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

Matthew Denny [*], **James ben-Aaron** [†], **Hanna Wallach** [† ‡], **and Bruce Desmarais** [*]

[*]Penn State University,[†]University of Massachusetts Amherst, and [‡]Microsoft Research NYC

In this paper, we study the role of gender in local government organizations. We analyze email data from seventeen county governments in North Carolina to understand the relationship between a department manager's gender and the emails they send and receive. First, we use descriptive statistics to identify aggregate gender-homophilous and gender-heterophilous patterns from the numbers of emails sent and received by all department managers. In contrast to previous research, we find no strong evidence of either gender homophily or heterophily. To investigate this finding, we therefore analyze department-to-department communication patterns. Here, we do find evidence of gender bias, with some departments exhibiting gender homophily and others exhibiting heterophily. Finally, to determine the extent to which these patterns are driven by communication content, we use a recently developed latent variable model to identify topic-specific communication subnetworks for each county. We find differing degrees of gender homophily and heterophily for different topics of communication. From a policy perspective, these findings suggest that a gender-equitable working environment cannot be created by hiring decisions alone, as gender bias in communication still exists independent of the positions held by men and women.

Gender in Organizations

Researchers have documented gender bias in both the public and private sectors. Women receive lower pay, hold less prestigious positions in these organizations, have reduced opportunity for advancement, and are excluded from the organizational locus of control [5, 3, 11, 2, 7]. As a result, organizations that aspire to a just, efficient, and sustainable professional culture strive to provide men and women with equal treatment in the workplace [9]. In practice, however, these organizations have historically found it difficult to understand the micro-level causes of gender bias due to the limited availability of primary source data. The data used in these studies [e.g., 6, 1, 8] has traditionally been observational, ethnographic, or self-reported. which can be biased and is often limited in scope.

However, with the increasing use of electronic communication, and the rise of transparency initiatives within government, scholars are now increasingly able to use public records requests to gather primary source email communication data. Using public records requests as a method for data collection also provides a major benefit of making the data collection process replicable. Therefore, we choose to study the relationship between a department manager's gender, and the emails they send and receive, in a sample of local governments. To do this, we collect and analyze a large scale email data set from a sample of seventeen county governments in North Carolina.

We begin by examining the aggregate patterns of email communication among department managers in these counties by gender, and the gender breakdown of managers for each department. Then we analyze interdepartmental email flows by gender of the email sender and its recipients. Finally, we apply a statistical model for communication network data to infer the topic-conditional structure of manager-to-manager email communication in these county governments. Our results indicate that while aggregate patterns of communication among department managers do not display significant gender bias, the topics of communication, and positions held by men and women differ significantly.

Data

The data used in this study were collected in 2013 via a series of public records requests made to all one hundred North

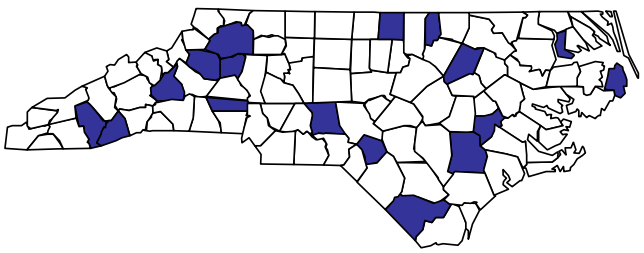

Fig. 1. North Carolina map with counties who provided useable data highlighted.

Table 1. Breakdown of department managers by gender, and the total number of emails sent by department managers, for each county in our sample.

	Manager Gender		
County	**Male**	**Female**	**Emails Sent**
Alexander	12	9	907
Caldwell	12	8	121
Chowan	12	11	2,027
Columbus	14	10	920
Dare	15	12	2,247
Duplin	13	14	1,914
Hoke	13	11	1,106
Jackson	18	6	1,499
Lenoir	15	5	560
Lincoln	15	7	573
McDowell	12	5	326
Montgomery	8	10	680
Nash	11	8	1,147
Person	12	9	1,491
Transylvania	16	4	1,857
Vance	10	8	185
Wilkes	15	2	303
Totals:	362	139	17,863

Table 2. Email statistics by manager gender.

	Manager Gender	
	Male	**Female**
Average # of emails sent	48.3	51
Average # of recipients per email sent	1.45	1.43
Average # of emails received	70.8	71.6

[1]Some counties provided too little data (i.e. 30 emails for the entire county), or provided data in paper form, which was not amenable to our analyses.

