## A. Proof of Lemma 9

**Lemma 9** *Given log-concave survival functions and concave hazard functions in the parameter(s) of the pairwise transmission likelihoods, then, a sufficient condition for the Hessian matrix $\mathcal{Q}^n$ to be positive definite is that the hazard matrix $X^n(\boldsymbol{\alpha})$ is non-singular.*

**Proof** Using Eq. 5, the Hessian matrix can be expressed as a sum of two matrices, $\mathbf{D}^n(\boldsymbol{\alpha})$ and $\mathbf{X}^n(\boldsymbol{\alpha})\mathbf{X}^n(\boldsymbol{\alpha})^\top$. The matrix $\mathbf{D}^n(\boldsymbol{\alpha})$ is trivially positive semidefinite by log-concavity of the survival functions and concavity of the hazard functions. The matrix $\mathbf{X}^n(\boldsymbol{\alpha})\mathbf{X}^n(\boldsymbol{\alpha})^\top$ is positive definite matrix since $\mathbf{X}^n(\boldsymbol{\alpha})$ is full rank by assumption. Then, the Hessian matrix is positive definite since it is a sum a positive semidefinite matrix and a positive definite matrix. ■

## B. Proof of Lemma 10

**Lemma 10** *If the source probability $\mathbb{P}(s)$ is strictly positive for all $s \in \mathcal{R}$, then, for an arbitrarily large number of cascades $n \to \infty$, there exists an ordering of the nodes and cascades within the cascade set such that the hazard matrix $\boldsymbol{X}^n(\boldsymbol{\alpha})$ is non-singular.*

**Proof** In this proof, we find a labeling of the nodes (row indices in $\mathbf{X}^n(\boldsymbol{\alpha})$) and ordering of the cascades (column indices in $\mathbf{X}^n(\boldsymbol{\alpha})$), such that, for an arbitrary large number of cascades, we can express the matrix $\mathbf{X}^n(\boldsymbol{\alpha})$ as $[T\, B]$, where $T \in \mathbb{R}^{p\times p}$ is an upper triangular with nonzero diagonal elements and $B \in \mathbb{R}^{p\times n-p}$. And, therefore, $\mathbf{X}^n(\boldsymbol{\alpha})$ has full rank (rank $p$). We proceed first by sorting nodes in $\mathcal{R}$ and then continue by sorting nodes in $\mathcal{U}$:

- **Nodes in $\mathcal{R}$**: For each node $u \in \mathcal{R}$, consider the set of cascades $C_u$ in which $u$ was a source and $i$ got infected. Then, rank each node $u$ according to the earliest position in which node $i$ got infected across all cascades in $C_u$ in decreasing order, breaking ties at random. For example, if a node $u$ was, at least once, the source of a cascade in which node $i$ got infected just after the source, but in contrast, node $v$ was never the source of a cascade in which node $i$ got infected the second, then node $u$ will have a lower index than node $v$. Then, assign row $k$ in the matrix $\mathbf{X}^n(\boldsymbol{\alpha})$ to node in position $k$ and assign the first $d$ columns to the corresponding cascades in which node $i$ got infected earlier. In such ordering, $\mathbf{X}^n(\boldsymbol{\alpha})_{mk} = 0$ for all $m < k$ and $\mathbf{X}^n(\boldsymbol{\alpha})_{kk} \neq 0$.
- **Nodes in $\mathcal{U}$**: Similarly as in the first step, and assign them the rows $d+1$ to $p$. Moreover, we assign the columns $d+1$ to $p$ to the corresponding cascades in which node $i$ got infected earlier. Again, this ordering satisfies that $\mathbf{X}^n(\boldsymbol{\alpha})_{mk} = 0$ for all $m < k$ and $\mathbf{X}^n(\boldsymbol{\alpha})_{kk} \neq 0$. Finally, the remaining columns $n-p$ can be assigned to the remaining cascades at random.

This ordering leads to the desired structure $[T\, B]$, and thus it is non-singular. ■

## C. Proof of Eq 7.

If the Hazard vector $\boldsymbol{X}(\mathbf{t}^c; \boldsymbol{\alpha})$ is Lipschitz continuous in the domain $\{\boldsymbol{\alpha} : \boldsymbol{\alpha}_S \geq \frac{\alpha^*_{\min}}{2}\}$,

$$\|\boldsymbol{X}(\mathbf{t}^c; \boldsymbol{\beta}) - \boldsymbol{X}(\mathbf{t}^c; \boldsymbol{\alpha})\|_2 \leq k_1\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2,$$

where $k_1$ is some positive constant. Then, we can bound the spectral norm of the difference, $\frac{1}{\sqrt{n}}(\boldsymbol{X}^n(\boldsymbol{\beta}) - \boldsymbol{X}^n(\boldsymbol{\alpha}))$, in the domain $\{\boldsymbol{\alpha} : \boldsymbol{\alpha}_S \geq \frac{\alpha^*_{\min}}{2}\}$ as follows:

$$\begin{aligned}
&|||\frac{1}{\sqrt{n}}\big(\mathbf{X}^n(\boldsymbol{\beta}) - \mathbf{X}^n(\boldsymbol{\alpha})\big)|||_2 \\
&= \max_{\|\boldsymbol{u}\|_2=1} \frac{1}{\sqrt{n}}\|\boldsymbol{u}\big(\mathbf{X}^n(\boldsymbol{\beta}) - \mathbf{X}^n(\boldsymbol{\alpha})\big)\|_2 \\
&= \max_{\|\boldsymbol{u}\|_2=1} \frac{1}{\sqrt{n}}\sqrt{\sum_{c=1}^{n}\langle \boldsymbol{u}, \mathbf{X}(\mathbf{t}^c; \boldsymbol{\beta}) - \mathbf{X}(\mathbf{t}^c; \boldsymbol{\alpha})\rangle^2} \\
&\leq \frac{1}{\sqrt{n}}\sqrt{k_1^2 n\|\boldsymbol{u}\|_2^2\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2} \\
&\leq k_1\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2.
\end{aligned}$$

