Supplementary Materials for "Network Inference Using the Hub Model and Variants"

Zhibing He[1], Yunpeng Zhao[1,*], Peter Bickel[2], Charles Weko[3], Dan Cheng[1] and Jirui Wang[4]

[1]Arizona State University

[2]University of California, Berkeley

[3]U.S. Army

[4]Medpace

[*]Corresponding author. Email: Yunpeng.Zhao@asu.edu

1 Proofs in Section 2.2

Proof of Theorem 1. Let $(\tilde{\rho}, \tilde{A}) \in \mathcal{P}$ be a set of parameters such that $\mathbb{P}(g|\rho, A) = \mathbb{P}(g|\tilde{\rho}, \tilde{A})$ for all g. For all $i = 1, \ldots, n_L$, $k = n_L + 1, \ldots, n$, consider the probability that only i

appears under parameterizations (ρ, A) and $(\tilde{\rho}, \tilde{A})$, respectively

$$\tilde{\rho}_i(1-\tilde{A}_{ik}) \prod_{j=1,\ldots,n, j\neq i, j\neq k} (1-\tilde{A}_{ij}) = \rho_i(1-A_{ik}) \prod_{j=1,\ldots,n, j\neq i, j\neq k} (1-A_{ij}),$$

and the probability that only i and k appear

$$\tilde{\rho}_i\tilde{A}_{ik} \prod_{j=1,\ldots,n, j\neq i, j\neq k} (1-\tilde{A}_{ij}) = \rho_i A_{ik} \prod_{j=1,\ldots,n, j\neq i, j\neq k} (1-A_{ij}).$$

As $A_{ij} < 1$ in condition (i), dividing the second equation by the first, we obtain $\tilde{A}_{ik}/(1-\tilde{A}_{ik}) = A_{ik}/(1-A_{ik})$ and hence $\tilde{A}_{ik} = A_{ik}$ for $i = 1, \ldots, n_L$, $k = n_L + 1, \ldots, n$.

For any $i = 1, \ldots, n_L$, $i' = 1, \ldots, n_L$, $i \neq i'$, suppose that k is the follower such that $A_{ik} \neq A_{i'k}$. Consider the probability that only i and i' appear

$$\begin{aligned}
&\tilde{\rho}_i\tilde{A}_{ii'}(1-\tilde{A}_{ik}) \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1-\tilde{A}_{ij}) + \tilde{\rho}_{i'}\tilde{A}_{i'i}(1-\tilde{A}_{i'k}) \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1-\tilde{A}_{i'j}) \\
=&\rho_i A_{ii'}(1-A_{ik}) \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1-A_{ij}) + \rho_{i'} A_{i'i}(1-A_{i'k}) \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1-A_{i'j}),
\end{aligned}$$

and the probability that i, i' and k appear

$$\begin{aligned}
&\tilde{\rho}_i\tilde{A}_{ii'}\tilde{A}_{ik} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1-\tilde{A}_{ij}) + \tilde{\rho}_{i'}\tilde{A}_{i'i}\tilde{A}_{i'k} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1-\tilde{A}_{i'j}) \\
=&\rho_i A_{ii'} A_{ik} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1-A_{ij}) + \rho_{i'} A_{i'i} A_{i'k} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1-A_{i'j}).
\end{aligned}$$

As $\tilde{A}_{ik} = A_{ik}$ for $i = 1, \ldots, n_L$, $k = n_L + 1, \ldots, n$, the above two equations become

$$\begin{aligned} &\tilde{\rho}_i \tilde{A}_{ii'}(1 - A_{ik}) \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - \tilde{A}_{ij}) + \tilde{\rho}_{i'} \tilde{A}_{i'i}(1 - A_{i'k}) \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - \tilde{A}_{i'j}) \\ =&\rho_i A_{ii'}(1 - A_{ik}) \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - A_{ij}) + \rho_{i'} A_{i'i}(1 - A_{i'k}) \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - A_{i'j}), \end{aligned} \tag{1}$$

$$\begin{aligned} &\tilde{\rho}_i \tilde{A}_{ii'} A_{ik} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - \tilde{A}_{ij}) + \tilde{\rho}_{i'} \tilde{A}_{i'i} A_{i'k} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - \tilde{A}_{i'j}) \\ =&\rho_i A_{ii'} A_{ik} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - A_{ij}) + \rho_{i'} A_{i'i} A_{i'k} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - A_{i'j}). \end{aligned} \tag{2}$$

(1) and (2) can be viewed as a system of linear equations with unknown variables

$$\tilde{\rho}_i \tilde{A}_{ii'} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - \tilde{A}_{ij}),$$

and

$$\tilde{\rho}_{i'} \tilde{A}_{i'i} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - \tilde{A}_{i'j}).$$

By condition (ii), as $A_{ik} \neq A_{i'k}$, the system has full rank and hence has one and only one solution:

$$\begin{aligned} \tilde{\rho}_i \tilde{A}_{ii'} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - \tilde{A}_{ij}) &= \rho_i A_{ii'} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - A_{ij}), \\ \tilde{\rho}_{i'} \tilde{A}_{i'i} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - \tilde{A}_{ij}) &= \rho_{i'} A_{i'i} \prod_{j=1,\ldots,n, j\neq i, j\neq i', j\neq k} (1 - A_{i'j}). \end{aligned} \tag{3}$$

Combining (3) with

$$\tilde{\rho}_i(1-\tilde{A}_{ii'}) \prod_{j=1,\dots,n, j\neq i, j\neq i', j\neq k} (1-\tilde{A}_{ij}) = \rho_i(1-A_{ii'}) \prod_{j=1,\dots,n, j\neq i, j\neq i', j\neq k} (1-A_{ij}),$$

we obtain $\tilde{A}_{ii'} = A_{ii'}$ for $i = 1,\dots,n_L, i' = 1,\dots,n_L$ by a similar argument to that at the beginning of the proof. It follows immediately that $\tilde{\rho}_i = \rho_i$ for $i = 1,\dots,n_L$. $\square$

Remark Neither conditions in Theorem 1 can be removed. That is, if either condition is removed, then there exists $(\rho, A) \in \mathcal{P}$ such that (ρ, A) is not identifiable. In fact,

$$\rho = (1/2, 1/2),\ \ A = \begin{pmatrix} 1 & 1/2 & 0 \\ 1 & 1 & 1/2 \end{pmatrix}$$

and

$$\rho = (1/4, 3/4),\ \ A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1/3 \end{pmatrix}$$

give the same probability distribution, which implies condition (i) is necessary.

Moreover,

$$\rho = (1/2, 1/2),\ \ A = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \end{pmatrix}$$

and

$$\rho = (1/4, 3/4),\ \ A = \begin{pmatrix} 1 & 0 & 1/2 \\ 2/3 & 1 & 1/2 \end{pmatrix}$$

give the same probability distribution, which implies condition (ii) is necessary.

Proof of Theorem 2. Let $(\tilde{\rho}, \tilde{A}) \in \mathcal{P}$ be a set of parameters of the hub model with the null component such that $\mathbb{P}(g|\rho, A) = \mathbb{P}(g|\tilde{\rho}, \tilde{A})$ for all g. Consider the probability that no one appears:

$$\tilde{\rho}_0 \prod_{j=1}^{n} (1 - \tilde{\pi}_j) = \rho_0 \prod_{j=1}^{n} (1 - \pi_j).$$

For $k = n_L + 1, \dots, n$, consider the probability that only k appears:

$$\tilde{\rho}_0 \tilde{\pi}_k \prod_{j=1,\dots,n, j \neq k} (1 - \tilde{\pi}_j) = \rho_0 \pi_k \prod_{j=1,\dots,n, j \neq k} (1 - \pi_j).$$

From the above equations, we obtain

$$\begin{aligned} \tilde{\pi}_k &= \pi_k, \quad k = n_L + 1, \dots, n, \\ \tilde{\rho}_0 \prod_{j=1}^{n_L} (1 - \tilde{\pi}_j) &= \rho_0 \prod_{j=1}^{n_L} (1 - \pi_j). \end{aligned} \tag{4}$$

By condition (iii), for $i = 1, \dots, n_L$, let k and k' be the nodes from $\{n_L + 1, \dots, n\}$ such that $\pi_k \neq A_{ik}$ and $\pi_{k'} \neq A_{ik'}$.

Consider the probability that i appears but no other nodes from $\{1, \dots, n_L\}$ appears

(the rest do not matter)

$$
\begin{aligned}
&\tilde{\rho}_0\tilde{\pi}_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\tilde{\pi}_j) + \tilde{\rho}_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\tilde{A}_{ij}) \\
=&\rho_0\pi_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\pi_j) + \rho_i \prod_{j=1,\ldots,n_L,j\neq i}(1-A_{ij});
\end{aligned} \tag{5}
$$

the probability that i and k appear but no other nodes from $\{1,\ldots,n_L\}$ appears (the rest do not matter)

$$
\begin{aligned}
&\tilde{\rho}_0\tilde{\pi}_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\tilde{\pi}_j)\pi_k + \tilde{\rho}_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\tilde{A}_{ij})\tilde{A}_{ik} \\
=&\rho_0\pi_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\pi_j)\pi_k + \rho_i \prod_{j=1,\ldots,n_L,j\neq i}(1-A_{ij})A_{ik};
\end{aligned} \tag{6}
$$

the probability that i and k' appear but no other nodes from $\{1,\ldots,n_L\}$ appears (the rest do not matter)

$$
\begin{aligned}
&\tilde{\rho}_0\tilde{\pi}_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\tilde{\pi}_j)\pi_{k'} + \tilde{\rho}_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\tilde{A}_{ij})\tilde{A}_{ik'} \\
=&\rho_0\pi_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\pi_j)\pi_{k'} + \rho_i \prod_{j=1,\ldots,n_L,j\neq i}(1-A_{ij})A_{ik'};
\end{aligned} \tag{7}
$$

and the probability that i, k and k' appear but no other nodes from $\{1,\ldots,n_L\}$ appears

(the rest do not matter)

$$
\begin{aligned}
&\tilde{\rho}_0\tilde{\pi}_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\tilde{\pi}_j)\pi_k\pi_{k'} + \tilde{\rho}_i \prod_{l=1,\ldots,n_L,j\neq i}(1-\tilde{A}_{ij})\tilde{A}_{ik}\tilde{A}_{ik'} \\
=&\rho_0\pi_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\pi_j)\pi_k\pi_{k'} + \rho_i \prod_{l=1,\ldots,n_L,j\neq i}(1-A_{ij})A_{ik}A_{ik'}. \qquad (8)
\end{aligned}
$$

Note that the above equations are not probabilities of a single realization g but are sums of multiple $\mathbb{P}(g)$. Moreover, we put $\pi_k, \pi_{k'}$ instead of $\tilde{\pi}_k, \tilde{\pi}_{k'}$ on the LHS of the equations, since we have proved $\tilde{\pi}_k = \pi_k, k = n_L+1, \ldots, n$.