[2]Requests to each county specified a random starting month for the request, between January and October, 2103.

Carolina county governments. We selected North Carolina for our data collection efforts because the state public records laws cover email data, and prevent counties from charging unreasonable fees for providing data. Twenty-three counties complied with our request, of which seventeen provided data was useable for our research[1]. Our records request to each county included all emails sent and received by all department managers (e.g. Health, Finance, Elections) over a three-month period in 2013[2].

We received approximately 500,000 emails from the seventeen county governments which provided useable data, of which approximately 17,000 were sent between county department managers, which we focus on in the analyses presented here. Figure 1 highlights the counties which responded to our requests for data. We also collected detailed metadata on the department and gender of the 362 department managers in our sample. We display some basic descriptive statistics for each county in table 1. Approximately 40% of department managers in our sample are women, although there is significant variation in the proportion of female department managers across counties. It is important to note that the counties in our sample are statistically indistinguishable from the rest of the counties in North Carolina on a number of demographic dimensions including population, per-capita income, and percent of the population that is white.

Descriptive Analysis

We begin our analysis of the relationship between a department manager's gender and the emails they send and receive, by looking for differences in the propensity for department managers to send emails to other managers of the same gender (*gender-homophilous* communication), and managers of the opposite gender (*gender-heterophilous* communication). Table 2 provides some basic descriptive statistics including the average number emails sent and received by male and female department managers in our sample. On average, male and female department managers send and receive a comparable number of emails, and emails sent by male and female department managers have a similar number of recipients. Therefore, if we observe gender differences in email communication, they are unlikely to be driven by some innate difference in the propensity for male and female department managers to send or receive emails.

To test whether the gender of an email sender is related to the gender of its recipients, in aggregate, we construct a contingency table of email sending and receiving by gender (see table 3). We then perform a χ^2 test for independence between the rows and columns. The χ^2 test statistic we obtain indicates that the gender of an email sender and its recipients is not independent ($\chi^2 = 92.9, p < 0.001$). However, inspection of the contingency table indicates that the test statistic is actually driven by mild gender heterophily in communication. Furthermore, the gender differences we observe are substantively quite small, as both male and female department managers send emails to recipients of each gender in rough proportion to their overall representation in the sample (60% to men, 40% to women).

Table 3. Each cell records the number of times a department manager of gender X was included as a recipient of an email sent by a department manager of gender Y. Statistics provided are calculated for all counties combined. Note that each email may have more than one recipient.

		Recipient		
		Male	**Female**	**Total**
Sender	**Male**	7,299	6,286	13,585
	Female	5,325	3,510	8,835
	Total	12,624	9,796	

Table 4. Number of male and female managers for each department.

	Emergency	*Health*	*IT*	*Manager*	*HR*	*Library*	*Plan/Dev*	*Deeds*	*Parks/Rec*	*Finance*	*Soc/Serv*	*Veterans*	*Util/Waste*
Male	15	5	11	15	3	3	11	6	9	5	8	5	11
Female	2	11	2	2	12	8	3	9	5	12	8	7	1

	Elections	*Sheriff*	*Info*	*Tax*	*Inspections*	*Animal*	*Maintenance*	*Seniors*	*Transport*	*Environment*	*Misc*	*Extension*
Male	2	16	2	10	11	9	5	2	6	7	5	8
Female	11	1	5	5	3	3	0	6	1	4	1	5

This result contradicts what we know about gender-driven interaction in the workplace, and is possibly a result of the vast majority of e-mails being substantively uninteresting (e.g. meeting reminders sent automatically from a manager's Microsoft Outlook account). Therefore, we need to partition e-mails according to content in order to observe substantively interesting patterns. Our data allow us to use the department of each manager as a rough proxy for the content of their email communication. We begin by looking for aggregate differences in the gender breakdown of department managers for each department, across our sample (see table 4). Departments were hand coded into one of twenty-five different categories based on the title given in the county directory, to group departments that perform a similar function. Note that not all departments are represented in each county.

