Minnesota Academy of Science Journal of Student Research

HELLO STRANGER: A BIG DATA APPROACH TO ONLINE DATING

E. Berman and R. Vijayakar
Breck School, Minneapolis, MN

Almost half of all single adults in the United States have used online dating, a still-growing field. Anonymity on dating sites is often marketed as a premium feature, but its effectiveness in different demographic groups has not been shown by published research. Therefore, our first goal was to obtain permission for a popular online dating site to access a dataset containing demographics and interactions of 100,000 de-identified users. We then ran "Big Data" analytics on the dataset to compare the quality and quantity of online-dating matches when a user uses a dating site anonymously or non-anonymously. The authors used R-programming language to clean the set and sort users, as well as run advanced statistical tests. Results showed that anonymity significantly decreases quantity of matches ($p < 0.0001$) and does not significantly increase quality of matches ($p > 0.05$), despite the fact that the dating site markets anonymity as a beneficial feature.

The second goal was motivated by a literature search that revealed the Gale-Shapley deferred acceptance matching algorithm, which generates mathematically stable matches. First, linear regressions were programmed in R to generate preference formulae for all users. Then, preference-assigning software was written in Java to apply these formulae to established matches. Finally, the algorithm was programmed and adapted to use the preference-assigning software to produce matches. The authors' application removes a massive limitation on the use of the Gale-Shapley algorithm, making it applicable to many fields, from matching residents and hospitals to students and colleges.

Keywords: Online dating, anonymity, matching algorithm, interest signaling
Abbreviations: VAS — valid, active, and straight

Funding and resources were provided by the University of Minnesota Carlson School of Management and Breck School. No conflicts of interest exist.

This paper received a 2017 Minnesota Academy of Science STEM Communicator Award. The judging process for this award satisfied the function of manuscript review.

INTRODUCTION

Millions of single people use the Internet to search for prospective romantic partners using online dating sites. In the United States, 46% of all single adults, or approximately 40 million adults, have used online dating. One in five new relationships and one in six new marriages have originated from the Internet (*1*, *2)*. Online dating is a valuable search method for potential partners, as it reduces social stresses and obligations inherent in offline search situations.
Most online dating sites allow users to browse other users' profiles and utilize a site-generated algorithm to recommend matches (*2)*. The algorithms that sites use are typically unavailable to the public, as they are proprietary, but most will employ some kind of machine-learning method, or neural networking, to be an adaptive system, incorporating feedback from users (*2)*. However, these search methods do not always generate viable matches, and the experience of online dating is not necessarily the same for all users, as some users may have different types of memberships on dating sites, such as paid or free. The business model of many online dating sites is often referred to as "freemium" (*1)*. This indicates that the site is free for users to join, but users can then choose to pay the site for additional features, including anonymity. Many sites market anonymity as

a premium feature, as it allows users to browse other profiles unseen by those whose profiles they view, providing the anonymous user more personal privacy. This study investigated whether anonymity affects the matching process, as anonymity has the potential to change the nature of interactions between users. The effect of anonymity on different demographics in online dating contexts was examined.

The majority of online dating sites use algorithms to suggest matches to users. One algorithm that is ideally suited to a marriage model is the Gale-Shapley deferred-acceptance algorithm (*3*). This algorithm, originally formulated to match high school students with colleges, takes into account a list of users, each with an individual list of preferences, and returns matches that are stable, meaning that no two unmatched users prefer each other over the person with whom they are matched. The Gale-Shapley matching algorithm was implemented in this study to improve the efficiency of online dating. A similarity or difference between the demographics of the algorithm-predicted match and the real-life outcome could indicate how predictable matches are.
Our study is based on recent research in decision sciences and social literature, as online dating is relatively new. Bapna et al. (2013) examines the overall effect of anonymity in online dating markets, specifically finding that it has an asymmetric effect across gender (*1*).

