EMPIRICAL ASSESSMENT OF PROGRAMS TO PROMOTE COLLABORATION: A NETWORK MODEL APPROACH[1]

BY KATHERINE R. MCLAUGHLIN AND JOSHUA D. EMBREE

Oregon State University and RAND Corporation

Collaboration networks are thought to be desirable to foster both individual and population productivity. Often programs are implemented to promote collaboration, for example, at academic institutions. However, few tools are available to assess the efficacy of these programs, and very few are data-driven. We carried out a survey at California State University, San Marcos during the 2012–2013 academic year to measure five types of collaboration ties among professors in five science departments at the university over time. During the time period of study, professors participated in NIH-sponsored curriculum development activities with members of other departments. It was hypothesized that participation in these activities would also foster overall collaboration between these departments.

This survey enables the exploration of several methodological and theoretical challenges in network research. In this paper we develop a statistical approach to assess the impact of programmatic interventions on collaboration using model-assisted social network analysis. We derive and implement a hierarchical Bayesian approach to modeling error-prone responses in surveys and examine the effect of an intervention on network structure. Based on this analysis we find an increase in educational collaboration over time after adjusting for the length of time each professor had to form collaborative ties at the university.

1. Introduction. It has long been recognized that many projects benefit from a collaborative approach. Every year, numerous grants are funded to promote cooperation and collaboration among a variety of entities. For example, efforts to promote healthy eating in schools may bring together food vendors, school officials, and nutritionists. In an academic setting, an interdisciplinary approach that brings together experts from a variety of fields is thought to be beneficial. These projects are often undertaken on the understanding that collaboration is not a one-time benefit, but instead something that once fostered has long-lasting merits. Such programs are promoted explicitly and implicitly by the National Science Foundation and National Institutes of Health [McCullough (1992)]. How do we evaluate

these claims? How can we tell if new relationships are formed, if they are necessary or beneficial, and if they are likely to continue promoting collaborative behavior?

This general question falls within the context of the assessment literature. Evaluating the efficacy of funding is a problem without an obvious solution [McCullough (1992)]. Efforts to tackle assessment have been largely based on subjective or anecdotal evidence, and not data-driven. We now discuss several of the data-driven methods that do exist.

Gajda (2004) creates an assessment tool called the Strategic Alliance Formative Assessment Rubric (SAFAR), where program evaluators rate collaboration at each stage of alliance development. A collaboration baseline was established at a focus group meeting, and then "integration scores" ranging from 1 to 5 were gathered after one and two years for a project intended to promote the collaboration of mental health, law enforcement, and social service agencies to aid children and families. This study did not have a rigorous quantitative approach, as rankings were somewhat arbitrary. Gajda (2004) notes that the integration score increased from 1 to 2.2 during the first year and from 2.2 to 2.6 over the second year. However, we have no way of knowing whether or not this was a significant increase.

Similarly, Gajda and Koliba (2007) design and use a tool called the Community of Practice: Collaboration Assessment Rubric (CoPCAR), which presents quality of interpersonal collaboration across a continuum of 1 to 6. In another scale method, Frey et al. (2006) introduce the Levels of Collaboration Scale, where respondents were asked to what extent they collaborated with each other grant partner on a scale from 0 to 5. Looking for simple increases over time misses more complex questions, such as localized collaboration within subgroups, or identifying particularly influential nodes. A more complex approach could begin to address these questions. Woodland and Hutton (2012) refine the approaches of Gajda (2004) and Gajda and Koliba (2007), among others, into the Collaboration Evaluation and Improvement Framework (CEIF). They examine a bipartite network between individuals and groups, and use social network analysis to look for bottlenecks, individuals that may be overextended or underextended, and groups that may be too big or too small.

Cross et al. (2009) used network data to assess the development of interagency relationships by evaluating changes in tie strength over time. All entities first met at a retreat where they were asked to rate on a scale from 1 to 5 their level of collaboration with all other participants. The same measures were recorded at a second retreat and two follow-up group meetings. Again, these ratings are subject to the same pitfalls as those of Gajda (2004). Cross et al. (2009) focus on the formation of communities within the network, including the homophilous behavior of such subgroups. They use modularity (clustering measured on edges, but not covariates) to assess goodness of fit for the clusters they created. Although this approach demonstrates one of the many promising ways in which network analysis can be incorporated into the assessment literature, it is subject to the same data quality issues and subjectivity of qualitative metrics.

McCullough (1992) focuses on the assessment efforts undertaken by the National Science Foundation (NSF). He notes that peer opinion is the most frequently utilized tool for evaluating research results, which commonly takes the form of the assessment of publications. However, although "bibliometrics" (i.e., examining the counts, content, and citations of publications) is quantitative, it is still quite subjective. Metrics are subject to methodological problems such as self-citation, circular citation among a small group of researchers, and the separation of one project into many small publications. Additionally, publications may not tell the entire story of a project or its intangible effects. McCullough (1992) notes that NSF-supported researchers in 3–5 fields produce "many more" publications than researchers in the same fields supported by other sources. However, causal conclusions should not be drawn from this observation. Perhaps NSF-supported researchers feel more obligated to publish, something that is not necessarily indicative of the impact of the research.

Eisenberg and Swanson (1996) incorporate network analysis to examine the efficacy of Connecticut's Healthy Start program. They looked at the position of this program within a network of groups providing similar services in several communities, and performed descriptive social network analysis. In particular, they considered measures of density and centrality, and attempted to identify subgroups or cliques within the network. Eisenberg and Swanson (1996) noted that since the Healthy Start programs tended to be "brokers" (have both high out-referrals and in-referrals), they served an important role in the community. While informative, this type of descriptive analysis relies heavily on the collected data and often overlooks potential shortcomings such as missing information and various types of survey bias, and is limited to the scope of the original project. For example, in this analysis, the networks were constructed using open-ended interviews and possibly incomplete directories.

An effective data-driven assessment, then, requires data that are collected from nonsubjective questions. For example, "Have you ever submitted a grant as co-PI with Person A," not "Rate your collaboration with Person A on a scale from 1 to 5." Direct surveying of the persons or entities involved is also preferable to indirect metrics, like publication counts. Finally, it is essential to correct for survey error, particularly if there is reason to believe responses are error-prone. Error could result from nonresponse, recall bias, social desirability bias, or a variety of other measurement errors.

Assessing and adjusting total survey error in network surveys is more complex than in traditional surveys because of the dependence between the observations [Marsden (1990)]. Handcock and Gile (2010) develop a modeling framework for sampled network data. Although sampling error is an important component of total survey error, in our survey we conducted a census and therefore focus on measurement error. Borgatti, Carley and Krackhardt (2006) evaluate network centrality measures when random errors are added to the network, including edge addition

and deletion, which we are concerned with in this paper. They find that the accuracy of centrality measures declines smoothly and predictably with the amount of error, and that dense networks were more robust to edge addition while sparse networks were more robust to edge deletion. To address measurement error, Borgatti, Carley and Krackhardt (2006) suggest creating confidence intervals for centrality measures and using caution when more than 5% of ties are erroneous. Wang et al. (2012) simulate a variety of measurement error scenarios, including false negative and false positive edges. They find that networks with low average clustering and less positively skewed degree distributions are most resistant to measurement error. They also find that false negative edges tend to affect network measures, but that most node-level network measures are relatively robust to false positive edges. These studies examined the effect of measurement error; our model provides a way to adjust for it for erroneous edges.

A variety of dynamic social network models exist, including the separable temporal exponential-family random graph model (STERGM) [Krivitsky and Handcock (2014)] and the stochastic actor-oriented model (SAOM) [Snijders (1996)]. Due to the nature of our survey, where data were collected retroactively at only one time point as described in Section 2, we cannot fit a dynamic model. In general we support dynamic network models for collaboration assessment, as they allow for more complex relationships to be evaluated over time. However, such models are not always possible. We focus on a model-assisted approach to adjust error-prone survey responses, and rely on a descriptive comparison to assess change.

In Section 2 we describe our survey design and implementation, and motivate the use of a network model to elucidate our initial findings. Then in Section 3 we describe the derivation and implementation of the latent network model used to improve upon this rudimentary analysis. We assess model performance through a simulation study in Section 4, apply it to our survey data in Section 5, and provide concluding statements in Section 6.

2. Survey design and description. Our motivating case study is a network of education collaboration ties among professors at California State University (CSU), San Marcos in five science departments. A grant was awarded to CSU, San Marcos, along with several other universities, aimed at improving the quantitative and computational (Q&C) skills of biology majors at the school. The steering committee for the grant at CSU, San Marcos decided to focus on cross-departmental curriculum development to achieve this goal.[2] Each year, a cohort of professors (ideally from a variety of departments) formed a panel to design and/or modify one of six courses in the biology department to create a more fully integrated curriculum. After implementing this approach, members of the Steering Committee were interested in a method of evaluating the efficacy of their approach.

---

[2]Other universities took very different approaches, for example using the money from the grant to build a computer laboratory and upgrade the technology available to students.

We designed a survey intended to measure changes in collaboration over time, which we gave to the complete set of professors among the five departments. This survey was implemented during the Spring of 2013. Professors were given a roster of all other names and asked to nominate anyone they had a certain type of relation with. Definitions were given for the different relations. Five potential relations were posited: grant-writing collaborations, educational project collaborations (such as the Q&C grant), research collaborations, seeking professional advice, and discussing personal matters. For this paper, we focus on the educational project collaborations network, as it is the one that most directly pertains to the grant. On the survey, an educational project collaboration was defined as "professional collaboration on an education-motivated project (e.g., curriculum development, pedagogy workshops and/or research)." Noneducational research was a separate question and is not considered in the education collaboration network.

Additionally, the professors were asked to give the academic term (Spring, Summer, or Fall) and year that the interaction began. Because we were only collecting data at the end of the grant retroactively (and not when it began in Fall 2008, or earlier), we had to rely on respondents' memories, which could be unreliable. In an attempt to obtain accurate timestamps, we encouraged the faculty to refer to their email or other records to attempt to ascertain when a given relation began. Despite our efforts, the large number of asymmetric ties in the data suggests the presence of both memory recall and perception bias. A person may not recall a relationship at all, or may recall that it began at a different time than it actually did. This could be due to simple forgetfulness or to a different definition of what constitutes the relationship forming. In order to account for these issues with memory recall and perception bias, we postulate the existence of a "true" or latent network that stochastically generates the observed network. We will model these sources of disparity between the observed and latent networks.

The survey also posed a variety of demographic questions to complement those aimed at gathering information on the network ties. The covariate most directly related to tie formation is department, as professors tended to form more collaborative ties within departments than between them. Additional variables were used to threshold the data, as explained below, and to adjust for possible differences in rates of tie formation.

To construct the final data sets used for analysis, several modifications were made. First, the survey was originally given to 46 professors among the five departments at CSU, San Marcos. We received 37 completed surveys (a response rate of 80.4%), with nine professors away from the university or otherwise unable to answer.[3] The surveys were administered in person by a member of the Steering Committee who was not affiliated with any of the departments of interest, and who was not a subject in the study. Only the 37 professors who returned

---

[3]Additionally, two of these nine people were the dean and assistant dean of the university, who perhaps should be ineligible in any case as they are not strictly research or teaching faculty.

completed surveys are eligible nodes in the networks considered here, so we disregarded nominations of professors who did not return the survey. We discussed potential nonresponse biases with the steering committee who were very familiar with all 46 professors we attempted to survey. They did not think that those who were missing would problematically bias our analysis, so we proceeded with the smaller network of 37 professors.