We can see that some departments are mostly managed by men: the county manager (the supervisor of all other department managers), the sheriff, and the emergency manager, for example. Other departments are mostly managed by women: the HR, finance, and health departments, for example. A a χ^2 test for the independence of department and gender confirms our qualitative finding that there is a strong bias in which departments tend to be managed by women ($\chi^2 = 130.8, p < 0.001$). This result indicates a bias in the gender breakdown of department managers for each department, across our sample, but cannot tell us about differences in email communication by men and women who manage the same department. To answer this question, we need to know whether the gender of an email sender and the gender of its recipients is independent of the their respective departments.

Table 2 displays the aggregate number of times each department manager was a recipient of an email sent by a manager of each other department. We see pronounced differences in email sending an receiving by department, but cannot graphically disentangle the role of gender. Next, we construct a contingency table of department dyad-types (where each "dyad-type" is a unique combination of sender and recipient department, eg. finance $\longrightarrow$ HR) against gender dyad-types (where each "dyad-type" is a unique combination of sender and recipient gender, eg. male $\longrightarrow$ female). We then perform a χ^2 test on this contingency table, which indicates that the gender of

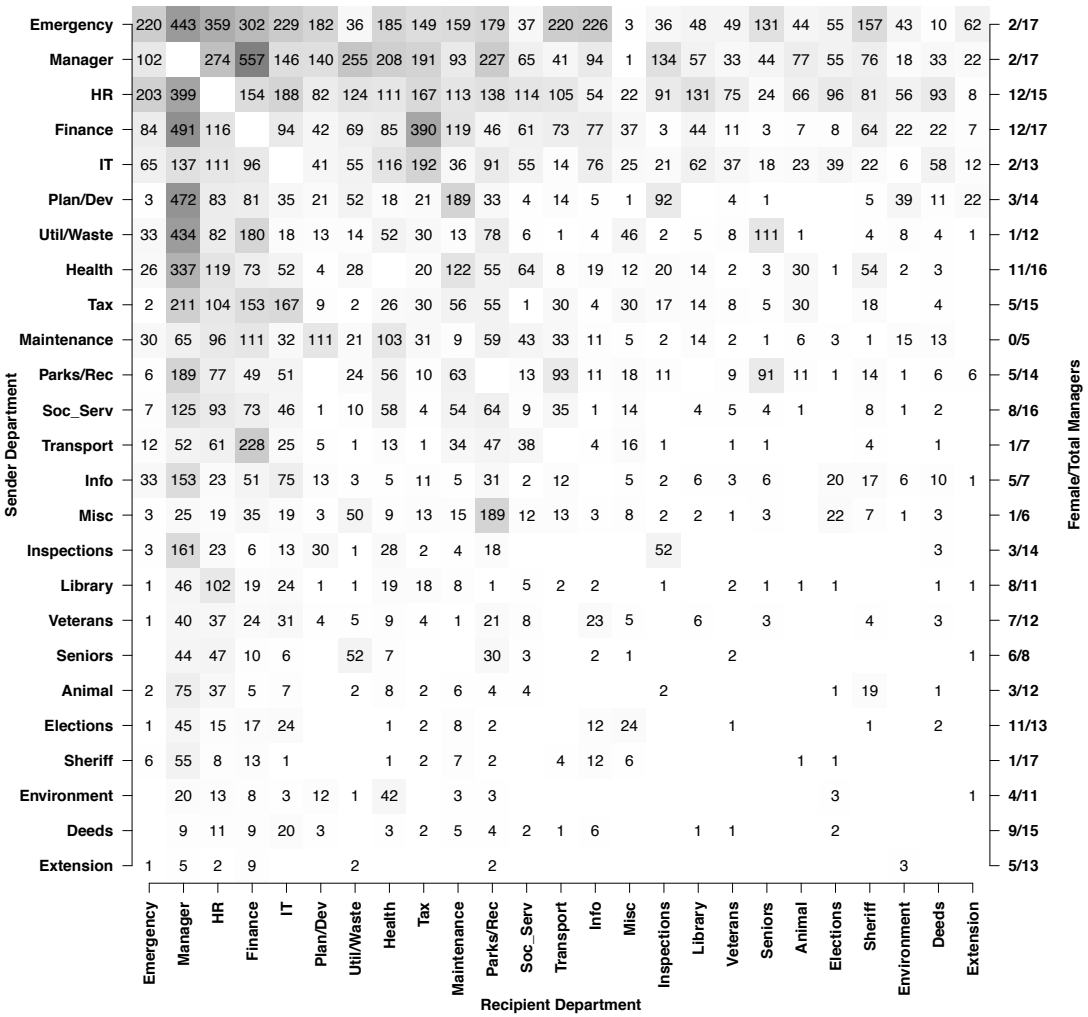

Fig. 2. Heat map depicting the total number of emails sent from the managers of each department to the managers of each other department, aggregated across counties. The right-hand y-axis displays the number of counties that had that department, and the number of those managers who were women. Departments are ordered by the total number emails their managers sent.

an email sender and its recipients is not independent of their respective departments ($\chi^2 = 37,404, p < 0.001$).

