# MATCHING IN CLUSTERS: ONLINE DATING AND INTERRACIAL MARRIAGES

OCTAVIA FEDERICA GHELFI

Abstract. Has online dating increased social integration? The aim of this paper is to analyze the interdependent relationship between matching and the network in which the matching takes place. In a highly clustered network, the expected probability of a match between agents belonging to two different clusters is low. On the other hand, a high number of matches between agents belonging to different clusters will contribute to the integration of the clusters in the following period. I develop a model of meeting and matching that sheds light on the patterns of ethnic homophily observed in the marriage market. The model is a two-stage game: in the first stage, agents engage in a population game to strategically increase their set of acquaintances, with the objective of maximizing their expected indirect utility in the second stage. In the second stage, the actual matching occurs, with the restriction that agents can only match with individuals they are connected to in the network. The model is then used to investigate how changes in the meeting technology, such as the introduction of online dating, affect matching frequencies and couples' assortativeness. In particular, the model is used to explain three empirical patterns of romantic relationships: (i) all ethnic groups are biased toward same-ethnicity partners, (ii) couples who meet online are more likely to be in an interracial relationship than those who meet offline, (iii) minorities who meet their partner online are significantly more likely to be in a relationship with a white person, but equally likely to be in a relationship with a member of another minority group other than their own. The model is estimated and bounds are found for the network effect on matching probabilities. The estimates show that online dating has increased integration between minorities and white people, but has not increased exposure among minorities. Moreover, its effect has not been uniform across the United States. Finally, the model estimates are used to form a prediction on the evolution of the clustering of the network over time with and without online dating.

1

*Date*: November 16th, 2020.

[1]I deeply thank Alfred Galichon, Alessandro Lizzeri, Alberto Bisin, and Erik Madsen for their invaluable support throughout this project. I also owe special thanks to Konrad Menzel, Daniel Waldinger, Milena Almagro, Fabio Maccheroni for their comments and suggestions that have significantly improved my paper. This paper has also benefited from comments by seminar participants at SAET, SEA, Columbia University, Kansas University, and NYU's Economic Theory and Applied Microeconomics seminars. Any comments and suggestions are welcome. Financial support from NSF grant DMS-1716489 is gratefully acknowledged.

## 1. Introduction

Traditionally matching is studied as an isolated phenomenon, abstracting from the context in which it takes place. This work is an attempt at embedding what we currently know around matching systems into the context of a network, by relaxing the assumption of completeness of the network and allowing for an uneven exposure to different types of agents.

Assortativeness is a well-known phenomenon both in the matching and network literaure, and it has been extensively studied across fields in economics, sociology and psychology. It refers to the tendency to associate with people who are similar to oneself along some dimension, for instance race, ethnicity or education. Several papers, such as McPherson, Smith-Lovin, and Cook (2001); Siow (2010); Hitsch, Hortaçsu, and Ariely (2010a); Greenwood, Guner, Kocharkov, and Santos (2014), provide evidence of this tendency both in friendships and marriages. Sorting patterns can however arise for two different reasons. First, assortativeness can be a consequence of the structure of the network, which affects the composition of acquaintances of a person by skewing it towards same-type individuals. On the other hand, even in a perfectly integrated social network, sorting can arise due to the individuals' preferences to associate with similar people. The relative contribution of these two forces on sorting patterns remains unclear.

In this paper I investigate the effects of the two forces that act on assortativeness in one-to-one matching systems: the composition of the choice set and individual preferences over the elements of such set. I start from the observation that our social networks present a high degree of clustering around race, ethnicity and education. Recent surveys report that 92% of the acquaintances of a white person in the United States are also white, and that about 75% of white Americans do not have links with members of minority groups in their core social network[2]. Assuming that the set of acquaintances of a person looking for a partner coincides with their choice set, it is easy to imagine how an uneven racial and educational composition of acquaintances is reflected on their matching probabilities, making them relatively more likely to marry someone with the same socio-economic background. From a theoretical standpoint, the existing literature on discrete choice (Luce (1959) and Dagsvik (2000)) informs us that the conditional choice probabilities of agents facing different choice sets are different, even when the systematic utilities of the two agents are identical. The degree of exposure that a person has towards people who are different to them has a major impact on their probability of marrying someone that is different to them. However, the composition of the choice set is not the only driving force of assortativeness: agents might have a preference to match with someone similar to them, in which case sorting arises from characteristics of preferences rather than characteristics of the choice set. In this paper, I propose a model of matching that occurs in an incomplete and clustered network in order to take into account both the effect of the network and the effect of preferences. In particular, I use the model to investigate how online dating has affected the relative contributions of the network structure and of preferences on matching probabilities.

[2] Source: American National Social Network Survey, August 2020.

The introduction of online dating in 1995 and its widespread adoption have changed the dynamics of meeting and dating. On one side, online dating makes it possible to reduce barriers to meeting distant people in the network. This would lead us to think that the meeting probabilities for people of different races, ethnicities and educational levels is higher on an online dating platform than it is offline. This characteristic of online dating has indeed the potential to attenuate the effects of a highly clustered network on matching probabilities. However, on the other side, by reducing search frictions, online dating allows its users to direct their search only towards potential partners that maximize their preferences. This second aspect makes the contribution of preferences on matching probabilities much more significant online than it is in an exogenous social network. Moreover, if individual preferences are biased toward same-type individuals, the effects of directed search can counteract the potential integration effect.

In recent years, there has been a growing debate around how online dating has impacted the diversity of the set of acquaintances of individuals who use it, and as a consequence their probability of entering an interracial marriage. In particular, attention has been devoted to studying the social integration effect that online dating can provide. Ortega and Hergovich (2017) observe that online dating has created a dramatic change in the structure of the network we live in by allowing the possibility for connections to be made between otherwise distant clusters. Using a simulation based model utilizing real-world data about racial preferences, they show that in a world where people are highly connected with others of their own race, but only poorly so with people from other races, even random links to perfect strangers will quickly increase the percentage of interracial marriages. Studies on neighborhood segregation such as Ananat (2011) explain how the presence of physical barriers, such as railroads and rivers, are major causes of segregation in both urban and rural areas. Geographically based matching services that present users to potential partners within a certain mile radius, have the potential of overcoming some of those barriers.

While it is true that online dating has contributed to the expansion and diversity of the pool of acquaintances, it is also true that the effect of directed search on the platform might be equally significant. Looking at a sample of American couples who have been surveyed in How Couples Meet and Stay Together (HCMST) in 2009 and in 2017, I observe three patterns. Firstly, the majority of individuals across races are in a relationship with someone of their own race. Secondly, couples who meet online are more likely to be in a relationship with a partner of a different race. Thirdly, the fraction of mixed couples with one of the partners being white and the other belonging to a minority group is significantly higher among couples who have met online than among those who have met offline. However, the fraction of couples of individuals belonging to two different minority groups is roughly the same in the online and offline dating samples. Excluding same ethnicity marriages, minorities form romantic relationships with white people at a rate that exceeds the relative fractions of the population, while they form romantic relationships with other minority groups at a much lower rate than the relative fractions of the population. These summary statistics seem to suggest that online dating has favored the formation of certain types of inter-ethnic relationships while disadvantaging or not affecting others, therefore pointing at the impact of directed search.

In Section 3, I present suggestive evidence of various stylized facts that justify the theoretical model, including the lower ethnic assortativeness of couples that met online. Descriptively, the data presents strong assortativeness along ethnic lines that has however steadily decreased since 1967, when in *Loving v. Virginia* the U.S. Supreme Court ruled that marriage across racial lines was legal throughout the country.

In order to shed light on the observed matching patterns, I develop a model of meeting and matching that allows us to investigate how changes in the meeting technology - and therefore the meeting frequencies - affect matching frequencies and assortativeness. The model is compatible with the patterns observed in the data. In this paper, I try to point at the two contrasting effects that online dating creates: on one side, an integration effect, as pointed out by Ortega and Hergovich (2017), given by the creation of new ties between agents belonging to distant clusters. On the other side, a segregation effect, due to the directed search users put in place when looking for a partner that they are interested in. Online dating alters the probabilities of meeting someone of one's own ethnic group versus a different ethnic group, but the overall effect is ambiguous due to the simultaneous presence of these two contrasting phenomena, an integration effect introduced by the platform and the directed search put in place by its users. I show how the overall effect that online dating has on the probability of two agents belonging to different clusters meeting depends jointly on agents' preferences and the current degree of clustering of the network. In certain cases I am able to predict the overall effect of online dating on the network, in other cases the effect is ambiguous and needs to be assessed on a case by case basis.

In order to understand the relationship between online dating and the societal network, in Sections 4 and 5 I model online dating as a population game. The model consists of a two stage game: in the first period, agents decide whether to download an online dating platform. If they do, they engage in the dating game with the objective of strategically expanding their set of acquaintances. The goal is to maximize their expected utility in the second period, when the actual matching occurs. The matching phase is described through a non transferable utility model à la Gale and Shapley. Agents do not know everyone in the economy and are therefore constrained to propose to or accept proposals from people that they are connected to in the network. Two people in the network can be connected either because they were originally connected in the exogenous network or because they met through the population game in the first period. The equilibrium of the population game consists of an optimal decision to use online dating or not, and a subsequent action profile describing the effort made by each agent to meet potential partners of the same or different types. The population game allows us to determine which factors contribute to directed search towards specific subgroups of the population. Some of the incentives to direct search include a low cost of crossing a racial boundary, a low expected cost of crossing an educational boundary, the size of the subgroup being large, and reciprocity (that is a correct expectation that the subgroup the agent is directing attention to will reciprocate their interest). Moreover, it is shown that depending on exogenous parameters, online dating can result either in an increased or decreased diversity of the set of acquaintances of its users. Given that individual decisions to use an online dating platform arise as an equilibrium object, the distribution of types on the online dating platform is described as part of the equilibrium. Section 4 describes the model and the equilibrium in general form, and establishes results of existence and uniqueness of an equilibrium. Section 5 applies the model to the case of interracial and inter-educational marriages, assuming both network clustering and same-type bias in preferences.

I study the matching phase of the game assuming that the matching outcome is a pairwise stable matching with non transferable utilities (NTU). Each agent has a strict preference ordering over individuals of the opposite side of the market, and associates a random utility to each potential partner. The utility is composed of a systematic part, which we will assume only depends on the observables of the agent, and a taste shifter in the form of an additive idiosyncratic shock to the systematic utility. Using matching theory, I am able to predict matching frequencies in the matching game. In particular it is possible to compare matching probabilities for agents that have used online dating with those of agents that have not. Clearly, the use of online dating will have an impact on matching probabilities. In particular we will be able to estimate the expected number of interracial marriages among agents that have used an online dating platform and agents who have not. Using such estimates, I analyze the effect that matching patterns have on the shape and clustering of the network. Indeed, the creation of newly formed connections between otherwise distant clusters will indirectly affect the matching in the following period by making the two clusters more connected. Thus, the close interdependence of networks and matching is studied in Section 8. I apply this logic to the marriage market, focusing especially on interracial marriages, and hingeing on the idea that interracial marriages are both the cause and the consequence of social integration.

In Section 6, adapting Menzel (2015) identification strategy to the model and following a similar logic to Jaffe and Weber (2018), I show that failing to take into account meeting probabilities into matching probabilities in an incomplete and clustered network leads to a biased estimation of couples' surpluses. The magnitude of the estimation bias is directly related to the sparseness of the network, and the exact value of the bias is derived. Finally, I propose an identification strategy that, if used on appropriate data, allows us to disentangle preferences from the network effect.

In Section 7 the model is estimated using data from How Couples Meet and How they Stay Together (HCMST) collected in 2009 and 2017. The data cover relationship information of a representative sample of the American population. The information contained in the database includes observables about the couple such as ethnicity, education and income, as well as information about when and how they met. I estimate couples' surpluses firstly based on their ethnicity and gender and subsequently also including education and in what region of the United States they are located. I adjust the estimates to take into account the network effects in the different regions of the country. The surplus estimates are then used to assess the effect of online dating on assortativeness, both educational and racial, in different parts of the country. Estimates show a statistically significant cost associated with crossing both an ethnic boundary and an educational boundary in a romantic relationship. Furthermore in some regions of the country, namely the Northeast and West, such "costs" appears to be lower for couples who met online, in particular if one of the spouses is white. This suggests that online dating has had an effect in diminishing racial and educational assortativeness in those areas of the country. However, no significant effect is detected in the Midwest and the South, implying that the effect of directed search there has trumped the integration effect of online dating in those areas. Since it is not possible to disentangle what part of the effect of online dating on mixed couples is due to a higher meeting frequency and what part is due to a lower bias, I derive bounds on the increase in exposure to different races that has been caused by online dating. According to my estimates, the probability for a white person of knowing a minority group member online is 11.6%, compared to the 8% offline.

Finally, in Section 8 I discuss the evolution of the network. A high number of matches between different clusters contributes to the network becoming more connected in the following period. I assume that after a match is created, the core social networks of the two people in the couple form new connections among themselves. I show how interracial marriages contribute to creating diversity in the set of acquaintances of the people connected to the couple, whereas same-race marriages exacerbate the clustering of the network. I derive the steady states of the evolution of the network, and I show that complete integration can be achieved even in the presence of preferences that are biased toward same-type individuals.

## 2. Related Literature

Homophily in human relationships is a widely known phenomenon extensively investigated in papers such as McPherson, Smith-Lovin, and Cook (2001), Currarini, Jackson, and Pin (2009). Homophily in friendships is investigated by Currarini, Jackson, and Pin (2009). Their paper similarly aims at explaining empirical observations, builds and estimates a model of friendship formation that is compatible with the empirical patterns observed in a sample of high school students in the United States. They reach the similar conclusion that patterns of homophily in friendship formation can be explained only by the simultaneous presence of a bias in meeting frequencies and a bias in preferences towards same-type friends. My model differs from theirs in two aspects: firstly, I look at the one-to-one matching of romantic relationships, rather than the many-to-many matching of friendships; secondly, I do not directly observe the network of acquaintances in which the matching takes place. I instead develop a method in this paper to derive information about the network entirely from matching frequencies.

In recent years there has also been a growing literature around patterns in online dating; papers like Ortega and Hergovich (2017) explore the effects on racial integration of online dating. They argue that by opening up a racially mixed pool of partners in places where social groups tend to be more homogenous, the internet has increased the number of mixed-race couples. Using a simulation based model based on real-world data about racial preferences, they show that in a world where people are highly connected with others of their own race, but only poorly so with people from other races, even random links to perfect strangers will quickly increase the percentage of interracial marriages. Their model takes the online network as exogenously given, and does not take into account the effects of directed search. As a result, their model predicts an even increase in the shares of interracial marriages for all possible combinations of races. As highlighted in the introduction, this result contrasts with the empirical patterns observed in our sample HCMST. Moreover, other papers, such as Hitsch, Hortaçsu, and Ariely (2010a) and Hitsch, Hortaçsu, and Ariely (2010b) find that agents sort along various attributes even when search frictions are minimal, as in the case of online dating. However, their results differ from the ones found in my paper mainly because in the difference degree of homophily that I observe in the offline sample compared to the online sample. Hitsch, Hortaçsu, and Ariely (2010a) argue that the sorting patterns observed online are qualitatively similar to the ones observed offline. I find however there are interesting quantitative differences that I extensively explore in my paper. Moreover those studies, and similar studies on speed dating such as Fisman, Iyengar, Kamenica, and Simonson (2006) and Fisman, Iyengar, Kamenica, and Simonson (2008), have been conducted over 10 years ago. Use of online dating platforms has since then skyrocketed, with most recent estimates reporting that 75% of relationships that started in 2019 have started on an online dating app. Such dramatic changes in attitude toward the use of online dating have most likely also had an impact on the way people use online dating platform and who they match with on these platforms.

As mentioned in the introduction, we will assume that utilities are non transferable. The matching outcome in the NTU setting was first studied by Gale and Shapley in Gale and Shapley (1962) and properties of pairwise stable matchings have been extensively studied in the matching literature (Roth and Sotomayor (1990) provide a comprehensive summary on the topic). An asymptotic approximation of matched observable characteristics in such setting is studied extensively by Menzel (2015). I use and adapt Menzel's identification strategy to my setting in an incomplete network.

## 3. Data

The data used in this paper are obtained from the databases "How Couples Meet and How They Stay Together" (HCMST) [3]) of 2009 and 2017. HCMST is a collection of survey answers about how Americans meet their spouses and romantic partners. The data consists of individual level observations which include detailed information about the subjects' current relationship, if any. The information in the data includes, but is not limited to, the race and ethnicity of

[3]This data is publicly available at https://data.stanford.edu/hcmst

the subject, their age, gender, sexual orientation, educational level, income, religion, political affiliation, geographical location in the United States, whether they were raised in the United States or not, questions about their family of origin, and relationship status. If the participant has a significant other at the time of the study, similar information is collected about their current partner. Furthermore, if the participant has a significant other at the time of the study, information about the relationship is collected. The information about the relationship that is recorded is on the nature of the relationship, in particular whether the couple is married, in a relationship or they are sexual partners. Moreover it records in what year and month they met, and in what year and month the relationship officially started. More importantly for this paper, for each relationship the data report how the couple met: online or offline. Other information is collected, such as the perceived quality of the relationship, that is not going to be used in this paper. If the participant is single, they are asked whether they used online dating services in the past year, which allows me to recover the shares of singles among the online community.

The sample is representative of the American population. The databases of 2009 and 2017 have respectively 4,002 and 3,510 data points. For the scope of the analysis, I will investigate patterns both in committed and casual relationships. I report results both considering as couples only those who reported being married or in a committed relationship, and including also individuals who reported having a sexual partner, but no stable relationship. In the two databases combined 5,686 of interviewed individuals were in a romantic relationship, while 1,826 were single. I restrict attention to couples who have met after 1995, the year Match.com was launched. Furthermore, in order to control for time effects on the perception of an interracial marriage and of online dating, I include in the sample only subjects between the ages of 18 and 75 at the time of the interview. This serves the double purpose of partially controling for the fact that preferences and attitudes toward interracial relationships have changed dramatically over the last few decades, and comparing a sample of individuals who had the possibility of using online dating if they wanted to. Table 1 provides summary statistics for the sample.

### 3.1. **Stylized Facts**

Before moving to the model and its estimation, I present some descriptive statistics of dating patterns. Dating dynamics have changed dramatically over the last decades, and two important trends have arisen. Firstly, after *Loving vs. Virgina* in 1967, when laws banning interracial marriages were declared unconstitutional, the number of interracial marriages started increasing rapidly. Despite interracial marriages still representing a small fraction of the total number of marriages, their number increased over fivefold since the 1970's. Secondly, another relevant change in dating dynamics has been brought upon by the introduction of online dating in the mid-1990's. The entry on the market of Match.com in 1995 marked the beginning of the rise of online dating services. Reports[4] show that the percentage of new relationships that begun online

[4] Tyro "The Dating Market: Thesis Overview"

Table 1 - Demographic Characteristics of the Sample

| | Online Sample | Offline Sample | US Population |
|---|---|---|---|
| Total Population | 699 | 3845 | 328.2 M |
| Percentage of Women | 0.48 | 0.53 | 0.51 |
| *Race* | | | |
| White | 0.72 | 0.71 | 0.71 |
| Black | 0.08 | 0.10 | 0.10 |
| Hispanic | 0.13 | 0.12 | 0.11 |
| Asian | 0.03 | 0.03 | 0.04 |
| Other | 0.04 | 0.04 | 0.03 |
| *Age* | | | |
| 17 to 20 | 0.04 | 0.05 | 0.06 |
| 21 to 30 | 0.27 | 0.27 | 0.23 |
| 31 to 40 | 0.25 | 0.21 | 0.22 |
| 41 to 50 | 0.25 | 0.17 | 0.20 |
| 51 to 60 | 0.17 | 0.15 | 0.21 |
| Over 60 | 0.06 | 0.13 | 0.08 |
| *Education* | | | |
| High School or less | 0.21 | 0.32 | 0.28 |
| Some College | 0.24 | 0.23 | 0.26 |
| College Degree | 0.39 | 0.32 | 0.28 |
| Graduate | 0.15 | 0.13 | 0.18 |

[1]

[1]*Source*: HCMST 2009-2017

in 2019 is around 75%. Facts 3 and 4 present descriptive statistics of how these two trends have progressively influenced dating dynamics. Fact 3 shows the increasing trend in the percentage of interracial relationships since the 1960's. Fact 4 provides evidence of the increasing relevance of online dating in the way couples meet. These changes in the dating scene, have had consequences on assortativeness patterns.

In this section I also present descriptive statistics of the American social network and of ethnic sorting patterns in romantic relationships. Facts 1 and 2 provide evidence of homophily patterns both in social and romantic relationships in the United States. Fact 2 further suggests that a considerable part of the observed homophily is driven by preferences, and that therefore, even in a world without search frictions, we could still witness homogamy.

Finally, I present comparisons in sorting patterns among the online and offline samples. Fact 5 shows that couples who meet online are more likely to be ethnically and racially diverse when compared to couples who met offline. Fact 6 suggests the presence of directed search on online dating markets. It does in fact show that when focusing on minorities, the sample of couples who met online is much more likely to be in a relationships with a white person, but equally or less likely to be in a relationship with a person belonging to a different minority group. Facts 5

https://2378nh2nfow32gm3mb25krmuyy-wpengine.netdna-ssl.com/wp-content/uploads/2019/11/Menratewomen.pdf and 6 jointly suggest that the network has a major effect on matching frequencies offline, and that directed search has a major effect on matching frequencies online.

The descriptive results presented here serve as strong suggestive evidence of the network effect on matching probabilities, and in turn of the effect that online dating has on the network effect. These facts provide support to the theoretical model by suggesting that agents on online dating platforms direct their search towards specific subgroups of the population.

***Fact 1: The social network in the United States is highly clustered***

Racial segregation among Americans social networks is still pervasive across ethnic and racial groups. More than 75% of white Americans report not having any connections in their core social network to non-white individuals and about 92% of the acquaintances of a white person are also white. The same pattern is observed to a lesser degree for black, asian and hispanic people who live in the United States. Despite efforts to promote social integration and diversity in education, in the work place and other institutions, patterns of composition of social networks have not changed much in the past few years. Studies conducted in 2020 by the American National Social Network Survey found essentially the same patterns of racial homogeneity found in 2013. The racial and ethnic segregation found in Americans' social networks is pronounced among white Americans regardless of educational background, age, or political affiliation. Table 2 summarizes information on the racial composition of acquaintances of individuals in the United States.

***Fact 2: All ethnic groups are biased towards same ethnicity partners***

Couples present a clear pattern of homophily, a tendency to match more often with other people of the same ethnic group. This phenomenon is widely studied in the literature, although it is still unclear what part of such pattern is due to preferences and what part is due to a higher meeting frequency between people of the same race. Table 3 displays the percentage composition of couples across different races. The first columns represents the marginals of each ethnicity in the general population.

Similar findings are reported in Hitsch, Hortaçsu, and Ariely (2010a), where the authors find sorting patterns in a sample of individuals from an online dating webiste.

Suggestive evidence of same-race bias in preferences is provided by Table 3.1 produced by OkCupid in 2004 and reported in the book Dataclysm by Rudder (2014). Users on OkCupid were asked to rate potential partners on a scale from 1 to 5, where 1 meant they did not find the proposed partner attractive and 5 meant they found them very attractive. The number in the box represents the percentage difference between the average score conditional on ethnicity and the average score across ethnicities. Users preferences display a clear tendency for ethnic homophily, especially among women.

Table 2 - Network composition

| | Marginals | White | Black | Hispanic | Asian |
|---|---|---|---|---|---|
| White | 0.71 | 0.92 | 0.15 | 0.33 | 0.27 |
| Black | 0.10 | 0.02 | 0.77 | 0.05 | 0.02 |
| Hispanic | 0.10 | 0.02 | 0.03 | 0.53 | 0.03 |
| Asian | 0.04 | 0.01 | 0.01 | 0.02 | 0.63 |
| Multi-race | 0.03 | 0.02 | 0.04 | 0.05 | 0.05 |
| Other | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 |

[2]

Table 3 - Romantic Relationships Composition

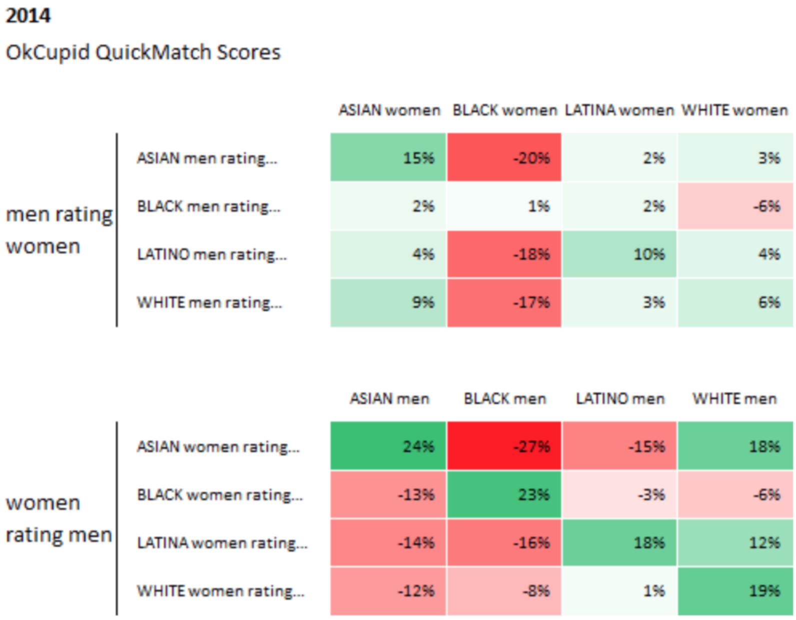
| | Marginals | White | Black | Hispanic | Asian |
|---|---|---|---|---|---|
| White | 0.71 | 0.87 | 0.14 | 0.42 | 0.33 |
| Black | 0.10 | 0.02 | 0.70 | 0.09 | 0.02 |
| Hispanic | 0.10 | 0.06 | 0.11 | 0.43 | 0.10 |
| Asian | 0.04 | 0.03 | 0.02 | 0.02 | 0.50 |
| Other | 0.05 | 0.02 | 0.03 | 0.04 | 0.04 |

[3]

[2]Source: American National Social Network Survey, August 2020.
[3]Source: HCMST 2009-2017

***Fact 3: The share of interracial marriages is small, but steadily increasing***

In the 1967 *Loving v. Virginia* the U.S. Supreme Court ruled that marriage across racial lines was legal throughout the country. That year the percentage of interracial marriages over total number of new marriages was 3%. In 2015, the percentage of interracial marriages has increased more than fivefold to 17%. Despite it still being a small fraction of the number of total marriages, the number of romantic relationships between individuals belonging to different ethnic groups has increased significantly and consistently over the last 60 years. The United States Census Bureau reports that the number of interracial marriages in 1970 was 310,000, compared to 2,340,000 in 2008. It is important to keep in mind that Census reports do not provide information on inter-ethnic couples, i.e. they do not keep track of marriages between a hispanic and a non-hispanic person if they belong to the same race. The sample in HCMST allows us to both control for ethnicity and to keep track of all romantic relationships, not only marriages. On the horizontal axis of the table below is the year the relationship between two people in the sample started. From the data in the sample, we see that the percentage of interracial relationships that started before 1960 is below 5% of the total number of relationships that started that year. Such percentage increased over the years to about 35% in 2017. Figure 3.2 shows the trend of the new interracial relationships that were formed each year in the sample HCMST.



