Project Summary
The electronic exchange of text (e.g., e-mail, text messaging, micro-blogging) now constitutes the cornerstone of the communication process in most organizations. Also, many organizations archive their communications for record-keeping purposes. Substantial recent research shows how important the patterns and structure of communication, formalized as communication networks, are to effective organizational and individual problem-solving. Domains studied include emergency responses to natural disasters, the dissemination of preventive health behaviors, national defense and problem-solving in behavioral experiments. Archives of textual communications offer substantial promise for measuring, diagnosing and optimizing intra-organizational communication networks. However, several challenges stand in the way of realizing this promise. Textual archives constitute complex, large and multi-scale data sources stored in varied and non-standard formats. On the machine learning end, effectively leveraging this data will require scalable algorithms that effectively organize and exploit the content, contextual, temporal and threading information associated with textual communications. From the social scientific perspective, effective methods will offer results that are insightful and interpretable to scholars of social organizations and practitioners alike.

Methodology: We propose to study the ways in which organizations' email corpora can be analyzed to improve intra-organizational communication networks. We will use North Carolina and Florida county government email archives acquired via public records requests and online data collection. We will design computational tools that incorporate (1) content, (2) the communication ties and (3) additional metadata such as temporal and threading information. The methods that we develop will be capable of insightful visualizations that permit the intuitive exploration of the content and context of intra-organizational communication networks, and will be anchored in realistically complex probabilistic models of real-world communication processes. The data we collect will permit extensive validation and innovative application of the algorithms we develop. 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 the computational (Wallach) and social scientific (Desmarais) aspects of textual communication network analysis.

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, including (a) the extension of a multinetwork spatial embedding model to leverage multiple node sets and temporal information, (b) develop scalable inference to capitalize on massive communication archives, and (c) model content-flow dynamics within and across textual modes. On the social science side, the methods we develop and data we collect will advance our understanding of government organizations. By precisely measuring dozens of government communication networks, we will offer an unprecedented assessment of the common structures that characterize intra-organizational communication patterns. Additionally, by integrating (1) communications in public fora, (2) intra-governmental communications, and (3) legislative activities, we will offer a novel micro-level characterization of the local government policymaking cycle.

Broader Impact: This project will provide essential tools for organizations maximize benefits of their internal communication patterns. This holds potential to, e.g., improve organizations' management of natural disasters, assist officials in coordinating to maintain national security, and aid in designing firms to usher in technological innovation.

Clear Connections: Modeling the Movement and Migration of Topics within and Across Communication Networks

1 Introduction

2 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