This dyadic χ^2 test tells us that there is an association between gender mixing patterns and department mixing patterns. Indeed, when we construct similar contingency tables for each department, we find that email senders in utilities and waste management departments exhibit gender-homophilous behavior ($\chi^2 = 10.49, p < 0.001$), senders in the tax department exhibit gender-heterophilous behavior ($\chi^2 = 4.54, p < 0.034$), and senders in the IT department seem unbiased ($\chi^2 = 1.23, p > 0.27$). However, we can't say from these results that the form of gender bias depends upon the domain of interaction since factors other than content are likely to vary with department labels, including the county in which the manager is working, the experience of the manager, and the manager's status in the organizational hierarchy. To disentangle these effects, we need explicitly model the topical content of emails. To do this, we extend a recently introduced joint model for communication topical content and network structure [12], and apply it to our email data.

A Model of Email Content

Here we provide a brief overview of our extension to the topic-partitioned multinetwork embedding (TPME) model [12]. In particular, we illustrate how this model lets us make inferences about the gender-specific patterns of communication in organizations, and how these patterns change with the topics of communication. Our model follows Krafft et al. [12] by integrating the latent space network model (LSM) [10] with latent Dirichlet allocation (LDA) [4] to jointly model the content and structure of an email network. This model is designed to be applied to email data where we observe each message sender and its recipients, as well as the email content (the words).

At a high level, our model assumes the following generative process for a message sent across the email network. First, we sample the content of the message following the generative process for LDA. Then, for a given message sender, we draw whether each other actor in the network is a recipient of that message following the LSM. However, different topics are associated with different latent spaces, so how likely each actor is to be a recipient of a particular message is dependent on the topical content of that message. We review this process in greater detail below.

The message content is assumed to be generated via LDA. Under this model, we assume that all unique words in our vocabulary are associated to varying degrees with each of T latent *topics*. Each topic is then a distribution over unique words (or *word types*), $\boldsymbol{\phi}^{(t)}$. For example *doctor*, *virus*, and *medicine* might all have high probability in a topic about hospitals, while *cat* and *music* might have low probability in that topic. We further assume that each topic distribution is drawn from a Dirichlet prior:

$$\boldsymbol{\phi}^{(t)} \sim \text{Dirichlet}(\beta, \boldsymbol{n}) \text{ for } t = \{1, ..., T\} \quad [1]$$

where the base measure $\boldsymbol{n}$ is the mean of the distribution, and the concentration parameter β controls how concentrated typical draws are around that mean.

To generate each of the D documents (messages) in our corpus, we first draw a document-specific distribution over topics:

$$\boldsymbol{\theta}^{(d)} \sim \text{Dirichlet}(\alpha, \boldsymbol{m}) \text{ for } d = \{1, ..., D\} \quad [2]$$

Once we have drawn our document-specific distribution over topics we can then generate each of N tokens (words) in the document via a two step process. First we draw the latent topic assignment for that token from the document specific distribution over topics.

$$z_n^{(d)} \sim \boldsymbol{\theta}^{(d)} \text{ for } n = \{1, ..., N\} \quad [3]$$

Then we draw the word type of that token from the topic specific distribution over word types.

$$w_n^{(d)} \sim \boldsymbol{\phi}^{(z_n^{(d)})} \text{ for } n = \{1, ..., N\} \quad [4]$$

We proceed in this manner for all tokens in all documents.

