An Empirical Investigation of Online Dating as a Marketplace

Final Paper

Andrea Domenico Antonacci
Student ID: 3004329

Academic Year 2017/2018

Advisor: Professor Francesco Candeloro Billari

Università Commerciale "Luigi Bocconi"
Bachelor of Science in International Economics and Management

To my unbelievably supportive parents.
For their love and tireless efforts in life that inspire me everyday.

Contents

Abstract

This thesis explores the efficiency gains in partner markets deriving from the diffusion of online dating. Following an introduction on its principles and functioning, we apply a rational choice approach to assess the consequences of online dating on mate choice processes from a theoretical standpoint. To give an illustration of the concept of competition for attention, and of the logics of exchange, network externalities, supply and demand, we take advantage of the "marketplace" metaphor. An empirical analysis on Pew Research Center's "Gaming, Jobs and Broadband" 2015 data set is run to test two hypotheses on the usage of online dating. Results are then discussed in the light of prior extensive research. We find that those who are more likely to be online daters are mostly young White males, with above-average socioeconomic status, a university degree and more liberal views. Thus, based solely on the available data, we dismiss the hypothesis of the primacy of online dating for those in a thin market for potential partners.

1. Introduction and Literature Review

1.1. A Solution to Offline Dating Limitations

The rise of the Internet and its effects on behaviour are undeniably pervading across all social strata. Online dating can be considered one of the foremost examples of a fundamentally altered social activity in the modern Internet era: finding a partner nowadays can be accomplished through Computer-Mediated Communication (CMC) in a faster – and usually more efficient – way with respect to traditional meeting venues. The Internet allows for free and instant access to worldwide information, going beyond the limits of physical space and time of offline activities. The same concept applies to Online Dating which, contrary to conventional meeting venues with soundly structured social spheres, allows its users to surpass social class, time and even geographical barriers.

This work is grounded in the assumption that the emergence of online dating is a natural consequence of the modernisation processes that lead to the rationalisation, marketisation[1] and ultimately to the "commodification of romance" – i.e., of choosing potential mates (Illouz, 1997) – that is quintessentially Western culture. This is reflected in the increasing significance of online dating in daily life, easily ascertainable by each own personal experiences. Official data on the usage of online dating apps is often unavailable mainly due to privacy concern, but it is out of doubt that online dating is penetrating all social classes and age groups: one in three marriages in the U.S. now[2] begins online (Cacioppo et al., 2013). Match Group, owning a portfolio of more than 45 online dating

[1] Please note that the marketisation of love and intimacy has not started with online dating. In fact, as Schmitz (2016) points out, common meeting venues have gradually shifted from family and neighbourhood to more explicit partner markets (like nightclubs) from the 50s onwards.

[2] This could be an underestimation, since data from 2013 are probably outdated nowadays.

brands – including Tinder, the most downloaded dating mobile app worldwide – reported $1.3 billion revenues in 2017 and boasts a current market capitalisation of $10.5 billion[3]. These data can be considered a proxy measure of this rapidly growing industry and a signal of these emerging social phenomena. Analysing the online partner market seems crucial to retrieve insights on usage, behaviours and mate choice preferences, especially since these new tools can shape how people date or fall in love in the digital era and hence the demography of our communities. Not only that, online dating can also help in surpassing the classical (micro-)economic theories of mate selection rooted in the pre-Internet era (Becker, 1991), when the problem of search in mate selection had no clear answers due to the absence of proper data (Rosenfeld and Thomas, 2012). Hence this work aims at providing a better understanding of the online partner market, and it is structured as follows: firstly, an introduction to online dating seen as an efficient solution to the issues of offline partner markets is elaborated; secondly, theories on its functioning – as a marketplace in particular – are discussed and its main characteristics are outlined; finally, an empirical analysis is run on a 2,001 U.S. adults data set, in order to shape common online dating user profiles and assess the determinants of its usage.

Online dating can be defined as the practice of using Internet to find a short-term sexual or a long-term romantic partner. Thereby an online dating platform is a website or mobile app that allows its users to find their best possible partner, through presenting themselves usually with a short biography, photos and other kinds of information, contacting potential mates via chat and picking promising profiles out of a large pool of contacts. Although different models can be identified, most online dating websites and mobile apps can be traced back to two main functioning standards that will be analysed later on (see 1.4.3). Often, very little differentiation can be found among them; however new trends of niche apps are emerging, with different concepts and targeting different market segments – homosexual dating, elderly dating, religious dating, etc. By definition, online dating could be extended to other websites and mobile apps like Facebook, as people could potentially flirt or fall in love on platforms with a great variety of purposes; for the sake

[3]Financial data retrieved from Bloomberg and Q1 2018 Investor Presentation of Match Group.

of simplicity, in this work we will only consider and refer to dating platforms that have been intentionally designed for this very purpose. It is important to notice that all the concepts apply both to online dating and mobile dating, as the difference between the two is becoming negligible (thus we will only refer to "online dating" in general from now on): as roughly 57% of Internet traffic comes from mobile and tablet devices nowadays[4], more and more people expect a mobile version of desktop websites as well as a mobile app, inducing dating services to become ubiquitous.

People used to meet their partners through mutual friends, workplace, church, neighbourhood, school and family. In the last 15-20 years, the Internet has partially displaced those conventional offline mating venues. It is common for people to partner with complete strangers[5] nowadays, as a consequence of the increased efficiency brought about by online dating platforms (Rosenfeld and Thomas, 2012). This unprecedented search efficiency is particularly convenient to those singles[6] who would otherwise lack an easy access to the partner market (see Figure 1.1 on page 5).

As social class and geographical barriers are often reduced by the way online dating platform are designed, their users can realise their preferences more efficiently. The effects of rationalisation and the intrusion of market-driven logics on intim acy and cyberdating lead the modern user to take everyday decisions based on cost-benefit ratios. User's actions on digital dating platforms are often driven by their own preferences and rational intentions. When compared to more traditional meeting venues – commonly based on structured social spheres[7] – online dating appears strongly influenced by market principles: exchange logics, objectification and competition for potential partners, a large base of "suppliers", lower transaction costs and conscious rational mate choices make it metaphorically similar to a marketplace (Schmitz, 2016). As a matter of fact, it can be considered rather close to the "ideal-typical" partner market from an economic standpoint. In addition to this,

[4]Data retrieved from BrightEdge (2017).

[5]That is to say with people to which they had no social ties previously.

[6]The definition of "single" is a rather vague. For the sake of simplicity, and unless otherwise noted, in this work we will adopt the statistical definition, which simply means unmarried, but leaves out plenty of family structures – e.g., couples.

[7]For example: bars and restaurants, schools, nightclubs and churches are often frequented by different people. Each meeting venue is usually frequented by a socially homogeneous group of people.

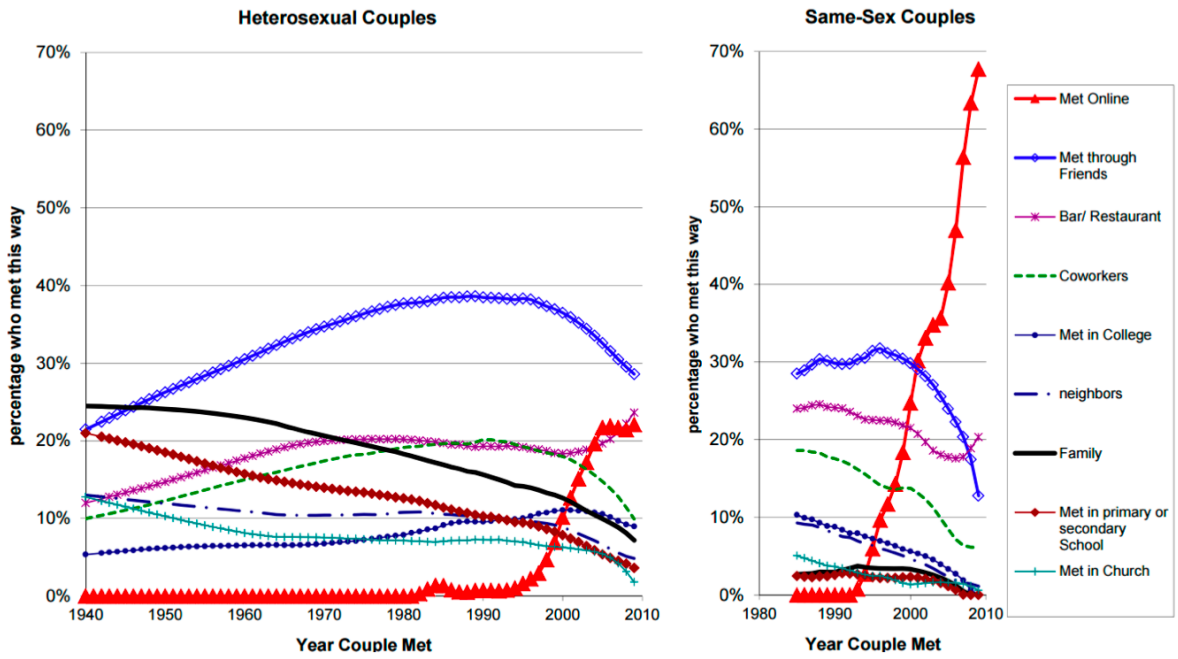

Figure 1.1.: The Changing Way Americans Meet Their Partners
N = 2,462 for heterosexual couples, N = 462 for same-sex couples.
"Individuals who face a thin market for potential partners, such as gays, lesbians, and middle aged heterosexuals, are especially likely to meet partners online."
Source: Rosenfeld and Thomas (2012).

online dating can be interpreted as a promising tool to demolish gender roles and norms: Scharlott and Christ (1995) argue that "the safety and anonymity the system offered" allows online daters to "break free from traditional sex role norms".

Following the realisation that these platforms are designed with the precise goal of efficient partner choice, some see the intrusion of market logics as "emotional capitalism", a rationalist threat to true romance that ultimately should not have economic motives (Illouz, 2007). However, in this paper the author tries to apply a value-free and unbiased perspective both to traditional and online dating concepts.

1.2. Early Stages in Social Network Analysis

In pursuance of explaining the efficiency gains that are said to be a strength of online dating, we start by deepening further the social network analysis.

In one of his well-known books, Moreno (1934) theorised the "sociogram" as a tool to

represent the formal properties of social configurations, built with diagrams whereby individuals are represented by points in space and their social relationships by lines. Nowadays this concept is well established and commonly used in academia. The concept of "balance" was later introduced in the sociogram after the development of the balance theory by Heider (1946). He argues that congruence among attitudes with other people is fundamental in order to achieve cognitive balance. That is to say that *ego's* balance is achieved only if in a set of attitudes towards people (*alter*), they are all similar in their sign – i.e., all attitudes are either positive or negative, at their simplest. In other words, given the fact that a person A likes[8] the person B and the person B likes the person C, balance in the triad is achieved only if A likes C as well. The concept of cognitive balance was later extended at interpersonal level in social groups, thus allowing researchers to study whole structures simultaneously and not from one *ego's* point of view at a time. A new framework developed by Cartwright and Harary (1956) builds on Moreno's work and applies the concepts of graph theory to the sociogram. They argue that the lines representing the relations among individuals in a graph could be provided with a sign + or – to indicate either a positive or a negative relation and an arrow to illustrate the direction of such relationship. For instance, it is possible to indicate that person A has a positive relation with person B (A likes B, or $A\mathbf{L}B$) but person B has a negative relation with person A (B dislikes A, or $B\sim \mathbf{L}A$). In the case of triadic relationships, the balanced state can be achieved only if all relations are positive, or if two are negative and one is positive. Thus ($A\mathbf{L}B$, $B\mathbf{L}C$, $A\mathbf{L}C$), ($A\sim \mathbf{L}B$, $B\sim \mathbf{L}C$, $A\mathbf{L}C$) are examples of balanced triads, while ($A\mathbf{L}B$, $B\mathbf{L}C$, $A\sim \mathbf{L}C$) is not (see Figure 1.2 on page 7). In undirected graphs (with no arrows) the relationships between A and B and between B and A are considered to be equal or perfectly reciprocal. This results in a (un)balanced state given only by the signs attached to the relations in the graph, irrespectively of the directions of such relations. These triads are of particular importance. According to Cartwright and Harary, actual and intricate social structures can be imagined as being composed of overlapping simple triads. Hence these triads are the elementary units of more complex

[8]Likes, loves, values or approves, as Cartwright and Harary (1956) put it.