Let

$$
\begin{aligned}
x &= \rho_0\pi_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\pi_j), \\
\tilde{x} &= \tilde{\rho}_0\tilde{\pi}_i \prod_{j=1,\ldots,n_L,j\neq i}(1-\tilde{\pi}_j), \\
y &= \rho_i \prod_{j=1,\ldots,n_L,j\neq i}(1-A_{ij}), \\
\tilde{y} &= \tilde{\rho}_i \prod_{l=1,\ldots,n_L,j\neq i}(1-\tilde{A}_{ij}).
\end{aligned}
$$

Then (5), (6) (7) and (8) become

$$
\begin{aligned}
\tilde{x}+\tilde{y} &= x+y, \\
\tilde{x}\pi_k + \tilde{y}\tilde{A}_{ik} &= x\pi_k + yA_{ik}, \\
\tilde{x}\pi_{k'} + \tilde{y}\tilde{A}_{ik'} &= x\pi_{k'} + yA_{ik'}, \\
\tilde{x}\pi_k\pi_{k'} + \tilde{y}\tilde{A}_{ik}\tilde{A}_{ik'} &= x\pi_k\pi_{k'} + yA_{ik}A_{ik'}.
\end{aligned}
$$

Plugging $\tilde{x} - x = y - \tilde{y}$ into the last three equations, we obtain

$$\tilde{y}\tilde{A}_{ik} = \tilde{y}\pi_k + y(A_{ik} - \pi_k), \tag{9}$$

$$\tilde{y}\tilde{A}_{ik'} = \tilde{y}\pi_{k'} + y(A_{ik'} - \pi_{k'}), \tag{10}$$

$$y\pi_k\pi_{k'} + \tilde{y}\tilde{A}_{ik}\tilde{A}_{ik'} = \tilde{y}\pi_k\pi_{k'} + yA_{ik}A_{ik'}. \tag{11}$$

Multiplying (11) by $\tilde{y}$, and plugging the right hand sides of (9) and (10) into the resulting equation, we obtain

$$\begin{aligned}
&y\tilde{y}\pi_k\pi_{k'} + \tilde{y}^2\pi_k\pi_{k'} + \tilde{y}\pi_k y(A_{ik'} - \pi_{k'}) + \tilde{y}\pi_{k'} y(A_{ik} - \pi_k) + y^2(A_{ik} - \pi_k)(A_{ik'} - \pi_{k'}) \\
&= \tilde{y}^2\pi_k\pi_{k'} + y\tilde{y}A_{ik}A_{ik'}, \\
\Rightarrow &y(A_{ik} - \pi_k)(A_{ik'} - \pi_{k'}) = \tilde{y}(A_{ik} - \pi_k)(A_{ik'} - \pi_{k'}).
\end{aligned}$$

Therefore, $\tilde{y} = y$ since $\pi_k \neq A_{ik}$ and $\pi_{k'} \neq A_{ik'}$. It follows that $\tilde{x} = x$, i.e.,

$$\tilde{\rho}_0\tilde{\pi}_i \prod_{j=1,\ldots,n_L, j\neq i} (1 - \tilde{\pi}_j) = \rho_0\pi_i \prod_{j=1,\ldots,n_L, j\neq i} (1 - \pi_j), \quad i = 1, \ldots, n_L.$$

Combining the above equation with (4), we obtain

$$\begin{aligned}
&\tilde{\pi}_i = \pi_i, \quad i = 1, \ldots, n_L, \\
&\tilde{\rho}_0 = \rho_0.
\end{aligned}$$

Note that $\mathbb{P}(g) = \mathbb{P}(g|z = 0)\mathbb{P}(z = 0) + \mathbb{P}(g|z \neq 0)\mathbb{P}(z \neq 0)$. So far we have proved parameters of $\mathbb{P}(g|z=0)$ and $\mathbb{P}(z=0)$ are identifiable. We only need to prove the identifiability of $\mathbb{P}(g|z\neq 0)$, which is the case of the asymmetric hub model and has been proved by Theorem 1. $\square$

Remark No conditions in Theorem 2 can be removed. Here we only give a counterexample when condition (iii) is not satisfied since the other two are similar to the case of Theorem 1. In fact,

$$\rho = (1/2, 1/2),\ \ A = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1 & 1/2 & 1/2 \end{pmatrix}$$

and

$$\rho = (1/4, 3/4),\ \ A = \begin{pmatrix} 0 & 0 & 1/2 \\ 1 & 1/3 & 1/2 \end{pmatrix}$$

give the same probability distribution.

2 Proofs in Section 2.3

We start by recalling notations defined in the main text. Recall that z_* is the true label assignment, z is an arbitrary label assignment, and $\hat{z}$ is the maximum profile likelihood estimator. Furthermore, $t_{i*} = \sum_t 1(z_*^{(t)} = i)$, and $t_i = \sum_t 1(z^{(t)} = i), t_{ii'} = \sum_t 1(z_*^{(t)} = i, \hat{z}^{(t)} = i')$.

Proof of Lemma 1. We first prove a fact: under H_1 and H_4, for $0 < \delta_1 < e^{-c_0}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{n_L}\left\{\frac{t_{ii}}{t_{i*}} \le \delta_1\right\}\right) \to 0.$$

Note that $\hat{z}$ must be feasible (the estimated hub must appear in the group as we assume $A_{ii} \equiv 1$), we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{t_{ii}}{t_{i*}} \le \delta_1 \middle| z_*\right) \\
\le &\mathbb{P}\left(\frac{1}{t_{i*}}\sum_{t=1}^{T} 1(z_*^{(t)} = i) \prod_{k\in\{1,\dots,n_L\},k\neq i}(1 - G_k^{(t)}) \le \delta_1 \middle| z_*\right).
\end{aligned} \tag{12}$$

Now since

$$\mathbb{E}\left[\prod_{k\in\{1,\dots,n_L\},k\neq i}(1 - G_k^{(t)}) \middle| z_*^{(t)} = i\right] = \prod_{k\in\{1,\dots,n_L\},k\neq i}(1 - A_{ik}) \ge (1 - c_0/n_L)^{n_L} \ge e^{-c_0},$$

by Hoeffding's inequality,

$$\begin{aligned}
(12) \le &\mathbb{P}\left(\frac{1}{t_{i*}}\sum_{t=1}^{T} 1(z_*^{(t)} = i)\left[\prod_{k\in\{1,\dots,n_L\},k\neq i}(1 - G_k^{(t)}) - \prod_{k\in\{1,\dots,n_L\},k\neq i}(1 - A_{ik})\right] \le \delta_1 - e^{-c_0} \middle| z_*\right) \\
\le &\exp\{-2t_{i*}(e^{-c_0} - \delta_1)^2\}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{i=1}^{n_L}\left\{\frac{t_{ii}}{t_{i*}} \leq \delta_1\right\}\middle| z_*\right) \\
=&\mathbb{P}\left(\bigcup_{i=1}^{n_L}\left\{\frac{t_{ii}}{t_{i*}} \leq \delta_1\right\}, \{t_{i*} \geq c_{\min}T/n_L, \text{for all } i\}\middle| z_*\right) \\
&+ \mathbb{P}\left(\bigcup_{i=1}^{n_L}\left\{\frac{t_{ii}}{t_{i*}} \leq \delta_1\right\}, \{t_{i*} < c_{\min}T/n_L, \text{for some } i\}\middle| z_*\right) \\
\leq& \sum_{i=1}^{n_L} \mathbb{P}\left(\frac{t_{ii}}{t_{i*}} \leq \delta_1 \middle| z_*\right) 1(t_{i*} \geq c_{\min}T/n_L) \\
&+ 1(t_{i*} < c_{\min}T/n_L, \text{for some } i) \\
\leq& n_L \exp\{-2c_{\min}T/(n_L)(e^{-c_0} - \delta_1)^2\} + 1(t_{i*} < c_{\min}T/n_L, \text{for some } i).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{i=1}^{n_L}\left\{\frac{t_{ii}}{t_{i*}} \leq \delta_1\right\}\right) \\
=&\mathbb{E}_{z_*}\left[\mathbb{P}\left(\bigcup_{i=1}^{n_L}\left\{\frac{t_{ii}}{t_{i*}} \leq \delta_1\right\}\middle| z_*\right)\right] \\
\leq& n_L \exp\{-2c_{\min}T/(n_L)(e^{-c_0} - \delta_1)^2\} + \mathbb{P}(t_{i*} < c_{\min}T/n_L, \text{for some } i) \to 0.
\end{aligned}$$

Therefore, $\frac{t_{ii}}{t_{i*}} \geq \delta_1$ for $i = 1, \ldots, n_L$ with probability approaching 1.

Let $\mathcal{E} = \{\frac{t_{ii}}{t_{i*}} \geq \delta_1 \text{ and } t_{i*} \geq c_{\min}T/n_L, i = 1, \ldots, n_L\}$. We have shown $\mathbb{P}(\mathcal{E}) \to 1$. The inequalities below are proved within the set $\mathcal{E}$, and thus hold with probability approaching 1.

