, he can enter the frequency of how often he drinks alcohol. There are two options, sometimes and often. A user can also declare whether or not he is smoking cigarette.

A.1. Educational Level

Another important question of basic profiles is about the education level of the users. There are six education levels defined in the system. The educational level is also quantified to graph and calculate the differences between users. The options are ordered by their number of years in education. The values for each educational level are

Table A.6. Options for job position

Value	Option
1	Inoccupied
2	Part time
3	Full time
4	Middle Level Manager
5	High Level Manager

Table A.7. Options for salary

Value	Option
1	$<$ 500 TL
2	500 - 1,000 TL
3	1,000 - 2,000 TL
4	2,000 - 3,000 TL
5	$>$ 3,000 TL

shown in Table A.5. The 8-year compulsory education is enacted in 1997. So most of the users in the system is graduated before the law, that is why the “Primary School” option is still available.

Another quantified property is “job position”. The options for the job position are inoccupied, part time, full time, middle level manager and high level manager. In Table A.6 shows the values of these options.

A.2. Salary

The last question of the basic profiles is about the salary. The users can select their salary from a predefined list. The options for salary question are between ranges. The maximum salary that can be selected in the system is “more than 3,000 TL”. The users who earns higher than 5,000 TL in a month, select the same answer for these question. So we can not know the exact salary of indivuals, however we can compare two users by their salaries. The salary is also quantified to be able to graph and analyze the users according to their salaries. The options for the salaries are shown in Table A.7. The values are in ascending order. There is also an option as “I do not want to answer”, however, in our analysis the users who have selected this option are discarded.

APPENDIX B: EXPRESS YOURSELF QUESTIONS

Express yourself questions are composed of several questions.

B.1. How do you express yourself

There are 24 available options for this question. A user may select more than one option. Some of the answers are open minded, optimist, friendly, etc... Full list can be found in Table B.1

B.2. What are the most important factors in a relationship

An individual may select more than one option, however, it is advised not to enter more than needed. Because it is asked to enter answers that are the most crucial ones for the user. Some of the answers to these questions are "sexual harmony", "educational conformity", "carreer conformity", etc... All available answers can be seen in Table B.2.

B.3. How long is your longest relationship

A user may select a single answer for this question. The available answers for this question can be listed as:

Table B.1. How do you express yourself?

Open minded	Domestic	Friendly	Reticent
Serious	Religious	Honest	Enjoyable
Introverted	Optimist	Funny	Talkative
Pessimistic	Adventurer	Marginal	Modern
Conservative	Sexy	Sympathetic	Jealous
Ashamed	Responsible	Creative	Smart

Table B.2. What are the most important factors in a relationship?

Sexual harmony	Educational Conformity	Intention to marry	Carreer conformity
Income conformity	Confidence	Tolerance	To have a lively time
Political Conformity	Romanticism	Faithfull	Respect
Love	Mental harmony		

Table B.3. What are your hobbies?

0-3 months	more than 1 years	more than 5 years
3-6 months	more than 2 years	no relationship
6-12 months	more than 3 years	
Housework	more than 4 years	

B.4. Hobbies

There are several hobbies that are defined in the system. Overall list can be found in Table B.4.

B.5. What are your favorite music genres

Available answers can be found in Table B.5.

Shopping	Photography	Reading	Sports
Computer	Night life	Art	Television
Dance	Web	Travelling	Theathre
Housework	Music	Cinema	Cooking

Table B.4. What are your hobbies?

Table B.5. What are your favorite music genres?

Arebesque	Latin	Turkish Art	Turkish Rock
Blues	New age	Turkish Pop	Pop
Jazz	Techno	Turkish Rap	Rap
Classical	Turkish Folk	Turkish Rock	Rock

Table B.6. Which sports are you interested in?

Athletism	Ice Skating	Football	Motor Sports
Basketball	Outdoor sports	Skiing	Tennis
Billiards	Fitness	Ping Pong	Volleyball
Bodybuilding	Sailing	Swimming	

B.6. Favorite Sports

More than 12 most popular sports are defined in the system. The available sports are shown in Table B.6

B.7. How often a person should bath

' The available answers are listed as:

- Several times a day
- Once a day
- Several times a week
- Once a week
- Several times a month

REFERENCES

1. Schulz, F., J. Skopek, D. Klein, and A. Schmitz, “Who is using internet dating platforms in Germany?”, *Zeitschrift Fur Familienforschung*, Vol. 20, No. 3, pp. 271–292, 2008.