**2014**

OkCupid QuickMatch Scores

| men rating women | ASIAN women | BLACK women | LATINA women | WHITE women |
|---|---|---|---|---|
| ASIAN men rating... | 15% | -20% | 2% | 3% |
| BLACK men rating... | 2% | 1% | 2% | -6% |
| LATINO men rating... | 4% | -18% | 10% | 4% |
| WHITE men rating... | 9% | -17% | 3% | 6% |

| women rating men | ASIAN men | BLACK men | LATINO men | WHITE men |
|---|---|---|---|---|
| ASIAN women rating... | 24% | -27% | -15% | 18% |
| BLACK women rating... | -13% | 23% | -3% | -6% |
| LATINA women rating... | -14% | -16% | 18% | 12% |
| WHITE women rating... | -12% | -8% | 1% | 19% |

FIGURE 3.1. Dataclysm
Rudder (2014)

***Fact 4: The fraction of relationships that begun online increased steadily since 1995***

In 1995 the first online dating website, Match.com, was launched. Online dating websites were initially mostly used by individuals who were facing "liquidity constraints" in the marriage market. These were the individuals who had few potential partners within their set of acquaintances. For instance, a socio-demographic group that largely utilized online dating platforms when they first came out were homosexual people. However, online dating services picked up quickly, and since 1995 the number of users on dating platforms and the number of relationships that started online has increased dramatically. According to a 2020 report by Pew Research, 30% of Americans adults reported having used an online dating website or app. The percentage of couples that met online over the total number of couples has in particular peaked after the entry in

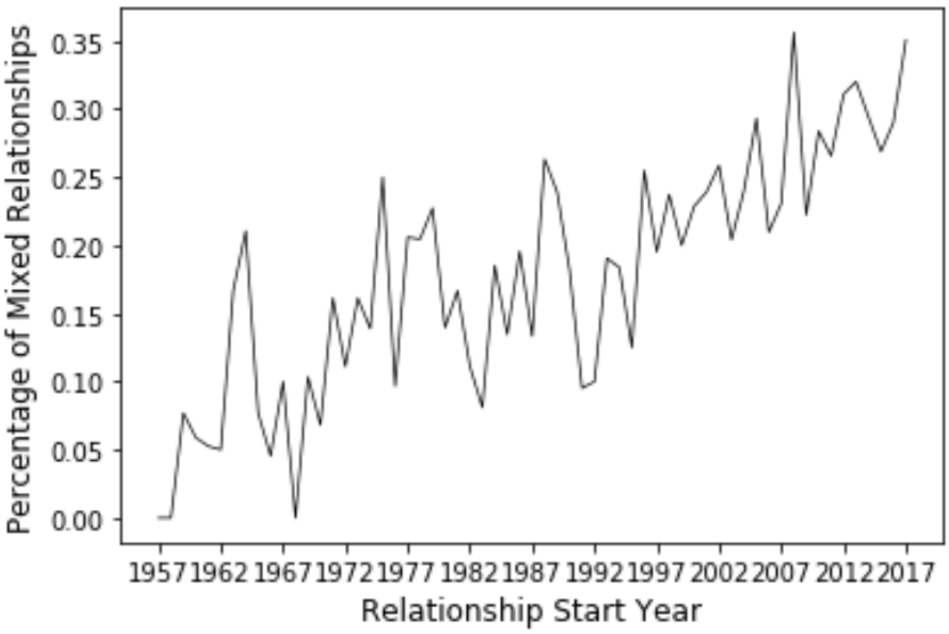
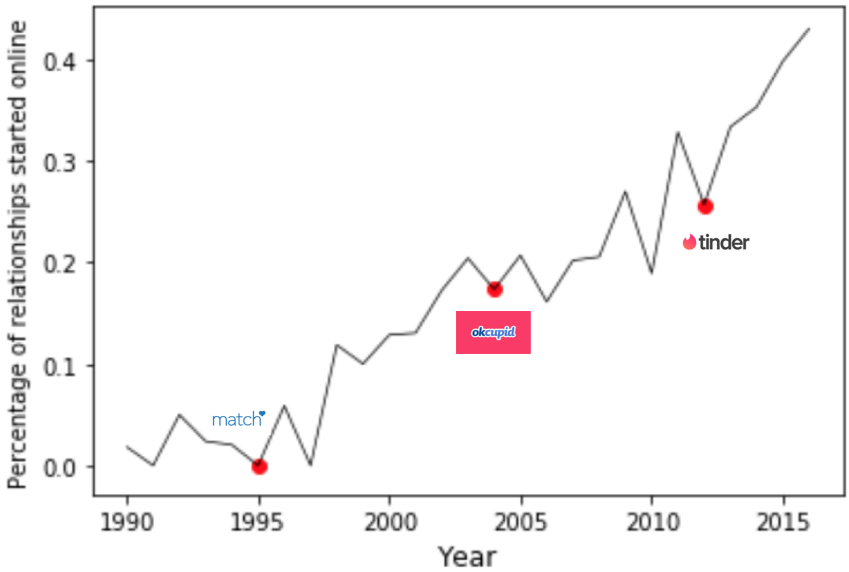



FIGURE 3.2. Percentage of newly formed mixed relationships per year in HCMST.
*Source*: HCMST 2009-2017

the market of three players: Match.com in 1995, OkCupid in 2004 and Tinder in 2012. As of September of 2019, Tinder reported having 7.86 million users. Focusing attention on the sample at hand, we can see that the percentage of couples that met online has increased from virtually 0 in the early 1990's to 43% in 2017. The percentage of relationships that started online has been increasing year after year, with market reports estimating that 75% of heterosexual relationships that began in 2019, began online. Figure 3.3 displays the percentage of couples who met online from 1990 to 2017.

***Fact 5: Couples that meet online are more likely to be ethnically diverse***

Splitting the database into couples who have met online and couples who have met offline, it is immediately noticeable that shares of inter and intra marriages are significantly different for the two groups. The number of couples that have met online in the sample is 648, which is 10.57% percent of the total number of couples recorded in the database. Tables 4 and 5 report the observed frequency of relationships conditional on ethnicity of the subject interviewed. For instance, 83% of white people who met their partner online have a white partner, compared to the 88% of white people who met their partner offline and are in a relationship with another white person. Figures along the diagonal are consistently higher for the sample of couples who




z

FIGURE 3.3. Percentage of newly formed relationship that started online per year. Source: HCMST 2009-2017

TABLE 4 - COMPOSITION OF COUPLES WHO MET ONLINE

| | Marginals | White | Black | Hispanic | Asian |
|---|---|---|---|---|---|
| White | 0.71 | 0.83 | 0.24 | 0.53 | 0.50 |
| Black | 0.10 | 0.04 | 0.60 | 0.05 | 0.05 |
| Hispanic | 0.10 | 0.07 | 0.08 | 0.32 | 0.05 |
| Asian | 0.04 | 0.04 | 0.06 | 0.02 | 0.36 |
| Other | 0.05 | 0.01 | 0.02 | 0.07 | 0.05 |

TABLE 5 - COMPOSITION OF COUPLES WHO MET OFFLINE

| | Marginals | White | Black | Hispanic | Asian |
|---|---|---|---|---|---|
| White | 0.71 | 0.88 | 0.12 | 0.38 | 0.27 |
| Black | 0.10 | 0.02 | 0.72 | 0.10 | 0.01 |
| Hispanic | 0.10 | 0.06 | 0.12 | 0.46 | 0.13 |
| Asian | 0.04 | 0.02 | 0.01 | 0.02 | 0.54 |
| Other | 0.05 | 0.02 | 0.03 | 0.04 | 0.04 |

[4]

[4] *Source*: HCMST 2009-2017

met offline, showing a higher degree of same-ethnicity concentration of relationships in the latter sample. The difference in frequencies of mixed couples in the online and offline population is statistically significant for all ethnicities with a $p$-value of less than 1%.

Table 6 - Minorities Relationships Composition

| | Online | Offline | $\Delta$ | Online | Offline | $\Delta$ |
|---|---|---|---|---|---|---|
| Same-Race Partner | 0.36 | 0.52 | $-0.16^{***}$ | 0.33 | 0.41 | $-0.08^{**}$ |
| Different-Race White Partner | 0.47 | 0.31 | $0.16^{***}$ | 0.44 | 0.24 | $0.20^{***}$ |
| Different-Race Minority Partner | 0.17 | 0.17 | 0.00 | 0.23 | 0.35 | $-0.12^{***}$ |
| Casual Relationships Included | N | N | N | Y | Y | Y |

*Note:* *p<0.1; **p<0.05; ***p<0.01

[5]

[5] *Source*: HCMST 2009-2017

***Fact 6: Minorities who meet their partner online are less or equally likely to be in a relationship with a member of a different minority group***

Comparing the sample of couples who met online with the sample of couples who met offline in Tables 4 and 5, it is possible to observe that white people who meet their partner online are more likely to be in a mixed relationship of any kind than their offline counterparts. Focusing on minorities, however, their probability of being in a mixed relationship with a white person is much higher than it is offline. In the case of Hispanic and Asians who meet their partner online, the assortative pattern reverts, making them more likely to be in a relationship with a white person than with someone of their own ethnic group. This difference is remarkable.

However, the percentage of relationships between members of different minority groups (i.e. hispanic and asian, or asian and black) is not significantly different for the online and offline samples. Looking only at committed relationships there is no statistical difference in the percentage of mixed-minority couples in the online compared to the offline samples. Including casual relationships (subjects who reported having a sexual partner, but not being in a relationship) in the sample, the percentage of mixed-minority relationships becomes significantly less in the online sample, compared to the offline sample. Grouping together all non-white observations and only considering subjects who reported being married or in a relationship, 52% of the offline sample relationships are same-race, compared to the 36% of the online sample. This represents a reversal of assortativeness, because the majority of relationships for the minority online sample is mixed. The percentage of relationships with white people goes from 31% in the offline sample, to 47% in the online sample. However, the percentage of relationships with other minority members belonging to a different minority is equal to 17% both in the online and offline sample. Table 6 summarizes these statistics: in the first three columns I am only considering marriages and committed relationships, while in the second three columns I am including in the sample casual relationships as well. The column denoted $\Delta$ represents the difference in percentage between the online and the offline sample for each type of relationship, and its significance level.

Concluding, the above mentioned facts point at the presence of two important trends in dating of the last few decades: a significant increase in the number of interracial couples and a significant increase in the number of couples who met online. The estimation part of this paper aims at determining whether the two trends are related, and in particular whether online dating has contributed to social integration in the form of an increase in the number of inter-racial and inter-educational couples. I will show in Section 7 that online dating has had a positive effect on the number of interracial couples, but only in certain parts of the United States, namely the Northeast and the West, and there is no significant evidence of it having an effect in the South and Midwest. Moreover, I will show that the rate on white-minority couples in the online sample is significantly higher, both before and after adjusting for network estimates. However, this is not the case for mixed-minority couples. The share of mixed-minority couples (individuals belonging to two different minority groups) is not significantly different in the online sample compared to the offline sample.

## 4. Model

Let $\mathcal{I}$ and $\mathcal{J}$ be two sides of the market. Individuals $i$ belonging to $\mathcal{I}$ can match with individuals $j$ in $\mathcal{J}$ or choose to remain single. Similarly for individuals in $\mathcal{J}$. Individuals on both sides of the market $\mathcal{I} \times \mathcal{J}$ are also nodes of an exogenous network of friendships. Using standard notation as in Jackson (2010), let the exogenous network be described by an adjacency matrix $\mathcal{N}$, where the elements $N_{ij}$ of $\mathcal{N}$ are

$$N_{ij} = \begin{cases} 1 & \text{if there is a direct link between } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$$

The network is undirected, therefore if $N_{ij} = 1$ it must be that $N_{ji} = 1$. The adjacency matrix $\mathcal{N}$ is therefore symmetric. We will denote with $\mathcal{N} + \{ij\}$ the network resulting from adding edge $ij$ to $\mathcal{N}$. In order to take into account the effects of the incompleteness of the network into matching probabilities, we assume that two individuals $i \in \mathcal{I}$ and $j \in \mathcal{J}$ can match only if they know each other, that is if $N_{ij} = 1$. Agents are aware of the set of people they are connected to and therefore aware of what matching partners are available to them. Formally, let the matching outcome be described by the matching matrix $\mu$. The entries $\mu_{ij}$ of the matching matrix are

$$\mu_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are matched} \\ 0 & \text{otherwise} \end{cases}$$

**Definition 1.** A feasible matching is a matrix $\mu$ such that for every $i$ and $j$

$$\mu_{ij} \in \{0,1\} \tag{4.1}$$

$$\sum_{i \cup \{0\}} \mu_{ij} = 1 \tag{4.2}$$

$$\sum_{j \cup \{0\}} \mu_{ij} = 1 \tag{4.3}$$

$$\mu_{ij} \leq N_{ij} \tag{4.4}$$

In this setting, (4.1) indicates that fractional assignments are not allowed; jointly, (4.2) and (4.3) express that the matching is one-to-one; (4.4) formalizes the assumption that a matching can occur only between two people who are connected in the network.

## 4.1. **Types, Preferences and Network**

Each individual is described by a vector of observable characteristics. Let $x_i \in \mathcal{X}$ and $y_j \in \mathcal{Y}$ denote the vectors of observable characteristics of individuals $i \in \mathcal{I}$ and $j \in \mathcal{J}$ respectively. Assume $\mathcal{X}$ and $\mathcal{Y}$ are bounded subset of some Euclidean space. The marginal distributions of types in $\mathcal{X}$ and $\mathcal{Y}$ are given respectively by $p(x)$ and $q(y)$. The observable characteristics of an individual affects simultaneously the systematic utility they obtain when they match with someone, and the type of people that they are connected to in the network. I consider a matching model with non transferable utilities (NTU), where individual $i$ derives utility $u_{ij}$ from being matched with individual $j$ and individual $j$ derives utility $v_{ji}$ from being matched with individual $i$. The utility $u_{ij}$ is modeled to be the sum of the systematic utility $u_{x_i y_j}$, that only depends on the observable types of $i$ and $j$, and an independent individual-specific chemistry shock $\varepsilon_{ij}$. Similarly for $v_{ji}$.

$$u_{ij} = u_{x_i y_j} + \varepsilon_{ij} \ \forall i \in \mathcal{I}$$
$$v_{ji} = v_{y_j x_i} + \eta_{ji} \ \forall j \in \mathcal{J}$$

We assume $\varepsilon, \eta$ are independent across $i$ and $j$'s and identically distributed as a Gumbel with mean 0 and variance $\sigma$. Chemistry shocks are realized upon meeting someone, which can only happen if those two individuals know each other. The systematic utility for the outside option, that is remaining single, is normalized to zero. The random utility obtained by $i$ and $j$ that remain unmatched is

$$u_{i0} = \varepsilon_{i0} \ \forall i \in \mathcal{I}$$
$$v_{j0} = \eta_{j0} \ \forall j \in \mathcal{J}$$

where, following Menzel (2015) $\varepsilon_{i0} = \max_{k=1,\ldots \lfloor, \sqrt{n} \rfloor} \{\varepsilon_{i0,k}\}$ and $\eta_{j0} = \max_{k=1,\ldots,\lfloor \sqrt{n} \rfloor} \{\eta_{j0,k}\}$ .

As mentioned above, the type of an individual will impact not only the utility the obtain from a potential match, but also the type of potential partners they are directly connected to in the network $\mathcal{N}$. Define the network degree of $i$ as

$$d(\mathcal{N};i) = \sum_{j \neq i} N_{ij}$$

And the type-conditional network degree of $i$ as

$$d(\mathcal{N},y;i) = \sum_{\substack{j \neq i; \\ y_j = y}} N_{ij}$$

In words, $d(\mathcal{N};i)$ is the size of the set of acquaintances of individual $i$, and is the size of the subset of acquaintances of $i$ that are of type $y$. Therefore, the fraction of acquaintances of type $y$ in $i'$'s network is described by $q(\mathcal{N},y;i)$

$$q(\mathcal{N},y;i) = \frac{d(\mathcal{N},y;i)}{d(\mathcal{N};i)}$$

In this model, same types have similar sets of acquaintances. Formally, we make the assumption that for every $y \in \mathcal{Y}$ and every $x \in \mathcal{X}$

$$d(\mathcal{N},y;i) = d(\mathcal{N},y;x) \text{ for all } i \text{ such that } x_i = x$$
$$d(\mathcal{N},x;j) = d(\mathcal{N},x;y) \text{ for all } j \text{ such that } y_i = y$$

For ease of notation we denote $\delta_x = d(\mathcal{N};x)$ and $\delta_{xy} = d(\mathcal{N},y;x)$. To model the clustering of the network, I will assume that for certain characteristics $x$, we have $\delta_{xx} > \delta_{xy}$ .

**Definition 2.** Let the coefficient of exposure $\beta$ be a value $\beta_{xy} \in \mathbb{R}_+$ such that

$$Pr\left\{N_{ij} = 1 \,|x_i = x, y_j = y\right\} = \frac{\delta_{xy}}{\delta_x} \cdot \frac{\delta_{yx}}{\delta_y} = \beta_{xy} p_x q_y$$

where $p_x$ and $q_y$ are the masses of $x$ and $y$ respectively.

The coefficient of exposure $\beta$ links the characteristics of the network to the relative masses of types in the economy. $\beta$ can be considered a Radon-Nykodim derivative measuring the degree of distortion in meeting probabilities, compared to the neutral benchmark of uniform meeting probabilities, independent of types. I am going to use both $\delta$ and $\beta$ as measures of clustering, depending on which one is more convenient to use. The measure $\frac{\delta_{xy}}{\delta_x} > q_y$ indicates over-exposure of types $x$ to types $y$ but does not provide information on the exposure of $y$ to $x$. On the other hand, $\beta_{xy} > 1$ indicates clustering of the network and an over-exposure of both $x$ to $y$ and $y$ to $x$.

Assume that agents have a strict preference order over potential partners in their set of acquaintances and that utilities derived from a match are non transferable (NTU). Once they have established their set of acquaintances, agents in the economy engage in a Deferred Acceptance algorithm à la Gale and Shapley resulting in a pairwise stable matching. A matching $\mu$ is pairwise stable if, given the preferences expressed by $u_{ij}$ and $v_{ji}$, each agent prefers their partner under the matching $\mu$ to any other achievable partner.

**Definition 3.** Let $\mu(i)$ and $\mu(j)$ indicate the partners matched with $i$ and $j$ respectively under matching $\mu$. The matching $\mu$ is stable if
(i) if $u_{ij} > u_{i\mu(i)}$ then $v_{j\mu(j)} \geq v_{ji}$ and
(ii) if $v_{ji} > v_{j\mu(j)}$ then $u_{i\mu(i)} \geq u_{ij}$.

The existing literature informs us that if preferences are strict, a pairwise stable matching always exists. It is important to notice that existence of a pairwise stable matching does not require that agents have perfect knowledge of all participants' preferences, but each matching partner needs to have awareness of what potential partners are available to her. Therefore in this context, individuals need to be aware of the agents in their personal network of acquaintances and in equilibrium they need to be aware of the indirect utility obtained by those agents.

Let $\mathcal{N}(i) = \{j \in \mathcal{J} : \; N_{ij} = 1\}$ be the set of acquaintances of individual $i$ in network $\mathcal{N}$. Let $Z$ be the random variable representing the indirect utility of agent $i$ when she has chosen the best candidate out of her choice set

$$Z_i = \max_{j \in O(i) \cap \mathcal{N}(i)} u_{x_i y_j} + \varepsilon_{ij}$$

For the next result recall that $\delta_{xy} = d(\mathcal{N}, y; x)$ is the number of individuals of type $y$ that an agent of type $x$ is directly connected to in the network, and that $\delta_x = \sum_y \delta_{xy}$ is the total number of individuals connected to her in the network.

**Proposition 1.** *The conditional expectation of the indirect utility for individual $i$ of type $x$, given the network $\mathcal{N}$ is*

$$\mathbb{E}\left[Z_i \,|x_i = x\right] = \gamma + \log\left(\sum_{y \in \mathcal{Y}} \delta_{xy} \exp\left\{u_{xy}\right\}\right)$$

*where $\gamma$ is the Euler-Mascheroni constant.*

*Proof.* In Appendix A

∎

The expression $\frac{1}{n}\sum_{y \in \mathcal{Y}} \exp\left\{u_{x_i y}\right\}$ is referred to in the literature as the inclusive value of agent $i$ of type $x$, and denoted $I_i(x)$. The inclusive value has the form of a sample average over the utility that potential partners are able to provide in case of a match. I will call inclusive value the modified version of it $\frac{1}{n}\sum_{y \in \mathcal{Y}} \frac{\delta_{xy}}{\delta_x} \exp\left\{u_{x_i y}\right\}$, which takes into account that agents are not connected to everyone else in the network. Notice that $\frac{1}{n}\sum_{y \in \mathcal{Y}} \exp\left\{u_{x_i y}\right\} \geq \frac{1}{n}\sum_{y \in \mathcal{Y}} \frac{\delta_{xy}}{\delta_x} \exp\left\{u_{x_i y}\right\}$, which implies that individual welfare is always higher in a complete network. Menzel (2015) shows that for large matching markets, when $n \to \infty$, the inclusive value $I_i(x)$ converges to its conditional expected value given type $x_i$. This implies that two agents with similar characteristics will have in expected value a similar indirect utility as the market grows large.

### 4.2. **A Population Game of Strategic Network Formation**

Now suppose that a meeting technology is introduced that allows agents to strategically modify their set of acquaintances. By modifying and expanding their set of acquaintances, agents can increase the conditional expectation of their indirect utility from a match. The new technology allows each $i$ to increase $d(\mathcal{N}, y; i)$ for each $y$, that is it allows the agent to increase the number of individuals of type $y$ that they are connected to in the network. However, such modifications are costly. Recall that we have assumed that the network is such that $d(\mathcal{N}, y; i)$ is type dependent but not individual dependent, and in particular that it is set to be equal to $\delta_{xy}$ for every $i$ of type $x$. Since the expectation of the indirect utility for an agent $i$ of type $x$ is $\gamma + \log\left(\sum_{y \in \mathcal{Y}} \delta_{xy} \exp\{u_{xy}\}\right)$, agents have an incentive to increase their exposure $\delta_{xy}$ towards individuals of those types $y$ that, in case of a match occurring, will provide them with a high utility. In particular, since the shock only occurs once the link in the network is created, agents will direct their search towards individuals who provide them with a high systematic utility $u_{xy}$, since they have no way of predicting ex-ante what will be the magnitude of the chemistry shock $\varepsilon_{ij}$.

Formally, agents in the economy engage in the following two stage simultaneous game: in the first stage players decide whether to purchase the new meeting technology or not. The cost of purchasing such technology is individual specific and equal to $\kappa_i$. We assume that $\kappa_i$ is independent across $i$'s and identically distributed with c.d.f. $F_\kappa(k)$. Purchasing the technology provides access to a meeting platform that allows connections to be formed among a pool of potential partners. The technology also allows agents to strategically expand their set of acquaintances towards types they prefer. If agents decide to not purchase the technology, their expected indirect utility is the one expected from matching in the exogenous social network as described in the previous section. If on the other hand, they do purchase the technology, in the second stage after entering into the meeting platform they make strategic decisions on how to expand and modify their network of acquaintances. Such modifications require effort and are, therefore, costly. The expected indirect utility derived if the agent enters the game is the convex combination of the exogenous expected utility and the inclusive value obtained online. To summarize, both purchasing the meeting technology and directing search using the technology are costly: there is an individual specific fixed cost associated with buying access to the online dating platform and a variable cost that depends on the effort made to meet specific types of individuals on the platform.

The technology provides access to the pool of potential partners who have also purchased the same technology. Assume that in equilibrium the technology introduces a number equal to $\tilde{\delta}_{xy}$ of types $y$ to an individual $i$ of type $x$. If agents purchase the technology, in equilibrium their expected indirect utility is the convex combination of the expected indirect utility in the exogenous social network and the expected indirect utility derived in the endogenous network of online dating. Formally:

$$\mathbb{E}\left[U_i \,|x_i = x,\ \text{purchasing tech}\right] = \gamma + (1-\theta)\log\left(\sum_{y\in\mathcal{Y}} \delta_{xy} \exp\left\{u_{xy}\right\}\right) + \theta \log\left(\sum_{y\in\mathcal{Y}} \tilde{\delta}_{xy} \exp\left\{u_{xy}\right\}\right) - \kappa_i$$

Denote $\tilde{p}(x)$ and $\tilde{q}(y)$ the marginal distributions of types $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively that have purchased the technology. Notice that $\tilde{p}(x)$ and $\tilde{q}(y)$ are endogenous. We will now make the extra assumption that the technology is type-neutral, meaning it introduces to each user a random set of $\lambda$ potential partners selected independently of their type. Namely, if $n$ people from $\mathcal{J}$ have purchased the technology, the probability of any of them being selected to be introduced to an agent $i$ from $\mathcal{I}$ who has also purchased the technology, is simply $\frac{1}{n}$. Therefore, the expected number of people of type $y$ that any agent will be introduced to is equal to $\tilde{\delta}_{xy} = \lambda \tilde{q}_y$ for all $x \in \mathcal{X}$, where $\tilde{q}_y$ is the fraction of types $y$ that have purchased the technology and $\lambda > 0$ is an exogenous parameter. We will call this assumption type-neutrality of the online dating platform.

**Assumption 1:** *The meeting technology is type-neutral, that is $\tilde{\delta}_{xy} = \lambda \tilde{q}_y$ for every $x \in \mathcal{X}$, and for a constant $\lambda > 0$.*

Assumption 1 implies that the platform does not favor certain types over others by presenting them to potential partners more often. The distribution of partners each agent is introduced to is equal to the distribution of types that are on the meeting platform. Under Assumption 1, the expected indirect utility of agent $i$ is in this case

(4.5)

$$\mathbb{E}\left[U_i \,|x_i = x,\ \text{purchasing tech}\right] = \gamma + (1-\theta)\log\left(\sum_{y\in\mathcal{Y}} \delta_{xy} \exp\left\{u_{xy}\right\}\right) + \theta \log\left(\lambda \sum_{y\in\mathcal{Y}} \tilde{q}_y \exp\left\{u_{xy}\right\}\right) - \kappa_i$$

Where we substituted $\tilde{\delta}_{xy} = \lambda \tilde{q}_y$ from the previous expression. Let us now focus only on the part of expected utility derived from purchasing the technology $\log\left(\lambda \sum_{y\in\mathcal{Y}} \tilde{q}_y \exp\left\{u_{xy}\right\}\right)$. In order to make the problem more tractable and without loss of generality we will consider the exponential of such expected utility, which we will refer to with an abuse of notation as the inclusive value online and denote $\tilde{I}(x)$ for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

$$\tilde{I}(x) = \lambda \sum_{y\in\mathcal{Y}} \tilde{q}_y \exp\left\{u_{xy}\right\}$$

$$\tilde{I}(y) = \lambda \sum_{x\in\mathcal{X}} \tilde{p}_x \exp\left\{v_{yx}\right\}$$

The above expressions represent the inclusive values and indirect utility of agents that have downloaded the platform, but used it passively, that is who have not directed their search. The technology allows agents to alter $\tilde{q}_y$ in the following way: if types $y$ direct their attention towards types $x$ while agent $i$ of type $x$ directs their attention towards types $y$ the derived exposure will be

$$(1 + a_{iy}) \sum_{j:y_j=y} (1 + a_{jx}) = (1 + a_{iy})(1 + a_{yx})\lambda\tilde{q}_y$$

If effort is 0 on both sides of the market, people on the platform randomly bump into each other following a uniform distribution. Therefore, exposure to any particular type will be proportional to the number of agents of that particular type that are present on the platform. If however $a > 0$ for both sides of the market and for a particular type combination $x$ and $y$, $x$ and $y$ will become over-exposed to each other. On the other hand $a < 0$ will cause types $x$ and $y$ to be under-exposed to each other. The action profile $a_i$ has cost $c(a_i)$. We assume that agents are utility maximizers, that is the problem that they aim to solve is

$$\max_{a_i \in A} U(a_i, a_{-i} \,|x_i = x\,) = \max_{a_i \in A} \lambda \sum_{y \in \mathcal{Y}} (1 + a_{iy})(1 + a_{yx})\tilde{q}_y \exp\{u_{xy}\} - c(a_i) - \kappa_i \tag{4.6}$$

$$\max_{a_j \in A} V(a_j, a_{-j} \,|y_j = y\,) = \max_{a_i \in A} \lambda \sum_{x \in \mathcal{X}} (1 + a_{jx})(1 + a_{xy})\tilde{p}_x \exp\{v_{yx}\} - c(a_j) - \kappa_i \tag{4.7}$$

A link between two agents forms if and only if the incremental benefit of that link for either individual exceeds the cost of forming it. Finally, agents are assumed to be rational. They purchase the technology if and only if the net increase in inclusive value minus the cost of directing search is higher than the fixed cost of purchasing the technology. That is agents purchase the technology if and only if the rationality constraint $(RC)$ is satisfied

$$(RC_x) \quad \max_{a_i \in A} \lambda \sum_{y \in \mathcal{Y}} (1 + a_{iy})(1 + a_{yx})\tilde{q}_y \exp\{u_{xy}\} - c(a_i) \geq \kappa_i$$

$$(RC_y) \quad \max_{a_i \in A} \lambda \sum_{x \in \mathcal{X}} (1 + a_{jx})(1 + a_{xy})\tilde{p}_y \exp\{v_{yx}\} - c(a_j) \geq \kappa_i$$

Since the chemistry shock is realized only upon meeting an individual, during the search period agents are maximizing their utility in expected value. I now define an equilibrium concept in this context.

