Abstract Outline

Matthew Denny   Friday 18th 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, extent, and nature of gender bias, but have not had access to primary source data. This has traditionally been observational, ethnographic and self reports, which can be biased. Limited in scope.

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 primary source data.

5. This also makes our analyses replicable.

6. Therefore, we choose to explore the relationship between gender and communication patterns in government organizations.

7. To do this, we collect and analyze large scale email data from a sample of 17 local governments.

8. What do we look for, how do we do it, what do we find – at a high level.

2 Data

1. North Carolina has robust public records laws which let us collect data.

2. 22 complied with our request but only 17 counties did so in a way that was useful to our research, allowing us to compare across organizations.

3. Provide some descriptive statistics of the data.

Figure 1: North Carolina county map.
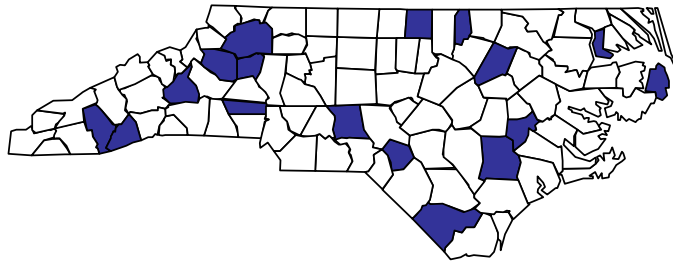
County	Mgrs.	Female	Internal	Total
Alexander	21	9	907	11,924
Caldwell	20	8	121	
Chowan	23	11	2,027	11,737
Columbus	24	10	920	12,707
Dare	27	12	2,247	
Duplin	27	14	1,914	
Hoke	24	11	1,106	5,565
Jackson	24	6	1,499	
Lenoir	20	5	560	10,499
Lincoln	22	7	573	8,727
McDowell	17	5	326	3,494
Montgomery	18	10	680	2,465
Nash	19	8	1,147	9,133
Person	21	9	1,491	14023
Transylvania	20	4	1,857	14,088
Vance	18	8	185	4,349
Wilkes	17	2	303	8,443
Totals:	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.

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

3   Descriptive Analysis

1. We want to see if there is a relationship between sender gender and recipient gender.

2. 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 Managers	61.6%	38.4%
Average # Emails Sent	48.3	51
Average # Recipients Per Email Sent	1.45	1.43
Average # Emails Received	70.8	71.6

Table 2: Manager email statistics by gender.

	Male Recipient	Female Recipient
Female Sender	62.2%	37.8%
Male Sender	60.7%	39.3%

Table 3: Manager email statistics by gender.

3. 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.

4. Male and female managers also do not communicate with their own gender at a higher rate.

5. We have replicated previous research, we don't see gender bias, we could stop here but because we have more information we are going to dig deeper.

6. No gender bias? One place the literature tells us to look for gender bias is in the roles/positions in the organization.