For each tie, the professor was asked to identify when that type of interaction began, in terms of an academic term (Spring, Summer, or Fall), and a year. We then created a new network for each academic year (i.e., Fall 2008, Spring 2009, and Summer 2009 are one academic year). The Q&C grant took effect during the Fall 2008 semester. For this paper, we consider the five years during the Q&C grant, and the five years preceding it. This allows us to look at the changes in the network over time both before the grant took effect and during it to examine trends in tie formation.

It is also worth making note of an implicit assumption here. Since we only gathered information about the formation of ties, and not the possible dissolution of existing ties, our network can only grow over time. We feel that this is a safe assumption to make regarding educational collaboration, as the dissolution of a collaborative tie would require a relatively significant event, such as a professional feud. Collaborative relationships can remain dormant—that is, even if two people are not currently engaged in a project together but have collaborated in the past, they may still collaborate in the future, so we would consider a tie to be present between them. Even a professor who left the university (which did not happen in our study) could still collaborate with colleagues from the original university, so the tie should still exist.

We also recorded the academic term in which a professor was first hired at CSU, San Marcos. No professors nominated each other before they had both begun working at the university. In cases when a professor nominated someone but did not provide a year for the tie formation, we by default assigned the year as the first year in which both professors worked at the university (i.e., the later of the two hire dates in the dyad). This is a conservative assumption because ties are attributed in most cases to a time before the start of the Q&C grant, rather than during it.

We chose to examine ten different time points: the five years during the Q&C grant, and the five years immediately preceding it. The dyad census for these 10 years for the education research collaboration network, showing the counts of mutual, asymmetric, and null ties, is displayed in Table 1, as well as the size of the network at each time.

Since we asked people about their education collaborations, we anticipated observing a relatively large proportion of mutual ties in the network, that is, two people who work together in some respect will nominate one another on the survey. To the contrary, we observe the opposite: many asymmetric ties and few mutual ties. The inconsistency in reported ties is likely an unfortunate byproduct of a retroactive survey. Specifically, people have a difficult time recalling all of their collaborative relationships over the course of up to 20 years. Moreover, individuals

TABLE 1
Observed dyad counts for 10 years. Times 1–5 are before the inception of the Q&C grant; times 6–10 are during the grant. The network size n of CSU, San Marcos professors who had been hired by each time is also given

Time	n	Mutual	Asymmetric	Null	Total
1	28	9	44	325	378
2	29	11	48	347	406
3	30	12	50	373	435
4	30	12	53	370	435
5	31	13	57	395	465
6	33	20	71	437	528
7	33	37	82	409	528
8	34	45	91	425	561
9	36	64	105	461	630
10	37	80	106	480	666

have different thresholds for what constitutes a collaboration, despite our efforts to provide a clear definition on the survey. Thus, differences in both recall accuracy and perception threshold can contribute to the reporting of asymmetric ties in cases where we believe we should only observe mutual ties. In the next section, we propose a model to address this issue.

3. A model for error-prone responses. To account for differences in recall accuracy and perception, we propose a Bayesian framework to model "false negative" and "false positive" rates within a network. By definition, a false negative is the omission of a tie in the observed network where we believe one truly exists in the latent network, and a false positive is the inclusion of a tie where one should not exist.[4]

3.1. *Statistical framework.* We assume that at each time point k there exists a true latent collaboration network which arises from some unobserved stochastic mechanism. The model can be fit on a single time point k, and is not dynamic by nature. This means that separate models are fit to each time point k and then evaluated as a collection after model fitting. This is a potential drawback of the work, but we note that our data necessitated this format and that a model for error-prone responses is also an important contribution to the literature. Let Θ_k denote

---

[4]These false positives and false negatives are "errors" only in the sense that there is a discrepancy between the latent and observed networks. We are not implying that a person's perception is incorrect.

the latent network at time k. Since we assume that collaboration is by definition mutual, Θ_k is an undirected binary network. Thus,

$$
(1) \qquad \Theta_{ij,k} = \Theta_{ji,k} = \begin{cases} 1 & (i, j) \in E(\Theta_k), \\ 0 & (i, j) \notin E(\Theta_k), \end{cases}
$$

where $E(\Theta_k)$ is the edge set of Θ_k. By definition, $i, j \in V(\Theta_k)$, the vertex set of Θ_k, and $\|V(\Theta_k)\| = n$ for a network Θ_k with n nodes. $\Theta_{ii,k} = 0$ by definition.

Let Y_k denote the observed network at time k. This network is potentially directed, that is, $Y_{ij,k} :\neq Y_{ji,k}$. We assume that $V(\Theta_k) = V(Y_k)$, that is, the vertex set is the same between the latent and observed networks at each time. Define

$$
(2) \qquad Y_{ij,k} = \begin{cases} 1 & (i, j) \in E(Y_k), \\ 0 & (i, j) \notin E(Y_k). \end{cases}
$$

For the San Marcos data, we use $k = 1, \ldots, 10$ to represent ten academic years (the five during the Q&C grant: 08/09, 09/10, 10/11, 11/12, and 12/13; and the five immediately preceding the Q&C grant: 03/04, 04/05, 05/06, 06/07, and 07/08); and $i, j \in \{1, \ldots, 37\}$ for the 37 professors in the network.

For a potential tie from actor i to actor j, there are four possible outcomes, summarized below.

$Y_{ij,k} = 1 \mid \Theta_{ij,k} = 1$	Correctly identified tie
$Y_{ij,k} = 1 \mid \Theta_{ij,k} = 0$	False positive
$Y_{ij,k} = 0 \mid \Theta_{ij,k} = 1$	False negative
$Y_{ij,k} = 0 \mid \Theta_{ij,k} = 0$	Correctly omitted tie

We ascribe the following probabilities to these nontrivial erroneous situations:

$$
p_{i,k} = P(Y_{ij,k} = 0 \mid \Theta_{ij,k} = 1) \qquad \text{false negative,}
$$

$$
q_{i,k} = P(Y_{ij,k} = 1 \mid \Theta_{ij,k} = 0) \qquad \text{false positive.}
$$

These are, respectively, the false negative and false positive rates. Note that $p_{i,k}$ and $q_{i,k}$ are only referenced by the subscripts i and k, that is, we are assuming each actor has a unique value of p and q at each time point, but it remains the same for all possible out-ties from that actor. That is, an actor is equally likely to falsely nominate any actors to whom she is not tied to in the latent network.

The likelihood is a Bernoulli mixture,

$$
(3) \qquad \begin{aligned} & P(Y_{ij,k} = y_{ij,k} \mid \Theta_{ij,k} = \theta_{ij,k}, p_{i,k}, q_{i,k}) \\ & = \big[(1 - p_{i,k})^{y_{ij,k}} (p_{i,k})^{1-y_{ij,k}}\big]^{\theta_{ij,k}} \big[(q_{i,k})^{y_{ij,k}} (1 - q_{i,k})^{1-y_{ij,k}}\big]^{1-\theta_{ij,k}} \end{aligned}
$$

and we assume general dyad independence so we can write the joint likelihood as

$$(4) \quad P(Y_k \mid \Theta_k, p_k, q_k) = \prod_{i=1}^{n} \prod_{j=1}^{n} \left[(1 - p_{i,k})^{y_{ij,k}\theta_{ij,k}} \right] \left[(p_{i,k})^{(1-y_{ij,k})\theta_{ij,k}} \right]$$

$$\times \left[(q_{i,k})^{y_{ij,k}(1-\theta_{ij,k})} \right] \left[(1 - q_{i,k})^{(1-y_{ij,k})(1-\theta_{ij,k})} \right]$$

$$(5) \qquad = \prod_{i=1}^{n} \left[(1 - p_{i,k})^{\sum_{j=1}^{n} y_{ij,k}\theta_{ij,k}} \right] \left[(p_{i,k})^{\sum_{j=1}^{n} (1-y_{ij,k})\theta_{ij,k}} \right]$$

$$\times \left[(q_{i,k})^{\sum_{j=1}^{n} y_{ij,k}(1-\theta_{ij,k})} \right]$$

$$\times \left[(1 - q_{i,k})^{\sum_{j=1}^{n} (1-y_{ij,k})(1-\theta_{ij,k})} \right].$$

3.2. *Specifying prior knowledge about reporting accuracy.* Following the work of Butts (2003) and referencing examples in Marcum, Bevc and Butts (2012) and Butts, Acton and Marcum (2012), we construct a hierarchical Bayesian model. We ascribe priors to the false negative rate, the false positive rate, and the latent network. We base our a priori knowledge of $p_{1,k}, \ldots, p_{n,k}$ and $q_{1,k}, \ldots, q_{n,k}$ on a beta distribution because of its generalizability. Further, since the actors are surveyed in isolation, we can consider the priors $p_{1,k}, \ldots, p_{n,k}$ and $q_{1,k}, \ldots, q_{n,k}$ independent. The distributions are explicitly defined below:

$$[p_{i,k} \mid \alpha_{p,k}, \beta_{p,k}] \sim \text{Beta}(\alpha_{p,k}, \beta_{p,k}),$$

$$[q_{i,k} \mid \alpha_{q,k}, \beta_{q,k}] \sim \text{Beta}(\alpha_{q,k}, \beta_{q,k}).$$

Thus, up to a normalizing constant,

$$(6) \qquad P(p_k \mid \alpha_{p,k}, \beta_{p,k}) \propto \prod_{i=1}^{n} (p_{i,k})^{\alpha_{p,k}-1} (1 - p_{i,k})^{\beta_{p,k}-1},$$

$$(7) \qquad P(q_k \mid \alpha_{q,k}, \beta_{q,k}) \propto \prod_{i=1}^{n} (q_{i,k})^{\alpha_{q,k}-1} (1 - q_{i,k})^{\beta_{q,k}-1}.$$

Finally, we need to put a prior on Θ_k. Again, we follow the recommendations of Butts (2003) and use an informative a priori value for the network density based on similar collaboration networks as the prior for $\Theta_{ij,k}$ when both persons i and j have been hired at time k. Therefore, the probability of a tie existing from i to j is

$$(8) \qquad P(\Theta_{ij,k} = 1) = \delta_{ij,k} = \begin{cases} d_k & \text{if } i, j \in V(\Theta_k), \\ 0 & \text{if } i \text{ and/or } j \notin V(\Theta_k), \end{cases}$$

where d_k is the informed prior network density at time k. Note that in this case d_k is not considered random or differentiated between respondents, but a natural extension would allow for these possibilities. In this form we can impose a structural zero for any professor who has not yet been hired at time k.

