ARE ALL DATING APPS CREATED EQUAL: COMPARATIVE SIMULATIONS OF MOBILE DATING APPLICATIONS THROUGH LEVELS OF MULTIPLICITY

Fatma Abdulwahab
Collin Henley
Steven Hoang
Liam McKenna
Joseph A.E. Shaheen, Ph.D

Department of Computational and Data Sciences
George Mason University
4400 University Drive
Fairfax, VA, USA
jshaheen@gmu.edu

ABSTRACT

This experiment seeks to compare the effectiveness of simple vs. complex mobile dating application interfaces in producing matches for users. Agent-based simulations were utilized to replicate today's most popular mobile dating applications based on varying levels of behavioral complexity. Agents in each simulation are assigned gender and randomized attributes. Male and female agents engage in pairing activity based on rule-sets drawn from academic literature. We observed a direct relationship between multiplicity in feature availability allowed by simulated platforms and matching quantity and observed differences in match network structure. We validate and expand our findings with a robust statistical model (Multiple Regression Quadratic Assignment Procedure) controlling for network dependencies.

Keywords: online dating, social networks, agent-based modeling, mobile applications, MRQAP.

1 INTRODUCTION

Mobile dating applications and sites demonstrated an exponential growth in user adoption in recent years, with 15 of the most popular instances acquiring 247 million downloads in 2018 alone (Koch 2019). This domain has also proven to be a high revenue generator for businesses, garnering over $2.1 billion in gross sales and was projected to reach $4.5 billion by 2020 (Lin).

The surge in mobile dating application's revenue is driven by several factors including changing social norms and a common acceptance of online dating. In a study conducted by Pew Research Center, 44% of Americans indicated that "online dating is a good way to meet people" in 2005, and 29% expressed that "people who use online dating sites are desperate" (McGrath 2015). When participants were asked the same set of questions in 2013, the narrative of mobile dating applications shifted, yielding 59% and 21%, respectively, and consequently indicating that "online dating has lost much of its stigma."

Today's mainstream mobile dating applications offer common features and a market-borne universal user interface, but continue to vary in levels of multiplicity–that is–the number of features available to users to achieve the same result; In the case of mobile dating applications, the desired result is often finding a suitable match.

One important example of a ubiquitous mobile dating application is Tinder with its well-known and simple interface, a Match Group Inc. dating application that is credited with bringing online dating into the mainstream. It is estimated that over 15 million matches occur on Tinder every day (Lin). The success of Tinder is due to its user-friendly interface: Typically after initial download, users are prompted to create a profile to showcase themselves to potential matches. These profiles consist of images uploaded by the user and a single field for profile summary. After users are satisfied with their profile, they choose matching preferences based on age, gender, distance, and location. Users sort through profiles of potential matches shown to them one at a time by the application's recommendation algorithm. Users then perform an action to either accept or reject the profile they have viewed. After each decision action has been made, the user is presented with another profile where the acceptance/rejection process is repeated. Consequently, when two users like each other, a notification is displayed. This is deemed a 'match' and the users are now able to communicate.

As Tinder's popularity grew, other mobile dating applications emerged with their own rule-sets and interfaces such as Hinge, which allows users multiple pathways to match, extending the functionality found on Tinder and providing more powerful filters. Users also have the option to create more detailed profiles that specify ethnicity, political views, desire to have children, and other attributes and preferences. Users can also enrich their profiles with clever prompts that enhance their appeal to visitors. Visitors can like or comment on those prompts. When mutual actions occur, users are then 'matched'. The central difference between Tinder-like applications and Hinge-like applications being that Tinder requires that users like each other before they 'match' and then communication may proceed, while Hinge allows users to send messages with their 'like' if they desire to do so.

These additional features differentiate Hinge from Tinder in important ways and represent additional features that create more pathways to reach a 'match' which we dub, *multiplicity*. We argue that this multiplicity of dating applications plays a role in the aggregate outcomes for each user and that users can navigate the range of available applications to maximize their dating potential and utility. For example, users that desire a more complex approach which offers more options can use CoffeeMeetsBagel, another mobile dating application which has all of the features of Tinder and Hinge with added layers of multiplicity, and that for some users this choice may be more productive.

Given the wide array of available mobile dating applications, each with their differing levels of multiplicity, analysis is necessary to determine whether more features generate more user matches. One of the primary key performance indicators of a successful dating application is the number of matches the application produces for each user. Consequently, multiplicity could be directly correlated with success. Thus, we have motivated our examination of the multiplicity of online dating applications.

In this paper, we construct two dating application models, one with a low multiplicity level representing applications such as Tinder, and another with a higher level of multiplicity representing applications such as Hinge. We dub them Multiplicity Level 1 (M1) and Multiplicity Level 2 (M2). These models were assigned rule-sets with the purpose of replicating their respective real-world counterparts. Our approach combines agent-based modeling and classical social network analysis as our go-to analytical frameworks. Our constructed networks contain various edge types that are set to reflect acceptance actions and clicks from each individual user as they use the interface. We define a 'like' as a one-sided acceptance of another's profile, a 'match' as a dyadic agreement, a 'dislike' as a rejection, and a 'message' as combination action (like with a higher probability of a response). Through implementation of simulations, we investigate the contribution of various mechanisms, or lack of, in generating matches.

This paper advances the state of the art in dating simulations in several ways. Primarily, we rely on social science theory to simulate an application's interface using well-proven social effects drawn from literature in the absence of relevant data which is simply unobtainable. Although this approach is not common, it is necessary for our usage. Data does exist but in two forms, each of which are either not useful or offer limited use for our purposes: The first is aggregate data from non-peer reviewed sources, which is not held to high empirical standards in collection or reporting. Furthermore, the connection between aggregate data sources and agent rule probabilities has been shown to be non-linear and mathematically intractable (Epstein 1999) rendering such data irrelevant for model development, and in the absence of complete aggregate data sets, insufficient for external validation. The second form–which we utilize in a stylized manner–is drawn from empirically reviewed sources Tyson2016, and offers agent attribute data, general distributional data and insight into the mechanics of mate selection. These datasets suffice in offering insight into our central question: How does the multiplicity of dating applications affect agent and aggregate "matching" outcomes? Thus, we make good use of this data.