In our extension of TPME, we further assume each topic is uniquely associated with one of C clusters, drawn from a discrete uniform distribution.

$$C_t \sim \text{U}(1, ..., C) \text{ for } t = \{1, ..., T\} \quad [5]$$

The intuition is that different broad content areas of communication (e.g. *planning*, *sports*, *meeting planning*) -each of which is a collection of more specific topics- will imply a different pattern of communication. Importantly, under this model, a message can be about a number of topics, and these topics can be associated with different clusters. Since each cluster is associated with a different pattern of communication, the probability that any actor is a recipient of that particular message depends on an admixture of their probabilities of being a recipient under those different communication patterns. For example, if a department manager in an organization sent an email message to schedule a budget meeting, they would likely include both staff whose job includes setting up meetings, and staff who needed to provide input on the budget as recipients.

Finally, for a given cluster c, the probability that an actor is selected as a recipient of a particular message is specified by the LSM [10]. Under this model, we assume there is some baseline propensity to include message recipients on any email, which is governed by a scalar intercept parameter.

$$b^{(c)} \sim \text{Normal}(\mu, \tau^2) \quad [6]$$

For example, messages about sensitive HR matters will probably include fewer recipients than messages announcing a department party. Second, some attributes X, of the sender and potential recipient (which we assume are observed) may also affect the probability of that actor being a recipient. The effects that each of these L different attributes have on the probability of an actor being included as a message recipient can then be parameterized by a vectors:

$$\gamma_l^{(c)} \sim \text{M. V. Normal}(\lambda, \eta^2) \text{ for } l = \{1, ..., L\} \quad [7]$$

These covariate effect parameter vectors may vary in size depending on whether the effect is only associated with the attribute of the email sender (only requiring a single parameter), or with the combination of covariate values for the sender and recipient. For example, there is a great deal of social science literature demonstrating gender homophily in communication. Our model could generate a gender-homophilous email communication pattern by drawing positive parameters for male–male and female–female sender–recipient dyads, and negative parameters for female–male and male–female sender–recipient dyads.

The additional variation in the probability that a message is sent between two actors is governed by how close they are in a k dimensional *latent social space*. To generate these latent positions, each actor a is assumed to have some position in the k-dimensional latent space

$$\mathbf{s}_a^{(c)} \sim \text{M. V. Normal}(0, \sigma^2) \text{ for } a = \{1, ..., A\} \quad [8]$$

and the probability that they send a message to a recipient $r \neq a$ is decreasing in their latent distance from r. Thus the probability of actor r being a recipient of a message from actor a under this model is:

$$P(y_{a,r}^c = 1) = \text{logisitic sigmoid}\left(b^c + \sum_L [\mathbf{X}_{a,r}\gamma_l^c] - |\mathbf{s}_a^c - \mathbf{s}_r^c|\right) \quad [9]$$

Edge values for an individual message are then drawn from a Bernoulli distribution. However, the probability that $y_{a,r} = 1$ for a given message d may be dependent on multiple latent spaces (because documents can be about multiple topics). The weight given to $P(y_{a,r}^{(c)})$ for each cluster is determined by the proportion of tokens in the message that are assigned to topics associated with cluster c.

$$P(y_{a,r}^{(d)} = 1) = \sum_C P(y_{a,r}^c = 1) \times * \quad [10]$$

where $*$ is the proportion of tokens in document d assigned to topics associated with cluster c. In this way, we condition the probability that any given actor r is a recipient of an email (document) from actor a on the content of that email, thus linking LDA and the LSM in the generative process.

Inference. Given that we do not directly observe the generative process, we have to perform inference for the posterior distribution of our model parameters given the data. We do this by *inverting* the generative process. This problem is analytically intractable, so we must approximate the posterior distribution via Markov chain Monte Carlo (MCMC) methods. We perform inference for this model via block Metropolis within Gibbs sampling[3]. A further discussion of this model is beyond the scope of this paper, but it provides a powerful and flexible framework to investigate the gender-specific patterns of communication in organizations.

However, it is important to briefly discuss the interpretation of the LSM parameters inferred using our model, as their meaning is somewhat counter-intuitive. Durring inference, the latent positions of actors capture all un-modeled factors associated with the propensity for any two actors to form a tie. In a typical social network this might include difficult or impossible to measure quantities like how nice a person is, or whether two people have *chemistry*. Additionally, it may capture observable traits that the researcher was unable to collect data

[3] The inference algorithm is currently implemented in a beta version as an R package, and is available here: github.com/matthewjdenny/ContentStructure

on (e.g. sexual orientation, or ethnicity). This makes the latent positions difficult to interpret when including covariates in the model (because some characteristics that would otherwise be considered latent are fixed by the covariate data), so great care should be taken when doing so. In the analysis that follows, we refrain from interpreting the inferred latent positions of actors.

