

## Project Summary

The massive quantities of textual communications generated within most organizations constitutes a largely untapped source for insightful, real-time organizational analytics. From understanding external demands placed on organizations to summarizing the pressing intra-organizational players and issues, most salient developments are documented in digitized text. The subject-matter content and socio-organizational context recorded in an organization’s textual record can be leveraged to understand and improve an organization’s performance. The basis of this project lies in two recent developments. First, research shows that the patterns and structure of communication, formalized as communication networks, are extremely important to effective organizational and individual problem-solving. Second, many organizations, particularly governments, have developed open textual input platforms in order to improve responsiveness to user (e.g., citizen, customer) needs. This project builds an analytical bridge between intra-organizational communication networks and streams of external input. Specifically, we will develop methods to parse and summarize the contents of (1) external input streams (2) intra-organizational communication networks in the same topic-space, and understand the relationships between the external and internal domains.

**Methodology:** We propose to study the ways in which government officials’ communications with those outside of government are related to intra-governmental communications and government outputs. We will use US local (i.e., county and city) government email archives acquired via public records requests and online data collection. We will design computational tools related to statistical topic modeling and network analysis that (1) identify topic-specific internal-external communication networks, (2) identify topic-specific internal communication networks and (3) learn the relationships between internal-external communications, intra-organizational communication networks and the contents of public policies. The methods we develop will track the migration of topics to, within and from government. As such, we will characterize the democratic process at a fine-grained, content and context specific level. The data we collect will permit extensive validation and innovative application of the algorithms developed. We will relate the core email data with additional publicly available data on county governments, including regulations/legislation and minutes from county legislatures. The proposed research will be conducted by an interdisciplinary team that brings expertise in computational (Wallach) and social scientific (Desmarais) fields.

**Intellectual Merit:** This project will offer important contributions to both computational and social sciences. In terms of computational approaches, we will enhance methods for the statistical analysis of text and network data. In particular, we will expand upon extant methods of textual network analysis in developing ways to learn (1) the topics that cut across network domains and (2) functions that characterize domain transfer of topics. On the social science side, the methods we develop and data we collect will advance organizations’ ability to connect streams of external input to their internal operations. Also, more directly, we will offer an unprecedented fine-grained assessment of government responsiveness at the local level in the US.

**Broader Impact:** This project will provide essential tools for organizations in providing timely and coherent responses to the demands of external constituencies. This holds potential to, e.g., improve the efficiency with which local governments manage public health needs, address environmental risks and establish revenue and spending policies. The contributions will be cross-disciplinary and will contribute to the broader scientific community. Also, we will provide an enormous data archive of government communication data to be tapped by other researchers.