In Bapna et al. (2013), it is hypothesized that anonymity decreases inhibitions when it comes to less socially acceptable matches; for example, anonymous users may be more likely to view users of the same sex or a different race than regular users (*1*). However, Bapna et al. (2013) never identified the differences in the effect of anonymity between demographics outside of gender (*1*). This study will observe the effect of anonymity on users across different age groups, races, genders, and combinations of these demographics.

For a user, both anonymity and non-anonymity may have associated costs and benefits. When a user is not using anonymous settings, other users may observe whether the focal user has viewed his or her profile, which has an inherent cost to the focal user, such as potentially leading to an unwanted interaction. Therefore, non-anonymity has a personal cost for the user. However, when a user is anonymous, other users will not be able to see that the focal user has viewed their profile. This does not encourage an interaction, thus there is a lost opportunity. Yet anonymity may lead to an increase in match quality, as it allows the focal user to be more selective (*1*).

Anonymity has the potential to increase the number and diversity of searches, as it should reduce inhibition in general, but also in what are considered socially taboo preferences, such as same-gender matches or interracial matches (*4*). This is a more specific example of the general reduction of search costs that anonymity can provide, which may also lead to users showing their true preferences, rather than those influenced by societal expectations. In this study, we will examine both the quantity and quality (as determined by whether an anonymous user is more likely to match with a more socially taboo match) of matches achieved with anonymity versus non-anonymity to determine whether the costs of either anonymity or non-anonymity outweigh the benefits.

Other studies examined the Gale-Shapley deferred-acceptance algorithm in online interactions. Hitsch et al. (2010) sought to establish whether economic matching models can explain observed online matching patterns, and measure the efficiency of a decentralized matching mechanism website (*5*). It was observed that the matching mechanism for the website returned approximately efficient results, as compared to the implementation of a deferred-acceptance Gale-Shapley algorithm. This study's purpose is to investigate anonymity's impact in online dating and implement the Gale-Shapley algorithm without explicit preferences as a method to increase the efficiency of online dating.

MATERIALS AND METHODS

Data set: A de-identified data set of online dating users was used. This data set included characteristics of 100,000 users collected from a popular online dating site, which will remain unidentified and be referred to as "monCherie.com" in this paper. The users in this data set were randomly divided into two groups of the same size: anonymous and non-anonymous, with 50,000 users in each group. The non-anonymous group was not gifted with anonymity, i.e. the ability to browse another user's profile without the other user seeing. The anonymous group was gifted with anonymity.

The following user demographics were recorded for each group: validity (whether the user is a robot or spammer, as determined by an in-site algorithm), activity (whether this user has been using the account recently), age, gender, ethnicity, race, sexual orientation, and attractiveness score. The use of validity and activity were imperative, as taking those variables into account prevented our data from being influenced by inactive users or robots, both of which are not representative of the general population. The variable attractiveness score served as a proxy for user attractiveness in our data, which was central in generating preferences later. The value of this variable was the average score that other users could secretly give the focal user, based on how attractive the other users found the focal user.

Table 1. Race and gender distribution in VAS data set (table by authors). This table displays the race and gender distribution of users present in the 24,664 user valid-active-straight (VAS) data set.

	Asian	Black	Indian	Latino	Middle Eastern	Native American	Other	Pacific	White	Total
Female	591	587	96	608	77	144	300	74	5,186	8,963
Male	747	812	240	1,061	170	253	558	116	8,879	15,701
Total	1,338	1,399	336	1,669	247	397	858	190	14,065	24,664

Table 2. Age group distribution in VAS data set (table by authors). This table displays the age group distribution of users present in the 24,664 user valid-active-straight (VAS) data set.

18–19	20–25	26–29	30–39	40–49	50–59	60+
1,477	6,732	6,484	5,972	2,530	1,082	387

This data set contains data recorded from each user's interactions with other users over a period of three months. In the first month, all users had the same privileges and were non-anonymous. During the second month, half of the users were given anonymity. For the third month, all users were once again non-anonymous. The following user interaction variables were recorded: number of unique profiles viewed, number of unique people who observed user's profile, number of messages sent, number of messages received, number of messages exchanged with one unique user, matches made when user initiated contact, and matches made when the other person initiated contact with the user.