For $i = 1, \ldots, n_L$, $k = 1, \ldots, n_L$, $k \neq i$,

$$\frac{t_{ik}}{t_k} = \frac{t_{ik}}{\sum_{k'=1}^{n_L} t_{k'k}} \leq \frac{t_{ik}}{t_{ik} + t_{kk}} = \frac{t_{ik}/t_{i*}}{t_{ik}/t_{i*} + t_{kk}/t_{k*} \cdot t_{k*}/t_{i*}} \leq \frac{1}{1 + \delta_1 \cdot c_{\min}/c_{\max}} = \delta_2 < 1.$$

Under H_2 and H_3, $\min_{i,i'=1,\ldots,n_L, i\neq i', j \in V_i} A_{ij} - A_{i'j} = \tau d$, where τ is bounded away from 0. Now we give a lower bound for $A_{ij} - \bar{A}_{kj}$ for $j \in V_i$ and $k \neq i$,

$$\begin{aligned} A_{ij} - \bar{A}_{kj} &= \frac{\sum_t (A_{ij} - P_j^{(t)}) 1(\hat{z}^{(t)} = k)}{t_k} \\ &= \frac{\sum_{k'=1}^{n_L} (A_{ij} - A_{k'j}) t_{k'k}}{t_k} \\ &\geq \frac{\tau d \sum_{k' \neq i} t_{k'k}}{t_k} \geq \tau(1 - \delta_2) d. \end{aligned} \tag{13}$$

Next, we show the following fact: if $p = \rho_1 d, q = \rho_2 d$ where $\rho_1 > \rho_2$ are fixed positive numbers, then there exists $\delta_3 > 0$ such that $\mathrm{KL}(p,q) \geq \delta_3 d$, where $\mathrm{KL}(p,q) = p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q}$.

$$\begin{aligned} \mathrm{KL}(p,q) &= p \log \frac{p}{q} + \log(1 - \frac{p-q}{1-q}) + p \log \frac{1-p}{1-q} \\ &= -p \log \frac{q}{p} + \frac{p-q}{1-q} + o(d) + \rho_1 d\, o(1) \\ &\geq -p \log \frac{q}{p} + (q - p) + o(d) \\ &= p \left[\frac{q-p}{p} - \log \left(1 + \frac{q-p}{p} \right) \right] + o(d) \\ &\geq \delta_3 d. \end{aligned}$$

The last line holds for sufficiently small δ_3 because $\frac{q-p}{p} - \log(1 + \frac{q-p}{p}) = c_{\rho_1, \rho_2} > 0$ where

$\frac{q-p}{p} \in (-1, 0)$ and c_{ρ_1,ρ_2} is a constant depending on ρ_1 and ρ_2.

As $\bar{A}_{kj} = \frac{\sum_t P_j^{(t)} 1(z^{(t)}=j)}{t_i} = [\sum_t A_{z_*^{(t)},j} 1(z^{(t)} = j)]/t_i \asymp d$, combining the above fact and (13), we have

$$
\begin{aligned}
L_P(z_*) - L_P(\hat{z}) &= \sum_t \sum_j \mathrm{KL}(P_j^{(t)}, \bar{A}_{\hat{z}^{(t)},j}) \\
&\geq \sum_{i=1}^{n_L} \sum_{k \neq i} \sum_{t: z_*^{(t)}=i, \hat{z}^{(t)}=k} \sum_{j \in V_i} \mathrm{KL}(A_{ij}, \bar{A}_{\hat{z}^{(t)},j}) \\
&\geq \sum_{i=1}^{n_L} \sum_{k \neq i} \sum_{t: z_*^{(t)}=i, \hat{z}^{(t)}=k} \sum_{j \in V_i} \tau(1-\delta_2)\delta_3 d \\
&\geq \tau(1-\delta_2)\delta_3 dvnT_e/n_L.
\end{aligned}
$$

Letting $\delta = 1/[\tau(1-\delta_2)\delta_3]$,

$$
\frac{\delta n_L}{dvnT}(L_P(z_*) - L_P(\hat{z})) \geq \frac{T_e}{T},
$$

with probability approaching 1. $\square$

To prove Theorem 3, we need the following lemma.

Lemma S1.

$$
\begin{aligned}
&\mathbb{P}(\max_z |L_G(z) - L_P(z)| \geq 2\eta) \leq \\
&\quad n_L^T (T/n_L + 1)^{n_L n} e^{-\eta} + 2n_L^T \exp\left\{ -\frac{\eta^2/4}{\sum_t \sum_j (\log \bar{A}_{ij})^2 \mathrm{Var}(G_j^{(t)}) + \max_{ij} |\log \bar{A}_{ij}| \eta/6} \right\} \\
&\quad + 2n_L^T \exp\left\{ -\frac{\eta^2/4}{\sum_{ij:\bar{A}_{ij}<1} \left((\log(1-\bar{A}_{ij}))^2 \sum_{t:z^{(t)}=i} \mathrm{Var}(G_j^{(t)}) \right) + \max_{ij:\bar{A}_{ij}<1} |\log(1-\bar{A}_{ij})| \eta/6} \right\}.
\end{aligned}
$$

Proof of Lemma S1.

$$
\begin{aligned}
L_G(z) - L_P(z) =& \left(\sum_{i=1}^{n_L} t_i \sum_j \hat{A}_{ij} \log \hat{A}_{ij} + (1 - \hat{A}_{ij}) \log(1 - \hat{A}_{ij}) \right) \\
&- \left(\sum_{i=1}^{n_L} t_i \sum_j \hat{A}_{ij} \log \bar{A}_{ij} + (1 - \hat{A}_{ij}) \log(1 - \bar{A}_{ij}) \right) \\
&+ \left(\sum_{i=1}^{n_L} t_i \sum_j \hat{A}_{ij} \log \bar{A}_{ij} + (1 - \hat{A}_{ij}) \log(1 - \bar{A}_{ij}) \right) \\
&- \left(\sum_{i=1}^{n_L} t_i \sum_j \bar{A}_{ij} \log \bar{A}_{ij} + (1 - \bar{A}_{ij}) \log(1 - \bar{A}_{ij}) \right) \\
=& \sum_{i=1}^{n_L} t_i \sum_j D(\hat{A}_{ij} | \bar{A}_{ij}) + B_{n_L,n,T}.
\end{aligned}
$$

To bound $\sum_{i=1}^{n_L} t_i \sum_j D(\hat{A}_{ij}|\bar{A}_{ij})$, we adopt the approach in Choi et al. (2012), which is based on a heterogeneous Chernoff bound in Dubhashi and Panconesi (2009). Let ν be any realization of $\hat{A}$.

$$
\mathbb{P}(\hat{A}_{ij} = \nu_{ij} | z_*) \leq e^{-t_i D(\nu_{ij}|\bar{A}_{ij})}.
$$

By the independence of $\hat{A}_{ij}$ conditional on z_*,

$$
\mathbb{P}(\hat{A} = \nu | z_*) \leq \exp\left\{ -\sum_{i=1}^{n_L} \sum_j t_i D(\nu_{ij}|\bar{A}_{ij}) \right\}.
$$

Let $\hat{\mathcal{A}}$ be the range of $\hat{A}$ for a fixed z. Then $|\hat{\mathcal{A}}| \leq \prod_{i=1}^{n_L} (t_i + 1)^n \leq \prod_{i=1}^{n_L} (t_i + 1)^n \leq (T/n_L + 1)^{n_L n}$, as $\hat{A}_{ij}$ can only take values from $0/t_i, 1/t_i, \ldots, t_i/t_i$.

For all $\eta > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{n_L}\sum_{j} t_i D(\hat{A}_{ij}|\bar{A}_{ij}) \geq \eta \,\middle|\, z_*\right)\\
=&\sum_{\nu\in\hat{\mathcal{A}}}\mathbb{P}\left(\hat{A}=\nu, \sum_{i=1}^{n_L}\sum_{j} t_i D(\nu_{ij}|\bar{A}_{ij}) \geq \eta \,\middle|\, z_*\right)\\
\leq&\sum_{\nu\in\hat{\mathcal{A}}}\exp\left\{-\sum_{i=1}^{n_L}\sum_{j} t_i D(\nu_{ij}|\bar{A}_{ij})\right\}1\left\{-\sum_{i=1}^{n_L}\sum_{j} t_i D(\nu_{ij}|\bar{A}_{ij}) \leq -\eta\right\}\\
\leq&\sum_{\nu\in\hat{\mathcal{A}}} e^{-\eta} \leq |\hat{\mathcal{A}}|e^{-\eta} \leq (T/n_L+1)^{n_L n}e^{-\eta},
\end{aligned}
$$

and then

$$
\mathbb{P}\left(\max_z \sum_{i=1}^{n_L}\sum_{j} t_i D(\hat{A}|\bar{A}_{ij}) \geq \eta\right) \leq n_L^T (T/n_L+1)^{n_L n}e^{-\eta}. \tag{14}
$$

Next, we bound $B_{n_L,n,T}$. Let $B_{n_L,n,T} = B_{1,n_L,n,T} + B_{2,n_L,n,T}$, where

$$
\begin{aligned}
B_{1,n_L,n,T} &= \sum_i \left(\sum_j \sum_{t:z^{(t)}=i} (G_j^{(t)} - P_j^{(t)}) \log \bar{A}_{ij}\right),\\
B_{2,n_L,n,T} &= \sum_i \left(\sum_j \sum_{t:z^{(t)}=i} (G_j^{(t)} - P_j^{(t)}) \log (1-\bar{A}_{ij})\right).
\end{aligned}
$$

