Roommates Problem with Correlated Preferences

Onur Burak Celik
Assistant Professor
Quinnipiac University, Department of Economics
Hamden, CT, 06518, USA
Email: oburakc@hotmail.com, Phone: (860) 830-6669

Abstract

All of the work done on the roommates problem, a one-sided matching market, is either studying the existence of a stable matching or searching for an efficient algorithm. This paper departs from the main stream of this literature and analyzes via simulations the effect of correlation in the preference lists on the aggregate satisfaction of the participants in roommates problem. The technique introduced by Celik and Knoblauch (2007) is used to create correlated preference lists. For the analysis, a measure is also defined to quantify the level of correlation. Results showed that correlation is an important factor on the aggregate satisfaction of the individuals. A higher correlation level among the preference lists leads to less satisfied participants.

Keywords: *One-sided Matching, Roommates Problem, Correlated Preferences, Gale-Shapley Algorithm.*

1. Introduction

The roommates problem, a one-sided matching model, was first introduced by Gale and Shapley (1962) as a generalization of the marriage problem, a two-sided matching model. Two-sided matching models have attracted more interest than one-sided models. Most of the matching theory research is devoted to two-sided matching models especially to the marriage model, the college admissions problem and the labor market for medical graduates[1]. In the roommates problem literature, since a stable matching does not need to exist, researchers are mostly focused either on the existence of a stable matching or on finding an efficient algorithm that produces one[2]. There is still a need for a better understanding on the implications of a stable outcome in a one-sided market.Preferences are the core elements in matching models and the assumption of random preferences is not very realistic.

This assumption might be misleading in the studies of the models and the results that depend on individuals' preferences. In general, there exists some common belief on the items to be preferred which creates correlation at some degree when a group of people make preferences over colleges or hospitals to which they are applying for admission, spouses to get married with, partners to work with or mates to share a room with. As more individuals are involved in these kinds of markets, the probability of having random preferences, where the correlation is measured to be zero, gets smaller. For this reason, the existence of the correlation in the preference lists cannot be ignored.Caldarelli and Capocci (2001) are the first researchers who recognized the importance of correlated preference lists in the matching markets. They introduced a formulation to create correlated preferences and studied its effects, via simulations, on the average rank of the matched partners in the marriage matching market. However their work was very limited. Their correlation measure cannot be applied to any preference list to measure the level of the correlation and they do not provide any statistical result so the significance of the effect of the correlation is not clear.

[1] Knuth (1976), McVitie and Wilson (1971), Roth (1984), Roth and Peranson (1999), Roth and Sotomayor (1990) are some of the most important and most cited studies on two-sided matching models.
[2] Tan (1991) and Chung (2000) are two good papers on roommates problem providing necessary and sufficient conditions on the existence of a stable matching.

An answer to this question is given by Celik and Knoblauch (2007). They develop a general method to study the effect of correlated preference lists on the aggregate satisfaction with the men-propose Gale and Shapley algorithm. Their correlation measure can be applied to any given preference profile. Therefore their methodology is more powerful than Caldarelli and Capocci's and has wide range of application in the field of two-sided matching. Moreover they provide significance tests along with the analysis of simulation data. A recent research on correlated preferences contributed to the matching literature by studying via simulations the effect of inter-correlated preferences, where men rank highly those women who rank them highly, on the aggregate satisfaction of men and women in a marriage matching model. Boudreau and Knoblauch (2010) used an extended version of Caldarelli and Capocci's correlation measure. Their findings also showed that the correlation in the preference lists is an important and significant factor on aggregate satisfaction of the participants. Boudreau and Knoblauch (2010) also emphasize the importance of the use of simulations in the models where theoretical progress is slow and difficult. In such cases, experiments and simulations are always good guides for the theoretical work. Results of this paper also rely on computer simulations to present some facts and provide motivation for further theoretical work.After all these studies, it is still a question mark how the correlation in the preference lists of the individuals will affect the outcome in a one-sided matching market. This paper studies via simulations the effect of the correlation of preferences on the aggregate satisfaction of the roommates from a matching. In this paper a similar methodology to that of Celik and Knoblauch (2007) is followed but the study here focuses on one-sided matching markets, namely roommates problem.