**Definition 4.** An equilibrium in the dating game $\Gamma = \left\langle \mathcal{I} \times \mathcal{J}, (S_i)_{i \in \mathcal{I} \times \mathcal{J}}, (U_i(\cdot))_{i \in \mathcal{I} \times \mathcal{J}} \right\rangle$ is

- A decision to Enter or Not Enter into the meeting platform for each player $i$
- An attention profile $(a_{iy})_{y \in \mathcal{Y}}$ for all $i \in \mathcal{I}$ and $y \in \mathcal{Y}$

– An attention profile $(a_{jx})_{x\in\mathcal{X}}$ for all $j \in \mathcal{J}$ and $x \in \mathcal{X}$

Such that

– The attention profile $(a_{iy})_{y\in\mathcal{Y}}$ maximizes the payoff in (4.6), given the distribution of types $\tilde{q}(y)$ in $\mathcal{J}$ who have entered the platform
– The attention profile $(a_{jx})_{x\in\mathcal{X}}$ maximizes the payoff in (4.7), given the distribution of types $\tilde{p}(x)$ in $\mathcal{I}$ who have entered the platform
– The decision to Enter or Not Enter is optimal given the indirect utility in the subgame after entry

The strategy space of each player $i$ is $S_i = \left\{(E, NE), (a_{iy})_{y\in\mathcal{Y}}\right\}$ where $(E, NE)$ is the decision of entering or not entering the meeting platform (purchasing or not purchasing the meeting technology) and $a_{iy} \in [\underline{a}, \bar{a}]$ is the amount of effort player $i$ directs towards a player of type $y$ on the online dating platform.

## 4.3. **Solution**

Assume that agents in the economy are expected utility maximizers. We have shown in the previous section how in this setting that is equivalent to being inclusive value maximizers. If agents enter the population game by purchasing the online dating technology they will choose an attention profile $a_i$ that maximizes the monotone transformation of their inclusive value $\tilde{U}(a_i, a_{-i}\,|x_i = x)$ of being on the platform. Moreover, they will purchase access to the platform if and only if the benefit of doing so exceeds its cost.

**Proposition 2.** *If the strategy space $A$ is a compact set and the cost function $c : A \to \mathbb{R}$ is continuous and convex in $A$, then a pure strategy Nash equilibrium exists in the subgame after entry.*

*Proof.* The strategy space $A$ is compact set for every $i$ and every $y$ and for every $j$ and every $x$. The inclusive values $\tilde{I}(a_i, a_{-i}\,|x_i = x)$ and $\tilde{I}(a_j, a_{-j}\,|y_j = y)$ are continuous in $a_{-i}$ and $a_{-j}$ respectively. Since $c$ is a continuous and convex function of $a_{iy}$ for every $y$ (and of $a_{jx}$ for every $x$), the payoff function $U_i(a_i, a_{-i})$ is continuous and concave in $a_i$ and similarly $V_j(a_j, a_{-j})$ in $a_j$. By Kakutani's theorem, a pure strategy Nash equilibrium exists in the subgame after buying the technology. Since the strategy space at the first stage is finite, a Nash Equilibrium exists in the game.

□

Let $a^{\star}=\left\{\left(a_{iy}^{\star}\right)_{y\in\mathcal{Y}},\left(a_{jx}^{\star}\right)_{x\in\mathcal{X}}\right\}_{i,j\in\mathcal{I}\times\mathcal{J}}$ be an equilibrium strategy profile and let $U_i(x)$ be the utility of an agent $i$ of type $x$ that enters into the game by downloading the online dating platform.

$$U_i(x)=\left[\lambda\sum_{y\in Supp(\tilde{q})}\left(1+\bar{a}_{yx}^{\star}\right)\left(1+a_{iy}^{\star}\right)\tilde{q}_y\exp\left\{u_{xy}\right\}-c(a_{iy}^{\star})\right]-\kappa_i$$

$$\bar{a}_{yx}^{\star}=\int_{\{j:y_j=y\}}a_{jx}d\tilde{q}$$

As seen in the previous section, agents who download the platform obtain an expected utility that is the convex combination of their inclusive value offline $I(x)$, and the expected value of utility they obtain online $U_i(x)$. Let $V_i\left(x\right)$ be the indirect utility of individual $i$ of type $x$ that is using an online dating platform. The value of $V_i(x)$ is given by

$$V_i(x)=\left(1-\theta\right)I\left(x\right)+\theta U_i(x)$$

Remember that if the agent does not download the platform their utility is fixed and equal to their inclusive value $I(x)$. Each individual $i$ is assumed to behave optimally, which implies that if they download the platform the effort exerted to try and meet potential partners of type $y$ is given by

$$a_{iy}^{\star}\in\text{argmax}_{a\in[0,1]}U_i(a_i,a_{-i})$$

In particular if $c$ is differentiable, a best reply $\left(a_{iy}^{\star}\right)_{y\in\mathcal{Y}}$ in the subgame after entry satisfies the following conditions

$$c'(a_{iy}^{\star})=\left[\lambda u_{xy}\tilde{q}_y(1_y+\bar{a}_{yx})\right]\tag{4.8}$$

$$a_{iy}\in[0,1]\tag{4.9}$$

Agents decide that it is worth it to download the platform if given they will use it optimally, its use improves on their inclusive value, that is if and only if $V_i(x)\geq I(x)$. This condition boils down to $U_i(x)\geq I(x)$. Notice that if it is rational and utility-maximizing for an agent to enter the platform, the $RC$ constraint is satisfied. In fact, since $I(x)\geq 0$

$$U_i(x)\geq I(x)\implies U_i(x)\geq 0\ (RC)$$

Moreover, in a large enough economy, $I(x)>0$ and the $(RC)$ condition holds strictly. Summarizing, individuals $i$ of type $x$ will use the online dating platform if they belong to the set $K_x$

$$K_x=\left\{i\in\mathcal{I}:\ \exists a_i^{\star}\in BR^x(a_{-i}^{\star}),\ \lambda\sum_{y\in\mathcal{Y}}\tilde{q}_y u_{xy}\left(1+\bar{a}_{yx}^{\star}\right)\left(1+a_{iy}^{\star}\right)-c(a_{iy}^{\star})-\kappa_i\geq I(x)\right\}\tag{4.10}$$

The probability of an individual $i$ of type $x$ downloading the platform is given by

$$Pr(i \in K_x) = Pr\left\{\kappa_i \leq \lambda \sum_{y\in\mathcal{Y}} u_{xy}\left(1 + a^{\star}_{yx}\right)\left(1 + a^{\star}_{iy}\right)\tilde{q}_y - c(a^{\star}_{iy}) - I(x)\right\}$$
$$= F_{\kappa}\left(\lambda \sum_{y\in\mathcal{Y}} u_{xy}\left(1 + \bar{a}^{\star}_{yx}\right)\left(1 + a^{\star}_{iy}\right)\tilde{q}_y - c(a^{\star}_{iy}) - I(x)\right)$$

Let $k_{x_i} = \lambda \sum_{y\in\mathcal{Y}} u_{xy}\left(1 + \bar{a}^{\star}_{yx}\right)\left(1 + a^{\star}_{iy}\right)\tilde{p}_y - c(a^{\star}_{iy}) - I(x)$. If $\kappa_i \leq k_{x_i}$ then $i$ will download the platform. Notice that the distribution of types on the platform $\tilde{q}$ is an equilibrium object, as it is the result of individual choices of entering the online dating game. In particular, the distribution of types present on the online dating platform is such that $\tilde{p}_x = p_x F_{\kappa}(k_{x_i})$ and $\tilde{q}_y = q_y F_{\kappa}(k_{y_j})$ where $k_{x_i} = \sum_{y\in\mathcal{Y}} u_{xy}\left(1 + \bar{a}^{\star}_{yx}\right)\left(1 + a^{\star}_{iy}\right)\tilde{q}_y - c(a^{\star}_{iy}) - I\left(x\right)$ and $k_{y_j} = \sum_{y\in\mathcal{Y}} u_{xy}\left(1 + \bar{a}^{\star}_{yx}\right)\left(1 + a^{\star}_{iy}\right)\tilde{p}_x - c(a^{\star}_{iy}) - I\left(x\right)$

## 4.4. Monomorphic Equilibria

Let us now restrict attention to monomorphic equilibria, that is equilibria where all agents of the same type adopt the same strategy. In particular we focus on equilibria that are monomorphic in the subgame after entry. Formally, in a monomorphic equilibrium the following condition holds:

$$a_{iy} = a_{xy} \quad \forall i \in I \text{ with } x_i = x \tag{4.11}$$

When (4.11) holds the dimensionality of the strategy profiles decreases, and for every $i$ whose type is $x$ the strategy space is $S_x = \left\{(E, NE), \left(a_{xy}\right)_{y\in\mathcal{Y}}\right\}$. Similarly, the utility in the subgame after entry will be the same for every player $i$ of type $x$ and we can simply write

$$U_i(a_x, a_{-x}) = \left[\sum_{y\in Supp(\tilde{q})} \left(1 + a_{yx}\right)\left(1 + a_{xy}\right)\tilde{q}_y \exp\left\{u_{xy}\right\} - c(a_{xy})\right] - \kappa_i \tag{4.12}$$

Where $a_{yx}$ now replaces $\bar{a}_{yx}$ as the two are equal. Let us now define formally a monomorphic equilibrium in this setting.

**Definition 5.** A monomorphic equilibrium in the dating game $\Gamma = \left\langle \mathcal{X} \times \mathcal{Y}, \left(S_x \times S_y\right)_{x\in\mathcal{X}, y\in\mathcal{Y}}, \left(U_x(\cdot), U_y(\cdot)\right)_{x\in\mathcal{X}, y\in\mathcal{Y}}\right\rangle$ is

- A map $\Phi_x : \kappa_i \to \{E, NE\}$ for every $x \in \mathcal{X}$ and a map $\Phi_y : \kappa_i \to \{E, NE\}$ for every $y \in \mathcal{Y}$
- A map $a_x : \tilde{q} \to [0,1]^{|\mathcal{Y}|}$ for every $x \in \mathcal{X}$ and $a_y : \tilde{p} \to [0,1]^{|\mathcal{X}|}$ for every $y \in \mathcal{Y}$

Such that:

- The attention profiles described by the maps $a_x: \ \tilde{q} \to [0,1]^{|\mathcal{Y}|}$ and $a_y: \tilde{p} \to [0,1]^{|\mathcal{X}|}$ maximize the utility $U_i$ in (4.12), given the distribution of types on the platform $\tilde{p}$ and $\tilde{q}$ induced by the entry maps $\Phi_x$ and $\Phi_y$.
- The entry maps $\Phi_x$ and $\Phi_y$ maximize the expected utility of a player of type $x$ and $y$, through a decision of entry or not on the online dating platform

Looking only at monomorphic equilibria greatly simplifies the problem, and allows us to derive more intuition from the solution. In particular, imposing further conditions on the cost function we are able to investigate the nature of the equilibrium.

**Assumption 2**: The cost function is differentiable

**Assumption 3:** The cost function is separable, $c(\sum_{y\in\mathcal{Y}} a_{iy}) = \sum_{y\in\mathcal{Y}} c(a_{iy})$ .

If Assumptions 2 and 3 hold, the optimality conditions in the subgame after entry for a type $x$ are as follows

$$\begin{aligned} c'(a^\star_{xy}) &= (1+a_{yx})\tilde{q}_y \exp\{u_{xy}\} \\ a_{xy} &\in [0,1] \end{aligned}$$

And the set of types $x$ that enters into the meeting platform is $K_x$

$$K_x = \left\{ i \in \mathcal{I}: \ \left(1+a^\star_{yx}\right)\left(1+a^\star_{xy}\right)\tilde{q}_y \exp\{u_{xy}\} - c(a^\star_{xy}) \geq \kappa_i + I(x) \right\}$$

If the function $c$ is differentiable and if two types $x$ and $y$ are on the online dating platform, the optimal level of attention they will give each other is determined by the two by two system of equations

$$\begin{cases} c'(a^\star_{xy}) = (1+a_{yx})\tilde{q}_y \exp\{u_{xy}\} \\ c'(a^\star_{yx}) = (1+a_{xy})\tilde{p}_x \exp\{v_{yx}\} \end{cases}$$

**Proposition 3.** *If the cost function $c : [\underline{a}, \bar{a}] \to \mathbb{R}_+$ is continuously differentiable with continuous derivative $c'$, there exists a monomorphic solution to the population game $\Gamma$*

*Proof.* Assume $x \in A_{yx}$ for every $a_{xy}$ and that $y \in A_{xy}$ for every $a_{yx}$. Notice that if $c$ is differentiable the optimal level of attention that two types $x$ and $y$ will give each other is the solution to the following system of equations

$$\begin{cases} c'(a^\star_{xy}) = (1+a_{yx})\tilde{q}_y \exp\{u_{xy}\} \\ c'(a^\star_{yx}) = (1+a_{xy})\tilde{p}_x \exp\{v_{yx}\} \end{cases}$$

If $c'$ is continuous and $u_{xy}$ ,$v_{yx}$ are bounded, by Brower's fixed point theorem, the system has a solution $(a^\star_{xy}, a^\star_{yx})$.

□

**Theorem 1.** *Let the cost function $c : \mathbb{R} \to \mathbb{R}_+$ be continuously differentiable and strictly convex and let $c'$ be the first derivative of $c$. If the inverse function of $c'$ is Lipschitz continuous with Lipschitz constant $q < \exp\left\{-\frac{1}{2}(u_{xy} + v_{yx})\right\}$, there exists a unique monomorphic solution to the population game $\Gamma$.*

*Proof.* Assume $x \in A_{yx}$ for every $a_{xy}$ and that $y \in A_{xy}$ for every $a_{yx}$. If $c$ is differentiable the optimal level of attention that two types $x$ and $y$ will give each other is the solution to the following system of equations

$$\begin{cases} c'(a^{\star}_{xy}) = \tilde{q}_y \exp\{u_{xy}\}(1 + a_{yx}) \\ c'(a^{\star}_{yx}) = \tilde{p}_x \exp\{v_{yx}\}(1 + a_{xy}) \end{cases}$$

Since $c$ is strictly convex, $c'$ is strictly increasing, and therefore invertible. Let $f = (c')^{-1}$. Then the above system can be rewritten as

$$\begin{cases} a^{\star}_{xy} = f\left([\tilde{q}_y \exp\{u_{xy}\}(1 + a_{yx})]\right) \\ a^{\star}_{yx} = f\left([\tilde{p}_x \exp\{v_{yx}\}(1 + a_{xy})]\right) \end{cases}$$

The existence of a solution boils down to the existence of a solution to the equation $a_{xy} = f\left(\tilde{q}_y \cdot \exp\{u_{xy}\}\left(1 + f\left(\tilde{p}_x \exp\{u_{yx}\}(1 + a_{xy})\right)\right)\right)$. Or in other words the existence of a fixed point for the mapping $\tilde{f}(x) = f\left(\tilde{q}_y \cdot \exp\{u_{xy}\}\left(1 + f\left(\tilde{p}_x \cdot \exp\{v_{yx}\}(1 + x)\right)\right)\right)$. Notice that the function $\tilde{f}$ is a contraction mapping. In fact

$$\begin{aligned} \left|\tilde{f}(x_1) - \tilde{f}(x_2)\right| &= \left|f\left(\exp\{u_{xy}\}\left(\tilde{p}_y + f\left(\tilde{p}_x \exp\{v_{yx}\}(1 + x_1)\right)\right)\right) - f\left(\exp\{u_{xy}\}\left(\tilde{p}_y + f\left(\tilde{p}_x \exp\{u_{yx}\}(1 + x_2)\right)\right)\right)\right| \\ &\leq q\left|\exp\{u_{xy}\}\left(\tilde{p}_y + f\left(\tilde{p}_x \exp\{v_{yx}\}(1 + x_1)\right)\right) - \exp\{u_{xy}\}\left(\tilde{p}_y + f\left(\tilde{p}_x \exp\{v_{yx}\}(1 + x_2)\right)\right)\right| \\ &= q \cdot \exp\{u_{xy}\} \cdot \left|f\left(\tilde{p}_x \exp\{v_{yx}\}(1 + x_1)\right) - f\left(\exp\{v_{yx}\}(1 + x_2)\right)\right| \\ &\leq q^2 \cdot \exp\{u_{xy}\} \cdot \left|\exp\{v_{yx}\}(1 + x_1) - \exp\{v_{yx}\}(1 + x_2)\right| \\ &= q^2 \exp\{u_{xy} + v_{yx}\} \cdot |x_1 - x_2| \end{aligned}$$

Since we assumed $q < \exp\left\{-\frac{1}{2}(u_{xy} + v_{yx})\right\}$ , $\left|\tilde{f}(x_1) - \tilde{f}(x_2)\right| < |x_1 - x_2|$. Since $\tilde{f} : \mathbb{R} \to \mathbb{R}$ is a contraction mapping, by Banach contraction theorem it has a unique fixed point.

□

It becomes now easier to characterize the group of individuals of type $x$ that will be on the online dating platform. An individual $i$ will download the platform if they belong to the set $K_x$

$$K_x = \left\{ i \in \mathcal{I} : \sum_{y \in \mathcal{Y}} u_{xy} \left(1 + a^\star_{yx}\right) \left(1 + a^\star_{xy}\right) \tilde{p}_y - c(a^\star_{xy}) - I(x) \geq \kappa_i \right\} \tag{4.13}$$

From the optimality condition of entry $\Phi$, I can derive the distribution of types on the platform in equilibrium. The mass of agents that enter the platform, conditional of being of type $x \in \mathcal{X}$ is

$$Pr(i \in K_x) = \int_{x_i = x} \mathbb{I}\left\{\Phi_x = E\right\} dF_\kappa$$

Given the cdf of the random variable $\kappa$, the probability of an individual $i$ downloading the platform given they are of type $x$ is

$$\begin{aligned} Pr(i \in K_x) &= Pr \left\{ \kappa_i \leq \sum_{y \in \mathcal{Y}} u_{xy} \left(1 + a^\star_{yx}\right) \left(1 + a^\star_{xy}\right) \tilde{p}_y - c(a^\star_{xy}) - I(x) \right\} \\ &= F_\kappa \left( \sum_{y \in \mathcal{Y}} u_{xy} \left(1 + a^\star_{yx}\right) \left(1 + a^\star_{xy}\right) \tilde{p}_y - c(a^\star_{xy}) - I(x) \right) \end{aligned}$$

We know that $\Phi_x(\kappa_i) = E$ if and only if $\kappa_i \leq k_x$ where $k_x = \sum_{y \in Supp(\tilde{q})} u_{xy} \left(1 + a^\star_{yx}\right) \left(1 + a^\star_{xy}\right) \tilde{q}_y - c(a^\star_{xy}) - I(x)$. This implies that if $p_x$ is the fraction of type $x$ agents in the entire economy, the mass of type $x$ agents that are going to be using online dating is given by $p_x F_\kappa(k_x)$. This in turn implies that the mass of types $x$ on the platform is $\tilde{p}_x = \frac{p_x F_\kappa(k_x)}{\sum_{x' \in \mathcal{X}} p_{x'} F_\kappa(k_x)}$. Notice that the behaviour on the platform is type specific, while the decision to enter the platform is individual specific.

In the following part I will explore two special cases for the cost function, namely linear and quadratic, that allow me to pin down a closed form solution for the equilibrium.

### 4.4.1. *Linear Cost Function*

Assume the cost function is linear, formally let Assumption 4a hold

**Assumption 4a:** The cost function is linear with functional form $c(a) = c_1 a + c_2$, and $c_1 > 0$.

In this case the population game boils down to a classical model of strategic network formation where a link is formed if the marginal cost $c_1$ of forming it is less than the benefit of having it. More precisely, let $a \in [\underline{a}, \bar{a}]$ and Assumption 4a hold. Then the best reply correspondence for this game is

$$a = \begin{cases} \bar{a} & \text{if } a_{yx} > \frac{c_1 \cdot \exp\{-u_{xy}\}}{q_y} - 1 \\ [0, \bar{a}] & \text{if } a_{yx} = \frac{c_1 \cdot \exp\{-u_{xy}\}}{q_y} - 1 \\ 0 & \text{otherwise} \end{cases}$$

In particular, if $u_{xy}$ is above or below a certain threshold the decision to direct attention or not does not depend on the action of the opponent. That is

$$a^{\star} = \begin{cases} \bar{a} & \text{if } u_{xy} \geq \ln(c) - \ln(\tilde{q}_y) \\ 0 & \text{if } u_{xy} \leq \ln(c) - \ln(\tilde{q}_y) - \ln(1+\bar{a}) \end{cases}$$

In the interval of $u_{xy}$ $[\ln(c) - \ln(\tilde{q}_y) - \ln(1+\bar{a}), \ln(c) - \ln(\tilde{q}_y)]$ the best reply depends on the action of the opponent: if the opponent is seeking types $x$, it is an incentive for types $x$ to make an effort to seek types $y$.

4.4.2. *Quadratic Cost Function*

The second special case is the one of a quadratic cost function with form $c(a) = \frac{1}{2}c_1 a^2 + c_2 a + c_3$. In particular, by Theorem 1 if $c_1 > \lambda\left(\tilde{p}_x \tilde{q}_y\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2}\left(u+v\right)\right\}$ there exists a unique monomorphic solution, and we are able to obtain a closed form solution for it. For the next part let Assumption 4 hold.

**Assumption 4:** The cost function is increasing in the domain and quadratic.

Assume the cost function has form $c(a) = \frac{1}{2}ca^2$. Then the objective function for an agent who has purchased the meeting technology is

$$U(a_i, a_{-i}\,|x_i = x) = \sum_{y\in\mathcal{Y}} \lambda(1+a_{iy})(1+a_{yx})\tilde{q}_y \exp\{u_{xy}\} - \frac{1}{2}c\cdot a_{iy}^2 - \kappa_i$$

$$V(a_j, a_{-j}\,|y_j = y) = \sum_{y\in\mathcal{Y}} \lambda(1+a_{jx})(1+a_{xy})\tilde{p}_x \exp\{v_{yx}\} - \frac{1}{2}c\cdot a_{jx}^2 - \kappa_i$$

Considering the subgame after entry and focusing once again on the monomorphic equilibrium, we obtain the following system of equations that holds in equilibrium

$$\begin{cases} a_{xy}^{\star} & = \frac{\lambda}{c}(1+a_{yx}^{\star})\tilde{q}_y \exp\{u_{xy}\} \\ a_{yx}^{\star} & = \frac{\lambda}{c}(1+a_{xy}^{\star})\tilde{p}_x \exp\{v_{yx}\} \end{cases}$$

Solving the system we obtain the following closed form equilibrium

$$\begin{cases} a_{xy}^{\star} & = \frac{\lambda(c+\lambda\tilde{p}_x\exp\{v_{yx}\})}{c^2-\lambda^2\tilde{p}_x\tilde{q}_y\exp\{u_{xy}+v_{yx}\}} \cdot \tilde{q}_y \exp\{u_{xy}\} \\ a_{yx}^{\star} & = \frac{\lambda(c+\lambda\tilde{q}_y\exp\{u_{xy}\})}{c^2-\lambda^2\tilde{p}_x\tilde{q}_y\exp\{u_{xy}+v_{yx}\}} \cdot \tilde{p}_x \exp\{v_{yx}\} \end{cases}$$

Looking at the closed form solution for the equilibrium we can make a few observations that allows us to gain intuition into the nature of the population game equilibrium.

(1) The best reply functions are decreasing in $c$. The higher the cost coefficient $c$ the lower the optimal level of effort $a^{\star}$

(2) The best response of a type $x$ depends both on the utility $u_{xy}$ that the player herself would obtain from a match with another type $y$, and on the utility $v_{yx}$ that they would be able to provide to the other person.
(3) The optimal level of attention directed to types $y$ is positively dependent on exposure to the group on the platform $\tilde{p}_y$: there is more of an incentive to direct attention to larger groups on the platform or groups that are largely present on the platform.

### 4.4.3. *Comparative Statics*

Looking at the characteristics of the equilibrium in the model, before making any assumption on the characteristics of the cost function, we can conclude that there are a few patterns worth noting.

(1) People with lower inclusive values $I(x)$ are ceteris paribus more likely to download the platform. The low inclusive value might be caused by either one of two factors: a sparse network, where exposure to appealing potential partners is low, or a low utility derived from the average acquaintance in the network. In either case, the probability of someone downloading the platform is negatively affected by their inclusive value in the exogenous network.
(2) The mass of people of type $x$ on the platform is determined by $F_\kappa(k_x)$. In particular the following factors contribute to a higher mass of individuals of type $x$ on the platform: A higher average utility obtained from matching with other people on the platform; being sought after by desirable people who are also on the platform; a low inclusive value, due to a sparse network or a network of undesirable partners.
(3) The more people are on the platform, the more people will want to be on the platform. More people on the platform is reflected by $\tilde{p}$ having a larger support, contributing to $\sum_{y\in Supp(\tilde{p})} \tilde{p}_y \exp\{u_{xy}\}\left(1+a^\star_{yx}\right)$ being larger. The larger $\sum_{y\in\mathcal{Y}}\sum_{y\in Supp(\tilde{p})} \tilde{p}_y \exp\{u_{xy}\}\left(1+a^\star_{yx}\right)$ the larger $k_x$ and the larger $k_x$ the larger $F_\kappa(k_x)$ and therefore $\tilde{p}_x$.
(4) The more selective an agent is, the least likely they are to download the platform. An agent that is planning to look for several categories of people is more likely to have an incentive to download the platform. In other words if $a^\star_{xy}$ is positive for several $y$'s, then a type $x$ is more likely to have an incentive to download the platform. The more selective agents are, the more unlikely to download the platform, because downloading it to only look for a specific and very small subgroup of the population might not be worth the downloading costs. In other words, if the distribution of $u(x,\cdot)$ tends to be uniform across $y's$ then the agent is more likely to download the platform. On the other hand, if the distribution of $u(x,y)$ across $y's$ has very high variance agents are less likely to download the platform. This implies that when the logic is applied to mixed couples, people that have a smaller ethnic biases are more likely to be willing to use an online dating platform.

