APPENDIX

A LIKELIHOOD OF RECRUITMENT TIME SERIES

We consider the recruitment of subject i. Recall that $R_u(i)$ denotes the set of recruiters of subject u just before time $\mathbf{t}_i$ and that $I_u(i)$ denotes the set of potential recruitees of recruiter u just before time $\mathbf{t}_i$.

We compute the likelihood of the i-th recruitment event (the recruitment of subject i) in the two cases: i enters the study via the recruitment of a subject already in the study (in this case, subject i is not a seed node, which is denoted by $i \notin M$) and via the direct recruitment of the researchers (in this case, subject i is a seed node, which is denoted by $i \in M$).

Suppose that $i \notin M$. The inter-recruitment time between i and its potential recruiter u is denoted $W_{ui} = \mathbf{t}_i - \mathbf{t}_u$ and is greater than $\mathbf{t}_{i-1} - \mathbf{t}_u$ conditional on previous recruitment of i. Let U be the random variable of next recruiter and X be the random variable of next recruitee, namely the subject that will be labeled as subject i. We would like to note here that subject i is in fact random. Let J denote the event $\forall j \in R(i), k \in I(i), W_{jk} > \mathbf{t}_{i-1} - \mathbf{t}_j$.

We first compute the probability that a certain subject $x \in I(i)$ is the next (i-th) recruitee, $u \in R_x(i)$ is its recruiter, and the inter-recruitment time between u and x is greater than or equal to $t - \mathbf{t}_u$, conditional on event J. Intuitively, t is the recruitment time of subject x and in fact we are computing the tail probability of W_{ux}. We condition on the event J because having observed the $(i-1)$-th recruitment event, we know that for all possible recruiter-recruitee pairs in the next (i-th) recruitment event, say $j \in R(i)$ and $x \in I(i)$, their inter-recruitment time W_{jk} should be greater than or equal to $\mathbf{t}_{i-1} - \mathbf{t}_j$ (otherwise, the event that subject j recruits subject x will happen before $\mathbf{t}_i$ and they will not appear in $R(i)$ and $I(i)$, respectively). We have

$$
\begin{aligned}
&\Pr\left[U = u, X = x, W_{ux} \geq t - \mathbf{t}_u \mid J\right] \\
=&\Pr\left[W_{ux} \geq t - \mathbf{t}_u, \mathbf{t}_j + W_{jk} > \mathbf{t}_u + W_{ux},\right. \\
&\left.\forall j \in R(i), k \in I(i), \{u, x\} \neq \{j, k\} \mid J\right]. \qquad (4)
\end{aligned}
$$

Since the i-th recruitment event is that u recruits x, the inter-recruitment time along this link must be minimum among those along all other links. Therefore, in Eq. (4) we consider W_{jk} for $\forall j \in R(i), k \in I(i), \{u, x\} \neq \{j, k\}$. We require that

$$\mathbf{t}_j + W_{jk} > \mathbf{t}_u + W_{ux},$$

which means exactly that the recruitment time of x ($\mathbf{t}_u + W_{ux}$) is minimum (smaller than the recruitment time of k for all k).

Then we marginalize the above probability in Eq. (4) over all possible combinations of x and u. Recall that any subject in $x \in I(i)$ could possibly be the subject i and any subject in $R_x(i)$ could be her recruiter; therefore we need to sum over all possible recruitee-recruiter combinations, i.e., sum over $x \in I(i)$ and $u \in R_x(i)$:

$$
\begin{aligned}
&\Pr\left[W_{Ui} \geq t - \mathbf{t}_U \mid J\right] \\
=& \sum_{x \in I(i)} \sum_{u \in R_x(i)} \Pr\left[W_{ux} \geq t - \mathbf{t}_u, \mathbf{t}_j + W_{jk} > \mathbf{t}_u\right. \\
&\left.+W_{ux}, \forall j \in R(i), k \in I(i), \{u, x\} \neq \{j, k\} \mid J\right] \\
=& \sum_{x \in I(i)} \sum_{u \in R_x(i)} \int_{t-\mathbf{t}_u}^{\infty} \rho_{\tau(u;i)}(s) \mathrm{d}s \Pr\left[W_{jk} > s + \mathbf{t}_u\right. \\
&\left.-\mathbf{t}_j, \forall j \in R(i), k \in I(i), \{u, x\} \neq \{j, k\} \mid J\right] \\
=& \sum_{x \in I(i)} \sum_{u \in R_x(i)} \int_{t-\mathbf{t}_u}^{\infty} \rho_{\tau(u;i)}(s) \cdot \\
&\frac{\prod_{j \in R(i)} \prod_{k \in I_j(i)} \left(1 - D_{\tau(j;i)}(s + \mathbf{t}_u - \mathbf{t}_j)\right)}{1 - D_{\tau(u;i)}(s)} \mathrm{d}s \\
=& \sum_{x \in I(i)} \sum_{u \in R_x(i)} \int_{t-\mathbf{t}_u}^{\infty} H_{\tau(u;i)}(s) \cdot \\
&\prod_{j \in R(i)} S_{\tau(j;i)}^{|I_j(i)|}(s + \mathbf{t}_u - \mathbf{t}_j) \mathrm{d}s.
\end{aligned}
$$

