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

Research Question

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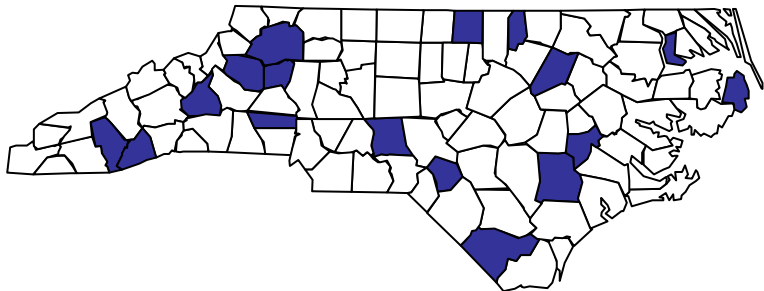
Previous Research

- 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

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

Hypothesis: women are 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
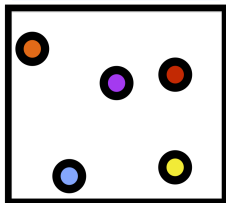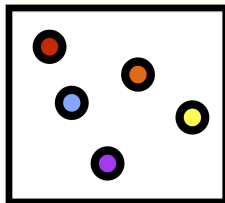
A Generative Model for Email Data
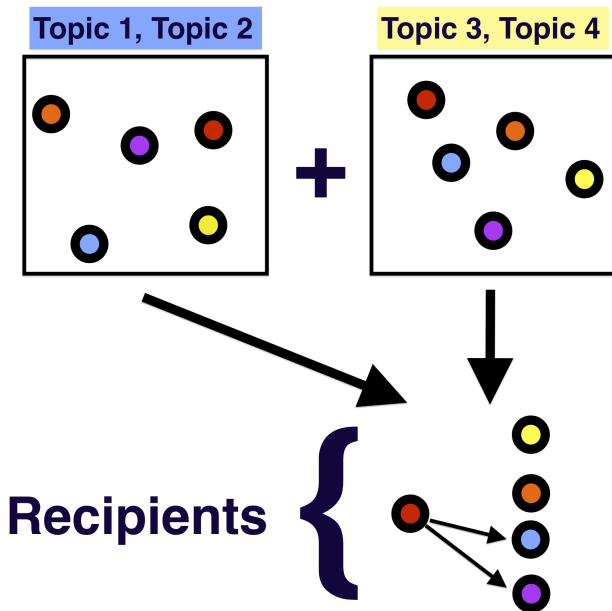
Specification and Methods

- Bayesian inference via MCMC.

- Separate model for each county.

- Include gender mixing covariates.

- 40 topics, 4 interaction patterns.

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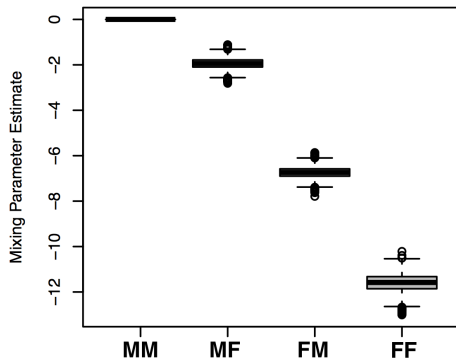
Coding	Topic top words
Sandy	will, track, winds, system, forecast
Sandy	storm, sandy, high, coastal, tides, night
Harbor	status, update, boat, today, weather
Planning	board, meeting, planning, seafood, hearing
Storm	box, planning, director, permit, building

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.

- Hypothesis: women are more active in majority female groups.

- Different operationalizations of domain.

- We find support.

Generative Process Pseudocode: Global Variables

1: **for** $t = 1$ to T **do**
2:     draw $\boldsymbol{\phi}^{(t)} \sim \text{Dir}(\beta, \boldsymbol{n})$
3: **end for**
4: **for** $c = 1$ to C **do**
5:     draw $b^{(c)} \sim \mathcal{N}(0, \sigma_1^2)$
6:     draw $\boldsymbol{\gamma}^{(c)} \sim \mathcal{N}(\boldsymbol{0}, \sigma_2^2 \, \text{I}_P)$
7:     **for** $a = 1$ to A **do**
8:         draw $\boldsymbol{s}_a^{(c)} \sim \mathcal{N}(\boldsymbol{0}, \sigma_3^2 \, \text{I}_K)$
9:     **end for**
10:     **for** $a = 1$ to A **do**
11:         **for** $r = 1$ to A **do**
12:             **if** $r \neq a$ **then**
13:                 set $p_{ar}^{(c)} = \sigma(b^{(c)} + \boldsymbol{\gamma}^{(c)\top} \boldsymbol{x}^{(ar)} - ||\boldsymbol{s}_a^{(c)} - \boldsymbol{s}_r^{(c)}||)$
14:             **else**
15:                 set $p_{ar}^{(c)} = 0$
16:             **end if**
17:         **end for**
18:     **end for**
19: **end for**
20: **for** $t = 1$ to T **do**
21:     draw $l_t \sim \text{Unif}(1, C)$
22: **end for**