3.3. *Derivation of posterior distributions.* The posterior distribution of $[\Theta_k, p_k, q_k \mid Y_k]$ results from 5, 6, 7, and 8. The full derivation is provided in Appendix A:

$$(9) \quad P(\Theta_k, p_k, q_k \mid Y_k) \propto P(Y_k \mid \Theta_k, p_k, q_k)$$

$$\times P(\Theta_k, p_k, q_k \mid \delta_k, \alpha_{p,k}, \beta_{p,k}, \alpha_{q,k}, \beta_{q,k})$$

$$\propto P(Y_k \mid \Theta_k, p_k, q_k)$$

$$\times P(p_k \mid \alpha_{p,k}, \beta_{p,k}) P(q_k \mid \alpha_{q,k}, \beta_{q,k}) P(\Theta_k \mid \delta_k)$$

$$(10) \qquad \propto \prod_{i=1}^{n} (p_{i,k})^{\alpha_{p,k} + \sum_{j=1}^{n}(1 - y_{ij,k})\theta_{ij,k} - 1}$$

$$\times \prod_{i=1}^{n} (1 - p_{i,k})^{\beta_{p,k} + \sum_{j=1}^{n} y_{ij,k}\theta_{ij,k} - 1}$$

$$\times \prod_{i=1}^{n} (q_{i,k})^{\alpha_{q,k} + \sum_{j=1}^{n} y_{ij,k}(1 - \theta_{ij,k}) - 1}$$

$$\times \prod_{i=1}^{n} (1 - q_{ij,k})^{\beta_{q,k} + \sum_{j=1}^{n}(1 - y_{ij,k})(1 - \theta_{ij,k}) - 1}$$

$$\times \prod_{i=1}^{n} \prod_{j=1}^{n} (\delta_{ij,k})^{\theta_{ij,k}} (1 - \delta_{ij,k})^{(1 - \theta_{ij,k})}.$$

This posterior is difficult to explore analytically, so we implement a Gibbs sampler to draw samples from the distribution.

3.4. *Computational algorithm.* The Gibbs sampler works by taking a series of draws from the full conditionals of the posterior, constructing a Markov chain whose equilibrium distribution converges to that of the joint posterior. Therefore, to draw samples from the posterior distribution in Equation (10) using a Gibbs sampler, we need to specify the full conditional distributions. The full derivations can be found in Appendix A.

For each $i \in V(\Theta_k)$, noting that $p_{1,k}, \ldots, p_{n,k}$ and $q_{1,k}, \ldots, q_{n,k}$ are independent,

$$(11) \quad [p_{i,k} \mid \Theta_k, Y_k]$$

$$\sim \text{Beta}\left(\alpha_{p,k} + \sum_{j=1}^{n}(1 - y_{ij,k})\theta_{ij,k}, \beta_{p,k} + \sum_{j=1}^{n} y_{ij,k}\theta_{ij,k}\right),$$

(12) $[q_{i,k} \mid \Theta_k, Y_k]$

$$\sim \text{Beta}\left(\alpha_{q,k} + \sum_{j=1}^{n} y_{ij,k}(1 - \theta_{ij,k}), \beta_{q,k} + \sum_{j=1}^{n}(1 - y_{ij,k})(1 - \theta_{ij,k})\right).$$

For each $i, j \in V(\Theta_k)$ such that $i \neq j$,

(13) $[\Theta_{ij,k} \mid Y_k, p_k, q_k] \sim \text{Bernoulli}(r_{ij,k}),$

where

(14) $$r_{ij,k} = \frac{P(\Theta_{ij,k} = 1 \mid Y_k, p_k, q_k)}{P(\Theta_{ij,k} = 1 \mid Y_k, p_k, q_k) + P(\Theta_{ij,k} = 0 \mid Y_k, p_k, q_k)}$$

and

(15) $P(\Theta_{ij,k} = 1 \mid Y_k, p_k, q_k) \propto \delta_{ij,k}(p_{i,k})^{(1-y_{ij,k})}(1 - p_{i,k})^{y_{ji,k}}$

$$\times (p_{j,k})^{(1-y_{ji,k})}(1 - p_{j,k})^{y_{ji,k}},$$

(16) $P(\Theta_{ij,k} = 0 \mid Y_k, p_k, q_k) \propto (1 - \delta_{ij,k})(q_{i,k})^{y_{ij,k}}(1 - q_{i,k})^{(1-y_{ij,k})}$

$$\times (q_{j,k})^{y_{ji,k}}(1 - q_{j,k})^{(1-y_{ji,k})}.$$

With the conditional distributions in 11, 12, and 13, we can iterate through the Gibbs sampler. The algorithm is given in Algorithm 1.

At each time point k, we have modeled the latent network Θ_k along with the false negative (p_k) and false positive (q_k) rates for each actor. A simulation study to assess the model is provided in Section 4 and application to the CSU, San Marcos education collaboration network is in Section 5. Note that in this model we cannot separate sources of error (e.g., recall accuracy vs. perception bias vs. filling in the wrong bubble on the survey table by mistake), but that is not of primary interest in this study. Instead, we seek to examine the structure of the latent networks.

4. Simulation study. To explore model performance, we conduct a simulation study. For each simulation, we generate an undirected latent ("true") network Θ with $n \sim \text{Unif}[25, 40]$ and $\delta \sim \text{Unif}[0.05, 0.2]$, representing a range of credible values. Then $\Theta_{ij} \sim \text{Bernoulli}(\delta)$, where $\Theta_{ij} = \Theta_{ji}$ and $\Theta_{ii} = 0$ by definition.

We then simulate an observed network Y from this latent network, which will contain asymmetric ties and some error. We draw $p \sim \text{Beta}(2, 8)$ and $q \sim \text{Beta}(2, 8)$, as these distributions have mean 0.2, which we believe a reasonable amount of error, while still having large enough variance to encompass a variety of scenarios. Y is then simulated such that $Y_{ij} \sim \text{Bernoulli}((1 - p)^{\Theta_{ij}} q^{1-\Theta_{ij}})$, where $Y_{ii} = 0$ by definition.

We then apply the model to Y using the hyperparameters $\alpha_p = \alpha_q = 2$, $\beta_p = \beta_q = 8$, and $d = 0.15$. We use a burn-in of 1000 and a thinning interval of

---

Algorithm 1: Gibbs sampler

Input: For each time k, observed adjacency matrix Y_k and hyperparameters
$\alpha_{p,k}, \beta_{p,k}, \alpha_{q,k}, \beta_{q,k}$, and δ_k
Output: Posterior distributions $\Theta_k^{(\cdot)}, p_k^{(\cdot)}, q_k^{(\cdot)}$

1 **for** $k = 1, \ldots, K$ **do**
2 $\quad$ Initialize $\Theta_k^{(0)}, p_k^{(0)}$, and $q_k^{(0)}$;
3 $\quad$ $l := 1$;
4 $\quad$ **for** $i = 1, \ldots, n$ **do**
5 $\quad\quad$ $p_{i,k}^{(l)} \sim \text{Beta}(\alpha_{p,k} + \sum_{j=1}^n (1 - y_{ij,k})\theta_{ij,k}^{(l-1)}, \beta_{p,k} + \sum_{j=1}^n y_{ij,k}\theta_{ij,k}^{(l-1)})$;
6 $\quad\quad$ $q_{i,k}^{(l)} \sim \text{Beta}(\alpha_{q,k} + \sum_{j=1}^n y_{ij,k}(1 - \theta_{ij,k}^{(l-1)})$,
$\quad\quad$ $\beta_{q,k} + \sum_{j=1}^n (1 - y_{ij,k})(1 - \theta_{ij,k}^{(l-1)}))$;
7 $\quad$ **end**
8 $\quad$ **for** $j = 2, \ldots, n$ *and* $i = 1, \ldots, j - 1$ **do**
9 $\quad\quad$ $\Theta_{ij,k}^{(l)} \sim \text{Bernoulli}(r_{ij,k}^{(l)})$ where $r_{ij,k}$ is defined in Equation (14);
10 $\quad\quad$ Set $\Theta_{ij,k}^{(l)} = \Theta_{ji,k}^{(l)}$ since Θ_k is undirected;
11 $\quad$ **end**
12 $\quad$ $l := l + 1$;
13 $\quad$ **Repeat until** $\Theta_k^{(\cdot)}, p_k^{(\cdot)}, q_k^{(\cdot)} \sim \Theta_k, p_k, q_k \mid Y_k$;
14 **end**

---

25 iterations to make 1000 draws from the posterior distribution of θ. That is, for each simulation we get 1000 draws for the error-corrected network, denoted $\theta^{(1)}, \ldots, \theta^{(1000)}$.

We performed 100 simulations as outlined above, and also the special case where $\Theta = Y$ (resulting from $p = q = 0$). To assess the model performance in these simulations, we considered several metrics. First, we examined how different the observed network Y was from the latent network Θ by looking at the proportion of edges that were incorrect: $\rho_y = \frac{\sum_{i=1}^n \sum_{j=1}^n (Y_{ij} - \Theta_{ij})^2}{n(n-1)}$. We also calculated this proportion for each of the $l = 1, \ldots, 1000$ posterior draws of $\theta^{(l)}$: $\rho_\theta^{(l)} = \frac{\sum_{i=1}^n \sum_{j=1}^n (\theta_{ij}^{(l)} - \Theta_{ij})^2}{n(n-1)}$. If the model is correcting error-prone responses, we would expect $\rho_\theta^{(l)} < \rho_y$.

Figure 1 compares ρ_y and $\rho_\theta^{(l)}$ for each of the 101 simulations, with each of the posterior draws l shown as a gray dot and the average of $\rho_\theta^{(l)}$, denoted $\bar{\rho}_\theta$, for that simulation shown as a short black line. The $y = x$ line is provided for comparison. Note that the model does a good job of correcting error-prone responses, especially for larger proportions of incorrect edges in the observed network. An "incorrect
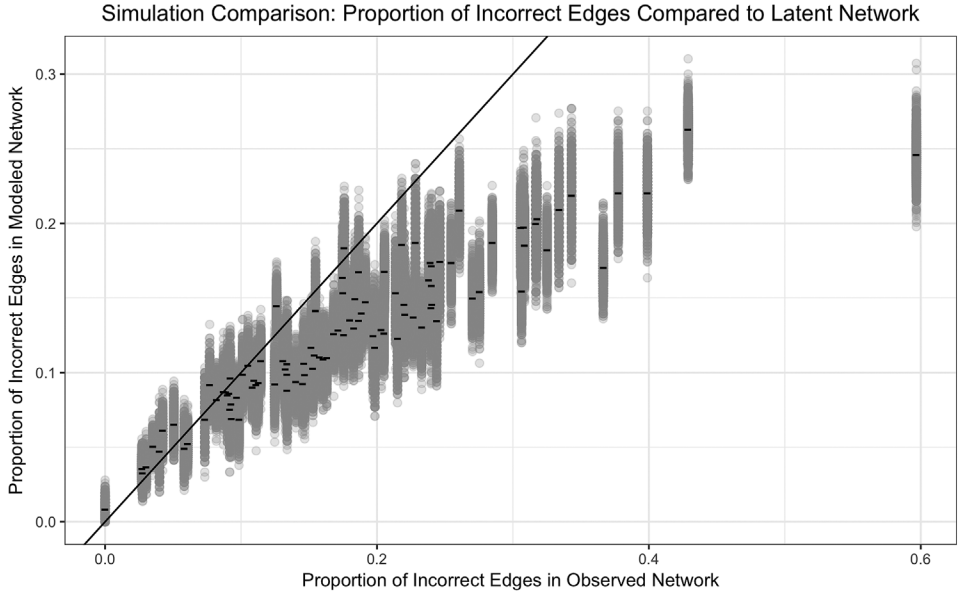
FIG. 1. *Comparison of the proportion of edges that were incorrect, where the x-axis compares the observed network to the latent network (ρ_y) and the y-axis compares each of the posterior draws from the model to the latent network $\rho_\theta^{(l)}$. The average of the posterior draws $\bar{\rho}_\theta$ for each simulation is shown as a short black line. The $y = x$ line is provided, demonstrating that the model does a good job of correcting error-prone responses, especially for larger proportions of incorrect edges in the observed network.*

edge" refers to an edge (i, j) in a given network that is different from Θ_{ij}; it could be either an edge that should exist but does not, or an edge that should not exist but does. For $\rho_y > 0.1$ (at least 10% of the edges in the observed network are incorrect, as compared to the latent network), $\bar{\rho}_\theta$ is almost always below the $y = x$ line, indicating the model is correcting some error-prone responses. And for $\rho_y > 0.25$, all posterior draws l have a smaller $\rho_\theta^{(l)}$ than ρ_y.