(5) The increase in probability of a match $\mu_{xy}$ through the online platform is non-trivial. Notice that the probability of a type $x$ and a type $y$ to match offline is given by

$$\mu_{xy} = \beta_{xy} p_x p_y \frac{\exp\{u_{xy} + v_{yx}\}}{I(x)\, I(y)}$$

On the other hand their probability to match online if they both are using the online dating platform is

$$\tilde{\mu}_{xy} = (1 + a_{xy})(1 + a_{yx})\, \tilde{p}_x \tilde{q}_y \frac{\exp\{u_{xy} + v_{yx}\}}{\tilde{I}(x)\tilde{I}(y)}$$

where $\tilde{p}_x = \frac{F_\kappa(k_x)p_x}{\sum_{x'\in\mathcal{X}} F_\kappa(k_x)p_x}$ , $\tilde{q}_y = \frac{F_\kappa(k_y)q_y}{\sum_{y'\in\mathcal{Y}} F_\kappa(k_y)q_y}$ and $k_x = \sum_{y\in Supp(\tilde{p})} u_{xy}\left(1 + a^\star_{yx}\right)\left(1 + a^\star_{xy}\right)\tilde{q}_y - c(a^\star_{xy}) - I(x)$. Also notice that if both $i$ and $j$ are on the platform it means that $\tilde{I}_i(x) > I(x)$ and $\tilde{I}_i(y) > I(y)$. The matching probability $\mu_{xy}$ of two individuals of type $x$ and $y$ is greater online than it is offline if and only if $\frac{\beta_{xy}\beta_{yx}p_x p_y}{I(x)I(y)} < \frac{(1+a_{xy})(1+a_{yx})\tilde{p}_x\tilde{q}_y}{\tilde{I}(x)\tilde{I}(y)}$ or rearranging and substituting for $\tilde{p}$, if and only if $\frac{\tilde{I}(x)\tilde{I}(y)}{I(x)I(y)} < \frac{F_\kappa(k_x)F_\kappa(k_y)\left(1+a^\star_{xy}\right)\left(1+a^\star_{yx}\right)}{\beta_{xy}\beta_{yx}}$. In other words for two people of type $x$ and $y$ respectively, their probability of matching is higher online than it is offline if the increase in exposure to each other that they obtain on the platform is greater than the relative increase in inclusive value that they obtain by using the platform. Overall, the probability of types $x$ and $y$ matching if they are online is $(1-\lambda)\mu_{xy} + \lambda\tilde{\mu}_{xy}$, the convex combination of their probability of matching offline and their probability of matching online.

In order to better understand how the model works and its implications, consider the following example.

**Example.** Suppose that there is only one type of individual in the set $\mathcal{I}$, that is $x_i = x$ for all $i \in \mathcal{I}$ and assume that the number of individuals in $\mathcal{I}$ is $n_x = 100$. Suppose there are two types of agents in the set $\mathcal{J}$, $\underline{y}, \bar{y} \in \mathcal{Y}$ and that $n_{\underline{y}} = n_{\bar{y}} = 50$. Let $U_{x\underline{y}} = 0$, $U_{x\bar{y}} = 2$, $V_{\underline{y}x} = 1$, $V_{\bar{y}x} = 2$ . Also assume that the cost function is $c(a) = \frac{3}{2}a^2$, $\kappa \sim U[90, 150]$ and assume for simplicity that $I(x) = I(\bar{y}) = I(\underline{y}) = 0$. Assume the matchmaker/platform proposes a fixed proportion $\alpha = \frac{1}{5}$ of the total number of participants in the platform, uniformly choosing among types.

Using backward induction to find the Nash equilibria in the subgame after entry it is easy to verify that $a^\star_{x\underline{y}} = 0$, $a^\star_{x\bar{y}} = \frac{4}{5}\lambda^\star_x + \frac{6}{5}\lambda^\star_{\bar{y}}$, $a^\star_{\underline{y}x} = \frac{1}{3}\lambda^\star_x$, $a^\star_{\bar{y}x} = \frac{6}{5}\lambda^\star_x + \frac{4}{5}\lambda^\star_{\bar{y}}$, where $\lambda^\star$ are the Poisson rates at which individual types are proposed to potential partners. Plugging the best replies into the payoff function we obtain the following

$$k_x = \frac{6}{25}\left(14\lambda_x^2 + 27\lambda_x\lambda_{\bar{y}} + 9\lambda_{\bar{y}}^2\right)$$
$$k_{\bar{y}} = \frac{6}{25}\left(9\lambda_x^2 + 27\lambda_x\lambda_{\bar{y}} + 14\lambda_{\bar{y}}^2\right)$$
$$k_{\underline{y}} = \lambda^\star_x\lambda^\star_{\underline{y}} + \frac{1}{6}\lambda_x^2$$

From these values we can determine $\lambda_x^\star, \lambda_{\bar{y}}^\star$ and $\lambda_{\underline{y}}^\star$ as fixed point solutions to the following equations

$$\lambda_x^\star = \frac{1}{5} F_\kappa(k_x) n_x \approx 2.32$$
$$\lambda_{\bar{y}}^\star = \frac{1}{5} F_\kappa(k_{\bar{y}}) n_{\bar{y}} \approx 3.74$$
$$\lambda_{\underline{y}}^\star = \frac{1}{4} F_\kappa(k_{\underline{y}}) n_{\underline{y}} \approx -59.09$$

Since the value of $\lambda_{\underline{y}}$ cannot be negative, it must be 0, or in other words the cost of entry is too high for low types. Therefore the platform will only have $x$'s and $\bar{y}$'s.

This implies optimal actions are

$$a_{x\underline{y}}^\star = 0$$
$$a_{x\bar{y}}^\star = 6.344$$
$$a_{\underline{y}x}^\star = 5.776$$
$$a_{\bar{y}x}^\star = 0.773$$

and

$$k_x = 104.5$$
$$k_{\bar{y}} = 114.8$$
$$k_{\underline{y}} = 2.7$$

From here we can calculate the expected utility of each type and distribution of types on the platform. For all $i$ and $j$, if $k_x < \kappa_i$ or $k_y < \kappa_j$, $I(x_i) = I(y_i) = 0$ because they do not enter the online dating game and we assumed the inclusive values in the exogenous network are 0. The distribution of types, we know that $\tilde{p}_x = \frac{F_\kappa(k_x)}{\sum_{x' \in \mathcal{X}} F_\kappa(k_{x'})}$. We have

$$F_\kappa(k_x) = 0.090$$
$$F_\kappa(k_{\bar{y}}) = 0.297$$
$$F_\kappa(k_{\underline{y}}) = 0$$

This means that about 9% of $x$ agents will be on the platform, 29.7% of $\bar{y}$ agents and 0% of $\underline{y}$ agents. Therefore the distribution of types on the platform is

$$\tilde{p}_x = 0.38$$
$$\tilde{p}_{\bar{y}} = 0.62$$
$$\tilde{p}_{\underline{y}} = 0$$

Whether online dating increases or decreases exposure to potential partners of a specific type depends on the shape of the network and the inclusive value offline. In particular if $\beta_{xy} n_y n_x <$

$\left(\lambda_x^\star + a_{xy}^\star\right)\left(\lambda_y^\star + a_{yx}^\star\right)$ then online dating increases exposure of types $x$ to types $y$. Otherwise it decreases it. In this particular case since we assume there are 0 connections on the exogenous network, the platform does increase exposure.

## 4.5. **Imperfect Directed Search**

So far we have assumed that when an agent downloads and uses an online dating platform they can specifically direct their efforts to meeting people of type $y$. However it might be the case that the platform does not allow for such a specific search, and only allows agents to direct search to subsets $\mathcal{Y}'$ of $\mathcal{Y}$. Formally suppose $\mathcal{E} = \{\mathcal{Y}_1, ..., \mathcal{Y}_n\}$ is a partition of the space of types and that players on the platform can direct their effort to meeting someone of any element of the partition, but not specifically to members of the partition. In other words their action profile must be an $\mathcal{E}$-measurable function. Let $a_{i\mathcal{Y}_j}$ be the effort made by individual $i$ to meeting individuals of the partition element $\mathcal{Y}_j$. The expected utility of downloading the platform and directing effort $a_{i\mathcal{Y}_j}$ to meet people of group $\mathcal{Y}_j$ is

$$U_i(a_i, a_{-i}) = (\tilde{p}_x + a_{i\mathcal{Y}'}) \sum_{y\in\mathcal{Y}'} U_{xy}\left(\tilde{p}_{y|\mathcal{Y}'} + \bar{a}_{y\mathcal{X}}\right)\mathbb{I}\{x \in \mathcal{X}\} - c(a_{i\mathcal{Y}}) - \kappa_i \tag{4.14}$$

where $\tilde{p}_{y|\mathcal{Y}'}$ is the conditional probability of $y$ given $\mathcal{Y}'$, $\bar{a}_{y\mathcal{X}}$ is the average attention that an individual of type $y$ will give to the whole category $\mathcal{X}$ to which $x$ belongs. We can reformulate the utility above as

$$U_i(a_i, a_{-i}) = (\tilde{p}_x + a_{i\mathcal{Y}'})\,\mathbb{E}\left[U_{xy}\,|y \in \mathcal{Y}'\right] - c(a_{i\mathcal{Y}}) - \kappa_i \tag{4.15}$$

where the expected utility $\mathbb{E}_{\tilde{\tilde{p}}}$ is taken with respect to the probability measure $\tilde{\tilde{p}}$ defined as $\tilde{\tilde{p}}_y = (F_\kappa(k_y) + a_{y\mathcal{X}})$. In other words the probability with respect $\mathbb{E}$ is taken is also an equilibrium object. The best reply for individuals of type $x$ satisfies

$$c'(a_{x\mathcal{Y}_j}) = \sum_{y\in\mathcal{Y}_j} U_{xy}\left(\tilde{p}_{y|\mathcal{Y}_j} + \bar{a}_{y\mathcal{X}_i}\right)$$

Focusing once again on the monomorphic equilibria, we impose $\bar{a}_{y\mathcal{X}} = a_{y\mathcal{X}}$. The equilibrium action profile determining the effort that groups $\mathcal{X}_i$ and $\mathcal{Y}_j$ will make to try and meet each other is determined by the following system of $|\mathcal{X}_i| \times |\mathcal{Y}_j|$ equations

$$\begin{cases} c'(a_{x\mathcal{Y}_j}) = \sum_{y\in\mathcal{Y}'} U_{xy}\left(\tilde{p}_{y|\mathcal{Y}_j} + a_{y\mathcal{X}}\right) & \forall x \in \mathcal{X}' \\ c'(a_{y\mathcal{X}_i}) = \sum_{x\in\mathcal{X}'} V_{yx}\left(\tilde{p}_{x|\mathcal{X}_i} + a_{x\mathcal{Y}}\right) & \forall y \in \mathcal{Y}' \end{cases}$$

Notice that when compared to the case of perfect directed search, the optimal action profiles depend on the category the individual is choosing to give attention to, but also and at the same time on the behavior of all individuals belonging to the same category as him or her, creating some sort of "reputation" effect.

Consider the value of $k_x^{\star\star}$ in the case of imperfect directed search.

$$k_x = \sum_{j=1}^{n}\sum_{y\in\mathcal{Y}_j} U_{xy}\left(\tilde{p}_{y|\mathcal{Y}_j} + a_{yx}\right)\left(\tilde{p}_{x|\mathcal{X}_i} + a_{xy}\right) - c(a_{x\mathcal{Y}_j}) - I(x)$$

The expected surplus of the game depends crucially on the grouping or partition. For types that are grouped together with lower types - or types that provide lower utility to the counterpart - the surplus decreases. Vice versa for types that are grouped together with individuals that provide a higher surplus to the counterpart.

## 5. Applications

5.1. **Interracial Marriages.** Let there be $K$ different racial groups $\{1,...,K\}$ and let $\{A_k\}_{k=1}^{K}$ be a finite racial partition of the population. Let the type of each individual $i$ be a vector $(k,x)$ where $k$ indicates $i$'s ethnic group and $x$ is the vector of other observables as defined in the previous sections. Let $p_{kx}$ indicate the mass of men of group $k$ and type $x$ and $q_{ky}$ the mass of women of group $k$ and type $y$. Let the characteristics of the network be dictated by ethnicity. In particular let $\beta_{kk'}$ indicate the coefficient of exposure of agents of race $k$ to agents of race $k'$, as formalized in Section 3. Formally,

$$\delta_{kk'} = \sum_{y\in\mathcal{Y}} \delta_{kx,k'y}$$

Since we have assumed that the shape of the network is only determined by race $k$, for every $x, x' \in \mathcal{X}$ , for every $y \in \mathcal{Y}$ and for every $k, k'$ we have

$$\delta_{kx,k'y} = \delta_{kx',k'y}$$

and

$$\delta_{kx,k'y} = \delta_{k,k'} \cdot q_y$$

where $q_y$ is the frequency of agents $y$ in the economy. This allows us to define $\delta_{k,k'y} = \delta_{kx,k'y}$ and to simplifies the above expression of exposure to be

$$\delta_{kk'} = \sum_{y \in \mathcal{Y}} \delta_{k,k'} \cdot q_y$$

From $\delta_{kk'}$ we define the coefficient of exposure the Radon-Nykodim derivative $\beta_{kk'}$ defined as

$$\frac{\delta_{kk'}}{\sum_{k''=1}^{K} \delta_{kk''}} = \beta_{kk'} p_k q_{k'}$$

The left-hand side represents the fraction of connections of anyone in group $k$ with anyone in group $k'$. We will further decompose $\beta \cdot p \cdot q$ into that the probability for a woman of group $k$ of knowing any man of group $k'$ is $\beta^f_{kk'} p_{k'}$ and we will sometimes refer to $\beta^f_{kk'}$ as the degree of exposure of a woman of group $k$ to a man of group $k'$. Similarly the probability for a man from $k'$ of knowing a woman of group $k$ will be $\beta^m_{k'k} q_k$. We will allow the degrees of exposure of men and women $\beta^m_{kk'}$ and $\beta^f_{kk'}$ to be different. Again, we keep the assumption from the previous section that two people can match only if they know each other.

Notice that we need to impose some natural bounds on this degree of exposure, and we will have that $\beta^m_{kk'} \in \left[0, \frac{1}{q_{k'}}\right]$ and $\beta^f_{kk'} \in \left[0, \frac{1}{p_{k'}}\right]$ , where a degree of exposure of 0 means that the two clusters $k$ and $k'$ are completely disconnected, so that no one in $k$ knows anyone in $k'$. On the other hand $\beta^f_{kk'}$ cannot be above $\frac{1}{p_{k'}}$ because an agent cannot know more people than there are in the economy. So when $\beta^f_{kk'}$ is equal to $\frac{1}{p_{k'}}$ it means that people from $k$ know everyone from $k'$. In this sense $\beta$ can be interpreted as a Radon-Nykodim derivative. To ease notation we will denote $\beta_{kk'} = \beta^m_{kk'} \beta^f_{k'k}$ . We denote with $\beta^m$ the $K \times K$ matrix

$$\begin{bmatrix} \beta^m_{11} & \beta^m_{12} & \dots & \beta^m_{1K} \\ \beta^m_{21} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \beta^m_{K1} & \dots & \dots & \beta^m_{KK} \end{bmatrix}$$

and similarly $\beta^f_{kk'}$. The probability that $i$ of type $k, x$ and $j$ of type $k', y$ know each other is

$$Pr\left\{ N_{ij} = 1 \,|x_i = (k, x), y_j = (k', y) \right\} = \beta_{kk'} p_{kx} q_{k'y}$$

In order to capture the ethnic clustering of the network we can assume that $\beta_{kk} > \beta_{kk'}$ for all $k' \neq k$. In terms of the utility that two agents obtain from matching we will model a heterogeneous preference towards ethnic homogamy, that is, agents in general prefer, other things being equal, to marry someone of their own race and educational group. Formally, let $i \in A_k$ and $j \in A_{k'}$. Utility from a match is

$$U_{ij} = u(x_i, y_j) - b_{kk'} + \varepsilon_{ij}$$
$$V_{ji} = v(x_i, y_j) - b_{k'k} + \eta_{ji}$$

We can interpret $b_{kk'}$ as the social cost sustained by someone from ethnic group $k$ marrying someone from ethnic group $k'$. Assume there is no cost in marrying someone of one's own race, that is $b_{kk} = 0$ for all $k$. As usual

$$\varepsilon_{ij} \sim Gumbel(0, Var)\ \forall i, j$$
$$\eta_{ji} \sim Gumbel(0, Var)\ \forall i, j$$

Agents' inclusive value offline is determined both by their type and the ethnic group they belong to, as the ethnic group affects both their network of acquaintances and their preferences

$$I(k, x) = \sum_{k'=1}^{K} \beta_{kk'} p_{k'y} \exp\{u_{xy} - b_{kk'}\}$$

In the exogenous network, the probability of a marriage of types $x$ and $y$ belonging to the same race is

$$\mu_{xy} = \beta_{kk} p_{kx} q_{ky} \frac{\exp\{u_{xy} + v_{yx}\}}{I(k, x) I(k, y)}$$

While the probability of a marriage of types $x$ and $y$ belonging to different races is

$$\mu_{xy} = \beta_{kk'} p_{kx} q_{ky} \frac{\exp\{u_{xy} + v_{yx} - b_{kk'} - b_{k'k}\}}{I(k, x) I(k', y)}$$

We can predict the probability of an interracial marriage being lower than the probability of a same race marriage, other things being equal, due to two factors: the clustering of the network $\beta_{kk'} < \beta_{kk}$, and the presence of a social cost $b_{kk'} + b_{k'k}$ of crossing an ethnic boundary.

5.1.1. *Heterogeneous Bias.* Now suppose that agents differ in their perceived social cost of crossing a boundary. In particular let utilities be now described by

$$U_{ij} = u(x_i, y_j) - b^i_{kk'} + \varepsilon_{ij}$$
$$V_{ji} = v(x_i, y_j) - b^j_{k'k} + \eta_{ji}$$

with $b^i_{kk'} \sim \mathcal{N}(-b_{kk'}, \sigma_b)$ and $b^j_{k'k} \sim F(\theta_{kk'})$ for all $k \neq k'$. Now the probability of a match between $i$ and $j$ when they belong to different races depends on their subjective cost of crossing a racial boundary. In particular

$$\mu_{ij} = \beta_{kk} p_{kx} q_{ky} \frac{\exp\left\{u_{xy} + v_{yx} - b^i_{kk'} - b^j_{k'k}\right\}}{I^i(k,x) I^i(k,y)}$$

and the predicted share of couples $x$ and $y$ is

$$\begin{aligned}\mu_{xy} &= \beta_{kk} p_{kx} q_{ky} \int \frac{\exp\left\{u_{xy} + v_{yx} - b^i_{kk'} - b^j_{k'k}\right\}}{I^i(k,x) I^i(k,y)} dF(\theta) \\ &= \beta_{kk} p_{kx} q_{ky} \exp\{u_{xy} + v_{yx}\} \int \frac{\exp\left\{-b^i_{kk'} - b^j_{k'k}\right\}}{I^i(k,x) I^i(k,y)} dF(\theta)\end{aligned}$$

5.2. **Ethnic and Educational boundaries.** Let us now add a layer to the model, by adding education as a factor impacting both the network and individual preferences. Let again $K$ be the number of different ethnic groups in the economy, and let $L$ be the number of different educational groups. Let $\{A_k\}_{k=1}^K$ be a finite racial partition of $A$ and $\{E_l\}_{l=1}^L$ is an educational partition of $A$. We can in general also introduce a gender partition $\{G_l\}_{l=1}^L$ of $A$, but for the sake of exposition in this paper I will assume that the gender partition is only composed of two groups, proposers and receivers of a proposal, which, again for the sake of the presentation, I will sometimes refer to as men and women, although it would not make a difference if it was otherwise. To simplify notation, I will sometimes write $i \in (k,l)$ in lieu of $i \in A_k \cap E_l$.

We will denote with $m$ the number of women and $n$ the number of men; in particular we will use the notation $n_{kl}$ to denote the number of men from group $k$ and educational attainment $l$, and similarly for women we will use $m_{kl}$. When we use the notation $n_k$ or $m_k$ we want to indicate the cardinality of the set of men and women, respectively, belonging to the ethnic group $k$, formally $n_k = |A_k \cap I|$ and $m_k = |A_k \cap J|$. Moreover we will indicate with $M$ and $N$ respectively the total number of women and men respectively in the economy. Similarly we will denote $p_{kl}$ and $q_{kl}$ the relative frequency of males and females respectively of type $(k,l)$, formally $p_{kl} = \frac{n_{kl}}{N}$ and $q_{kl} = \frac{m_{kl}}{M}$. Each individual $i$ in the economy is also characterized by a type $x_i$ for men and $y_i$ for women that can be any discrete or continuous variable indicating characteristics other than ethnic group and educational attainment, like wealth, income, political affiliation, appearance, weight and so on. Each individual in the economy forms a node in the network and is connected to other individuals through an arc in the network; two people are connected in the network $\mathcal{N}$ if they know each other; in particular we will assume that each agent in the economy is more well-connected to other nodes that are similar to her both on an ethnic or educational level. In this sense the network presents 2-dimensional clusters, each of which is represented by an element of the partition described above. We introduce now the variables $\beta(t)$ and $\gamma(t)$ to describe the degree of connectedness of the clusters in the network at a given time $t$. We will then attach a similar interpretation to $\gamma_{ll'}$ which will indicate the degree of exposure of people with educational attainment $l$ to people with educational attainment $l'$. Thus, we will model the probability of $i \in (k,l)$ and $j \in (k',l')$ being in each other's set of acquaintances as

$$Pr\left\{N_{ij} = 1 \,|x_i = (k,l,x), y_j = (k',l',y)\right\} = \beta_{kk'}\gamma_{ll'}p_{klx}q_{k'l'y}$$

The probability of any two types $x$ and $y$ knowing each other, given they belong to $(k,l)$ and $(k',l')$ respectively is

$$\begin{aligned} Pr\left\{N_{ij} = 1 \,|x_i = (k,l,\cdot), y_j = (k',l',\cdot)\right\} &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \beta_{kk'}\gamma_{ll'}p_{klx}q_{k'l'y} \\ &= \beta_{kk'}\gamma_{ll'}p_{kl}q_{k'l'} \end{aligned}$$

We will allow the shape of the network to change as a result of social interaction, therefore we will denote $\beta(t)$ and $\gamma(t)$ when necessary. Finally, in order to capture the clustering of the network we can impose

$$\beta_{kk} \geq \beta_{kk'} \quad \forall k' \in \{1,...,K\}$$

$$\gamma_{ll} \geq \gamma_{ll'} \quad \forall l' \in \{1,...,L\}$$

The degree of exposure $\beta_{kk'}$ and $\gamma_{ll'}$ can be further decomposed into the respective degrees of exposure of men and women, $\beta_{kk'} = \beta^m_{kk'}\beta^f_{k'k}$ and $\gamma_{ll'} = \gamma^m_{ll'} \cdot \gamma^f_{l'l}$ so that the probability that $i \in (k,l)$ and $j \in (k',l')$ know each other can be written as

$$Pr\left\{N_{ij} = 1 \,|x_i = (k,l,\cdot), y_j = (k',l',\cdot)\right\} = \left(\beta^m_{kk'}\beta^f_{k'k}\right)\left(\gamma^m_{ll'}\gamma^f_{l'l}\right)p_{kl}q_{k'l'}$$

This notation will make the analysis of the dating game easier to formalize.

Utility is again non transferable and preferences over potential partners are described by a systematic part, given by the type of the other person, and a random shock that we will sometimes refer to as a chemistry shock. Moreover we assume people have a preference towards homogamy, that is, they prefer, other things being equal, to marry someone of their own race and educational group. Formally, let $i \in (k,l)$ and $j \in (k',l')$

$$\begin{aligned} U_{ij} &= u(x_i,y_j) - b_{kk'}\mathbb{I}\left\{k \neq k'\right\} - d_{ll'}\mathbb{I}\left\{l \neq l'\right\} + \varepsilon_{ij} \\ V_{ji} &= v(y_j,x_i) - b_{k'k}\mathbb{I}\left\{k \neq k'\right\} - d_{l'l}\mathbb{I}\left\{l \neq l'\right\} + \eta_{ji} \end{aligned}$$

Where $b_{kk'}$ is the social cost of someone from ethnic group $k$ to marry someone from ethnic group $k'$, and $d_{ll'}$ is the social cost for someone with education $l$ to marry someone with education $l'$. We allow both $b$ and $d$ to be either positive or negative. We will assume that the functions $u$ and $v$ are bounded above. As usual

$$\varepsilon_{ij} \sim Gumbel(0, Var)\ \forall i, j$$
$$\eta_{ji} \sim Gumbel(0, Var)\ \forall i, j$$

5.3. **The Dating Game: Strategic Network Formation.** Suppose now agents are given the option of playing the game described in Section 3. Agents have now the possibility to purchase access to a dating app that allows them to take an action $a^i \in \mathbb{R}^K$ to increase their exposure to potential partners of a certain race. I model the game in a way that allows agents to filter for race but not for education or other characteristics of the potential partner, although the model can be easily extended to allow for a more or less coarse search, as described in the general model in section 3. Let Assumption 1 hold, the technology is type-neutral. In this context, it implies that when downloading the platform the probability of bumping into a potential partner of race $k$, educational level $l$ and type $y$ is given by $\tilde{p}_{kly}$, where $\tilde{p}_{kly}$ is the fraction of such agents that are present on the platform. Therefore, the ex-ante degrees of exposure on the platform, $\beta$ and $\gamma$ are assumed to be equal to 1 for all agents of all types. This is equivalent to Assumption 3 of the previous section. Agents can increase or decrease their individual exposure to agents of race $k$ to $(1 + a_{ik})$ by exerting effort $a_{ik}$. Notice that this might cause every individual $i$ to form a different distribution of acquaintances. It is assumed that if individual $i \in k'$ exerts effort $a_{ik}$ to meet potential partners of group $k$ and individuals in group $k$ exert effort $a_{jk'}$ to meet potential partners of $k'$ the resulting exposure to one another will be given by

$$(1 + a_{ik})(1 + \frac{1}{|A_k|} \sum_{j \in A_k} a_{jk'})q_{k'}$$

In particular, in a monomorphic equilibrium $a_i = a_{x_i}$ for all $i \in (k, l)$ and $a_j = a_{y_j}$ for all $j \in (k', l')$. Therefore in a monomorphic equilibrium the resulting exposure in the game of types $(k, l, x)$ to types $(k', l', y)$ is

$$(1 + a_{klx,k'})(1 + \sum_{l',y} \frac{n_{k'l'y}}{N} a_{k'l'y,k})$$

If agents decide to play the game their total exposure to a certain race $k$ is given by the convex combination of their exposure offline and their exposure online. Let $\tilde{\beta}_{kk'}$ be the exposure of group $k$ to group $k'$ if they decided to enter the online dating game.

$$\tilde{\beta}_{kk'} = (1 - \lambda)\beta_{kk'} + \lambda(1 + a_{kk'})(1 + a_{k'k})$$

The objective for the agent is to maximize her inclusive value. Intuitively the agent's objective is to increase the number of acquaintances of ethnic and educational categories of people with whom they will obtain a high utility from a match.