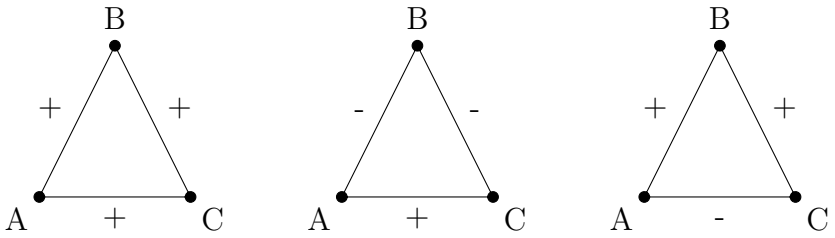

Figure 1.2.: Some triadic structures examples.

social structures seen in real life, and should be analysed in order to infer the properties of complex networks. For instance, a large network is balanced if all its elementary triads are balanced. Large interpersonal networks can be studied with a great variety of metrics. In his social network development analysis, Scott (1991) gathers and classifies these metrics following the codification work of Mitchell (1969). For the sake of this work we put emphasis on *homophily*, that is the extent of actors to forge relationships with (dis)similar individuals; *propinquity*, the propensity to tie with physically close actors; *reciprocity*, the degree to which actors are mutually interested and interact with each other; *intensity*, that refers to the strength of obligations arisen in a relationship; *density*, which stands for the proportion of ties present in a network compared to the number of all potentially possible ties – i.e., the completeness of the network.

In the next section we build on these foundations to elaborate efficiency theories based on the strength of ties in social networks.

1.3. Weak and Absent Ties Hypotheses

Strength is another essential characteristic of interpersonal ties in social networks. Granovetter (1973) defines the tie strength as:

> "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterise the tie[9]."

Consequently, three main categories of tie strengths can be identified: strong ties, weak ties and absent ties. Typically a strong tie is correlated with high degrees of homophily

[9]Assuming positive and symmetric ties.

and propinquity, as opposed to weak ties that serve as "bridges" between clumps of individuals with strong ties. Starting from his revolutionary "The Strength of Weak Ties", Granovetter (1973) radically influenced the concept of ties in social networks, posing particular attention to weak and absent ties[10]. The weak tie hypothesis suggests that, according to probability reasoning, if A has a strong tie both with B and C, then a (probably weak) tie between B and C is also present – or is very unlikely to be absent (see Figure 1.3). If that was not the case, a *forbidden triad* would arise[11].

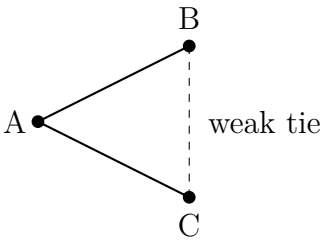

Figure 1.3.: An example of weak tie in a triadic relationship.

Moreover, Granovetter argues that weak ties are those responsible for transmitting most information across individuals and through social networks (see Figure 1.4). Surprisingly, more novel information is said to flow through weak ties rather than through strong ties[12].

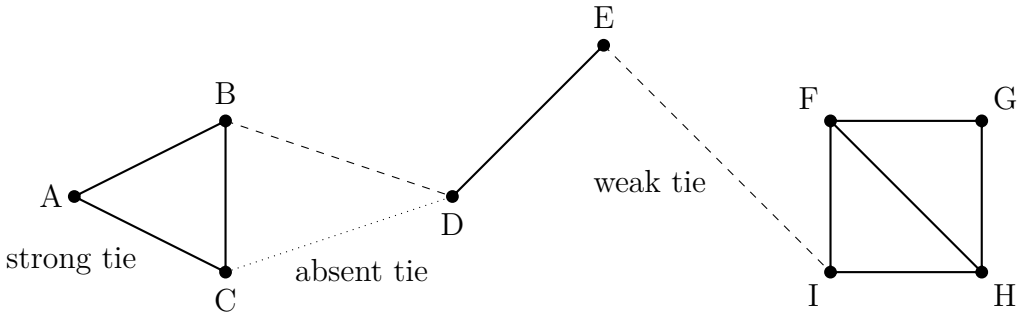

Figure 1.4.: Some examples of ties in a network.

With his "Getting a Job", Granovetter applied those concepts to explore how people find a new job, or more specifically, how they collect information about job opportunities.

[10]Those ties with no particular significance: e.g., the neighbour you wave when back home.

[11]Imagine that person A likes and values the celebrity B. If B likes and endorses a product C, then it is very likely that A will start liking C as well – or disliking B – to achieve balance.

[12]To illustrate this better think of close friends or family members: they are all in the same circles that we are in, thus it is rare to receive novel information from them, since it would overlap with the information we already know. On the other hand, acquaintances know different people from our circles, thus it is easier to receive novel information from them.

Results from his work were consistent with the weak tie hypothesis. It showed that the least important people in providing new job opportunities information were friends and family. Since they interact with the same overlapping circles of contacts, they tend to possess the same information. This is not to say that strong ties are not important. In fact, when a new piece of information reaches anyone in the circle, it spreads easily and is very likely to pass on everyone with a strong tie. However, these people are unlikely to be the source of new information coming from distant nodes in the network. On the contrary, weak ties with acquaintances far in the network – e.g., in different work situations – make these people far more likely to pass on new information (Granovetter, 1974).

Ortega and Hergovich (2017) have built upon these notions and applied them to investigate social integration outcomes resulting from online dating. Their work focuses on explaining the effects of online dating on racial diversity, by virtue of the "strength of absent ties". They investigate how interracial marriages, a widely-accepted measure of social distance (Qian, 2002; Wong, 2003), have changed as a result of online dating penetration in the American society. Their resulting model not only predicts almost complete racial integration with the rise of online dating, but also slightly stronger marriages in society overall[13]. This theoretical framework is backed up by empirical U.S. data: there is evidence that the percentage of interracial marriages has significantly increased with the diffusion of online dating platforms (see Figure 1.5 on page 10), and that divorce rate of marriages started online has also diminished. Although the authors do not claim that these phenomena are directly caused by the spreading of online dating, they also highlight that this observed increase in interracial marriages cannot be solely explained by the recent changes in the American population composition. Notwithstanding that a simple theoretical model could fail at grasping all facets of complex social processes and at providing explanation of causality effects, given the above, it can be concluded that the diversity of society should anyway increase dramatically with the diffusion of online dating.

[13]That is to say that, "on average, marriages created when online dating becomes available last longer than those created in societies without this technology" (Ortega and Hergovich, 2017).

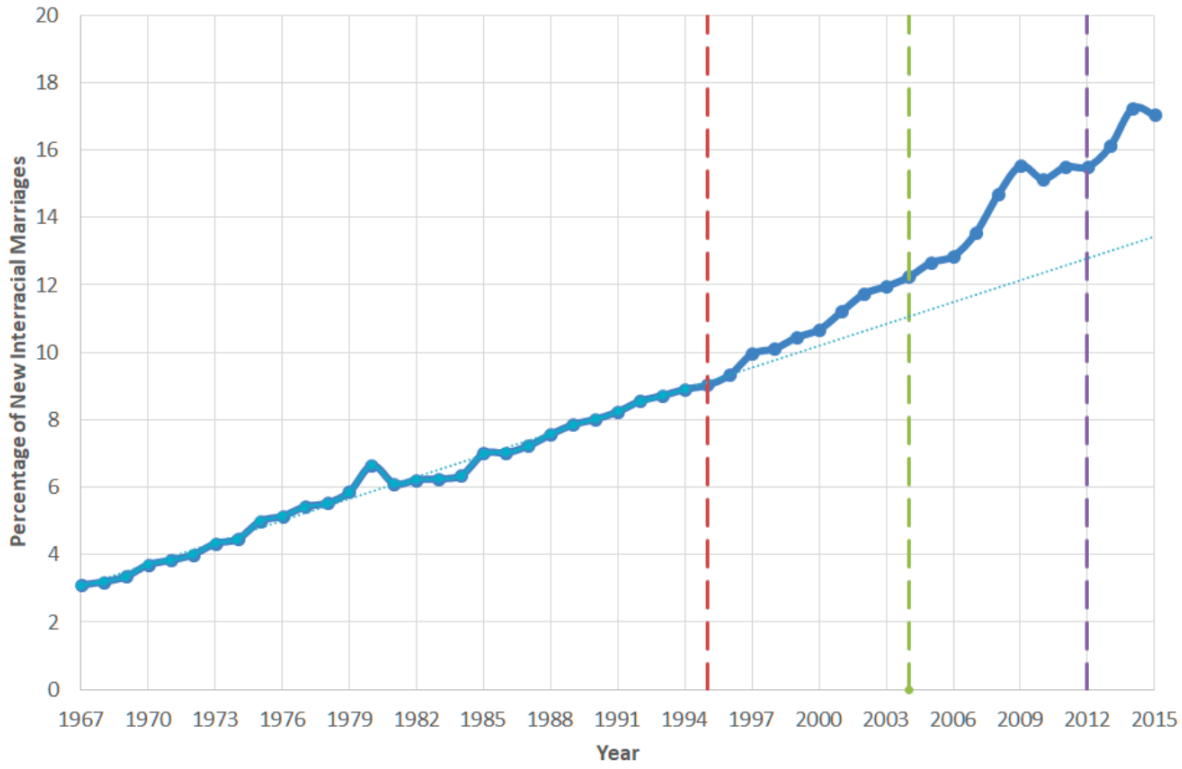

Figure 1.5.: Percentage of new interracial marriages in the U.S.
"The red, green, and purple lines represent the creation of Match.com, OKCupid, and Tinder, three of the largest dating websites. The blue line represents a linear prediction for 1996 – 2015 using the data from 1967 to 1995."
Data retrieved from Pew Research Center analysis of 2008-2015 American Community Survey and 1980, 1990 and 2000 decennial censuses (IPUMS).
Source: Ortega and Hergovich (2017).

Having scrutinised the efficiency gains at societal levels, in the next sections we investigate the efficiency of online dating at micro-level, by inspecting the its basic characteristics and comparing it to a metaphorical marketplace.

1.4. Principles of Online Dating

What we elaborate in this section is supposed to refer to general dating platforms that potentially represent all social strata existing in society. This helpful simplification allows us to ignore specialised dating platforms that involve only few segments of the population.

Two opposing underlying logics can be identified in the mate search process: subconscious decisions and instrumental rationality. Willis and Todorov (2006) found that when draw-

ing "trait inferences from the facial appearance of other people", people form a judgement and a first impression in only 100 milliseconds. Thus we would be induced to reject the assumptions of instrumental rationality in the process of mate selection. However this should not be necessarily the case, since online dating seems to be heavily influenced by such logics of choice, exchange and competition in a market, firmly relying on the rationale of supply and demand (see Section 1.5).

Physicality is a crucial factor in explicit partner markets like dating platforms or nightclubs. In the context of online dating, it is primarily expressed through the use of profile pictures, to which is given great relevance by the user experience design. Due to the fierce competition for attention, users must apply selective criteria of choice and reduce complexity in decision making, both verbally and visually, hence the paramount power that profile pictures exercise on mate choice processes. For instance, when users are looking for a partner, they might be tempted to flip through profiles on a dating platform, de-facto giving substantial priority to physicality through profile pictures[14]. Physicality is often related to the concept of erotic capital, that Hakim (2011) defines as a fourth personal asset (added to economic, cultural, and social capital) that one should exploit not only in mating markets, but also to powerfully advance within society. Women allegedly have more erotic capital than men because "they work harder at it" (Hakim, 2010), but this is not the only observable gender difference in partner markets, especially in mate preference matters.

