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 content and 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, recent 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 government entities, 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 the contents of (1) input streams from external sources and (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 Florida and North Carolina county 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 tcomputational (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 eternal 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.

Organizational Responsiveness to External Demands: A Modeling Approach based on Statistical Text and Network Analysis

1 Introduction

Nearly every organization strives to respond in a timely and accurate manner to the needs and demands of some external constituency. Firms respond to customers, governments respond to citizens and educational institutions respond to students. The rapid advancement in communications technology over the last two decades has forever transformed the nature, volume and sources of input and feedback available to organizations. Also, electronic communications have drastically improved the ability of organizations to document and communicate their internal developments. These complimentary developments have had a transformative impact on governance - moving to what **CITE** call 'we government'. Most elected officials can be directly contacted electronically through simple internet pools. Citizens can advertise and sign petitions on the web and attend internet 'town meetings' with their representatives. Regarding the internal activities of government; citizens can access electronic communications of their officials through public records requests, access meeting minutes on the web and, e.g., watch the floor activities of the US House of Representatives on HouseLive.gov.

In this project we will develop and apply methods for identifying the cycle of input, response and feedback that leaves its fingerprint in the electronic communications record. We will focus on the nexus between government organizations and their constituents, but the methods we develop will be portable to other types of organizations. Government responsiveness to citizen input offers an ideal venue within which to model the relationship between streams of textual records. First, in democratic societies there is a common expectation that the government will respond to public demands. Second, most of the input modes on which we will focused were designed precisely for the objective of providing input to which public officials could respond. Third, and perhaps of greatest practical importance, due to the scope of freedom of information laws in the US, we as researchers can access the public input and internal communications data associated with a multitude of government organizations.

We frame this project by associating different phases in the cycle of governance with four different types of textual streams - public input (e.g., emails from citizens to government officials, informal internal communications (e.g., emails among officials), formal deliberations (e.g., legislative meeting minutes) and policy outputs (e.g., regulations, laws). We seek to understand these textual themes through the lens of statistical topic models **CITES**. We will develop and apply models that permit the identification of the ways in which topics rise and fall within domains and, crucially, are related across domains. The result will be an analytical approach that permits an organization to distill and investigate the dynamics of input, responsiveness and feedback through a common framework of statistical text analysis. The methods we develop will offer answers regarding several pertinent questions about organizational management of outside input, e.g., is organizational attention to a topic proportional to its attention in outside input, how does an organization adapt to the rise of issues that are novel relative to its current foci, is responsiveness timely?

Topic models infer discrete topics from a corpus of documents. A topic is simply a relative frequency distribution of words and each document is probabilistically associated with each topic identified **CITES**. Statistical topic models provide a dually qualitative and quantitative inferential summary of textual corpora. Qualitative in that the textual content of a corpus is maintained and

words themselves form the basis of the quantitative analysis. Dynamic topic models provide an excellent framework within which to understand input to, output from and feedback to organizations that document their activities at various stages in a textual format. Since the seminal work on statistical topic models **CITE**, the basic framework has been extended and adapted to focus on several aspects of textual corpora **CITE**, including author-specific attributes of text **CITE**, dyadic (i.e., author-recipient) aspects of messages **CITE**, dynamics, the underlying communication network, and joint text-metadata models of documents **CITE**. In the current project, we will undertake an ambitious set of extensions that integrate several of these extensions - jointly modeling separate streams of text that influence each other, are informed by rich meta-data, incorporate the underlying communication network, and characterize the over-time aspect of the text streams.

Figure 1 Illustrates the cycle of organizational responsiveness that we intend to model through the guise of co-evolving textual streams. Considering the case of governance, substantial research exists that focuses on parts of this cycle. For example, a large body of research exists that documents recent developments in tools for citizens to provide precise, timely and voluminous input to government officials **CITES**. There is also a large body of research focusing on legislative adaptation to broad ideological trends among constituents **CITES**. And, yet another literature that addresses the processes by which topics rise from informal awareness among officials and outside parties to the legislative agenda **CITES**.. However, due to the historical inaccessibility of timely and common data modes related to each component of the governance cycle, little research has endeavored to connect all of the dots. We will provide such a complete picture, leveraging the common, available, and timely mode of text streams.

2 Motivating Example

A couple paragraphs on an external-input system.

3 Problem Statement

Here we distill problem to abstract structure.

4 What has been done so far

Here we describe the pilot work, publication and data collection, as well as related research.

5 Proposed New Work

Here we describe the methods we intend to develop, datasets we intend to create and applications we intend to develop.

6 Timeline and Division of Labor

Here we give a clear breakdown of how the work will be divided and what will be completed.

7 Broader Impact

Here we discuss broader scientific impact, training and education, as well as potential non-academic impacts.