As $\left|(G_j^{(t)} - P_j^{(t)}) \log \bar{A}_{ij}\right| \leq |\log \bar{A}_{ij}|$, by Bernstein's inequality, we have

$$
\begin{aligned}
\mathbb{P}(|B_{1,n_L,n,T}| \geq \eta/2) &\leq 2\exp\left\{-\frac{\eta^2/4}{\sum_t\sum_j (\log \bar{A}_{ij})^2 \mathrm{Var}(G_j^{(t)}) + \max_{ij}|\log \bar{A}_{ij}|\eta/6}\right\},\\
\mathbb{P}(\max_z |B_{1,n_L,n,T}| \geq \eta/2) &\leq 2n_L^T\exp\left\{-\frac{\eta^2/4}{\sum_t\sum_j (\log \bar{A}_{ij})^2 \mathrm{Var}(G_j^{(t)}) + \max_{ij}|\log \bar{A}_{ij}|\eta/6}\right\}.
\end{aligned} \tag{15}
$$

In addition, if $\bar{A}_{ij} = 1$, $\sum_{t:z^{(t)}=i}(G_j^{(t)} - P_j^{(t)}) \equiv 0$, which implies the term $\sum_{t:z^{(t)}=i}(G_j^{(t)} - P_j^{(t)}) \log(1-\bar{A}_{ij})$ in $B_{2,n_L,n,T}$ can be dropped. As $\left|(G_j^{(t)} - P_j^{(t)}) \log(1 - \bar{A}_{ij})\right| \leq |\log(1-\bar{A}_{ij})|$, by Bernstein's inequality,

$$
\begin{aligned}
&\mathbb{P}(|B_{2,n_L,n,T}| \geq \eta/2) \\
&\leq 2\exp\left\{-\frac{\eta^2/4}{\sum_{ij:\bar{A}_{ij}<1}\left((\log(1-\bar{A}_{ij}))^2 \sum_{t:z^{(t)}=i} \mathrm{Var}(G_j^{(t)})\right) + \max_{ij:\bar{A}_{ij}<1} |\log(1-\bar{A}_{ij})|\eta/6}\right\}, \\
&\mathbb{P}(\max_z |B_{2,n_L,n,T}| \geq \eta/2) \\
&\leq 2n_L^T \exp\left\{-\frac{\eta^2/4}{\sum_{ij:\bar{A}_{ij}<1}\left((\log(1-\bar{A}_{ij}))^2 \sum_{t:z^{(t)}=i} \mathrm{Var}(G_j^{(t)})\right) + \max_{ij:\bar{A}_{ij}<1} |\log(1-\bar{A}_{ij})|\eta/6}\right\}.
\end{aligned}
\tag{16}
$$

Finally, combining (14), (15) and (16), we obtain

$$
\begin{aligned}
&\mathbb{P}(\max_z |L_G(z) - L_P(z)| \geq 2\eta) \\
&\quad \leq \mathbb{P}\left(\max_z \sum_{i=1}^{n_L}\sum_j t_i D(\hat{A}|\bar{A}_{ij}) \geq \eta\right) + \mathbb{P}(\max_z |B_{1,n_L,n,T}| \geq \eta/2) + \mathbb{P}(\max_z |B_{2,n_L,n,T}| \geq \eta/2) \\
&\quad \leq n_L^T (T/n_L + 1)^{n_L n} e^{-\eta} + 2n_L^T \exp\left\{-\frac{\eta^2/4}{\sum_t \sum_j (\log \bar{A}_{ij})^2 \mathrm{Var}(G_j^{(t)}) + \max_{ij} |\log \bar{A}_{ij}|\eta/6}\right\} \\
&\quad + 2n_L^T \exp\left\{-\frac{\eta^2/4}{\sum_{ij:\bar{A}_{ij}<1}\left((\log(1-\bar{A}_{ij}))^2 \sum_{t:z^{(t)}=i} \mathrm{Var}(G_j^{(t)})\right) + \max_{ij:\bar{A}_{ij}<1} |\log(1-\bar{A}_{ij})|\eta/6}\right\}.
\end{aligned}
$$

$\square$

Proof of Theorem 3. First we show the following fact: under $H_1 - H_4$, if $n_L^2 \log T/(dTv) \to$

0, $(\log d)^2 n_L^2 \log n_L/(dnv^2) \to 0$ and $(\log T)^2 n_L^2 \log n_L/(dnv^2) \to 0$, then

$$\max_z \frac{n_L}{dvnT}|L_P(z) - L_G(z)| = o_p(1), \quad \text{as } n_L \to \infty, n \to \infty, T \to \infty. \tag{17}$$

Letting $\eta = dvnT\epsilon/n_L$, the LHS in Lemma S1 becomes $\mathbb{P}(\max_z \frac{n_L}{dvnT}|L_G(z) - L_P(z)| \geq 2\epsilon)$. To prove the above fact, we need to show each term in the RHS of Lemma S1 goes to 0.

For the first term, it is easy to check that if $n_L \log n_L/(dvn) \to 0$ and $n_L^2 \log T/(dvT) \to 0$, then

$$n_L^T (T/n_L)^{n_L n} e^{-\frac{dvnT\epsilon}{n_L}} \to 0.$$

Under H_2, $A_{ij} \asymp d$ and $|\log \bar{A}_{ij}| = O(|\log d|)$ for $i \neq j$. We can therefore find a constant C_1 such that

$$\mathbb{P}(|B_{1,n_L,n,T}| \geq dvnT\epsilon/(2n_L)) \leq 2\exp\left\{-\frac{d^2v^2n^2T^2\epsilon^2/(4n_L^2)}{C_1^2 Tn(\log d)^2 d + C_1|\log d| dvnT\epsilon/(6n_L)}\right\},$$

and

$$\mathbb{P}(\max_z |B_{1,n_L,n,T}| \geq dvnT\epsilon/(2n_L)) \leq 2n_L^T \exp\left\{-\frac{d^2v^2n^2T^2\epsilon^2/(4n_L^2)}{C_1^2 Tn(\log d)^2 d + C_1|\log d| dvnT\epsilon/(6n_L)}\right\}.$$

Then if $(\log d)^2 n_L^2 \log n_L/(dnv^2) \to 0$,

$$\mathbb{P}(\max_z |B_{1,n_L,n,T}| \geq dvnT\epsilon/(2n_L)) \to 0.$$

For the third term, when $\bar{A}_{ij} < 1$, we have

$$\bar{A}_{ij} \leq \frac{(t_i - 1) + P_j^{(t)}}{t_i},$$
$$1 - \bar{A}_{ij} \geq \frac{1 - P_j^{(t)}}{t_i} \geq \frac{1 - P_j^{(t)}}{T},$$

which imply $|\log(1 - \bar{A}_{ij})| \leq C_2 \log T$ for some constant $C_2 > 0$. Therefore,

$$\mathbb{P}(\max_z |B_{2,n_L,n,T}| \geq dvnT\epsilon/(2n_L)) \leq 2n_L^T \exp\left\{-\frac{d^2v^2n^2T^2\epsilon^2/(4n_L^2)}{C_2^2(\log T)^2Tnd + C_2(\log T)dnvT\epsilon/(6n_L)}\right\}.$$

Furthermore, if $(\log T)^2 n_L^2 \log n_L/(dnv^2) \to 0$,

$$\mathbb{P}(\max_z |B_{2,n_L,n,T}| \geq dvnT\epsilon/(2n_L)) \to 0.$$

Combining the inequalities of the above three terms, we have proved (17).

Finally, for all $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\frac{T_e}{T} \geq \epsilon\right) =& \mathbb{P}\left(\frac{T_e}{T} \geq \epsilon, \frac{\delta n_L}{dvnT}(L_P(z_*) - L_P(\hat{z})) \geq \frac{T_e}{T}\right) \\
&+ \mathbb{P}\left(\frac{T_e}{T} \geq \epsilon, \frac{\delta n_L}{dvnT}(L_P(z_*) - L_P(\hat{z})) < \frac{T_e}{T}\right) \\
=& \mathbb{P}\left(\frac{\delta n_L}{dvnT}(L_P(z_*) - L_P(\hat{z})) \geq \epsilon\right) + o(1) \quad \text{(by Lemma 1)} \\
=& \mathbb{P}\left(\frac{\delta n_L}{dvnT}\left[(L_P(z_*) - L_G(z_*)) + (L_G(z_*) - L_G(\hat{z})) + (L_G(\hat{z}) - L_P(\hat{z}))\right] \geq \epsilon\right) + o(1) \\
\leq& \mathbb{P}\left(\frac{\delta n_L}{dvnT}\left(|L_P(z_*) - L_G(z_*)| + |L_G(\hat{z}) - L_P(\hat{z})|\right) \geq \epsilon\right) + o(1) \\
\to& \, 0.
\end{aligned}$$

$\square$

We now give the result of label consistency for fixed n_L. We make the following assumptions similar to $H_1 - H_4$.

H_1': $c_{\min}T \leq t_{i*} \leq c_{\max}T$ for $i = 1, \ldots, n_L$.

H_2': $A_{ij} = s_{ij}d$ for $i = 1, \ldots, n_L, j = 1, \ldots, n$ and $i \neq j$ where s_{ij} are unknown constants satisfying $0 < s_{\min} \leq s_{ij} \leq s_{\max} < \infty$ while d goes to 0 as n goes to infinity.

H_3': There exists a set $V_i \subset \{n_L + 1, \ldots, n\}$ for $i = 1, \ldots, n_L$ with $|V_i| \geq vn$ such that $\tau = \min_{i,i'=1,\ldots,n_L, i \neq i', j \in V_i}(s_{ij} - s_{i'j})$ is bounded away from 0.

H_4': $A_{ii'}$ is bounded away from 1 for $i = 1, \ldots, n_L, i' = 1, \ldots, n_L$ and $i \neq i'$.

Theorem 3′. *Under* H_1'–H_4', *if* $\log T/(dTv) = o(1)$, $(\log d)^2/(dnv^2) = o(1)$ *and* $(\log T)^2/(dnv^2) = o(1)$, *then*

$$T_e/T = o_p(1), \quad \text{as } n \to \infty, T \to \infty.$$

We omit all the proofs for fixed n_L because they are trivial corollaries of the results for growing n_L.