It has been shown that both men and women strongly prefer similarity along most attributes – in particular for the educational level (Blossfeld, 2009) – but women tend to prefer income over physical attributes more than men (Hitsch et al., 2010). In fact, Regan et al. (2000) demonstrated that men emphasise "attributes related to sexual desirability" more than women, who by contrast value "characteristics pertaining to social status" more than men. These tendencies hold true also when it comes to online dating (Abramova et al., 2016). Educational homophily seems a determining factor in the online mate choice

[14]Please notice that attractiveness is an important factor also for mate choice processes that happen offline. However, other intangible traits (and hopefully qualities) are far less likely to be perceived online by a potential partner because of the platform design.

process; it is however second to other attributes, like racial homophily for instance. Lin and Lundquist (2013) showed that "education does not mediate the observed racial preferences" in the U.S. partner market. To put it in another way, white people with a college degree are more likely to reach or respond to other white contacts with lower educational degree than to black people with a college degree. Moreover, it has been shown that homophily is a strong determinant for attributes of life course (i.e., the marital history, children etc.) and self-reported physicality (attractiveness) and smoking habits (Fiore and Donath, 2005; Fiore, 2004). Lastly, the concept of perfect age homophily has been partially dismissed – within the socially acceptable ranges of age difference according to societal norms. Skopek et al. (2011) found in their study that age preference shifts with age, but quite differently between men and women. For heterosexual men, heterophily increased with age (i.e., men prefer younger women as they age); by contrast, for heterosexual women, age homophily increased with age (i.e., they look for men within their same age group). All things considered, we assume that homophily certainly plays a role in shaping mate preferences, but it is definitely not the only significant pattern, thus we reject the idea of one and only general logic behind online dating interactions.

Generally, men contact women more than vice versa on dating platforms; however, a minority of women also reportedly contacted far more than average men (Scharlott and Christ, 1995). An exploratory analysis unveiled six main motivations to use Tinder: love, casual sex, ease of communication, self-worth validation, thrill of excitement and trendiness (Sumter et al., 2016). Even though the love motivation broadly outclassed the casual sex one, in line with previous studies (e.g., Regan et al., 2000), Sumter et al. found that men are more likely to admit a casual sex motivation than women, whereas the latter are more prone to romance and long-term relationships. As emerged from the work of Timmermans and Caluwé (2017) on the motives to use Tinder based on the Five-Factor Model, singles who use the app are more extraverted, open to new experiences and less conscientious than those who do not. However, shy users are more likely than extroverts to admit that they are using the platform to look for love or sex, suggesting that shy users are helped by Computer-Mediated Communication and matchmaking platforms to overcome

their inhibitions and relationship-initiation barriers (Scharlott and Christ, 1995).

Due to the excess of potentially available partners, the logics of competition (for attention) can be applied to online dating. In fact, the actual probability of a first date are quite low. Not to mention that the chances of early termination of communication are extremely high, especially since not replying to a message is more broadly and socially accepted when compared to offline social venues, like school or family. This easy way out of a conversation becomes necessary in a highly competitive context where actions are lead by the principle of maximisation. However, as researches reveal, people "search surprisingly little for available marriage partners" and they often evaluate attributes in a strongly biased way (Frey and Eichenberger, 1996). The question of whether the same applies to online partner markets or online dating platforms help to mitigate such issues is yet to be answered.

1.4.1. Online Dating Functioning and its Effects on Behaviour

Fiore (2004) proved that the number of contact requests sent by online daters significantly depends on the number of contacts received. Moreover, the latter number increases the user's mate value (see 1.5.1). Thus it is clear why laymen crave to be contacted by as many potential daters as possible. As a consequence, self-presentation seems crucial to attract other users on dating platforms but often degenerates, inducing pressure to present themselves the most attractive as possible. To a certain extent, this concept could be even traced back to old "personal ads" in newspapers, studied in "People as Products" by (Hirschman, 1987). However, this is not to say that a good first impression is unnecessary in traditional dating environments, but it is especially important in online dating due to technical design and low attention span.

Combined with the above reasons, the high level of uncertainty about the authenticity of the other party enables a wide range of deceptive practices (Wang and Lu, 2007). These are very frequent on online dating platforms and usually range from small discrepancies to completely false profiles, even though the extent of such practices is insignificant in

the majority of cases (Hancock et al., 2007). False self-presentation is in fact a common rational strategy put in place to avoid competitive disadvantage, as everyone wishes to develop the most attractive profile. Thus it is not a result of psychological traits, but rather a deliberate practice to defeat competition. Those less endowed with relevant resources (i.e., with a low mate value, see 1.5.1) are more likely to adopt deceptive practices in their profiles. Nonetheless, women generally deceive less than men, and the latter are more likely to take "specific compensation of disadvantageous chances" when referring to body characteristics (e.g small men misrepresenting their height), according to Zillmann et al. (2011).

One final remark, which is omnipresent both in laymen and scholars debates, relates to infidelity. Extradyadic behaviour is common practice, both online and offline, especially among young adults. Individuals can engage in emotional infidelity or sexual infidelity – or both (Blow and Hartnett, 2005). Some researchers explored online infidelity through Tinder and found that the app effectively facilitated the engagement with extraneous relationship partners (Weiser et al., 2017). However, the debate is still open: despite these findings, Rosenfeld (2017) not only rejected the idea of online dating as a source of relationship instability – couples met online and offline have similar rates of breakup – but also found that heterosexual couples met online have a quicker transition to marriage than those who met offline.

1.4.2. Differences with Respect to Traditional Venues

When compared to traditional partner markets, online dating seems defined by smaller search friction, an expanded set of potential mates, a higher degree of anonymity and less subject to particular social norms[15]. Thus all matches and potential encounters are mainly guided by personal preferences and rational intentions. Moreover, this is magnified by the explicit function of online dating, that is to form new couples. While in many traditional contexts, like school or neighbourhood, couple formation is an "unintended side-effect of

[15]For instance, dating multiple partners offline (i.e., having interactions with more than potential partner at a time) is often a socially disapproved behaviour.

context-specific practice" that does not foster "a seller-buyer logic of interactions", the main motive behind dating platforms is to look for a partner (Schmitz, 2016). Due to its explicit nature, dating platforms differ further from conventional venues usually frequented by homogeneous groups of people, as they allow users to interact with people from socially heterogeneous groups as well. Furthermore, in case two people who interacted offline and failed at forming a couple or relationship might continue to interact in offline contexts, in online dating the chances of future meetings are modest. Finkel et al. (2012) concluded that indeed online dating significantly altered access, communication and matching compared to offline dating, but it does not necessarily generate superior romantic outcomes, despite some of its valuable advantages over offline partner markets.

1.4.3. Matchmaking and Scroll-based Models

The online dating landscape is very diverse and many different functioning models of dating platforms can be identified. A first classification relates to the degree of specialisation: mainstream services coexist along with more specialised services that only target a niche audience (e.g., ethnics, single parents, seniors or younger demographics etc.) or particular interests (e.g., adult dating, fetish, hook-ups etc.). However, a more interesting distinction can be further outlined between matchmaking systems and scroll-based platforms ("online dating" in general among scholars). According to Schmitz (2014), in the first model a new user is required to answer a comprehensive list of questions regarding his or her personality traits and lifestyle – along with other questions on smoke habits, children and marriages. An algorithm then computes a one-dimensional factor based on these pieces of information and the system uses it to suggest potential partners to the new user. The rationale behind this matching factor is, in most of the cases, the similarity between the user profile with those of potential others. A second model refers to those platforms in which potential couples are not matched by an algorithm, but rather require the active[16] involvement of both parties in searching[17] and contacting the profiles

[16]In reality, a passive participation is also possible if the user only awaits for incoming contacts.
[17]I.e., browsing, hence the "scroll" action required on mobile apps.

that they personally consider interesting. In the latter model usually new users provide only brief personal (socio-demographic) information and the desired characteristics of a potential partner.

1.5. Online Dating as a Marketplace

From an economic standpoint, a market is a medium by which the exchange of a certain good is facilitated with the interaction of buyers and sellers who establish a price. A Weberian definition of market assumes its existence whenever there is competition for the opportunities of exchange. Becker (1974) supports this thesis and applies it to couple formation processes: "since many men and women compete as they seek mates, a market in marriages can be presumed to exist". By virtue of extensive information about a wide pool of potential mates, it seems that online dating can be seen as strongly affected by competition that manifest itself mainly through competition for attention and potential partners. The marketplace metaphor, interpreted as "a place where people go to shop for potential romantic partners and to sell themselves in hopes of creating a successful romantic relationship", is particularly emphasised by the functioning and design of dating platforms that often resemble e-commerce websites (Heino et al., 2010). Economic-based metaphors by which looking for a mate is seen in terms of "shopping" and consumption, whereas potential mates as products, are common and useful to explicate the exchange processes that happen in partner markets (Bernard and Adelman, 1990).

This resonates with the principle of maximisation that guide actions on online dating platforms. In fact, if users have clear intentions when looking for a partner, they might be particularly inclined to apply cost-benefit calculations: in other words, the search process is grounded upon trait-oriented choices that ultimately force the comparison of a large pool of alternative profiles (Illouz and Finkelman, 2009). As a consequence, we might argue that online dating is characterised by relatively low transaction and search costs. The ultimate decision of picking a partner is thus conceived through a (quasi-)formal model of expected utility maximisation, by fulfilling personal mate preferences while minimising

the costs – e.g., opportunity costs, search costs or costs of a particular date or relationship. The choice of offline continuation over the termination of contacts is thus realised if the costs do not exceed the benefits. Because of the homogeneous good exchanged[18], the inherent logics of supply and demand, exchange, competition and conscious rational mate preferences, online dating can be seen as a highly efficient partner market, and unusually close to the ideal one (Schmitz, 2016).

1.5.1. The Partner Market and Mate Choice

A partner market is usually referred to the user's field of social interactions directed to eligible partners and pursued for mating goals. However, in practice, identifying such a market is no easy task, because of the intrinsic difficulty in determining what is the object of trading in such a market. In fact, there is no uniform and universal exchange entity in partner markets – as opposed to money in financial markets. Rationalist approaches require a general utility unit that should be maximised in mate choice processes, but assuming a common metric for all users is a pitfall. Different users may be interested in different attributes (i.e., goods) in potential partners[19]. To put it another way, it is not clear whether the (symbolic) goods being marketed are the traits of potential mates, the potential partners themselves, relationships or the exchange chances on dating platforms (Schmitz, 2016). Moreover, there could be more than one kind of goods being bartered on the same market[20].

One solution could be of thinking about the "price" of an actor as "a function of actively and passively approved and rejected offers" (Schmitz, 2016). Thus it would be possible to attach a mate value to each user of a dating platform, defined according to his or her chances of attention in the partner market. In some attempts at computing this metric the number of potential partners has been used, but researchers failed at recognising that

[18] To a certain extent, online dating forces a standardising influence on the users, since they are expected to appear as attractive and interesting as possible.

[19] For instance, it would be erroneous to assume that a certain physical attribute (e.g., being tall) is equally desirable by all users.

[20] Could we consider looking for a marriage, sexual satisfaction and the fame deriving from coupling with a celebrity all the same kind of goods?

it is too ambiguous (e.g., Pawowski and Dunbar, 1999). Schmitz (2016) acknowledged that contacts received by users with a low market value are worth less than those received by users with a higher market value. Therefore he elaborated a new function of mate value which takes into account the fact that different contacts have different market values (see Equation 1.1 and Figure 1.6).