8 Results From Prior NSF Support

Data Management Plan

A. Project Information

B. General Data Management Plan Information

C. Policies

D. Legal Guidelines and Requirements

E. Access, Sharing and Re-use of Data

F. Data Standards and Capture

G. Security, Storage, Management and Back-Up of Data

H. Preservation, Review and Long-Term Management of Data

Biographical Sketch: Bruce A. Desmarais

(a) Professional Preparation:

– Eastern Connecticut State University, Economics and Public Policy, B.A. (2002)

– University of North Carolina at Chapel Hill, Political Science, M.A. (2008)

– University of North Carolina at Chapel Hill, Political Science, Ph.D. (2010)

(b) Appointments:

– *University of Massachusetts Amherst* Assistant Professor, 2010 - Present

(c) Publications

(i) Publications Directly Related to the Proposed Project

– Cranmer, Skyler J. and Bruce A. Desmarais. 2011. "Inferential Network Analysis with Exponential Random Graph Models." *Political Analysis.* 19(1): 66-86.

– Desmarais, Bruce A. and Skyler J. Cranmer. 2012. "Statistical Inference for Valued-Edge Networks: The Generalized Exponential Random Graph Model" *PLoS-ONE.* 7(1):e30136.

– Desmarais, Bruce A. and Skyler J. Cranmer. 2012. "Statistical Mechanics of Networks: Estimation and Uncertainty." *Physica A* 391(4): 1865-1876.

– Desmarais, Bruce A. and Skyler J. Cranmer. 2012. "Micro-Level Interpretation of Exponential Random Graph Models with Application to Estuary Networks" *Policy Studies Journal.* 40(3): 402-434.

– Cranmer, Skyler J., Tobias Heinrich, and Bruce A. Desmarais. Accepted 2012. "Reciprocity and the Structural Determinants of the International Sanctions Network." *Social Networks.*

(iii) Other Significant Publications

– Desmarais, Bruce A. 2012. "Lessons in Disguise: Multivariate Predictive Mistakes in Collective Choice Models." *Public Choice.* 151(3–4): 719-737..

– Cranmer, Skyler J., Bruce A. Desmarais, and Elizabeth J. Menninga. 2012. "Complex Dependencies in the Alliance Network" *Conflict Management and Peace Science* 23(3).

– Cranmer, Skyler J., Bruce A. Desmarais, and Justin H. Kirkland. 2012. "Towards a Network Theory of Alliance Formation" *International Interactions.* 38(3): 295-324.

– Harden, Jeffrey J. and Bruce A. Desmarais. 2011. "Linear Models with Outliers: Choosing Between Conditional Mean and Conditional Median Methods" *State Politics and Policy Quarterly.* 11(4): 371-389.

– Desmarais, Bruce A. and Jeffrey J. Harden. 2012. "Comparing Partial Likelihood and Robust Estimation Methods for the Cox Regression Model" *Political Analysis.* 20(1): 113-135.

(d) Synergistic Activities

    – Participated in the founding and administration of the Triangle Political Methodology Group (2009–2010) – `http://www.unc.edu/depts/polisci/methods/`.

    – Co-organized an interdisciplinary speaker series in Computational Social Science at UMass Amherst (2010–Present) – `http://cssi.umass.edu/seminars.html`.

    – Editorial board member, *State Politics & Policy Quarterly*, (2011–Present).

    – Member of the fellowship committee for the 2012 Political Networks Conference.

(e) Collaborators and other Affiliations

  (i) Collaborators

    – Skyler Cranmer, University of North Carolina at Chapel Hill
    – Jeffrey J. Harden, University of North Carolina at Chapel Hill
    – Hanna Wallach, University of Massachusetts Amherst
    – Brian Schaffner, University of Massachusetts Amherst
    – Vincent Moscardelli, University of Connecticut
    – Tobias Heinrich, Rice University
    – Allison Freeman, Center for Community Capital (UNC Chapel Hill)
    – Elizabeth Menninga, University of North Carolina Chapel Hill
    – Justin Kirkland, University of North Carolina, Chapel Hill
    – Rachel Shorey, University of Massachusetts Amherst
    – Stuart Benjamin, Duke University
    – Peter Krafft, University of Massachusetts Amherst

  (ii) Graduate and Post-Doctoral Advisors

    – Thomas Carsey, University of North Carolina Chapel Hill
    – Skyler Cranmer, University of North Carolina Chapel Hill
    – James Stimson, University of North Carolina Chapel Hill
    – Kevin McGuire, University of North Carolina Chapel Hill
    – Isaac Unah, University of North Carolina Chapel Hill

  (ii) Thesis Advisor

    – Rachel Shorey, UMass Amherst Computer Science M.S. Student
    – Peter Krafft, UMass Amherst Computer Science M.S. Student
    – James Aaron, UMass Amherst Political Science Ph.D. Student
    – Michael Kowal, UMass Amherst Political Science Ph.D. Student

  Total number of graduate students advised: 4

Budget Justification

Senior Personnel

Other Personnel

Fringe Benefits

Travel

Other Direct Costs

Indirect Costs

Facilities, Equipment and Other Resources

Laboratory

Clinical

Computing

Office

Major Equipment

Other Resources