First, a correlation measure that can be used for any one-sided market is defined in order to quantify the correlation among preference lists of individuals. Secondly, computer simulations for a market of 100 participants are run using an extended version of the Gale-Shapley algorithm. And finally, for each run aggregate satisfaction of the participants is calculated where individual satisfaction is measured by the position of a person's matched mate in his preference list, and the aggregate satisfaction is the sum of all individual satisfaction levels. Notice that a higher value for the aggregate satisfaction measure means less satisfied individuals. Our results showed that the correlation is an important and significant determinant on the aggregate satisfaction of the roommates where a higher degree of correlation among the preference lists leads to a higher aggregate satisfaction measure, equivalently less satisfied roommates. This study also provides an estimation of a cubic equation that calculates the expected aggregate satisfaction level when the correlation level is given. The algorithm used here picks the stable matching with the lowest aggregate satisfaction measure from the set of all stable matchings. Therefore, the estimated curve gives the lowest possible aggregate satisfaction measure for any correlation level. All other stable matchings will result in a higher aggregate satisfaction measure, that is less happy participants.The rest of the essay is organized as follows. Section 2 explains the roommates problem, the algorithm used and the minimum regret stable matching, section 3 introduces the technique to generate correlation and the formulation to measure the correlation, section 4 presents the results, and section 5 concludes the essay.

2. Roommates Problem, Algorithm and the Minimum Regret Stable Matching

Gale and Shapley (1962) introduced the roommates problem as a general case of a marriage problem. In the same paper, with a simple proof, they also showed that contrary to the marriage problem, a stable matching does not need to exist in a roommates problem. The roommates problem is simply forming unordered pairs of n people coming from the same set to make partners or roommates according to their preferences where the preferences are strict, complete and transitive. A matched pair can be denoted by $\{i,j\}$ or equivalently $\{j,i\}$. In the roommates problem, the definition of "stable matching" is the same as in the marriage model. A matching μ is *stable* if there are no two persons where each prefers the other to his partner in the matching μ. If a roommates problem has at least one stable matching, then it is said to be *solvable*, otherwise *unsolvable*. Since this paper focuses on the aggregate satisfaction of roommates when preference lists are correlated, the solvable roommates problems are taken into consideration and unsolvable ones are ignored. Also if there exists more than one stable matching, the algorithm picks the one that gives the lowest aggregate satisfaction measure.

Gusfield and Irving (1989) give a detailed explanation of the algorithm used in this essay which is an extended version of the Gale and Shapley algorithm with two phases. First phase is similar to the algorithm defined by Gale and Shapley (1962) for the college admissions problem. The first phase of the algorithm starts with each person proposing to his highest ranked mate in his preference list. If person *i* receives a proposal from person *j*, in any stable matching *i* will not be matched by someone he gives a rank lower than *j* and in the worst case person *j* and person *i* will be matched. Therefore, there is no harm in deleting anybody ranked lower than *j* in *i*'s preference list. In order to have consistency among preference lists, whenever a person is deleted from *i*'s list, *i* should also be deleted from that person's list. At this point *j* is said to be semi-engaged to *i*. It is one-sided engagement since *i* can still be free or semi-engaged to someone else he proposed to who has not rejected him yet. During this phase, if *i* receives another proposal from, let's say, person *k* ranked higher than *j* in his preference list, once again everybody below *k*, including *j,* will be removed from *i*'s list resulting in *i* being removed from the list of everybody below *k,* including *j*. Then, *j* will become free and the algorithm continues with *j* making a proposal to the next person on his list that he has not made a proposal yet. At the end of the first phase, when proposals and deletions are completed, each person will be semi-engaged to the first person in their reduced preference lists, and the last person in their list will be semi-engaged to them.

During the first phase of the algorithm, if any of the preference lists becomes empty then there exists no stable matching. For example, with preferences below

A	**B**	**C**	**D**
B	C	A	A
C	A	B	B
D	D	D	C

after the first phase, A is semi-engaged to B, B to C and C to A, and D's reduced preference list is empty.