Let $i \in (k,l)$ belong to $\mathcal{I}$. Agent $i$'s inclusive value without the meeting technology is

$$I\left[i\,|x_i = x, k, l\right] = \sum_{k'}\sum_{l'} \beta_{kk'}\gamma_{ll'}q_{k'l'}\mathbb{E}\left[\exp\left\{u(x_i,y_j) - b_{kk'}\mathbb{I}\left\{k \neq k'\right\} - d_{ll'}\mathbb{I}\left\{l \neq l'\right\}\right\}|j \in (k',l')\right]$$

Notice at this stage agents are abstracting from taking into account whether potential partners are going to be willing to match with them or not. Although this is a simplifying assumption to make the problem tractable, we will see that no important intuition is lost, and the equilibrium best reply for each agent will take into account the utility that their potential partner would derive by matching with them. If agents download the online dating platform, they can act to maximize their inclusive value online. The resulting inclusive value will be a convex combination of their inclusive value online and their inclusive value offline. Introducing the possibility for agents to exert a certain effort to modify their distribution of acquaintances, the inclusive value online depends on actions in the following way

$$I\left[i;a^i\right] = \sum_{k'}\sum_{l'}\left(1+a^i_{k'}\right)\frac{m_{k'l'}}{M}\mathbb{E}\left[\left(1+a^j_k\right)\exp\left\{u(x_i,y_j) - b_{kk'}\mathbb{I}\left\{k \neq k'\right\} - d_{ll'}\mathbb{I}\left\{l \neq l'\right\}\right\}|j \in \left(k',l'\right)\right]$$

The search action has a cost

$$c(a^i) = \sum_{k=1}^{K} c(a^i_k)$$

where $a^i = \left[a^i_1, .., a^i_K\right]$. Notice that the total effect of taking an action is also determined by the action of agents in the other set, and their type $y$. Notice now that $a^i$ has some natural bounds: in fact $a^i_{k'}$ cannot be below $-1$, because total exposure to another group cannot be less than 0. Similarly, and $a^i_{k'}$ cannot be above $\frac{1}{\tilde{p}_{k'}}$ because $i$ cannot know more people than there are in set $k'$. In other words, $a^i$ needs to be constrained in the closed interval

$$a^i \in \times_{k' \in K}\left[-1,\ \frac{1}{\tilde{p}_{k'}}\right]$$

Summarizing the objective function of agent $i$ is

$$\max_{a^i}(1-\lambda)I[i] + \lambda I[i,a^i] - c(a^i)$$
$$s.t.\ -1 \leq a^i_{k'} \leq \frac{1}{\tilde{p}_{k'}}$$

Now let's estimate the optimal action $a^i$ when it is an interior point. The objective is to maximize

$$\sum_{k'} e^{-b_{kk'}\mathbb{I}\{k\neq k'\}}\frac{n_{k'}}{N}\left(1+a^i_{k'}\right)\mathbb{E}\left[\left(1+a^j_k\right)\left\{u(x_i,y_j) - d_{ll'}\mathbb{I}\left\{l \neq l'\right\}\right\}|j \in \left(k',l'\right)\right] - \sum_{k'} c(a^i_{k'})$$

Taking first order conditions we have

$$c'(a_k^i) = e^{-b_{kk'}\mathbb{I}\{k\neq k'\}}\frac{n_k}{N}\mathbb{E}\left[\left(1+a_k^j\right)\{u(x_i,y_j) - d_{ll'}\mathbb{I}\{l\neq l'\}\}\,|j\in(k',l')\right]$$

The first observations that we can make are:

(1) A lower bias towards other ethnic groups will cause a higher effort to meet them. This points to the segregation effect. Other things being equal, there is a higher incentive to look for one's own ethnic group, since this will imply not paying the social cost of crossing a racial boundary.
(2) Higher optimal effort to meet bigger populations - high values of $n_k$.
(3) Race will be utilized by users as a proxy for education. Indeed, $i$ has an incentive to exert more effort in equilibrium to meet a population that has a similar educational level to his own. This is reflected in the fact that a population $k'$ with a similar educational level to $i$'s will provide $i$ with a higher value of the expected utility $\mathbb{E}\left[\left(1+a_k^j\right)\{u(x_i,y_j) - d_{ll'}\mathbb{I}\{l\neq l'\}\}\,|j\in(k',l')\right]$ because he has to pay the cost $d_{ll'}$ will lower frequency.
(4) The higher the expected surplus from the match, the higher the effort $i$ is going to put in
(5) Low cost of search will incentivize a more intense search.

**Example 6.** Consider a very simple example where systematic utilities are flat $u(x,y) = u$ and $v(y,x) = v$, everyone has the same educational achievement, and there are two ethnic groups, 1 and 2. Crossing an ethnic boundary and marrying with someone of the other group has a social cost equal to $b$ for both groups and both sides of the market. Assume that the cost function $c(a)$ is quadratic and separable and in particular assume $c(a) = \frac{1}{2}ca_1^2 + \frac{1}{2}ca_2^2$. The best replies are summarized by

$$\begin{aligned}
a_{11}^m &= \frac{1}{c}q_1(1+a_{11}^f)e^u \\
a_{12}^m &= \frac{1}{c}q_2(1+a_{21}^f)e^{u-b} \\
a_{21}^f &= \frac{1}{c}p_1(1+a_{12}^m)e^{v-b} \\
a_{22}^f &= \frac{1}{c}p_2(1+a_{22}^m)e^{v}
\end{aligned}$$

And the solution to the game is

$$
\begin{aligned}
a_{11}^{m\star} &= q_1 e^{u} \cdot \frac{c + p_1 e^{v}}{c^2 - p_1 q_1 e^{u+v}} \\
a_{12}^{m\star} &= q_2 e^{u-b} \frac{c + p_1 e^{v-b}}{c^2 - p_1 q_2 e^{u+v-2b}} \\
a_{21}^{f\star} &= p_1 e^{v-b} \frac{c + q_2 e^{u-b}}{c^2 - p_1 q_2 e^{u+v-2b}} \\
a_{22}^{f\star} &= p_2 e^{v} \frac{c + q_2 e^{u}}{c^2 - p_1 q_2 e^{u+v}}
\end{aligned}
$$

Notice how, other things being equal the effort made to meet someone of the other group is made lower by the presence of the bias $b$.

The resulting exposure to different ethnic groups after playing the game is given by $(1+a_{kk'})(1+a_{k'k})\tilde{p}_k$ . Whether the game increases or decreases exposure to different groups of agents depends jointly on the bias towards a different group and the current clustering of the network. The game results in overall higher exposure when $(1 + a_{kk'})(1 + a_{k'k})\tilde{p}_k\tilde{q}_{k'} > \beta_{kk'} p_k q_k$. To summarize, a general picture of the effect of online dating on exposure and integration is given in the following table.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Network Clustering</td></tr>
<tr><td>Low</td><td>High</td></tr>
<tr><td rowspan="2">Bias</td><td>Low</td><td>No<br>Ambiguous total effect</td><td>Yes<br>Integration Effect $>$ Directed Search</td></tr>
<tr><td>High</td><td>No<br>Integration Effect $<$ Directed Search</td><td>Ambiguous total effect</td></tr>
</table>

People can only match with people they know. At each $t$ the people from cohort $t$ match following a Gale and Shapley algorithm. People are divided into proposers and receivers of the proposal. For a proposer $i \in (k, l)$, a strategy $s^i = \{y_1, ..., y_n, 0\}$ in the matching game is an ordered list of potential partners to propose to among the ones he's connected to in the network. The last element of the list is always a 0, meaning remaining unmatched.

For a receiver $j \in (k', l')$, a strategy is a contingent plan indicating whom to retain from the ones that proposed to her at a given round.

At each moment in time, an individual $i$ considers the set of people he is connected to in the network, meaning the set of people he knows. This set will be composed of $\frac{\beta_{kk}\frac{n_k}{N}}{\sum_{k'}\beta_{kk'}\frac{n_k}{N}}$ men from their own group and $\frac{\beta_{kk'}\frac{n_{k'}}{N}}{\sum_{k'}\beta_{kk'}\frac{n_k}{N}}$ people from other groups $k'$.

A matching system $\mu$ is called stable if for all $i$ and $j$ such that $\mu_{ij} > 0$,
$U_{i\mu(i)} + \varepsilon_{i\mu(i)} \geq U_{ij} + \varepsilon_{ij}$ for all $j \in \mathbb{Y}(i)$ and $V_{j\mu(j)} + \eta_{j\mu(j)} \geq V_{ji} + \eta_{ji}$ for all other $i \in \mathbb{X}(j)$. Or in other words, if $U_{ij} + \varepsilon_{ij} > U_{i\mu(i)} + \varepsilon_{i\mu(i)}$ it must be $V_{ji} + \eta_{ji} < V_{j\mu(j)} + \eta_{j\mu(j)}$ and vice versa, ruling out the existence of blocking pairs.

**Proposition 4.** *The probability of $i \in (k, l)$ and $j \in (k', l')$ matching in the exogenous network is given by*

$$\mu_{ij} = \beta_{kk'}\gamma_{ll'}p_{kly}q_{k'l'y}\frac{\exp\left\{u(x_i,y_j)+v(y_j,x_i)-B_{kk'}\mathbb{I}\left\{k\neq k'\right\}-D_{ll'}\mathbb{I}\left\{l\neq l'\right\}\right\}}{I[i]I[j]}$$

*where* $I\left[i\right] = \sum_{k'}\sum_{l'}\beta_{kk'}\gamma_{ll'}\exp\left\{-b_{kk'}\mathbb{I}\left\{k \neq k'\right\} - d_{ll'}\mathbb{I}\left\{l \neq l'\right\}\right\}\sum_{j\in(k',l')}\exp\left\{u(x_i, y_j)\right\}$, $B_{kk'} = b_{kk'} + b_{k'k}$ *and* $D_{ll'} = d_{ll'} + d_{l'l}$

Notice that if groups $k$ and $k'$ are educationally similar, that is the value $\gamma_{ll} \cdot p_{k'l}$ is large compared to others, there is a higher chance of interracial marriages between the two groups. Compare now the probabilities of different matches.

– Same ethnicity and educational level

$$\mu_{ij} = \beta_{kk}\gamma_{ll}\frac{n(k,l)}{N}\frac{m(k,l)}{M}\cdot\frac{\exp\left\{u(x_i,y_j)+v(y_j,x_i)\right\}}{I_iI_j}$$

– Same ethnicity but different educational level

$$\mu_{ij} = \beta_{kk}\gamma_{ll'}\frac{n(k,l)}{N}\frac{m(k,l')}{M}\cdot\frac{\exp\left\{u(x_i,y_j)+v(y_j,x_i)-D_{ll'}\right\}}{I_iI_j}$$

– Different ethnicity but same educational level

$$\mu_{ij} = \beta_{kk'}\gamma_{ll}\frac{n(k,l)}{N}\frac{m(k',l)}{M}\cdot\frac{\exp\left\{u(x_i,y_j)+v(y_j,x_i)-B_{kk'}\right\}}{I_iI_j}$$

– Different ethnicity and different educational level

$$\mu_{ij} = \beta_{kk'}\gamma_{ll'}\frac{n(k,l)}{N}\frac{m(k',l')}{M}\cdot\frac{\exp\left\{u(x_i,y_j)+v(y_j,x_i)-B_{kk'}-D_{ll'}\right\}}{I_iI_j}$$

Notice how the probability of a match between people belonging to different clusters is lower, due to two reasons: the first one is the bias, or social cost, of crossing a racial or an educational boundary. Notice that these costs $b$ and $d$ can in principle be positive or negative. The second reason is the network effect: people belonging to different clusters have a lower probability of matching each other. However for instance two agents that belong to different racial clusters but same educational cluster can compensate the crossing of a racial boundary with the non-crossing of an educational one - not to mention, this increases their probability of knowing each other.

**Proposition 5.** *The expected number of interracial marriages between $i \in k$ and $j \in k'$ in the exogenous network is*

$$\Lambda_{kk'} = \beta_{kk'} e^{-B_{kk'}\mathbb{I}\{k \neq k'\}} \sum_{l'} \gamma_{ll'} \frac{m_{k'l'}}{M} \frac{n_{kl}}{N} \mathbb{E}\left[ e^{-D_{ll'}\mathbb{I}\{l \neq l'\}} \cdot \frac{\exp\{u(x_i, y_j) + v(y_j, x_i)\}}{I[j]I[i]} \,|j \in (k', l') \right]$$

*Proof.*

To see this

$$\begin{aligned}
\Lambda_{kk'} &= \sum_{i \in k} Pr\left\{\mu_{ij} > 0 \,|j \in k'\right\} \\
&= \sum_{i \in k} \beta_{kk'} \exp\left\{-B_{kk'}\mathbb{I}\left\{k \neq k'\right\}\right\} \sum_{l'} \delta_{ll'} \frac{m_{k'l'}}{M} \frac{1}{I[i]} \mathbb{E}\left[ \frac{\exp\left\{u(x_i, y_j) + v(y_j, x_i) - D_{ll'}\mathbb{I}\{l \neq l'\}\right\}}{I[j]} \,|j \in (k', l') \right] \\
&= \beta_{kk'} \exp\left\{-B_{kk'}\mathbb{I}\left\{k \neq k'\right\}\right\} \sum_{l'} \delta_{ll'} \frac{m_{k'l'}}{M} \sum_{i \in k} \frac{1}{I[i]} \mathbb{E}\left[ \frac{\exp\left\{u(x_i, y_j) + v(y_j, x_i)\right\}}{I[j]} \,|j \in (k', l') \right] \\
&= \beta_{kk'} \exp\left\{-B_{kk'}\mathbb{I}\left\{k \neq k'\right\}\right\} \sum_{l'} \delta_{ll'} \frac{m_{k'l'}}{M} \frac{n_{kl}}{N} \mathbb{E}\left[ \frac{\exp\left\{u(x_i, y_j) + v(y_j, x_i)\right\}}{I[j]I[i]} \,|j \in (k', l') \right]
\end{aligned}$$

□

Notice that populations with similar educational distributions will have higher expected number of interracial marriages, due to a higher value of $\gamma_{ll'} m_{k'l} n_{kl}$ . Recall that $\gamma_{ll} > \gamma_{ll'}$. Suppose that the distribution of educational levels for populations $k$ and $k'$ are similar. This will cause the product $m_{k'l} n_{kl}$ to be higher when multiplied by $\gamma_{ll}$. Intuitively, if people of different ethnic groups have similar educational levels, they have a higher chance of knowing each other, and therefore of marrying each other.

The effect of the online dating game on the number of any particular type of marriages is not trivial. This is because a change in meeting probabilities affects both the probability of a particular match, but also the inclusive value of each agent. Intuitively, the meeting technology introduced by online dating changes both meeting frequencies and the overall competition that anyone faces for obtaining a match. Since we have assumed that players download the online dating platform only if it increases their inclusive value, in this model online dating increases competition for a partner. In order to obtain some insight, let us first restrict attention to a simplified case.

### 5.5. **The effect of a change in exposure in matching frequencies**

Suppose that utilities are constant in $x$ and $y$, so that $u(x,y) = u$ and $v(y,x) = v$. Suppose also that there are $K$ racial groups, and the cost of crossing a racial boundary is the same for everyone, that is $b_{kk'}$ for all $k, k'$. Also, let $p_k$ be the fraction of same race individuals an agent from $k$ knows and $(1-p_k)$ the total number of agents from different groups. Notice that in such case, the inclusive value for agents is equal to

$$I_k^f(t_0) = n^f + p_k e^u + (1-p_k)e^{u-b} = n^f + e^u \left[p_k + (1-p_k)e^{-b}\right]$$
$$I_k^m(t_0) = n^m + q_k e^v + (1-q_k)e^{v-b} = n^m + e^v \left[q_{k'} + (1-q_{k'})e^{-b}\right]$$

Therefore the predicted shares of matches between $i$ and $j$ where $i$ belongs to $k$ and $j$ belongs to $k'$ is equal to

$$\mu_{kk'}(t_0) = \frac{(1-p_k)(1-q_{k'})e^{u+v-2b}}{(n^f + p_k e^u + (1-p_k)e^{u-b})\,(n^m + q_k e^v + (1-q_k)e^{v-b})}$$

**Proposition 6.** *Let $b$ be the bias, and let $p_k$ and $q_{k'}$ be the meeting frequencies for agents of the same the cluster of $k$ and $k'$. A change $\varepsilon$ in meeting frequencies causes a change in inter-cluster marriages of*

$$\Delta\mu_{kk'} = \tfrac{\varepsilon}{1+\varepsilon} \cdot \frac{\left(e^{-b}+(1-e^{-b})p_k\right)+\left(e^{-b}+(1-e^{-b})q_k\right)-\frac{\varepsilon}{1+\varepsilon}}{\left(e^{-b}+(1-e^{-b})p_k-\frac{\varepsilon}{1+\varepsilon}\right)\left(e^{-b}+(1-e^{-b})q_k-\frac{\varepsilon}{1+\varepsilon}\right)}$$

*Proof.* In Appendix A

□

Notice how the effect of a change in exposure is non-trivial, as it affects both meeting probabilities and inclusive values. In particular, higher values for the bias $b$ will cause the change in inter-cluster relations to be lower

## 6. Identification Strategy

In matching models with a complete network the probability of a match between two individuals of type $x$ and $y$ respectively will be given by

$$\mu_{xy} = \frac{\exp\{u(x,y) + v(y,x)\}}{I[x]I[y]} \tag{6.1}$$

Where the inclusive value $I[x]$ is defined as $I[x] = \sum_{y\in O(x)} \exp\{U_{xy}\}$ with $O(x)$ being the opportunity set of $x$. Relaxing the assumption that the network is complete we need to enrich the matching probability in (5.1) with the probability that a type $x$ and a type $y$ will know each other, that is with their probability of being connected in the network. Thus, letting $Pr\{(x,y) \in \mathcal{N}\}$ being the probability of a link existing between a type $x$ and a type $y$, the resulting matching probability is

$$\mu_{xy} = \Pr\{(x,y) \in \mathcal{N}\} \frac{\exp\{u(x,y) + v(y,x)\}}{I[x]I[y]}$$

Substituting the probability $Pr\{(x,y) \in \mathcal{N}\}$ with the probability as we modeled it in section 3 we obtain

$$\begin{aligned} \mu_{xy} &= p_{xy} \frac{\exp\{u(x,y) + v(y,x)\}}{I[x]I[y]} \\ &= \beta_{xy} p_x q_y \frac{\exp\{u(x,y) + v(y,x)\}}{I[x]I[y]} \end{aligned}$$

with $I[x] = \left(n_w + \sum_{y\in\mathcal{Y}} \beta^f_{xy} q_y \exp\{u(x,y)\}\right)$ and $I[y] = \left(n_m + \sum_{x\in\mathcal{X}} \beta^m_{yx} p_x \exp\{v(y,x)\}\right)$. Following Menzel (2015), the inclusive values $I[x]$ and $I[y]$ can be recovered by looking at the shares of singles and in particular

$$\begin{aligned} \mu_{x0} &= \exp\{-I[x]\} \\ \mu_{0y} &= \exp\{-I[y]\} \end{aligned}$$

So that

$$\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}} = \beta_{xy} p_x q_y \exp\{u(x,y) + v(y,x)\}$$

Notice that it is at this point impossible to separately identify the surplus of a match between a type $x$ and a type $y$ from the probability of the two types knowing each other. This in particular means that if we did not take into account the effect of the network on the matching probabilities and two types $x$ and $y$ are poorly connected in the network, we would end up concluding that the surplus of their match $u(x,y) + v(y,x)$ is lower than it is in reality.

**Proposition 7.** *If the clustering of the network is not taken into account for surplus estimation, the surplus estimates of a couple $xy$ will be biased by $-\ln(\beta_{kk'} p_x q_y)$.*

*Proof.* Let $s(x,y)$ be the pseudo-surplus of a couple, with $s(x,y) = u(x,y) + v(y,x)$. Since $\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}} = \beta_{xy} p_x q_y \exp\{u(x,y) + v(y,x)\}$ the real value of the pseudo-surplus is

$$s(x,y) = \ln\left(\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) - \ln\left(\beta_{xy} p_x q_y\right)$$

However, disregarding the clustering and incompleteness of the network we estimate

$$\hat{s}(x,y) = \ln\left(\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right)$$

The result follows.

□

Notice in particular that if we are estimating the surplus of a match of two types $x$ and $y$ with similar mass in the population, $p(x) = p(y)$ but that are poorly connected to each other in the network, that is $\beta_{xx} > \beta_{xy}$

$$s(x,x) - s(x,y) = \ln\left(\frac{\mu_{xx}}{\mu_{xy}} \cdot \frac{\mu_{0y}}{\mu_{0x}}\right) - \ln(\beta_{xx} - \beta_{xy})$$

The more segregated and clustered the network is, the bigger the difference $\beta_{xx} - \beta_{xy}$, and therefore the bigger the estimation bias if we were to not take into account the shape of the network.

Now suppose that there are two types of people, the ones who will use an online dating platform to look for a partner and those who will not. The composition of acquaintances and exposure to other types for these two types of people will be different. The goal is to disentangle the bias from the network effect. By comparing shares of couples who met online with the shares of couples who did not meet online we can obtain the network effect induced by online platforms.

Consider the share of couples with types $x$ and $y$ respectively that met online and the share of couples with the same characteristics that met offline. Let them be denoted by $\tilde{\mu}_{xy}$ and $\mu_{xy}$ respectively. We will assume that the bias is the same for both, as we think of it as an objective social cost incurred by people when marrying someone different. Since degrees of exposure of the online types are different than the ones of the offline types, their respective inclusive values will also be different. In particular, for the offline type we will have

$$\mu_{xy} = p_{xy} \frac{\exp\{U(x,y) + V(y,x) - b_{xy}\}}{I[x]I[y]} \tag{6.2}$$

For the online type

$$\tilde{\mu}_{xy} = \tilde{p}_{xy} \frac{\exp\{U(x,y) + V(y,x) - b_{xy}\}}{\tilde{I}[x]\tilde{I}[y]} \tag{6.3}$$

Next following Menzel (2015) the inclusive values for singles can be easily recovered by looking at the shares of singles among the online and offline people, $\mu_{x0}$ and $\tilde{\mu}_{x0}$ respectively. In particular

$$\mu_{x0} = \exp\{-I(x)\} \quad \tilde{\mu}_{x0} = \exp\left\{-\tilde{I}(x)\right\}$$
$$\mu_{0y} = \exp\{-I(y)\} \quad \tilde{\mu}_{0y} = \exp\left\{-\tilde{I}(y)\right\}$$

Therefore we can rewrite $\mu_{xy}$ and $\tilde{\mu}_{xy}$ as

$$\mu_{xy} = p_{xy}\mu_{xo}\mu_{0y}\exp\{U(x,y) + V(y,x) - b_{xy}\}$$
$$\tilde{\mu}_{xy} = \tilde{p}_{xy}\tilde{\mu}_{xo}\tilde{\mu}_{0y}\exp\{U(x,y) + V(y,x) - b_{xy}\}$$

Or

$$\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}} = p_{xy}\exp\{U(x,y) + V(y,x) - b_{xy}\} \tag{6.4}$$

$$\frac{\tilde{\mu}_{xy}}{\tilde{\mu}_{x0}\tilde{\mu}_{0y}} = \tilde{p}_{xy}\exp\{U(x,y) + V(y,x) - b_{xy}\} \tag{6.5}$$

Taking the ratio of the above expressions we obtain the odds of $x$ and $y$ meeting on an online dating platform versus them meeting in the exogenous network offline

$$\frac{\tilde{p}_{xy}}{p_{xy}} = \frac{\tilde{\mu}_{xy}}{\mu_{xy}} \cdot \frac{\mu_{x0}\mu_{0y}}{\tilde{\mu}_{x0}\tilde{\mu}_{0y}}$$

Since $\sum_{y\in\mathcal{Y}} \tilde{p}_{xy} = 1$ and $\sum_{y} p_{xy} = 1$ we can pin down $\tilde{p}_{xy}$ and $p_{xy}$. Having these two probabilities we can now estimate the cost of being in a relationship with a different type. By plugging $\hat{\tilde{p}}_{xy}$ and $\hat{p}_{xy}$ into expression 3.6 we obtain

$$\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}} = \hat{p}_{xy}\exp\{U(x,y) + V(y,x) - b_{xy}\}$$

Which finally leads to estimation of the surplus

$$U(x,y) + V(y,x) - b_{xy} = \log\left(\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}} \cdot \frac{1}{\hat{p}_{xy}}\right)$$

*Note:* If the bias represents the social cost of crossing an ethnic boundary, then such bias $b_{kk'}$ would be simply estimated by subtracting the surplus obtained for a couple of type $x$ and $y$ that belongs to the same ethnicity with the surplus obtained for a couple with the same characteristics but belonging to two different ethnic groups.

## 7. Estimation

Using the identification method proposed by Menzel (2015), firstly I estimate the pseudo-surplus for each couple in the sample using the following equation discussed in Section 6 that links observables to primitives

$$\ln\left(\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) = \ln p_{xy} + u(x,y) + v(y,x) \tag{7.1}$$

Notice that when $u$ and $v$ are constant in $x$ and $y$, the left-hand side of the equation captures merely meeting frequencies. On the other hand, when $p_{xy}$ is constant across $x$ and $y$, the left-hand side captures the sum of systematic utilities for $x$ and $y$. I assume the specification of the surplus to be as in Section 3, that is

$$\text{pseudo-surplus} = \ln\left(\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) = \ln\left(\beta_{kk'}p_x p_y\right) + u(x,y) + v(y,x) - b_{kk'}\mathbb{I}\{k \neq k'\} - d_{ll'}\mathbb{I}\left\{k \neq k'\right\}$$

Where recall that $\beta_{kk'}$ is a coefficient of clustering around race and the type vectors $x$ and $y$ include race, education, region of the United States, age and whether the couple met online or offline. The assumption here is that the race or ethnicity of an individual dictates both her preferences and the composition of her set of acquaintances. Moreover I will assume for the network adjustment that the composition of acquaintances of an individual is jointly and uniquely determined by their race and the region of the United States in which they reside, therefore disregarding the effect of education on the structure of the network. After estimating the pseudo-surplus for each couple, I adjust it by taking into account network estimates in different parts of the country from the American National Social Network Survey. I use such estimates as an approximation for $\beta_{kk'}p_x p_y$. In particular, since I use different network estimates for different regions of the country, I am approximating $\beta_{kk'}p_x p_y$ with $p^r_{kk'}$, where $p_{kk'}$ is the probability that two people from ethnic groups $k$ and $k'$ respectively know each other, and $r$ indicates the region of the United States (South, Northeast, West, Midwest). I will call the resulting estimate the adjusted pseudo-surplus.

$$\text{adj. pseudo-surplus} = \left(\ln\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) - \ln\left(p^r_{kk'}\right) = u(x,y) + v(y,x) - b_{kk'}\mathbb{I}\{k \neq k'\} - d_{ll'}\mathbb{I}\left\{k \neq k'\right\}$$

The adjustment increases the estimated surplus of mixed couples, taking into account that their meeting frequencies are lower compared to same-race couples. Notice that in the approximation we are assuming that the probability of two types $x$ and $y$ of knowing each other is entirely dependent on their race and the region where they live. This is clearly a significant limit of the estimation, as the clustering of the network depends on other factors as well, such as education and income. In order to take into account also the dimension of clustering dictated by other factors such as educational levels, I would need more detailed data on the social network structure.

As presented in Section 6, in order to be able to estimate utilities we need to have the shares of singles $\mu_{x0}$ for each type $x$ that is taken into consideration. The shares of singles divided by ethnicity among the online and offline dating community - $\tilde{\mu}_{x0}$ and $\mu_{x0}$ respectively - will be derived from the information provided in question w6_otherdate_app in the questionnaire ("In the Past Year have you ever used an app on your phone (such as Tinder or Grindr) to meet someone in person, for dating, romance, or sex?"). Using this information we are able to distinguish among single people who used online dating and people who did not in the year before the survey.

The resulting estimated utilities thus obtained are in line with the results published by online dating companies such as OkCupid regarding the behavior of their users when it comes to race and ethnicity. Firstly, in general, there seems to be a remarkable preference for ethnic homogamy both for men and women. Secondly, the adjusted psuedo-surplus of mixed-race and mixed-education couples who met online is consistently higher than the adjusted pseudo-surplus of mixed couples who met offline, suggesting once again the likely impact of a network effect on matching probabilities. Thirdly, minorities who use online dating seem to focus their search for a partner primarily on other individuals of the same ethnicity or the majority group of white people. A possible explanation for this phenomenon is provided by the theoretical model presented in the previous sections. Finally, online dating seems to have contributed to the creation of a higher number of mixed-race-same-education or same-race-mixed-education couples, but not mixed-race-mixed-education couples. These results are summarized in Table 1.

As mentioned above, the utility derived from an ethnically homogamous match seems higher across ethnicities and this seems particularly true for minorities. However, since the formula for the pseudo-surplus includes both the network effect and the effect of preferences, the higher utility estimate obtained among the ethnically homogamous couples that met offline might be simply due to a higher probability of two people of the same ethnicity knowing each other compared to their probability of knowing someone of another group. In other words, the higher estimate might simply reflect a higher degree of exposure $\beta_{kk}$. This problem is resolved by estimating the adjusted pseudo-surplus and taking into account network probabilities.