Proof of Theorem 4. First we show the following fact: under the conditions in Theorem 4,

$$n_L T_e/T = o_p(1), \text{ as } n_L \to \infty, n \to \infty, T \to \infty.$$

According to the proof in Theorem 3, we need

$$\mathbb{P}\left(\frac{\delta n_L^2}{dvnT}\left(|L_P(z_*)-L_G(z_*)|+|L_G(\hat{z})-L_P(\hat{z})|\right)\geq\epsilon\right)\to 0,$$

which holds if we can show

$$\max_z \frac{n_L^2}{dvnT}|L_G(z)-L_P(z)| = o_p(1).$$

As in the proof of Lemma S1, this holds by letting $\eta = dvnT\epsilon/n_L^2$.

Then we bound $\left|\hat{A}_{ij}^{\hat{z}}-\hat{A}_{ij}^{z_*}\right|$:

$$\begin{aligned}
|\hat{A}_{ij}^{\hat{z}}-\hat{A}_{ij}^{z_*}| &= \left|\frac{\sum_t G_j^{(t)}1(\hat{z}^{(t)}=i)}{t_i}-\frac{\sum_t G_j^{(t)}1(z_*^{(t)}=i)}{t_{i*}}\right| \\
&\leq \left|\frac{\sum_t G_j^{(t)}1(\hat{z}^{(t)}=i)}{t_i}-\frac{\sum_t G_j^{(t)}1(\hat{z}^{(t)}=i)}{t_{i*}}\right| + \left|\frac{\sum_t G_j^{(t)}1(\hat{z}^{(t)}=i)}{t_{i*}}-\frac{\sum_t G_j^{(t)}1(z_*^{(t)}=i)}{t_{i*}}\right| \\
&\leq \left|\frac{t_{i*}-t_i}{t_{i*}}\right| + \frac{\sum_t\left|1(\hat{z}^{(t)}=i)-1(z_*^{(t)}=i)\right|}{t_{i*}} \leq \delta n_L T_e/T,
\end{aligned}$$

where δ is a constant. The last line holds by H_1'.

Furthermore,

$$\begin{aligned}
&\mathbb{P}\left(\max_{ij}\left|\hat{A}_{ij}^{\hat{z}}-A_{ij}\right|\geq\epsilon\right) \\
\leq&\mathbb{P}\left(\max_{ij}\left|\hat{A}_{ij}^{\hat{z}}-\hat{A}_{ij}^{z_*}\right|\geq\epsilon/2\right)+P\left(\max_{ij}\left|\hat{A}_{ij}^{z_*}-A_{ij}\right|\geq\epsilon/2\right) \\
\leq&\mathbb{P}\left(\delta n_L T_e/T\geq\epsilon\right)+P\left(\max_{ij}\left|\hat{A}_{ij}^{z_*}-A_{ij}\right|\geq\epsilon/2\right).
\end{aligned}$$

The second term vanishes by Hoeffding's inequality: for all $\epsilon > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\left|\hat{A}_{ij}^{z_*} - A_{ij}\right| \geq \epsilon/2 \middle| z_*\right) \\
=&\mathbb{P}\left(\left|\sum_t 1(z_*^{(t)} = i)(G_j^{(t)} - A_{ij})\right| \geq \epsilon t_{i*}/2 \middle| z_*\right) \\
\leq& 2\exp\{-\epsilon^2 t_{i*}/2\}.
\end{aligned}
$$

Therefore, if $n_L \log n/T \to 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\max_{ij}\left|\hat{A}_{ij}^{z_*} - A_{ij}\right| \geq \epsilon/2\right) \\
\leq& 2nn_L \exp\{-\epsilon^2 c_{\min} T/(2n_L)\} + \mathbb{P}(t_{i*} < c_{\min}T/n_L, \text{for some } i) \to 0.
\end{aligned}
$$

$\square$

The following theorem is on estimation consistency for fixed n.

Theorem 4′. *Under H_1'–H_4', if $\log n/T = o(1)$, $\log T/(dTv) = o(1)$, $(\log d)^2/(dnv^2) = o(1)$ and $(\log T)^2/(dnv^2) = o(1)$, then*

$$
\max_{i\in\{1,\ldots,n_L\}, j\in\{1,\ldots,n\}} \left|\hat{A}_{ij}^{z_*} - A_{ij}\right| = o_p(1), \quad \text{as } n \to \infty, T \to \infty.
$$

Finally, we give the simplest version of the estimation consistency result, which only considers the rates of n and T but treats n_L, d, and v as fixed.

Theorem 4″. *Under $H_1' - H_4'$, for fixed d and v, if $\log n/T = o(1)$ and $(\log T)^2/n = o(1)$,* *then*

$$\max_{i \in \{1,\ldots,n_L\}, j \in \{1,\ldots,n\}} \left| \hat{A}_{ij}^{z_*} - A_{ij} \right| = o_p(1), \quad \text{as } n \to \infty, T \to \infty.$$

The first condition means n can grow faster than T as long as $\log n/T \to 0$. Such a condition is common in the literature of high-dimensional statistics. The second condition is more of a technical one: for proving the label consistency, we need an upper bound of the growth rate of T due to the concentration bound in Lemma S1.

Proof of Lemma 2. By the proof of Lemma 1, there exists $\delta_1 > 0$ such that

$$t_{ii} + t_{i0} \geq \delta_1 t_{i*}, \quad i = 1, \ldots, n_L, \tag{18}$$

$$t_{00} \geq \delta_1 t_{0*}, \tag{19}$$

with probability approaching 1.

Therefore[1], for $i = 1, \ldots, n_L$, $j \in V_i$,

$$\begin{aligned} A_{ij} - \bar{A}_{0j} &= \frac{\sum_t (A_{ij} - P_j^{(t)}) 1(\hat{z}^{(t)} = 0)}{t_0} \\ &= \frac{\sum_{k=0}^{n_L} (A_{ij} - A_{kj}) t_{k0}}{t_0} \\ &\geq \frac{(A_{ij} - A_{0j}) t_{00}}{t_0} \geq \tau d \frac{t_{00}}{T} \geq \tau d \frac{t_{00}}{(n_L + 1) t_{0*}/c_{\min}} \geq \frac{\tau d c_{\min} \delta_1}{n_L}. \end{aligned}$$

[1]Some inequalities below hold with probability approaching 1. We omit this sentence occasionally.

Using the same argument in Lemma 1, it follows that

$$
\begin{aligned}
L_P(z_*) - L_P(\hat{z}) &= \sum_t \sum_j \mathrm{KL}(P_j^{(t)}, \bar{A}_{\hat{z}^{(t)},j}) \\
&\geq \max_{i=1,\ldots,n_L} \sum_{t: z_*^{(t)}=i, \hat{z}^{(t)}=0} \sum_{j \in V_i} \mathrm{KL}(A_{ij}, \bar{A}_{0j}) \\
&\geq \max_{i=1,\ldots,n_L} \frac{\tau d c_{\min} \delta_1 \delta_3}{n_L} \frac{vn}{n_L} t_{i0} \\
&\geq \max_{i=1,\ldots,n_L} \frac{\tau d c_{\min} \delta_1 \delta_3}{n_L} \frac{vn}{n_L} \frac{t_{i0}}{t_{i*}} \frac{c_{\min} T}{n_L} \\
&\geq \max_{i=1,\ldots,n_L} \tau \epsilon \frac{dvnT}{n_L^3} \frac{t_{i0}}{t_{i*}},
\end{aligned} \tag{20}
$$

where ϵ is a positive constant and τ is bounded away from 0.

Next, we show the following fact: under the conditions in Lemma 2,

$$
\max_z \frac{n_L^3}{dvnT} |L_G(z) - L_P(z)| = o_p(1).
$$

As in the proofs of Lemma S1 and Theorem 3, the above statement holds by letting $\eta = dvnT\epsilon/n_L^3$. Combining (20) and the above fact, by the same argument in Theorem 3, we have

$$
\mathbb{P}\left(\max_{i=1,\ldots,n_L} \frac{t_{i0}}{t_{i*}} \leq \eta \right) \to 1. \tag{21}
$$

□

Proof of Theorem 5. Due to (18) and (21), there exists $\delta_2 > 0$ such that

$$t_{ii} \geq \delta_2 t_{i*} \quad \text{for } i = 0, \dots, n_L,$$

with probability approaching 1. By the same argument in Lemma 1,

$$\begin{aligned} L_P(z_*) - L_P(\hat{z}) =& \sum_{t=1}^{T} \sum_{j=1}^{n} \mathrm{KL}(P_j^{(t)}, \bar{A}_{\hat{z}^{(t)},j}) \\ \geq& \sum_{i=1}^{n_L} \sum_{0 \leq k \leq n_L, k \neq i} \sum_{t: z_*^{(t)} = i, \hat{z}^{(t)} = k} \sum_{j \in V_i} \mathrm{KL}(A_{ij}, \bar{A}_{kj}) \\ \geq& \frac{vn}{n_L} \sum_{i=1}^{n_L} \sum_{0 \leq k \leq n_L, k \neq i} t_{ik} \tau (1 - \delta_2) \delta_3 d, \end{aligned}$$

which implies that there exists $\delta > 0$ such that with probability approaching 1,

$$\frac{\delta n_L}{dvnT} (L_P(z_*) - L_P(\hat{z})) \geq \sum_{i=1}^{n_L} \sum_{0 \leq k \leq n_L, k \neq i} \frac{t_{ik}}{T}. \tag{22}$$

By the same argument in Theorem 3, this further implies

$$\sum_{i=1}^{n_L} \sum_{0 \leq k \leq n_L, k \neq i} \frac{t_{ik}}{T} = o_p(1), \quad \text{as } n_L \to \infty, n \to \infty, T \to \infty, \tag{23}$$

if $n_L^2 \log T / (dvT) = o(1)$, $n_L^2 (\log T)^2 \log n_L / (dnv^2) = o(1)$ and $n_L^2 (\log d)^2 \log n_L / (dnv^2) = o(1)$.