The monCherie.com data set was refined to include only users who met three criteria: the user is valid, the user is active, and the user is straight. These criteria were designed to ensure that the users analyzed were not robots/spammers, were actively using the website, and would match with someone of the opposite gender (a prerequisite for the Gale Shapley portion of our research). The valid-active-straight (VAS) data set, consisting of 24,664 users, was then analyzed in RStudio using R Programming Language. The gender and age groups distributions in VAS (Tables 1, 2) are not unusual for online dating sites (*6*, *1*).

Statistics on anonymity: Exploratory statistical analyses were used to get a sense of the data set. Whether data distributions were normal or skewed was observed through frequency histograms, Q-Q plots, and simulations of normal data in RStudio. JMP Student Edition was used to perform t-tests and ANOVAs on subsets of size n ≤ 5,000.

Unpaired t-tests were used to compare the control group to the anonymous group for each variable to investigate any significant differences for both demographic or user interaction-related variables. Though distributions were not normal, the use of parametric statistical tests were still valid because the data sets contained more than 30 users. To create visual representations of the data, the statistical tool Tableau Public was used. The two groups (anonymous and non-anonymous) were split up into sub-groups based on gender, age, and race. Means, medians, and variances were calculated for each group and subgroup using JMP distributions (Table 3). These groups' variables were further analyzed using the same statistical tests we ran on the two large groups, including t-tests and regression algorithms ran in JMP Student Edition first. However, JMP Student Edition only had the capacity for 5,000 users. For groups and subgroups with more than 5,000 users, R programs were written and run to analyze significant differences in any match variables across anonymity.

Table 3. Gender and average match variables (table by authors). This table displays the match variables (the variables describing user match-behavior on the site, e.g. messages exchanged, views exchanged, and ultimately matches exchanged) for each gender. Both the total, T (top line of each), and the average, A (line below total), of each variable are displayed in the table below.

Gender	Match Sent Count	Match Received Count	Message Sent Count	Message Received Count	View Sent Count	View Received Count
Female (T/A) 8,963	4,940 0.511	15,529 1.733	15,694 1.751	105,583 11.780	171,892 19.178	692,147 77.223
Male (T/A) 15,701	14,462 0.921	3,646 0.232	106,594 6.789	10,891 0.694	650,849 41.453	121,324 7.727

Gale-Shapley algorithm: Code was written in R to perform regressions on the VAS group. The linear regressions were run on the VAS data set in RStudio and two formulae were found for each gender. One formula predicted a match's age and one predicted a match's attractiveness score (Table 4).

A program was coded in Java to assign preferences for each user, which is necessary for the Gale Shapley algorithm. This program accounted for each user's gender, age, attractiveness score, and race. A set of "ideal scores" for age and attractiveness were calculated based on the formulae found from the linear regressions run. Ideal gender was assumed to be the opposite of the user's gender as the users in this data set reported their sexuality as straight, and ideal race was the same race as the user based on assumptions of similarity. The Preferences program prioritized the ideal age and ideal attractiveness score variables in ordering the users in a preference list; however, given a comparison of two users with similar numbers, the program also took race into account (Figure 1). After the Preferences program was run, a CSV file was generated with each user and a complete list of the user's match preferences.

The Gale-Shapley algorithm was programmed in Java and ran using the Preferences CSV file as a parameter (Figure 2). The algorithm iterates through the list of users and returns an individual, stable match for each user based on each user's preferences list. Each user was matched with another user.

Table 4. Preferences formulae (table by authors). The formulae returning the "ideal" attractiveness score and age predicted for a match are displayed below for men and women. These formulae were found by running linear regressions in RStudio.

