The Role of Optimal Distinctiveness and Homophily in Online Dating

Danaja Maldeniya★,1 , Arun Varghese ★,2, Toby Stuart†,3, Daniel Romero★,4
★ University of Michigan, †University of California, Berkeley
1 dmal@umich.edu, 2 arunv@umich.edu, 3 tstuart@berkeley.edu, 4 drom@umich.edu

Introduction

- Online dating site users compete for the attention of potential partners.
- What determines whether a message will receive a reply?
- In this work we explore the effects of two social mechanisms on the likelihood of a response:
  Homophily*:* similarity between a user and their potential partner.
  Distinctiveness*:* similarity between a user and their competition.

Homophily and Distinctiveness

- Homophily suggests that users will be successful if they are perceived as similar to the users they message.
- Theories of optimal distinctiveness suggest that a user will benefit from standing out from the competition.

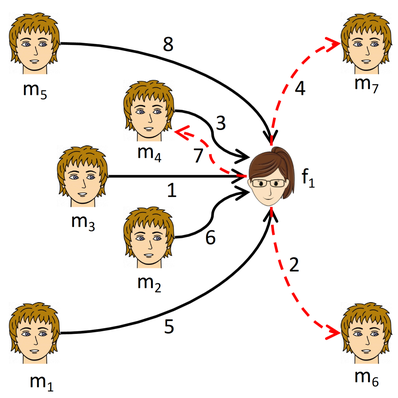
Online Dating Site Data

- We used a dataset of anonymized user activity from a popular U.S dating site.
- The data contained user profiles, messages, clicks and ratings over a 3 month period for approximately 410,000 active users.
- Males initiated 86% of communication and account for 62% of all messages

Types of Competition

Competition sets of a male *u* who initiates contact with a female *v*:

- **Market Level**: other males who also contacted female *v*.
- **Female Choice**: other males who also contacted female v before *u* did.
- **Profile View**: the set of males whose profiles were viewed by *v* prior to being contacted by *u.*



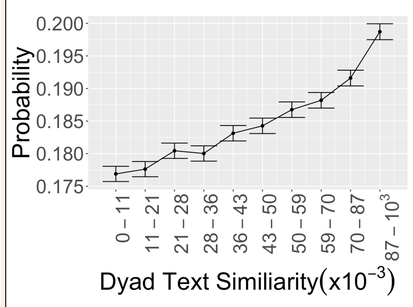
Market level = {m2,m3,m4,m5}
Female Choice ={m3,m4}
Profile view={m6,m7}

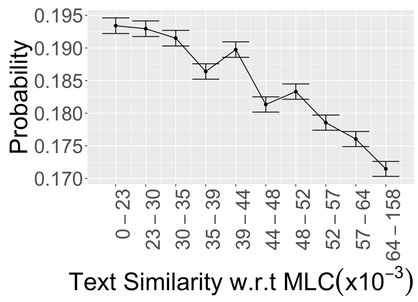
Profile Text Similarity

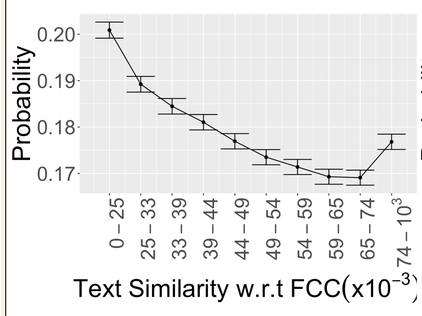
- Member profiles include free text responses to a set of questions.
- We use this text to measure the similarity of self-presentation between users.
- For each user, we generate a TF-IDF vector from the bag-of-words derived by taking all responses and removing stop words.
- For users **A** and **B** with TF-IDF vectors V_A and V_B, similarity between the profile text of A and B is estimated as the cosine similarity of V_A and V_B
- We consider profile text similarity between users in two ways
  i. **Dyad text similarity:** similarity between the male and the female he initiates contact with.
  ii. **Competition text similarity:** average similarity of the male who initiates contact with a female and his competition.

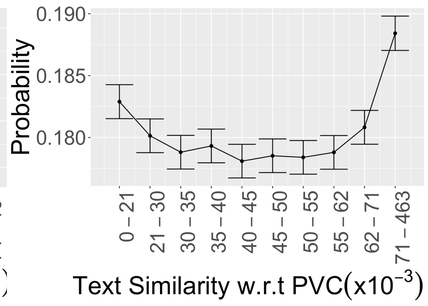
Variation of reply probability with Text Similarity

- As the dyad-similarity increases, so does probability of response from a female. Homophily has an effect.
- As **Market Level** and **Female Choice** similarity increases, the likelihood of a response from a female decreases. Being distinct from those competing for the attention of a female improves chances of success.
- As **Profile View** similarity increases, initially the female response probability decreases but later increases as the similarity increases further. A male has better odds with a female if his profile is either distinct from or very similar to the profiles of males viewed by her.









Modeling reply probability

We use a Linear Probability Model to measure the effect of dyad homophily and competition distinctiveness on the likelihood of a male *u* receiving a reply from female *v*.

Control Variables

i. Dyad similarity/distance measure for various profile properties (12 variables)
ii. Indicator of whether *u* and *v* live in the same city.
iii. The difference in attractiveness (as measured by ratings) between *u* and the average among his competition sets.
iv. Age of *v*.
v. Percentage of message *v* replied to.

Results

- Across all competition sets, a male is more likely to receive a response when his profile text is different from the competition – net of variables that account for effects of homophily.
- A female is more likely to respond when height, body-type, age and attractiveness of the male have "desirable" values.
- All other control variables support homophily.

	MLC	FCC	PVC
TEXT SIMILARITY			
MLC text similarity	-0.0289***	-	-
FCC text similarity	-	-0.0222***	-
PVC text similarity	-	-	-0.0074***
CONTROL VARIABLES			
MLC attractiveness	0.1162***	-	-
FCC attractiveness	-	0.1354***	-
PVC attractiveness	-	-	0.1118***
female message response rate	0.3720***	0.3683***	0.3509***
age of female	-0.0319***	-0.0264***	-0.0294***
dyad age difference	-0.0379***	-0.0406***	-0.0416***
dyad height difference	0.0126***	0.0130***	0.0177***
dyad language similarity	0.0065***	0.0078***	0.0060***
dyad sought relationship type similarity	0.0009	-0.0001	-0.0001
dyad physical distance indicator	0.0070***	0.0074***	0.0098***
dyad preferred age interval overlap	0.0179***	0.0174***	0.0179***
dyad body-type similarity	-0.0028***	-0.0035***	-0.0037***
dyad drugs similarity	0.0041***	0.0032***	0.0027***
dyad ethnicities similarity	0.0192***	0.0185***	0.0227***
dyad smokes similarity	0.0076***	0.0075***	0.0062***
dyad attractiveness difference	0.0640***	0.0339***	0.0599***
dyad text similarity	0.0243***	0.0169***	0.0108***
Intercept	-4.04e-15	1.447e-15	7.635e-15

* p <0.05 ** p <0.01 *** p <0.005

Conclusions

- **Homophily**: our results suggest that male-female text and profile similarity increases chances of a response.
- **Distinctiveness**: we find that being textually different from the competition, improves a male's odds of a response.
- An optimal strategy will need to balance exhibiting common interests with the female while standing out from the competition.