As in the proof of Theorem 4,

$$\sum_{i=1}^{n_L} \sum_{0 \le k \le n_L, k \ne i} (n_L + 1) \frac{t_{ik}}{T} = o_p(1), \quad \text{as } n_L \to \infty, n \to \infty, T \to \infty, \tag{24}$$

if $n_L^3 \log T/(dvT) = o(1)$, $n_L^4 (\log T)^2 \log n_L/(dnv^2) = o(1)$ and $n_L^4 (\log d)^2 \log n_L/(dnv^2) = o(1)$.

Now we bound t_{0i}, $i = 1, \ldots, n_L$. From (24), $\sum_{1 \le k \le n_L, k \ne i} t_{ki} = o_p(T/(n_L + 1))$. And from $\delta_2 T c_{\min}/(n_L + 1) \le \delta_2 t_{i*} \le t_{ii}$, $\sum_{1 \le k \le n_L, k \ne i} t_{ki} \le t_{ii}$, with probability approaching 1. Moreover, from (19), $t_{0i} \le (1 - \delta_1) t_{0*}$.

Therefore, there exists $\delta_4 > 0$ such that for $i = 1, \ldots, n_L$, $j \in V_i$,

$$\begin{aligned}
A_{ij} - \bar{A}_{ij} &= \frac{\sum_t (A_{ij} - P_j^{(t)}) 1(\hat{z}^{(t)} = i)}{t_i} \\
&\ge \frac{(A_{ij} - A_{0j}) t_{0i}}{t_i} \\
&\ge \frac{\tau d t_{0i}}{t_{0i} + t_{ii} + \sum_{1 \le k \le n_L, k \ne i} t_{ki}} \\
&\ge \frac{\tau d t_{0i}}{(1 - \delta_1) t_{0*} + 2 t_{ii}} \\
&\ge \frac{\tau d t_{0i}}{(1 - \delta_1) t_{0*} + 2 t_{i*}} \ge \frac{\tau d n_L t_{0i}}{\delta_4 T}.
\end{aligned}$$

It follows that

$$\begin{aligned}
L_P(z_*) - L_P(\hat{z}) &\geq \max_{i=1,\ldots,n_L} \sum_{t: z_*^{(t)}=i, \hat{z}^{(t)}=i} \sum_{j \in V_i} \mathrm{KL}(A_{ij}, \bar{A}_{ij}) \\
&\geq \max_{i=1,\ldots,n_L} \frac{\tau d n_L t_{0i} \delta_3}{\delta_4 T} \frac{vn}{n_L+1} t_{ii} \\
&\geq \max_{i=1,\ldots,n_L} \frac{d}{\delta_4} \frac{n_L t_{0i}}{T} \frac{vn}{n_L+1} \tau \delta_2 \delta_3 t_{i*} \\
&\geq \max_{i=1,\ldots,n_L} \frac{d}{\delta_4} \frac{n_L t_{0i}}{T} \frac{vn}{n_L+1} \tau \delta_2 \delta_3 T \frac{c_{\min}}{n_L+1} \\
&\geq \max_{i=1,\ldots,n_L} \frac{dvnT}{n_L^2} \frac{n_L t_{0i}}{T} \delta, \qquad (25)
\end{aligned}$$

where $\delta = \tau \delta_2 \delta_3 c_{min} / \delta_4$ is positive constant.

By using the same argument in Theorem 3,

$$\max_{i=1,\ldots,n_L} \frac{n_L t_{0i}}{T} = o_p(1), \qquad (26)$$

if $n_L^4 \log T / (dvT) = o(1)$, $(\log T)^2 n_L^6 \log n_L / (dnv^2) = o(1)$ and $n_L^6 (\log d)^2 \log n_L / (dnv^2) = o(1)$. It follows that

$$\sum_{i=1}^{n_L} \frac{t_{0i}}{T} = o_p(1).$$

Combining (23) and (26),

$$\frac{T_e}{T} = o_p(1), \quad \text{as } n \to \infty, T \to \infty.$$

$\square$

For label consistency under the hub model with the null component with fixed n_L, we make the following assumptions:

$H_1^{*\prime}$: $Tc_{\min}/n_L \leq t_{i*} \leq Tc_{\max}/n_L$ for $i = 0, \ldots, n_L$.

$H_2^{*\prime}$: $A_{ij} = s_{ij}d$ for $i = 0, \ldots, n_L, j = 1, \ldots, n$ and $i \neq j$ where s_{ij} are unknown constants satisfying $0 < s_{\min} \leq s_{ij} \leq s_{\max} < \infty$ while d goes to 0 as n goes to infinity.

$H_3^{*\prime}$: There exists a set $V_i \subset \{n_L + 1, \ldots, n\}$ for $i = 1, \ldots, n_L$ with $|V_i| \geq vn$ such that $\tau = \min_{i=1,\ldots,n_L, i'=0,\ldots,n_L, i\neq i', j\in V_i}(s_{ij} - s_{i'j})$ is bounded away from 0.

$H_4^{*\prime}$: $A_{ii'}$ is bounded away from 1 for $i = 0, \ldots, n_L, i' = 1, \ldots, n_L$ and $i \neq i'$.

Theorem 5′. *Under $H_1^{*\prime}$–$H_4^{*\prime}$, if $\log T/(dTv) = o(1)$, $(\log d)^2/(dnv^2) = o(1)$ and $(\log T)^2/(dnv^2) = o(1)$, then*

$$T_e/T = o_p(1), \quad \text{as } n \to \infty, T \to \infty.$$

Proof of Theorem 6. By the same argument in Theorem 4, it is sufficient to show

$$\frac{(n_L + 1)T_e}{T} = o_p(1), \quad \text{as } n_L \to \infty, n \to \infty, T \to \infty. \tag{27}$$

From (24), we have shown

$$\sum_{i=1}^{n_L} \sum_{0\leq k\leq n_L, k\neq i} \frac{(n_L + 1)t_{ik}}{T} = o_p(1), \quad \text{as } n_L \to \infty, n \to \infty, T \to \infty.$$

From (25), there exists $\delta' > 0$ such that

$$L_P(z_*) - L_P(\hat{z}) \geq \max_{i=1,\ldots,n_L} \frac{dvnT}{\delta' n_L^3} \frac{n_L(n_L+1)t_{0i}}{T},$$

which further implies

$$\max_{i=1,\ldots,n_L} \frac{n_L(n_L+1)t_{0i}}{T} = o_p(1),$$

if $n_L^5 \log T/(dTv) = o(1)$, $(\log d)^2 n_L^8 \log n_L/(dnv^2) = o(1)$ and $(\log T)^2 n_L^8 \log n_L/(dnv^2) = o(1)$.

It follows that

$$\sum_{i=1}^{n_L} \frac{(n_L+1)t_{0i}}{T} = o_p(1).$$

Eq. (27) is therefore proved and so is the theorem. $\square$

Finally, we give the result for estimation consistency under the hub model with the null component with fixed n_L:

Theorem 6′. *Under $H_1^{*\prime} - H_4^{*\prime}$, if $\log n/T = o(1)$, $\log T/(dTv) = o(1)$, $\log T/(dTv) = o(1)$, $(\log d)^2/(dnv^2) = o(1)$ and $(\log T)^2/(dnv^2) = o(1)$, then*

$$\max_{i\in\{0,\ldots,n_L\}, j\in\{1,\ldots,n_L\}} |\hat{A}_{ij}^{\hat{z}} - A_{ij}| = o_p(1), \quad \text{as } n\to\infty, T\to\infty.$$

3 Additional Discussion of the Hub Model with the Null Component and Unknown Hub Set

We give a new identifiability result for the hub model with the null component and unknown hub set. Recall that V_0 is the true hub set with $|V_0| = n_L$. Let $\tilde{V}_0$ be another potential hub set with the corresponding parameters $(\tilde{\rho}, \tilde{A}) \in \mathcal{P}$ such that $\mathbb{P}(g|\rho, A) = \mathbb{P}(g|\tilde{\rho}, \tilde{A})$.

Theorem S1. *The parameters (ρ, A) of the hub model with the null component and unknown hub set are identifiable under the following conditions:*

(i′) $A_{ij} < 1$ *for* $i \in V_0 \cup \{0\}$ *and* $\tilde{A}_{ij} < 1$ *for* $i \in \tilde{V}_0 \cup \{0\}, j = 1, \ldots, n, j \neq i$*;*

(ii′) *for all* $i \in V_0, i' \in V_0, i \neq i'$*, there exists* $k \in V \setminus V_0$ *such that* $A_{ik} \neq A_{i'k}$*;*

(iii′) *for all* $i \in V_0$*, there exist* $k, k' \in V \setminus V_0$ *and* $k \neq k'$ *such that* $\pi_k \neq A_{ik}$ *and* $\pi_{k'} \neq A_{ik'}$*;*

(iv′) *there exists* $k \notin V_0 \cup \tilde{V}_0$ *such that for any* $i \in V_0$*,* $\pi_k \neq A_{ik}$*, and for any* $l \in \tilde{V}_0$*,* $\tilde{\pi}_k \neq \tilde{A}_{lk}$*.*

Conditions (i') - (iii') are identical to those in Theorem 1 and Theorem 2. Condition (iv') requires there exists at least one node that can only play a role as a follower.

Proof of Theorem S1. Theorem 2 shows when $V_0 = \tilde{V}_0$, the parameters in the hub model with null component are identifiable. Therefore, we only need to show $V_0 = \tilde{V}_0$ if $\mathbb{P}(g|\rho, A) = \mathbb{P}(g|\tilde{\rho}, \tilde{A})$ for all g.