A	**B**	**C**	**D**
$\boxed{B}$	$\boxed{C}$	$\boxed{A}$	~~A~~
C	A	B	~~B~~
~~D~~	~~D~~	~~D~~	~~C~~

If all preference lists are reduced to a single entry then a stable matching is achieved and it is unique. Otherwise, first phase is completed when everybody has made a proposal and is not free. Then, we continue with the second phase of the algorithm.

The second phase is the successive deletion of rotations embedded in the reduced preference profile P remaining after the first phase.

Let $f_P(i)$ and $s_P(i)$ be the first and second choices in *i*'s preference list for any remaining preference profile P in hand after the first phase and/or the deletion of previous rotations. A rotation is a list of pairs in P, $\{i_0, j_0\}$, $\{i_1, j_1\}$… $\{i_{r-1}, j_{r-1}\}$, such that for all t ($0 \leq t \leq r-1$), $j_t = f_P(i_t)$ and $j_{t+1} = s_P(i_t)$, where t+1 is taken modulo r.

For example, with preferences

A	**B**	**C**	**D**
B	C	D	A
C	D	A	B
D	A	B	C

after the first phase we have A semi-engaged to B, B to C, C to D and D to A. Reduced preferences are the same as the original preferences. The rotation is {A, B}, {B,C}, {C,D}, {D,A}. To remove the rotation, cross B off A's list. Then A must be crossed off B's list. After A proposes to C, B and C are crossed off each other's list. After B proposes to D, C and D are crossed off each other's list. After C proposes to A, A and D are crossed off each other's list.

Preferences are now A:B B:D C:A D:B, and the matching is complete.

A	**B**	**C**	**D**
~~B~~	~~C~~	~~D~~	~~A~~
$\boxed{C}$	$\boxed{D}$	$\boxed{A}$	$\boxed{B}$
~~D~~	~~A~~	~~B~~	~~C~~

In the second phase there might be more than one rotation. Depending on which rotation is picked to be deleted, each different path followed in the second phase will produce a different stable matching.

Algorithm used in this essay will produce a "*minimum regret*" stable matching which tries to make the least happy person as happy as possible, as described in Gusfield and Irving (1989). The regret of a person is defined as the position of his matched mate in his original preference list. Similarly, at any point of phase two, the regret of a person is measured as the original position of the last entry of his reduced preference list. In the minimum regret stable matching, regret is as small as possible.When more than one rotation exists, the algorithm picks the rotation that requires the deletion of the last entry of the preference list of the person with the maximum regret. Therefore, the least happy person is happier now. The algorithm stops when a preference list becomes empty, which means there is no stable matching, or all lists have single entry, which is the stable matching. The stable matching found here is the one with minimum regret, equivalently minimum aggregate satisfaction measure.

3. Generating and Measuring Correlated Preferences

Students who will be paired to share a room are grouped into disjoint sets according to their popularity. For example, for n=100, grouping (10, 30, 60) means that the first 10 students are very popular so everybody wants be a roommate with any of them. Therefore these 10 students will be ranked randomly in the top 10 in all preference lists. Then everybody will rank the next 30 students randomly between 11^{th} and 40^{th} places since they are less popular. And finally, the least popular 60 will be placed randomly between the 41^{st} and 100^{th} ranks. Since a student cannot place himself in his own preference list, after generating all preference lists as described above, each student will be removed from his own preference list. Hence each list will have n-1 entries. Using this technique, with different groupings, different levels of correlation can be achieved for any one-sided matching market.

In order to quantify the correlation among preference lists, we define a measure. Our measure is,

$$\rho = \frac{\sum_{i=1}^{n}\left[\mathrm{Ave}(i) - \frac{n}{2}\right]^2}{\sum_{i=1}^{n}\left[\frac{(n-2)i+1}{n-1} - \frac{n}{2}\right]^2}, \qquad (1)$$

where *Ave (i)* is the average ranking of student *i* in all the others' preference lists.
Therefore,

$$Ave(i) = \frac{1}{n-1}\sum_{j=1,\, j\neq i}^{n} r_j(i), \qquad (2)$$

where $r_j(i)$ is the ranking of student *i* in student *j*'s preference list.

