Reading between the Emails: Gendered Patterns of Communication in Local Government

Matthew J. Denny

Penn State University — mdenny@psu.edu
www.mjdenny.com
@MatthewJDenny

Collaborators: James ben Aaron, Hanna Wallach, Bruce A. Desmarais

Supported by NSF Grants: DGE-1144860, and CISE-1320219

Research Question

How does the relationship between gender and the pattern of communication vary across organizational domains?

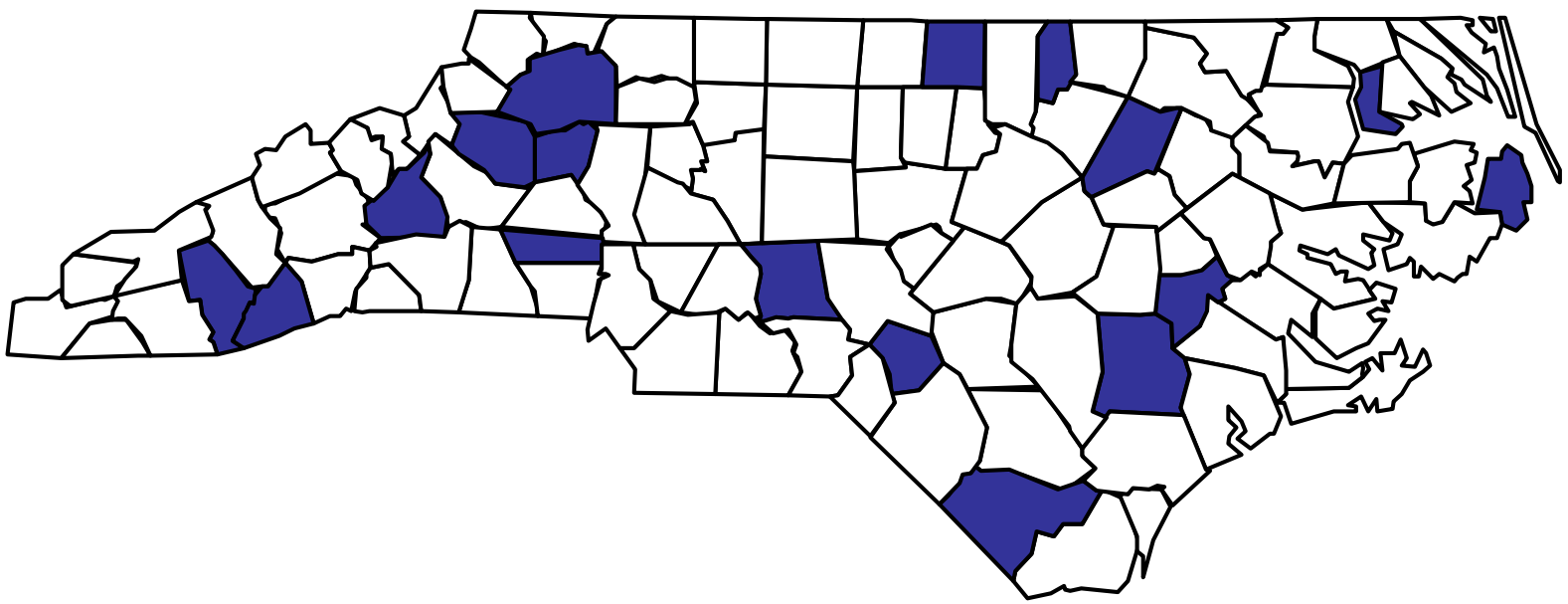
Previous Research

- Administrative Data
  - Salary, position, education, etc.
- Self Reports
  - Surveys
  - Interviews

Public Domain Email Data

County Government Email Data

- Public records requests.
- Department manager email data from 17 North Carolina Counties.
- 500,000 emails, 17,000 between managers.