Suppose there exist $(\tilde{\rho}, \tilde{A}) \neq (\rho, A)$ such that $\mathbb{P}(g|\rho, A) = \mathbb{P}(g|\tilde{\rho}, \tilde{A})$ for any g. Let $B_1 = \tilde{V}_0 \setminus V_0$ and $B_2 = V \setminus (V_0 \cup \tilde{V}_0)$. First, we consider the probability that no node appears

$$\rho_0 \prod_{j=1}^{n}(1-A_{0j}) = \tilde{\rho}_0 \prod_{j=1}^{n}(1-\tilde{A}_{0j}), \tag{28}$$

and the probability that only $k \in B_2$ appears,

$$\rho_0 A_{0k} \prod_{j\neq k}^{n}(1-A_{0j}) = \tilde{\rho}_0 \tilde{A}_{0k} \prod_{j\neq k}^{n}(1-\tilde{A}_{0j}). \tag{29}$$

Dividing (29) by (28), since $A_{0k} < 1$, we have $A_{0k} = \tilde{A}_{0k}$ for any $k \in B_2$.

Next we show that $B_1 = \tilde{V}_0 \setminus V_0 = \emptyset$. Suppose $B_1 \neq \emptyset$. By condition (iv'), for any $i \in B_1$, there exists a $k \in B_2$ such that $\tilde{A}_{0k} \neq \tilde{A}_{ik}$. Consider the probability that only i appears,

$$\tilde{\rho}_0 \tilde{A}_{0i} \prod_{j=1,\ldots,n, j\neq i}(1-\tilde{A}_{0j}) + \tilde{\rho}_i \prod_{j=1,\ldots,n, j\neq i}(1-\tilde{A}_{ij}) = \rho_0 A_{0i}(1-A_{0k}) \prod_{j=1,\ldots,n, j\notin\{i,k\}}(1-A_{0j}), \tag{30}$$

and the probability that only i and k appear

$$\tilde{\rho}_0 \tilde{A}_{0i} A_{0k} \prod_{j=1,\ldots,n, j\notin\{i,k\}}(1-\tilde{A}_{0j}) + \tilde{\rho}_i \tilde{A}_{ik} \prod_{j\notin\{i,k\}}(1-\tilde{A}_{ij}) = \rho_0 A_{0i} A_{0k} \prod_{j\notin\{i,k\}}(1-A_{0j}). \tag{31}$$

Let

$$\tilde{x} = \tilde{\rho}_0 \tilde{A}_{0i} \prod_{j=1,\ldots,n, j\notin\{i,k\}}(1-\tilde{A}_{0j}),$$

$$\tilde{y} = \tilde{\rho}_i \prod_{j=1,\ldots,n, j\notin\{i,k\}}(1-\tilde{A}_{ij}).$$

Then (30) and (31) can be viewed as a system of linear equations with unknown variables

$\tilde{x}$ and $\tilde{y}$:

$$\begin{pmatrix} A_{0k} & \tilde{A}_{ik} \\ 1 - A_{0k} & 1 - \tilde{A}_{ik} \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} \rho_0 A_{0i} A_{0k} \prod_{j=1,\ldots,n, j \notin \{i,k\}} (1 - A_{0j}) \\ \rho_0 A_{0i} (1 - A_{0k}) \prod_{j=1,\ldots,n, j \notin \{i,k\}} (1 - A_{0j}) \end{pmatrix}.$$

Since $A_{0k} = \tilde{A}_{0k} \neq \tilde{A}_{ik}$, the system is full rank and hence has a unique solution:

$$\begin{aligned} \tilde{\rho}_0 \tilde{A}_{0i} \prod_{j=1,\ldots,n, j \notin \{i,k\}} (1 - \tilde{A}_{0j}) &= \rho_0 A_{0i} \prod_{j=1,\ldots,n, j \notin \{i,k\}} (1 - A_{0j}), \\ \tilde{\rho}_i \prod_{j=1,\ldots,n, j \notin \{i,k\}} (1 - \tilde{A}_{ij}) &= 0. \end{aligned}$$

Combining with (28), we have

$$\tilde{\rho}_0 (1 - \tilde{A}_{0i}) \prod_{j=1,\ldots,n, j \notin \{i,k\}} (1 - \tilde{A}_{0j}) = \rho_0 (1 - A_{0i}) \prod_{j=1,\ldots,n, j \notin \{i,k\}} (1 - A_{0j}).$$

As $\tilde{A}_{ij} < 1$, $A_{0i} = \tilde{A}_{0i}$ for any $i \in B_1 \subset \tilde{V}_0$ and $\tilde{\rho}_i = 0$, which contradicts the assumption that $0 < \tilde{\rho}_i < 1$ for any $i \in \tilde{V}_0$. Therefore, $\tilde{V}_0 \setminus V_0 = \emptyset$ implies that $\tilde{V}_0$ does not contain any redundant component.

By the same argument, we obtain $A_{0i} = \tilde{A}_{0i}$ for any $i \in V_0 \setminus \tilde{V}_0$ and $\rho_i = 0$, which contradicts the assumption $0 < \rho_i < 1$ for $i \in V_0$. Therefore, $V_0 \setminus \tilde{V}_0 = \emptyset$. Hence, $V_0 = \tilde{V}_0$. By Theorem 2, we have $(\tilde{\rho}, \tilde{A}) = (\rho, A)$. □

We close this section by a discussion on how the penalized log-likelihood function (Section 3.2 in the main text) can result in sparse solutions. Maximizing the Lagrangian form of the penalized log-likelihood function is equivalent to maximizing $L(A, \rho)$ under the following constraints

$$\rho_i \geq 0,\ i = 0, 1, \ldots, M,\ \sum_{i=0}^{M} \rho_i = 1,\ \sum_{i=1}^{M} [\log(\epsilon + \rho_i) - \log \epsilon] \leq t.$$

To show how the constraints can result in sparse solutions, we consider a toy model containing only two nodes, both of which are potential hub set members, that is, $M = 2$. The constraints become

$$\rho_1 \geq 0,\ \rho_2 \geq 0,\ \rho_1 + \rho_2 \leq 1, \quad (32)$$
$$\log(1 + \frac{\rho_1}{\epsilon}) + \log(1 + \frac{\rho_2}{\epsilon}) \leq t.$$

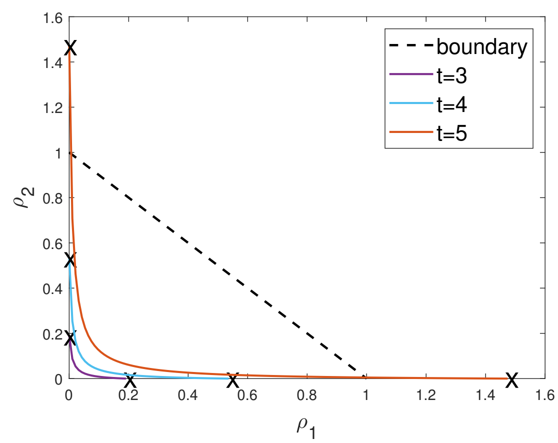

Figure 1: Feasible regions of the log penalty with different values of t.

Figure 1 shows the feasible regions of the log penalties for $t = 3, 4, 5$ and $\epsilon = 0.01$, where the crosses mark the intersection of $\log(1 + \rho_1/\epsilon) + \log(1 + \rho_2/\epsilon) = t$ and the axes, and the dashed line indicates $\rho_1 + \rho_2 = 1$. For $t = 3$ and 4, $\hat{\rho}_1$ (resp. $\hat{\rho}_2$) can potentially reach 0 with $\hat{\rho}_2$ (resp. $\hat{\rho}_1$) being non-zero, indicated by the cross markers within the region defined by (32). For $t = 5$ (corresponding to a smaller λ), this cannot happen because $\log(1 + \rho_1/\epsilon) + \log(1 + \rho_2/\epsilon) = 5$ intersects with the axes outside of the region defined by (32).

4 Additional Simulation Results

To further study the performance of the estimates under the setting of sparse A, we introduce a scale factor α to control the density of A. Specifically, $A_{ij} \sim U(0.2\alpha, 0.4\alpha)$ for $j \in V_i$ and $A_{ij} \sim U(0, 0.2\alpha)$ for $j \notin V_i$, where $\alpha = 0.1, 0.2, \ldots, 1$. We study how the ratios of the RMSEs when the hub labels are unknown to those when the hub labels are known i.e., $\mathrm{RMSE}(\hat{A}_{ij})/\mathrm{RMSE}^*$, change with the degree of sparsity. We present the results for the case when $n = 100$. Other simulation settings are the same with those in Section 4.1.

Figure 2 and 3 show the results of ratio versus α for the asymmetric hub model and the hub model with the null component, respectively. As α decreases, the ratio typically first increases and then decreases. This suggests that the estimators in both cases perform well when A is dense, and the problem becomes more difficult for the estimator with unknown hubs as A becomes sparser. However, when A becomes too sparse, the matrix A cannot be well estimated even for the case of known hub labels (i.e., the baseline).