	Women	Men
Ideal attractiveness score	1.824729 + (0.27931 • actual attractiveness score) – (0.003326 • actual age)	2.5278803 + (0.3289965 • actual attractiveness score) – (0.0007682 • actual age)
Ideal age	4.2709 + (0.2459 • actual attractiveness score) + (0.8834 • actual age)	3.9611+ (0.2175 • actual attractiveness score) + (0.7878 • actual age)

RESULTS

Table 5 shows the results of unpaired t-tests, comparing the match sent and received counts across anonymity by age group. The impact of anonymity was far more significant in match received count than match sent count. T-test results were analyzed in JMP. Subgroups for race, gender, and age group, as well as more specific combinations of several divisions (e.g. twenty-year-old-men, or Latino-thirty-year-olds), were analyzed. All results for race are displayed in Table 6.

Table 5. Significance of anonymity on number of matches with respect to age group (table by authors). This table displays the findings of t-tests assuming unequal variance. Significance of mean match counts for different age groups, across anonymity are shown. Statistically significant variables ($p < 0.05$) are noted in the APA style.

Age group (years)	Mean match sent count (A/NA)	Mean match sent count p-value (significance level)	Mean match received count (A/NA)	Mean match received count p-value (significance level)
18–19	0.7489/0.7803	0.8009 (ns)	1.247/1.278	0.8732 (ns)
20–25	0.7309/0.7455	0.7994 (ns)	0.7661/1.000	0.0001039 (***)
26–29	0.8428/0.7838	0.3660 (ns)	0.6080/0.7626	0.0007622 (***)
30–39	0.8081/0.8279	0.8017 (ns)	0.6573/0.8030	0.007875 (**)
40–49	0.8561/0.8115	0.7701 (ns)	0.6285/0.7146	0.3264 (ns)
50–59	0.6200/0.7324	0.4213 (ns)	0.5652/0.7215	0.1618 (ns)
60+	0.4541/1.126	0.02435 (*)	0.2755/0.644	0.007244 (**)

Table 6. Significance of anonymity with respect to race (table by authors). This table displays the findings of t-tests assuming unequal variance. Significance of mean match counts for different race groups, across anonymity are shown. Statistically significant variables ($p < 0.05$) are noted in the APA style.

	Messages sent	Messages received	Views sent	Views received	Matches sent	Matches received
Asian	0.2079 (ns)	0.5744 (ns)	0.1061 (ns)	0.7354 (ns)	0.5571 (ns)	0.1216 (ns)
Black	0.0007 (***)	0.8215 (ns)	0.0002 (***)	0.6900 (ns)	0.0067 (**)	0.4934 (ns)
Indian	0.9940 (ns)	0.7001 (ns)	0.4367 (ns)	0.7864 (ns)	0.2443 (ns)	0.7209 (ns)
Latino	0.3933 (ns)	0.0900 (ns)	0.4904 (ns)	0.0489 (*)	0.9185 (ns)	0.0036 (**)
Middle Eastern	0.3740 (ns)	0.3872 (ns)	0.1293 (ns)	0.2330 (ns)	0.6726 (ns)	0.9626 (ns)
Native American	0.6860 (ns)	0.6426 (ns)	0.3911 (ns)	0.8791 (ns)	0.6912 (ns)	0.2020 (ns)
Other race	0.8785 (ns)	0.4479 (ns)	0.4284 (ns)	0.0602 (ns)	0.7674 (ns)	0.9042 (ns)
Pacific	0.0251 (*)	0.2344 (ns)	0.1229 (ns)	0.2475 (ns)	0.1048 (ns)	0.7807 (ns)
White	0.3974 (ns)	0.03923 (*)	0.0102 (*)	0.2033 (ns)	0.9454 (ns)	0.0000002649 (***)

For each subgroup, significant match variables (matches, views, and messages; sent or received) were recorded. Significant variables were found by t-tests run in either JMP or R, depending on whether the subgroup contained more than 5,000 users. Significance was set at $p < 0.05$. All statistically significant groups are displayed in Table 7.