	Manager Gender		
County	Male	Female	# Emails
Alexander	12	9	907
Caldwell	12	8	121
Chowan	12	11	2,027
Columbus	14	10	920
Dare	15	12	2,247
⋮	⋮	⋮	⋮
Transylvania	16	4	1,857
Vance	10	8	185
Wilkes	15	2	303
Total	223	139	17,863

Department Affiliation as Domain

Individual Department Gender Breakdown

	Emergency Manager	HR	Finance	IT	Health	Plan/Dev Util/Waste	Tax	Parks/Rec	Soc_Serv	Transport	Info	Misc	Inspections
Male	29 15	3	5	11	6	17 15	11	9	8	8	2	5	13
Female	3 2	12	12	2	11	6 2	7	5	10	1	6	2	3
Total	32 17	15	17	13	17	23 17	18	14	18	9	8	7	16

Male dominated | Female dominated

Does Gender Composition Matter?

Dasgupta et al. (2015). "Female peers in small work groups enhance womens motivation, verbal participation, and career aspirations in engineering". *PNAS*

Are women more active in female dominated domains?

Department-Dyad Regression Analysis

- Model department interaction.
- Base Model: Single Intercept
- Gender Model: Separate intercepts
  - MM, MF, FM, FF
- Likelihood-Ratio Test.

Female-Centric Dyads

FF > FM > MF > MM
HR & Health
HR & Information Technology
HR & County Manager
Planning & HR
Register of Deeds & HR
Parks and Recreation & HR
Finance & HR
Finance & Parks and Recreation
Social Services & HR
Solid Waste and Recycling & HR
Tax Administrator & HR

Email Content as Domain

A Generative Model for Email Data

Topic 1: finance, budget, year, pay, fiscal

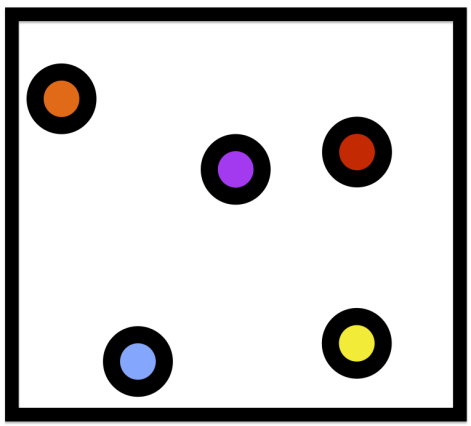
Topic 2: shots, flu, health, senior, hospital

Topic 3: police, judge, jail, firearm, testify

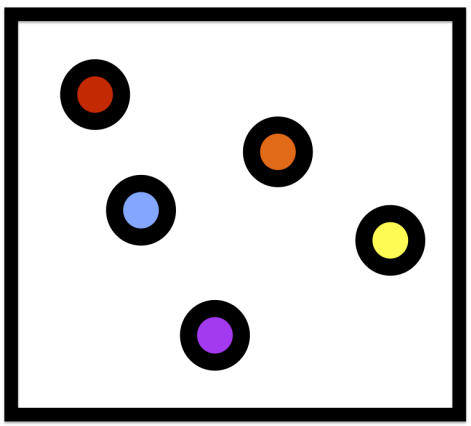
Topic 4: meeting, office, schedule, afternoon