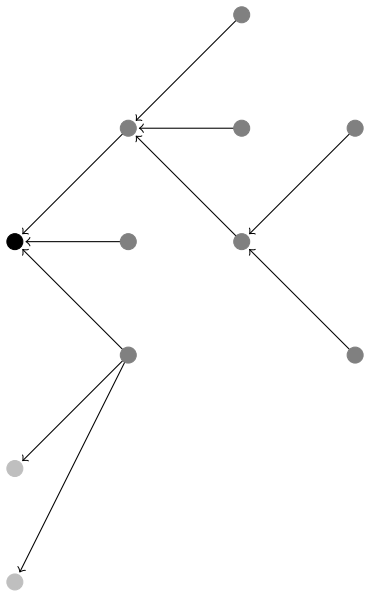

Figure 1.6.: *Ego*'s (black) incoming contact network.

Thus the mate prestige indicator (MP) is a function of the ranks of other users contacting *ego*. In fact, the number of ingoing contacts increases *ego*'s MP, but higher *alter* MP values increase the former even more. That is to say that "it is good for one's mating chances to get a lot of offers; it is even better if the offers are from potential mates who also have good mating chances".

$$MP_{IN}(A) = (1-d) + d \cdot \sum_{i=1}^{n} \frac{MP_{IN}(T_i)}{C_{IN}(T_i)} \tag{1.1}$$

with:

$MP_{IN}(A)$ the mate prestige value of individual A

$MP_{IN}(T_i)$ the mate prestige of individuals T_i, who contacted A

$C_{IN}(T_i)$ the total number of contacts that were established by T_i

d a damping factor between 0 and 1

Nonetheless, readers should bear in mind that this model is nothing more than a helpful oversimplification of complex social processes. In reality, mental processes are far from being fully rational and rarely follow the logics of maximisation, but rather the principle of "satisficing", by which a potential partner is picked if sufficiently meets the user's preferences (Simon, 1956).

1.5.2. Choice Overload Issues

As previously noted, when compared to traditional dating, online dating makes available a larger pool of potential partners at once, in an environment explicitly designed to commence communication with many of them. Few studies have been conducted to examine whether this aspect improved chances of romantic outcomes or lead to choice overload, laziness in comparison strategies and ultimately to poor decisions. In fact, it has been shown that a large choice set can overwhelm users, leading to less efficient decisions or even in a state of choice overload with so many alternatives that people avoid taking a decision in the first place, rather than going through the mentally laborious process of comparing and selecting among all the available options (Finkel et al., 2012). This paradox of choice is a trade-off between the frustrating "maximisers' desires to find the best solution"[21] and the "advantage of choice" often advertised as the greatest feature of online dating platforms that can offer thousands of potential partners (Rosenfeld, 2017).

1.5.3. Network Externality in a Two-sided Market

One final remark on the functioning of online dating platforms as marketplaces concerns indirect network effects. Network goods are those whose utility deriving from their consumption is affected by the number of other people consuming them (direct network effect) or by the availability of complementary goods or networks, that depends on the number of potential buyers (indirect network effect). In other words, the utility increases when

[21] As Rosenfeld (2017) puts it: "without the possibility of knowing they have found the optimal solution, maximisers become unhappy and (out of frustration) make hasty choices that make them even more unhappy."

more people are using these products. Thus we infer that the utility U is a function of the number of users n and can be approximately computed as $U(n_i) = s_i + \alpha \cdot n_i$, with s_i the stand-alone value of the good, α the strength of the network effect and n_i the number of users n in the network i.

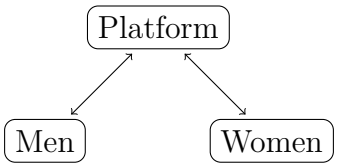

Figure 1.7.: An example of heterosexual online dating platform as a two-sided market.

Many definitions of a two-sided market can be encountered in academic research. For instance, Evans, 2003 defines it as a market in which a firm (the platform) sells two different products to two groups of consumers, recognising that the demand of one group depends on the demand of the other one (indirect network effect) – and vice versa. More generally, we define a multi-sided platform as an organisation that creates value primarily by enabling direct interactions between two (or more) distinct types of affiliated customers (Hagiu and Wright, 2015). The price structure is also peculiar, as the allocation of the total price is different between the different sides of the market – often one side is subsidised to profit from the other (Rochet and Tirole, 2003). Since online dating platforms seem to fulfil these criteria, we are comfortable at considering them as two-sided markets (see Figure 1.7). In fact, there are two sides or groups of customers (men and women in the following heterosexual example, or group A and group B more generally); the value of such platforms strictly depends on the number of people using them and on the capacity to connect the two sides; the usage is influenced by the allocation of prices, although price discrimination is often ethically and logically discouraged. For instance, by raising the price only for women on an heterosexual dating platform, less women would use it. With less women on board, the overall value of the service is decreased for men (and thus for the company itself as well). Since the platform values less but the price for men has not been changed, more men would leave, de-facto igniting a negative feedback loop since the value of the platform would decrease also for women as a consequence of more men leaving, and so on.

2. Methods and Data

2.1. Research Design and Data Set

In accordance with the above-mentioned insights, we elaborate some hypotheses on online dating users and usage to test in this study. As described in section 1.1, the increased search efficiency is especially convenient to those people who lack an easy access to a partner market (e.g., Stevenson and Wolfers, 2007), thus a first hypothesis can be outlined.

Hypothesis 1: Online daters are more likely to be in a thin market for partners or in a minority group (e.g., race, sexual orientation, religion, disability, etc.).

On the other hand, we investigate the online dating platforms penetration among all social strata and the upper ones in particular. In fact, such demographic groups (young adults, graduates and those from high-income households) are those who use the Internet the most[22]. Thus, consistent with prior research (e.g., Cacioppo et al., 2013), we expect the following traits to be confirmed:

Hypothesis 2: Young, white males, who are not already married nor with children, with a higher socioeconomic status and liberal views are those who use online dating more often.

Because official data from online dating private companies is not made publicly available, all analyses in this thesis are based on secondary data from Pew Research Center's Internet, Science & Technology Project "Spring Tracking" Survey of 2015. Among the many topics covered – such as gaming, home broadband and smartphone usage – online dating questions were also included.

[22] As reported in the Internet/Broadband Fact Sheet by Pew Research Center (2018).

The survey was conducted between June and July 2015 and provides a national sample of 2,001 adults (18 years of age or older) living in the United States. 701 of these respondents have been interviewed on a landline telephone, while the remaining 1,300 were interviewed on a cellphone, all via random digit dialing. Therefore, respondents in the landline sample have been randomly picked by asking for the youngest adult who was at home at the moment of calling. Conversely, those in the cellphone sample were personally interviewed when they answered the phone, if they were at least 18 years old. With this method, fraudulent or deceptive responding is minimised: for instance, no multiple surveys could be completed from the same respondent, nor they could complete it too quickly to reflect valid data (i.e., random responses).

Data collection processes were conducted with strict methodology, and the final data set has been provided with full documentation. As a result of the rigorous techniques put in place, the data set is considered to be fully representative of all residents in each U.S. state. In order to correct for the stratification sampling design at the state level and to overcome over-sampling problems due to different population densities and other known biases (on parameters such as age, race, gender, etc.), we conduct all analyses weighting data accordingly. Moreover, the weights take into account that people in large households and that people with both a landline and a cellphone have respectively lower and higher chances of being selected. All analyses and computations are processed with IBM SPSS software.

As we wish to test the aforementioned hypotheses by measuring the partial (ceteris paribus) effects of each independent variable on the variation in the usage of online dating, we run a Multiple Regression Model (MRM) along with some preliminary descriptive statistics analyses. We develop different models to explain the online dating usage frequency, progressively adding up variables to reach our final and complete model (see Table 3.3 on page 33). All analyses are conducted on the same observations and with the same dependent variable, that is thoroughly discussed in the following section.

2.2. Online Dating Usage: Dependent Variable

We want to investigate the impact of different independent variables on the actual usage of online dating platforms. Respondents were asked two questions that particularly fit our needs: DATE1A ("Have YOU, personally, ever used an online dating site such as Match.com, eHarmony, or OK Cupid?"), that has been asked to all Internet users (EMINUSE[23]=1 or INTMOB[24]=1), and DATE2A ("Have you ever used a dating app on your cell phone?"), that has been asked to every smartphone owner (SMART1[25]=1). Both variables can take either value 1 ("Yes"), 2 ("No"), 8 ("Don't know") or 9 ("Refused"). As explained later, they have been jointly used to create a new dummy (ISONLINEDATER) that is included in the MRM as the dependent variable.

As we would expect, they both showed high percentages of missing values: 13.4% for DATE1A and 32.4% for DATE2A (see Table 2.1 and A.1; Figures A.1 and A.2). However, in the light of the fact that there is no commonly accepted way to handle missing data, and that a generally high percentage of them can be found in social science research papers, we decided to proceed by keeping all cases in further analyses. To clarify, we could have adopted other solutions: for instance, we could have omitted all cases with missing values for both DATE1A and DATE2A (listwise deletion) – and thus keeping those cases that reported a 1 ("Yes") in at least one of these variables[26] – which should be considered as "online daters" anyway; filtering out all respondents who reported not to use the Internet at least occasionally (EMINUSE≠1) could have also been a quick and similar fix. However, none of these proved to be sound choices. In fact, we want to take into account in our analysis also people who do not frequently use the Internet (namely the elderly). We suspect that the great majority of missing values is not due to an error, but rather to the fact that these questions have been asked only to a restricted pool of respondents: those

[23] EMINUSE (asked all): "Do you use the internet or email, at least occasionally?"

[24] INTMOB (asked all): "Do you access the internet on a cell phone, tablet or other mobile handheld device, at least occasionally?"

[25] SMART1 (asked if have a cell phone): "Some cell phones are 'smartphones' because of certain features they have. Is your cell phone a smartphone such as an iPhone, Android, Blackberry or Windows phone, or are you not sure?"

[26] I.e., meaning that the respondent had used an online dating site at least once but never a smartphone dating app (or vice versa), hence he or she should be considered as an "online dater" anyway.

who had disclosed that they use the Internet or own a smartphone. In fact, the Little's MCAR Test is significant at 1%, thus we reject the null hypothesis that values are missing completely at random (see Table A.1). That said, we do not exclude the occurrence of any other potential issue a priori, in fact some improbable values have also been observed, but their frequency seems negligible (see Table A.2). Since no other issues are noticeable, we proceed by describing the actual dependent variable of our model.

In order to run a regression on the usage of online dating, we computed a new dummy variable ISONLINEDATER that takes value 1 if either DATE1A or DATE2A is true, 0 otherwise. In this manner, we are considering all missing values from DATE1A and DATE2A as if they were coded a 2 ("No")[27]. Thus, from here on, we define as "online dater" anyone who has used a dating site or a mobile dating app at least once. As a result, 15.5%[28] of respondents is defined as "online dater" (see Table A.3 and Figure A.3). It follows that ISONLINEDATER has been employed as the outcome variable in our final regression model. Table 2.1 collects some statistics for the variables in question.

		DATE1A	DATE2A	ISONLINEDATER
N	**Valid**	5428	4236	6267
	Missing	839	2032	0
Mean		1.87	1.89	.1549
Median		2.00	2.00	.0000
Std. Deviation		.442	.445	.36181
Skewness		4.643	4.982	1.908
Std. Error of Skewness		.033	.038	.031
Kurtosis		88.310	81.147	1.642
Std. Error of Kurtosis		.066	.075	.062
Minimum		1	1	.00
Maximum		9	8	1.00

Table 2.1.: Univariate Statistics for the dependent variables.

[27] This provides a conservative estimate since we are not taking into account potential respondents who lied by not declaring their actual usage of online dating.

[28] Notice that this slightly higher percentage of 1 ("Yes") in ISONLINEDATER, compared to DATE1A and DATE2A, is due to the fact that some respondents who reported a 2 ("No") in one of these variables are being considered anyway "online daters", if they reported a 1 ("Yes") in the other one.

