Content-Conditioned Hierarchical Latent Space Models for Textual Communication Networks

 $Matthew \ J. \ Denny^1 - {\tt mdenny@polsci.umass.edu}$

Collaborators: James ben Aaron¹, Bruce Desmarais¹, and Hanna Wallach^{1,2}

¹University of Massachusetts Amherst ²Microsoft Research NYC

June 8, 2015

Work supported by NSF grant CISE-1320219 (Wallach and Desmarais, PIs)



うつん 川 エー・エー・ エー・シック

Communication Network Structure Matters



(ロ) (目) (目) (目) (目) (日) (日)

Organizational Bias



Content Matters



▲□▶ ▲圖▶ ▲≣▶ ▲≣▶ = 差 = のへで

${\rm Different}\ {\rm Content}\ \longrightarrow\ {\rm Different}\ {\rm Structure}$



Study Objective

Simultaneously model communication content and network structure.

・ロト ・ 日 ・ モ ト ・ モ ・ うへぐ

Latent Dirichlet Allocation

▲□▶ ▲圖▶ ▲臣▶ ▲臣▶ ―臣 … 釣べで

Content: Statistical Topic Modeling

support position

fill

budget funds money budgeted including cost salary amount revenues debt

services care child information system community nurse completed provided pregnancy

Probability

staff desk service customer begin duties vacancy

▲□▶ ▲圖▶ ▲臣▶ ▲臣▶ 三臣 - のへで



・ロト ・母 ・ ・ キャ ・ キャ ・ ちゃくらう

Latent Space Models

◆ロト ◆御ト ◆注ト ◆注ト 注目 のへで





Estimate x and y Coordinate for each node











Our Approach

Partition the communication network by content area.

Data: Network of Interpersonal Messages (e.g., email)



◆□▶ ◆□▶ ◆三▶ ◆三▶ ○ 三 ○ ○ ○

Use content and network structure to classify edges and locate nodes in two-dimensional space



Cluster topics based on shared structure.

Constitutes content specific network analysis



 $\mathcal{O} \mathcal{Q} \mathcal{O}$

æ

Generative Model



Latent Space Model



Latent Dirichlet Allocation

Data

◆□▶ ◆□▶ ◆目▶ ◆目▶ 目 のへで

North Carolina County Government Email Data



- Department manager emails for 3 month period in 2012/2013
- ► FOIA request field experiment 17 counties provided data.

North Carolina County Government Email Data

County	Mgrs.	Female	Internal	Non-Spam
Alexander	21	9	907	11,924
Chowan	23	11	2,027	11,737
Columbus	24	10	920	12,707
:	:	:	:	÷
Totals:	362	139	17,863	117,154

 \blacktriangleright 442,429 raw eamils, 17 counties, 3 months

 Heavily spam-filtered, cleaned, soon publicly available

- ロ ト - 4 回 ト - 4 □

Results

・ロト ・ 日 ・ ・ ヨ ・ ・ 日 ・ うへぐ

Application: Performance \leftrightarrow Structure



Example: Dare County Email Network

- ▶ September November 2012
- Hurricane Sandy passes NC October 26-27th



Communication Content

monday, wx, winds, tonight, east, flooding, night, effect, areas, highway,

county, dare, nc, director, box, po, elizabeth, manteo, fax, phone,

sandy, winds, track, wx, storm, mph, forecast, north, saturday, shore,

fw, storm, point, power, coastal, sandy, weekend, fwd, updated, potential,

◆□▶ ◆□▶ ◆□▶ ◆□▶ → □ ● ● ● ●

wx, updates, winds, sandy, force, track, duration, view, storm, current,

system, wx, coast, forecast, east, track, threat, surf, ts, atlantic,

Structure of Communication



Mixing Parameter Estimates



 $\mathcal{O} \land \mathcal{O}$

Research Agenda

- ▶ R package in development
- ▶ Cross organizational email corpus.
- ▶ Cross org. and temporal extensions.
- Contact: mdenny@polsci.umass.edu
 www.mjdenny.com
 @MatthewJDenny

(日) (日) (日) (日) (日) (日) (日)