Table 7. Statistically significant test groups across anonymity (table by authors). This table displays the variables found to be statistically significant (using t-tests, $p < 0.05$) in each test group in which a statistically significant variable exists.

Test group	Significant variable
VAS (entire set)	Matches received
Black	Messages sent Views sent Matches sent
Latino	Messages received Matches received
Pacific	Views sent
White	Matches received
Over 60	Matches received Matches sent
Women in teens	Views sent
Women in late 20s	Messages received Matches received
Women in 40s	Messages sent
Men in late 20s	*Every match variable*
Men in 30s	*Every match variable*
Men in 50s	Views sent
Men above sixty	Messages received Matches sent Matches received

DISCUSSION AND CONCLUSIONS

This study examined the impact of anonymity on match variables across demographic groups. No previous study has examined how anonymity affects specific demographic groups.

Despite conventional thinking, we found that anonymity does not have a significant positive impact on the number of matches. In fact, anonymity is shown to have a negative effect on several different demographic groups. In these demographic groups, matches actually *decreased* with the addition of anonymity.

The overall VAS data set showed a significant decrease in matches. Of the different groups by race, the black, latino, and white groups experienced a significant decrease in matches across anonymity. Additionally, users in their twenties, thirties, and above sixty experienced a significant decrease in matches across anonymity ($p < 0.05$). Some more specific groups, such as men in their late twenties and women in their late twenties, also experienced a significant decrease in matches ($p < 0.05$). One reason for this may be that anonymity decreases weak-signalling, a phenomenon in which a user is alerted to the interest of the focal user when they see that the focal user has viewed their profile.

Anonymity eliminates the weak signal of interest made when viewing a user's page; thus fewer matches are created, as the viewed user can no longer observe this action and interest. The significance of this has been shown to be more impactful for received matches than sent matches, as is shown in Table 5. In this table, the number of matches sent and received are shown for age groups both with and without anonymity. Anonymity significantly impacted the number of matches received for four of the seven groups, while only significantly impacting the match sent number of one group. This may be explained by the fact that received matches are initiated by messaging from the other user, rather than the focal user. When the focal user is anonymous, signals are not sent, reducing the likelihood of the other user sending a message, and creating a received match.

Another notable trend is the number of views sent. Several groups, including black and white, experienced a significant difference in number of views sent across anonymity. This may be addressed by the online disinhibition effect. This theory asserts that people will act with less discretion and restraint when they can separate their online actions from their offline lives. That dissociation allows people to disown their behavior by not requiring them to acknowledge it (Suler, 2004). If an anonymous user knows that their viewing practices cannot be observed, they are more likely to view people they would not otherwise view, thus increasing the total number of views sent.

Additionally, we successfully implemented the Gale-Shapley algorithm on a set of VAS users who we know had matched with another user online. The output of the Gale-Shapley algorithm provided a list of potential couples, creating a suggestion engine. This engine can be used to increase the efficiency of online dating by providing users with possible partners. That allows users to increase the number of matches, as any given user cannot look through all other users who match their desired characteristics. By increasing the number of matches, the probability of finding a long-term partnership may also increase, thus improving the outcome of online dating.

LIMITATIONS

One setback to this research was lack of sufficient technology to process large volumes of data. Statistical softwares such as JMP Student Edition, which can only hold a certain number of rows, were sometimes unable to hold the entire set. Thus, two different statistical programs — in JMP and RStudio — were used to perform statistical tests on the data. This may cause possible error as the statistical softwares may use slightly different algorithms to perform the same statistical tests. An additional limiting factor was the capacity of the computers to process very large data sets (billions of data points).

FUTURE WORK

The next steps will involve applying the analyses used in our research to other online dating sites that feature anonymity. This will allow the impact of anonymity to be more fully understood. Additionally, quantifying the correlation between achieved matches and the results of the Gale-Shapley algorithm will help determine the stability of online dating outcomes produced by the Gale-Shapley algorithm. Another area of future work is to investigate correlation between the contents of various messages that result in matches and combine those findings with the Gale-Shapley algorithm to make the match search process more accurate.