1.1 Behavioral Background

Homophily in Social Networks

The principle of network homophily implies that "people's personal networks are homogeneous with regard to many socio-demographic, behavioral, and intrapersonal characteristics" (McPherson, Smith-Lovin, and Cook 2001). This social dynamic implies that people prefer those who are similar to themselves. As a cognitive process, it has been shown to be correlated with a sociological concept McPherson calls "con-structuralism", the possibility that shared knowledge, beliefs and attributes incentivize positive interactions. In the context of online dating, perceived similarity has been shown to correlate with attraction. Fiore and Donath (2005) measure the effects of homophily by sampling a pool of users on an unnamed online dating application and analyze their preferences for potential partners that are similar to themselves. This included similarities like desire for children, level of education, and physical appearance which were found to account for a better likelihood of a match than random. We rely on this sociological framework and incorporate it in our models, though admittedly, other sociological processes are likely at play. Nonetheless, homophily is a strong foundational driver for our virtual experiments. Specifically, we incorporate general notions of age, physical attractiveness, ethnicity and gender. And, since the features of these attributes are a strong mix of discrete (ethnicity, gender), continuous (age, attractiveness), uniform (age, attractiveness, ethnicity, and gender) and, Gaussian (overall compatibility score), they represent the minimum required to investigate our central question.

2 MATERIALS AND METHODS

2.1 Model Mechanics

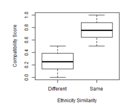
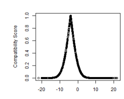
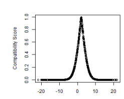
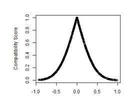
We use the Python module, NetworkX, to build our model and run our simulations. To issue a direct comparison, the number of agent attributes were kept to a minimum and consistent across both models. We instantiate our synthetic population with a binary gender (male and female), a generic ethnicity chosen from 4 levels, physical attractiveness (between 0 and 1), and age (18-40). All agents attributes were assigned uniformly. In both models, we restrict population size to 1000 agents. In every turn of the simulation, agents are presented with up to 40 agents of the opposite sex and within 10 years of age for consideration. Each agent then evaluates the compatibility score of the considered agent. The compatibility score is a

Table 1: Summary of Input Distributions

Input	Variable Type	Assignment
Ethnicity ($\mathscr{E}$)	Discrete Uniform	U[0,1,2,3]
Age ($\mathscr{A}$)	Continuous Uniform	U[18,40]
Physical Attractiveness ($\mathscr{P}$)	Continuous Uniform	U[0,1]

score based on the attributes of the considered agent. If the compatibility score exceeds ego's compatibility threshold, the probability of a "like" is set to maximum parameter value. If this threshold is not reached the probability of a "like" is minimized. We chose probability parameters arbitrarily in the absence of data. If both agents record a bi-directional "like", a "match" is created. This *threshold* assignment is consistent with social threshold models (Granovetter 1978) and their dating counterparts (Hitsch, Hortaçsu, and Ariely 2010). We summarize all agent rule calculations in Table 2.

Table 2: Rules Compatibility Score Calculation

Model Attribute	Condition	Value	Description
Compatibility Score (ethnicity)	$\mathscr{E}_M \top \mathscr{E}_F$ $\mathscr{E}_M \perp \mathscr{E}_F$	$S_e = (\frac{\mathscr{X}}{2}) + 0.5$ $S_e = (\frac{\mathscr{X}}{2})$	 $\mathscr{X} \in [0,1]$
Compatibility Score (age)	$\mathscr{P} == M$	$S_M^{\mathscr{A}} = 1 - (2)(\lvert\mathscr{N}(\mathscr{A})_{cdf}\rvert - 0.5)$ $\mathscr{A} = \frac{1}{2}[1 + erf(\frac{x-\mu}{\sigma\sqrt{2}})$	 $\mu = -3.996$ $\sigma = 3.317$ $x = \mathscr{A}_M - \mathscr{A}_F$
Compatibility Score (age)	$\mathscr{P} == F$	$S_M^{\mathscr{A}} = 1 - (2)(\lvert\mathscr{N}(\mathscr{A})_{cdf}\rvert - 0.5)$ $\mathscr{A} = \frac{1}{2}[1 + erf(\frac{x-\mu}{\sigma\sqrt{2}})$	 $\mu = 2.046$ $\sigma = 2.906$ $x = \mathscr{A}_F - \mathscr{A}_M$
Compatibility Score (attractiveness)		$S_M^{\mathscr{P}} = 1 - (2)(\lvert\mathscr{N}(\mathscr{P})_{cdf}\rvert - 0.5)$ $\mathscr{P} = \frac{1}{2}[1 + erf(\frac{x-\mu}{\sigma\sqrt{2}})$	 $\mu = 0$ $\sigma = 0.341$ $x = \mathscr{A}_M - \mathscr{A}_F$
Total Score		$S_{Total}^M = \beta_1^M S_e^M + \beta_2^M S_M^{\mathscr{P}} + \beta_3^M S_M^{\mathscr{A}}$	$\beta_1^M = 0.3$ $\beta_1^F = 0.4$ $\beta_2^M = 0.3$ $\beta_2^F = 0.3$ $\beta_3^M = 0.4$ $\beta_3^F = 0.3$

2.2 Agent Preferences

In both of our models, agents will favor homophily in selection as referenced in McPherson et al's (2001) based on ethnicity and attractiveness, while holding differential preferences in age (male agents preferring younger female agents and female agents preferring older male agents). As agents are randomly selected, agents will evaluate whether to "like" each other based on these attributes and an arbitrarily set threshold.

