

---

# Abstract Outline

MATTHEW DENNY

FRIDAY 18<sup>TH</sup> SEPTEMBER, 2015

---

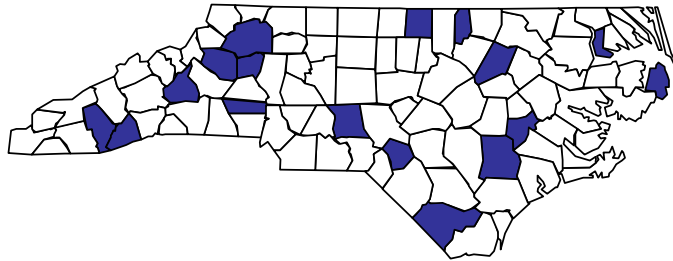
## 1 Gender in Organizations

1. Gender bias in organizations is well documented in terms of pay, prestige, position, and social interaction.
2. Gender equity is normatively important.
3. Scholars have sought to understand the roots of gender bias, but have not had access to much behavioral data.
4. With the increasing use of electronic communication, and the rise of e-government/ transparency, we are now able to use government email to study gender bias using behavioral data.
5. Therefore, we choose to explore the relationship between gender and communication patterns in government organizations.
6. To do this, we collect and analyze large scale email data from a sample of 17 local governments.

## 2 Data

1. North Carolina has robust public records laws which let us collect data.
2. We did this as part of a transparency-by-conformity field experiment.
3. 17 counties participated, allowing us to compare across organizations.
4. Provide some descriptive statistics of the data.

Figure 1: North Carolina county map.



County	Mgrs.	Female	Internal	Total
Alexander	21	9	907	11,924
Caldwell	20	8	121	
Chowan	23	11	2,027	11,737
Columbus	24	10	920	12,707
Dare	27	12	2,247	
Duplin	27	14	1,914	
Hoke	24	11	1,106	5,565
Jackson	24	6	1,499	
Lenoir	20	5	560	10,499
Lincoln	22	7	573	8,727
McDowell	17	5	326	3,494
Montgomery	18	10	680	2,465
Nash	19	8	1,147	9,133
Person	21	9	1,491	14,023
Transylvania	20	4	1,857	14,088
Vance	18	8	185	4,349
Wilkes	17	2	303	8,443
<b>Totals:</b>	362	139	17,863	117,154

Table 1: Participating county email statistics. **Mgrs.** is the total number of department managers in a county, **Female** is the number of female managers in that county, **Internal** is the number of emails sent between managers in a county, and **Total** is the total number of emails sent and received by department managers in each county in our sample. Note that in this study, we only make use of the internal email data. Some email **Total**'s are omitted due to challenges in determining which emails (not sent by managers) were valid in these counties.

5. These counties are a representative sample and we have tons of data so lets start analyzing it.

### 3 Descriptive Analysis

1. If we want to know about gendered patterns of communication in organizations, we should start by looking at aggregate statistics by gender:

	Male	Female
Proportion of Sample	61.6%	38.4%
Average Emails Sent	48.3	51
Average Emails Received	70.8	71.6
Average Recipients Per Email Sent	1.45	1.43
Recipients of Emails Sent By Female Managers	62.2%	37.8%
Recipients of Emails Sent By Male Managers	60.7%	39.3%

Table 2: Manager email statistics by gender.

2. The gender split is 60% men, 40% women and they both send and receive a comparable number of emails on average, so no volume difference.
3. Male and female managers also do not communicate with their own gender at a higher rate.
4. No gender bias? One place the literature tells us to look for gender bias is in the roles/positions in the organization.



Figure 2: Heat map depicting the number of emails sent from the row department to the column department aggregated across counties. Departments were also hand coded into one of 25 different categories based on given titles, to group departments that perform a similar function. The right margin displays the number of counties that had a manager of type X, and the number of those managers who were women.

5. We can see that some departments are mostly men – the county manager (the boss), the sheriff, and the emergency manager, for example. Others are mostly women – the HR, finance, health, for example.
6. So there is some evidence of gendering in positions, but what about men and women who perform the same position? Are they treated differently by the county manager?
7. Lets look at an example of finance managers because every county has one. First off, there are only two female county managers and the finance managers in their counties are both women, so we focus instead on counties with male managers. Of these 15 counties, 10 have a female finance manager and 5 have a male finance manager. How do they communicate with the county manager?

	Finance Mgr. Gender Male	Female
Emails Sent to County Manager	48.4% (151/312)	20.5% (334/1628)
Percent of Emails Sent by County Manager to Finance Manager	21.2% (182/859)	24.2% (372/1535)

Table 3: County manager – finance manager communication for male county managers. Out of 15 counties with a male county manager, 10 had female finance managers and 5 had male finance managers. The (Fraction) displayed under each percentage is the raw number of emails sent to that manager divided by the total number of emails sent, aggregated over counties.

8. Male county managers do not communicate preferentially with finance managers of either gender, so no evidence for biased treatment by supervisor within positions in this case.
9. Female finance managers are sending lots more emails on average than their male counterparts. Are they being assigned different tasks?
10. More broadly, are female department managers performing different tasks than their male counterparts?
11. To understand this, we need to look at email content.
12. A quick way to start is by searching for things like question marks, exclamation points, and "Thanks" to see if we observe differences by gender.

Email Content	Male	Female	t-test
Question Marks per Email Sent	0.218	0.235	0.033
Exclamation Marks per Email Sent	0.088	0.363	0.001
"Thanks" per Email Sent	0.340	0.425	0.028

Table 4: Manager-to-manager email communication statistics by gender across all departments and all counties. The **t-test** column reports Welch 2-sample t-test  $p$  values for difference in means where applicable. When calculating content statistics, only the body text was used so as to avoid double counting subject lines in responses.

13. We do see significant differences, but why?
14. If we want to do this in a more principled way, and for all departments at once, we might want to use a topic model to categorize emails, then look at gender mixing using the LSM in each topic.
15. However, the gender mixing parameters we infer using this approach would be confounded by selection effects. As a toy example, either men prefer to talk to male coworkers over female coworkers about football, or women prefer not to talk about football at work, so do not participate in those conversations.
16. To disentangle these effects, we need a joint model for email topical content, and the structure of communication.

## 4 A Model of Email Content

1. Motivation for building a model for email data – who you send an email to depends on what it is about. And what you say depends on who you are talking to.
2. Our solution: a generative model for email topics and recipients. Then discuss the existing TPME model and why we build on it.
3. Overview of the generative process in plain english.
4. Describe the generative process for LDA part of model.
5. Model based topic clustering.

6. Explain how draws of message recipients are conditioned on the email topics.
7. Describe the generative process for the latent space portion of model.
8. Summarize the generative process and lead into inference.

## 4.1 Inference

1. We have to invert the generative process to perform inference on the model parameters.
2. We use block Metropolis Hastings within Gibbs sampling.
3. A beta R package is available for those interested.
4. Discuss our model specification and justify our hyper-parameter choices.

## 5 Gender Mixing

