Misogynistic Language on Twitter and Sexual Violence

Rachael Fulper[1], Giovanni Luca Ciampaglia, Emilio Ferrara, Yong-Yeol Ahn, Alessandro Flammini, Filippo Menczer
School of Informatics and Computing
Indiana University, Bloomington, USA

Bryce Lewis
Bryn Mawr College
Bryn Mawr, PA

Kehontas Rowe
Mills College
Oakland, CA

ABSTRACT
Studies have demonstrated that social media may offer insights into social behaviors. Here we investigate the potential of social media in predicting criminal behavior, in particular rape and sexual abuse. Traditional approaches for studying sexual violence are effective but laborious, although often limited to small sample sizes and coarse temporal resolutions. Additionally, the sensitive nature of sexual violence and stigmas against victims result in serious under reporting of rape crime statistics. The factors contributing to rape are not fully agreed upon, but research shows that the acceptance of, and willingness to commit rape are highly correlated with sex-role stereotyping, rape myth beliefs, and misogyny. Here we explore whether social media can be used as an indicator of sexual violence in the US, by tracking misogynistic tweets. We compared the number of tweets and rape crime statistics for each state, and found a significant association. Our work paves the way to the design of a 'social sensor' system to detect rape and other abuse by monitoring social streams.

Author Keywords
Social media, human behavior, social media language analysis

ACM Classification Keywords
[Human-centered computing**]: Collaborative and social computing |** Social media**; [**Information systems**]: World Wide Web |** Social networks**; [**Networks**]: Network types |** Social media networks

INTRODUCTION

Online social data can offer insight about real world phenomena and societal behaviors. By mining massive datasets provided by social media services and other databases, previous research has studied positive behaviors such as friendship [16], team production [14], evolution of organizations [4], political discourse [10], knowledge production [5] and the diffusion of trends, fashions and fads [13].

Collective behaviors with a negative connotation have been investigated as well, using a mix of social and conventional data, including criminal activities [11, 22], terrorist attacks [7], and financial fraud [20]. Sexual violence is a major societal issue, but its ramifications in the online world have not been given much attention yet.

In the social sciences, traditional approaches for studying violence, and sexual violence in particular, rely on official criminal records, surveys, interviews, and field research [9]. While these methods can provide in-depth accounts, they are also costly and laborious, and thus often limited to small sample sizes and coarse temporal resolutions. Moreover, the sensitive nature of sexual violence, and the stigma victims are often faced with, result in serious under reporting of sexual crimes. According to the Rape, Abuse and Incest National Network, 60% of the rapes in the US go unreported. Of the rapes reported, only 25% of those cases ever lead to an arrest [18].

The individual and societal factors contributing to rape are not fully agreed upon. Research suggests that a combination of rape-tolerant beliefs and demographics are responsible for the under reporting of rape [3, 12]. Psychological surveys have shown that the acceptance of, and willingness to commit rape are highly correlated with sex-role stereotyping and rape myth beliefs [3, 6]. Misogyny is often portrayed by sexual objectification, discrimination, denigration, and violence against women

[1] Corresponding author: rfulper@umail.iu.edu

[8]. Many factors contributing to a rape-tolerant society are thus also indicators of misogyny, or the hatred of women. Some characteristics and personality traits are known to be reflected in language usage on social media [1, 15, 19].

Here we explore whether social media can be used as a 'societal sensor' of sexual violence, by tracking the occurrence of misogynistic language on Twitter. We analyzed a sample of tweets from January and February in 2012 and rape statistics from the 2012 FBI Uniform Crime Reports, and found a significant association between per capita rates of rape and misogynistic language. This preliminary result suggests that increased misogynistic language is associated with increased instances of sexual violence. Thus language use on social media may provide a useful signal about the occurrence of sexual violence in the real world. Because the users of social media services, and Twitter in particular, are known to react in almost real time to external events, such signal may thus complement in a timely fashion data from traditional survey-based research in monitoring the incidence of sexual violence in the population.

METHODS

The FBI Uniform Crime Reports provides rape statistics in the US at the state level (including Washington DC) [21]. This data does not break down across victim gender, but because the overwhelming majority of rapes (over 90%) are committed against females [18] and prior research has prevalently focused on this type of sexual violence, here we use it to explore the connection with misogynistic language.