## D. Proof of Lemma 3

By Lagrangian duality, the regularized network inference problem defined in Eq. 4 is equivalent to the following constrained optimization problem:

$$\begin{array}{ll}
\text{minimize}_{\boldsymbol{\alpha}_i} & \ell^n(\boldsymbol{\alpha}_i) \\
\text{subject to} & \alpha_{ji} \geq 0,\ j = 1, \ldots, N, i \neq j, \\
& ||\boldsymbol{\alpha}_i||_1 \leq C(\lambda_n)
\end{array} \tag{20}$$

where $C(\lambda_n) < \infty$ is a positive constant. In this alternative formulation, $\lambda_n$ is the Lagrange multiplier for the second constraint. Since $\lambda_n$ is strictly positive, the constraint is active at any optimal solution, and thus $||\boldsymbol{\alpha}_i||_1$ is constant across all optimal solutions.

Using that $\ell^n(\boldsymbol{\alpha}_i)$ is a differentiable convex function by assumption and $\{\boldsymbol{\alpha} : \alpha_{ji} \geq 0, ||\boldsymbol{\alpha}_i||_1 \leq C(\lambda_n)\}$ is a convex set, we have that $\nabla\ell^n(\boldsymbol{\alpha}_i)$ is constant across optimal primal solutions (Mangasarian, 1988). Moreover, any optimal primal-dual solution in the original problem must satisfy the KKT conditions in the alternative formulation defined by Eq. 20, in particular,

$$\nabla\ell^n(\boldsymbol{\alpha}_i) = -\lambda_n\mathbf{z} + \boldsymbol{\mu},$$

where $\boldsymbol{\mu} \geq 0$ are the Lagrange multipliers associated to the non negativity constraints and $\mathbf{z}$ denotes the subgradient of the $\ell 1$-norm.

Consider the solution $\hat{\boldsymbol{\alpha}}$ such that $||\hat{\mathbf{z}}_{\mathbf{S^c}}||_\infty < 1$ and thus $\nabla_{\alpha_{S^c}} \ell^n(\hat{\boldsymbol{\alpha}}_i) = -\lambda_n \hat{\mathbf{z}}_{\mathbf{S^c}} + \hat{\boldsymbol{\mu}}_{S^c}$. Now, assume there is an optimal primal solution $\tilde{\boldsymbol{\alpha}}$ such that $\tilde{\alpha}_{ji} > 0$ for some $j \in S^c$, then, using that the gradient must be constant across optimal solutions, it should hold that $-\lambda_n \hat{z}_j + \hat{\mu}_j = -\lambda_n$, where $\tilde{\mu}_{ji} = 0$ by complementary slackness, which implies $\hat{\mu}_j = -\lambda_n(1 - \hat{z}_j) < 0$. Since $\hat{\mu}_j \geq 0$ by assumption, this leads to a contradiction. Then, any primal solution $\tilde{\boldsymbol{\alpha}}$ must satisfy $\tilde{\boldsymbol{\alpha}}_{S^c} = 0$ for the gradient to be constant across optimal solutions.

Finally, since $\boldsymbol{\alpha}_{S^c} = 0$ for all optimal solutions, we can consider the restricted optimization problem defined in Eq. 17. If the Hessian sub-matrix $[\nabla^2 L(\hat{\boldsymbol{\alpha}})]_{SS}$ is strictly positive definite, then this restricted optimization problem is strictly convex and the optimal solution must be unique.

## E. Proof of Lemma 4

To prove this lemma, we will first construct a function

$$G(\mathbf{u}_S) := \ell^n(\boldsymbol{\alpha}^*_S + \mathbf{u}_S) - \ell^n(\boldsymbol{\alpha}^*_S) + \lambda_n(\|\boldsymbol{\alpha}^*_S + \mathbf{u}_S\|_1 - \|\boldsymbol{\alpha}^*_S\|_1).$$

whose domain is restricted to the convex set $\mathcal{U} = \{\mathbf{u}_S : \boldsymbol{\alpha}^*_S + \mathbf{u}_S \geq \mathbf{0}\}$. By construction, $G(\mathbf{u}_S)$ has the following properties

1. It is convex with respect to $\mathbf{u}_S$.
2. Its minimum is obtained at $\hat{\mathbf{u}}_S := \hat{\boldsymbol{\alpha}}_S - \boldsymbol{\alpha}^*_S$. That is $G(\hat{\mathbf{u}}_S) \leq G(\mathbf{u}_S), \forall \mathbf{u}_S \neq \hat{\mathbf{u}}_S$.
3. $G(\hat{\mathbf{u}}_S) \leq G(\mathbf{0}) = 0$.