2.3. Independent Variables

Unless otherwise noted, all the covariates have been recoded to provide better interpretable results. The following input variables have been standardised for consistency reasons, and only pertinent answers have been included in our analyses. In other words, common entries such as 98 ("Don't know") or 99 ("No answer") have been recoded as system-missing values.

As regards the controlling variables, we included some basic regressors in accordance to preexisting literature. GENDER is a dummy variable that takes value 0 if male, 1 if female. UNIVERSITY and MINRACE also follow the same logic: they take value 1 in case the respondent – respectively – graduated at least at a Bachelor level and is from a racial minority (i.e., is not White), 0 otherwise. Since most of the covariates are categorical, nominal or ordinal variables (AGE is the only numerical variable included in our model), new dummies have been coded for them, following the general principle of assigning a 1 for the presence of a level, and a 0 for its absence. Some anchored scales are also present. INTERNETFREQ[29] ("About how often do you use the internet?") and POLVIEW ("In general, would you describe your political views as...") are both 5-point scales that range from, respectively, "Less often" to "Almost constantly" and from "Very conservative" to "Very liberal". Because we did not want to lose the information in their ordering, we opted for treating them as numeric. However, this was not possible for the original variable INC ("Last year - that is in 2014 - what was your total family income from all sources, before taxes?"), which is also an anchored scale, but it ranges from 1 ("Less than $10,000") to 9 ("$150,000 or more") with different numerical distance between each category. Thus, we introduced new dummies, INCOME1, INCOME2 and INCOME3, that split respondents into three different income categories, from the lowest to the highest.

Table 2.2 provides a full list of selected variables included in the final regression model. Since all these variables directly refer to respondent's individual traits and preferences, and interpretation seems straightforward, we do not deepen further into them.

[29] INTERNETFREQ is the inverted anchor scale of the original variable INTFREQ, that has been recoded for more easily interpretable results.

Label	Description or original question	Values admitted (recoded)
ISONLINEDATER	Have you used a dating site or a mobile dating app at least once?	0=No; 1=Yes.
GENDER	Respondent's sex.	0=Male; 1=Female.
AGE	What is your age?	*numeric*
UNIVERSITY	Is true if respondent graduated at least at a Bachelor level.	0=No university degree; 1=Yes.
MINRACE	Is true if respondent is from a racial minority (not White).	0=White; 1=Else.
INCOME1	Was your family income last year from less than $10,000 to under $30,000?	0=No; 1=Yes.
INCOME2	Was your family income last year from $30,000 to under $75,000?	0=No; 1=Yes.
INCOME3	Was your family income last year $75,000 or more?	0=No; 1=Yes.
MARRIED	Are you currently married or living with a partner?	0=Never been married; 1=Married/living with a partner.
ISPARENT	Are you parent of any children under 18?	0=No; 1=Yes.
POLVIEW	Would you describe your political view as...	1=Very conservative; 2=Conservative; 3=Moderate; 4=Liberal; 5=Very liberal.
URBAN	Community type is urban or suburban.	0=Rural; 1=Urban/Suburban
PHYSICALLABOUR	Respondent's job involves manual labour.	0=No; 1=Yes.
DISABILITY	Respondent has any handicap or disability.	0=No; 1=Yes.
INTERNETFREQ	How often do you use the Internet?	1=Less often; 2=Several times a week; 3=About once a day; 4= Several times a day; 5=Almost constantly.
ATTITUDE1	Online dating is a good way to meet people.	0=Disagree; 1=Agree.
ATTITUDE2	Online dating is easier and efficient.	0=Disagree; 1=Agree.

Table 2.2.: Description of variables selected in the final model.

2.4. Model Construction

Among all the variables provided in the data set, we decided to include only those ones that best could influence individual adoption or usage of online dating in our opinion. The first model is a very basic one, with only some controlling variables (socio-demographic) included: GENDER, AGE, UNIVERSITY, MINRACE, INCOME2 and INCOME3. To avoid the dummy variable trap, we kept INCOME1 out from our models as a reference category. Our second model, takes into account also some attributes of the life course, like marital history or dependent children (underage). The third model builds upon the first ones and considers political ideologies and potential handicaps as well. It includes POLVIEW and DISABILITY. The final[30] and most comprehensive model includes these previous variables, along with other key ones to account for individual attitudes and preferences toward the Internet and technology in general. Thus it turned out to be based upon the following regressors:

$$\begin{aligned} Y = {} & \beta_0 + \beta_1 \cdot GENDER + \beta_2 \cdot AGE + \beta_3 \cdot UNIVERSITY + \beta_4 \cdot MINRACE \\ & + \beta_5 \cdot INCOME2 + \beta_6 \cdot INCOME3 + \beta_7 \cdot MARRIED + \beta_8 \cdot ISPARENT \\ & + \beta_9 \cdot POLVIEW + \beta_{10} \cdot DISABILITY + \beta_{11} \cdot INTERNETFREQ \\ & + \beta_{12} \cdot ATTITUDE1 + \beta_{13} \cdot ATTITUDE2 + \varepsilon_i \end{aligned} \tag{2.1}$$

One simpler model has also been developed, specifically to test the thin market *hypothesis 1*. Notwithstanding that very few useful variables can be found in the data set for this purpose – questions on religious beliefs and sexual orientation, for instance, could have definitely improved this model – we included DISABILITY and MINRACE. We also decided to add URBAN, in order to assess whether living in a rural community might push inhabitants to resort to online dating for their mating purposes, since the low population density could be equated with being in a thin market for potential partners.

[30] Actually, a further model with PHYSICALLABOUR is also presented. The rationale for including it is to assess to what extent having a manual or physical intensive job position affects the outcome variable in question. However, as it is explained later, it lacks of significance despite a higher R^2, thus we discourage considering this model as the definitive one.

3. Results

3.1. Descriptive Statistics

We observe that females in our sample are 51.5% and that the average age is 46.7. Both AGE and POLVIEW seem normally distributed – the absolute value of skewness and kurtosis are < 1 and few outliers have been observed – while INTERNETFREQ seems left-skewed (see Figures A.4, A.5, A.6 and Table A.4 on page 46): it has a mean of 3.76 but skewness is slightly problematic (-1.087). We encountered the same kind of problem with other variables. However, due to the very large sample size, we considered that the Central Limit Theorem could be applied and therefore we refused to drop these important variables.

No significant issues emerged from a Missing Value Analysis (see Table 3.1 on page 29), with few notable exceptions: INTERNETFREQ (14.1% of missing values), PHYSICALLABOUR (41.4%), POLVIEW (9.6%) and MARRIED (19.2%). However, these high percentages could be explained by the fact that these questions have not been asked to all respondents. In fact, INTERNETFREQ has been asked only if the respondent had previously admitted to use the Internet or own a cell phone; similarly, PHYSICALLABOUR has been asked only to currently employed people (leaving out not only the elderly, but also students or recent graduates). On the other hand, questions on sensitive matters, like political orientation, usually entail a higher than average percentage of missing values. MARRIED, instead, has been recoded to take value 1 ("Yes") only if the respondent is currently married or living with a partner, and 0 ("No") if he or she has never been married before, thus leaving out all divorced, separated and widowed. Although the adequate percentage of missing values ranges from 5% to 8% at most, we decided to keep

			Std.	Missing		N. of Ext.[a b]	
	N	**Mean**	**Deviation**	**Count**	**Percent**	**Low**	**High**
AGE	5318	46.2232	17.84838	100	1.8	0	0
GENDER	5418	.5172	.49975	0	.0	0	0
UNIVERSITY	5383	.2631	.44033	35	.6	0	0
MINRACE	5291	.2657	.44177	127	2.3	.	.
INCOME1	5418	.3073	.46142	0	.0	0	0
INCOME2	5418	.2999	.45827	0	.0	0	0
INCOME3	5418	.2527	.43459	0	.0	0	0
MARRIED	4378	.6942	.46082	1040	19.2	0	0
ISPARENT	5365	.2939	.45561	53	1.0	.	.
POLVIEW	4897	2.9579	1.05925	521	9.6	0	0
URBAN	5418	.8319	.37403	0	.0	.	.
PHYSICALLABOUR	3173	.4901	.49998	2245	41.4	0	0
DISABILITY	5399	.1721	.37748	19	.4	.	.
INTERNETFREQ	4652	3.7623	1.13325	766	14.1	136	0
ATTITUDE1	5094	.6307	.48265	324	6.0	0	0
ATTITUDE2	4985	.5047	.50003	433	8.0	0	0

Table 3.1.: Missing Value Analysis.

[a] Number of extreme cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).
[b] . indicates that the inter-quartile range (IQR) is zero.

these variables due to their considerable high number of observations and their crucial importance in our model.

Table 3.2 on page 30 displays the bivariate (linear) correlations between all pairs of variables selected in our model. This reveals some interesting and statistically significant links among the variables, that to a certain extent seem to confirm our *hypothesis 2*. For instance, we notice that online dating is positively correlated with being male, young, highly educated, with liberal views, single and with no children. Interestingly enough, some insights on U.S. (and sometimes worldwide) common social issues emerge as well.

		ISON.DAT.	GEN.	AGE	UNI	M.RACE	INC.1	INC.2	INC.3	MARR.	ISPAR.	POLVIEW	URB.	PHY.LAB.	DISA.	INT.FREQ	ATT.1	ATT.2
ISON.DAT.	Pearson C.	1																
	Sig.																	
GEN.	Pearson C.	-.048**	1															
	Sig.	.000																
AGE	Pearson C.	-.211**	.045**	1														
	Sig.	.000	.000															
UNI	Pearson C.	.056**	.013	.044**	1													
	Sig.	.000	.311	.001														
M.RACE	Pearson C.	-.018	-.011	-.143**	-.048**	1												
	Sig.	.167	.372	.000	.000													
INC.1	Pearson C.	.000	.046**	-.002	-.253**	.129**	1											
	Sig.	.991	.000	.880	.000	.000												
INC.2	Pearson C.	.045**	-.019	-.032*	-.005	-.023	-.429	1										
	Sig.	.000	.137	.012	.704	.074	.000											
INC.3	Pearson C.	.023	-.057**	-.052**	.301**	-.075**	-.388**	-.386**	1									
	Sig.	.067	.000	.000	.000	.000	.000	.000										
MARR.	Pearson C.	-.282**	.038**	.478**	.126**	-.145**	-.177**	.038**	-.386**	1								
	Sig.	.000	.000	.000	.000	.000	.000	.008	.000									
ISPAR.	Pearson C.	-.047**	.046**	-.263**	.020	.056**	-.006	.000	.090**	.249**	1							
	Sig.	.000	.000	.000	.119	.000	.637	.983	.000	.000								
POLVIEW	Pearson C.	.103**	.049**	-.145**	.096**	.084**	.033*	-.009	.025	-.126**	-.010	1						
	Sig.	.000	.000	.000	.000	.000	.014	.499	.058	.000	.435							
URB.	Pearson C.	.047**	.018	-.117**	.101**	.109**	-.045**	-.041**	-.078**	-.109**	-.049**	.091**	1					
	Sig.	.000	.147	.000	.000	.000	.000	.001	.000	.000	.000	.000						
PHY.LAB.	Pearson C.	-.041*	-.107**	-.123**	-.453**	.098**	.222**	.070**	-.260**	-.134**	.021	-.004	-.080**	1				
	Sig.	.012	.000	.000	.000	.000	.000	.000	.000	.000	.208	.830	.000					
DISA.	Pearson C.	-.015	.046**	.226**	-.157**	.011	.250**	-.077**	-.184**	.017	-.120**	-.022	-.043**	.016	1			
	Sig.	.231	.000	.000	.000	.390	.000	.000	.000	.228	.000	.100	.001	.334				
INT.FREQ	Pearson C.	.152**	-.012	-.344**	.162**	.013	-.195**	.017	.178**	-.077**	.101**	.111**	.105**	-.243**	-.174**	1		
	Sig.	.000	.360	.000	.000	.351	.000	.220	.000	.000	.000	.000	.000	.000	.000			
ATT.1	Pearson C.	.167**	-.096**	-.161**	.074**	-.079**	-.091**	.031*	.112**	-.097**	-.015	.054**	.021	-.084**	-.051**	.156**	1	
	Sig.	.000	.000	.000	.000	.000	.000	.019	.000	.000	.243	.000	.109	.000	.000	.000		
ATT.2	Pearson C.	.113**	-.062**	-.062**	.062**	-.071**	-.036**	-.008	.067**	-.030*	-.021	.018	.019	-.131**	.024	.110**	.400**	1
	Sig.	.000	.000	.000	.000	.000	.007	.550	.000	.041	.113	.201	.145	.000	.067	.000	.000	

Table 3.2.: Bivariate correlation matrix.