The preceding comparison examined differences between the observed and latent networks, and between the modeled and latent networks. We also compare the modeled networks to the observed network on edges where the observed network differs from the latent network. For all the edges (i, j) such that $\Theta_{ij} \neq Y_{ij}$, we calculate the proportion of them that are fixed by the model such that $\Theta_{ij} = \theta_{ij}^{(l)}$ for each simulation and each posterior draw. More precisely, let

$$\rho_f^{(l)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} ((Y_{ij} - \Theta_{ij})^2 (1 - (\theta_{ij}^{(l)} - \Theta_{ij})^2))}{\sum_{i=1}^{n} \sum_{j=1}^{n} (Y_{ij} - \Theta_{ij})^2}.$$

Figure 2(a) shows the proportion of incorrect edges fixed by the model for each of 1000 posterior draws for each of 100 simulations. The proportion of incorrect
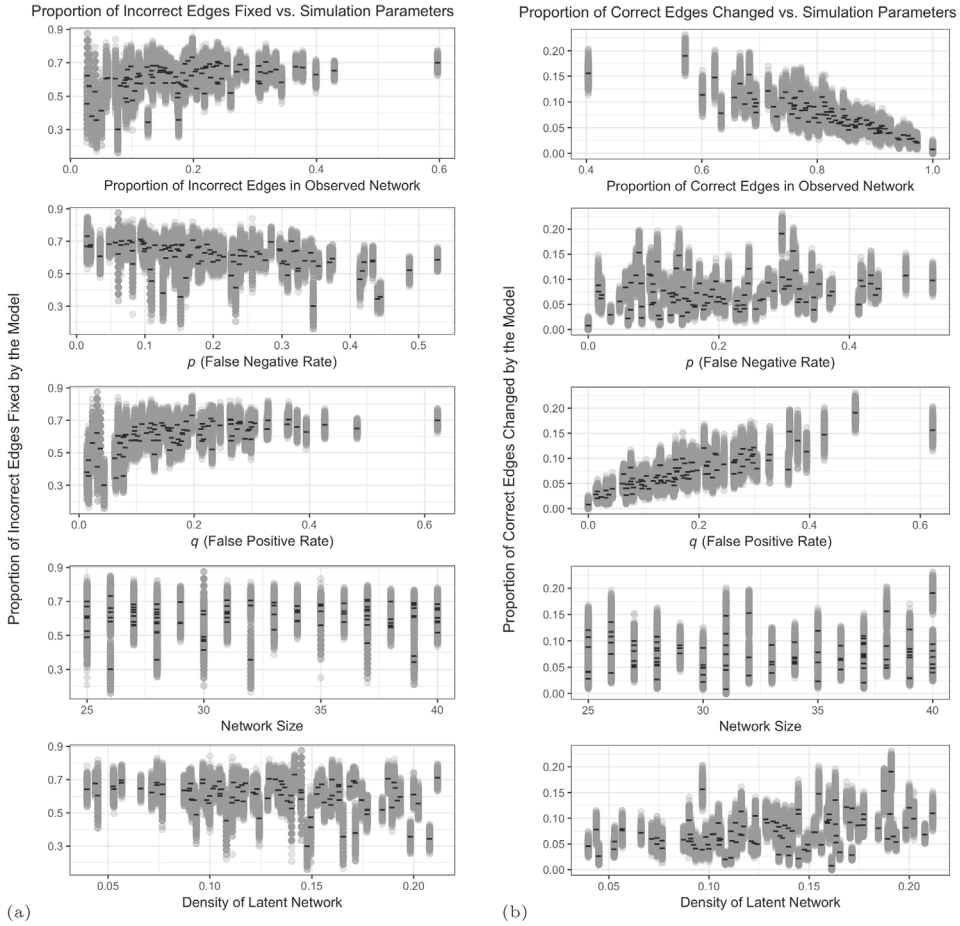
FIG. 2. *Comparison of* (a) *proportion of incorrect edges fixed*, $\rho_f^{(l)}$, *and* (b) *proportion of correct edges changed*, $\rho_c^{(l)}$, *to various simulation parameters, for* 1000 *posterior draws for each of* 100 *simulations. The average of the posterior draws*, $\bar{\rho}_f$ *and* $\bar{\rho}_c$, *respectively, for each simulation is shown as a short black line.*

edges fixed by the model ($\rho_f^{(l)}$) is plotted against five simulation parameters: the proportion of incorrect edges in the observed network (ρ_y), the false negative rate (p), the false positive rate (q), the network size (n), and the density of the latent network (δ). The $\Theta = Y$ simulation is not included since it has no erroneous ties. The black horizontal lines show the mean of $\rho_f^{(l)}$, denoted $\bar{\rho}_f$ for each simulation. The smallest $\bar{\rho}_f$ is 0.3014, indicating that on average at least 30% of erroneous edges in the observed network were fixed by the model. Most values of $\bar{\rho}_f$ are between 0.5 and 0.7, demonstrating the strength of the model. Interestingly, values of $\bar{\rho}_f$ do not show strong associations with ρ_y, p, n, or δ. There may be a slight positive association between $\bar{\rho}_f$ and q, although it is less pronounced for $q > 0.1$.

To make sure that the model is not doing more harm than good, we also examine the proportion of edges that were correct in the observed network but incorrect after the model was applied. Let this proportion be

$$\rho_c^{(l)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} ((1 - (Y_{ij} - \Theta_{ij})^2)(\theta_{ij}^{(l)} - \Theta_{ij})^2)}{\sum_{i=1}^{n} \sum_{j=1}^{n} (Y_{ij} - \Theta_{ij})^2}.$$

The proportion of correct edges changed by the model is shown in Figure 2(b) against the same simulation parameters as in (a). We see that as the proportion of correct edges increases to 1, the proportion of correct edges changed decreases. Note that the proportion of correct edges changed is also generally lower than the proportion of incorrect edges fixed.

We further examine this by calculating the ratio of the number of incorrect edges fixed to the number of correct edges changed for each posterior draw of each simulation. This ratio is

$$\phi^{(l)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} ((Y_{ij} - \Theta_{ij})^2 (1 - (\theta_{ij}^{(l)} - \Theta_{ij})^2))}{\sum_{i=1}^{n} \sum_{j=1}^{n} ((1 - (Y_{ij} - \Theta_{ij})^2)(\theta_{ij}^{(l)} - \Theta_{ij})^2)},$$

where values greater than 1 indicate that the model is fixing more incorrect edges than changing correct ones. $\phi^{(l)}$ are shown in Figure 3, compared with the propor-

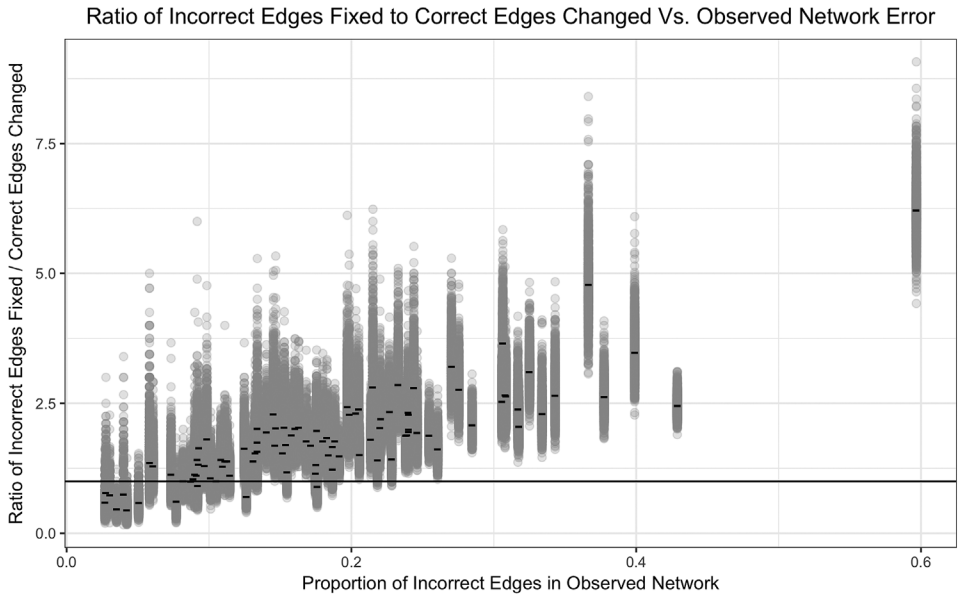FIG. 3. *Comparison of the ratio of the number of incorrect edges fixed to the number of correct edges changed, $\phi^{(l)}$ for 1000 posterior draws for each of 100 simulations. Points above the horizontal line $y = 1$ indicate that more incorrect edges are being fixed than correct edges changed for a given modeled network. The average of the posterior draws, $\bar{\phi}$, for each simulation is shown as a short black line.*

tion of incorrect ties in the observed network. The horizontal line $y = 1$ is added for reference. 85.34% of the 10,000 $\phi^{(l)}$ values are above 1, indicating that the model tends to fix more incorrect ties than it changes correct ones. Additionally, $\phi^{(l)}$ tends to increase as the proportion of incorrect edges in the observed network increases.

These simulations demonstrate that the proposed model is a useful tool when there is reason to believe the observed network contains errors. When the observed network is nearly error-free, the model does not provide much benefit and may even increase error. However, the model can substantially reduce error in observed networks with a larger proportion of incorrect edges. Even though the latent network is not observed in practice, it is still possible to identify observed networks that likely have many incorrect edges. For example, the latent San Marcos collaboration network, which we turn to in the next section, is undirected but we observe many asymmetric ties. We therefore have reason to believe the observed network contains many incorrect edges as either $Y_{ij} \neq \Theta_{ij}$ or $Y_{ji} \neq \Theta_{ji}$ for each asymmetric tie.

5. Analysis of modeled San Marcos networks. We apply the latent network model to the networks of collaboration on education-related projects among professors. We used the following hyperparameters to specify respondent accuracy: $\alpha_{p,k} = \alpha_{q,k} = 2$, $\beta_{p,k} = \beta_{q,k} = 8$, and $d_k = 0.15$ for all k. We chose these values as parameters for the prior false negative and false positive rates because they are centered at a reasonable value (0.2) and have enough variance to allow for extreme cases to occur. Plots for the beta distributions formed with these priors can be seen in Figure 4. Similarly, the prior used for the density is a reasonable value of the network density based on these and similar observed networks. Similar results are achieved by using other plausible values for the priors.