Based on property 1 and 3, we deduce that any point in the segment, $\mathbb{L} := \{\tilde{\mathbf{u}}_S : \tilde{\mathbf{u}}_S = t\hat{\mathbf{u}}_S + (1-t)\mathbf{0}, t \in [0,1]\}$, connecting $\hat{\mathbf{u}}_S$ and $\mathbf{0}$ has $G(\tilde{\mathbf{u}}_S) \leq 0$. That is

$$\begin{aligned} G(\tilde{\mathbf{u}}_S) &= G(t\hat{\mathbf{u}}_S + (1-t)\mathbf{0}) \\ &\leq tG(\hat{\mathbf{u}}_S) + (1-t)G(\mathbf{0}) \leq 0. \end{aligned}$$

Next, we will find a sphere centered at $\mathbf{0}$ with strictly positive radius $B$, $\mathbb{S}(B) := \{\boldsymbol{u}_S : \|\boldsymbol{u}_S\|_2 = B\}$, such that function $G(\boldsymbol{u}_S) > 0$ (strictly positive) on $\mathbb{S}(B)$. We note that this sphere $\mathbb{S}(B)$ can not intersect with the segment $\mathbb{L}$ since the two sets have strictly different function values. Furthermore, the only possible configuration is that the segment is contained inside the sphere entirely, leading us to conclude that the end point $\hat{\mathbf{u}}_S := \hat{\boldsymbol{\alpha}}_S - \boldsymbol{\alpha}^*_S$ is also within the sphere. That is $\|\hat{\boldsymbol{\alpha}}_S - \boldsymbol{\alpha}^*_S\|_2 \leq B$.

In the following, we will provide details on finding such a suitable $B$ which will be a function of the regularization parameter $\lambda_n$ and the neighborhood size $d$. More specifically, we will start by applying a Taylor series expansion and the mean value theorem,

$$\begin{aligned} G(\mathbf{u}_S) = {} & \nabla_S \ell^n(\boldsymbol{\alpha}^*_S)^\top \mathbf{u}_S \\ & + \mathbf{u}_S^\top \nabla^2_{SS} \ell^n(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S)\mathbf{u}_S \\ & + \lambda_n(\|\boldsymbol{\alpha}^*_S + \mathbf{u}_S\|_1 - \|\boldsymbol{\alpha}^*_S\|_1), \end{aligned} \tag{21}$$

where $b \in [0,1]$. We will show that $G(\mathbf{u}_S) > 0$ by bounding below each term of above equation separately.

We bound the absolute value of the first term using the assumption on the gradient, $\nabla_S \ell(\cdot)$,

$$\begin{aligned} |\nabla_S \ell^n(\boldsymbol{\alpha}^*_S)^\top \mathbf{u}_S| &\leq \|\nabla_S \ell\|_\infty \|\mathbf{u}_S\|_1 \\ &\leq \|\nabla_S \ell\|_\infty \sqrt{d} \|\mathbf{u}_S\|_2 \\ &\leq 4^{-1}\lambda_n B\sqrt{d}. \end{aligned} \tag{22}$$

We bound the absolute value of the last term using the reverse triangle inequality.

$$\lambda_n |\|\boldsymbol{\alpha}^*_S + \mathbf{u}_S\|_1 - \|\boldsymbol{\alpha}^*_S\|_1| \leq \lambda_n \|\mathbf{u}_S\|_1 \leq \lambda_n \sqrt{d} \|\mathbf{u}_S\|_2. \tag{23}$$

Bounding the remaining middle term is more challenging. We start by rewriting the Hessian as a sum of two matrices, using Eq. 5,

$$\begin{aligned} q = {} & \min_{\mathbf{u}_S} \mathbf{u}_S^\top \mathbf{D}^n_{SS}(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S)\mathbf{u}_S \\ & + n^{-1}\mathbf{u}_S^\top \mathbf{X}^n_S(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S)\mathbf{X}^n_S(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S)^\top \mathbf{u}_S \\ = {} & \min_{\mathbf{u}_S} \mathbf{u}_S^\top \mathbf{D}^n_{SS}(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S)\mathbf{u}_S + \|\mathbf{u}_S^\top \mathbf{X}^n_S(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S)\|_2^2. \end{aligned}$$

Now, we introduce two additional quantities,

$$\begin{aligned} \Delta\mathbf{D}^n_{SS} &= \mathbf{D}^n_{SS}(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S) - \mathbf{D}^n_{SS}(\boldsymbol{\alpha}^*_S) \\ \Delta\mathbf{X}^n_S &= \mathbf{X}^n_S(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S) - \mathbf{X}^n_S(\boldsymbol{\alpha}^*_S), \end{aligned}$$

and rewrite $q$ as

$$\begin{aligned} q \;= {} & \min_{\mathbf{u}_S} \big[\mathbf{u}_S^\top \mathbf{D}^n_{SS}(\boldsymbol{\alpha}^*_S)\mathbf{u}_S + n^{-1}\|\mathbf{u}_S^\top \mathbf{X}^n_S(\boldsymbol{\alpha}^*_S)\|_2^2 \\ & + n^{-1}\|\mathbf{u}_S^\top \Delta\mathbf{X}^n_S\|_2^2 + \mathbf{u}_S^\top \Delta\mathbf{D}^n_{SS}\mathbf{u}_S \\ & + 2n^{-1}\langle \mathbf{u}_S^\top \mathbf{X}^n_S(\boldsymbol{\alpha}^*_S), \mathbf{u}_S^\top \Delta\mathbf{X}^n_S\rangle\big]. \end{aligned}$$