**Correlation is significant at 0.01 level (2-tailed).
*Correlation is significant at 0.05 level (2-tailed).

Namely, the gender pay gap seems noticeable, by observing the correlations between GENDER and different income levels. There is also a significant linear relationship between being in a racial minority and earning less and marrying less often. Finally, we note not only that INTERNETFREQ is positively correlated with ISONLINEDATER, but also that it averagely follows its same pattern of correlations with other variables (i.e., it has a positive relation with being male, young, richer, etc.). By contrast, the thin market *hypothesis 1* finds very little support – if any – from this bivariate analysis since, for instance, living in a urban area is positively correlated with ISONLINEDATER, and having a disability is negatively correlated with the frequency of Internet usage. However, more meaningful and precise results on actual dependency links are addressed with the following multiple regression models.

3.2. Multivariate Analysis

Table 3.3 on page 33 displays all the resulting models from the multiple regressions that we have conducted. F test is significant at 1% in all models (see Table A.5 on page 47). Unless otherwise noted, all variables are statistically significant at the 0.01 level as well, and no particular issues of multicollinearity have been detected – VIF scores never higher than $\simeq 1.8$.

The first basic model only included six controlling variables. Each coefficient brought significant results ($P-value < 0.01$), except for INCOME2 ($P-value < 0.05$) and INCOME3 (which becomes however significant in other models). The overall model accounted for approximately 5.5% of the variance of online dating usage ($R^2 = .055$). Moreover, all the variables' coefficients seem consistent with the theoretical framework built for *hypothesis 2* (except for INCOME3): thus it seems that being male, younger, with a university degree, White and in the upper social classes, predicts higher chances of being an online dater. In order to boost the explanatory power of the model, we included some more variables.

Model 2, taking into consideration some attributes of the life course, provides a considerable higher explanatory power ($Adjusted\ R^2 = 0.111$). As expected, we find that one

standard deviation decrease in MARRIED has the strongest effect (23.7% in Model 2 and 25.9% in Model 5) on our dependent variable compared to one standard deviation variation in any other covariate. A third and a fourth more robust models are subsequently developed, adding regressors on political views, disability and then Internet attitudes. Model 4 reaches an R^2 of .123 thanks to the contribution of the Internet factors, that are all statistically significant at 1% and with a positive sign. The latter model seems to us the most useful and easily interpretable one. In fact, in Model 5, PHYSICALLABOUR has also been added, and although it further increased R^2 (.126), we find more covariates to be statically not significant at all (e.g., AGE or ISPARENT) and generally more counterintuitive results.

In pursuit of more powerful predictions, we also built a sixth model (not shown in Table 3.3) with some interaction terms: GENDERXUNIVERSITY and GENDERXMINRACE. However, despite the higher R^2 (.137), it failed at providing just as many significant beta coefficients, thus we dismissed it. To clarify, we value more those lower R-square models with higher P-values rather than vice versa. If the only goal of the model was to make an accurate prediction, this would be a flawed idea. Yet this is not the case, since we are more interested in assessing whether a (perhaps small) reliable relationship truly exists among the considered variables. Even though the model does not explain much of the variation of the data, it still is significant. Moreover, it is quite common to find this kind of R-square percentages in social science research. Outcomes could be dictated by many other latent factors at play that may affect other facets of the model – especially if the data deal with human behaviour, attitudes, thoughts, preferences and feelings, which are often object of measurement errors.

One final remark on the model H_1 that has been designed to test the thin market hypothesis. It is overall significant and no issues have been detected as in the previous models. The three key variables chosen to test the hypothesis – MINRACE, DISABILITY and URBAN – are all significant at 0.01 level. Some valuable insights emerged and they are extensively addressed in the following section.

	Model **1**	Model **2**	Model **3**	Model **4**	Model **5**	Model $\mathbf{H_1}$
GENDER (*female*)	-.038**	-.048**	-.044**	-.047**	-.063**	-.041**
AGE	-.221**	-.131**	-.121**	-.032	.001	-.229**
UNIVERSITY	.064**	.078**	.063**	.046**	.047*	.068**
MINRACE	-.045**	-.051**	-.051**	-.048**	-.060**	-.050**
INCOME2 V. INCOME1	.031*	.076**	.082**	.062**	.060**	
INCOME3 V. INCOME1	-.006	.052**	.056**	.035	.060*	
MARRIED		-.237**	-.239**	-.259**	-.243**	
ISPARENT		-.052**	-.045**	-.025	-.069**	
POLVIEW (*liberal*)			.076**	.066**	.122**	
DISABILITY			.028	.037*	-.011	.049**
INTERNETFREQ				.062**	.052*	
ATTITUDE1				.084**	.104**	
ATTITUDE2				.055**	.037	
PHYSICALLABOUR					-.009	
URBAN						.030*
No. of observations	6010	4849	4468	3678	2556	6000
R^2	.055	.112	.119	.123	.126	.056
Adjusted R^2	.054	.111	.117	.120	.121	.055

Table 3.3.: Multiple regression models on the likelihood of online dating.

DV: ISONLINEDATER. We are reporting the standardised beta coefficients only.
**Significant at 0.01 level.
*Significant at 0.05 level.

4. Discussion

4.1. Common User Profiles

Despite its poor robustness, model H_1 reveals some interesting and rather contrasting insights. We partially reject *hypothesis 1* since the model predicted higher chances of being online daters if White and living in a urban area – i.e., not being in a racial minority and living in a rural community far from the large pool of potential partners of a big city. The positive sign of URBAN, by violating the first hypothesis, actually validates the concept of network effects, since people seem to use online dating more in highly dense environments. Moreover, AGE in this model has the strongest effect among all covariates ($-.229^{**}$) and this goes further against the first hypothesis. Older singles are, in fact, in a thin dating market, since most of their own kind are already partnered in their 30s and 40s usually (Rosenfeld and Thomas, 2012). By contrast, DISABILITY appears to confirm the thin market hypothesis with a standardised beta coefficient of .049, significant at 1%. This suggests that people with a handicap are in fact more likely to resort to online dating than those who do not. Assessing whether one is in a thin market for potential partners obviously requires far more variables to be taken into account, in order to uncover the latent factors behind such diverse human behaviours, attitudes and preferences. All things considered, we do not have enough (and appropriate) data to produce unambiguous findings on this topic, thus we dismiss the first hypothesis.

As regards *hypothesis 2*, results from the models seem consistent with the theoretical framework previously discussed and in accordance with the extensive studies already conducted on this topic. As we would expect, most online daters are generally the youngest

White males with a higher income and education level than average. They are, in fact, those who also spend more time on the Internet and have more progressive views towards it. These people also tend to lean more on democratic and liberal views in politics. According to our models, online daters are generally more likely to have no children nor a stable partner yet. Since only 5.8% of married[31] respondents also reported having utilised online dating platforms, we infer that the vast majority of them found a stable partner through conventional mating markets (and perhaps those in the minority are just some residual cases of self-reported infidelity). It seems plausible to assume that most of them married before the diffusion of online dating in the United States, thus it would be erroneous to interpret these data as a defeat of such platforms. Given the above, we find enough empirical evidence to support and confirm the second hypothesis, which however begs for more in-depth investigations.

4.2. Implications for Business and Further Research

Studies like this are far from being purely theoretical. Online dating analyses can help to grasp insights on the economic mechanisms behind match formation and marriages. The consequences of these results for businesses are twofold. First and foremost, it is of absolute importance that companies interpret current socio-demographic trends, in order to understand men, women and families of the future. The way people interact, get to know potential partners, fall in love, forge profound relationships and eventually create new families, has a dramatic impact on everyday trades and society as a whole. In other words, understanding the logics behind mate choice processes also helps to understand the logics behind consumer choices. Market metaphors are in this sense useful to move companies closer to people and to what would be otherwise considered naturally extraneous to exchange logics. A practical example of taking advantage of these studies could be the more precise and efficient use of programmatic advertising or, more generally, of targeted ads on online dating platforms.

[31] I.e., 5.8% of MARRIED. It includes all married people and those living with a partner.

One second key managerial concept mainly regards online dating companies. Those market characterised with increasing returns of adoption, such as network effects, steadily tend towards the dominance of a single technology or product, usually in ways very difficult to reverse because of lock-in effects. It is rare that more than a single product profitably co-exist in the long term[32]. Since very little differentiation can currently be found among different online dating services in this phase, it is likely that in the future one single dominant platform (or a dominant design) will emerge. Hence, it seems crucial at this point to adopt measures safeguarding the standalone value of one's own dating platform – or to attract as many users as possible.

This work certainly raises more questions than it answers. Our models work well for conveying simple concepts, yet they fail at exploring all facets of intricate human behaviours and providing strong predictions. Undoubtedly, actual user behaviours in both online and offline dating markets cannot be exactly described nor predicted by models. Nevertheless, we suggest further research to be carried out based on primary data and taking into account more individual traits and distinctions, in order to deeply explore the development of modern romantic relationships. Additional analyses could not only inspect online dating market dynamics, mate choice logics and marriages in complex networks, but they could also qualitatively unravel the human stories behind those matches.

[32]E.g., think of the dominance of Facebook among social media.

5. Conclusion

This thesis offers insight into the principles of online dating and explores the determinants of its usage through simple models. We investigated two – rather opposing – views and tested their hypotheses. While we found little empirical evidence supporting the hypothesis of online dating primacy for those in a thin market for romantic mates, we recognise that some common distinguishing features can be traced among its users. Through our final model we were able to identify the key demographics, traits and attitudes of those who reported having utilised online dating at least once. In the light of the resulting quite weak (yet significant) models, we would suggest grounding future research on more recent data, as well as taking into account personality traits and further distinctions among cases (e.g., sexual orientation, engagement in religious practices, intentions and desired outcomes from the dating platforms above all).

Ultimately, online dating exists to allow a more efficient search for a partner, but it is paramount to acknowledge that "the Internet complements, rather than displaces, existing behaviour patterns" (DiMaggio et al., 2001), and by extension, so do the tools it made available, like online dating platforms. In conclusion, we believe that more accurate research should be carried out, notwithstanding the fact that it will take generations, perhaps, to assess the impact of such practices at societal level (Stevenson and Wolfers, 2007). The Internet might be able to change the marriage market by promoting better matches, but it still is too early to appraise its effects on divorce and marriage rates, for instance, and ultimately on society of the generations to come.

Acknowledgements

I would like to thank my supervisor, Prof. Francesco Candeloro Billari, for his guidance, immense knowledge and support of this work. I am also deeply grateful to my lecturers Prof. Nicoletta Balbo, who instilled in me the passion for empirical analysis, and Prof. Thorsten Grohsjean, who encouraged this idea from the very beginning.

This thesis would not have been possible without the aid and support of my friends and valuable proofreaders Edoardo, Manuele, Marco and Valentina. I owe them much.

Last but by no means least, I wish to thank my parents, my great emotional (and financial...) supporters who every day encourage me in all of my passions and inspire me to chase my dreams.