Generally, threshold levels determine the likelihood of "like" or "dislike" events occurring and are summarized in Figures 1 and 2.

We base our assumptions on several important findings from peer-reviewed literature on online dating. To begin, male agents were given a preference for younger female agents where the mean difference was 4.00 years, and female agents were given a preference for older males with a mean difference of 2.05 years (Conway, Noë, Stulp, and Pollet 2015). The compatibility score is intended to ensure that maximum score contributed by age resides at this mean, and the greater the difference from this mean difference he lower the contribution of age to the total compatibility score. Physical attractiveness and ethnicity are based on similar principles. Agents evaluate ethnicity based on the concept of homophily, with much higher scores associated with agents who consider agents of the same ethnicity, and agents who are similar in physical attractive levels. For physical attractiveness and ethnicity we rely on arbitrary parameter values but intuitive rule calculations in the absence of data.

We also utilized a weighing mechanism for both male and female agents as genders may value attributes differently. In line with Greenlees and McGrew's (1994), male agents were given a higher weight ($\beta_3^M = 0.4$) for attractiveness preference than for female agents ($\beta_3^F = 0.3$) as males value physical features more than women. Females place greater weight in finding a potential match of their own ethnicity (Fisman, Iyengar, Kamenica, and Simonson 2006) and so we adjust the weighing and contribution for ethnicity to account for this ($\beta_1^F = 0.4$, $\beta_1^M = 0.3$). Finally, the weighing of scores on age by both genders remained the same (0.3) as there was little evidence that the contribution to a total score was different from both genders even in light of different age preferences for males ($\mu = -4.0$) than for females ($\mu = 2.05$).

Since our model is subject to matching process fundamentals (Mortensen 1982), synthetic population structure and heterogeneity play an important role in model specification. The choice of male to female agent ratios could–if chosen unwisely–create unintended consequences and artifacts in our model. Thus, we rely on McGrath (2015): The author reports that Tinder's user base is roughly 60% male and 40% female (McGrath 2015). Though this statistic is a rough estimate drawn from non-peer-reviewed sources, therefore we consider it a guide (and not a rule) to our implementation and model specification. As a result, we chose population parameters consistent with his overall notion (that there are many more men than women on dating applications) at a rate of 68% men and 32% women in line with a general notion of a 2:1 ratio.

Once agents are presented with (up to 40) other agents of a different gender to evaluate they proceed with said evaluation ($S \in [0,1]$) against a pre-set parameter to determine whether they will "like" the considered agent(s). These threshold parameters are assigned arbitrarily in the absence of data but follow the following intuitive patterns:

- Female agents threshold for a "like" are higher than male agents (Bruch and Newman 2018) [M1 & M2].
- The probability of a "like" when the compatibility threshold is not achieved ($S_{compatibility} < S_{Threshold}$) is not zero. It is reduced but a "like" may still occur. [M1 & M2]
- A "like with a message" is still subject to the same compatibility thresholds but increases the probability of a like. [M2]

All thresh-holds, probabilities, and decision junctures are summarized in Figures 1 and 2. And, while threshold levels and probabilities are homogeneous, each agent's compatibility score for every other considered agent is heterogeneous. This ensures that agents' internal standard for acceptance of a potential mate are heterogeneous and that agents maintain an individual preference in agent-agent interactions.

To illustrate, consider a female agent F_1 with ethnicity = E, age = 30, and attractiveness = 0 or {E, 30, 0} who evaluates a male agent M_1 with attributes {E, 32.046, 0}. Given our model's rules F_1 will calculate that M_1's compatibility with her is $0.4\left(0.5+\frac{x \in U[0,1]}{2}\right)+0.3(1)+0.3(1)$. If we assume that $x \in U[0,1]$ = 0.5, then F_1 will view her compatibility with M_1 as $0.3+0.3+0.3=0.9$ or 90%. The M1 model specifies that since her compatibility score is greater than the designated threshold, 0.5 (Figure 1), then there is an 80% probability of a "like" occurring. If the compatibility score was less than 0.5, the probability of a "like" would be 30%.

This previous example illustrates the general mechanics of the M1 model (e.g. Tinder) where a message may not be sent with each "like" interaction. In model M2's specification (e.g. Hinge) two key differences exist. Firstly, agents may choose to send a "like" or a "like with a message". We assume that a "like with a message" increases the probability of a reciprocated like (we assume messages are positive and well-suited). As a consequence it is then obvious that the second key difference between our specified M1 and M2 Models is that ego in M2 consider the "like" of the alter in their decision. This latter effect is known as reciprocity (Altman 1973) and is a direct result of the difference in our hypothetical application's user interfaces and multiplicity–the difference being that applications such as Hinge (M2) allow users to attach a message with a like prior to matching, while applications such as Tinder (M1) do not. This forces users to evaluate incoming interactions not only based on the sender's attributes, but on the message attached to the "like" as well. Finally, we should report that in the absence of data we specified that all decision points are Bernoulli tests in (U[0,1]) and that coefficients attached to our formulaic rules (Table 1) scale the compatibility score to be within [0,1] to ensure a one-to-one comparison scheme.

3 MODEL COMPARISON

As we have noted, our objective was to compare our two models with varied multiplicity levels in terms of aggregate outcomes. We chose to focus our effort on matches that agents accumulate in each model. We carry out a formal inquiry of the statistical parameters and distributions of likes, dislikes, messages, and matches as well as evaluate conditional distributions though a network regression–the Multiple Regression Quadratic Assignment Procedure–a suitable method for our analysis that accounts for our model's rule dependency without inappropriate assumptions.

