Uncovering Latent Community Issues Through Social Media

PI: Dr. Leah C. Windsor (Leah.Windsor@memphis.edu)
Co-PI: Dr. Alistair Windsor (awindsor@memphis.edu)

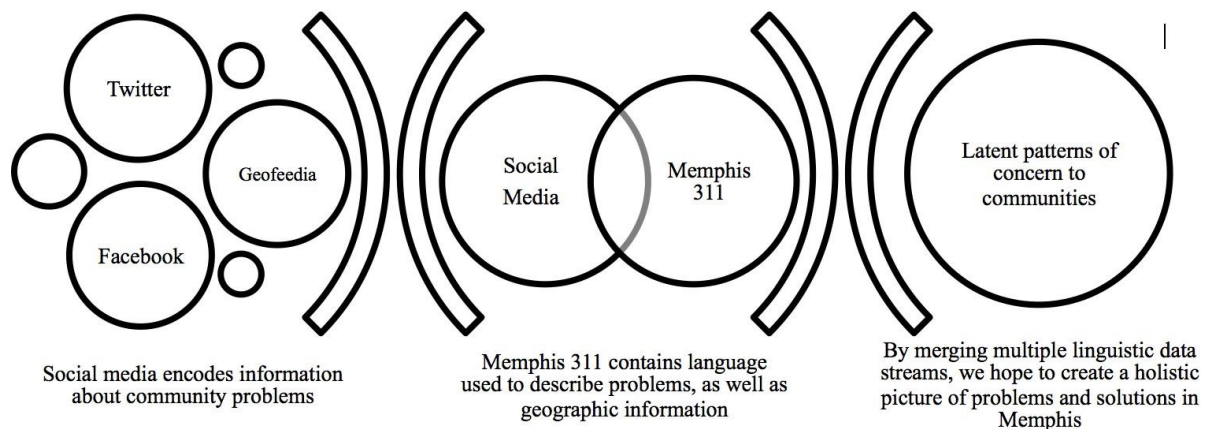## Defining the Issues

We identify two key issues in our proposal. First, latent patterns of problems (blight, litter, crime) go undetected because many people do not report incidents to the police through formal channels. Thus no data is generated for public officials to investigate. As a result resources are not allocated to these problems because of the lack of formal reporting, resulting in citizens feeling that needs are not being met. Citizens are increasingly using social media sites like Twitter to alert friends and neighbors about local problems (Gil de Zúñiga, Jung, & Valenzuela, 2012; Jiang, Bazarova, & Hancock, 2013). However, there is a disconnect between the information they post, and those who can do things about it - policy makers and police. Building better communities means more than just fixing broken windows (Kelling & Coles, 1998). Second, people report community problems through the Memphis 311 system about blight, neglect, and other neighborhood issues that are likely not cross-referenced with the unofficial social media reports, resulting in undetected trends. Thus, by analyzing the content of the Memphis 311 reports and merging the information gleaned from social media and the Memphis 311 system we hope to shed light on latent patterns, increasing the efficiency of city government and community agency responsiveness, thus creating happier and safer communities.

Words encode meaning, and we can learn more about what communities need by evaluating what people are saying in unofficial forums. Computational linguistics can better leverage the information in social media and the Memphis 311 system by uncovering patterns in the content of the language as well as the location. To increase government efficiency and improve citizens' well being, we propose examining social media to detect latent patterns of important community issues. Information about latent problems is encoded in the language that citizens and neighbors use to communicate with each other via social media and through the 311 reporting system. By systematically evaluating the language they use to describe these problems, we hope to reveal latent patterns that will allow government officials to better respond to citizens' issues. By studying unofficial reporting methods via social media, we can create a more holistic picture of problematic trends and behaviors.

