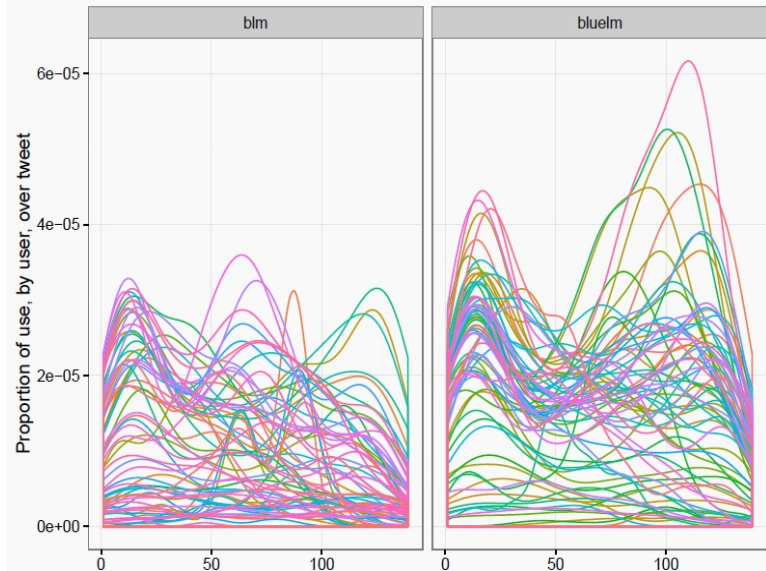## Social Identity and Punctuation Variation
in the #BlueLivesMatter and #BlackLivesMatter Twitter Communities

Background: Twitter is a microblogging platform that allows users to publicly share short written messages (Tweets). Previous work has found that Tweets exhibit much of the same sociolinguistic variation observed in spoken language, including lexical (Jones 2015, Shapp 2015), phonetic (Eisenstein 2013, Tatman 2016) and syntactic (Grieve 2016) variation. Other work examines stylistic variation with respect to features unique to computer-mediated communication, such as nose inclusion in emoticons in Tweets (Schnoebelen 2012). The present work builds on this research by investigating how the pattern of punctuation use in Tweets varies with users' social identity.

#BlackLivesMatter and #BlueLivesMatter: #BlackLivesMatter is a hashtag associated with an activist movement which, among other goals, seeks to stop police killings of African Americans (Garza 2014). #BlueLivesMatter was formed in opposition to #BlackLivesMatter to support police officers, characterizing the latter movement as anti-police. The use of these hashtags indexes directly oppositional social identities, and in our data set, there is no overlap between self-identifying members of these communities.

Data: We used the Twitter browser interface to search for accounts where the term #BlackLivesMatter or #BlueLivesMatter appeared either in the user's self-written 'bio' or display name. From these 188 accounts, up to 50 of the most recent tweets were sampled using Twitter's public API. The final data set consisted of 5170 total tweets, 2031 from #BlackLivesMatter affiliated users & 3139 from #BlueLivesMatter affiliated users.

Results: In order to compare not only the total amount of punctuation used by users in each of these groups, but also the overall pattern of use, the density of all punctuation characters over the length of the Tweet was calculated. This can be seen in the figure to the right, with #BlackLivesMatter on the left and #BlueLivesMatter on the right. Each colored line in the graph represents an individual user, and distance from the x-axis represents the probability that the character at that index of the Tweet is a punctuation mark (where a Tweet is

≤150 characters in length). From this figure, it can be clearly seen that both the amount of punctuation and its placement in Tweets differ between these communities. The #BlueLivesMatter tweets use much more punctuation overall, and are also more likely to use it towards the end of the Tweet—which is only true of a few users in the #BlackLivesMatter community.

Discussion: The robust differences in the use of punctuation between these two communities are of interest to both sociolinguists and computational linguists. For the former, it suggests that variation in punctuation use by different online communities may be a fruitful area for future study. For the latter, it suggests that extra-lexical features of written language such as

punctuation may provide robust features for social classification of users. This has applications in areas such as automatically recommending new accounts for Twitter users to follow and identifying impostors or trolls in online communities by quantifying how members diverge from the community's stylistic norm.

Works Cited

Eisenstein, J. (2013, June). Phonological factors in social media writing. In *Proc. of the Workshop on Language Analysis in Social Media* (Vol. 3, pp. 11-19).

Garza, A. (2014). A herstory of the #blacklivesmatter movement.

Grieve, J. (2016). *Regional Variation in Written American English*. Cambridge University Press.

Jones, T. (2015). Toward a Description of African American Vernacular English Dialect Regions Using "Black Twitter". *American Speech*, *90*(4), 403-440.

Schnoebelen, T. (2012) Do you smile with your nose? Stylistic variation in Twitter emoticons. University of Pennsylvania Working Papers in Linguistics 18(2):14.

Shapp, A. (2015) "Gender Variation in the Pragmatic Uses of Twitter Hashtags." Annual Meeting of the Linguistic Society of America, Portland, Oregon.

Tatman, R. (2016) " 'I'm a spawts guay': Comparing the Use of Sociophonetic Variables in Speech and Twitter" Selected Papers from NWAV 44