The Multiple Regression Quadratic Assignment Procedure (MRQAP) is used to test for significance in an observed correlation (Krackardt 1987) where dependency between two or more dyadic relations may exist. It is a non-parametric, permutation-based test that preserves the integrity of the observed structures (Krackardt 1987). This approach was originally created by Mantel (1967) to identify geographic clustering of diseases (Mantel 1967). It is well known that geographic data is highly dependent (not i.i.d) and thus is unsuitable for regression models that assume independent observations. Since then, the method has been developed and deployed as a mainstream network analysis tool. MRQAP is particularly useful when calculating coefficient magnitudes and parameter estimates, and as a corollary, the strength of social effects–through permutations of network statistics. Since our synthetic population samples, rules, attributes, and mechanisms are not identically and independently distributed (i.i.d), we must rely on this powerful analytical tool. We omit a comprehensive discussion of the nature of this method at this time, but direct the reader to aforementioned citations.

MRQAP is well suited for our comparative aims: Consider that M1 (Figure 1) emulates the simplest dating application interface–an example we cite is Tinder. We do so by framing likes and matches as edges in a network (ties), and thus we frame our output analysis as a network analysis with edges within both simulations' code-base reflecting interactions through the application's interface. Framing our analysis in this way allows for the use of MRQAP and its powerful permutation-based statistical method.

To compare M1 to M2 by using the regression's output, we consider their shared and common output. In our case the common output is matches. We ensure that inputs among the two models are also shared for a one-to-one comparison. A synthetic population instantiaed for M1 can apply M2's behavioral rules without loss of generality. We use ethnicity, age, physical attractiveness as input attributes, and likes, dislikes as input variables. Clearly, if there is a statistical difference in the models' parameter estimates then since we have randomly generated synthetic populations in both models according to the same rules, then this difference must be due to the difference in agent behaviors between the two models. In our case, the difference between M1 and M2 lies squarely in an increase in multiplicity where M2 allows for the sending of a message with a like and M1 does not–hence, reciprocity as a social effect. In Section 4, we present typical results from a representative run of n=1000 nodes for both model types.

3.1 Internal Validation

Internal validation of our model parmeters, inputs and ouputs were sufficient to answer our simple question. Our primary method of internal validation was to ensure that all synthetic populations were generated to our strict requirements of uniformity for attractiveness [0,1], age [18-40], and ethnicity (categorical) [A, B, C. D]. Both Tables 3 and 3 show the results of a representative comparison between generations of a synthetic population. It agrees with our expectations and supplies confidence that our code-base is reflective of our model's specifications. Since our goal is to quantify the difference of the models based on the given multiplicity of dating apps or the *structure* of this particular system, no further statistical validation was necessary since the outputs-given-structural-differences is what we had hoped to determine.

To elaborate further, we would expect that the mean attractiveness for our population would be roughly 0.5 since we designated this variable as U[0,1], 29 years of age as the mean value between 18 and 40, and for ethnicity–a categorical variable–to be equally scaled among the $0^{th}, 25^{th}, 50^{th}$, and 75^{th} percentiles (lower bounds). This is clearly shown in our reported tables.

Table 3: Validation of Simulation Inputs for M1 Model

Statistic	N	Mean	0^{th}	25^{th}	50^{th}	75^{th}	100^{th}
Attractiveness	1,000	0.521	0.003	0.282	0.512	0.778	0.999
Age	1,000	28.960	18	23	29	35	40
Ethnicity	1,000	2.509	1	2	3	4	4

Table 4: Validation of Simulation Inputs for M2 Model

Statistic	N	Mean	0^{th}	25^{th}	50^{th}	75^{th}	100^{th}
Attractiveness	1,000	0.516	0.001	0.257	0.501	0.769	1.000
Age	1,000	29.057	19	24	29	34	39
Ethnicity	1,000	2.446	1	1	2	3	4