Notice that a student cannot place himself in his own preference list but appears in all the other lists. Therefore, in contrast to two-sided matching markets, perfect correlation does not refer to having exactly the same preference lists for all students in roommates problem.

For example, in a one-sided market, perfectly correlated preference lists will look like the one below.

1	**2**	**3**	**4**	**5**	**6**	….	**N**
2	1	1	1	1	1	….	1
3	3	2	2	2	2	….	2
4	4	4	3	3	3	….	3
5	5	5	5	4	4	….	4
6	6	6	6	6	5	….	5
.	.	.	.	.	.	….	.
.	.	.	.	.	.	….	.
.	.	.	.	.	.	….	.
n	n	n	n	n	n	….	n-1

For perfectly correlated preferences, from (2), $\mathrm{Ave}(i) = \frac{(n-2)i+1}{n-1}$. When preferences are uncorrelated, each student will appear only once at each ranking from 1st to (n-1)th places in a preference profile consisting of n preference lists as in the example below. Then $\mathrm{Ave}(i) = \frac{n}{2}$.

1	**2**	**3**	**4**	**5**	**6**	….	**N**
2	3	4	5	6	7	….	1
3	4	5	6	7	8	….	2
4	5	6	7	8	9	….	3
5	6	7	8	9	10	….	4
6	7	8	9	10	11	….	5
.	.	.	.	.	.	….	.
.	.	.	.	.	.	….	.
.	.	.	.	.	.	….	.
n-1	n	1	2	3	4	….	n-2
n	1	2	3	4	5	….	n-1

In our correlation measure, the numerator calculates the deviation of the preference profile from being completely uncorrelated. If we have an uncorrelated preference profile, the numerator in (1) will be zero. Hence we will get $\rho = 0$. If the preference profile is perfectly correlated, then the numerator will be equal to the denominator in (1) resulting in $\rho = 1$.

4. Results:

In this section, we present our analysis for the market size n=100. For each grouping shown in Table 1, we ran 200 simulations.[3] In each simulation, the correlation level and the aggregate satisfaction are calculated. The average values of the correlation and the aggregate satisfaction for each grouping are given in Table 1. It is clear that higher correlation level among the preference lists leads to less satisfied individuals.We also estimate a cubic equation, Aggregate Satisfaction = $c(0) + c(1)*\rho + c(2)*\rho^2 + c(3)*\rho^3$. The estimated coefficients and the significance levels are shown in Table 2. Using this estimation, without spending time on running the algorithm for very large markets one can easily calculate the overall satisfaction level for the minimum regret stable matching once the correlation level is known. The value estimated by this equation is the average of the maximum satisfaction level that can be achieved for a given correlation level.With R^2 being almost equal to 1, Figure 1 shows how well the estimated cubic equation fits the data points. By setting the second derivative equal to zero, we find that $\rho^*=0.38066$ is the inflection point of the estimated equation where it changes from being a concave function to a convex function. This tells us that as the correlation level rises over ρ^*, the increase in the satisfaction measure is higher.

[3] Simulation results for different market sizes are also consistent with and support the results for n=100. But here, we present the results only for n=100.

5. Conclusion

The roommates problem is the least studied matching model in the literature. Researchers are focused mostly on the stability problem of the model. Using computer simulations, this paper explores the effect of correlation among preference lists on the satisfaction level of the participants. Our analysis showed that higher correlation in the preference lists result in less satisfied individuals. High correlation in the preference lists leads to competition for the few popular roommates. Therefore a few people will be very happy for being matched with those most desired roommates but the majority will be matched with roommates who are ranked much lower in their lists. Thus, overall satisfaction with correlated preferences is lower than the one with random preferences. As the correlation among the preference lists increases, the satisfaction and the happiness of the participants decreases. An estimation of a cubic equation is also provided which helps to find out the satisfaction level of the roommates once the correlation level is known. Correlation might be at different levels in different matching markets but this does not lessen the importance of the correlation in the preference lists. Previous studies on the matching outcome with correlated preferences mentioned in this study and this work also show the importance of the correlation. Thus, it should be on the top of the check list of market designers.