For the sampler, we used a burn-in of 1000 iterations and a thinning interval of 25. In total, we made 10,000 draws from the posterior distribution. Some model-fit diagnostics for these draws are provided below, followed by results.
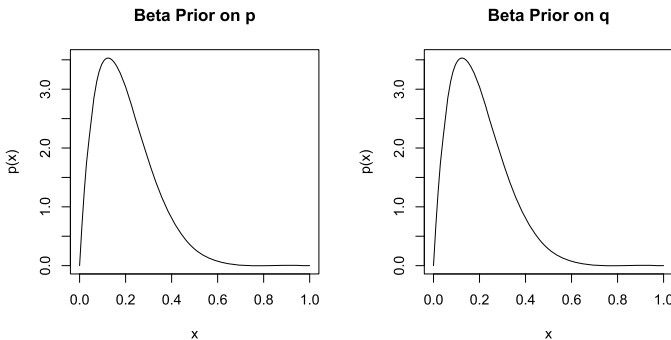

FIG. 4. *Prior distributions of $p_{i,k}$ and $q_{i,k}$. The mean for each distribution is at* 0.2.

TABLE 2
Average percent of ties by dyad type in 10,000 *posterior draws, by year. For example, in Year* 1*, on average* 2960 *of* 10,000 *posterior draws among dyads with an observed asymmetric tie had a mutual tie in that simulated latent network. The latent draws only contain mutual ties by design*

	Mutual	**Asymmetric**	**Null**
1	0.913	0.296	0.006
2	0.908	0.274	0.006
3	0.903	0.263	0.006
4	0.901	0.263	0.006
5	0.903	0.282	0.007
6	0.870	0.235	0.008
7	0.853	0.205	0.007
8	0.826	0.185	0.008
9	0.785	0.177	0.009
10	0.786	0.171	0.009

5.1. *Dyad census for latent collaboration.* To examine the model parameters, we consult draws from the posterior distribution provided by the sampler. First we investigate the ties present in each sample. Recall that by design all ties in the latent network models are mutual, while many dyads in the observed networks were asymmetric. We can classify the observed dyads at each time point into the mutually exclusive categories: mutual, asymmetric, and null. Then, again at each time point, we can aggregate over $L = 10,000$ draws from the posterior distribution for the latent network and all dyads of each type to find the proportion of those dyads that have a mutual tie (are realized) in the latent network. These proportions are shown in Table 2. Thus, at Year 1, among all dyads that were asymmetric in the observed network, we observe a mutual tie in the latent network 29.6% of the time. We classify these as *weaker* ties, and thus they are not picked up as frequently in the posterior draws. By comparison, for the same year among mutual ties, we observe a mutual tie in the latent network 91.3% of the time, indicating we have a *stronger* tie. The proportion of null ties that are modeled as mutual ties in the latent network is very low (0.6% for that year), a satisfying result.

5.2. *Visual comparisons of observed and latent networks.* This same phenomenon, where we observe mutual ties more strongly than asymmetric ties, is illustrated in Figures 5 and 6. Figure 5 shows the observed and latent networks during Year 5, the last year before the inception of the Q&C grant, while Figure 6 depicts Year 10, the most recent year on record. In each figure, the left panel shows the observed collaboration network at that time, with the nodes colored by department. Thicker lines represent mutual ties, while thinner lines show asym-
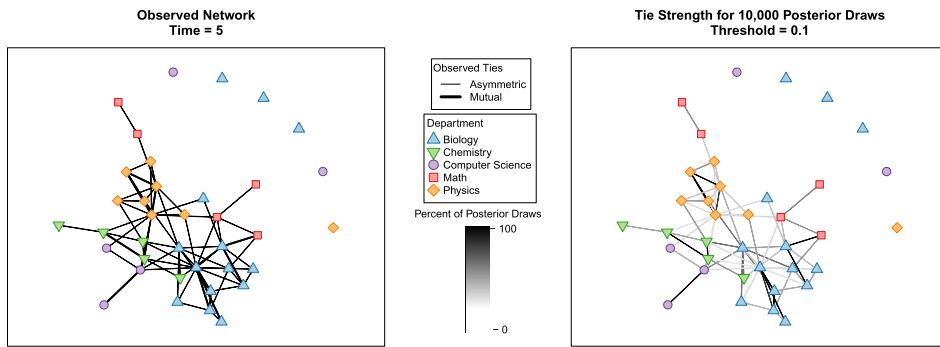
FIG. 5. *Observed network at time k = 5 (2007/2008), the year prior to the Q&C grant. The left panel shows the observed network with nodes colored by department, and thicker ties representing mutual collaborative relationships. The right panel shows the aggregated results of 10,000 posterior draws, with the shade of the tie representing the proportion of times it was present in a posterior draw. Darker shades represent more common ties. All ties were mutual by design.*

metric ties.[5] The right panel shows the aggregated results of 10,000 draws from the posterior distribution of the Bayesian hierarchical latent network model. The tie strength in this graph is indicated by the grayscale shade, with lines closer to black representing ties that are realized close to 100% of the time, while lines that are closer to white are hardly realized. We set a threshold of 0.1 for this graph, that is, only ties that are realized in at least 1000 of the 10,000 posterior draws are depicted. The ties that are darker in the latent graph tend to coincide with those that
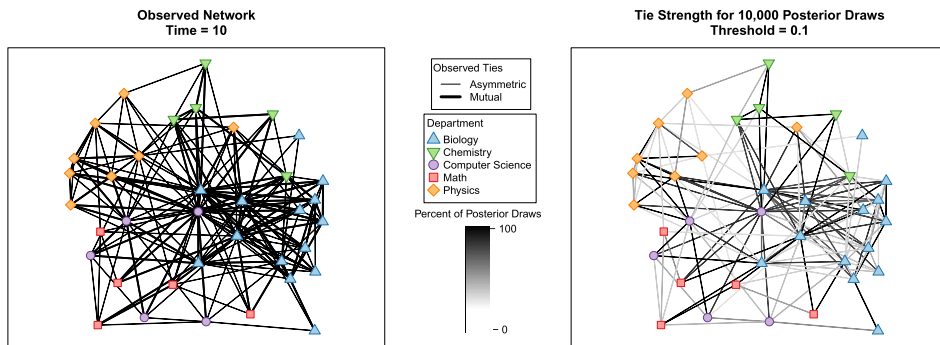
FIG. 6. *Observed network at time k = 10 (2012/2013), the year the survey was conducted. The left panel shows the observed network with nodes colored by department, and thicker ties representing mutual collaborative relationships. The right panel shows the aggregated results of 10,000 posterior draws, with the shade of the tie representing the proportion of times it was present in a posterior draw. Darker shades represent more common ties. All ties were mutual by design.*

[5]Note that the direction of the tie is not indicated in this figure.

are thicker (mutual) in the observed graph. The nodes are given the same coordinates in both panels to make differences between the two graphs more apparent.

Note also in Figure 5, the earlier set of graphs, the relative clustering within department. Not only are there comparatively higher rates of ties within departments than between departments, many of the observed mutual ties exist between members of the same department. By the latter time point, shown in Figure 6, we see a more equal distribution of ties within and between departments.

Also note the increase in density from Year 5 to Year 10 as well as the incorporation of all nodes into the network. This change is also picked up in the latent network.

Using these metrics we can state that the latent networks we modeled seem to be good representations of the observed networks. We found similar counts and locations of strong and weak ties, and did not observe artifacts that could be induced by too many null ties. Visually, the networks have similarly placed tie locations and exhibit similar departmental clustering.

5.3. *Analysis of reporting accuracy.* We can also examine the values of $p_{i,k}$ and $q_{i,k}$, respectively the false negative and false positive rates for each individual at each time. These are summarized in Appendix B, in Tables 4 and 5, where the rows are the nodes and the columns are the year intervals. The value given is the average value of $p_{i,k}$ or $q_{i,k}$ over 10,000 posterior draws from the Bayesian hierarchical latent network model, with the standard deviation shown parenthetically.

Let $p_{i,k} = \frac{1}{L} \sum_{l=1}^{L} p_{i,k}^{(l)}$ and $q_{i,k} = \frac{1}{L} \sum_{l=1}^{L} q_{i,k}^{(l)}$, where $L = 10,000$. Observe that in general the values for the $q_{i,k}$ are lower than those for the $p_{i,k}$, that is, the false positive rate is generally lower than the false negative rate. This is understandable given that recall inaccuracy tends toward omission, that is, people are more likely to forget a tie than they are to cite one that never existed. The latter situation we would consider an "error" in the truer sense of the word, while the former could be the result of differences in cognition, perception, or memory, which are not necessarily a mistake. (Perhaps, if someone cannot remember a collaborative relationship, it should not be termed as such.)

These values of $p_{i,k}$ and $q_{i,k}$ can also be seen graphically in Figures 7 and 8. Each connected line represents a single person, with the color designating department. Therefore, there are 37 lines on each plot. Use of the same axes allows us to compare the false negative and false positive rates. Most of the lines for most of the years show a lower error rate for q than for p.

Recall that we used the same priors for p_k and q_k, so the fact that we observe a lower false positive than false negative rate is indicative of the truth captured by the data. In this case, our model has picked up a subtlety in data collection that would not be captured by descriptive methods or existing models used for intervention assessment.

FIG. 7. *Plot of average values of p (false negative rate) per person over time, colored by department. Averages are from* 10,000 *posterior draws. Standard deviations provided in Table* 4.

5.4. *Assessment of the Q&C grant.* Now that we have a model which accounts for error-prone responses, we can examine the effect of the intervention (in this case, the Q&C grant) on the network to check for evidence of increased collaboration during the period of the grant.
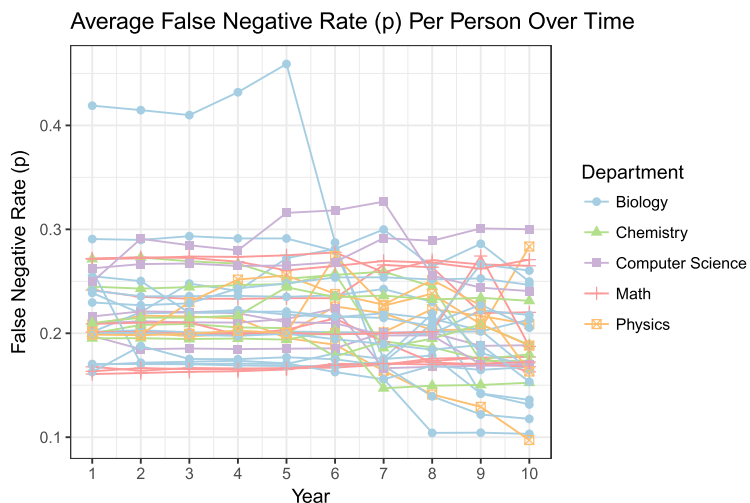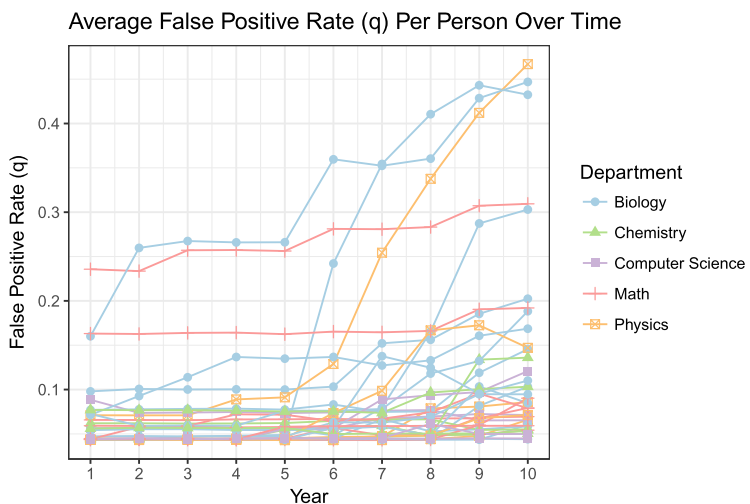

FIG. 8. *Plot of average values of q (false positive rate) per person over time, colored by department. Averages are from* 10,000 *posterior draws. Standard deviations provided in Table* 5.