Aggregate Interdepartmental Email Volume Across 17 NC Counties

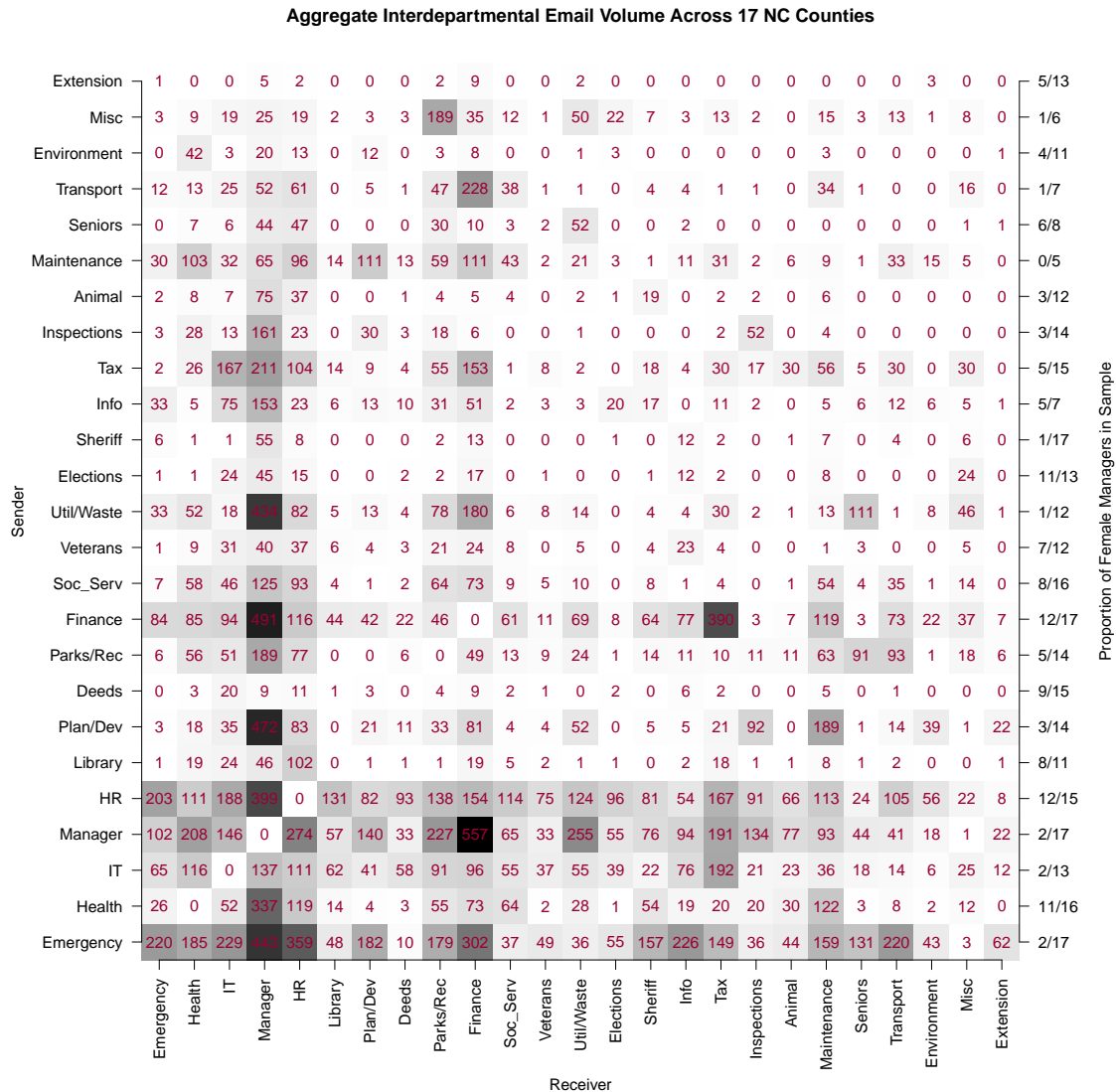| Sender \ Receiver	Emergency	Health	IT	Manager	HR	Library	Plan/Dev	Deeds	Parks/Rec	Finance	Soc_Serv	Veterans	Util/Waste	Elections	Sheriff	Info	Tax	Inspections	Animal	Maintenance	Seniors	Transport	Environment	Misc	Extension	Prop. Female Managers
Extension	1	0	0	5	2	0	0	0	2	9	0	0	2	0	0	0	0	0	0	0	0	0	3	0	0	5/13
Misc	3	9	19	25	19	2	3	3	189	35	12	1	50	22	7	3	13	2	0	15	3	13	1	8	0	1/6
Environment	0	42	3	20	13	0	12	0	3	8	0	0	1	3	0	0	0	0	0	3	0	0	0	0	1	4/11
Transport	12	13	25	52	61	0	5	1	47	228	38	1	1	0	4	4	1	1	0	34	1	0	0	16	0	1/7
Seniors	0	7	6	44	47	0	0	0	30	10	3	2	52	0	0	2	0	0	0	0	0	0	0	1	1	6/8
Maintenance	30	103	32	65	96	14	111	13	59	111	43	2	21	3	1	11	31	2	6	9	1	33	15	5	0	0/5
Animal	2	8	7	75	37	0	0	1	4	5	4	0	2	1	19	0	2	2	0	6	0	0	0	0	0	3/12
Inspections	3	28	13	161	23	0	30	3	18	6	0	0	1	0	0	0	2	52	0	4	0	0	0	0	0	3/14
Tax	2	26	167	211	104	14	9	4	55	153	1	8	2	0	18	4	30	17	30	56	5	30	0	30	0	5/15
Info	33	5	75	153	23	6	13	10	31	51	2	3	3	20	17	0	11	2	0	5	6	12	6	5	1	5/7
Sheriff	6	1	1	55	8	0	0	0	2	13	0	0	0	1	0	12	2	0	1	7	0	4	0	6	0	1/17
Elections	1	1	24	45	15	0	0	2	2	17	0	1	0	0	1	12	2	0	0	8	0	0	0	24	0	11/13
Util/Waste	33	52	18	434	82	5	13	4	78	180	6	8	14	0	4	4	30	2	1	13	111	1	8	46	1	1/12
Veterans	1	9	31	40	37	6	4	3	21	24	8	0	5	0	4	23	4	0	0	1	3	0	0	5	0	7/12
Soc_Serv	7	58	46	125	93	4	1	2	64	73	9	5	10	0	8	1	4	0	1	54	4	35	1	14	0	8/16
Finance	84	85	94	491	116	44	42	22	46	0	61	11	69	8	64	77	390	3	7	119	3	73	22	37	7	12/17
Parks/Rec	6	56	51	189	77	0	0	6	0	49	13	9	24	1	14	11	10	11	11	63	91	93	1	18	6	5/14
Deeds	0	3	20	9	11	1	3	0	4	9	2	1	0	2	0	6	2	0	0	5	0	1	0	0	0	9/15
Plan/Dev	3	18	35	472	83	0	21	11	33	81	4	4	52	0	5	5	21	92	0	189	1	14	39	1	22	3/14
Library	1	19	24	46	102	0	1	1	1	19	5	2	1	1	0	2	18	1	1	8	1	2	0	0	1	8/11
HR	203	111	188	399	0	131	82	93	138	154	114	75	124	96	81	54	167	91	66	113	24	105	56	22	8	12/15
Manager	102	208	146	0	274	57	140	33	227	557	65	33	255	55	76	94	191	134	77	93	44	41	18	1	22	2/17
IT	65	116	0	137	111	62	41	58	91	96	55	37	55	39	22	76	192	21	23	36	18	14	6	25	12	2/13
Health	26	0	52	337	119	14	4	3	55	73	64	2	28	1	54	19	20	20	30	122	3	8	2	12	0	11/16
Emergency	220	185	229	443	359	48	182	10	179	302	37	49	36	55	157	226	149	36	44	159	131	220	43	3	62	2/17