Finally I use the pseudo-surplus and the adjusted pseudo-surplus estimates to establish the existence of a social cost of crossing an ethnic and an educational boundary and of a network effect that influences the probabilities of a match between individuals in different regions of the United States. I compare estimates of the social cost of crossing a racial boundary using the pseudo-surplus and the adjusted pseudo-surplus in order to show the magnitude of the estimation bias in the first case. I then assess whether the way people meet has a significant effect on their degree of assortativeness. The estimation shows the existence of significant costs of crossing both racial and educational boundaries across all areas of the United States. The cost of crossing a racial boundary appears to be higher in the South and the Midwest when compared to the West and Northeast of the country. Moreover, interracial and intereducational couples who have met online in the Northeast and the West appear to have a higher pseudo-surplus and adjusted pseudo-surplus than the couples who have met offline. This suggests that the way a couple has met influences their degree of assortativeness, and in particular that in the Northeast and the West the degree of ethnic and educational assortativeness is lower for couples who have met online. Since I do not observe meeting probabilities online, the difference might be due to an increase of exposure to other races online or it might be due to a selection effect onto the platform. On the other hand, in the South and the Midwest I find no significant difference in surplus estimates for mixed couples in the online and offline sample, both before and after network adjustments. This empirical finding is compatible with the theoretical model presented in Section 4, in the sense that it appears to be the case that in the Southern and Midwestern areas of the United States, due to a higher cost of crossing a racial boundary in a romantic relationship, the effect of directed search on online dating platforms is greater than the integration effect introduced by those same platforms. Results are summarized in Tables 2 and 3.

Finally, estimates are used to establish the increase in exposure to different races on online dating platforms. Table 4 reports the percentage of acquaintances of a different race for white people and minorities in the United States as a whole, and then in the Northeast and West of the country, where we know online dating has contributed to increasing diversity of the set of acquaintances. The estimates show that for an average white person in the United States the probability of knowing someone of a different race has increased from 8% to 11.6% as a result of the introduction of online dating. Effects in the Northeast and the West are even more remarkable.

For the sake of estimation I will define types according to whether they met online or offline, their ethnicity, the ethnicity of their partner, their education and their partner's education, in what region of the United States they live, their age group, and in the last part further distinguishing between females and males in the sample. In regards to ethnicity, we will consider a partition of the sample into White, Black, Hispanic and Asian or other minorities. The number of Asian subjects in the sample is too small (especially for the subgroup of individuals who met their partner online) to be able to derive a reliable utility estimation as a stand alone group. For this reason I group together Asian subjects and all other minorities recorded in the sample that are not already accounted for in another category. For the sake of completeness I include in the Appendix utility estimations that consider Asian and other minorities individuals as two separate groups.

### 7.1. Race and Education in Online Dating Markets

In this section I study the effects of online dating on the relative rates of crossing a racial or an educational boundary, or both. In regressions (1) and (2) I restrict attention only to couples who are either married or in a relationship, while for regressions (3) and (4) I include subjects who reported seeing someone more casually. For regressions (1) and (3) I regress the pseudo-surplus, without adjusting for network estimates. That is the dependent variable is

$$\ln\left(\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) = \ln(p_{xy}) + u(x,y) - b_{kk'}\mathbb{I}\left\{k \neq k'\right\} - d_{ll''}\mathbb{I}\left\{l \neq l'\right\}$$

And the specification for the regression is

$$\begin{aligned}\ln\left(\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) &= \beta_0 + \beta_1 \cdot mixed\ ethn + \beta_2 \cdot mixed\ ed + \beta_3 \cdot mixed\ ethn \times mixed\ ed + \\ &+ \beta_4 \cdot mixed\ ethn \times metonline + \beta_5 \cdot mixed\ ed \times metonline + \\ &+ \beta_6 \cdot mixed\ ethn \times mixed\ ed \times metonline + \beta_7 \cdot FE + \varepsilon\end{aligned}$$

For regressions (2) and (4), I adjust the left-hand side to take into account meeting probabilities for agents belonging to different ethnic groups. This allows me to reduce the bias in the estimation

Table 7.1

| | Dependent variable: adjusted pseudo-surplus | |
|---|---|---|
| | (1) | (2) |
| mixed ethn | -1.359*** | |
| | (-1.463 , -1.255) | |
| mixed ethn online | 0.286*** | |
| | (0.094 , 0.477) | |
| mixed ethn white | | -1.350*** |
| | | (-1.452 , -1.248) |
| mixed ethn minority | | -1.397*** |
| | | (-1.607 , -1.187) |
| mixed ethn white online | | 0.552*** |
| | | (0.347 , 0.757) |
| mixed ethn minority online | | 0.098 |
| | | (-0.145 , 0.340) |
| mixed educ | -0.819*** | |
| | (-1.217 , -1.113) | |
| mixed educ online | 0.321*** | |
| | (0.293 , 0.562) | |
| mixed educ same ethn | | -0.886*** |
| | | (-0.955 , -0.817) |
| mixed educ same ethn online | | 0.385*** |
| | | (0.229 , 0.542) |
| mixed educ mixed ethn | | -0.595*** |
| | | (-0.765 , -0.426) |
| mixed educ mixed ethn online | | 0.111 |
| | | (-0.203 , 0.426) |
| Observations | 2,844 | 2,844 |
| $R^2$ | 0.805 | 0.805 |
| Adjusted $R^2$ | 0.804 | 0.805 |
| Online FE | Y | Y |
| Race, Educ FE | Y | Y |
| Geog FE | Y | Y |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

of the cost of crossing a racial boundary. Formally, let $p^r_{kk'}$ the probability that an agent of race or ethnicity $k$ is connected in the network to an agent of race $k'$ in region $r$. The dependent variable is as follows

$$\left(\ln \frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) - \ln\left(p^r_{kk'}\right) = u(x,y) + v(y,x) - b_{kk'}\mathbb{I}\{k \neq k'\} - d_{ll'}\mathbb{I}\left\{k \neq k'\right\}$$

If most of the information on whether two people know each other is dictated by their race and the region in which they live, then we have a good estimate of their utility from a match. However, if that is not the case, the estimator will be biased. I will then use the adjusted pseudo-surplus $\ln\left(\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) - \ln\left(p^r_{kk'}\right)$ as a dependent variable and estimate

$$\begin{aligned}\left(\ln\frac{\mu_{xy}}{\mu_{x0}\mu_{0y}}\right) - \ln\left(p^r_{kk'}\right) = \beta_0 + \beta_1 \cdot mixed\ ethn + \beta_2 \cdot mixed\ ed + \beta_3 \cdot mixed\ ethn \times mixed\ ed + \\ + \beta_4 \cdot mixed\ ethn \times metonline + \beta_5 \cdot mixed\ ed \times metonline + \\ + \beta_6 \cdot mixed\ ethn \times mixed\ ed \times metonline + \beta_7 \cdot FE + \varepsilon\end{aligned}$$

as in the equation above. In both regressions fixed effects include race of the subject, race of the partner, geographical region of the United Stated, educational levels of the subject and their partner, and age group.

Network adjustments in regressions (2) and (4) diminish the estimated coefficients associated with the dummy variable of a mixed race couple. This reflects the fact that what we are capturing in regressions (1) and (3) is the sum of the network effect and the effect of homogamous preferences, as described in Section 3. On the other side, the network effect increases in absolute value the estimated cost of crossing an educational boundary. This hints at the presence of directed search in online dating markets. Finally, notice how the estimated coefficient for the dummy variable for couples that are both mixed on an educational and a racial level is negative, and online dating has no significant effect on these type of couples. This again could be a reflection of the effects of directed search on an online dating market. There exists a tradeoff between crossing a racial and an educational boundary, and crossing both boundaries is expensive. Online dating allows users to select what boundaries to cross, if any, and the cost of crossing both boundaries simultaneously is high. The theoretical model explains how this reflects on behaviour on online dating platform, that in this case will have, less of an integration effect and more of a directed search effect.

In terms of the estimation, given that the sample is small, I use K-modes clustering to group interviewed subjects and partners into different groups. In particular, I cluster subjects into 15 clusters, and I split these into online and offline sample, so that I obtain a total of 30 clusters. The variables used for clustering are ethnicity, education, and region of the United States. On the side of the partners, I cluster partners into 6 groups, based on their educational level and ethnicity. A full description of the clusters is provided in the Appendix.

In Tables 3 and 4 I regress the pseudo-surplus, both non adjusted and adjusted with network probabilities, on indicators of the 4 macro regions of the United States, to assess whether online dating has had an effect on the relative frequencies of interracial and inter-educational couples in different parts of the country. The regions I consider are the South, Midwest, Northeast and West. The first two columns exclude subjects that have reported seeing someone, but not being in a stable relationship with that person yet, while the second two columns include in the regression also subjected who reported having a casual relationship or that just started dating someone. Table 3 shows that adjusting for network estimates, the estimated cost of crossing a racial boundary decreases, and the effect of online dating increases. In all regions of the country the cost associated with crossing a racial boundary is significant. The estimated cost is, as expected, lower in regressions (2) and (4) due to network adjustments. Moreover such cost is higher in the South and Midwest when compared to the Northeast and the West. Subsequently, due to the joint effect of integration effect and directed search on online dating platforms, in the Northeast and West online dating seems to have a positive effect on the number of interracial couples. On the contrary, in the South and Midwest online dating does not have a significant effect on interracial couples neither before nor after adjusting for meeting frequencies. This is the case of the lower left part of the table presented in Section 4.3, where due to stark preferences for homogamy, the effect of directed search on online dating platform is higher than the integration effect. From the point of view of education, the cost of crossing an educational boundary is also significant across all regions. Online dating has had a positive impact on the rates on inter-educational couples in all regions except for the South. Non-coincidentially, the South is also the region where the cost of crossing an educational boundary is the lowest.

The coefficient associated with the variable mixed ethnicity online and mixed education online could be capturing two different characteristics of agents on online dating sites: either an increased exposure towards agents of different races and educational levels, or a lower cost of crossing a racial or educational boundary for people that self-select into the platform. Formally, let $p_{kk'}$ and $\tilde{p}_{kk'}$ be the meeting frequencies of agents belonging to $k$ and $k'$ offline and online respectively, and let $b$ and $\tilde{b}$ be the costs associated with crossing a racial boundary for the offline and online samples respectively. Then

$$\beta_4 = \ln\left(\frac{\tilde{p}_{kk'}}{p_{kk'}}\right) + b - \tilde{b}$$

Therefore we have two extreme case scenarios. In the first case, the online sample and the offline sample are identical from a preference standpoint. Their difference in meeting patterns is therefore reflective of an increased exposure to a more diverse set of potential partners for those who decided to use online dating platforms. In such case, the increase in meeting frequencies is summarized in Table 4. In the other case, there is a difference between the perceived cost of crossing a racial boundary among the online and offline populations. However, notice that it cannot be the case that the network formed online is identical to the exogenous offline network, the reason being that, if that were to be the case, agents with a lower racial bias would not have an incentive to enter into the platform. Thefore, in the upper bound the coefficient $\beta_4$ captures a combination of the effect of self-selection into the platform, and the increased exposure obtained through directed search by users on the platform. In order to consider this case, I will use the optimality conditions from the population game in Section 3 to determine the resulting exposure in the online dating platform. In order to do so, I assume that the equilibrium achieved on the platform during the population game is a monomorphic equilibrium, as described in the previous sections, and that the cost function is quadratic. In this second case, results are found through calibration of the model. In such case, the estimated difference in bias among the online and offline samples would be 0.141, while as a result of directed search, exposure to different races for a white person would be increased by 1.45% and for a minority group would be increased by 5.6%.

Table 7.2. Ethnicity

| | (1) pseudo-surplus | (2) adjusted psuedo-surplus | (3) pseudo-surplus | (4) adjusted pseudo-surplus |
|---|---|---|---|---|
| South Mixed | -1.227*** | -0.797*** | -1.198*** | -0.767*** |
| | (-1.371 , -1.083) | (-0.941 , -0.654) | (-1.342 , -1.055) | (-0.911 , -0.624) |
| Northeast Mixed | -0.903*** | -0.353*** | -0.908*** | -0.357*** |
| | (-1.070 , -0.736) | (-0.519 , -0.186) | (-1.076 , -0.741) | (-0.525 , -0.190) |
| West Mixed | -0.545*** | -0.381*** | -0.538*** | -0.373*** |
| | (-0.678 , -0.412) | (-0.513 , -0.248) | (-0.669 , -0.407) | (-0.503 , -0.242) |
| Midwest Mixed | -0.987*** | -0.571*** | -1.016*** | -0.600*** |
| | (-1.160 , -0.814) | (-0.744 , -0.399) | (-1.191 , -0.842) | (-0.774 , -0.426) |
| Mixed metonline S | 0.178 | 0.178 | 0.151 | 0.151 |
| | (-0.091 , 0.447) | (-0.091 , 0.447) | (-0.114 , 0.415) | (-0.114 , 0.415) |
| Mixed metonline N | 0.654*** | 0.654*** | 0.655*** | 0.655*** |
| | (0.305 , 1.003) | (0.305 , 1.003) | (0.307 , 1.002) | (0.307 , 1.003) |
| Mixed metonline W | 0.335*** | 0.335*** | 0.311*** | 0.312*** |
| | (0.120 , 0.550) | (0.120 , 0.550) | (0.097 , 0.526) | (0.097 , 0.526) |
| Mixed metonline M | 0.405 | 0.405 | 0.397 | 0.397 |
| | (-0.083 , 0.893) | (-0.083 , 0.893) | (-0.078 , 0.872) | (-0.078 , 0.872) |
| Observations | 2,844 | 2,844 | 2,941 | 2,941 |
| $R^2$ | 0.871 | 0.854 | 0.867 | 0.850 |
| Adjusted $R^2$ | 0.870 | 0.853 | 0.867 | 0.849 |
| Online FE | Y | Y | Y | Y |
| Race, Educ FE | Y | Y | Y | Y |
| Geog FE | Y | Y | Y | Y |
| Casual Relationships | N | N | Y | Y |

*Note:* *p<0.1; **p<0.05; ***p<0.01

## Table 7.3. Education

| | (1) pseudo-surplus | (2) adjusted pseudo-surplus | (3) pseudo-surplus | (4) adjusted pseudo-surplus |
|---|---|---|---|---|
| South Mixed Educ | -0.756*** | -0.757*** | -0.734*** | -0.734*** |
| | (-0.804 , -0.709) | (-0.804 , -0.710) | (-0.782 , -0.686) | (-0.781 , -0.687) |
| Northeast Mixed Educ | -0.788*** | -0.793*** | -0.780*** | -0.786*** |
| | (-0.860 , -0.716) | (-0.865 , -0.722) | (-0.851 , -0.709) | (-0.857 , -0.716) |
| West Mixed Educ | -0.853*** | -0.858*** | -0.849*** | -0.852*** |
| | (-0.917 , -0.789) | (-0.922 , -0.793) | (-0.912 , -0.785) | (-0.915 , -0.788) |
| Midwest Mixed Educ | -0.937*** | -0.935*** | -0.944*** | -0.940*** |
| | (-1.022 , -0.852) | (-1.018 , -0.852) | (-1.027 , -0.860) | (-1.021 , -0.859) |
| Mixed Educ metonline S | 0.095 | 0.103 | 0.079 | 0.087 |
| | (-0.045 , 0.235) | (-0.036 , 0.243) | (-0.060 , 0.217) | (-0.051 , 0.225) |
| Mixed Educ metonline N | 0.445*** | 0.467*** | 0.434*** | 0.458*** |
| | (0.216 , 0.674) | (0.233 , 0.701) | (0.205 , 0.663) | (0.224 , 0.692) |
| Mixed Educ metonline W | 0.388*** | 0.371*** | 0.385*** | 0.366*** |
| | (0.249 , 0.527) | (0.233 , 0.508) | (0.247 , 0.523) | (0.229 , 0.502) |
| Mixed Educ metonline M | 0.587*** | 0.584*** | 0.593*** | 0.589*** |
| | (0.383 , 0.790) | (0.384 , 0.784) | (0.391 , 0.795) | (0.391 , 0.787) |
| Observations | 2,844 | 2,844 | 2,941 | 2,941 |
| $R^2$ | 0.871 | 0.858 | 0.868 | 0.855 |
| Adjusted $R^2$ | 0.870 | 0.857 | 0.867 | 0.854 |
| Online FE | Y | Y | Y | Y |
| Race, Educ FE | Y | Y | Y | Y |
| Geog FE | Y | Y | Y | Y |
| Casual Relationships | N | N | Y | Y |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Table 7.4. Increased Meeting Frequencies

| | (1) Offline $p_d$ | (2) Online $p_d$ | (3) Difference |
|---|---|---|---|
| U.S. white | 0.087 | 0.116*** | 0.029*** |
| | | (0.111, 0.121) | (0.024, 0.034) |
| U.S. minority | 0.243 | 0.323*** | 0.080*** |
| | | (0.318, 0.328) | (0.078, 0.082) |
| Northeast white | 0.072 | 0.138*** | 0.066*** |
| | | (0.125, 0.151) | (0.053, 0.079) |
| Northeast minority | 0.177 | 0.340*** | 0.163*** |
| | | (0.327, 0.353) | (0.150, 0.176) |
| West white | 0.153 | 0.214*** | 0.061*** |
| | | (0.208, 0.220) | (0.055, 0.067) |
| West minority | 0.285 | 0.398*** | 0.113*** |
| | | (0.392, 0.404) | (0.107, 0.119) |

## 8. Evolution of the Network

Let us now move on to consider the effects that a marriage, whether it is same-race or mixed-race, has on the social network. Assume that agents in the economy have a fixed number of people that they are connected to in the network. Such number is assumed to only depend on the agent's type, as modeled in section 3. The set of acquaintances of each agent can vary in terms of racial composition, as we have described in the previous sections.

Assume that after a match is created, the individuals that a couple is directly connected to in the network meet. For example, assume that both the bride and the groom know 10 people each. For simplicity assume that all 10 friends of the groom know each other and all 10 friends of the bride know each other. To explain the evolution of the network, assume that the 10 friends on each side meet each other after the match has occurred, creating 100 new links. In this sense, we can say that a marriage creates an initial link between two parts of the network, and more links are created as a result of it. The dynamic of network formation arising from a marriage is represented in Figures 7.1, 7.2 and 7.3. In the figures, red lines represent links among members of group $x$, blue lines represent links among members of group $y$, and green links represent diverse links, that is links between a member of group $x$ and and member of group $y$. From the figures one can see how in a clustered network an interracial marriage will contribute more to the creation of diverse links than a same-race marriage, and how same-race marriages can, under certain conditions, exacerbate the clustering of the network. This is formalized below.

Let types be fully described by race. Moreover, assume that there are only two racial groups, denoted $x$ and $y$. Let the composition of acquaintances of people from group $x$ be described by the vector $[\delta_{xx}, \delta_{xy}]$, and the composition of acquaintances of people from group $y$ be described by the vector $[\delta_{yx}, \delta_{yy}]$. As in Section 3, $\delta_{xx}$ is the number of acquaintances of type $x$ that a type $x$ has, and $\delta_{xy}$ is their number of acquaintances of type $y$. Notice that every marriage between agents of different groups $\mu_{xy}$, contributes to the number of new links that are created between people belonging to different groups with $\delta_{xx}\delta_{yy} + \delta_{xy}\delta_{yx}$ new links. That is

$$\mu^t_{xy} > 0 \implies \delta^{t+1}_{xy} = \delta^t_{xx}\delta^t_{yy} + \delta^t_{xy}\delta^t_{yx} \tag{8.1}$$

Following the same logic, for every within-cluster marriage between agents of group $x$, the fraction of mixed links $(x, y)$ that are created is equal to $2\delta_{xx}\delta_{xy}$, that is

$$\mu^t_{xx} > 0 \implies \delta^{t+1}_{xy} = 2\delta^t_{xx}\delta^t_{xy} \tag{8.2}$$

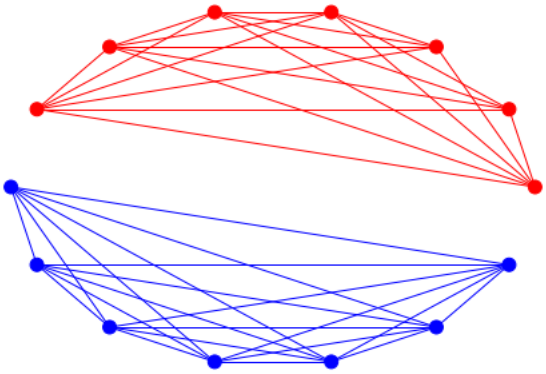
Finally, for every within-cluster marriage between agents of group $y$ the fraction of mixed links that are created is equal to

$$\mu^t_{yy} > 0 \implies \delta^{t+1}_{xy} = 2\delta_{yy}\delta_{yx} \tag{8.3}$$

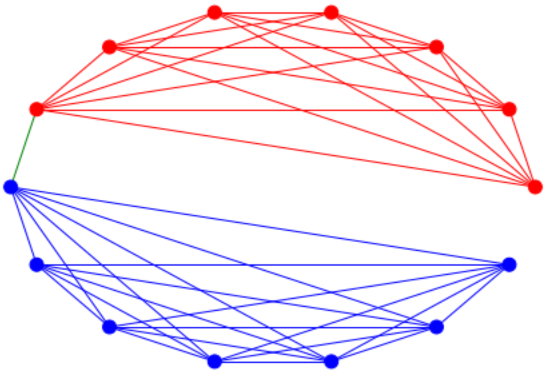
The total number of new mixed links $(x, y)$ that are created in the network as a consequence of marriages is

$$\left(\delta^t_{xx}\delta^t_{yy} + \delta^t_{xy}\delta^t_{yx}\right)\mu^t_{xy} + \left(2\delta^t_{xx}\delta^t_{xy}\right)\mu^t_{xx} + \left(2\delta_{yy}\delta_{yx}\right)\mu^t_{yy} \tag{8.4}$$

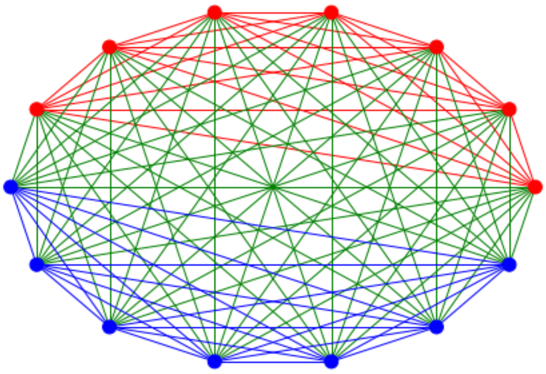
where $\mu^t_{xy}$ is the total fraction of different-type marriages $xy$ at time $t$, and $\mu^t_{xx}, \mu^t_{yy}$ are the total fractions of same-type marriages $xx$ and $yy$ respectively at time $t$.



(A) Before Interracial Matching



(B) Matching

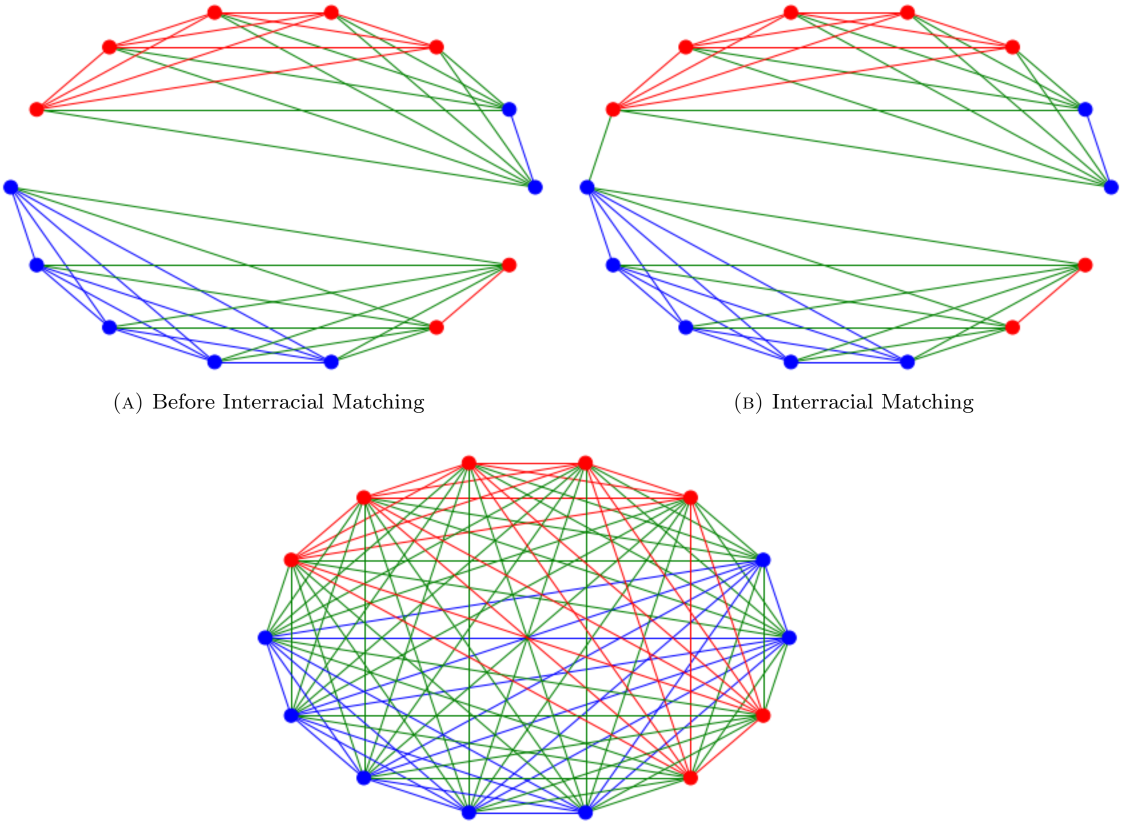


(C) After Interracial Matching

FIGURE 8.1. Interracial Marriages in a Totally Clustered Network

Following the same logic, a mixed marriage creates $\delta^t_{xx}\delta^t_{yy} + \delta^t_{xy}\delta^t_{yx}$ new mixed links in the network, $\delta_{xx}\delta_{yx}$ new links among types $x$'s, and $\delta_{xy}\delta_{yy}$ new links among types $y's$. A marriage among types $x$'s creates $2\delta^t_{xx}\delta^t_{xy}$ new mixed links, $\left(\delta^t_{xx}\right)^2$ new links between types $x'$s, and $\left(\delta^t_{xy}\right)^2$ new links between types $y$'s. And finally, a marriage among types $y$'s creates $2\delta_{yy}\delta_{yx}$ new inter-cluster links, $\left(\delta^t_{yx}\right)^2$ new links among types $x'$s, and $\left(\delta^t_{yy}\right)^2$ new links among types $y's$. The following table summarizes the resulting links created by each type of marriage.

| | link $(x, y)$ | link $(x, x)$ | link $(y, y)$ |
|---|---|---|---|
| marriage $(x, y)$ | $\delta_{xx}\delta_{yy} + \delta_{xy}\delta_{yx}$ | $\delta_{xx}\delta_{yx}$ | $\delta_{xy}\delta_{yy}$ |
| marriage $(x, x)$ | $2\delta_{xx}\delta_{xy}$ | $\delta^2_{xx}$ | $\delta^2_{xy}$ |
| marriage $(y, y)$ | $2\delta_{yy}\delta_{yx}$ | $\delta^2_{yx}$ | $\delta^2_{yy}$ |



(A) Before Interracial Matching

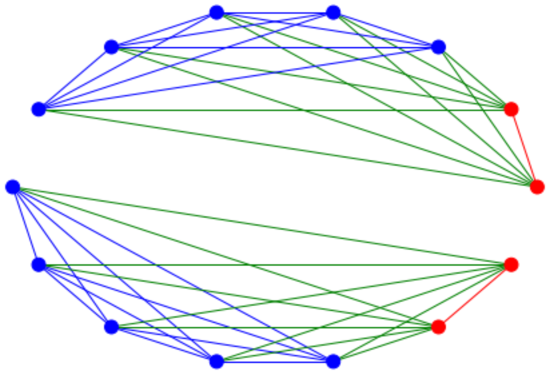
(B) Interracial Matching

(C) After Interracial Matching

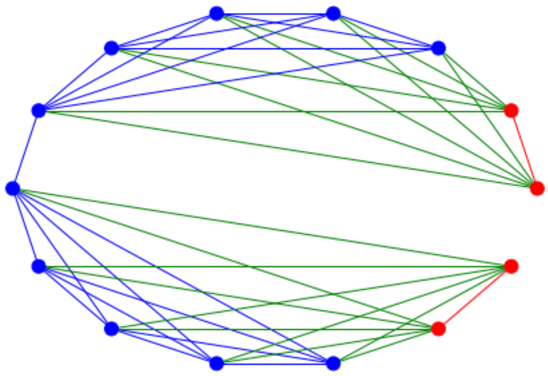
FIGURE 8.2. Interracial Marriages in Partially Clustered Network

It is easy to see how in a clustered network where $\delta_{xx} > \delta_{xy}$ and $\delta_{yy} > \delta_{yx}$, an interracial marriage contributes much more to the creation of new diverse links than a same-race marriage does. In particular, if $\delta_{xx} > 2\delta_{xy}$ and $\delta_{yy} > 2\delta_{yx}$, same-type marriages exacerbate the clustering of the network, by creating more same-type links than different-type links. In probabilistic terms, let $\delta_x = \delta_{xx} + \delta_{xy}$ and $p = \frac{\delta xx}{\delta x}$. If $p > \frac{2}{3}$, same type marriages create more same-type links than different-type links, thus perpetuating the imbalance and clustering of the network.