Next, we use dependency condition,

$$\begin{aligned} q \geq {} & C_{\min}B^2 - \max_{\mathbf{u}_S} |\underbrace{\mathbf{u}_S^\top \Delta\mathbf{D}^n_{SS}\mathbf{u}_S}_{T_1}| \\ & - \max_{\mathbf{u}_S} 2|\underbrace{n^{-1}\langle \mathbf{u}_S^\top \mathbf{X}^n_S(\boldsymbol{\alpha}^*_S), \mathbf{u}_S^\top \Delta\mathbf{X}^n_S\rangle}_{T_2}|, \end{aligned}$$

and proceed to bound $T_1$ and $T_2$ separately. First, we bound $T_1$ using the Lipschitz condition,

$$\begin{aligned} |T_1| &= |\sum_{k\in S} u_k^2[\mathbf{D}^n_k(\boldsymbol{\alpha}^*_S + b\mathbf{u}_S) - \mathbf{D}^n_k(\boldsymbol{\alpha}^*_S)]| \\ &\leq \sum_{k\in S} u_k^2 k_2 \|b\mathbf{u}_S\|_2 \\ &\leq k_2 B^3. \end{aligned}$$

Then, we use the dependency condition, the Lipschitz condition and the Cauchy-Schwartz inequality to bound $T_2$,

$$\begin{aligned} T_2 &\leq \frac{1}{\sqrt{n}}\|\mathbf{u}_S^\top \mathbf{X}_S^n(\boldsymbol{\alpha}_S^*)\|_2 \frac{1}{\sqrt{n}}\|\mathbf{u}_S^\top \Delta \mathbf{X}_S^n\|_2 \\ &\leq \sqrt{C_{\max}} B \frac{1}{\sqrt{n}}\|\mathbf{u}_S^\top \Delta \mathbf{X}_S^n\|_2 \\ &\leq \sqrt{C_{\max}} B \|\mathbf{u}_S\|_2 \frac{1}{\sqrt{n}} |||\Delta \mathbf{X}_S^n|||_2 \\ &\leq \sqrt{C_{\max}} B^2 k_1 \|b\mathbf{u}_S\|_2 \\ &\leq k_1 \sqrt{C_{\max}} B^3, \end{aligned}$$

where we note that applying the Lipschitz condition implies assuming $B < \frac{\alpha_{min}}{2}$. Next, we incorporate the bounds of $T_1$ and $T_2$ to lower bound $q$,

$$q \geq C_{\min} B^2 - (k_2 + 2k_1\sqrt{C_{\max}})B^3. \tag{24}$$

Now, we set $B = K\lambda_n\sqrt{d}$, where $K$ is a constant that we will set later in the proof, and select the regularization parameter $\lambda_n$ to satisfy $\lambda_n\sqrt{d} \leq 0.5C_{\min}/K(k_2 + 2k_1\sqrt{C_{\max}})$. Then,

$$\begin{aligned} G(\mathbf{u}_S) &\geq -4^{-1}\lambda_n\sqrt{d}B + 0.5C_{\min}B^2 - \lambda_n\sqrt{d}B \\ &\geq B(0.5C_{\min}B - 1.25\lambda_n\sqrt{d}) \\ &\geq B(0.5C_{\min}K\lambda_n\sqrt{d} - 1.25\lambda_n\sqrt{d}). \end{aligned}$$

In the last step, we set the constant $K = 3C_{\min}^{-1}$, and we have

$$G(\mathbf{u}_S) \geq 0.25\lambda_n\sqrt{d} > 0,$$

as long as

$$\begin{aligned} \sqrt{d}\lambda_n &\leq \frac{C_{\min}^2}{6(k_2 + 2k_1\sqrt{C_{\max}})} \\ \alpha_{\min}^* &\geq \frac{6\lambda_n\sqrt{d}}{C_{\min}}. \end{aligned}$$

Finally, convexity of $G(\mathbf{u}_S)$ yields

$$\|\hat{\boldsymbol{\alpha}}_S - \boldsymbol{\alpha}_S^*\|_2 \leq 3\lambda_n\sqrt{d}/C_{\min} \leq \frac{\alpha_{\min}^*}{2}.$$

## F. Proof of Lemma 5

Define $z_j^c = [\nabla g(\mathbf{t}^c; \boldsymbol{\alpha}^*)]_j$ and $z_j = \frac{1}{n}\sum_c z_j^c$. Now, using the KKT conditions and condition 4 (Boundedness), we have that $\mu_j^* = \mathbb{E}_c\{z_j^c\}$ and $|z_j^c| \leq k_3$, respectively. Thus, Hoeffding's inequality yields

$$P(|z_j - \mu_j^*| > \frac{\lambda_n\varepsilon}{4(2-\varepsilon)}) \leq 2\exp\left(-\frac{n\lambda_n^2\varepsilon^2}{32k_3^2\,(2-\varepsilon))^2}\right),$$

and then,

$$P(\|\boldsymbol{z} - \boldsymbol{\mu}^*\|_\infty > \frac{\lambda_n\varepsilon}{4(2-\varepsilon)}) \leq 2\exp\left(-\frac{n\lambda_n^2\varepsilon^2}{32k_3^2\,(2-\varepsilon))^2} + \log p\right).$$

## G. Proof of Lemma 6

We start by factorizing the Hessian matrix, using Eq. 5,

$$R_j^n = \left[\nabla^2\ell^n(\bar{\boldsymbol{\alpha}}_j) - \nabla^2\ell^n(\boldsymbol{\alpha}^*)\right]_j^\top (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) = \omega_j^n + \delta_j^n,$$

where,

$$\begin{aligned} \omega_j^n &= \left[\mathbf{D}^n(\bar{\boldsymbol{\alpha}}_j) - \mathbf{D}^n(\boldsymbol{\alpha}^*)\right]_j^\top (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \\ \delta_j^n &= \frac{1}{n}\boldsymbol{V}_j^n(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \\ \boldsymbol{V}_j^n &= [\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)]_j \mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)^\top - [\mathbf{X}^n(\boldsymbol{\alpha}^*)]_j \mathbf{X}^n(\boldsymbol{\alpha}^*)^\top. \end{aligned}$$