A. Appendix

A.1. Descriptive Statistics for the Dependent Variable

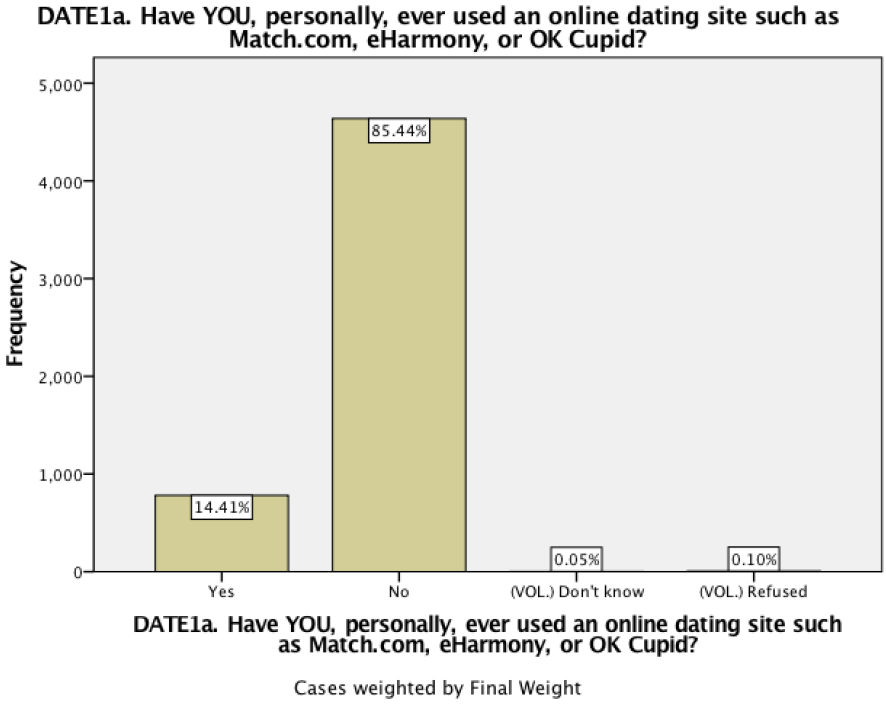

Figure A.1.: DATE1A: frequency bar charts.

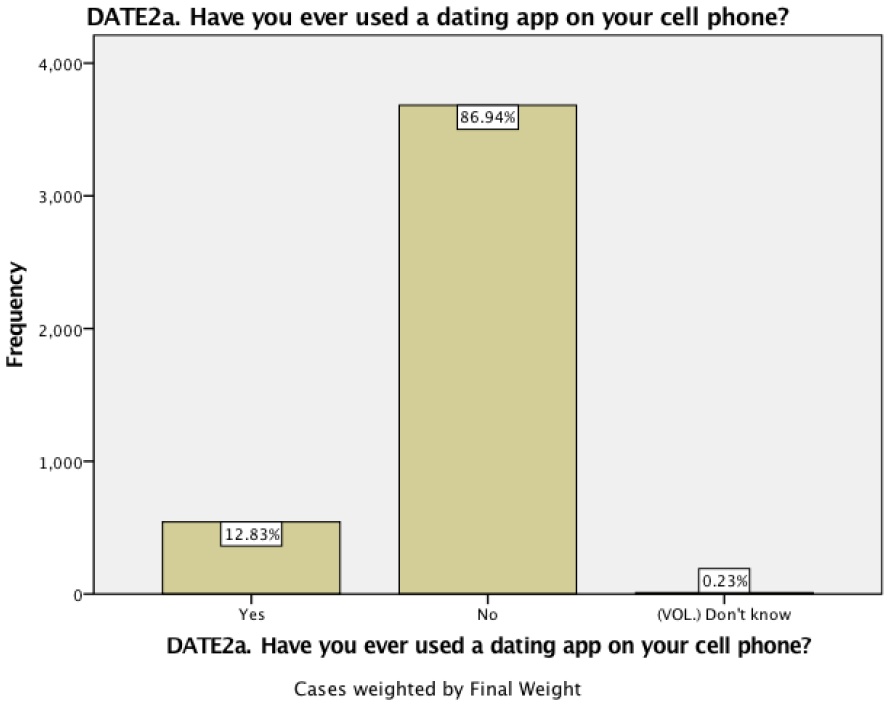

Figure A.2.: DATE2A: frequency bar charts.

			Std.	Missing		N. of Ext.[ab]		EM
	N	Mean	Deviation	Count	Percent	Low	High	Means[c]
DATE1A	4681	1.86	.443	737	13.6	.	.	1.86
DATE2A	3664	1.88	.456	1754	32.4	.	.	1.89

Table A.1.: Missing Value Analysis.

[a]Number of extreme cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).
[b]. indicates that the inter-quartile range (IQR) is zero.
[c]Little's MCAR test: Chi-Square = 10,057, DF = 2, Sig. = ,007

			EMINUSE		
			Yes	**No**	**Total**
DATE1A	**Yes**	Count	769	13	782
		% of Tot.	14.2%	0.2%	14.4%
	No	Count	4387	250	4637
		% of Tot.	80.8%	4.6%	85.4%
	Don't know	Count	3	0	3
		% of Tot.	0.1%	0.0%	0.1%
	Refused	Count	6	0	6
		% of Tot.	0.1%	0.0%	0.1%
	Total	Count	5165	263	5428
		% of Tot.	95.2%	4.8%	100.0%

Table A.2.: Crosstabs: DATE1A and EMINUSE.

		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	**No**	5297	84.5	84.5	84.5
	Yes	971	15.5	15.5	100.0
	Total	6267	100.0	100.0	

Table A.3.: ISONLINEDATER: frequency table.

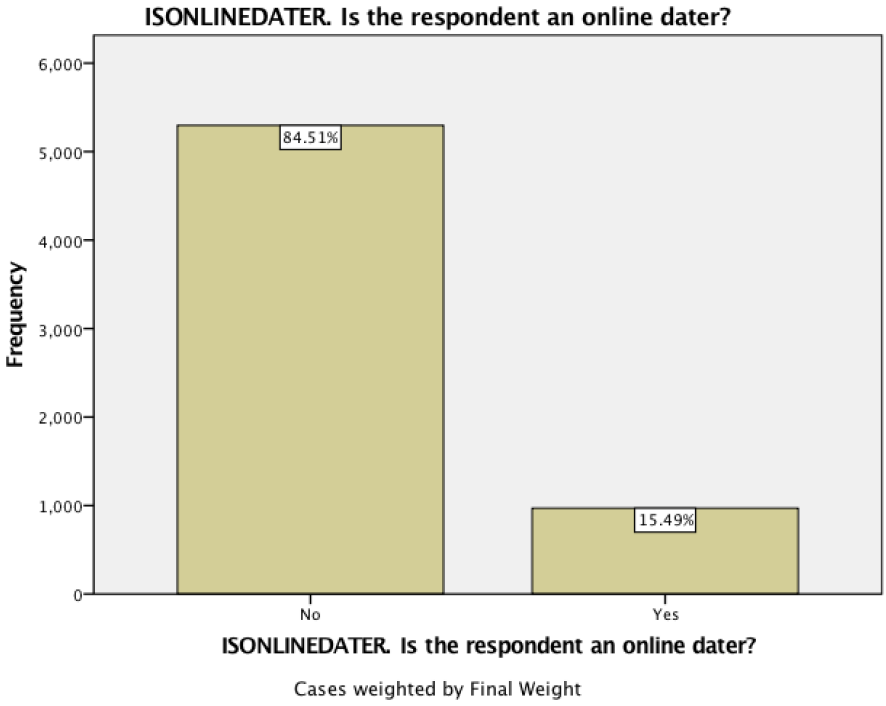

Figure A.3.: ISONLINEDATER: frequency bar charts.

A.2. Descriptive Statistics for Independent Variables

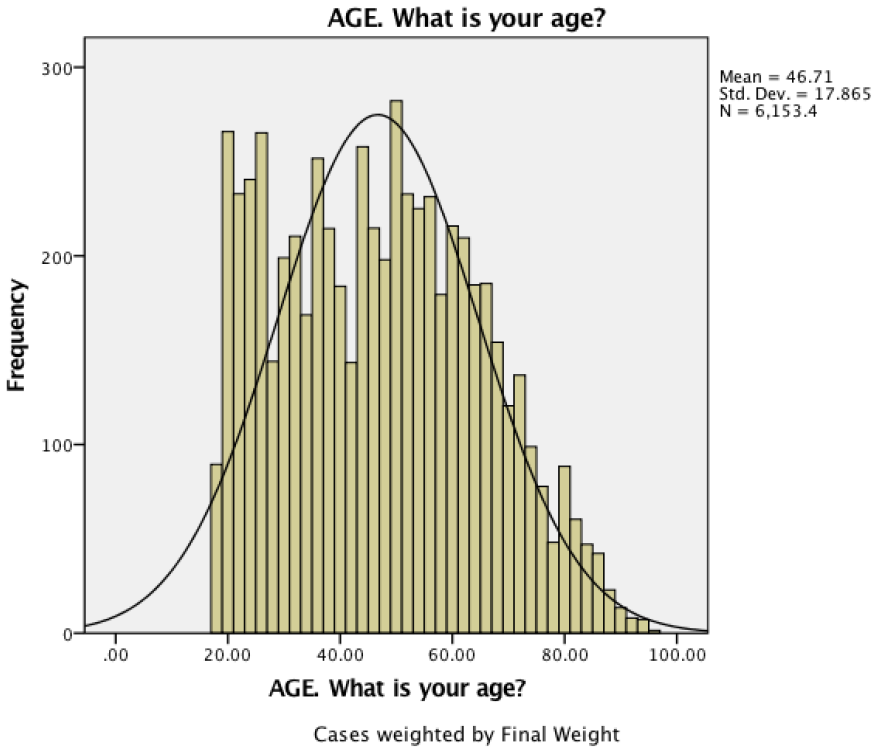

Figure A.4.: AGE: histogram.

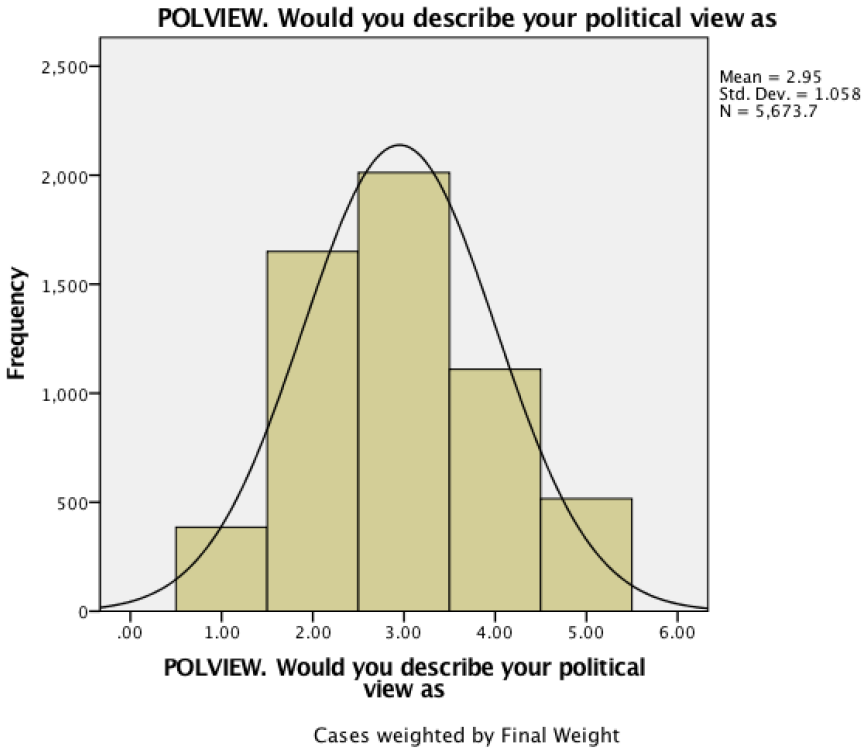

Figure A.5.: POLVIEW: histogram.