Our social media data comes from a 10% sample of the Twitter stream produced during 2012. We manually compiled a list of 90 terms that are commonly used as misogynistic insults. The term adoption analysis revealed that the most common one — 'bitch' — had a very broad meaning due to its frequent colloquial use, therefore we decided to exclude it.

We geo-coded tweets that contain misogynistic language and location (either latitude/longitude or a free-form location string) and mapped them to states, using state boundary data from the US Census Bureau. Our final data set contains roughly 170 million geo-located tweets, of which 1.2 million contain misogynistic language (accounting for 0.68% of the total amount).

Because a state with a larger population will see a larger number of both rape cases and misogynistic tweets, we normalized each variable. We used per capita rape rates and the ratio of misogynistic tweets to the total number of tweets from that state. We used a generalized Gaussian linear model to assess the degree of association between the two variables.

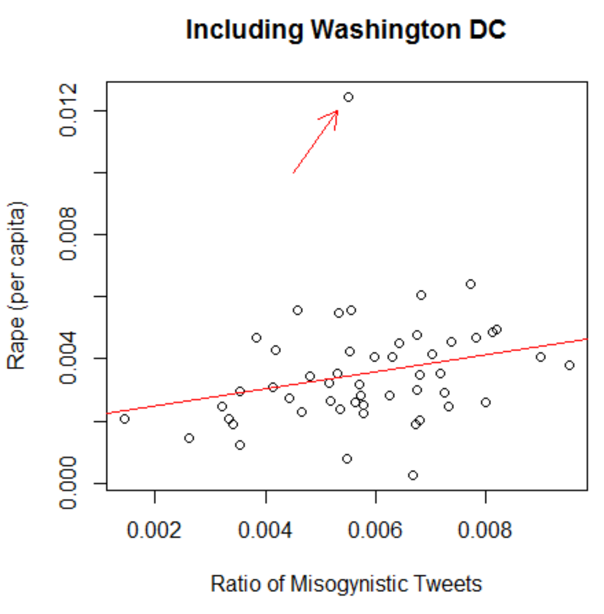

Figure 1: When considering all 50 states and DC, we found a Pearson correlation of *r* = 0.25 and a *p*-value of 0.09. The arrow indicates Washington D.C.

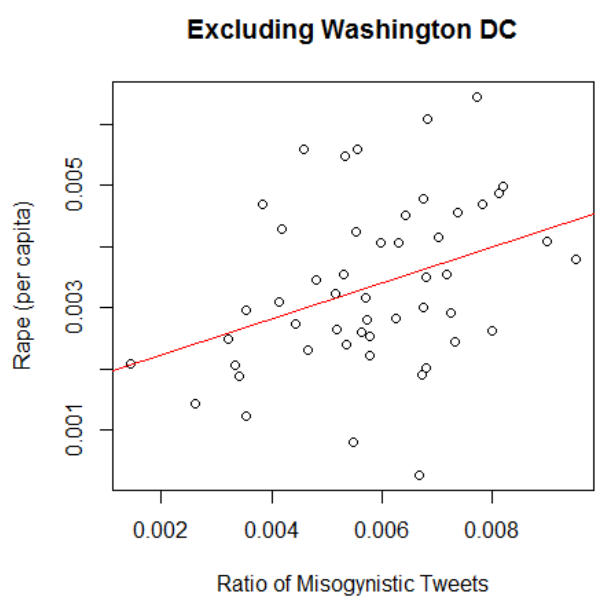

Figure 2: When Washington DC is removed, the Pearson correlation rises to *r* = 0.36 with a *p*-value of 0.01

RESULTS

We find that the yearly per capita rate of rape and misogynistic language are correlated. When considering all 50 states and Washington DC, we found a Pearson correlation of $r = 0.25$ and a p-value of 0.09 (see Fig. 1). However, in the figure we can see that a single outlier, Washington DC, is present. When this outlier is removed, the Pearson correlation become stronger and significant ($r = 0.36$ and p-value=0.01, see Fig. 2).

DISCUSSION

Traditional rape statistics suffer from severe under-reporting due to a number of factors, such as victim blaming, justification, downplaying, and narrow definitions of sexual violence crimes. Taken together, all these factors may prove challenging to overcome, and this contributes hugely to the societal problem of sexual violence. Our preliminary results show that use of misogynistic language on social media is correlated with higher number of rapes per capita at the level of states. Tracking language on social media and other public arenas on the Internet could thus provide an alternative source of information about the level of sexual violence in the population, one that may be less affected by the above biases.