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.

7. 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.

8. So there is some evidence of gendering in positions, but what about men and women who perform the same position? We can form a contingency table that records the number of emails sent by each gender combination against every department communication dyad.

9. 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.

10. However, the gender mixing parameters we infer using this approach would be confounded by selection effects. As a toy example, we would not be able to tell whether men prefer to talk to male coworkers over female coworkers about football, or women prefer not to talk about football at work, and so they do not participate in those conversations.

11. 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

1. Since we have modeled content and structure together, we have broken up the confounding we were concerned about, and we can now interpret the gender mixing parameters that come out of our model with more confidence.

2. Lets return to the question we posed at the end of our descriptive analysis section: Are women being excluded from conversations by their male coworkers, or are they simply not discussing the same things via email?

3. Lets form some expectations about the relationship between gender mixing parameters we estimate.

4. If we observe a positive relationship between the female-female mixing parameters and the male-female or female-male mixing parameters we estimate, this is evidence that there are male and female topics, and if you want to talk about a female (male) topic, you send an email to a female (male) colleague. This is evidence for the gendering of communication content. As a toy example, if you want to email about football, you email a male colleague.

5. If we observe a negative relationship, this is evidence that there are some things that men send or receive emails about, and some things women send or receive emails about. This is evidence for gender bias in organizational roles.
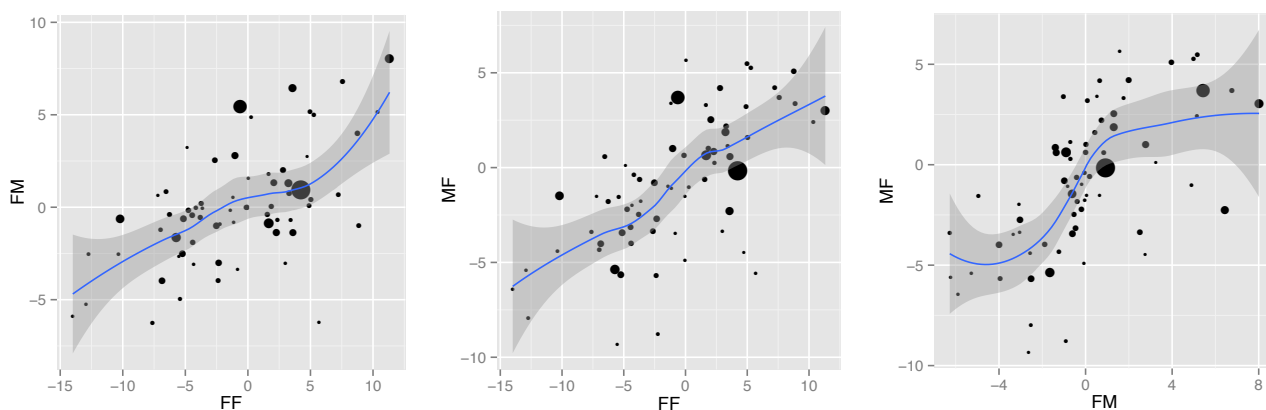
6. We observe a pattern that is consistent with a gendering of communication content.