We compare the five years before the beginning of the grant to the five years during it. A strict measure of network density (i.e., a count of the number of ties that exist divided by the total number of possible ties that could be formed at a given time) is naïve for several reasons. First, it does not allow for comparisons across time as different numbers of people were in the network at each time. Second, it does not account for the differential exposure to the social space. We would naturally expect professors who were hired one year ago to have fewer ties than those who were hired 20 years ago. Density alone as a metric for longitudinal changes in collaboration simply cannot capture this phenomenon.

Instead of density, we use hire-adjusted out-degree.[6] Adjustments are made as follows: for each of the posterior draws at each time, we calculate the proportion of possible out-ties each person forms as

$$(17) \qquad \rho_{i,k}^{(l)} = \frac{\sum_{j=1}^{37} \Theta_{ij,k}^{(l)}}{h_k - 1},$$

where h_k is the number of people who have been hired at time k. Let $\tau_{i,k}$ be the tenure-year of each person during each year, that is, the number of years they have been employed at the university during time k. Thus, if a professor was hired in 2000, their tenure-year in 2002 is 2. By definition, $\tau_{i,k+1} = \tau_{i,k} + 1$ for each i. Using tenure-year adjusts for the fact that people naturally form ties over time. Without adjustment, we would obviously expect more ties to be present at $k = 10$ than $k = 5$, independent of the Q&C grant. Table 3 shows the average proportion of possible out-ties ($\bar{\rho}$) formed for each hire year τ, split by whether the year occurred in the five years preceding the grant or the 5 years during it. In the 5 years preceding the grant, $n = 5$ people were in their first hire year, and on average over $l = 10,000$ posterior draws for $\Theta^{(l)}$ these people formed 0.2% of possible relationships. In the five years during the grant, $n = 4$ people were in their first hire year, and formed 0.24% of possible relationships. The larger $\bar{\rho}$ for each τ is shown in bold, with most occurring during the Q&C grant.

In another comparison, for each $l \in \{1, \dots, L\}$, $L = 10,000$, we fit two linear models. For each, use the tenure-year $\tau_{i,k}$ as the predictor and the proportion of possible out-ties $\rho_{i,k}^{(l)}$ as the response. For the first model, aggregate over years 1–5 (before the grant) and for the second model, aggregate over years 6–10 (during the grant). For both models, we constrain the line to pass through the origin since an individual's proportion of possible out-ties is zero when they have not yet been

---

[6]Note that for the posterior draws out-degree, in-degree, and degree are interchangeable up to a constant since all ties are mutual.

TABLE 3
Average proportion of possible out-ties ($\bar{\rho}$) by hire year (τ), with larger value in bold. In the five years preceding the grant, $n = 5$ people were in their first hire year, and on average over $l = 10{,}000$ posterior draws for $\Theta^{(l)}$ these people formed 0.2% of possible relationships. In the five years during the grant, $n = 4$ people were in their first hire year, and formed 0.24% of possible relationships

	Before Grant		During Grant	
Hire Year (τ)	n	$\bar{\rho}$	n	$\bar{\rho}$
1	5	0.0020	4	**0.0024**
2	7	**0.0556**	6	0.0484
3	9	0.0629	6	**0.0691**
4	12	**0.0660**	3	0.0632
5	11	0.0669	5	**0.0717**
6	11	0.0661	5	**0.0998**
7	9	0.0508	7	**0.0926**
8	8	0.0392	9	**0.0922**
9	6	0.0346	12	**0.1067**
10	6	0.0385	11	**0.1157**
11	9	0.0500	11	**0.1280**
12	9	0.0542	9	**0.1028**
13	13	0.0832	8	**0.0994**
14	12	**0.0846**	6	0.0649
15	11	0.0897	6	**0.0999**
16	8	0.0911	9	**0.1086**
17	6	0.0966	9	**0.1227**
18	0	–	13	0.1295
19	0	–	12	0.1295
20	0	–	11	0.1295
21	0	–	8	0.1022
22	0	–	6	0.1083

hired by the university. Denote the slope of the first model by $m_B^{(l)}$ and the slope of the second by $m_A^{(l)}$.

For each l, we take the ratio of the slopes

(18)
$$\xi^{(l)} = \frac{m_A^{(l)}}{m_B^{(l)}}.$$

For the purpose of visualization, we fit a density curve through these values using a local likelihood method [Loader (2013)]. The plot of this density, along with basic summary statistics, is found in Figure 9. The ratios should be compared with the value 1, the case in which the slopes before and during the grant were identical, that
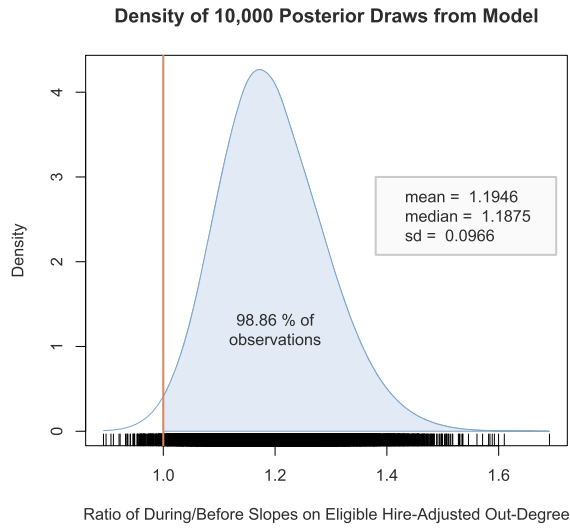
Density of 10,000 Posterior Draws from Model

FIG. 9.   *Density of* 10,000 *ratios of the slope during the grant divided by the slope before the grant. Each slope is calculated from a regression of proportion of possible out-ties formed on tenure-year. Summary statistics are given.*

is, there was no change in tie formation between these two time periods. Values less than one represent a greater propensity to form ties before the grant, while values greater than one represent a greater propensity to form ties during the grant. If the grant is having a positive effect on collaboration, we would expect to see many more values greater than one than less than it. Indeed this is the case, with only 114 of 10,000 cases (or 1.14%) less than one.[7]

Even after accounting for error-prone responses in the survey data and differential rates of tie-formation depending on tenure-year, we observe a clear increase in collaboration on education-related projects during the Q&C grant implemented at CSU, San Marcos. This appears to be a validation of the methods implemented by the Steering Committee.

Several limitations of this study and these analyses should be noted. First, our model focuses on adjusting for error-prone responses, but is not a true dynamic model. Each time point is modeled separately, and the collection of modeled networks is then analyzed. This was necessitated by the retroactive data collection, which also did not allow us to observe "active" collaborations. We could only collect data on whether individuals had ever collaborated, not if

---

[7]Note that although the use of a *z*-test in this case would violate standard assumptions, the density appears approximately normal and we can reject the null hypothesis that the true mean of the ratio is one at the $\alpha = 0.05$ significance level.

they were currently collaborating at a particular point in time. This means we were not able to study tie dissolution, and networks grow naturally over time as more ties form. Ideally, data would have been collected throughout the time period of interest and dynamic models could be used. The metrics $\bar{\rho}$ and $\xi^{(l)}$ used to compare collaboration rates during the Q&C grant to those before it are ad hoc. Both attempt to account for naturally increasing network size over time as more people are hired and differential rates of tie formation based on how long someone had been at the university, but are in no way definitive measures.

6. Conclusion. Social network surveys are prone to many types of errors, especially when retroactively collecting longitudinal data. Subjects exhibit differential recall accuracy as well as varying thresholds for the perception of ties. To address these issues, we constructed a latent network model and developed inference for it within a Bayesian framework. We applied this method to a network of collaboration on education-related activities among professors from five science departments at a small teaching university. We found that while draws from the simulated posterior distribution of the false positive rate (reporting a collaborative tie that does not exist in reality) remain low, the false negative rate (omitting a collaborative tie that does in reality exist) can be larger. Additionally, we find a conclusive increase in educational collaboration over time after adjusting for the length of time each professor had to form ties at the university. While we do not advocate a retroactive survey over a truly longitudinal one, we provide a framework for the assessment of an intervention on collaboration networks when the latter is unavailable. The algorithm was implemented in the R programming language, using some tools in the statnet package [Handcock et al. (2016)]. The code is available from the first author's GitHub in the CollabAssess R package at https://github.com/krmclaughlin/CollabAssess.

APPENDIX A: DERIVATIONS

The derivation of the posterior distribution in Section 3.3. Note that the subscript k is omitted from each term for brevity.

$$P(\Theta, p, q \mid Y) \propto P(Y \mid \Theta, p, q) P(\Theta, p, q \mid \delta, \alpha_p, \beta_p, \alpha_q, \beta_q)$$

$$\propto P(Y \mid \Theta, p, q) P(p \mid \alpha_p, \beta_p) P(q \mid \alpha_q, \beta_q) P(\Theta \mid \delta)$$

$$\propto \prod_{i=1}^{n} \left[(1 - p_i)^{\sum_{j=1}^{n} y_{ij}\theta_{ij}} p_i^{\sum_{j=1}^{n}(1-y_{ij})\theta_{ij}} \right.$$

$$\left. \times q_i^{\sum_{j=1}^{n} y_{ij}(1-\theta_{ij})} (1-q_i)^{\sum_{j=1}^{n}(1-y_{ij})(1-\theta_{ij})} \right]$$

$$\times \prod_{i=1}^{n}[p_i^{\alpha_p-1}(1-p_i)^{\beta_p-1}]\prod_{i=1}^{n}[q_i^{\alpha_q-1}(1-q_i)^{\beta_q-1}]$$

$$\times \prod_{i=1}^{n}\prod_{j=1}^{n}[\delta_{ij}^{\theta_{ij}}(1-\delta_{ij})^{(1-\theta_{ij})}]$$

$$= \prod_{i=1}^{n}[(p_i)^{\alpha_p+\sum_{j=1}^{n}(1-y_{ij})\theta_{ij}-1}(1-p_i)^{\beta_p+\sum_{j=1}^{n}y_{ij}\theta_{ij}-1}]$$

$$\times \prod_{i=1}^{n}[(q_i)^{\alpha_q+\sum_{j=1}^{n}y_{ij}(1-\theta_{ij})-1}(1-q_i)^{\beta_q+\sum_{j=1}^{n}(1-y_{ij})(1-\theta_{ij})-1}]$$

$$\times \prod_{i=1}^{n}\prod_{j=1}^{n}[(\delta_{ij})^{\theta_{ij}}(1-\delta_{ij})^{(1-\theta_{ij})}].$$

Using this posterior distribution, we can derive the full conditional distributions for p, q, and Θ. For p, collecting only the terms that contain p, we have the form of a beta distribution.