} **Topics**

Topic 1, Topic 2

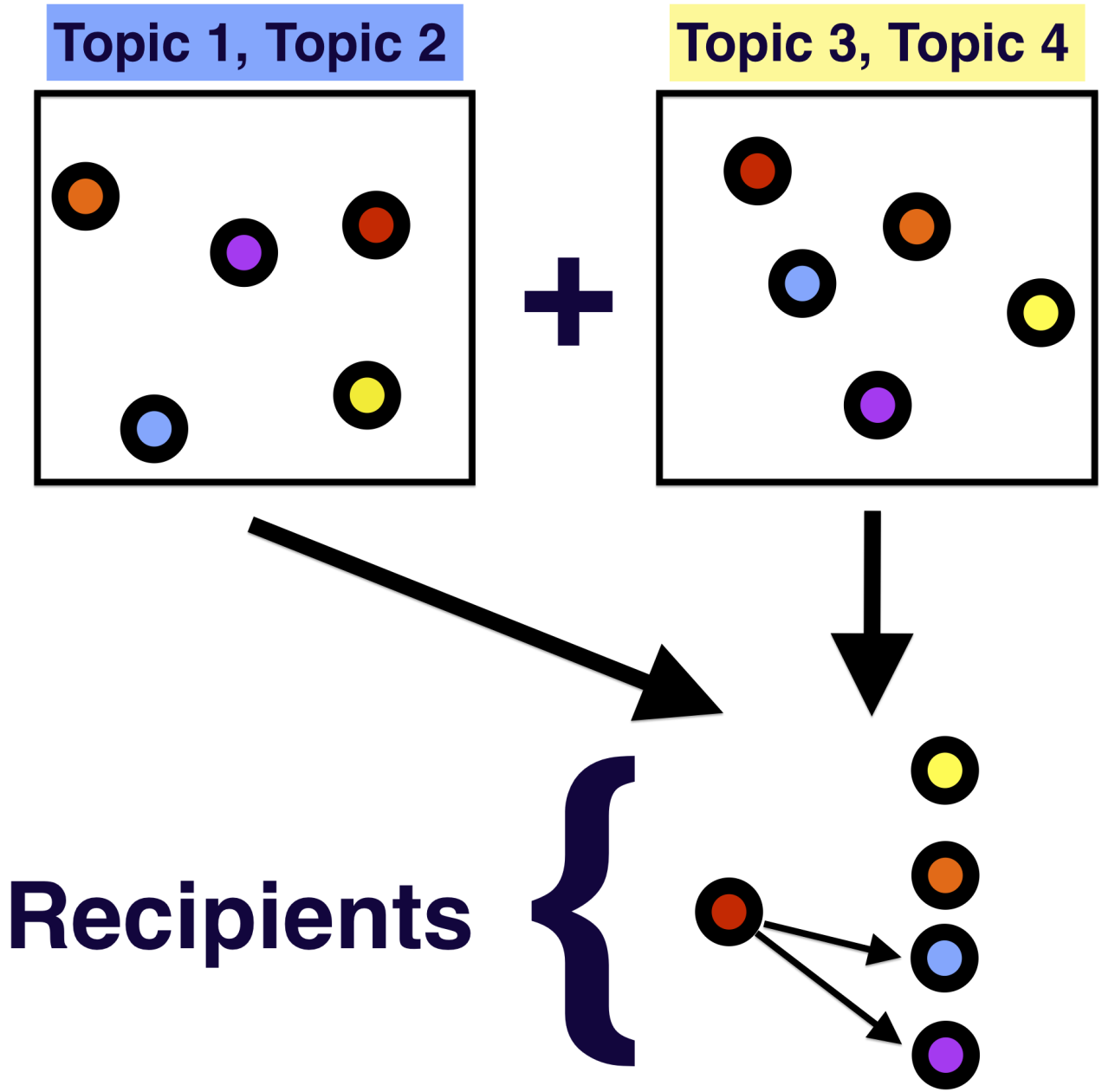


Topic 3, Topic 4



} **Interaction Patterns**

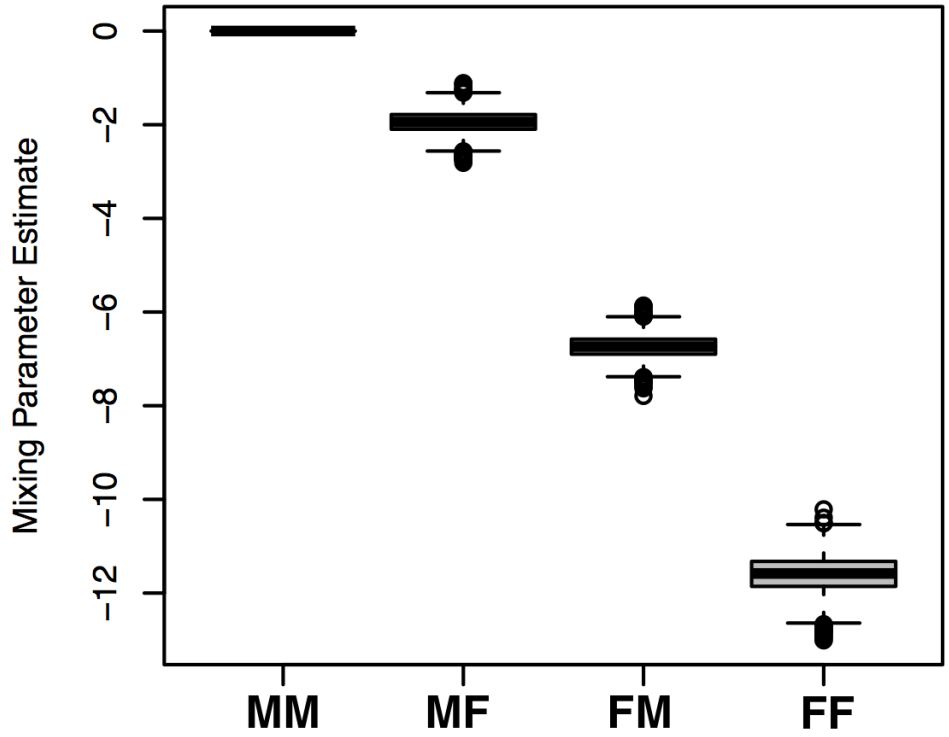
A Generative Model for Email Data



Specification and Methods

- Bayesian inference via MCMC.
- Separate model for each county.
- Include gender mixing covariates.
- 40 topics, 4 interaction patterns.

Example Model Output



Coding	Topic top words
Sandy	will, track, winds, system, forecast
Sandy	storm, sandy, high, coastal, tides, night
Harbor	status, update, boat, today, weather
Planning	board, meeting, planning, seafood, hearing
Storm	box, planning, director, permit, building

Female-Centric Topics of Communication

Coding	FF > FM > MF > MM
Finance	order, time, good, april, attached, requests
Finance	budget, phone, finance, media, ext, department
Health	meeting, going, fyi, tricaster, health, project
Finance	meeting, box, fax, finance, attached, resolution
Finance	equity, fax, debt, refunding, time, finance, call
Finance	debt, box, fax, finance, policies, contract, audit
Finance	learn, leader, director, washington, dream
Finance	fax, ext, phone, finance, director, street
Health	public, health, email, contact, disclosure