These results are thus promising but further investigation is needed. Our statistical analysis, for instance, showed that the Pearson correlation coefficient improves when the outlier corresponding to Washington DC is removed. Previous research has shown that cities tend to have super-linear rate of activity of crime in population size [2], and this could explain why Washington DC is an outlier, when compared to other states. For future work, we plan to break down our geographic analysis to the city/county level to see if other cities show the same effect.

We believe that with further study and data analysis, a method to identify 'rape-tolerant' cultures from data about language use can be found. Our approach could be further expanded to design a 'social sensor' system to detect and monitor rape and other abuse by means of real-time social media stream analysis. Databases about past criminal events are increasingly mined to predict and track "hotspots" of crime [17], and recently Twitter has been used for a similar purpose [22]. However, this analysis focuses on language directly related to the crime or unlawful behavior of interest, without considering terms more broadly associated with the root social bias.

This work could also be expanded to find insights into other crimes or wide-spread negative social phenomena. Racism and hate crimes have a similar societal link as misogyny and rape. Mentions of drinking and substance abuse might shed light on drug-related crimes.

ACKNOWLEDGMENTS

We are grateful to The Kinsey Institute for Research in Sex, Gender, and Reproduction for feedback. This work was supported in part by NSF grant CCF-1101743.

REFERENCES

1. Agha, A.. Language and social relations (No. 24). Cambridge University Press (2007).
2. Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104, 17 (2007), 7301-7306.
3. Burt, M.R. Cultural myths and supports for rape. *Journal of Personality and Social Psychology*, 38 (1980), 217-230.
4. Burt, R. S. Structural holes and good ideas. *American Journal of Sociology*, 110, 2 (2004), 349-399.
5. Capocci, A., Servedio, V. D., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S., & Caldarelli, G. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E*, 74, 3 (2006).
6. Check, J., Malamuth, N. M. Sex role stereotyping and reactions to depictions of stranger versus acquaintance rape. *Journal of Personality and Social Psychology*, 45 (1983), 344-356.
7. Clauset, A., & Gleditsch, K. S. The developmental dynamics of terrorist organizations. *PloS ONE*, 7, 11 (2012).
8. Code, L. Definition of 'misogyny' *Encyclopedia of Feminist Theories,* (2000), 346.
9. Collins, R. Violence: A micro-sociological theory. Greenwood Publishing Group (2009).
10. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A. Political polarization on twitter. In *Proc*. *ICWSM* (2011).

11. Ferrara, E., De Meo, P., Catanese, S., Fiumara, G. Detecting criminal organizations in mobile phone networks. *Expert Systems with Applications* (2014).

12. Frese, B., Moya, M., Megias, J. L. Social Perception of Rape: How Rape Myth Acceptance Modulates the Influence of Situational Factors. *Journal of Interpersonal Violence*, 19 (2004), 143-161.

13. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A. Information diffusion through blogspace. In *Proc. WWW 2004*. ACM Press (2004), 491-501.

14. Guimerà, R., Uzzi, B., Spiro, J., Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308, 5722 (2005), 697-702.

15. Kosinski, M., Stillwell, D., Graepel, T.. Private traits and attributes are predictable from digital records of human behavior. *PNAS,* 110.15 (2013): 5802-5805.

16. Kumar, R., Novak, J., Tomkins, A. Structure and evolution of online social networks. *Link mining: models, algorithms, and applications.* Springer New York (2010). 337-357.

17. Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., Tita, G. E. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106, 493 (2011).

18. Rape, Abuse & Incest National Network Statistics. https://www.rainn.org/statistics.

19. Schwartz, H. Andrew, et al. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PloS ONE* (2013).

20. Shetty, J., Adibi, J. Discovering important nodes through graph entropy the case of Enron email database. *In Proc. 3rd International Workshop on Link discovery*. ACM Press (2005) 74-81.

21. Uniform Crime Reports from the Federal Bureau of Investigations (2012). http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012

22. Wang, X., Gerber, M. S., Brown, D. E.. Automatic crime prediction using events extracted from twitter posts. *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer Berlin Heidelberg (2012) 231-238.