$$P(p \mid \Theta, q, Y) \propto P(Y \mid \Theta, p, q)P(p \mid \alpha_p, \beta_p)$$

$$\propto \prod_{i=1}^{n}[(1-p_i)^{\sum_{j=1}^{n}y_{ij}\theta_{ij}}p_i^{\sum_{j=1}^{n}(1-y_{ij})\theta_{ij}}$$

$$\times q_i^{\sum_{j=1}^{n}y_{ij}(1-\theta_{ij})}(1-q_i)^{\sum_{j=1}^{n}(1-y_{ij})(1-\theta_{ij})}]$$

$$\times \prod_{i=1}^{n}[p_i^{\alpha_p-1}(1-p_i)^{\beta_p-1}]$$

$$\propto \prod_{i=1}^{n}(p_i)^{\alpha_p+\sum_{j=1}^{n}(1-y_{ij})\theta_{ij}-1}(1-p_i)^{\beta_p+\sum_{j=1}^{n}y_{ij}\theta_{ij}-1},$$

$$[p_i \mid \Theta, Y] \sim \text{Beta}\left(\alpha_p + \sum_{j=1}^{n}(1-y_{ij})\theta_{ij}, \beta_p + \sum_{j=1}^{n}y_{ij}\theta_{ij}\right).$$

Similarly for q,

$$P(q \mid \Theta, p, Y) \propto P(Y \mid \Theta, p, q)P(q \mid \alpha_q, \beta_q)$$

$$\propto \prod_{i=1}^{n}(q_i)^{\alpha_q+\sum_{j=1}^{n}y_{ij}(1-\theta_{ij})-1}(1-q_i)^{\beta_q+\sum_{j=1}^{n}(1-y_{ij})(1-\theta_{ij})-1},$$

$$[q_i \mid \Theta, Y] \sim \text{Beta}\left(\alpha_q + \sum_{j=1}^{n}y_{ij}(1-\theta_{ij}), \beta_q + \sum_{j=1}^{n}(1-y_{ij})(1-\theta_{ij})\right).$$

For Θ we impose undirected ties,

$$P(\Theta \mid p, q, Y)$$

$$\propto P(Y \mid \Theta, p, q) P(\Theta \mid \delta)$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{n} p_i^{(1-y_{ij})\theta_{ij}} (1-p_i)^{y_{ij}\theta_{ij}} q_i^{y_{ij}(1-\theta_{ij})} (1-q_i)^{(1-y_{ij})(1-\theta_{ij})}$$

$$\times \delta_{ij}^{\theta_{ij}} (1-\delta_{ij})^{(1-\theta_{ij})}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{n} \left[\delta_{ij} p_i^{(1-y_{ij})} (1-p_i)^{y_{ij}}\right]^{\theta_{ij}} \left[(1-\delta_{ij}) q_i^{(1-y_{ij})} (1-q_i)^{(1-y_{ij})}\right]^{1-\theta_{ij}}$$

$$\propto \prod_{i=1}^{j-1} \prod_{j=2}^{n} \left[\delta_{ij} p_i^{(1-y_{ij})} (1-p_i)^{y_{ij}} p_j^{(1-y_{ji})} (1-p_j)^{y_{ji}}\right]^{\theta_{ij}}$$

$$\times \left[(1-\delta_{ij}) q_i^{(1-y_{ij})} (1-q_i)^{(1-y_{ij})} q_j^{(1-y_{ji})} (1-q_j)^{(1-y_{ji})}\right]^{1-\theta_{ij}}.$$

This is a Bernoulli mixture with

$$P(\Theta_{ij} = 1 \mid Y, p, q) = \delta_{ij} p_i^{(1-y_{ij})} (1-p_i)^{y_{ij}} p_j^{(1-y_{ji})} (1-p_j)^{y_{ji}}$$

and

$$P(\Theta_{ij} = 0 \mid Y, p, q) = (1-\delta_{ij}) q_i^{(1-y_{ij})} (1-q_i)^{(1-y_{ij})} q_j^{(1-y_{ji})} (1-q_j)^{(1-y_{ji})}.$$

APPENDIX B: TABLES

The mean of the 10,000 posterior draws for each $p_{i,k}$ is provided in Table 4, with the standard deviations shown parenthetically. The mean of the 10,000 posterior draws for each $q_{i,k}$ is provided in Table 5, with the standard deviations shown parenthetically.

TABLE 4

Mean of 10,000 *posterior draws for p (false negative) for each person with standard deviations at each time, where $p_i, k \sim \text{Beta}(\alpha_p = 2, \beta_p = 8)$. A plot of these values is provided in Figure* 7

	1	2	3	4	5	6	7	8	9	10
1	0.42 (0.14)	0.41 (0.15)	0.41 (0.15)	0.43 (0.15)	0.46 (0.14)	0.29 (0.12)	0.16 (0.09)	0.10 (0.07)	0.10 (0.07)	0.10 (0.07)
2	0.17 (0.11)	0.17 (0.11)	0.17 (0.11)	0.17 (0.10)	0.17 (0.11)	0.18 (0.11)	0.18 (0.11)	0.18 (0.11)	0.14 (0.09)	0.14 (0.09)
3	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.23 (0.13)	0.22 (0.13)	0.20 (0.12)	0.20 (0.12)	0.28 (0.14)
4	0.17 (0.11)	0.16 (0.10)	0.17 (0.10)	0.17 (0.10)	0.17 (0.10)	0.17 (0.10)	0.17 (0.11)	0.17 (0.11)	0.18 (0.11)	0.18 (0.11)
5	0.21 (0.12)	0.22 (0.13)	0.22 (0.13)	0.21 (0.12)	0.20 (0.12)	0.19 (0.11)	0.16 (0.10)	0.14 (0.09)	0.13 (0.08)	0.10 (0.06)
6	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.22 (0.13)	0.22 (0.13)	0.21 (0.12)
7	0.20 (0.12)	0.22 (0.13)	0.25 (0.14)	0.24 (0.14)	0.27 (0.14)	0.28 (0.14)	0.30 (0.14)	0.26 (0.13)	0.29 (0.14)	0.25 (0.13)
8	0.29 (0.14)	0.29 (0.14)	0.29 (0.14)	0.29 (0.14)	0.29 (0.14)	0.28 (0.14)	0.18 (0.10)	0.22 (0.11)	0.14 (0.09)	0.13 (0.08)
9	0.21 (0.12)	0.21 (0.12)	0.21 (0.13)	0.21 (0.12)	0.21 (0.13)	0.22 (0.13)	0.17 (0.10)	0.17 (0.10)	0.17 (0.10)	0.17 (0.10)
10	0.20 (0.12)	0.18 (0.11)	0.19 (0.11)	0.18 (0.11)	0.19 (0.11)	0.19 (0.11)	0.20 (0.12)	0.20 (0.12)	0.19 (0.11)	0.19 (0.11)
11	0.24 (0.13)	0.24 (0.13)	0.23 (0.13)	0.23 (0.13)	0.23 (0.13)	0.23 (0.13)	0.27 (0.14)	0.26 (0.14)	0.22 (0.12)	0.22 (0.12)
12	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)	0.17 (0.10)	0.17 (0.10)	0.17 (0.10)	0.18 (0.11)	0.18 (0.11)	0.16 (0.10)
13	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.21 (0.12)
14	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.22 (0.13)	0.22 (0.13)
15	0.20 (0.11)	0.21 (0.12)	0.21 (0.12)	0.21 (0.12)	0.20 (0.12)	0.18 (0.11)	0.19 (0.11)	0.19 (0.11)	0.18 (0.11)	0.18 (0.11)
16	0.17 (0.10)	0.17 (0.11)	0.17 (0.11)	0.17 (0.11)	0.17 (0.11)	0.16 (0.10)	0.16 (0.10)	0.17 (0.10)	0.16 (0.10)	0.17 (0.10)
17	0.16 (0.10)	0.19 (0.11)	0.18 (0.10)	0.18 (0.10)	0.18 (0.11)	0.17 (0.10)	0.17 (0.10)	0.14 (0.09)	0.12 (0.08)	0.12 (0.08)
18	0.21 (0.12)	0.21 (0.12)	0.21 (0.12)	0.20 (0.12)	0.20 (0.12)	0.21 (0.12)	0.19 (0.11)	0.17 (0.10)	0.17 (0.10)	0.17 (0.10)
19	0.27 (0.14)	0.27 (0.14)	0.27 (0.14)	0.27 (0.14)	0.25 (0.13)	0.26 (0.14)	0.26 (0.14)	0.24 (0.13)	0.17 (0.11)	0.17 (0.10)
20	0.24 (0.13)	0.22 (0.12)	0.22 (0.12)	0.22 (0.12)	0.22 (0.12)	0.22 (0.12)	0.22 (0.12)	0.21 (0.12)	0.20 (0.11)	0.16 (0.10)
21	0.27 (0.14)	0.27 (0.14)	0.27 (0.14)	0.27 (0.14)	0.27 (0.14)	0.28 (0.14)	0.26 (0.13)	0.27 (0.14)	0.27 (0.14)	0.26 (0.14)
22	0.24 (0.13)	0.24 (0.13)	0.24 (0.14)	0.25 (0.14)	0.25 (0.14)	0.26 (0.14)	0.19 (0.11)	0.19 (0.12)	0.21 (0.12)	0.19 (0.11)
23	0.24 (0.13)	0.24 (0.13)	0.24 (0.13)	0.24 (0.13)	0.24 (0.13)	0.24 (0.13)	0.24 (0.14)	0.23 (0.13)	0.18 (0.11)	0.17 (0.10)
24	0.22 (0.13)	0.22 (0.13)	0.22 (0.13)	0.22 (0.13)	0.21 (0.12)	0.21 (0.12)	0.20 (0.12)	0.20 (0.12)	0.17 (0.11)	0.17 (0.11)
25	0.20 (0.12)	0.20 (0.12)	0.23 (0.13)	0.25 (0.14)	0.26 (0.14)	0.24 (0.13)	0.23 (0.13)	0.25 (0.13)	0.23 (0.12)	0.16 (0.10)
26	0.26 (0.14)	0.17 (0.11)	0.17 (0.11)	0.17 (0.10)	0.17 (0.11)	0.17 (0.10)	0.17 (0.11)	0.21 (0.12)	0.23 (0.12)	0.21 (0.12)
27	0.16 (0.10)	0.17 (0.10)	0.17 (0.10)	0.17 (0.10)	0.17 (0.10)	0.17 (0.11)	0.17 (0.11)	0.17 (0.10)	0.17 (0.11)	0.17 (0.11)
28	0.20 (0.12)	0.20 (0.12)	0.19 (0.12)	0.19 (0.12)	0.19 (0.12)	0.20 (0.12)	0.15 (0.09)	0.15 (0.09)	0.15 (0.09)	0.15 (0.09)
29	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.27 (0.14)	0.19 (0.11)
30	0.23 (0.13)	0.23 (0.13)	0.23 (0.13)	0.24 (0.14)	0.25 (0.14)	0.25 (0.14)	0.25 (0.14)	0.25 (0.14)	0.25 (0.14)	0.25 (0.14)
31	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.19 (0.12)	0.19 (0.12)	0.18 (0.11)	0.19 (0.11)	0.15 (0.09)
32	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.20 (0.12)	0.27 (0.14)	0.23 (0.13)	0.24 (0.13)	0.21 (0.12)	0.19 (0.11)
33	0.25 (0.14)	0.25 (0.14)	0.22 (0.12)	0.22 (0.13)	0.22 (0.12)	0.22 (0.12)	0.21 (0.12)	0.21 (0.12)	0.27 (0.14)	0.26 (0.14)
34	0.21 (0.12)	0.22 (0.12)	0.22 (0.12)	0.22 (0.12)	0.24 (0.13)	0.23 (0.12)	0.24 (0.13)	0.23 (0.13)	0.23 (0.12)	0.23 (0.13)
35	0.27 (0.14)	0.27 (0.14)	0.27 (0.14)	0.27 (0.14)	0.26 (0.14)	0.26 (0.14)	0.27 (0.14)	0.27 (0.14)	0.26 (0.14)	0.27 (0.14)
36	0.26 (0.14)	0.27 (0.14)	0.27 (0.14)	0.26 (0.14)	0.27 (0.14)	0.27 (0.14)	0.29 (0.14)	0.29 (0.14)	0.30 (0.14)	0.30 (0.15)
37	0.25 (0.14)	0.29 (0.15)	0.28 (0.14)	0.28 (0.14)	0.32 (0.15)	0.32 (0.15)	0.33 (0.15)	0.26 (0.13)	0.24 (0.13)	0.24 (0.13)