Finance	good, time, increase, call, pay, office, today
Budget	manager, street, main, fax, office, east, budget
Budget	fund, budget, balance, year, funds, pay, original

Summary

- Micro-level behavioral data.
- Are women more active in female dominated domains?
- Different operationalizations of domain.
- Our results suggest this might be the case.

Generative Process Pseudocode: Global Variables

for $t = 1$ to T **do**
  draw $\boldsymbol{\phi}^{(t)} \sim \text{Dir}(\beta, \boldsymbol{n})$
end for
for $c = 1$ to C **do**
  draw $b^{(c)} \sim \mathcal{N}(0, \sigma_1^2)$
  draw $\boldsymbol{\gamma}^{(c)} \sim \mathcal{N}(\mathbf{0}, \sigma_2^2 \, \text{I}_P)$
  for $a = 1$ to A **do**
    draw $\boldsymbol{s}_a^{(c)} \sim \mathcal{N}(\mathbf{0}, \sigma_3^2 \, \text{I}_K)$
  end for
  for $a = 1$ to A **do**
    for $r = 1$ to A **do**
      if $r \neq a$ **then**
        set $p_{ar}^{(c)} = \sigma(b^{(c)} + \boldsymbol{\gamma}^{(c)\top} \boldsymbol{x}^{(ar)} - ||\boldsymbol{s}_a^{(c)} - \boldsymbol{s}_r^{(c)}||)$
      else
        set $p_{ar}^{(c)} = 0$
      end if
    end for
  end for
end for
for $t = 1$ to T **do**
  draw $l_t \sim \text{Unif}(1, C)$
end for