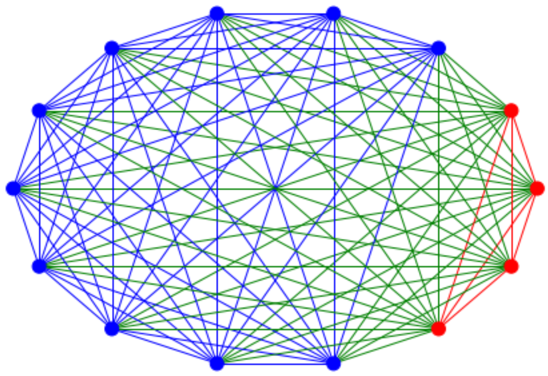
On the other hand, an interracial marriage in a clustered network always contributes to the diversity of the network by causing the formation of different-type links. In particular, let $p = \frac{\delta xx}{\delta x}$ and $q = \frac{\delta yy}{\delta y}$. Assume the two network clusters for types $x$ and types $y$ are equally segregated, with $p = q$. Then, an interracial marriage contributes to network diversity by creating more diverse links than same-type links, as long as $p > \frac{1}{2}$. Thus, as long as the network is clustered and people of the same-type have a higher number of connections to other agents of the same type than to different agents, interracial marriages contribute to the de-clustering of the network.



(A) Before Same Race Matching



(B) Same Race Matching



(C) After Same Race Matching

FIGURE 8.3. Same Race Marriage in a Partially Clustered Network

I assume that the exogenous network at time $t+1$ is the convex combination of the network at time $t$ and the new links formed between different and same types as a consequence of marriages. In particular assume that the law of motion for the number and composition of acquaintances is

$$\delta_{xx}^{t+1} = \lambda\delta_{xx}^{t} + (1-\lambda)\left(\delta_{xx}^{t}\delta_{yx}^{t}\mu_{xy}^{t} + \left(\delta_{xx}^{t}\right)^{2}\mu_{xx}^{t} + (\delta_{yx})^{2}\mu_{yy}^{t}\right) \tag{8.5}$$

$$\delta_{yy}^{t+1} = \lambda\delta_{yy}^{t} + (1-\lambda)\left(\delta_{xy}^{t}\delta_{yy}^{t}\mu_{xy}^{t} + (\delta_{xy}^{t})^{2}\mu_{xx}^{t} + \left(\delta_{yy}^{t}\right)^{2}\mu_{yy}^{t}\right) \tag{8.6}$$

$$\delta_{xy}^{t+1} = \lambda\delta_{xy}^{t} + (1-\lambda)\left(\left(\delta_{xx}^{t}\delta_{yy}^{t} + \delta_{xy}^{t}\delta_{yx}^{t}\right)\mu_{xy}^{t} + 2\delta_{xx}^{t}\delta_{xy}^{t}\mu_{xx}^{t} + 2\delta_{yy}^{t}\delta_{yx}^{t}\mu_{yy}^{t}\right) \tag{8.7}$$

$$\delta_{yx}^{t+1} = \lambda\delta_{yx}^{t} + (1-\lambda)\left(\left(\delta_{xx}^{t}\delta_{yy}^{t} + \delta_{xy}^{t}\delta_{yx}^{t}\right)\mu_{xy}^{t} + 2\delta_{xx}^{t}\delta_{xy}^{t}\mu_{xx}^{t} + 2\delta_{yy}^{t}\delta_{yx}^{t}\mu_{yy}^{t}\right) \tag{8.8}$$

## 8.1. Steady States

To have an understanding of how the evolution of the network works, let us start with a study of the steady states of this dynamic system. Notice that if the network has two completely disconnected clusters, one for $x$ and one for $y$, the system is in a steady state. Formally, if the two clusters of agents of types $x$ and $y$ are completely disconnected we have $\delta_{xy} = \delta_{yx} = 0$

Firstly, notice that since the total share of marriages between $x$ and $y$ is equal to $\mu_{xy} = \delta_{xy}^{t} e^{-B_{xy}\frac{\exp\{u(x,y)+v(y,x)\}}{I^{t}[x]I^{t}[y]}}$ , if $\delta_{xy} = 0$ then $\mu_{xy} = 0$. Therefore the law of motion of the system in (8.5),(8.6) and (8.7) becomes

$$\begin{aligned}
\delta_{xx}^{t+1} &= \lambda\delta_{xx}^{t} + (1-\lambda)\left(\delta_{xx}^{t}\right)^{2}\mu_{xx}^{t}\\
\delta_{yy}^{t+1} &= \lambda\delta_{yy}^{t} + (1-\lambda)\left(\delta_{yy}^{t}\right)^{2}\mu_{yy}^{t}\\
\delta_{xy}^{t+1} &= 0\\
\delta_{yx}^{t+1} &= 0
\end{aligned}$$

Secondly, notice that since $x$ and $y$ are the only types in the economy, $\delta_{xy}^{t} = \delta_{yx}^{t} = 0$ imply $\delta_{xx}^{t} = \delta_{x}^{t}$ and $\delta_{yy}^{t} = \delta_{y}$. Thus within-type acquaintances are all the acquaintances an agent has. Therefore the system boils further down to

$$\begin{aligned}
\delta_{x}^{t+1} &= \lambda\delta_{x}^{t} + (1-\lambda)\left(\delta_{x}^{t}\right)^{2}\mu_{xx}^{t}\\
\delta_{y}^{t+1} &= \lambda\delta_{y}^{t} + (1-\lambda)\left(\delta_{y}^{t}\right)^{2}\mu_{yy}^{t}\\
\delta_{xy}^{t+1} &= 0\\
\delta_{yx}^{t+1} &= 0
\end{aligned}$$

Recall that from proposition 12 the expected number of homogamous marriages is $\mu_{xx}^{t} = \delta_{xx}^{t} \cdot \frac{\exp\{u(x,x)+v(x,x)\}}{I^{t}[x]I^{t}[x]}$. Let $S_{xx}^{t} = \frac{\exp\{u(x,x)+v(x,x)\}}{I^{t}[x]I^{t}[x]}$ and $S_{yy}^{t} = \left[\frac{\exp\{u(y,y)+v(y,y)\}}{I^{t}[y]I^{t}[y]}\right]$, so that we can rewrite the total number of marriages as $\mu_{xx}^{t} = \delta_{xx}^{t}S_{xx}^{t}$ and $\mu_{yy}^{t} = \delta_{yy}^{t}S_{yy}^{t}$ respectively. Notice that $I\left[x\right]$ and $I\left[y\right]$ depend on $\delta_{xx}^{t}, \delta_{yy}^{t}$ and $\delta_{xy}^{t}$ also. In this case, $I^{t}\left[x\right] = \delta_{xx}\exp\left\{u(x,x)\right\}$ and $I^{t}\left[y\right] = \delta_{yy}\exp\left\{v(y,y)\right\}$. Putting everything together we obtain

$$\begin{aligned}
\delta_{x}^{t+1} &= \lambda\delta_{x}^{t} + (1-\lambda)\left(\delta_{x}^{t}\right)^{3}S_{xx}^{t} = \lambda\delta_{x}^{t} + (1-\lambda)\delta_{x}^{t}\\
\delta_{y}^{t+1} &= \lambda\delta_{y}^{t} + (1-\lambda)\left(\delta_{y}^{t}\right)^{3}S_{yy}^{t} = \lambda\delta_{y}^{t} + (1-\lambda)\delta_{y}^{t}\\
\delta_{xy}^{t+1} &= 0\\
\delta_{yx}^{t+1} &= 0
\end{aligned}$$

This proves that the system is in steady state in two ways: first of all, if the two clusters $x$ and $y$ are completely separated, no new links are going to be formed between the two different clusters in the future, at least through marriages. That is

$$\delta_{xy}^{t}=\delta_{yx}^{t}=0\implies\delta_{xy}^{t'}=\delta_{yx}^{t'}=0\ \forall t'\geq t$$

Moreover, in this case the system is also in steady state in the sense that the total number of acquaintances does not change over time. Formally

$$\begin{aligned}\delta_{xy}^{t}=0&\implies\delta_{x}^{t}=\delta_{x}^{t'}\ \ \forall t'\geq t\\ \delta_{yx}^{t}=0&\implies\delta_{y}^{t}=\delta_{y}^{t'}\ \ \forall t'\geq t\end{aligned}$$

In conclusion one steady state is the totally segregated one, with $\delta_x=[\delta_{xx},\ 0]$ and $\delta_y=[0,\delta_{yy}]$. Network clusters are completely segregated and they stay that way.

Other steady states are found when $\delta_{xy}>0$. However these steady states are more complicated to derive in closed form, as they depend both on the exposures $\delta_{xx},\delta_{yy},\delta_{yx}$ and on preferences $u,v$ and $b$. The steady states are characterized by the following system of equations

$$\begin{aligned}\delta_{xx}&=e^{u+v}\left(\frac{\delta_{xx}\delta_{yx}^{2}e^{-2b}}{I[x]\,I[y]}+\frac{\delta_{xx}^{3}}{I[x]^{2}}+\frac{\delta_{yx}^{2}\delta_{yy}}{I[y]^{2}}\right)\\ \delta_{yy}&=e^{u+v}\left(\frac{\delta_{xy}^{2}\delta_{yy}e^{-2b}}{I[x]\,I[y]}+\frac{\delta_{xy}^{2}\delta_{xx}}{I[x]^{2}}+\frac{\delta_{yy}^{3}}{I[y]^{2}}\right)\\ \delta_{xy}&=e^{u+v}\left((\delta_{xx}\delta_{yy}+\delta_{xy}\delta_{yx})\frac{\delta_{xy}e^{-2b}}{I[x]\,I[y]}+\frac{2\delta_{xx}^{2}\delta_{xy}}{I[x]^{2}}+\frac{2\delta_{yy}^{2}\delta_{yx}}{I[y]^{2}}\right)\\ \delta_{yx}&=e^{u+v}\left((\delta_{xx}\delta_{yy}+\delta_{xy}\delta_{yx})\frac{\delta_{yx}e^{-2b}}{I[x]\,I[y]}+\frac{2\delta_{xx}^{2}\delta_{xy}}{I[x]^{2}}+\frac{2\delta_{yy}^{2}\delta_{yx}}{I[y]^{2}}\right)\end{aligned}$$

Where $I[x]=\delta_{xx}e^{u}+\delta_{xy}e^{u-b}$ and $I[y]=\delta_{yx}e^{u-b}+\delta_{yy}e^{u}$. This system allows for several other steady states that cannot all be characterized in closed form. However, making some extra assumptions we can get some intuition into the nature of this steady state. Assume that the networks for the two populations $x$ and $y$ are perfectly symmetric, formally $\delta_{xx}=\delta_{yy}$ and $\delta_{xy}=\delta_{yx}$. Therefore we can simply call $\delta_s=\delta_{xx}=\delta_{yy}$ for same-type number of links and $\delta_d=\delta_{yx}=\delta_{xy}$ for different-type number of links. The simplifying implication of this is that $I[x]=I[y]$. The system is in a partial steady state when

$$\delta_s=\frac{e^{2b}}{2}\delta_d \tag{8.9}$$

The steady state is partial in the sense that if condition (8.9) is satisfied, the ratio between $\delta_s$ and $\delta_d$ remains constant, but their values can be increasing over time. Notice that for $b=\ln(2)$, the number of same-type and different-type links is the same. For $b>\ln(2)$, $\delta_s>\delta_d$. The reason why the threshold for perfect integration of $b$ is not 0 is that interracial marriages have a larger effect on the diversity of the set of aquaintances than same-race marriages. The system is in complete steady state equilibrium - in the sense that both the total and the relative numbers of acquaintances remain constant in time - when $b=0$.

Concluding, the two steady states of the system are described by (8.10) and (8.11).

$$\delta_s=\delta\quad\delta_d=0 \tag{8.10}$$

$$\delta_s=\delta\quad\delta_d=2e^{-2b}\delta \tag{8.11}$$

## 8.2. The Impact of Online Dating on the Evolution of the Network

I have shown in the previous sections how online dating has likely contributed to the increase in the number of interracial marriages. Regardless of whether the cause of it has been an increase in exposure as explained in Table 4 or access to a more diverse pool of potential partners for agents with a lower cost of crossing racial and educational boundaries, the end result has been an increase in the number of interracial marriages. Since interracial marriages contribute more to the creation of mixed links in the network than same race marriages, an increase in the number of interracial marriages brought upon by online dating contributes to the diversity of the network. For the next result, assume again that $\delta_{xx} = \delta_{yy} = \delta_s$ and $\delta_{xy} = \delta_{yx} = \delta_d$.

**Proposition 8.** *An exogenous increase of magnitude $\varepsilon$ at time $t$ in the share of mixed marriages $\mu^t_{xy}$ causes in the following period an increase of size $(1-\lambda)(\delta_s - \delta_d)^2 \varepsilon$ in the number $\delta_d$ of mixed links in the network and a decrease of size $(1-\lambda)\left(\delta_d^2 - \delta_s^2\right)\frac{\varepsilon}{2}$ in the number $\delta_s$ of same type links.*

*Proof.* Consider two states of the world where the number of mixed marriages are respectively $\mu^1_{xy}$ and $\mu^0_{xy}$. Let $\Delta\mu_{xy} = \mu^1_{xy} - \mu^0_{xy} = \varepsilon$ and suppose that $\Delta\mu_{xx} = \Delta\mu_{yy} = -\frac{\varepsilon}{2}$ where $\Delta\mu_{xx}$ and $\Delta\mu_{yy}$ are similarly defined. Let $\delta^{t+1}_{d,1}$ and $\delta^{t+1}_{d,0}$ be the number of diverse links in the network at time $t+1$ in the two different states of the world, and assume $\delta^t_{d,0} = \delta^t_{d,1}$ Then

$$\begin{aligned}
\delta^{t+1}_{d,1} - \delta^{t+1}_{d,0} &= (1-\lambda)\left[\left(\delta_s^2 + \delta_d^2\right)\left(\mu^1_{xy} - \mu^0_{xy}\right) + 2\delta_s\delta_d\left(\mu^1_{xx} - \mu^0_{xx}\right) + 2\delta_s\delta_d\left(\mu^1_{yy} - \mu^0_{yy}\right)\right] \\
&= (1-\lambda)\left[\left(\delta_s^2 + \delta_d^2\right)\varepsilon - 2\delta_s\delta_d\frac{\varepsilon}{2} - 2\delta_s\delta_d\frac{\varepsilon}{2}\right] \\
&= (1-\lambda)\,\varepsilon\left[\delta_s^2 + \delta_d^2 - 2\delta_s\delta_d\right] \\
&= (1-\lambda)(\delta_s - \delta_d)^2\,\varepsilon
\end{aligned}$$

And similarly

$$\begin{aligned}
\delta^{t+1}_{s,1} - \delta^{t+1}_{s,0} &= (1-\lambda)\left(\delta_d^2\left(\mu^1_{xy} - \mu^0_{xy}\right) + \delta_s^2\left(\mu^1_{xx} - \mu^0_{xx}\right) + \delta_d^2\left(\mu^1_{yy} - \mu^0_{yy}\right)\right) \\
&= (1-\lambda)\left(\delta_d^2\varepsilon - \delta_s^2\frac{\varepsilon}{2} - \delta_d^2\frac{\varepsilon}{2}\right) \\
&= (1-\lambda)\left(\delta_d^2 - \delta_s^2\right)\frac{\varepsilon}{2}
\end{aligned}$$

□

Therefore the impact of an exogenous increase in interracial marriages depends both on the magnitude of the increase and on the current clustering of the network. The higher the network clustering, the higher the impact of an increase in the number of interracial marriages both on $\delta_d$ and $\delta_s$. If there is no clustering, that is if $\delta_s = \delta_d$, an increase in the number of interracial marriages does not bear any effecs on the network.

Using the result from Proposition 13, jointly with the estimates from Section 7, I can infer that the increase in the number of interracial marriages that was due to higher exposure to different races through online dating websites is equal to $\Delta\mu_{xy} = 0.067$. The remaining 12% increase in interracial marriages since the launch of online dating can be explained in my model through the increase in exposure due to the evolution of the network without online dating. Clearly, it is likely that there are other factors, such as the evolution of preferences, that have played a role in such increase. I am abstracting from those in my model.

If online dating was never launched, the estimated number of connections between different clusters of the network would be lower by $-(1-\lambda)\cdot 0.046$ and the estimated number of connections within cluster would be higher by $(1-\lambda)\cdot 0.028$. Therefore for instance, if $\lambda = 0.5$, the percentage of different-type connections would be lower by 2.3%, while the number of same-type connections would be higher by 1.4%.

## 9. Conclusion

In this paper I investigate the effects of online dating on society, both from a theoretical and empirical perspective. From the point of view of theory, I model online dating as a population game that allows agents to manipulate their network in order to maximize the expected utility of a match in the marriage market. The type of search the online dating allows an agent to do can be more or less coarse, which has an impact on their best reply. In the application, I consider a technology that allows to filter for race but not for education. The results of the theoretical model predict that agents will direct their search primarily to individuals of the same race, they will use race as an indicator for education (and other characteristics), they will focus their attention on larger groups and on groups that reciprocate their interest. The choice to enter or not in an online dating platform depends on the magnitude of the bias towards other races but also on the value of the exogenous inclusive value: if an agent has a sparse network, they are more likely to join the platform. If they have a low bias or derive a similar utility from many potential matches, they are also more likely to join. I show how the total effect on the network is not straightforward, and online dating could lead, from a theoretical perspective, both to social integration or social segregation. The total effect is determined by the magnitude of the racial and educational bias and on the current clustering of the network. When the bias is low and the network clustering is high, online dating is likely to have positive effects on social integration.

In the empirical part of the paper, I exploit Menzel (2015) identification strategy, enriching it to take into account meeting probabilities. I derive the estimation bias in the pseudo-surplus when the pseudo-surplus is estimated without taking into account the shape of the network and its impact on matching frequencies. I then move to estimation using data from How Couples Meet and they Stay Together. I group observations using k-mode clustering methods, using race, education, gender and region of the United States as variables to perform the clustering. I then use network estimates from the American National Social Network Survey to adjust the pseudo-surplus to take into account different meeting probabilities between agents belonging to the same race as opposed to different races in different parts of the country. The results show the presence of a significant cost of crossing a racial and an educational boundary across all regions of the United States, even after adjusting for network effects. Moreover, it is shown that online dating has a positive effect on the shares of both interracial and intereducational couples in the Northeast and the West. However that is not the case in the South and the Midwest for what concerns race, and no effect is found in the South when it comes to education. This might be suggesting the presence of an intense directed search put in place by users in those places. Furthermore, assuming that the observed difference is due to an increase in exposure on online dating platforms, I pin down the magnitude of such increase. Finally, I study the evolution of the network comparing scenarios with and without online dating, pinning down the resulting change in exposure to different races and in the number of interracial marriages that is caused by online dating.

## References


Adachi, H. (2000): "On a characterization of stable matchings," *Economic Letters*, (68), 43–49.

——— (2003): "A Search Model of Two-Sided Matching under Nontransferable Utility," *Journal of Economic Theory*, 113(2), 182–198.

Agarawal, N. (2015): "An Empirical Model of the Medical Match," *American Economic Review*, 105(7), 1939–1978.

Almagro, M., and T. Dominguez-Iino (2020): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," .

Ananat, E. O. (2011): "The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverity and Inequality," *American Economic Journal: Applied Economics*, 3(2), 34–66.

Baerveldt, C., M. Duijin, L. Vermeij, and D. Hemert (2004): "Ethnic Boundaries and Personal Choice: Assessing the Influence of Individual Inclinations to Choose Intra-Ethnic Relationships on Pupils' Networks," *Social Networks*, 26(55-74), 1005–1009.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–846.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The Rand Journal of Economics*, 25(2), 242–262.

Bisin, A., and G. Tura (2019): "Marriage, Fertility, and Cultural Integration in Italy," *CEPR Discussion Paper*, (No. DP14179).

Blackwell, D., and D. T. Lichter (2004): "Homogamy among Dating, Cohabiting, and Married Couples," *Sociological Quarterly*, 45(4), 719–737.

Choo, E., and A. Siow (2006): "Who marries whom and why," *Journal of Political Economy*, 114(1), 175–201.

Christakis, N., G. Fowler, G. Imbens, and K. Kalyanaraman (2010): "An Empirical Model for Strategic Network Formation," *Working Paper*.

Ciscato, E., A. Galichon, and M. Gousse (2018): "Like Attract Like? A Structural Comparison of Homogamy Across Same-Sex and Different-Sex Households," .

Currarini, S., M. Jackson, and P. Pin (2009): "An Economic Model of Friendship: Homophily, Minorities and Segregation," *Econometrica*, 77(4), 1003–1045.

Dagsvik, J. (1994): "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes," *Econometrica*, 62, 1179–1205.

——— (2000): "Aggregation in Matching Markets," *International Economic Review*, 41(1), 27–57.

Decker, C., E. Lieb, R. McCann, and B. Stephens (2013): "Unique Equilibria and Substitution Effects in a Stochastic Model of the Marriage Market," *Journal of Economic Theory*, 148(901), 778–792.

Dupuy, A., A. Galichon, S. Jaffe, and S. Kominers (2017): "Taxation in Matching Markets," *International Economic Review*.

Echenique, F., and J. Oviedo (2006): "A theory of stability in many-to-many matching markets," *Theoretical Economics*, 1, 233–273.

Echenique, F., and J. R.G. Fryer (2007): "A Measure of Segregation Based on Social Interactions," *Quarterly Journal of Economics*, 122(441-485), 1011.

Eeckout, J. (2000): "On the uniqueness of stable marriage matchings," *Economics Letters*, 69, 1–8.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006): "Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment," *Quarterly Journal of Economics*, 121(2), 673–697.

——— (2008): "Racial Preferences in Dating," *Review of Economic Studies*, 75(1), 117–132.

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

Galichon, A. (2016): *Optimal Transport Methods in Economics*. Princeton University Press.

Galichon, A., and A. Dupuy (2014): "Personality Traits and the Marriage Market," *Journal of Political Economy*, 122(6), 1271–1319.

Galichon, A., A. Dupuy, and Y. Sun (2016): "Estimating matching affinity matrix under low-rank constraints," .

Galichon, A., and B. Salanie (2020): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," .

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2014): "Marry Your Like: Assortative Mating and Income Inequality," *NBER Workin Paper*, (19829).

Hitsch, G. J., A. Hortaçsu, and D. Ariely (2010a): "Matching and Sorting in Online Dating," *American Economic Review*, 100(1), 130–163.

——— (2010b): "What makes you click? - Mate preferences in online dating," *Quantitative Marketing and Economics*, 8, 393–427.

Jackson, M. O. (2010): *Social and Economic Networks*. Princeton University Press.

Jacskon, M., and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71(1), 44–74.

Jaffe, S., and S. Weber (2018): "The Effects of Meeting Rates on Matching Outcomes," *Economic Theory*.

Luce, R. (1959): *Individual Choice Behavior*. Wiley.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001): "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–444.

Menzel, K. (2015): "Large matching markets as two-sided demand systems," *Econometrica*, 83(3), 897–941.

——— (2016): "Inference for Games with Many Players," *The Review of Economic Studies*, 1(294), 306–337.

Ortega, J. (2018): "Social Integration in Two-Sided Matching Markets," *Journal of Mathematical Economics*, 78(2002), 119–126.

——— (2019): "The losses from integration in matching markets can be large," *Economic Letters*, 174, 48–51.

Ortega, J., and P. Hergovich (2017): "The strength of absent ties: social integration via online dating," *Working Paper*.

Purcar, C. (2010): "On the wrong side of the track: railways as urban boundaries in the towns of the First Transylvanian Railway," *Urban History*, 37(1), 66–89.

Qian, Z., and S. H.Preston (1993): "Changes in American Marriage, 1972 to 1987: Availability and Forces of Attraction by Age and Education," *American Sociological Review*, 58(4), 482–495.

Roth, A., and M. O. Sotomayor (1990): *Two-Sided Matching*. Cambridge University Press.

Rudder, C. (2014): *Dataclysm: Love, Sex, Race, and Identity - What Our Online Lives Tell Us about Our Offline Selves*. Crown.

Schwartz, C. R., and R. D. Mare (2005): "Trends in Educational Assortative Marriage from 1940 to 2003," *Demography*, 42(4), 621–646.

Siow, A. (2010): "Testing Becker's Theory of Positive Assortative Matching," *Journal of Labor Economics*, 33(2), 409–441.

## Appendix A

### Proof of Proposition 2, Section 4.2

*Proof.* Recall that $\varepsilon$ is distributed as a Gumbel(0,1) which has c.d.f. $G(x) = \exp\{-\exp\{-x\}\}$. Therefore

$$\begin{aligned}
P(Z_i \leq z) &= P\left(\max_{j\in\mathcal{N}(i)} u_{x_i y_j} + \varepsilon_{ij} \leq z\right) \\
&= P\left(u_{x_i y_1} + \varepsilon_{i1} \leq z, ..., u_{x_i y_{\delta_x}} + \varepsilon_{i\delta_x} \leq z\right) \\
&= \prod_{j\in\mathcal{N}(i)} P\left(u_{x_i y_j} + \varepsilon_{ij} \leq z\right) \\
&= \prod_{j\in\mathcal{N}(i)} P\left(\varepsilon_{ij} \leq z - u_{x_i y_j}\right) \\
&= \prod_{j\in\mathcal{N}(i)} G(z - u_{x_i y_j})
\end{aligned}$$

where the equality between the second and third lines follows from independence across $j$'s of the shocks $\varepsilon_{ij}$. Plugging in the c.d.f. and taking the log of the probability

$$\begin{aligned}
\log P(Z_i \leq z) &= -\sum_{j\in O(i)\cap\mathcal{N}(i)} \exp\left\{u_{x_i y_j} - z\right\} \\
&= -\exp\left\{-z + \log \sum_{j\in\mathcal{N}(i)} \exp\left\{u_{x_i y_j}\right\}\right\} \\
&= -\exp\left\{-z + \log \sum_{y\in\mathcal{Y}} \sum_{\substack{j\in\mathcal{N}(i)\\ y_j = y}} \exp\left\{u_{x_i y}\right\}\right\} \\
&= -\exp\left\{-z + \log \sum_{y\in\mathcal{Y}} \delta_{xy} \exp\left\{u_{x_i y}\right\}\right\}
\end{aligned}$$

Notice that the last line is equal to the ln of the c.d.f. of a Gumbel with location parameter $\lambda = \log \sum_{y\in\mathcal{Y}} \delta_{xy} \exp\left\{u_{x_i y_j}\right\}$. Therefore the expected value of $Z_i$ is equal to $\mathbb{E}\left[Z_i \,|x_i = x\right] = \gamma + \log \sum_{y\in\mathcal{Y}} \delta_{xy} \exp\left\{u_{x_i y_j}\right\}$.