Generative Process Pseudocode: Local Variables

```
23: for d = 1 to D do
24:     draw θ^(d) ~ Dir(α, m)
25:     set N̄^(d) = max(1, N^(d))
26:     for n = 1 to N̄^(d) do
27:         draw z_n^(d) ~ θ^(d)
28:         if N^(d) ≠ 0 then
29:             draw w_n^(d) ~ φ^(z_n^(d))
30:         end if
31:     end for
32:     for c = 1 to C do
33:         set N̄^(c|d) = Σ_{n=1}^{N̄^(d)} δ(l_{z_n^(d)} = c)
34:     end for
35:     for r = 1 to A do
36:
37:         draw y_r^(d) ~ Bern(Σ_{c=1}^{C} (N̄^(c|d))/(N̄^(d)) p_{a^(d)r}^(c))
38:     end for
39: end for
```

$$23: \textbf{ for } d = 1 \text{ to } D \textbf{ do}$$
$$24: \quad \text{draw } \boldsymbol{\theta}^{(d)} \sim \text{Dir}(\alpha, \boldsymbol{m})$$
$$25: \quad \text{set } \bar{N}^{(d)} = \max(1, N^{(d)})$$
$$26: \quad \textbf{for } n = 1 \text{ to } \bar{N}^{(d)} \textbf{ do}$$
$$27: \quad\quad \text{draw } z_n^{(d)} \sim \boldsymbol{\theta}^{(d)}$$
$$28: \quad\quad \textbf{if } N^{(d)} \neq 0 \textbf{ then}$$
$$29: \quad\quad\quad \text{draw } w_n^{(d)} \sim \boldsymbol{\phi}^{(z_n^{(d)})}$$
$$30: \quad\quad \textbf{end if}$$
$$31: \quad \textbf{end for}$$
$$32: \quad \textbf{for } c = 1 \text{ to } C \textbf{ do}$$
$$33: \quad\quad \text{set } \bar{N}^{(c|d)} = \sum_{n=1}^{\bar{N}^{(d)}} \delta(l_{z_n^{(d)}} = c)$$
$$34: \quad \textbf{end for}$$
$$35: \quad \textbf{for } r = 1 \text{ to } A \textbf{ do}$$
$$37: \quad\quad \text{draw } y_r^{(d)} \sim \text{Bern}(\sum_{c=1}^{C} \frac{\bar{N}^{(c|d)}}{\bar{N}^{(d)}} p_{a^{(d)}r}^{(c)})$$
$$38: \quad \textbf{end for}$$
$$39: \textbf{end for}$$

Inference Pseudocode

```
1: for i = 1 to I do
2:     for d = 1 to D do
3:         for n = 1 to N^(d) do
4:             z_n^(d) ~ P(z_n^(d) | B, Γ, S, L, Z_\d,n, W, Y, X, A)
5:         end for
6:     end for
7:     for t = 1 to T do
8:         l_t ~ P(l_t | B, Γ, S, L_\t, Z, X, A)
9:     end for
10:    B, Γ, S ~ P(B, Γ, S | L, Z, Y, X, A)
11: end for
```

1: **for** $i = 1$ to I **do**
2: $\quad$ **for** $d = 1$ to D **do**
3: $\quad\quad$ **for** $n = 1$ to $N^{(d)}$ **do**
4: $\quad\quad\quad$ $z_n^{(d)} \sim P(z_n^{(d)} \,|\, \mathcal{B}, \Gamma, \mathcal{S}, \mathcal{L}, \mathcal{Z}_{\setminus d,n}, \mathcal{W}, \mathcal{Y}, \mathcal{X}, \mathcal{A})$
5: $\quad\quad$ **end for**
6: $\quad$ **end for**
7: $\quad$ **for** $t = 1$ to T **do**
8: $\quad\quad$ $l_t \sim P(l_t \,|\, \mathcal{B}, \Gamma, \mathcal{S}, \mathcal{L}_{\setminus t}, \mathcal{Z}, \mathcal{X}, \mathcal{A})$
9: $\quad$ **end for**
10: $\quad$ $\mathcal{B}, \Gamma, \mathcal{S} \sim P(\mathcal{B}, \Gamma, \mathcal{S} \,|\, \mathcal{L}, \mathcal{Z}, \mathcal{Y}, \mathcal{X}, \mathcal{A})$
11: **end for**

- 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