Generative Process Pseudocode: Local Variables

for $d = 1$ to D **do**
  draw $\boldsymbol{\theta}^{(d)} \sim \mathrm{Dir}(\alpha, \boldsymbol{m})$
  set $\bar{N}^{(d)} = \max(1, N^{(d)})$
  for $n = 1$ to $\bar{N}^{(d)}$ **do**
    draw $z_n^{(d)} \sim \boldsymbol{\theta}^{(d)}$
    if $N^{(d)} \neq 0$ **then**
      draw $w_n^{(d)} \sim \boldsymbol{\phi}^{(z_n^{(d)})}$
    end if
  end for
  for $c = 1$ to C **do**
    set $\bar{N}^{(c|d)} = \sum_{n=1}^{\bar{N}^{(d)}} \delta(l_{z_n^{(d)}} = c)$
  end for
  for $r = 1$ to A **do**

    draw $y_r^{(d)} \sim \mathrm{Bern}(\sum_{c=1}^{C} \frac{\bar{N}^{(c|d)}}{\bar{N}^{(d)}} p_{a^{(d)}r}^{(c)})$
  end for
end for

Inference Pseudocode

for $i = 1$ to I **do**
  for $d = 1$ to D **do**
    for $n = 1$ to $N^{(d)}$ **do**
      $z_n^{(d)} \sim P(z_n^{(d)} \,|\, \mathcal{B}, \Gamma, \mathcal{S}, \mathcal{L}, \mathcal{Z}_{\setminus d,n}, \mathcal{W}, \mathcal{Y}, \mathcal{X}, \mathcal{A})$
    end for
  end for
  for $t = 1$ to T **do**
    $l_t \sim P(l_t \,|\, \mathcal{B}, \Gamma, \mathcal{S}, \mathcal{L}_{\setminus t}, \mathcal{Z}, \mathcal{X}, \mathcal{A})$
  end for
  $\mathcal{B}, \Gamma, \mathcal{S} \sim P(\mathcal{B}, \Gamma, \mathcal{S} \,|\, \mathcal{L}, \mathcal{Z}, \mathcal{Y}, \mathcal{X}, \mathcal{A})$
end for

Human Coding Topics Used in Analysis

- Validate interpretation of top words
- Top ten emails.
  - Rank: $\frac{(\text{tokens assigned to topic})^2}{\text{tokens in email}}$
- Label: common thread among emails.

Male-Centric Dyads

MM > FM > MF > FF
Planning & Information Technology
Solid Waste and Recycling & Health
Sheriff & Health
Tax Administrator & Planning
Tax Administrator & Social Services
Inspections & Tax Administrator
Animal Control & Finance
Environment & Health
Environment & Solid Waste and Recycling

Male-Centric Topics of Communication

Coding	MM > FM > MF > FF
Manager	office, email, time, staff, meeting, work, good
IT	electronic, mail, intended, email, message
Health	health, department, project, email, code, garden
Comments	jail, mobile, inmates, ago, money, jails
Planning	east, planning, street, court, administrator
Public Works	public, nashville, suite, washington
Public Works	email, energy, carolina, north, address
Planning	description, director, street, church, suite
Planning	description, fax, phone, director, street, church
Zoning	board, meeting, planning, amendment
Emergency	operations, emergency, director, lines, street
Development	description, director, development, projects