Moreover, Figure 2 and 3 show that the turning point, i.e., the maximizer of the ratio, comes earlier when A is more difficult to estimate, which corresponds to the cases with larger n_L, smaller T, and the hub model with the null component. The turning point corresponds to the α value that gives the largest gap between the RMSE for the estimator

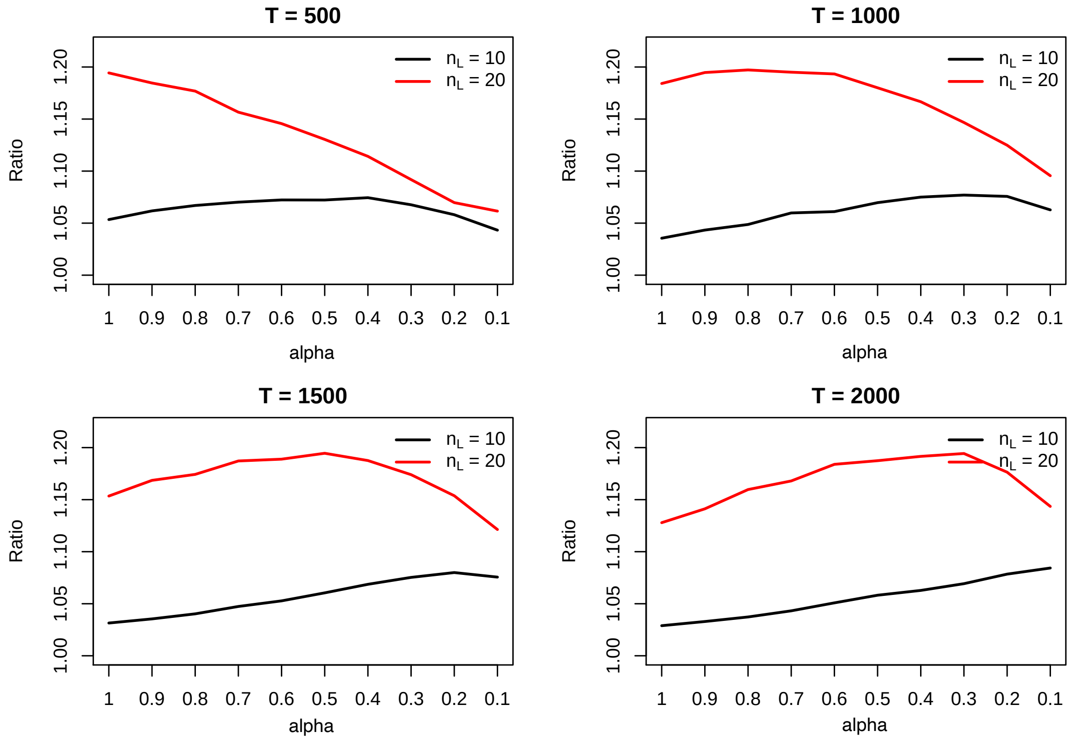

Figure 2: The asymmetric hub model results. The ratio is $\mathrm{RMSE}(\hat{A}_{ij})/\mathrm{RMSE}^*$.

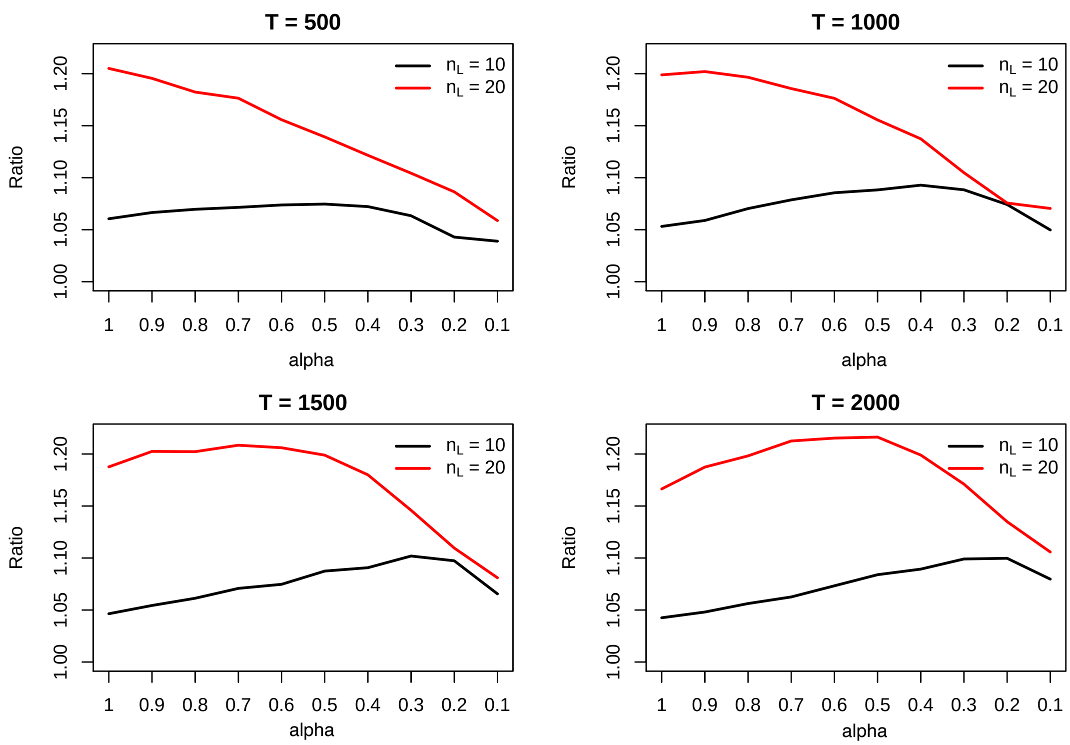

Figure 3: The hub model with the null component results. The ratio is $\mathrm{RMSE}(\hat{A}_{ij})/\mathrm{RMSE}^*$.

with unknown hub labels and the baseline, and when the settings become more difficult, the estimator with unknown hub labels starts to face challenges on a denser graph.

5 Additional Analysis of Passerine Data

We bootstrap 1,000 samples from the original data to evaluate the stability of the proposed hub set selection method. Specifically, we perform our method on each bootstrapped sample under λ from 0.045 to 0.065 and compute the proportion of each node being selected as a hub set member. Table 1 demonstrates the stability of the proposed method: the majority of the birds are not selected as a hub set member in any bootstrap sample, and v_9, v_{30} and v_{42}, the three birds identified from the original data dominate in the selection proportions across the bootstrapped samples.

6 Analysis of Extended Bakery Data

We apply the hub model with the null component to the extended bakery dataset (available at http://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery) to find the hub items and relationships among all the items. The dataset is a collection of purchases in a chain of bakery stores. The stores provide 50 items including 40 bakery goods (1-40) and 10 drinks (41-50). The goods can be divided into five categories: cakes (1-10), tarts (11-20), cookies (21-30) and pastries (31-40). Each purchase contains a collection of items bought together.

The extended bakery data was used as a benchmark dataset to test certain machine learning methods. For example, Agarwal and Nanavati (2016) used association rule mining

Table 1: Selection proportion from bootstrap

λ	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}	v_{15}
0.045	0	0	0	0.045	0	0	0.81	0	0.995	0.870	0	0	0	0	0
0.050	0	0	0	0.050	0	0	0	0	1	0.600	0	0	0	0	0
0.055	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0.060	0	0	0	0.005	0	0	0	0	0.965	0	0	0	0	0	0
0.065	0	0	0	0	0	0	0	0	0	0.005	0	0	0	0	0
λ	v_{16}	v_{17}	v_{18}	v_{19}	v_{20}	v_{21}	v_{22}	v_{23}	v_{24}	v_{25}	v_{26}	v_{27}	v_{28}	v_{29}	v_{30}
0.045	0.01	0	0	0	0.810	0	0	0	0.010	0.095	0.115	0.025	0	0	0.890
0.050	0	0	0	0	0.600	0	0	0	0.015	0.075	0.100	0.015	0	0	0.625
0.055	0	0	0	0	0.005	0	0	0	0	0.005	0	0.005	0	0	0.945
0.060	0	0	0	0	0.025	0	0	0	0	0	0.015	0.005	0	0	0.830
0.065	0	0	0	0	0.010	0	0	0	0	0	0.005	0	0	0	0.015
λ	v_{31}	v_{32}	v_{33}	v_{34}	v_{35}	v_{36}	v_{37}	v_{38}	v_{39}	v_{40}	v_{41}	v_{42}	v_{43}	v_{44}	v_{45}
0.045	0	0	0.825	0	0	0	0.830	0	0	0	0	0.965	0	0.005	0
0.050	0	0	0.625	0	0	0	0.105	0	0	0	0	0.935	0	0	0
0.055	0	0	0.010	0	0	0	0.015	0	0	0	0	0.985	0	0	0
0.060	0	0	0.040	0	0	0	0.020	0	0	0	0	0.910	0	0	0
0.065	0	0	0.045	0	0	0	0.050	0	0	0	0	0.080	0	0	0
λ	v_{46}	v_{47}	v_{48}	v_{49}	v_{50}	v_{51}	v_{52}	v_{53}	v_{54}	v_{55}					
0.045	0.845	0	0	0	0	0	0	0	0	0					
0.050	0.235	0	0	0	0	0	0	0	0	0					
0.055	0	0	0	0	0	0	0	0	0	0					
0.060	0	0	0	0	0	0	0	0	0	0					
0.065	0	0	0	0	0	0	0	0	0	0					

to extract the hidden relationships of items and Negahban et al. (2018) applied a multinomial logit (MNL) model to address the problem of collaboratively learning representations of the users and the items in recommendation systems.

In our experiment, we use the 5,000 receipts in the dataset. Since drinks are typically purchased as affiliated items of food, we use the 40 bakery goods as the potential hub set, i.e., $\bar{V}_0 = \{1, \ldots, 40\}$. We use $\lambda = 0.025, 0.030, \ldots, 0.045$ to estimate the hub set.

Table 2: Estimated hub set for extended bakery data

λ	Selected hub nodes								
0.025	1	4	5	6	12	13	25	29	33
0.030	1	4	5	15	23	29	33		
0.035	5	15	23	29	34				
0.040	15	16	23	29	34				
0.045	15	23	29	34					

Table 2 shows the estimated hub sets. As λ increases, nodes are removed gradually from the hub set. According to the BIC criteria, the optimal λ is 0.045, at which the estimated hub set contains v_{15}, v_{23}, v_{29} and v_{34}, where v_{15} is tart, v_{23} and v_{29} are cookies, and v_{34} is pastry.

In addition, if the data was fitted by the hub model without the null component, then the entire node set has to be used as the hub set. In fact, each of the 50 items was purchased individually for at least once, and therefore must serve as a hub if the hubless groups are not assumed. When the hub model with the null component is used, the corresponding items may be removed from the hub set, which greatly reduces the model complexity.

References

Agarwal, A. and Nanavati, N. (2016). Association rule mining using hybrid ga-pso for multi-objective optimisation. In *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–7. IEEE.

Choi, D. S., Wolfe, P. J., and Airoldi, E. M. (2012). Stochastic blockmodels with growing number of classes. *Biometrika*, 99(2):273–284.

Dubhashi, D. P. and Panconesi, A. (2009). *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press.

Negahban, S., Oh, S., Thekumparampil, K. K., and Xu, J. (2018). Learning from comparisons and choices. *The Journal of Machine Learning Research*, 19(1):1478–1572.