2. Valkenburg, P. M. and J. Peter, “Who visits online dating sites? Exploring some characteristics of online daters”, *Cyberpsychology & Behavior*, Vol. 10, No. 6, pp. 849–852, Dec 2007.

3. Gunter, B., “Internet dating: a British survey”, *Aslib Proceedings*, Vol. 60, No. 2, pp. 88–98, 2008.

4. Raban, D. R. and E. Rabin, “Statistical inference from power law distributed web-based social interactions”, *Internet Research*, Vol. 19, No. 3, pp. 266–278, 2009.

5. Stephen, A. T. and O. Toubia, “Explaining the power-law degree distribution in a social commerce network”, *Social Networks*, Vol. 31, No. 4, pp. 262–270, OCT 2009.

6. Smith, A., “Exploring online dating and customer relationship management”, *Online Information Review*, Vol. 29, No. 1, pp. 18–33, 2005.

7. Glasser, C. L., B. Robnett, and C. Feliciano, “Internet Daters’ Body Type Preferences: Race-Ethnic and Gender Differences”, *Sex Roles*, Vol. 61, No. 1-2, pp. 14–33, Jul 2009.

8. Laeng, B., R. Mathisen, and J.-A. Johnsen, “Why do blue-eyed men prefer women with the same eye color?”, *Behavioral Ecology And Sociobiology*, Vol. 61, No. 3, pp. 371–384, Jan 2007.

9. Thayer, S. E. and S. Ray, “Online communication preferences across age, gender,

and duration of Internet use", *Cyberpsychology & Behavior*, Vol. 9, No. 4, pp. 432–440, Aug 2006.

10. Panzarasa, P., T. Opsahl, and K. M. Carley, "Patterns and Dynamics of Users' Behavior and Interaction: Network Analysis of an Online Community", *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 5, pp. 911–932, May 2009.

11. Holme, P., C. Edling, and F. Liljeros, "Structure and time evolution of an Internet dating community", *Social Networks*, Vol. 26, No. 2, pp. 155–174, May 2004.

12. Luo-Sheng, W., Y. Xiao-Fan, and Z. Jiang, "Two-sex epidemic spreading on bipartite scale-free networks", *Acta Physica Sinica*, Vol. 57, No. 8, pp. 4794–4799, AUG 2008.

13. Ergun, G., "Human sexual contact network as a bipartite graph", *Physica A*, Vol. 308, No. 1-4, pp. 483–488, May 15 2002.

14. Ajrouch, K., A. Blandon, and T. Antonucci, "Social networks among men and women: The effects of age and socioeconomic status", *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, Vol. 60, No. 6, pp. S311–S317, Nov 2005, 55th Annual Meeting of the Gerontological-Society-of-America, BOSTON, MA, NOV 22-26, 2002.

15. Suna, T., M. Hardey, J. Huhtinen, Y. Hiltunen, K. Kaski, J. Heikkonen, and M. Ala-Korpela, "Self-organising map approach to individual profiles: Age, sex and culture in Internet dating", *Sociological Research Online*, Vol. 11, No. 1, Mar 30 2006.

16. Greenlees, I. and W. McGrew, "Sex and Age-Differences in Preferences and Tactics of Mate Attraction - Analysis of Published Advertisements", *Ethology and Sociobiology*, Vol. 15, No. 2, pp. 59–72, Mar 1994.

17. Stephure, R. J., S. D. Boon, S. L. MacKinnon, and V. L. Deveau, "Internet Initiated Relationships: Associations Between Age and Involvement in Online Dating", *Journal of Computer-mediated Communication*, Vol. 14, No. 3, pp. 658–681, Jul 2009.

18. Ghoshal, G. and P. Holme, "Attractiveness and activity in Internet communities", *Physica A*, Vol. 364, pp. 603–609, May 15 2006.

19. Hu, H., D. Han, and X. Wang, "Individual popularity and activity in online social systems", *Physica A*, Vol. 389, No. 5, pp. 1065–1070, Mar 1 2010.

20. Wouter de Nooy, V. B., Andrej Mrvar, *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, 2005.

21. Alpaydın, E., *Introduction to Machine Learning 2e*, The MIT Press, 2010.

22. Ronald E. Walpole, S. L. M. K. Y., Raymond H. Myers, *Probability and Statistics for Engineers and Scientists, 8th Edition*, Prentice-Hall (Pearson), 2006.

23. National Heart, L. and B. Institute, "Calculate Your BMI", 2010, `http://www.nhlbisupport.com/bmi/`.