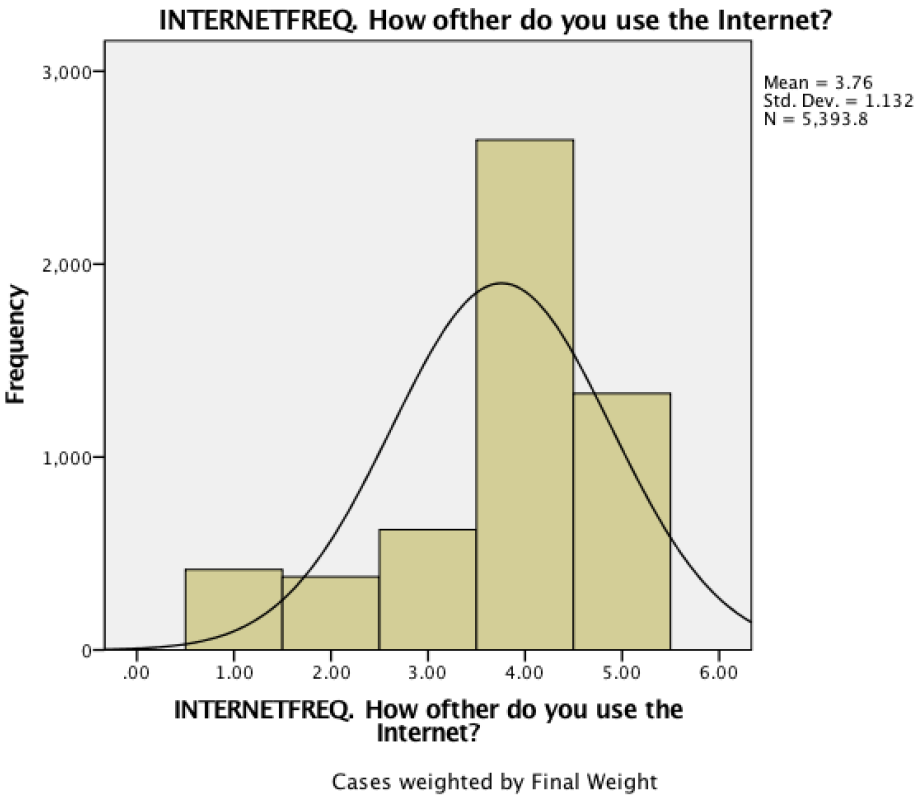

Figure A.6.: INTERNETFREQ: histogram.

	N				Std.		Std. Error		Std. Error		
	Valid	**Missing**	**Mean**	**Median**	**Deviation**	**Skewness**	**of Skewness**	**Kurtosis**	**of Kurtosis**	**Min.**	**Max.**
AGE	6153	114	46.7150	46.0000	17.8647	.230	.031	-.839	.062	18.00	96.00
POLVIEW	5674	594	2.9508	3.0000	1.05832	.215	.033	-.557	.065	1.00	5.00
INTERNETFREQ	5394	874	3.7582	4.0000	1.13188	-1.087	.033	.520	.067	1.00	5.00

Table A.4.: Descriptive Statistics for AGE, POLVIEW and INTERNETFREQ.

A.3. Multiple Regression Models

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	43.363	6	7.227	57.813	.000
	Residual	750.379	6003	.125		
	Total	793.742	6009			
2	Regression	69.431	8	8.679	76.651	.000
	Residual	548.963	4840	.113		
	Total	617.494	4848			
3	Regression	70.150	10	7.015	59.967	.000
	Residual	521.430	4457	.117		
	Total	591.580	4467			
4	Regression	67.372	13	5.182	36.670	.000
	Residual	478.591	3664	.131		
	Total	545.963	3677			
5	Regression	51.080	14	3.649	26.201	.000
	Residual	353.778	2541	.139		
	Total	404.857	2555			
H_1	Regression	44.642	6	7.440	59.542	.000
	Residual	748.862	5993	.125		
	Total	793.504	5999			

Table A.5.: ANOVA of the different models.

Dependent Variable: ISONLINEDATER. Is the respondent an online dater?
Predictors: see Table 3.3.

Bibliography

Abramova, O., Baumann, A., Krasnova, H., and Buxmann, P. (2016). Gender Differences in Online Dating: What Do We Know So Far? A Systematic Literature Review. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*.

Becker, G. S. (1974). A Theory of Marriage. *Economics of the Family: Marriage, Children, and Human Capital*.

Becker, G. S. (1991). *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Bernard, A. and Adelman, M. B. (1990). Market Metaphors For Meeting Mates. *Advances in Consumer Research*, 17.

Blossfeld, H.-P. (2009). Educational Assortative Marriage in Comparative Perspective. *Annual Review of Sociology*, 35:513–530.

Blow, A. J. and Hartnett, K. (2005). Infidelity in committed relationships II: A substantive review. *Journal of Marital and Family Therapy*, 31(2):217–233.

BrightEdge (2017). 57% of Search Traffic is Now Mobile. Technical report, BrightEdge.

Cacioppo, J. T., Cacioppo, S., Gonzaga, G. C., Ogburn, E. L., and VanderWeele, T. J. (2013). Marital satisfaction and break-ups differ across on-line and off-line meeting venues. *Proceedings of the National Academy of Sciences*, 110.

Cartwright, D. and Harary, F. (1956). Structural Balance: A Generalization of Heider's Theory. *The Psychological Review*, 63(5).

DiMaggio, P., Hargittai, E., Neuman, W. R., and Robinson, J. P. (2001). Social Implications of the Internet. *Annual Review of Sociology*, 27:307–336.

Evans, D. S. (2003). Some Empirical Aspects of Multi-sided Platform Industries. *Review of Network Economics*, 2(3):191–209.

Finkel, E. J., Eastwick, P. W., Karney, B. R., Reis, H. T., and Sprecher, S. (2012). Online Dating: A Critical Analysis From the Perspective of Psychological Science. *Psychological Science in the Public Interest*, 13(1):3–66.

Fiore, A. T. (2004). *Romantic Regressions. An Analysis of Behavior in Online Dating Systems.* PhD thesis, Massachusetts Institute of Technology.

Fiore, A. T. and Donath, J. S. (2005). Homophily in Online Dating: When Do You Like Someone Like Yourself? In *Proceedings of the 2005 Conference on Human Factors in Computing Systems, CHI 2005.*

Frey, B. S. and Eichenberger, R. (1996). Marriage Paradoxes. *Rationality and Society*, 8(2):187–206.

Granovetter, M. S. (1973). The Strenght of Weak Ties. *American Journal of Sociology*, 78:1360–1380.

Granovetter, M. S. (1974). *Getting a Job.* Harvard University Press, Cambridge.

Hagiu, A. and Wright, J. (2015). Multi-Sided Platforms. *Harvard Business School Working Paper*, 15-037.

Hakim, C. (2010). Erotic Capital. *European Sociological Review*, 26(5):499–518.

Hakim, C. (2011). *Erotic Capital: the Power of Attraction in the Boardroom and the Bedroom.* New York: Basic Books.

Hancock, J. T., Toma, C., and Ellison, N. (2007). The truth about lying in online dating profiles. In *Proceedings of the 2007 Conference on Human Factors in Computing Systems, CHI 2007.*

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology: Interdisciplinary and Applied*, 21:107–112.

Heino, R. D., Ellison, N. B., and Gibbs, J. L. (2010). Relationshopping: Investigating

the market metaphor in online dating. *Journal of Social and Personal Relationships*, 27(4):427–447.

Hirschman, E. C. (1987). People as Products: Analysis of a Complex Marketing Exchange. *Journal of Marketing*, 51(1):98–108.

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2010). What Makes You Click? – Mate Preferences in Online Dating. *Quantitative Marketing and Economics*, 8(4):393–427.

Illouz, E. (1997). *Consuming the Romantic Utopia: Love and the Cultural Contradictions of Capitalism*. University of California Press.

Illouz, E. (2007). *Cold Intimacies: The Making of Emotional Capitalism*. Polity Press.

Illouz, E. and Finkelman, S. (2009). An Odd and Inseparable Couple: Emotion and Rationality in Partner Selection. *Theory and Society*, 38(4):401–422.

Lin, K.-H. and Lundquist, J. (2013). Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education. *American Journal of Sociology*, 119(1):183–215.

Mitchell, J. C. (1969). *Social Networks in Urban Situations*. Manchester University Press.

Moreno, J. L. (1934). *Who Shall Survive?* Beacon Press.

Ortega, J. and Hergovich, P. (2017). The strength of absent ties: Social integration via online dating. *arXiv.org*.

Pawowski, B. and Dunbar, R. I. M. (1999). Impact of market value on human mate choice decisions. *Proceedings of the Royal Society of London B*, 266(281-285).

Qian, Z. (2002). Race and social distance: intermarriage with non-latino whites. *Race & Society*, 5(1):33–47.

Regan, P. C., Levin, L., Sprecher, S., Christopher, F. S., and Gate, R. (2000). Partner Preferences: What Characteristics Do Men and Women Desire in Their Short-Term Sexual and Long-Term Romantic Partners? *Journal of Psychology & Human Sexuality*, 12(3):1–21.

Rochet, J.-C. and Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, 1(4):990–1029.

Rosenfeld, M. J. (2017). Marriage, Choice, and Couplehood in the Age of the Internet. *Sociological Science*, 4:490–510.

Rosenfeld, M. J. and Thomas, R. J. (2012). Searching for a Mate: The Rise of the Internet as a Social Intermediary. *American Sociological Review*, 77(4):523–547.

Scharlott, B. W. and Christ, W. G. (1995). Overcoming Relationship-Initiation Barriers: The Impact of a Computer-Dating System on Sex Role, Shyness, and Appearance Inhibitions. *Computers in Human Behavior*, 11(2):191–204.

Schmitz, A. (2014). The Online Dating Market: Theoretical and Methodological Considerations. *Economic Sociology The European Electronic Newsletter*, 16(1):11–24.

Schmitz, A. (2016). *The Structure of Digital Partner Choice: A Bourdieusian Perspective*. Springer.

Scott, J. (1991). *Social Network Analysis. A handbook*. Sage Publications Ldt.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138.

Skopek, J., Schmitz, A., and Blossfeld, H.-P. (2011). The gendered dynamics of age preferences – Empirical evidence from online dating. *Journal of Family Research*, 23(3).

Stevenson, B. and Wolfers, J. (2007). Marriage and Divorce: Changes and Their Driving Forces. *The Journal of Economic Perspectives*, 21(2):27–52.

Sumter, S. R., Vandenbosch, L., and Ligtenberg, L. (2016). Love me Tinder: untangling emerging adults' motivations for using the dating application Tinder. *Telematics and Informatics*.

Timmermans, E. and Caluwé, E. D. (2017). To Tinder or not to Tinder, that's the question: An individual differences perspective to Tinder use and motives. *Personality and Individual Differences*, 110:74–79.

Wang, H. and Lu, X.-A. (2007). Cyberdating: Misinformation and (Dis)trust in Online Interaction. *Informing Science Journal*, 10.

Weiser, D. A., Niehuis, S., Flora, J., Punyanunt-Carter, N. M., Arias, V. S., and Baird, R. H. (2017). Swiping right: Sociosexuality, intentions to engage in infidelity, and infidelity experiences on Tinder. *Personality and Individual Differences*, 133:29–33.

Willis, J. and Todorov, A. (2006). First Impressions. Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science*, 17(7):592–598.

Wong, L. Y. (2003). Why Do Only 5.5% of Black Men Marry White Women? *International Economic Review*, 44(3):803–826.

Zillmann, D., Schmitz, A., and Blossfeld, H.-P. (2011). Liars have short legs. The relationship between deceptive self-presentation and mating opportunities in online dating. *Zeitschrift fur Familienforschung*, 23(3):291–318.