We apply our model separately to the email data from each county and then pool our model results. To perform inference on real data, we must first select a number of model hyperparameters. In particular, we must select the number topics, number of clusters, and topic model hyper-parameters[4] to be used by our model, which we hold constant across counties. We choose to include manager gender as the only covariate in our specification, and include the full complment of gender mixing parameters in our model[5]. In addition, we fix the male-male mixing parameter at zero for our analysis, to aid in directly interpreting the other gender mixing parameters. We select 40 topics and 4 clusters, to provide reasonable granularity in capturing variation in the content of communication, while improving the interpretability of the latent space model results by constraining the number of possible patterns of communication. This choice has a practical advantage of ensuring that enough data will be available to fit each clusters latent space with reasonably low uncertainty in the parameter estimates.

To perform inference, we must also select a number of iterations for our MCMC sampler. In the first step of inference, we alternate between one iteration of Gibbs sampling for the LDA parameters, and 1,000 iterations of Metropolis Hastings sampling for the LSM parameters, as the Metropolis Hastings algorithm explores the parameter space much more slowly than Gibbs sampling. We did this for a total of 4,000 iterations of Gibbs sampling, until Geweke statistics indicated convergence in the un-normalized LDA model log likelihood for all counties. We then ran the LSM component of our model for an additional 10,000,000 iterations, holding the LDA parameters fixed, to ensure that all latent space parameter estimates had converged.

Gender Mixing

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.

Lets return to the question we posed at the end of our descriptive analysis section: is there gender bias in the patterns of communication in these county governments.

Add in Bruce's MV KS test results for each county here.

Interpret the results.

Take as a case study, a county with the highest degree of gender bias, as identified by this method.

interpret the topic model output from two different clusters and discuss

ACKNOWLEDGMENTS. This work was supported by US National Science Foundation Grant CISE-1320219 (Hanna Wallach and Bruce A. Desmarais, PIs)

References

1. Susan M Adams, Atul Gupta, Dominique M Haughton, and John D Leeth. Gender differences in ceo compensation: evidence from the usa. *Women in Management Review*, 22(3):208–224, 2007.
2. James Albrecht, Anders Björklund, and Susan Vroman. Is there a glass ceiling in Sweden? *Journal of Labor Economics*, 21(1):145–177, 2003.
3. William T. Bielby and James N. Baron. Men and Women at Work : Sex Segregation and Statistical Discrimination. *American Journal of Sociology*, 91(4):759–799, 1986.
4. DM Blei, AY Ng, and MI Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.
5. Daniel J. Brass. Men's and Women's Networks: A Study of Interaction Patterns and Influence in an Organization. *Academy of Management Journal*, 28(2):327–343, 1985.
6. Emilio J Castilla. Gender, race, and meritocracy in organizational careers. In *Academy of Management Proceedings*, volume 2005, pages G1–G6. Academy of Management, 2005.
7. Colin Duncan and Wendy Loretto. Never the Right Age? Gender and Age-Based Discrimination in Employment. *Gender, Work & Organization*, 11(1):95–115, 2004.
8. Kim M Elsesser and Janet Lever. Does gender bias against female leaders persist? quantitative and qualitative data from a large-scale survey. *Human Relations*, 64(12):1555–1578, 2011.
9. Robin J Ely and Debra E Meyerson. Theories of gender in organizations: A new approach to organizational analysis and change. *Research in organizational behavior*, 22:103–151, 2000.
10. Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, December 2002.
11. Herminia Ibarra. Homophily and Differential Returns : Sex Differences in Network Structure and Access in an Advertising Firm. *Administrative Sceince Quarterly*, 37:422–447, 1992.
12. Peter Krafft, Juston Moore, Bruce A Desmarais, and Hanna Wallach. Topic-partitioned multinetwork embeddings. In *Advances in Neural Information Processing Systems Twenty-Five*, 2012.

[4] We use uniform base measures $\mathbf{m}$ and $\mathbf{n}$, and set $\alpha = 1$ and β equal to 0.01 times the length of the vocabulary for each county. This is standard practice in the literature using LDA, and provides good performance.

[5] male–male, male–female, female–male, and female–female