Research Question

What are the differences across racial backgrounds (white vs. non-white) in how heterosexual males on dating sites in the United States describe themselves to potential romantic partners?

Importance

Self-concept and self-representation have long served as grounds of debate in cognitive and positive psychology (Bruning et al, 1999) as well social anthropology (Goffman, 1975). The recent adoption of social networking and its specific affordances have allowed individuals to build out different online 'selves' (Papacharissi, 2010). Arguably, this tendency merely replicates an offline parallel- the need to adapt behaviour- and by extension, identity- depending on the specific needs of a social situation.

One such critical scenario may be that of mate selection, which several economists and sociologists have likened this to 'marriage marketplace'. A number of online dating service providers in developed countries may facilitate expansion of potential mates beyond the limits of even extended offline social networks (Piskorski, 2014). Caciopo et al (2013) assert that as many of one in three marriages in the United States is facilitated through these portals- a finding that has been referenced by several players in the industry as evidence of deep, fulfilling relationships on offer.

Heino et al (2010) argue that these avenues further entrench the economic dimension through an acute, implicit awareness of 'relationshopping'. Herein, potential partners are reduced to entries in a catalogue to be scrolled through. In this sense, they suggest an emerging conscientiousness of 'marketing', with the product being themselves, and the potential mate assuming the role of a buyer (ibid). This perception thus links the private worlds of romantic intimacy with those of mass consumption and broader perceived appeal to the opposite sex. Self-concept is therefore further adapted to meet the requirements of not just the social objectives of the situation per se, but also the demands and constraints of its technological milieu..

The inclusion of race infuses political undertones into this largely sociological and psychological exploration, especially as the United States evolves from a biracial to multiracial society (Lin & Linquist, 2013). These authors leverage data from an unidentified online dating platform and discover preferences of racial homophily across male and female participants on dating platforms. However, the women were found to respond to messages sent by men belonging to a group with equal or higher perceived racial status. Moreover, Fiore et al (2010) find that photos play the largest role in determining overall profile attractiveness- where race can be easily determined visually. The essay components then allow for a self-representation that attempts to- though not necessarily successfully- represent the non-racial dimensions of their attractiveness.

 80% more interactions were initiated by men relative to women (ibid). While the mechanics of which messages get read and responded to are beyond the scope of this study, we may assume that the attractiveness of the men's profile plays some role in helping women determine whether would choose to proceed with the interaction. Moreover, Emory (2017) finds that women in the US typically accord 50% higher importance to their suitors' descriptions relative to men (who place greater weight on the photos). Since these exchanges are then seen to lead to a number of marriages, this research aims to address key questions on how an initial online representation

affects the likelihood of interracial marriages, and by extension, the structure of American society itself.

Unique Contribution

Lin and Lindquist (2013) provide a comprehensive review of the racial dynamics of interactions initiated and sustained across users from different backgrounds. However, the specific factors behind the decision to engage in such an interaction- the attractiveness of profiles- are not thoroughly explored. Fiore et al (2010) attempted to capture these influences through 50 sample profiles on Yahoo Personals. They also delve into the self-description component, which was largely absent in Lin and Lindquist (2013). However, the limited sample size cast doubts on the large-scale generalizability of their findings. Moreover, they do not touch on racial background.

To fill these gaps, this study will leverage publicly open and anonymized user profile data for 59,946 users of OkCupid from the San Francisco region that were extracted with permission (Kim & Escobedo-Land, 2012). The data set includes "typical user information, lifestyle variables, and text responses to 10 essay questions" (ibid). Among these, we will focus on the very first essay on how the users would describe themselves ('About Me'). To the best of my knowledge, this research would be the first time that topic modelling has never been applied to online dating profiles.

The study would face specific limitations. Without data on the eventual online matches and offline meetings from these dates, we do not know which profiles were selected more often/ rated as more attractive. Without follow-up interviews or, it is impossible to measure whether the specific choice of words was aimed at authenticity, or matches with an awareness of 'relationshopping'. The latter could have potentially been inferred by the date of creation of the profile, with longer-term users perhaps gaining more 'experience' and conscientiousness of what achieves their self-designated 'dating success'.