Next, we proceed to bound each term separately. Since $[\bar{\boldsymbol{\alpha}}_j]_S = \theta_j\hat{\boldsymbol{\alpha}}_S + (1-\theta_j)\boldsymbol{\alpha}_S^*$ where $\theta_j \in [0,1]$, and $\|\hat{\boldsymbol{\alpha}}_S - \boldsymbol{\alpha}_S^*\|_\infty \leq \frac{\alpha_{\min}^*}{2}$ (Lemma 4), it holds that $[\bar{\boldsymbol{\alpha}}_j]_S \geq \frac{\alpha_{\min}^*}{2}$. Then, we can use condition 3 (Lipschitz Continuity) to bound $\omega_j^n$.

$$\begin{aligned} |\omega_j^n| &\leq k_1\|\bar{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}^*\|_2 \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2 \\ &\leq k_1\theta_j\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2 \\ &\leq k_1\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2. \end{aligned} \tag{25}$$

However, bounding term $\delta_j^n$ is more difficult. Let us start by rewriting $\delta_j^n$ as follows.

$$\delta_j^n = (\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 + \boldsymbol{\Lambda}_3)(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*),$$

where,

$$\begin{aligned} \boldsymbol{\Lambda}_1 &= [\mathbf{X}^n(\boldsymbol{\alpha}^*)]_j (\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)^\top - \mathbf{X}^n(\boldsymbol{\alpha}^*)^\top) \\ \boldsymbol{\Lambda}_2 &= \{[\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)]_j - [\mathbf{X}^n(\boldsymbol{\alpha}^*)]_j\}(\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)^\top - \mathbf{X}^n(\boldsymbol{\alpha}^*)^\top) \\ \boldsymbol{\Lambda}_3 &= \big([\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)]_j - [\mathbf{X}^n(\boldsymbol{\alpha}^*)]_j\big)\mathbf{X}^n(\boldsymbol{\alpha}^*)^\top. \end{aligned}$$

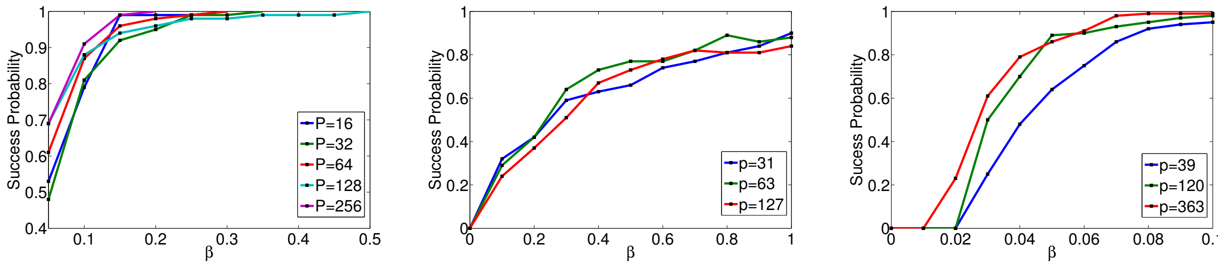
Next, we bound each term separately. For the first term, we first apply Cauchy inequality,

$$\begin{aligned} |\boldsymbol{\Lambda}_1(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)| \leq \|&[\mathbf{X}^n(\boldsymbol{\alpha}^*)]_j\|_2 \\ &\times |||\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)^\top - \mathbf{X}^n(\boldsymbol{\alpha}^*)^\top|||_2 \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2, \end{aligned}$$

and then use condition 3 (Lipschtiz Continuity) and 4 (Boundedness),

$$\begin{aligned} |\boldsymbol{\Lambda}_1(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)| &\leq nk_4k_1\|\bar{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}^*\|_2 \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2 \\ &\leq nk_4k_1\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2. \end{aligned}$$

For the second term, we also start by applying Cauchy in-




(a) Chain ($d_i = 1$)

(b) Stars with different # of leaves ($d_i = 1$)

(c) Tree ($d_i = 3$)

*Figure 5.* Success probability vs. # of cascades. Different super-neighborhood sizes $p_i$.

equality,

$$|\mathbf{\Lambda}_2(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)| \leq \|[\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)]_j - [\mathbf{X}^n(\boldsymbol{\alpha}^*)]_j\|_2 \times |||\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)^\top - \mathbf{X}^n(\boldsymbol{\alpha}^*)^\top|||_2 \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2,$$

and then use condition 3 (Lipschtiz Continuity),

$$|\mathbf{\Lambda}_2(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)| \leq n k_1^2 \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2.$$

Last, for third term, once more we start by applying Cauchy inequality,

$$|\mathbf{\Lambda}_3(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)| \leq \|[\mathbf{X}^n(\bar{\boldsymbol{\alpha}}_j)]_j - [\mathbf{X}^n(\boldsymbol{\alpha}^*)]_j\|_2 \times |||\mathbf{X}^n(\boldsymbol{\alpha}^*)^\top|||_2 \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2,$$

and then apply condition 1 (Dependency Condition) and condition 3 (Lipschitz Continuity),

$$|\mathbf{\Lambda}_3(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)| \leq n k_1 \sqrt{C_{\max}} \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2$$

Now, we combine the bounds,

$$\|\boldsymbol{R}^n\|_\infty \leq K \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2,$$

where

$$K = k_1 + k_4 k_1 + k_1^2 + k_1 \sqrt{C_{\max}}.$$

Finally, using Lemma 4 and selecting the regularization parameter $\lambda_n$ to satisfy $\lambda_n d \leq C_{\min}^2 \frac{\varepsilon}{36K(2-\varepsilon)}$ yields:

$$\begin{aligned} \|\boldsymbol{R}^n\|_\infty / \lambda_n &\leq 3K \lambda_n d / C_{\min}^2 \\ &\leq \frac{\varepsilon}{4(2-\varepsilon)} \end{aligned}$$