TABLE 5

Mean of 10,000 *posterior draws for q* (*false positive*) *for each person with standard deviations at each time, where* $q_i, k \sim \text{Beta}(\alpha_q = 2, \beta_q = 8)$. *A plot of these values is provided in Figure* 8

	1	2	3	4	5	6	7	8	9	10
1	0.05 (0.14)	0.05 (0.15)	0.05 (0.15)	0.05 (0.15)	0.05 (0.14)	0.24 (0.12)	0.35 (0.09)	0.41 (0.07)	0.44 (0.07)	0.43 (0.07)
2	0.10 (0.11)	0.10 (0.11)	0.10 (0.11)	0.10 (0.10)	0.10 (0.11)	0.10 (0.11)	0.15 (0.11)	0.16 (0.11)	0.19 (0.09)	0.20 (0.09)
3	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.05 (0.13)	0.05 (0.13)	0.08 (0.12)	0.08 (0.12)	0.09 (0.14)
4	0.24 (0.11)	0.23 (0.10)	0.26 (0.10)	0.26 (0.10)	0.26 (0.10)	0.28 (0.10)	0.28 (0.11)	0.28 (0.11)	0.31 (0.11)	0.31 (0.11)
5	0.07 (0.12)	0.07 (0.13)	0.07 (0.13)	0.09 (0.12)	0.09 (0.12)	0.13 (0.11)	0.25 (0.10)	0.34 (0.09)	0.41 (0.08)	0.47 (0.06)
6	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.13)	0.05 (0.13)	0.07 (0.12)
7	0.04 (0.12)	0.04 (0.13)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.07 (0.14)	0.05 (0.13)	0.06 (0.14)	0.06 (0.13)
8	0.06 (0.14)	0.06 (0.14)	0.06 (0.14)	0.06 (0.14)	0.06 (0.14)	0.05 (0.14)	0.08 (0.10)	0.17 (0.11)	0.29 (0.09)	0.30 (0.08)
9	0.06 (0.12)	0.06 (0.12)	0.06 (0.13)	0.06 (0.12)	0.06 (0.13)	0.06 (0.13)	0.07 (0.10)	0.07 (0.10)	0.07 (0.10)	0.07 (0.10)
10	0.09 (0.12)	0.07 (0.11)	0.07 (0.11)	0.07 (0.11)	0.07 (0.11)	0.07 (0.11)	0.08 (0.12)	0.08 (0.12)	0.07 (0.11)	0.07 (0.11)
11	0.04 (0.12)	0.06 (0.13)	0.06 (0.13)	0.06 (0.13)	0.06 (0.13)	0.06 (0.13)	0.06 (0.14)	0.05 (0.14)	0.07 (0.12)	0.07 (0.12)
12	0.07 (0.10)	0.07 (0.10)	0.07 (0.10)	0.07 (0.10)	0.07 (0.10)	0.07 (0.10)	0.07 (0.10)	0.07 (0.11)	0.07 (0.11)	0.08 (0.10)
13	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.06 (0.12)
14	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.08 (0.13)	0.10 (0.13)
15	0.06 (0.11)	0.06 (0.12)	0.06 (0.12)	0.06 (0.12)	0.06 (0.12)	0.06 (0.11)	0.07 (0.11)	0.07 (0.11)	0.05 (0.11)	0.06 (0.11)
16	0.08 (0.10)	0.08 (0.11)	0.08 (0.11)	0.08 (0.11)	0.08 (0.11)	0.08 (0.10)	0.07 (0.10)	0.08 (0.10)	0.12 (0.10)	0.15 (0.10)
17	0.16 (0.10)	0.26 (0.11)	0.27 (0.10)	0.27 (0.10)	0.27 (0.11)	0.36 (0.10)	0.35 (0.10)	0.36 (0.09)	0.43 (0.08)	0.45 (0.08)
18	0.06 (0.12)	0.06 (0.12)	0.06 (0.12)	0.07 (0.12)	0.07 (0.12)	0.06 (0.12)	0.07 (0.11)	0.07 (0.10)	0.10 (0.10)	0.08 (0.10)
19	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.05 (0.13)	0.06 (0.14)	0.06 (0.14)	0.05 (0.13)	0.13 (0.11)	0.14 (0.10)
20	0.07 (0.13)	0.06 (0.12)	0.06 (0.12)	0.06 (0.12)	0.08 (0.12)	0.08 (0.12)	0.08 (0.12)	0.12 (0.12)	0.13 (0.11)	0.19 (0.10)
21	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.06 (0.14)	0.05 (0.13)	0.05 (0.14)	0.05 (0.14)	0.05 (0.14)
22	0.04 (0.13)	0.04 (0.13)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.05 (0.11)	0.05 (0.12)	0.05 (0.12)	0.05 (0.11)
23	0.05 (0.13)	0.06 (0.13)	0.06 (0.13)	0.05 (0.13)	0.05 (0.13)	0.06 (0.13)	0.06 (0.14)	0.05 (0.13)	0.10 (0.11)	0.08 (0.10)
24	0.04 (0.13)	0.04 (0.13)	0.04 (0.13)	0.04 (0.13)	0.06 (0.12)	0.06 (0.12)	0.09 (0.12)	0.09 (0.12)	0.10 (0.11)	0.12 (0.11)
25	0.04 (0.12)	0.04 (0.12)	0.04 (0.13)	0.04 (0.14)	0.04 (0.14)	0.07 (0.13)	0.10 (0.13)	0.17 (0.13)	0.17 (0.12)	0.15 (0.10)
26	0.07 (0.14)	0.09 (0.11)	0.11 (0.11)	0.14 (0.10)	0.13 (0.11)	0.14 (0.10)	0.13 (0.11)	0.13 (0.12)	0.16 (0.12)	0.17 (0.12)
27	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)	0.17 (0.11)	0.16 (0.11)	0.17 (0.10)	0.19 (0.11)	0.19 (0.11)
28	0.08 (0.12)	0.08 (0.12)	0.08 (0.12)	0.08 (0.12)	0.08 (0.12)	0.08 (0.12)	0.07 (0.09)	0.10 (0.09)	0.10 (0.09)	0.10 (0.09)
29	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.06 (0.14)	0.09 (0.11)
30	0.04 (0.13)	0.04 (0.13)	0.04 (0.13)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)
31	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.06 (0.12)	0.14 (0.12)	0.12 (0.11)	0.10 (0.11)	0.11 (0.09)
32	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.12)	0.04 (0.14)	0.05 (0.13)	0.05 (0.13)	0.07 (0.12)	0.07 (0.11)
33	0.04 (0.14)	0.04 (0.14)	0.05 (0.12)	0.05 (0.13)	0.05 (0.12)	0.06 (0.12)	0.06 (0.12)	0.07 (0.12)	0.09 (0.14)	0.09 (0.14)
34	0.06 (0.12)	0.06 (0.12)	0.06 (0.12)	0.06 (0.12)	0.06 (0.13)	0.05 (0.12)	0.05 (0.13)	0.05 (0.13)	0.05 (0.12)	0.05 (0.13)
35	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.06 (0.14)	0.06 (0.14)	0.06 (0.14)	0.06 (0.14)	0.06 (0.14)	0.06 (0.14)
36	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.04 (0.14)	0.05 (0.14)	0.04 (0.15)
37	0.04 (0.14)	0.05 (0.15)	0.04 (0.14)	0.04 (0.14)	0.05 (0.15)	0.04 (0.15)	0.05 (0.15)	0.06 (0.13)	0.05 (0.13)	0.05 (0.13)

REFERENCES

BORGATTI, S. P., CARLEY, K. M. and KRACKHARDT, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Soc. Netw.* **28** 124–136.

BUTTS, C. T. (2003). Network inference, error, and informant (in)accuracy: A Bayesian approach. *Soc. Netw.* **25** 103–140.

BUTTS, C. T., ACTON, R. M. and MARCUM, C. S. (2012). Interorganizational collaboration in the hurricane Katrina response. *J. Soc. Struct.* **13**.

CROSS, J. E., DICKMANN, E., NEWMAN-GONCHAR, R. and FAGAN, J. M. (2009). Using mixed-method design and network analysis to measure development of interagency collaboration. *Am J. Eval.* **30** 310–29.

EISENBERG, M. and SWANSON, N. (1996). Organizational network analysis as a tool for program evaluation. *Evaluation & the Health Professions* **19** 488–506.

FREY, B. B., LOHMEIER, J. H., LEE, S. W. and TOLLEFSON, N. (2006). Measuring collaboration among grant partners. *Am J. Eval.* **27** 383–392.

GAJDA, R. (2004). Utilizing collaboration theory to evaluate strategic alliances. *Am J. Eval.* **25** 65–77.

GAJDA, R. and KOLIBA, C. (2007). Evaluating the imperative of intraorganizational collaboration: A school improvement perspective. *Am J. Eval.* **28** 26–44.

HANDCOCK, M. S. and GILE, K. J. (2010). Modeling social networks from sampled data. *Ann. Appl. Stat.* **4** 5–25. MR2758082

HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M., KRIVITSKY, P. N., BENDER-DEMOLL, S. and MORRIS, M. (2016). statnet: Software tools for the statistical analysis of network data. The Statnet Project (http://www.statnet.org). R package version 2016.9.

KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 29–46. MR3153932

LOADER, C. (2013). locfit: Local regression, likelihood and density estimation. R package version 1.5-9.1.

MARCUM, C. S., BEVC, C. A. and BUTTS, C. T. (2012). Mechanisms of control in emergent interorganizational networks. *The Policy Studies Journal* **40** 516–546.

MARSDEN, P. V. (1990). Network data and measurement. *Annu. Rev. Sociol.* **16** 435–463.

MCCULLOUGH, J. (1992). Evaluating grants programs for basic scientific research: Difficult measurement problems, few techniques and slow progress. *Am J. Eval.* **13** 95–101.

SNIJDERS, T. A. B. (1996). Stochastic actor-oriented models for network change. *J. Math. Sociol.* **21** 149–172.

WANG, D. J., SHI, X., MCFARLAND, D. A. and LESKOVEC, J. (2012). Measurement error in network data: A re-classification. *Soc. Netw.* **34** 396–409.

WOODLAND, R. H. and HUTTON, M. S. (2012). Evaluating organizational collaborations: Suggested entry points and strategies. *Am J. Eval.* **33** 366–383.

DEPARTMENT OF STATISTICS
OREGON STATE UNIVERSITY
CORVALLIS, OREGON 97331
USA
E-MAIL: mclaugka@oregonstate.edu

RAND CORPORATION
SANTA MONICA, CALIFORNIA 90401
USA
E-MAIL: jembree@rand.org