Nonetheless, the essays themselves can provide critical insights into the heterosexual male users' perceptions and beliefs around self-representation. This gains additional importance given when Fiore et al (2010) find that fixed-choice responses do not correlate strongly with overall profile attractiveness. This may suggest that the user's ability to express themselves may be included in the factors for mate selection.

<u>**Methodology**</u>

This paper specifically harnesses the computational techniques of unsupervised learning with text data, specifically topic modelling. Each observation's essay assessed as one 'document'. We will then assess the differences in the essays for self-description across white and non-white heterosexual male users. Two key variables to control for in this analysis would be age and education-levels, which may strongly influence levels of self-awareness and communicative skill.

The study will begin with a cleaning of the data. All HTML tags, such as break lines and hyperlinks would need to be removed. Spelling discrepancies will not be corrected for, as the level of language fluency and correctness may factor into how such profiles are evaluated. The occurrences of such errors will first be checked for across white and non-white male profiles. Whether it will serve as one of the major findings remains to be seen. Another complication is the listing of 218 distinct ethnicities, which will need to suitably grouped and documented.

Next, the study will use some initial textual analysis. We will explore word frequency and collocations. The most common self-descriptors will be checked across racial groups, analogous to the comparison across genders in Kim and Escobedo-Land (2012). We could also consider recurring themes in the other essays, such as interests. However, Fiore et al (2010) found no association between the perceived attractiveness thereof with the length of the text itself, use of self-reference, or the proportion of emotionally-charged words. Taking a cue from these findings, we will not concern ourselves with analytical methods such as sentiment analysis, or part of speech recognition (to detect first-person pronouns). However, this may be explored if time permits, given that Fiore et al (2010) may not have been sufficiently representative of the sample under study here.

The topic modelling itself will evolve with experience and experimentation. Using Latent Dirichlet Allocation models, we will calibrate different values for the number of topics. As rigorously as possible, the study will aim to document which 'topics' seem most coherent and the reason for selection of a certain number. Perplexity will also be harnessed as a quantitative measure in this effort. Finally, packages like PyLDAvis will be employed to visualize this self-representation and any emerging patterns. Collectively, these analyses may offer a small, but highly relevant contribution to the emerging debates on identity, self-representation, race and romantic partner selection.

BIBLIOGRAPHY

Bruning, R. H., Schraw, G. J., & Ronning, R. R. (1999). *Cognitive psychology and instruction*. Prentice-Hall, Inc., One Lake Street, Upper Saddle River, NJ 07458.

'Marital outcomes from on-line meetings' John T. Cacioppo, J, Stephanie Cacioppo, Gian C. Gonzaga, Elizabeth L.Ogburn, Tyler J. VanderWeele. Proceedings of the National Academy of Sciences Jun 2013, 110 (25) 10135-10140; DOI:10.1073/pnas.1222447110

Curington, C. V., Lin, K. H., & Lundquist, J. H. (2015). Positioning multiraciality in cyberspace: Treatment of multiracial daters in an online dating website. American Sociological Review, 80(4), 764-788.

Emory, Leah(2017), Bustle. 'How Many People Who Meet On Dating Apps Get Married? Swiping Isn't Just For Hookups', Retrieved from - https://www.bustle.com/p/how-many-people-who-meet-on-dating-apps-get-married-swiping-isnt-just-for-hookups-44359

Fiore, A. T., Taylor, L. S., Mendelsohn, G. A., & Hearst, M. (2008, April). Assessing attractiveness in online dating profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 797-806). ACM.

Goffman, Erving. (1959). The presentation of self in everyday life. Garden City, N.Y. :Doubleday,

Heino, R. D., Ellison, N. B., & Gibbs, J. L. (2010). Relationshopping: Investigating the market metaphor in online dating. *Journal of Social and Personal relationships*, *27*(4), 427-447.

Lin, K. H., & Lundquist, J. (2013). Mate selection in cyberspace: The intersection of race, gender, and education. *American Journal of Sociology*, *119*(1), 183-215.

Papacharissi, Z. (Ed.). (2018). *A Networked Self and Love*. Routledge.

Piskorski, M. J. (2014). *A social strategy: How we profit from social media*. Princeton University Press.

Kim, Albert Y. & Escobedo-Land , Adriana (2015) OkCupid Data for Introductory Statistics and Data Science Courses, Journal of Statistics Education, 23:2, , DOI: 10.1080/10691898.2015.11889737