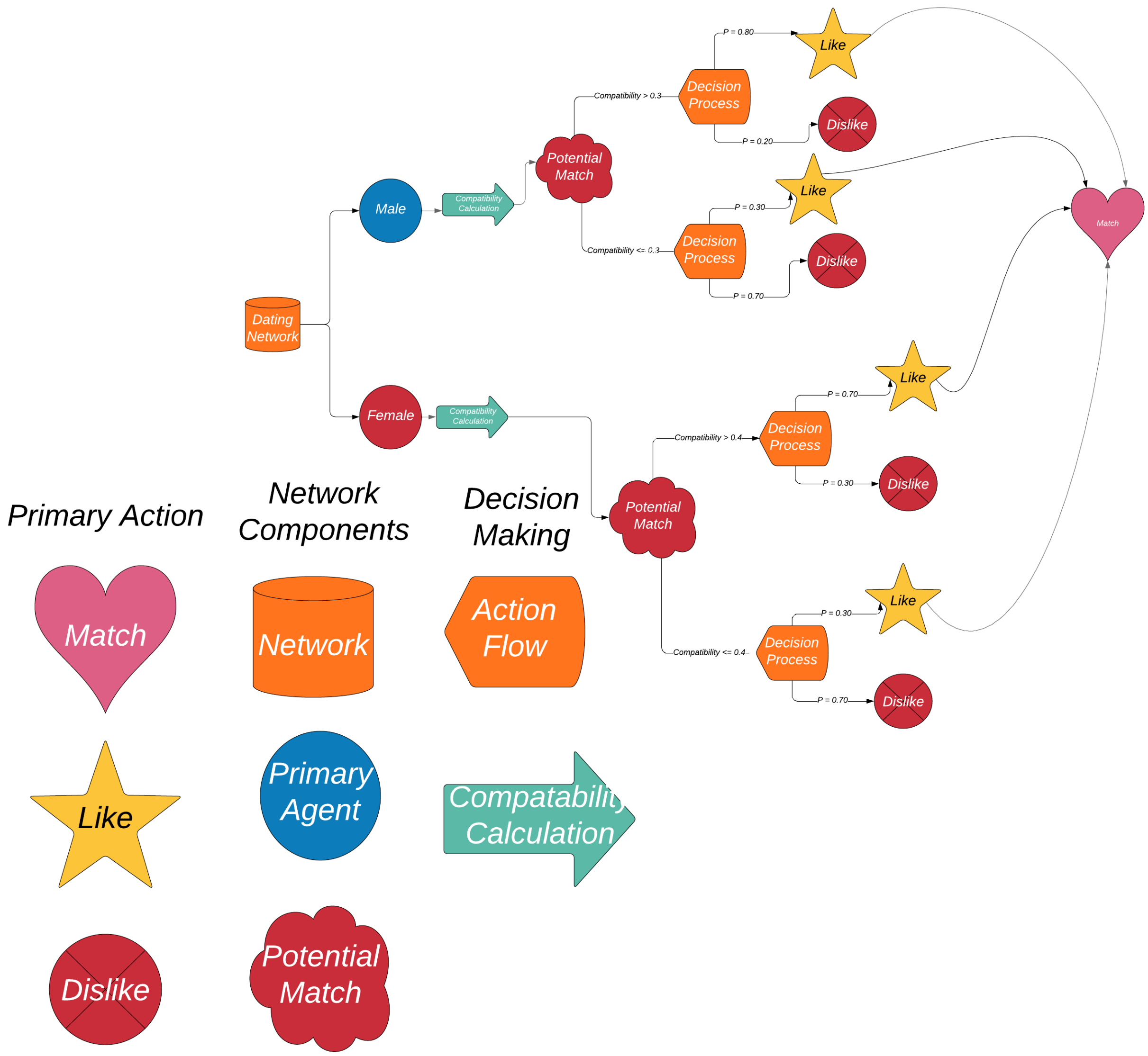


Figure 1: Multiplicity Level One (M1) Decision Diagrams

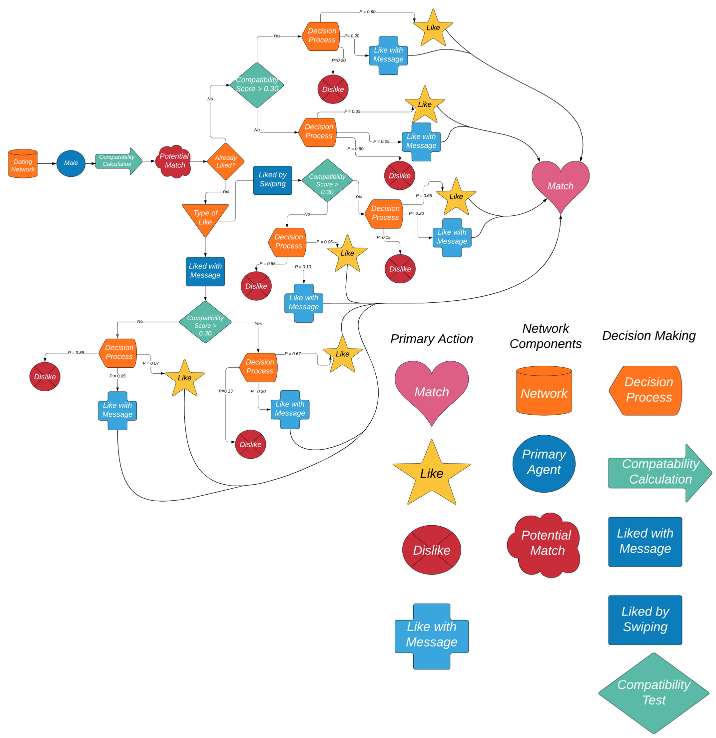


Figure 2: Multiplicity Level Two (M2) Decision Diagrams (Males)

4 RESULTS

4.1 Gender-Based Distributions of 'Likes', 'Dislikes', 'Matches' and 'Messages'

Figures 1 and 2 summarize the frequency distributions for both M1 (eg. Tinder) and M2 (e.g. Hinge) model runs. We found similar overall behavior in both M1 and M2, specifically in the general shape and location of the frequency distributions. However, there are some telling differences. Male and female agents in both models liked and disliked other agents according to some bell-shaped distribution. This was anticipated and is likely a result of the use of Gaussian differential inputs in age and attractiveness. On average, female agents liked and disliked more male agents as the simulations were initiated with more male than female agents.[1] Because both models can essentially be viewed as matching processes, lower counts of female agents mean that they will 'dislike' more but also 'like' more–since the male agents possess greater diversity (variance) and thus more likely to overcome female agents' "threshold".

However, while female agents (in both models) were more active, the matching distribution did not reflect this activity. The majority of male and female agents received no matches. And while more female than male agents in both models (M1 and M2) received a singular match, more male agents received 2 or more matches in M1 (Figure 3c) or roughly an equal amount (Figure 4c) in M2. A reasonable interpretation is that female agents were more likely to find a match of any quality while there were many male agents unable to find any matches, but that male agents who were highly valued receive many matches. This entailed that both models' marginal distribution of matches held a right-skew. Finally, the messaging distribution was both skewed and followed similar patterns to the match distribution. Overall, we observed similar stylized facts in our analysis when conditioning on gender between M1 and M2 models.

4.2 Overall Comparison of M1 and M2 Using a Paired t-test

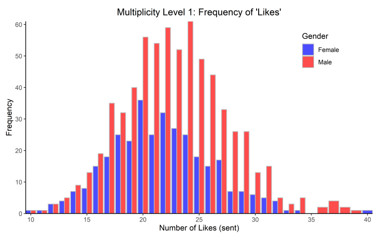
M1, with its simpler features and focus on generating likes and matches exceeded M2 in scaling activity. Tables 5 and 6 provide summary statistics for both M1 and M2. It's important to note that the average number of likes in M1 was 4 times the number of likes in M2, but nonetheless that the number of likes necessary to produce a match in M1 was approximately 20 and in M2 was 17. That is–while the number of likes occurring in M1 was quadruple the number of likes in M2, M1's agent rules failed to produce more matches per like though it did produce more matches overall. Furthermore, because agents in M2 considered more features and more ways to match, the number of 'dislikes' reflected this behavior.