Additionally, given the preferences determined by the linear regressions conducted, a more specific formula based on other factors investigated via linear regression, such as user demographics, observed user behavior, and the users' stated preferences, may be added to online dating sites to better suggest matches for users.

This research has applications in any field that requires matching two groups. Our study examined anonymity more extensively than in any previous study, and results show that despite current marketing of anonymity as a premium feature, anonymity neither improved quality or quantity of matches in an online dating market. This research also involved developing an implementation of the Gale-Shapley matching algorithm in Java, and creating a preference algorithm. Using big data analysis techniques in R, a set of discrete preferences for each user can be created. From resident physicians and hospitals to college dormitory roommates, this technique can be applied to create stable pairings. This combination of observing and modeling general trends, and applying algorithms creates a powerful matching engine. The development of such an engine helps maximize the efficiency of the online dating market by providing users with more potential matches than they could find without assistance.

Figure 1

Preferences program pseudo code (figure by authors). This section of pseudo code displays a method that compares two users—users 1 and 2— to find which should be higher on the focal user's preferences list; i.e., which would be the better match for the focal user. This method is a part of the larger Preferences program that assigns each user a list of preferences and writes a CSV file with this information.

```
private User compareUsers(FocalUser, User1,  User2){
if((difference in age < 3 years) and (difference in
attract score < 1)){ if(User1 race = FocalUser race
and User2 race != FocalUser race){
                    return User1;
        }
else if((User2 race = FocalUser race and User1
race != FocalUser race){ return User2;
        }
}
if(User1 age and attract is closer than User2's to
Focal User's ideal scores){
               return User1;
          }
         else{
             return User2;
         }
}
```

Figure 2

Gale-Shapley algorithm pseudo code (figure by authors). This section of pseudo code displays the "matching" section of the Gale-Shapley algorithm, in which the program iterates through each user and finds matches until every user is matched.

```
public  void stableMatching(){
        while(not everyone is matched){ for(the next
        unmatched User  u){
        if(the next user v on u's list is not
        matched){ Match u and v;
        }
        else if (the next user v on u's list is
        already matched){
        if(u is a better match than the user that v
        is already matched with){ Undo v's current
        match;
        Match u and v;
                       }
               }
             }
         }
}
```

Figure 3

Female oriented Gale-Shapley output (figure by authors). This figure displays a sample of the list of pairs generated by the Gale-Shapley matching algorithm. The ID numbers correspond to users in the dataset, while the Male and Female columns show the arbitrary values we assigned to users (1–4625). To obtain the above table, the Male and Female columns were alternatingly sorted in ascending order, and the ID columns were imported from the input preference files.

	Male ID	Female ID	Female	Male
1	15153037061203273728	13325296975221372928	1	3863
2	15509529069832589312	10923024243019030528	2	112
3	10198722645928280064	16385243983350679552	3	902
4	14021935717173622784	4282847657055940608	4	187
5	16402657647941773312	6772964349186309120	5	3494
6	1496524942308025088	2157420545613496320	6	803
7	5046152145306383360	15149247685632915456	7	2010
8	2861721233560519680	863647449290792960	8	4213
9	3085374849888153600	15203789973946116096	9	70
10	1165680894543920896	10975772619063453696	10	3283
11	5412100349194756096	4975911604624098304	11	5
12	578276748255402112	8235858399768688640	12	3695
13	8773855434978464768	7214652321543727104	13	4236
14	14807391001863923712	16842969534017529856	14	38
15	16127647271730335744	16634704243699949568	15	35
16	11006071297587103744	10843165866588538880	16	13
17	1514459441141798912	8708793886347833344	17	46

Figure 4

Male oriented Gale-Shapley output (figure by authors). This figure displays a sample of the list of pairs generated by the Gale-Shapley matching algorithm. The ID numbers correspond to users in the dataset, while the Male and Female columns show the arbitrary values we assigned to users (1–4625). To obtain the above table, the Male and Female columns were alternatingly sorted in ascending order, and the ID columns were imported from the input preference files.