Using the notation that we introduced in Section 2 to rewrite and simplify the above expression, we obtain the likelihood of the i-th recruitment event for $i \notin M$:

$$
\begin{aligned}
&\sum_{x \in I(i)} \sum_{u \in R_x(i)} H_{\tau(u;i)}(\mathbf{t}_i - \mathbf{t}_u) \prod_{j \in R(i)} S_{\tau(j;i)}^{|I_j(i)|}(\mathbf{t}_i - \mathbf{t}_j) \\
=& \prod_{j \in R(i)} S_{\tau(j;i)}^{|I_j(i)|}(\mathbf{t}_i - \mathbf{t}_j) \sum_{u \in R(i)} |I_u(i)| H_{\tau(u;i)}(\mathbf{t}_i - \mathbf{t}_u)
\end{aligned}
$$

Now suppose that $i \in M$, which means that subject i is recruited into the study directly by the researchers. Therefore the inter-recruitment time of any possible recruiter-recruitee pairs of the i-th recruitment event, say j and k, should be greater than or equal to $t - \mathbf{t}_j$, where t is the recruitment time of subject i. In the terminology of survival analysis, all these potential recruitment links are censored. So we compute the probability that all these links are censored:

$$
\begin{aligned}
&\Pr\left[W_{jk} \geq t - \mathbf{t}_j, \forall j \in R(i), k \in I_j(i) \mid J\right] \\
=& \prod_{j \in R(i)} \prod_{k \in I_j(i)} \left(1 - D_{\tau(j;i)}(t - \mathbf{t}_j)\right) \\
=& \prod_{j \in R(i)} S_{\tau(j;i)}^{|I_j(i)|}(t - \mathbf{t}_j).
\end{aligned}
$$

Plugging in the observed recruitment time of subject i (denoted by $\mathbf{t}_i$), we obtain the likelihood of the i-th recruitment event for $i \in M$:

$$\prod_{j \in R(i)} S_{\tau(j;i)}^{|I_j(i)|}(\mathbf{t}_i - \mathbf{t}_j).$$

So far we have obtained the likelihood of the i-th recruitment event for both cases ($i \notin M$ and $i \in M$); Multiplying the likelihoods with i running from 1 to n, we have the entire likelihood:

$$\prod_{i=1}^{n} \left(\sum_{u \in R(i)} |I_u(i)| H_{\tau(u;i)}(\mathbf{t}_i - \mathbf{t}_u) \right)^{1\{i \notin M\}} \cdot \prod_{j \in R(i)} S_{\tau(j;i)}^{|I_j|}(\mathbf{t}_i - \mathbf{t}_j), \quad (5)$$

where $1\{i \notin M\}$ the indicator random variable for the event that $i \notin M$.

B LOG-LIKELIHOOD OF RECRUITMENT TIME SERIES

According to Eq. (5) in Appendix A , the log-likelihood is

$$\sum_{i=1}^{n} \left[1\{i \notin M\} \log \left(\sum_{u \in R(i)} |I_u(i)| H_{\tau(u;i)}(\mathbf{t}_i - \mathbf{t}_u) \right) + \sum_{j \in R(i)} |I_j(i)| \log S_{\tau(j;i)}(\mathbf{t}_i - \mathbf{t}_j) \right].$$

The number of recruitees of recruiter u just before time $\mathbf{t}_i$ is given by

$$|I_u(i)| = \mathbf{C}_{ui} \left(\sum_{k=i}^{n} \mathbf{A}_{uk} + \mathbf{u}_u \right). \quad (6)$$

The cardinality of $I_u(i)$ is zero if and only if recruiter u has at least one coupon just before $\mathbf{t}_i$, i.e., $\mathbf{C}_{ui} = 1$. Therefore there is a factor $\mathbf{C}_{ui}$ in Eq. (6). When recruiter u has at least one coupon, the number of recruitees of recruiter u just before time $\mathbf{t}_i$ is

$$\sum_{k=i}^{n} \mathbf{A}_{uk} + \mathbf{u}_u,$$

where $\sum_{k=i}^{n} \mathbf{A}_{uk}$ is the number of recruitees in the final sample and $\mathbf{u}_u$ is the number of recruitees outside the final sample. In the expression $\sum_{k=i}^{n} \mathbf{A}_{uk}$, we sum over k from i to n since subjects $i, i+1, \ldots, n$ are those in the final sample and recruited at and after time $\mathbf{t}_i$ and they contribute one to the sum if they are adjacent to subject u (namely $\mathbf{A}_{uk} = 1$).