$\square$

### Proof of Proposition 13, Section 5.5

*Proof.* Suppose that at time $t_1$ there is a change in exposure $\varepsilon$ to people of different groups, and in particular assume that $p'_k = p_k(1+\varepsilon) - \varepsilon$ and $1 - p'_k = (1-p_k)(1+\varepsilon)$. The inclusive values for $i$ and $j$ after the change are

$$
\begin{aligned}
I_i^f(t_1) &= e^u \left[p_k(1+\varepsilon) - \varepsilon + (1-p_k)(1+\varepsilon)\, e^{-b}\right] \\
&= e^u (1+\varepsilon) \left[p_k + (1-p_k)\, e^{-b}\right] - \varepsilon e^u \\
&= (1+\varepsilon)\, e^u \left[p_k + (1-p_k)\, e^{-b}\right] - \varepsilon e^u \\
&= (1+\varepsilon) I_i^f(t_0) - \varepsilon e^u
\end{aligned}
$$

And similarly

$$
I_i^m(t_1) = (1+\varepsilon) I_i^m(t_0) - \varepsilon e^v
$$

This implies that the predicted share of matches after the change in exposure is given by

$$
\begin{aligned}
\mu_{ij}(t_1) &= \frac{(1-p_k)(1-q_{k'})(1+\varepsilon)^2 e^{u+v-2b}}{\left((1+\varepsilon) I_i^f(t_0) - \varepsilon e^u\right)\left((1+\varepsilon) I_i^m(t_0) - \varepsilon e^v\right)} \\
&= \frac{(1+\varepsilon)^2 (1-p_k)(1-q_{k'}) e^{u+v-2b}}{(1+\varepsilon)^2 \left(I_i^f(t_0) - \frac{\varepsilon}{1+\varepsilon} e^u\right)\left(I_i^m(t_0) - \frac{\varepsilon}{1+\varepsilon} e^v\right)} \\
&= \frac{(1-p_k)(1-q_{k'}) e^{u+v-2b}}{\left(I_i^f(t_0) - \frac{\varepsilon}{1+\varepsilon} e^u\right)\left(I_i^m(t_0) - \frac{\varepsilon}{1+\varepsilon} e^v\right)}
\end{aligned}
$$

Consider the relative change in matching frequencies

$$
\Delta\mu_{kk'} = \frac{\mu_{kk'}(t_1) - \mu_{kk'}(t_0)}{\mu_{kk'}(t_0)}
$$

After the change,

$$
\begin{aligned}
\Delta\mu_{kk'} &= \left(\frac{(1-p_k)(1-q_{k'}) e^{u+v-2b}}{\left(I_i^f(t_0) - \frac{\varepsilon}{1+\varepsilon} e^u\right)\left(I_i^m(t_0) - \frac{\varepsilon}{1+\varepsilon} e^v\right)} - \frac{(1-p_k)(1-q_{k'}) e^{u+v-2b}}{I_i^f(t_0) I_i^m(t_0)}\right) \times \frac{I_i^f(t_0) I_i^m(t_0)}{(1-p_k)(1-q_{k'}) e^{u+v-2b}} \\
&= \left(\frac{(1-p_k)(1-q_{k'}) e^{-2b}}{\left(e^{-b} + (1-e^{-b}) p_k - \frac{\varepsilon}{1+\varepsilon}\right)\left(e^{-b} + (1-e^{-b})\, q_{k'} - \frac{\varepsilon}{1+\varepsilon}\right)} - \frac{(1-p_k)(1-q_{k'}) e^{-2b}}{\left(e^{-b} + (1-e^{-b}) p_k\right)\left(e^{-b} + (1-e^{-b})\, q_{k'}\right)}\right) \\
&\times \frac{\left(e^{-b} + (1-e^{-b}) p_k\right)\left(e^{-b} + \left(1-e^{-b}\right) q_{k'}\right)}{(1-p_k)(1-q_{k'}) e^{-2b}} \\
&= \frac{\left(e^{-b} + (1-e^{-b}) p_k\right)\left(e^{-b} + \left(1-e^{-b}\right) q_{k'}\right) - \left(e^{-b} + (1-e^{-b}) p_k - \frac{\varepsilon}{1+\varepsilon}\right)\left(e^{-b} + \left(1-e^{-b}\right) q_{k'} - \frac{\varepsilon}{1+\varepsilon}\right)}{\left(e^{-b} + (1-e^{-b}) p_k - \frac{\varepsilon}{1+\varepsilon}\right)\left(e^{-b} + (1-e^{-b})\, q_{k'} - \frac{\varepsilon}{1+\varepsilon}\right)}
\end{aligned}
$$

Let $\tilde{I}_k = \left(e^{-b} + (1-e^{-b}) p_k\right)$ and notice that $\tilde{I}_k = I_k e^{-u}$

$$
\begin{aligned}
\Delta\mu_{kk'} &= \frac{\tilde{I}_k\tilde{I}_{k'} - (\tilde{I}_k - \frac{\varepsilon}{1+\varepsilon})(\tilde{I}_k - \frac{\varepsilon}{1+\varepsilon})}{(\tilde{I}_k - \frac{\varepsilon}{1+\varepsilon})(\tilde{I}_k - \frac{\varepsilon}{1+\varepsilon})} \\
&= \frac{\tilde{I}_k\tilde{I}_{k'} - \tilde{I}_k\tilde{I}_{k'} + \frac{\varepsilon}{1+\varepsilon}\tilde{I}_k + \frac{\varepsilon}{1+\varepsilon}\tilde{I}_{k'} - \left(\frac{\varepsilon}{1+\varepsilon}\right)^2}{(\tilde{I}_k - \frac{\varepsilon}{1+\varepsilon})(\tilde{I}_k - \frac{\varepsilon}{1+\varepsilon})} \\
&= \frac{\varepsilon}{1+\varepsilon} \cdot \frac{\tilde{I}_k + \tilde{I}_{k'} - \frac{\varepsilon}{1+\varepsilon}}{(\tilde{I}_k - \frac{\varepsilon}{1+\varepsilon})(\tilde{I}_k - \frac{\varepsilon}{1+\varepsilon})}
\end{aligned}
$$

□

**Proof of Steady State in Section 8.1**

Imposing $\delta_{yx} = \delta_{xy}$, we can further simplify the above expressions to

$$
\begin{aligned}
\delta_{xx} &= \left( \frac{\delta_{xx}\delta_{yx}^2 e^{-2b}}{(\delta_{xx} + \delta_{xy}e^{-b})\,(\delta_{yy} + \delta_{yx}e^{-b})} + \frac{\delta_{xx}^3}{(\delta_{xx} + \delta_{xy}e^{-b})^2} + \frac{\delta_{yx}^2\delta_{yy}}{(\delta_{yy} + \delta_{yx}e^{-b})^2} \right) \\
\delta_{yy} &= \left( \frac{\delta_{xy}^2\delta_{yy} e^{-2b}}{(\delta_{xx} + \delta_{xy}e^{-b})\,(\delta_{yy} + \delta_{yx}e^{-b})} + \frac{\delta_{xy}^2\delta_{xx}}{(\delta_{xx} + \delta_{xy}e^{-b})^2} + \frac{\delta_{yy}^3}{(\delta_{yy} + \delta_{yx}e^{-b})^2} \right) \\
\delta_{xy} &= \left( (\delta_{xx}\delta_{yy} + \delta_{xy}\delta_{yx})\, \frac{\delta_{xy} e^{-2b}}{(\delta_{xx} + \delta_{xy}e^{-b})\,(\delta_{yy} + \delta_{yx}e^{-b})} + \frac{2\delta_{xx}^2}{(\delta_{xx} + \delta_{xy}e^{-b})^2} + \frac{2\delta_{yy}^2}{(\delta_{yy} + \delta_{yx}e^{-b})^2} \right) \\
\delta_{yx} &= \left( (\delta_{xx}\delta_{yy} + \delta_{xy}\delta_{yx})\, \frac{\delta_{yx} e^{-2b}}{(\delta_{xx} + \delta_{xy}e^{-b})\,(\delta_{yy} + \delta_{yx}e^{-b})} + \frac{2\delta_{xx}^2\delta_{xy}}{(\delta_{xx} + \delta_{xy}e^{-b})^2} + \frac{2\delta_{yy}^2\delta_{yx}}{(\delta_{yy} + \delta_{yx}e^{-b})^2} \right)
\end{aligned}
$$

9.1. **Estimation.**

The following tables provide similar analysis for different grouping of ethnicities, in particular keeping Asian separate from other minorities and distinguishing between men and women in the estimation. In the first case I have estimated 52 utilities, defining subjects' categories based on ethnicity, gender and whether they met their partner online, and partners' types based on their ethnicity. Although the utility estimation might be noisy due to the limited amount of observations in certain ethnic groups (especially for asian and other and black people in the online bin) we can use the previous more robust estimation to intuit whether estimates are on the right track.

Table 9.1

| | *Dependent variable: Utility* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| White f partner | 2.125*** (1.802 , 2.449) | 1.837*** (1.547 , 2.126) | 2.062*** (1.790 , 2.334) |
| Non White f partner | 2.288*** (1.957 , 2.619) | | |
| Hispanic f partner | | 2.402*** (2.111 , 2.693) | 2.315*** (2.035 , 2.595) |
| Black f partner | | 1.915*** (1.599 , 2.230) | 1.830*** (1.521 , 2.138) |
| Asian f partner | | 2.588*** (2.299 , 2.877) | 2.499*** (2.224 , 2.774) |
| White m partner | 2.168*** (1.845 , 2.491) | 1.881*** (1.592 , 2.171) | 2.107*** (1.835 , 2.379) |
| Non White m partner | 2.227*** (1.895 , 2.559) | | |
| Hispanic m partner | | 1.840*** (1.543 , 2.137) | 1.733*** (1.447 , 2.020) |
| Black m partner | | 2.664*** (2.359 , 2.969) | 2.580*** (2.285 , 2.876) |
| Asian m partner | | 2.303*** (2.015 , 2.592) | 2.228*** (1.952 , 2.504) |
| White | 0.622*** (0.299 , 0.944) | 0.915*** (0.625 , 1.204) | 0.683*** (0.411 , 0.954) |
| Hispanic | 1.550*** (1.214 , 1.886) | 1.577*** (1.282 , 1.872) | 1.696*** (1.410 , 1.981) |
| Black | 1.497*** (1.155 , 1.840) | 1.404*** (1.096 , 1.712) | 1.529*** (1.226 , 1.833) |
| Asian or Other | 1.759*** (1.426 , 2.092) | 1.661*** (1.367 , 1.954) | 1.779*** (1.498 , 2.060) |
| Mixed | | -1.675*** (-1.723 , -1.626) | |
| Mixed Minority | -1.934*** (-2.008 , -1.860) | | -1.968*** (-2.032 , -1.904) |
| Mixed White | -1.306*** (-1.416 , -1.195) | | -1.288*** (-1.396 , -1.180) |
| Metonline | 0.865*** (0.847 , 0.883) | 0.879*** (0.858 , 0.899) | 0.879*** (0.858 , 0.900) |
| Mixed Metonline | 0.394*** (0.311 , 0.477) | 0.417*** (0.331 , 0.503) | 0.406*** (0.321 , 0.492) |
| Observations | 5,686 | 5,686 | 5,686 |
| $R^2$ | 0.988 | 0.991 | 0.991 |
| Adjusted $R^2$ | 0.988 | 0.991 | 0.991 |
| Residual Std. Error | 0.320(df = 5674) | 0.278(df = 5671) | 0.273(df = 5670) |
| F Statistic | 1789557.378*** (df = 12.0; 5674.0) | 1015374.521*** (df = 15.0; 5671.0) | 1383881.031*** (df = 16.0; 5670.0) |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Table 9.2

| | *Dependent variable: Utility* | |
|---|---|---|
| | (1) | (2) |
| White | 1.506*** | 0.963*** |
| | (1.220 , 1.792) | (0.804 , 1.121) |
| Hispanic | 2.447*** | 2.275*** |
| | (2.163 , 2.731) | (2.160 , 2.391) |
| Black | 1.917*** | 1.760*** |
| | (1.627 , 2.206) | (1.633 , 1.887) |
| Asian or Other | 2.363*** | 2.196*** |
| | (2.079 , 2.647) | (2.083 , 2.309) |
| White Partner | 1.552*** | 2.089*** |
| | (1.266 , 1.839) | (1.931 , 2.248) |
| Non White Partner | 2.080*** | 2.280*** |
| | (1.795 , 2.365) | (2.164 , 2.396) |
| Mixed | -1.904*** | |
| | (-1.945 , -1.863) | |
| Mixed White | | -1.491*** |
| | | (-1.596 , -1.385) |
| Mixed Minority | | -2.216*** |
| | | (-2.289 , -2.144) |
| Metonline | 0.830*** | 0.831*** |
| | (0.812 , 0.848) | (0.812 , 0.849) |
| Mixed Metonline | 0.450*** | 0.438*** |
| | (0.371 , 0.528) | (0.364 , 0.512) |
| Observations | 5,686 | 5,686 |
| $R^2$ | 0.897 | 0.903 |
| Adjusted $R^2$ | 0.897 | 0.903 |
| Residual Std. Error | 0.236(df = 5677) | 0.228(df = 5676) |
| F Statistic | nan*** (df = 8.0; 5677.0) | nan*** (df = 9.0; 5676.0) |

*Note:* *p<0.1; **p<0.05; ***p<0.01

9.2. **K-modes Clustering.**

The following table describes the clustering used for estimation in this paper, both for interviewed subjects and their partners. This particular clustering yields the results presented here, but the several other types of clustering have been used yielding similar results.

| | Name | Ethnicity | | | | | Education | | | $n$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | White | Black | Hispanic | Asian | Other | Low | Med | High | |
| Cluster 1 | NE-W Minority | 0.000 | 0.078 | 0.734 | 0.095 | 0.093 | 0.694 | 0.180 | 0.125 | 399 |
| Cluster 2 | NE Low Ed | 0.806 | 0.133 | 0.000 | 0.019 | 0.042 | 1.000 | 0.000 | 0.000 | 360 |
| Cluster 3 | NE High Ed | 0.718 | 0.099 | 0.092 | 0.063 | 0.028 | 0.000 | 0.000 | 1.000 | 142 |
| Cluster 4 | M White Low Ed | 0.855 | 0.086 | 0.000 | 0.011 | 0.049 | 1.000 | 0.000 | 0.000 | 571 |
| Cluster 5 | S Black | 0.000 | 0.884 | 0.000 | 0.033 | 0.083 | 0.681 | 0.232 | 0.087 | 276 |
| Cluster 6 | S White High Ed | 0.881 | 0.000 | 0.000 | 0.057 | 0.063 | 0.000 | 0.000 | 1.000 | 159 |
| Cluster 7 | S White Low E | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 540 |
| Cluster 8 | M Hispanic Med-Low Ed | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.432 | 0.405 | 0.162 | 37 |
| Cluster 9 | W White | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.783 | 0.000 | 0.217 | 474 |
| Cluster 10 | W Med Ed | 0.789 | 0.034 | 0.000 | 0.089 | 0.089 | 0.000 | 1.000 | 0.00 | 327 |
| Cluster 11 | NE Med Ed | 0.781 | 0.090 | 0.079 | 0.022 | 0.029 | 0.000 | 1.000 | 0.00 | 278 |
| Cluster 12 | S Hispanic Med-Low Ed | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.588 | 0.342 | 0.070 | 187 |
| Cluster 13 | S White Med Ed | 0.946 | 0.000 | 0.000 | 0.020 | 0.034 | 0.000 | 1.000 | 0.000 | 355 |
| Cluster 14 | M Med Ed | 0.873 | 0.084 | 0.000 | 0.022 | 0.022 | 0.000 | 1.000 | 0.000 | 322 |
| Cluster 15 | M High Ed | 0.906 | 0.043 | 0.000 | 0.034 | 0.017 | 0.000 | 0.000 | 1.000 | 117 |

9.2.1. *5 Ethnic Groups.*

For the sake of completeness, I will include here OLS regression for utility calculated on the sample based on the following characteristics: ethnicity and partner ethnicity, which include White, Black, Hispanic, Asian and Other Minorities, and a dummy variable for whether they met online or offline. In the regression, we regress the values of utility $U$ so estimated on dummy variables for the ethnicity of the subject and their partner, a dummy variable for whether they are an ethnically mixed couple and a dummy variable for whether they have met online. The coefficient associated with a couple being mixed is negative, seemingly suggesting the existence of both a network effect and a social cost in crossing an ethnic boundary in a romantic relationship.

Note however that due to the limited size of the sample at hand, the utility estimates might be noisy for some specific combinations of couples. While we can consider the utility estimates for White people and for the offline community reliable given the number of observations in each bin, the utility of specific combinations of couples among Black, Hispanic and Asian individuals in the sample who have met their partner online might be noisy. As suggestive as the estimates are, this might be problematic in that it will lead to measurement errors in the regressions assessing specific effects of ethnicity and online dating in the generated surplus.

With that in mind, the table below provides interesting evidence of the network effect and social cost associated with crossing an ethnic boundary, and how the network effect might be mitigated by the presence of online dating.

*Table 4: Estimated Utility for Women Online*

| | White | Hispanic | Black | Asian | Other |
|---|---|---|---|---|---|
| White | 3.026 | 2.296 | 2.107 | 2.322 | 2.140 |
| Hispanic | 2.675 | 4.501 | 2.322 | -$inf$ | 3.914 |
| Black | 1.151 | 1.539 | 3.835 | 3.032 | 2.339 |
| Asian | 3.015 | -$inf$ | -$inf$ | 5.407 | 4.021 |
| Other | 2.728 | 3.221 | 2.339 | -$inf$ | 3.328 |

*Table 5: Estimated Utility for Men Online*

| | White | Hispanic | Black | Asian | Other |
|---|---|---|---|---|---|
| White | 2.946 | 2.677 | 1.284 | 3.882 | 1.174 |
| Hispanic | 2.977 | 4.077 | 2.183 | 3.865 | 2.479 |
| Black | 1.438 | 2.589 | 3.940 | 2.983 | -$inf$ |
| Asian | 2.560 | 3.172 | 2.983 | 6.456 | -$inf$ |
| Other | 2.427 | 3.172 | 3.388 | 5.070 | 3.278 |

| | *Dependent variable: Utility* |
|---|---|
| | (1) |
| White | 1.489$^{***}$<br>(1.295 , 1.683) |
| Hispanic | 2.430$^{***}$<br>(2.234 , 2.626) |
| Black | 1.985$^{***}$<br>(1.777 , 2.194) |
| Asian or Other | 2.285$^{***}$<br>(2.089 , 2.481) |
| White Partner | 1.569$^{***}$<br>(1.375 , 1.763) |
| Hispanic Partner | 2.110$^{***}$<br>(1.916 , 2.303) |
| Black Partner | 1.976$^{***}$<br>(1.767 , 2.185) |
| Asian or Other Partner | 2.257$^{***}$<br>(2.063 , 2.450) |
| Mixed | -1.881$^{***}$<br>(-1.918 , -1.845) |
| Metonline | 0.830$^{***}$<br>(0.811 , 0.848) |
| Mixed Metonline | 0.446$^{***}$<br>(0.377 , 0.515) |
| Observations | 5,686 |
| $R^2$ | 0.996 |
| Adjusted $R^2$ | 0.996 |
| Residual Std. Error | 0.202(df = 5675) |
| F Statistic | 3949686.646$^{***}$ (df = 11.0; 5675.0) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 6: *Estimated Utility for Women Offline*

| | White | Hispanic | Black | Asian | Other |
|---|---|---|---|---|---|
| White | 3.120 | 1.881 | 0.811 | 1.690 | 1.898 |
| Hispanic | 2.132 | 4.520 | 2.737 | 2.509 | 2.039 |
| Black | 0.083 | 2.149 | 4.462 | 2.320 | 2.187 |
| Asian | 2.622 | 2.915 | $-inf$ | 6.112 | 1.923 |
| Other | 2.493 | 2.039 | 2.880 | 4.002 | 3.627 |

Table 7: *Estimated Utility for Men Offline*

| | White | Hispanic | Black | Asian | Other |
|---|---|---|---|---|---|
| White | 3.125 | 1.944 | 0.159 | 2.894 | 1.966 |
| Hispanic | 2.623 | 4.528 | 1.846 | 3.327 | 2.634 |
| Black | 0.629 | 2.493 | 4.080 | 1.634 | 2.328 |
| Asian | 1.435 | 2.740 | 0.941 | 6.207 | 3.029 |
| Other | 2.128 | 1.536 | 1.347 | 3.316 | 4.281 |

| | *Dependent variable: U* | |
|---|---|---|
| | OLS | Robust |
| | (1) | (2) |
| White | 1.431*** | 1.676*** |
| | (1.336 , 1.526) | (1.676 , 1.676) |
| Hispanic | 2.250*** | 2.531*** |
| | (2.153 , 2.347) | (2.531 , 2.531) |
| Black | 1.953*** | 2.408*** |
| | (1.855 , 2.052) | (2.408 , 2.408) |
| Asian or Other | 2.382*** | 2.662*** |
| | (2.284 , 2.481) | (2.662 , 2.662) |
| White Partner | 1.584*** | 1.331*** |
| | (1.488 , 1.679) | (1.331 , 1.331) |
| Hispanic Partner | 2.222*** | 1.877*** |
| | (2.125 , 2.318) | (1.877 , 1.877) |
| Black Partner | 1.979*** | 1.771*** |
| | (1.880 , 2.077) | (1.771 , 1.771) |
| Asian or Other Partner | 2.069*** | 1.820*** |
| | (1.972 , 2.166) | (1.820 , 1.820) |
| Mixed | -1.825*** | -1.676*** |
| | (-1.842 , -1.808) | (-1.676 , -1.676) |
| Metonline | 1.583*** | 1.582*** |
| | (1.565 , 1.600) | (1.582 , 1.582) |
| Observations | 5,686 | 5,686 |
| $R^2$ | 0.996 | |
| Adjusted $R^2$ | 0.996 | |
| Residual Std. Error | 0.215(df = 5676) | 0.000(df = 5676) |
| F Statistic | 129798.359*** (df = 10.0; 5676.0) | (df = 9.0; 5676.0) |
| *Note:* | | S=1 *p<0.1; **p<0.05; ***p<0.01 |

| | | White (1) | White (2) | White (3) | White (4) | Hispanic (1) | Hispanic (2) | Hispanic (3) | Hispanic (4) | Black (1) | Black (2) | Black (3) | Black (4) | Asian (1) | Asian (2) | Asian (3) | Asian (4) | Other | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| White | (1) | 3.655 | 3.000 | 2.590 | 2.237 | 2.255 | 1.812 | 1.645 | 1.251 | 0.973 | 0.820 | -0.964 | -0.161 | 1.556 | 2.586 | 1.225 | 1.711 | 2.270 | 2.404 | 0.654 | 0.900 |
| | (2) | 2.620 | 2.985 | 2.864 | 2.440 | 1.958 | 2.088 | 1.529 | 1.454 | 0.857 | 1.582 | -0.761 | -inf | 1.759 | 2.383 | 2.526 | 1.913 | 0.527 | 2.925 | 1.550 | -inf |
| | (3) | 2.439 | 2.764 | 3.777 | 3.663 | 1.520 | 1.749 | 2.665 | 3.060 | -0.126 | 1.579 | 1.316 | 0.732 | 2.044 | 3.633 | 3.622 | 1.910 | 1.623 | 1.910 | 2.107 | 2.486 |
| | (4) | 1.692 | 2.588 | 3.700 | 4.589 | 0.599 | 1.617 | 2.533 | 2.746 | 0.252 | -inf | 0.020 | 2.769 | 2.135 | 3.164 | 2.902 | 4.640 | 1.714 | -inf | 2.331 | 2.982 |
| Hispanic | (1) | 2.834 | 2.344 | 1.607 | 1.852 | 5.156 | 4.839 | 4.040 | 4.002 | 2.781 | 2.435 | -inf | 2.435 | -inf | 3.391 | 2.435 | 2.921 | 2.633 | 2.921 | 1.865 | 3.496 |
| | (2) | 1.955 | 2.445 | 2.491 | 2.022 | 4.285 | 4.519 | 4.470 | 3.073 | 2.189 | 3.452 | 2.650 | -inf | -inf | 4.408 | 3.452 | -inf | 2.957 | -inf | 2.882 | 3.127 |
| | (3) | 1.886 | 2.222 | 3.297 | 2.939 | 4.146 | 4.307 | 4.971 | 4.577 | 1.496 | 2.354 | 2.873 | 3.452 | 3.378 | 4.813 | 3.858 | -inf | 2.552 | -inf | -inf | 3.127 |
| | (4) | 0.846 | 1.741 | 2.837 | 3.969 | 3.154 | 2.380 | 4.172 | 3.286 | -inf | -inf | -inf | 3.665 | -inf | -inf | -inf | -inf | 2.764 | -inf | 3.095 | -inf |
| Black | (1) | 0.521 | -0.124 | 0.567 | 0.658 | 2.675 | 2.189 | 2.189 | 2.401 | 4.562 | 4.321 | 3.637 | 2.493 | 2.014 | -inf | -inf | -inf | 2.286 | -inf | -inf | 2.455 |
| | (2) | -inf | 1.022 | 1.243 | 1.110 | 2.723 | -inf | 2.759 | -inf | 4.627 | 4.738 | 4.818 | 3.862 | 2.872 | -inf | -inf | -inf | 3.144 | 3.431 | -inf | -inf |
| | (3) | -0.964 | 0.338 | 0.335 | 1.407 | -inf | 3.343 | 2.873 | -inf | 3.300 | 4.818 | 4.312 | 4.254 | -inf | 4.198 | 3.242 | 3.035 | 1.648 | 3.035 | -inf | 2.917 |
| | (4) | -0.161 | 0.041 | 1.830 | 0.823 | -inf | -inf | 2.066 | 2.972 | 2.781 | 4.199 | 4.053 | 5.223 | -inf | -inf | -inf | 3.837 | -inf | 3.144 | 2.781 | -inf |
| Asian | (1) | 1.844 | -inf | 2.267 | 1.442 | 3.747 | 2.685 | 2.685 | -inf | -inf | -inf | -inf | -inf | 5.283 | 6.312 | 5.069 | -inf | -inf | -inf | -inf | -inf |
| | (2) | 2.181 | 2.383 | 2.380 | 2.471 | 3.391 | -inf | -inf | -inf | -inf | 3.901 | -inf | -inf | 4.926 | 5.956 | 5.000 | -inf | -inf | -inf | -inf | -inf |
| | 3 | 1.225 | 2.121 | 2.678 | 3.308 | 3.128 | 2.759 | 2.759 | -inf | -inf | -inf | 2.549 | -inf | 5.069 | 6.099 | 6.817 | 6.833 | -inf | 3.837 | 4.167 | -inf |
| | (4) | -inf | 2.319 | 3.414 | 3.793 | -inf | -inf | 3.245 | -inf | -inf | -inf | -inf | -inf | -inf | 5.486 | 7.303 | 7.724 | -inf | -inf | -inf | 5.591 |
| Other | (1) | 2.464 | 2.473 | 0.930 | -inf | 1.534 | 1.859 | -inf | -inf | 2.573 | 2.045 | -inf | 2.451 | -inf | 4.099 | 3.144 | -inf | 4.035 | 3.629 | 2.573 | 3.511 |
| | (2) | 1.711 | 1.626 | 2.721 | 2.694 | 2.921 | 3.245 | 2.552 | -inf | 1.880 | 3.837 | 2.341 | 3.144 | -inf | 5.486 | 3.837 | 4.322 | 3.629 | 4.728 | 3.266 | 4.205 |
| | (3) | 1.348 | 2.243 | 2.934 | 3.025 | 1.865 | 2.189 | -inf | 3.095 | 1.517 | 3.069 | 1.979 | 3.474 | -inf | 4.430 | 4.860 | 3.960 | 3.266 | 4.876 | 4.695 | -inf |
| | (4) | 2.286 | 1.795 | 2.891 | 4.368 | -inf | -inf | 3.127 | 4.033 | 2.455 | 3.314 | 2.917 | 3.719 | -inf | -inf | -inf | 5.996 | 3.511 | 4.205 | 4.535 | 4.780 |