	Female ID	Male ID	Male	Female
1	7836923167268478976	4325361397502581248	1	3140
2	7709004595083494400	15476794657087627264	2	2938
3	1652241294539887360	6067347760800848896	3	775
4	6701967818028038144	14421022821277478912	4	310
5	4975911604624098304	5412100349194756096	5	11
6	416488717648925376	15323922805471891456	6	241
7	9856905134407237632	15091190949684985856	7	1267
8	12176471665600792576	3932630622077568512	8	54
9	17562293213922150400	15840834001204623360	9	3519
10	14235458645414453248	1485347107298391040	10	179
11	10537788366163218432	8579807113210845184	11	257
12	12418889353342955520	10372950844657530880	12	640
13	10843165866588538880	11006071297587103744	13	16
14	15589755977017094144	3681721830418909696	14	148
15	8559833892108968960	7780505105744662528	15	4553
16	9492348207707852800	1778909973900882432	16	3018
17	66135957067730544	13910493629082226688	17	231

ACKNOWLEDGEMENTS

This independent research was conducted and reported as a joint effort by both authors. The work was completed under the mentorship of Dr. Ravi Bapna (Curtis L. Carlson Chair in Business Analytics and Information Systems; Academic Director, Carlson Analytics Lab; Program Director of MS-Business Analytics in Information & Decision Sciences) at the University of Minnesota Carlson School of Management and Ms. Princesa VanBuren Hansen, our teacher. We appreciate the assistance Dr. Bapna provided by helping us obtain permission to study the popular online dating-site data set. Ms. Princesa VanBuren Hansen advised us on our paper-writing process.

REFERENCES CITED

1. Bapna R, Ramprasad J, Shmueli G, Umyarov A. One-Way Mirrors in Online Dating: Randomized Field Experiment. *Management Science*. 2016: 1-37.

2. Gelles D. Inside Match.com. http://www.ft.com/cms/s/2/f31cae04-b8ca-11e0-8206-00144feabdc0.html. Published July 29, 2011, accessed May 25, 2016.

3. Gale D, Shapley LS. College Admissions and the Stability of Marriage. The American Mathematical Monthly. 1962; 69(1): 9–15.

4. Suler J. The Online Disinhibition Effect. CyberPsychology & Behavior. 2004; 7(3): 321–325.

5. Hitsch GJ, Hortaçsu A, Ariely D. Matching and Sorting in Online Dating. *The American Economic Review.* 2010; 100(1): 130-163.

6. Fiore AT, Taylor LS, Zhong X, Mendelsohn G.A., Cheshire C. Who's Right and Who Writes: People, Profiles, Contacts, and Replies in Online Dating. *HICSS.* 2010: 1-10.

OTHER REFERENCES

Augur H. Can Big Data Save Your Love Life? Online Dating Apps Say "Yes." http://dataconomy.com/can-big-data-save-your-love-life/. Published February 9, 2016, accessed June 29, 2016.

Finkel EJ, Eastwick PW, Karney BR, Harry TR, Sprecher S. Online Dating: A Critical Analysis From the Perspective of Psychological Science. Association for Psychological Science. 2012; 13(1): 3–66.

Paumgartner N. Looking for Someone: Love, sex, and loneliness in the time of the Internet. http://www.newyorker.com/magazine/2011/07/04/looking-for-someone. Published July 4, 2011, accessed May 25, 2016.

Reis HT, Aron A, Clark MS, Finkel EJ. Ellen Berscheid, Elaine Hatfield, and the Emergence of Relationship Science. Perspectives on Psychological Science. 2013; 8(5): 558–572.

Smith L The Marriage Model with Search Frictions. Journal of Political Economy. 2006; 114(6):1124–1144.

Tilly J. Algorithms for Matching Markets in R and C++. https://github.com/jtilly/matchingR. Published January 5, 2016, accessed August 18, 2016.