## H. Proof of Lemma 7

We will first bound the difference in terms of nuclear norm between the population Fisher information matrix $\mathcal{Q}_{SS}$ and the sample mean cascade log-likelihood $\mathcal{Q}_{SS}^n$. Define $z_{jk}^c = [\nabla^2 g(\mathbf{t}^c; \boldsymbol{\alpha}^*) - \nabla^2 \ell^n(\boldsymbol{\alpha}^*)]_{jk}$ and $z_{jk} = \frac{1}{n} \sum_{c=1}^n z_{jk}^c$. Then, we can express the difference between the population Fisher information matrix $\mathcal{Q}_{SS}$ and the sample mean cascade log-likelihood $\mathcal{Q}_{SS}^n$ as:

$$\begin{aligned} |||\mathcal{Q}_{SS}^n(\boldsymbol{\alpha}^*) - \mathcal{Q}_{SS}^*(\boldsymbol{\alpha}^*)|||_2 &\leq |||\mathcal{Q}_{SS}^n(\boldsymbol{\alpha}^*) - \mathcal{Q}_{SS}^*(\boldsymbol{\alpha}^*)|||_F \\ &= \sqrt{\sum_{j=1}^d \sum_{k=1}^d (z_{jk})^2}. \end{aligned}$$

Since $|z_{jk}^{(c)}| \leq 2k_5$ by condition 4, we can apply Hoeffding's inequality to each $z_{jk}$,

$$P(|z_{jk}| \geq \beta) \leq 2 \exp\left(-\frac{\beta^2 n}{8k_5^2}\right), \tag{26}$$

and further,

$$P(|||\mathcal{Q}_{SS}^n(\boldsymbol{\alpha}^*) - \mathcal{Q}_{SS}^*(\boldsymbol{\alpha}^*)|||_2 \geq \delta) \leq 2 \exp\big(-K \frac{\delta^2 n}{d^2} + 2 \log d\big) \tag{27}$$

where $\beta^2 = \delta^2/d^2$. Now, we bound the maximum eigenvalue of $\mathcal{Q}_{SS}^n$ as follows:

$$\begin{aligned} \Lambda_{\max}(\mathcal{Q}_{SS}^n) &= \max_{\|x\|_2=1} x^\top \mathcal{Q}_{SS}^n x \\ &= \max_{\|x\|_2=1} \{x^\top \mathcal{Q}_{SS}^* x + x^\top (\mathcal{Q}_{SS}^n - \mathcal{Q}_{SS}^*) x\} \\ &\leq y^\top \mathcal{Q}_{SS}^* y + y^\top (\mathcal{Q}_{SS}^n - \mathcal{Q}_{SS}^*) y, \end{aligned}$$

where $y$ is unit-norm maximal eigenvector of $\mathcal{Q}_{SS}^*$. Therefore,

$$\Lambda_{\max}(\mathcal{Q}_{SS}^n) \leq \Lambda_{\max}(\mathcal{Q}_{SS}^*) + |||\mathcal{Q}_{SS}^n - \mathcal{Q}_{SS}^*|||_2,$$

and thus,

$$P\big(\Lambda_{\max}(\mathcal{Q}_{SS}^n) \geq C_{\max} + \delta\big) \leq \exp\left(-K \frac{\delta^2 n}{d^2} + 2 \log d\right).$$

Reasoning in a similar way, we bound the minimum eigen-

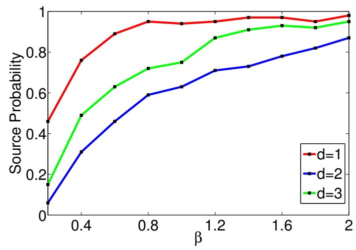



*Figure 6.* Success probability vs. # of cascades. Different in-degrees $d_i$.

value of $\mathcal{Q}^n_{SS}$:

$$P\big(\Lambda_{\min}(\mathcal{Q}^n_{SS}) \leq C_{\min} - \delta\big) \leq \exp\left(-K\frac{\delta^2 n}{d^2} + 2\log d\right)$$

## I. Proof of Lemma 8

We start by decomposing $\mathcal{Q}^n_{S^cS}(\boldsymbol{\alpha}^*)(\mathcal{Q}^n_{S^cS}(\boldsymbol{\alpha}^*))^{-1}$ as follows:

$$\mathcal{Q}^n_{S^cS}(\boldsymbol{\alpha}^*)(\mathcal{Q}^n_{S^cS}(\boldsymbol{\alpha}^*))^{-1} = A_1 + A_2 + A_3 + A_4,$$

where,

$$\begin{aligned} A_1 &= \mathcal{Q}^*_{S^cS}[(\mathcal{Q}^n_{S^cS})^{-1} - (\mathcal{Q}^*_{S^cS})^{-1}], \\ A_2 &= [\mathcal{Q}^n_{S^cS} - \mathcal{Q}^*_{S^cS}][(\mathcal{Q}^n_{S^cS})^{-1} - (\mathcal{Q}^*_{S^cS})^{-1}] \\ A_3 &= [\mathcal{Q}^n_{S^cS} - \mathcal{Q}^*_{S^cS}](\mathcal{Q}^*_{SS})^{-1}, \\ A_4 &= \mathcal{Q}^*_{S^cS}(\mathcal{Q}^*_{SS})^{-1}, \end{aligned}$$