Table 5: M1 Model Run Summary Statistics

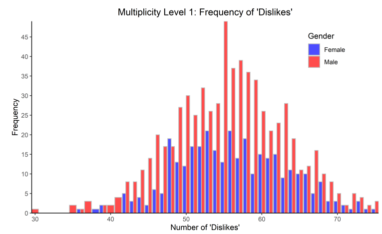
Statistic	N	μ	σ	0^{th}	25^{th}	75^{th}	100^{th}
'Like' Count	1,000	22.449	4.662	10	19	25	40
'Dislike' Count	1,000	55.878	7.360	30	51	61	75
'Match' Count	1,000	1.129	1.152	0	0	2	6
'Like Message' Count	1,000	0.609	0.917	0	0	1	5

Table 7 summarized the comparison of basic statistics using a paired t-test. We found distinct differences between the models. All results are significant. As shown, M1 (e.g Tinder) exceeded M2 (e.g Hinge) in producing more likes, matches, and messages, but less dislikes.

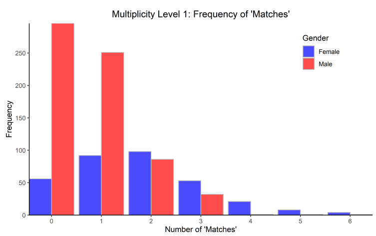
[1] In the methods section we discussed our choices for gender proportions.



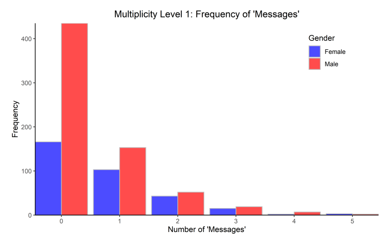
(a) Frequency of 'Likes' for males (red) and females (blue)



(b) Frequency of 'Dislikes' for males (red) and females (blue)

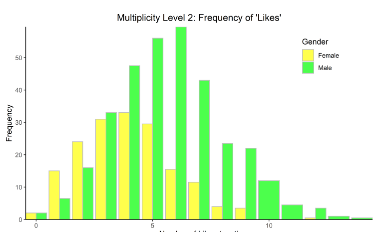


(c) Frequency of 'Matches' for males (red) and females (blue)

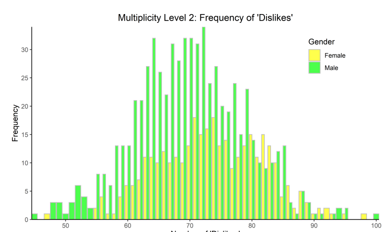


(d) Frequency of 'Messages' for males (red) and females (blue)

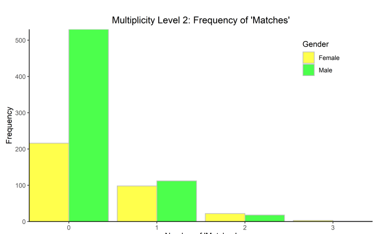
Figure 3: Summary: 'like', 'dislike', 'match', and 'message' distributions for Multiplicity Level 1 models



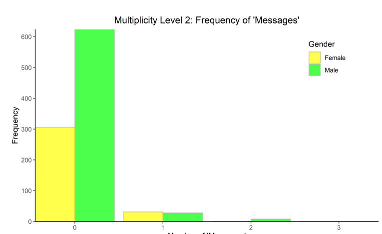
(a) Frequency of 'Likes' for males (green) and females (yellow)



(b) Frequency of 'Dislikes' for males (green) and females (yellow)



(c) Frequency of 'Matches' for males (green) and females (yellow)



(d) Frequency of 'Messages' for males (green) and females (yellow)

Figure 4: Summary: 'like', 'dislike', 'match', and 'message' distributions for Multiplicity Level 2 models

Table 6: M2 Model Run Summary Statistics

Statistic	N	μ	σ	0^{th}	25^{th}	75^{th}	100^{th}
'Like' Count	1,000	5.123	2.386	0	3	7	14
'Dislike' Count	1,000	71.141	8.766	45	65	77	100
'Match' Count	1,000	0.302	0.561	0	0	1	3
'Like Message' Count	1,000	0.083	0.326	0	0	0	3

Table 7: Statistical Comparison M1 vs. M2 using paired t-test

Relation Type	estimate	statistic	p.value	DOF	conf.low	conf.high
'Like'	17.326	107.8	0	999	17.01	17.64
'Dislike'	-15.263	-43.1	2.11e-230	999	-15.957	-14.56
'Match'	0.827	20.4	5.04e-78	999	0.74	0.90
'Message'	0.526	17.0	2.802e-57	999	0.46	0.58

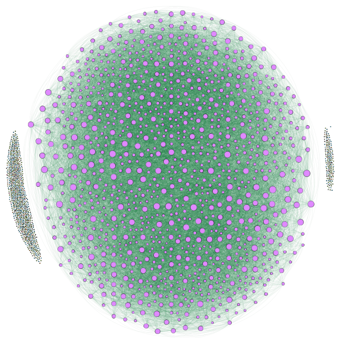
4.3 Network Results and Regression Model

We also issued a direct comparison between the M1 and M2 models resultant 'like' and 'match' networks shown in Figures 5 and 6. As might be expected, since much of the mechanism of the underlying rules was set to follow random behavior, the like networks of both contained little non-random structure (by visual inspection). Since M1 generated more activity than M2, the density of M1 was higher than M2. More interestingly, the match networks for both models did contain non-random structure (figure 5b and 6b). In the M1 model the match network produced a large component and many isolates, suggesting that from simple rules can emerge match networks with structures indicative that small preferences (in age for example) can produce clusters of high matching probability. The M2 match network showed highly dyadic structure (rather than transitive) suggesting that agents matched with one or two other agents but not within larger clusters of agents as in M1. We remind the reader that that the main difference between behaviors in our models is simply that in M2 the user considers more features.