References

Boudreau, J.W., Knoblauch, V. (2010). Marriage Matching and Intercorrelation of Preferences, Journal of Public Economic Theory,12/3, 587-602.

Caldarelli, G., Capocci, A. (2001). Beauty and distance in the stable marriage problem, Physica, A 300, 325-331.

Celik, O. B., Knoblauch, V. (2007). Marriage Matching with Correlated Preferences, Working Paper Series, University of Connecticut, 16.

Chung, K. (2000). On the existence of stable roommate matchings, Games and Economic Behavior, 33, 206-230.

Gale, D., Shapley, L.S. (1962). College admissions and the stability of marriage, American Mathematical Monthly, 69, 9-15.

Gusfield, D., Irving, R.W. (1989). The stable marriage problem: structure and algorithms, The MIT Press, Cambridge, 1989.

Knuth, Donald E. (1976). Mariages Stables et leurs relations avec d' autres problémes combinatoires, Les Presses de l' Université de Montréal, Montréal.

McVitie, D., Wilson, L.B. (1971). The stable marriage problem, Communications of the A.C.M., 14, 486-490.

Roth A.E. (1984). The evolution of the labor market for medical interns and residents: a case study in game theory, Journal of Political Economy, 92, 991-1016.

Roth, A.E., Peranson, E. (1999). The redesign of the matching market for American physicians: some engineering aspects of economic design, The American Economic Review, 89, 748-780.

Roth, A. E., and Sotomayor, M. (1990). Two-sided matching: A study in game-theoretic modeling and analysis, Cambridge University Press, Cambridge.

Tan, J. J. M. (1991). A necessary and sufficient condition for the existence of a complete stable matching, Journal of Algorithms, 12, 154-178.

Figure 1: Aggregate satisfaction vs. Correlation

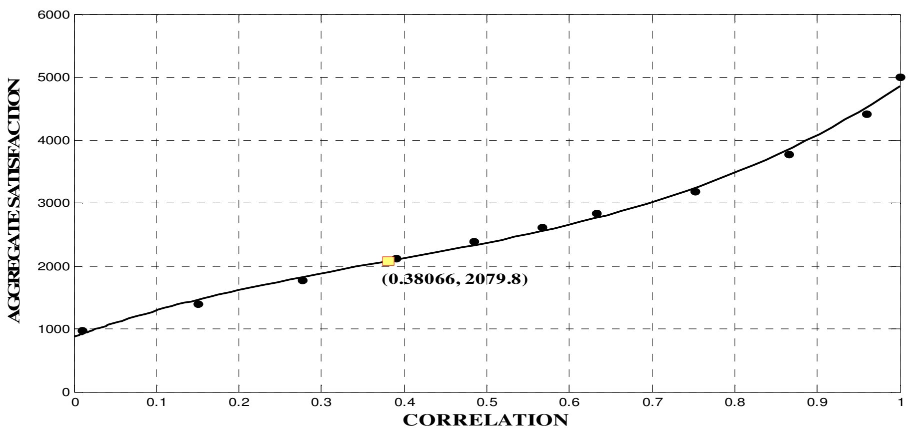

Table 1: Average values of the correlation and the aggregate satisfaction from 200 simulations for each grouping used in the analysis for n=100.

GROUPING	CORRELATION	SATISFACTION
Random	0.0101	976.74
5, 95	0.1511	1389.97
10, 90	0.2773	1769.79
5, 10, 85	0.3910	2122.91
20, 80	0.4852	2386.92
25, 75	0.5670	2611.67
30, 70	0.6337	2830.47
50, 50	0.7526	3188.44
20, 40, 40	0.8654	3770.98
20, 20, 20, 20, 20	0.9605	4419.50
Perfect Correlation	1	5000

Table 2: Regression results for Aggregate Satisfaction=c(0)+c(1)*ρ+c(2)*ρ²+c(3)*ρ³ where n=100.

Variable	Coefficient
Constant	874.4429*** (97.355)
ρ	4766.298*** (856.589)
ρ^2	-6304.610** (2021.3)
ρ^3	5520.744*** (1301.646)
R^2	0.9958

* significant at 10%
** significant at 5%
*** significant at 1%