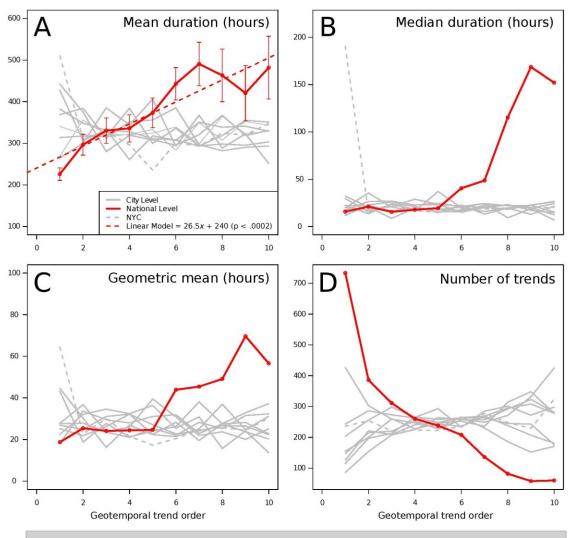
As a digital humanities scholar, my research applies data science tools and methods to the study of rhetoric and writing in digital networks. While concepts like *network*, *trending*, and *virality* have motivated extensive theoretical inquiry into the cultural effects of digital media, we have barely glimpsed the possibilities for exploring the material dimensions and operational functions of these concepts for humanities research. My forthcoming article in *Digital Humanities Quarterly*, titled "Attention Ecology: Trend Circulation and the Virality Threshold," exemplifies the type of research afforded by data science methods. Through a study of trend data mined from Twitter, the article identifies a *virality threshold* for Twitter trends, and argues that such a concept has broader implications for social network research in the digital humanities. For this study we collected data on 17,343 unique trends over 73 days through Twitter's API, and conducted a geotemporal analysis of the data. As the figure below displays, trends that quickly appear at the national level tend to be volatile in nature—quick to rise, and quick to fall. By contrast, topics that increase in attention slowly, growing and spreading to more locations prior to trending at a national level, are much more likely to achieve critical mass:



The lines in the figure above display geotemporal trend sequence. The red line represents the trends that appeared in all other locations in our analysis prior to trending at the national level. The grey lines depict trends that appeared at a national level prior to reaching all 10 locations.

A follow-up study that adds to this research is already underway. This study extends the methods and replicates the results from Eunsong Kim's Model View Culture article: "The Politics of Trending." In order to show the problematic bias of Twitter's trending algorithm, Kim used Topsy's free cloud analytics tool to show the total tweets that trends like #ferguson (a #blacklivesmatter trend) received in comparison to other #hashtags on Twitter. As Kim's analysis shows, #ferguson produced substantially more tweets in 2015 compared to other trends, but trends receiving far less attention were shown to be trending by Twitter (appearing within Twitter's top trend list)—#ferguson was not. However, by the end of 2015, Apple purchased Topsy for \$225 million and soon thereafter shutdown the online service. As a result, the analysis Kim provides in her article cannot be easily replicated or applied to other similar investigations. Using MassMine, we are currently working to collect and archive the massive sample stream of all tweets, as well as collecting and archiving the trending lists Twitter provides through its API (refreshed every 5 minutes). We will conduct an analysis similar to the one Kim produced (but on a much larger scale, showing changes over time), comparing the most frequent terms and phrases in the streaming sample dataset to those which Twitter claims are trending within the network. As recent upheavals at Facebook have shown, human curation also plays a significant role in what social networks show to as trending within their systems, and we hope that our comparisons will shed light on this aspect of trend filtering as well.

My dissertation, Writing Macroscopes: Data Mining Methods for Networked Writing, Digital Rhetoric, and Attention Ecology, draws on computational approaches from the digital humanities to develop macroscopic methods for studying rhetoric and writing within social networks. Whereas microscopic analysis of networked writing focuses on the digital artifacts shared/produced by individual users/writers and how those artifacts may be received by their peers within a specific network, macroscopic research into networked writing relies on the wider lens (provided by data science methods) to consider the massive flow and aggregation of artifacts within and across networks as a whole. Other macroscopic endeavors within the digital humanities have played a crucial role in the development of this project—Franco Moretti's "distance reading" concept, Matthew Jockers' Text Analysis with R for Students of Literature and his Macroanalysis: Digital Methods and Literary History, and Exploring Big Historical Data: The Historian's Macroscope by Shawn Graham, Ian Milligan, and Scott Weingart. However, the issues involved in collecting, curating, archiving, and analyzing data from social networks presents a different set of challenges than those posed by the macroscopic study of literary and historical texts. Some of those challeges include issues of privacy rights, the ethics of the surveillance economy, and the various accessibility issues associated with different forms of social network data. To address these challenges, Writing Macroscopes explores how data mining and social network analysis (1) bring attention to valuable trends and important digital artifacts that have been systematically ignored, (2) inform the invention and discovery of new networks that do not surveil, manipulate, and control the users and artifacts within their networks, and (3) develop pedagogical approaches where students contribute to this work and take an active role in shaping the future of digital networks.

>>>>>>>

While the research described here focuses on those projects concerned with applying data science methods to the study of rhetoric and writing in social networks, I remain interested in broad interdisciplinary applications of data science and digital humanities methods for many various forms of research and technology development.