$\mathcal{Q}^* = \mathcal{Q}^*(\boldsymbol{\alpha}^*)$ and $\mathcal{Q}^n = \mathcal{Q}^n(\boldsymbol{\alpha}^*)$. Now, we bound each term separately. The fourth term, $A_4$, is the easiest to bound, using simply the incoherence condition:

$$|||A_4|||_\infty \leq 1 - \varepsilon.$$

To bound the other terms, we need the following lemma:

**Lemma 11** *For any $\delta \geq 0$ and constants $K$ and $K'$, the following bounds hold:*

$$P[|||\mathcal{Q}^n_{S^cS} - \mathcal{Q}^*_{S^cS}|||_\infty \geq \delta] \leq 2\exp\left(-K\frac{n\delta^2}{d^2} + \log d + log(p-d)\right) \quad (28)$$

$$P[|||\mathcal{Q}^n_{SS} - \mathcal{Q}^*_{SS}|||_\infty \geq \delta] \leq 2\exp\left(-K\frac{n\delta^2}{d^2} + 2\log d\right) \quad (29)$$

$$P[|||(\mathcal{Q}^n_{SS})^{-1} - (\mathcal{Q}^*_{SS})^{-1}|||_\infty \geq \delta] \leq 4\exp\left(-K\frac{n\delta}{d^3} - K'\log d\right) \quad (30)$$

**Proof** We start by proving the first confidence interval. By definition of infinity norm of a matrix, we have:

$$\begin{aligned} &P[|||\mathcal{Q}^n_{S^cS} - \mathcal{Q}^*_{S^cS}|||_\infty \geq \delta] \\ &\quad = P\big[\max_{j\in S^c}\sum_{k\in S}|z_{jk}| \geq \delta\big] \\ &\quad \leq (p-d)P\big[\sum_{k\in S}|z_{jk}| \geq \delta\big], \end{aligned}$$

where $z_{jk} = [\mathcal{Q}^n - \mathcal{Q}^*]_{jk}$ and, for the last inequality, we used the union bound and the fact that $|S^c| \leq p - d$. Furthermore,

$$\begin{aligned} P\big[\textstyle\sum_{k\in S}|z_{jk}| \geq \delta\big] &\leq P[\exists k \in S||z_{jk}| \geq \delta/d] \\ &\leq dP[|z_{jk}| \geq \delta/d]. \end{aligned}$$

Thus,

$$P[|||\mathcal{Q}^n_{S^cS} - \mathcal{Q}^*_{S^cS}|||_\infty \geq \delta] \leq (p-d)dP[|z_{jk}| \geq \delta/d].$$

At this point, we can obtain the first confidence bound by using Eq. 26 with $\beta = \delta/d$ in the above equation. The proof of the second confidence bound is very similar and we omit it for brevity. To prove the last confidence bound, we proceed as follows:

$$\begin{aligned} &|||(\mathcal{Q}^n_{SS})^{-1} - (\mathcal{Q}^*_{SS})^{-1}|||_\infty \\ &= |||(\mathcal{Q}^n_{SS})^{-1}[\mathcal{Q}^n_{SS} - \mathcal{Q}^*_{SS}](\mathcal{Q}^*_{SS})^{-1}|||_\infty \\ &\leq \sqrt{d}|||(\mathcal{Q}^n_{SS})^{-1}[\mathcal{Q}^n_{SS} - \mathcal{Q}^*_{SS}](\mathcal{Q}^*_{SS})^{-1}|||_2 \\ &\leq \sqrt{d}|||(\mathcal{Q}^n_{SS})^{-1}|||_2|||\mathcal{Q}^n_{SS} - \mathcal{Q}^*_{SS}|||_2|||(\mathcal{Q}^*_{SS})^{-1}|||_2 \\ &\leq \frac{\sqrt{d}}{C_{\min}}|||\mathcal{Q}^n_{SS} - \mathcal{Q}^*_{SS}|||_2|||(\mathcal{Q}^n_{SS})^{-1}|||_2. \end{aligned}$$

Next, we bound each term of the final expression in the above equation separately. The first term can be bounded using Eq. 27:

$$P\big[|||\mathcal{Q}^n_{SS} - \mathcal{Q}^*_{SS}|||_2 \geq C^2_{min}\delta/2\sqrt{d}\big] \leq 2\exp\big(-K\frac{n\delta^2}{d^3} + 2\log d\big),$$

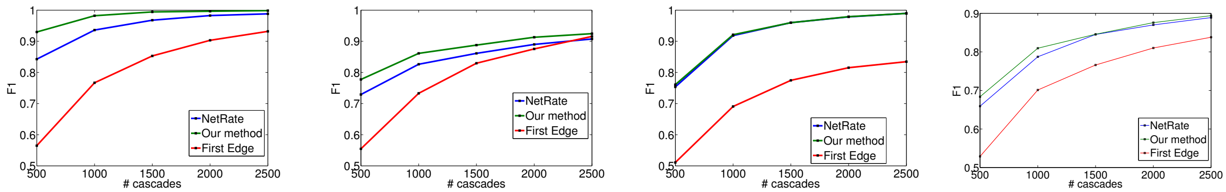
The second term can be bounded using Lemma 6:

$$\begin{aligned} &P\big[|||(\mathcal{Q}^n_{SS})^{-1}|||_2 \geq \frac{2}{C_{\min}}\big] \\ &\quad = P\big[\Lambda_{\min}(\mathcal{Q}^n_{SS}) \leq \frac{C_{\min}}{2}\big] \\ &\quad \leq \exp\left(-K\frac{n}{d^2} + B\log d\right). \end{aligned}$$