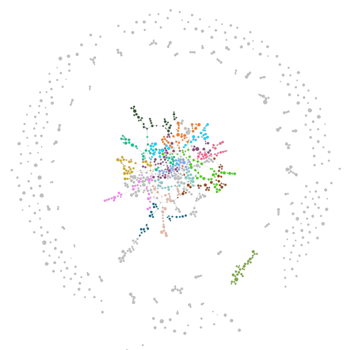
Aside from summary network statistics and visual model inspection, we utilized the network regression technique, Multiple Regression Quadratic Assignment Procedure (MRQAP) (Mantel 1967) to investigate the relationship of our monadic and dyadic covariates to the production of a match. MRQAP is best used on dyadic relations and so in order to utilize it effectively we converted out monadic covariates, such as attractiveness, age, and ethnicity into difference adjacency matrices. We also included the 'like' and 'dislike' networks as independent variables. Our question, simply put, is what are the effects of our covariates on producing a match, given the multiplicity of our two models? Table 8 provides statistically significant models considered over 100 iterations that answer this question.

As might be expected almost all parameter estimates are both small and negative. This is due to obvious factors, primarily, since to ensure a proper baseline of comparison, we set the majority of our randomization tests to follow uniformity–this, by-design–would ensure that parameter estimates are small. If we were to use real world input distributions for frequency of likes, dislikes and others, we may find greater effect sizes.[2] However, the difference in order of magnitude for each parameter estimate is telling on which effects

[2]However, this data is not made available publicly and is considered highly proprietary.

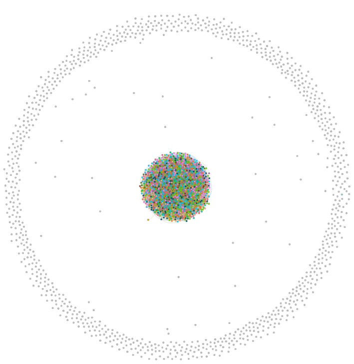


(a) M1 ’Like’ Network

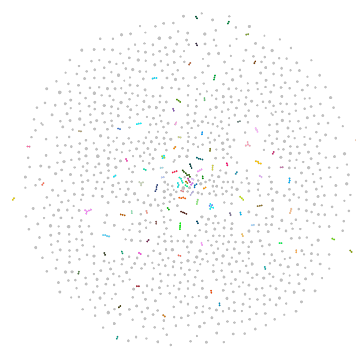


(b) M1 ’Match’ Network

Figure 5: M1 cross-sectional resultant networks for a single run including isolates



(a) M2 ’Like’ Network



(b) M2 ’Match’ Network

Figure 6: M2 cross-sectional resultant networks for a single run including isolates

Table 8: MRQAP Linear Regression Comparison Using Monadic and Dyadic Covariates

	Dependent Network: 'Match' Network	
	Multiplicity Level 1	Multiplicity Level 2
'like' relation	$-1.227880e-03$	$-7.184141e-02$
Pr(<=b)	(0.0)	(0.0)
Pr(>=b)	(1.0)	(1.0)
Pr(>=|b|)	(0.0)	(0.0)
'dislike' relation	$-1.227854e-03$	$-7.183297e-02$
Pr(<=b)	(0.0)	(0.0)
Pr(>=b)	(1.0)	(1.0)
Pr(>=|b|)	(0.0)	(0.0)
attractiveness (difference)	$-6.221186e-05$	$-6.871735e-04$
Pr(<=b)	(0.07)	(0.37)
Pr(>=b)	(0.93)	(0.63)
Pr(>=|b|)	(0.19)	(0.8)
age (difference)	$-5.836387e-06$	$-5.691259e-05$
Pr(<=b)	(0.0)	(0.32)
Pr(>=b)	(1.0)	(0.68)
Pr(>=|b|)	(0.0)	(0.59)
ethnicity (difference)	$1.629981e-04$	$-5.710848e-04$
Pr(<=b)	(0.98)	(0.22)
Pr(>=b)	(0.02)	(0.78)
Pr(>=|b|)	(0.08)	(0.4)
intercept	$1.104061e-03$	$7.226399e-02$
Pr(<=b)	(1.0)	(1.0)
Pr(>=b)	(0.0)	(0.0)
Pr(>=|b|)	(0.0)	(0.0)
Replication(s)	1,00	1,00
R^2	0.0001038	0.0006752
Adjusted R^2	9.88e-05	0.0006702
Residual Std. Error (df = 998994)	0.0336	0.2571
F Statistic (df = 5; 998994)	20.74	135
Note:		*both models are significant at the* p<0.0001 *level*

contribute more to the production of match edges. Secondly, while parameter estimates may be small, their negative sign (with exception to ethnicity) signifies the reduced density of the match network when compared to the like network. Thus, both models presented in table 8 seem to possess parameter estimates that pass out litmus tests. Finally, since MRQAP does not produce traditional p-values, we have included the proportion of simulated networks that meet lower-end one-tailed, upper-end one-tailed, and two-tailed tests. We will proceed in explaining and interpreting each effect.[3]

In both M1 and M2 models, liking and disliking produces statistically significant effects. In fact, for 100 iterations considered by MRQAP, no networks produced an effect as large as the ones presented (one-tailed test) in the table for the 'like' and 'dislike' effects. Interestingly, the magnitude of the effect was an order larger for M1 than it was for M2, suggesting that 'liking' and 'disliking' have a greater influence on matching dynamics in M1 than in M2. Attractiveness difference and age difference followed a similar pattern, with the M1 effect influencing matching more than the M2 effect. However, these effects were 2 and 3 orders of magnitude smaller than the 'like' and 'dislike' effects. That is, in both models 'like' relations are much more closely related to who you ultimately match with than age difference or attractiveness difference. Finally, ethnicity (difference) for both models was a positive effect and 98% of iterations were below this effect - in essence this covariate is significant on the lower-tail test, and since its sign is positive, it reveals that *similarity*, not difference is what plays a role; otherwise known as homophily. That is, the more similarity exists between agents engaged in a like interaction, the more likely a match will occur. Also of interest is the size of this effect which is roughly 3.5 times larger for M2 than for M1.