Latent Issue Patterns and Social Media

To address these information gaps, our team proposes to probe existing social media outlets like Twitter, as well as the Memphis 311 system, with the hope of eventually building an automated analytical tool that will aggregate and analyze information from these sites to detect latent patterns and problems. By revealing patterns in problematic areas, or emerging patterns in new areas, community resources can be better allocated, increasing government efficiency, and improving citizens' relationship to their neighborhoods and their city officials.

Social media encodes information about community problems | Memphis 311 contains language used to describe problems, as well as geographic information | By merging multiple linguistic data streams, we hope to create a holistic picture of problems and solutions in Memphis

Action Plan

We intend to use computational linguistics methodology and public policy expertise to better understand the needs of Memphis' citizens. We will analyze social media posts to uncover latent patterns that may not be evident in the venues where formal reporting takes place, like the Memphis Police Department and the Memphis 311. To do this, we will scrape posts from social media and Memphis 311 sites to generate an overview with geolocated information about the issues. We will compare the trends related to crime, litter, blight, and government inefficiency with actual crime data collected by the Memphis Police Department. We hope to eventually build a computational tool to automate this process and generate reports about latent problems in the community, which can be utilized by the proper government agencies.

Tools and Computational Methods

To accomplish these goals, we will use the following computational programs:
· Coh-Metrix, a lexical and semantic program that reveals language patterns;
· LIWC, a sentiment analysis tool;
· MALLET, a computer program that uses Latent Dirichlet Allocation to generate sets of topics from within documents.

Data Requests

We are interested in analyzing social media discourse to detect latent trends that go unreported in the community. We hope that the city will be able to grant us access to several of its data sources. We intend to analyze data from the following sources:

- Twitter, a 140-character social media platform, available via public API.
- The 311 database, both from online entries as well as from the call center; • Geofeedia, the social media data collected for the Memphis Police Department;
- Memphis crime data.
·

Project Team

Dr. Leah Windsor will serve as the PI on this pilot initiative. She earned her BS in Linguistics from Georgetown (1998), her Master's in Political Science from The University of Memphis (2005), and her Ph.D. in Political Science from The University of Mississippi (2012). She recently joined

the Institute for Intelligent Systems as a Research Assistant Professor and serves as the PI on a Minerva Initiative grant from the Department of Defense. Her work focuses on the political implications of both public and private discourse related to credible threats and bluffs, security, onset of radical extremism, and opaque environments where preferences and beliefs are difficult to observe.

Dr. Alistair Windsor is an Associate Professor of Mathematics. He has considerable experience analyzing data.

Tristan Nixon will serve as a consultant on the project. He is a staff programmer at the University of Memphis specializing in cloud-scalable data processing pipelines and machine learning.

Nia Dowell is a cognitive psychology doctoral candidate at the Institute for Intelligent Systems at the University of Memphis. Her research broadly focuses on using language and discourse to uncover the dynamics of socially significant, cognitive, and affective processes.

Grayson Cupit is an undergraduate student in the Electrical and Computer Engineering Department and a research assistant at the Institute for Intelligent Systems. His specializations include data scraping, data mining, and web development.

Project Management
Our team will hold biweekly meetings in the Institute for Intelligent Systems in the FedEx Institute of Technology to discuss the project's goals, updates, findings, and challenges. We will generate a final report that summarizes our findings and advances next steps for generating an automated tool to analyze community-generated data.

Works Cited

Gil de Zúñiga, H., Jung, N., & Valenzuela, S. (2012). Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication*, *17*(3), 319–336. https://doi.org/10.1111/j.10836101.2012.01574.x

Jiang, L. C., Bazarova, N. N., & Hancock, J. T. (2013). From Perception to Behavior: Disclosure Reciprocity and the Intensification of Intimacy in Computer-Mediated Communication. *Communication Research*, *40*(1), 125–143. https://doi.org/10.1177/0093650211405313

Kelling, G. L., & Coles, C. M. (1998). *Fixing Broken Windows: Restoring Order And Reducing Crime In Our Communities* (unknown edition). New York: Free Press.