Then, the third confidence bound follows. ■

*Control of $A_1$.* We start by rewriting the term $A_1$ as

$$A_1 = \mathcal{Q}^*_{S^cS}(\mathcal{Q}^*_{SS})^{-1}[(\mathcal{Q}^*_{SS}) - (\mathcal{Q}^n_{SS})](\mathcal{Q}^n_{SS})^{-1},$$




(a) Kronecker hierarchical, EXP (b) Kronecker hierarchical, RAY (c) Forest Fire, POW (d) Forest Fire, RAY

*Figure 7.* $F_1$-score vs. # of cascades.

and further,

$$|||A_1|||_\infty \leq |||\mathcal{Q}^*_{S^cS}(\mathcal{Q}^*_{SS})^{-1}|||_\infty \\ \times |||(\mathcal{Q}^*_{SS}) - (\mathcal{Q}^n_{SS})|||_\infty |||(\mathcal{Q}^n_{SS})^{-1}|||_\infty.$$

Next, using the incoherence condition easily yields:

$$|||A_1|||_\infty \leq (1-\varepsilon)|||(\mathcal{Q}^*_{SS}) - (\mathcal{Q}^n_{SS})|||_\infty \\ \times \sqrt{d}|||(\mathcal{Q}^n_{SS})^{-1}|||_2$$

Now, we apply Lemma 6 with $\delta = C_{\min}/2$ to have that $|||(\mathcal{Q}^n_{SS})^{-1}|||_2 \leq \frac{2}{C_{\min}}$ with probability greater than $1 - \exp(-Kn/d^2 + K' \log d)$, and then use Eq. 30 with $\delta = \frac{\varepsilon C_{\min}}{12\sqrt{d}}$ to conclude that

$$P\big[|||A_1|||_\infty \geq \frac{\varepsilon}{6}\big] \leq 2\exp\left(-K\frac{n}{d^3} + K' \log d\right).$$

*Control of* $A_2$. We rewrite the term $A_2$ as

$$|||A_2|||_\infty \leq |||\mathcal{Q}^n_{S^cS} - \mathcal{Q}^*_{S^cS}|||_\infty |||(\mathcal{Q}^n_{SS})^{-1} - (\mathcal{Q}^*_{SS})^{-1}|||_\infty,$$

and then use Eqs. 28 and 29 with $\delta = \sqrt{\varepsilon/6}$ to conclude that

$$P\big[|||A_2|||_\infty \geq \frac{\varepsilon}{6}\big] \leq \\ 4\exp\left(-K\frac{n}{d^3} + \log(p-d) + K'\log p\right).$$

*Control of* $A_3$. We rewrite the term $A_3$ as

$$\begin{aligned} |||A_3|||_\infty &= \sqrt{d}|||(\mathcal{Q}^*_{SS})^{-1}|||_2 |||\mathcal{Q}^n_{S^cS} - \mathcal{Q}^*_{S^cS}|||_\infty \\ &\leq \frac{\sqrt{d}}{C_{\min}}|||\mathcal{Q}^n_{S^cS} - \mathcal{Q}^*_{S^cS}|||_\infty. \end{aligned}$$

We then apply Eq. 28 with $\delta = \frac{\varepsilon C_{\min}}{6\sqrt{d}}$ to conclude that

$$P\big[|||A_3|||_\infty \geq \frac{\varepsilon}{6}\big] \leq \exp\left(-K\frac{n}{d^3} + \log(p-d)\right),$$

and thus,

$$P\big[|||\mathcal{Q}^n_{S^cS}(\mathcal{Q}^n_{SS})^{-1}|||_\infty \geq 1 - \frac{\varepsilon}{2}\big] \\ = \mathcal{O}\left(\exp(-K\frac{n}{d^3} + \log p)\right).$$

## J. Additional experiments

**Parameters** $(n, p, d)$**.** Figure 5 shows the success probability at inferring the incoming links of nodes on the same type of canonical networks as depicted in Fig. 2. We choose nodes the same in-degree but different super-neighboorhod set sizes $p_i$ and experiment with different scalings $\beta$ of the number of cascades $n = 10\beta d \log p$. We set the regularization parameter $\lambda_n$ as a constant factor of $\sqrt{\log(p)/n}$ as suggested by Theorem 2 and, for each node, we used cascades which contained at least one node in the super-neighborhood of the node under study. We used an exponential transmission model and time window $T = 10$. As predicted by Theorem 2, very different $p$ values lead to curves that line up with each other quite well.

Figure 6 shows the success probability at inferring the incoming links of nodes of a hierarchical Kronecker network with equal super neighborhood size ($p_i = 70$) but different in-degree ($d_i$) under different scalings $\beta$ of the number of cascades $n = 10\beta d \log p$ and choose the regularization parameter $\lambda_n$ as a constant factor of $\sqrt{\log(p)/n}$ as suggested by Theorem 2. We used an exponential transmission model and time window $T = 5$. As predicted by Theorem 2, in this case, different $d$ values lead to noticeably different curves.

**Comparison with NETRATE and First-Edge.** Figure 7 compares the accuracy of our algorithm, NETRATE and First-Edge against number of cascades for different type of networks and transmission models. Our method typically outperforms both competitive methods. We find especially striking the competitive advantage with respect to First-Edge, however, this may be explained by comparing the sample complexity results for both methods: First-Edge needs $O(Nd \log N)$ cascades to achieve a probability of success approaching 1 in a rate polynomial in the number of cascades while our method needs $O(d^3 \log N)$ to achieve a probability of success approaching 1 in a rate exponential in the number of cascades.