5 DISCUSSION

Both models were statistically significant, though that evidence in isolation is simply indicative of the dependencies we built into the model in terms of agent-based rules. However, effect sizes considered in light of the number of options available to agents to achieve a match (multiplicity) are instructive. Consider that for the M1 model, the order of importance by effect size is 'like' relation, 'dislike' relation (an analogue to 'like'), ethnic similarity, attractiveness difference, and finally age difference. Given the same effects in the network regression model, the ordering for M2 diverges from this with 'like' and 'dislike', and age difference occupying the same positions with an order (or more) of magnitude difference. In M2, ethnicity similarity and attractiveness difference are of roughly the same magnitude and certainly the same numerical order. This is direct evidence supporting our hypothesis–that holding all else equal, multiplicity allows for more pathways in achieving a 'match' and consequently, agents must consider those new pathways. In so doing, social effects are transformed. In our specific case, the importance of attractiveness increased to the level of the importance of ethnicity in M2, and earlier we showed that even the total number of matches differed in a significant way from both simulations.

6 CONCLUSIONS

In this paper, we have shown, through agent-based simulations how additional features (multiplicity) in online dating applications can cause differences in overall matches and how the underlying social effects can be transformed by those same features. There are many directions we could take this area of inquiry, including calibrating our social effect, models, adding weighted dyadic relationships and most importantly using some reference data set. The latter represents our work's greatest limitation. We hope to be able to pursue these avenues in future papers.

[3]As a reminder, our dependent variable is the match network, and our independent variables are the like network, dislike network, monadic attractiveness converted to dyadic difference in attractiveness, monadic age converted to dyadic age difference, and monadic ethnicity converted to dyadic difference in ethnicity.

REFERENCES

Altman, I. 1973. “Reciprocity of Interpersonal Exchange”. *Journal for the Theory of Social Behaviour* vol. 3 (2), pp. 249–261.

Bruch, E. E., and M. E. J. Newman. 2018, aug. “Aspirational pursuit of mates in online dating markets”. *Science Advances* vol. 4 (8), pp. eaap9815.

Conway, J. R., N. Noë, G. Stulp, and T. V. Pollet. 2015. “Finding your Soulmate: Homosexual and heterosexual age preferences in online dating”. *Personal Relationships* vol. 22 (4), pp. 666,678.

Epstein, J. M. 1999, may. “Agent-based computational models and generative social science”. *Complexity* vol. 4 (5), pp. 41–60.

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* vol. 121 (2), pp. 673–697.

Granovetter, M. 1978. “Threshold Models of Collective Behavior”. *The American Journal of Sociology* vol. 83 (6), pp. 1420–1443.

Hitsch, G. J., A. Hortaçsu, and D. Ariely. 2010. “What makes you click?-mate preferences in online dating”. *Quantitative Marketing and Economics* vol. 8 (4), pp. 393–427.

Koch, Lucy 2019. “Dating App Growth Slows, but Advertisers Shouldn’t Ignore Hopeful Singles”. Retrieved from https://www.emarketer.com/content/dating-app-growth-slows-but-advertisers-shouldnt-ignore-hopeful-single-users.

Krackardt, D. 1987. “QAP partialling as a test of spuriousness”. *Social Networks* vol. 9 (2), pp. 171,186.

Lin, Melissa. “Online Dating Industry: The Business of Love”. Retrieved from https://www.toptal.com/finance/business-model-consultants/online-dating-industry, accessed 2019-10-11.

Mantel, N. 1967. “The Detection of Disease Clustering and a Generalized Regression Approach”. *Cancer Research* vol. 27 (2 Part 1), pp. 209–220.

McGrath, Felim 2015. “What to Know About Tinder in 5 Charts”. Retrieved from https://blog.globalwebindex.com/trends/what-to-know-about-tinder/, accessed 2019-09-25.

McPherson, M., L. Smith-Lovin, and J. M. Cook. 2001. “BIRDS OF A FEATHER: Homophily in Social Networks”. *Annual Review of Sociology*.

Mortensen, D. T. 1982. “The Matching Process as a Noncooperative Bargaining Game”. In *The Economics of Information and Uncertainty*, Volume I, pp. p.173–203. National Bureau of Economic Research.

AUTHOR BIOGRAPHIES

LIAM MCKENNA is a senior at George Mason University approaching his final semester majoring in Computational and Data Science and minoring in Data Analysis. He worked as a data science intern this past summer at Noblis and plans to return to work full time after graduation. His email is lmckenna@gmu.edu.

FATMA ABDULWAHAB is a graduating senior completing a Bachelor of Science in Computational and Data Science with a double minor in Mathematics and Sociology at George Mason University. Her email address is fabdulwa@gmu.edu.

STEVEN HOANG is a junior at George Mason University pursuing a Bachelor of Science in Computational and Data Science with a minor in Biology. He is currently working as a Data Analyst for the Vietnamese Boat People Podcast. His email address is shoang@gmu.edu

COLLIN HENLEY is an upcoming college senior and working on his Bachelor's in Computational and Data Science with a minor in statistics at George Mason University. Most recently, he worked as an Operations Research Analyst at Fort Belvoir. His email address is chenley4@gmu.edu.

JOSEPH A.E. SHAHEEN, PHD is a postdoc fellow and adjunct professor at George Mason University. He supervised this undergraduate project and these exceptionally bright students. His email address is jshaheen@gmu.edu.