Recall that $\mathbf{B}$ is the Hadamard product of $\mathbf{C}$ and $\mathbf{H}$, which yields that $\mathbf{B}_{ui} = \mathbf{C}_{ui}\mathbf{H}_{ui}$, and that

$$\mathbf{H}_{ui} = H_{\tau(u;i)}(\mathbf{t}_i - \mathbf{t}_u).$$

Therefore the term

$$\sum_{u \in R(i)} |I_u(i)| H_{\tau(u;i)}(\mathbf{t}_i - \mathbf{t}_u)$$

in the log-likelihood can be written as

$$\begin{aligned} & \sum_{u \in R(i)} |I_u(i)| H_{\tau(u;i)}(\mathbf{t}_i - \mathbf{t}_u) \\ =& \sum_{u} \mathbf{C}_{ui} \left(\sum_{k=i}^{n} \mathbf{A}_{uk} + \mathbf{u}_u \right) \mathbf{H}_{ui} \\ =& (\mathbf{B}'\mathbf{u} + \mathrm{LowerTri}(\mathbf{AB})' \cdot \mathbf{1})_i. \end{aligned}$$

Recall the definition of the matrix $\mathbf{S}$:

$$\mathbf{S}_{ji} = \log S_{\tau(j;i)}(\mathbf{t}_i - \mathbf{t}_j),$$

and that $\mathbf{D}$ is the Hadamard product of $\mathbf{C}$ and $\mathbf{S}$, which yields that $\mathbf{D}_{ji} = \mathbf{C}_{ji}\mathbf{S}_{ji}$. Similarly, the term

$$\sum_{j \in R(i)} |I_j(i)| \log S_{\tau(j;i)}(\mathbf{t}_i - \mathbf{t}_j)$$

in the log-likelihood is given by

$$\begin{aligned} & \sum_{j \in R(i)} |I_j(i)| \log S_{\tau(j;i)}(\mathbf{t}_i - \mathbf{t}_j) \\ =& \sum_{j \in R(i)} \mathbf{C}_{ji} \left(\sum_{k=i}^{n} \mathbf{A}_{jk} + \mathbf{u}_j \right) \mathbf{S}_{ji} \\ =& (\mathbf{D}'\mathbf{u} + \mathrm{LowerTri}(\mathbf{AD})' \cdot \mathbf{1})_i. \end{aligned}$$

Thus the log-likelihood is

$$\sum_{i=1}^{n} [1\{i \notin M\} \log (\mathbf{B}'\mathbf{u} + \mathrm{LowerTri}(\mathbf{AB})' \cdot \mathbf{1})_i + (\mathbf{D}'\mathbf{u} + \mathrm{LowerTri}(\mathbf{AD})' \cdot \mathbf{1})_i] = \mathbf{m}'\beta + \mathbf{1}'\delta.$$

C PROOF OF THEOREM 1

In this section, we will show that $\log \tilde{L}(\boldsymbol{\gamma})$ is submodular in $\boldsymbol{\gamma}$. We have

$$\begin{aligned} & \log \tilde{L}(\boldsymbol{\gamma}) \\ =& \log L(\mathbf{t}|\mathbf{A}, \theta) + \log \pi(\mathbf{A}) + \log \phi(\theta) \\ =& \mathbf{m}'\beta + \mathbf{1}'\delta - \psi(\max\{\mathbf{u} + \mathbf{A} \cdot \mathbf{1} - \mathbf{d}, \mathbf{0}\}) + \log \phi(\theta). \end{aligned}$$

Later we will show that it is submodular part by part.

First, we need to prove that $-\psi(\max\{\mathbf{u}+\mathbf{A}\cdot\mathbf{1}-\mathbf{d},\mathbf{0}\}\})$ is submodular. We temporarily view them as real-valued vectors and matrices rather than binary vectors and matrices. In light of the fact that

$$\mathbf{u} = \boldsymbol{\mu} \cdot \begin{pmatrix} 2^0 & 2^1 & 2^2 & \dots & 2^{\lceil \log_2 u_{\max} \rceil - 1} \end{pmatrix}',$$

we know that $\mathbf{u}$ is a linear function of $\boldsymbol{\mu}$, which yields that $\mathbf{u} + \mathbf{A} \cdot \mathbf{1} - \mathbf{d}$ is a linear function of $\boldsymbol{\mu}$. We know that if $g(\mathbf{x})$ is a linear function, then $f(\mathbf{x}) = \max\{g(\mathbf{x}), 0\}$ is a convex function (see Section 3.2.3 in (Boyd and Vandenberghe, 2004)). Therefore, every entry of $\max\{\mathbf{u} + \mathbf{A} \cdot \mathbf{1} - \mathbf{d}, \mathbf{0}\}$ is convex in $\mathbf{u}$. Since ψ is a convex function and non-decreasing in each argument whenever this argument is non-negative, thus $\psi(\max\{\mathbf{u} + \mathbf{A} \cdot \mathbf{1} - \mathbf{d}, \mathbf{0}\}\})$ is convex in $\mathbf{u}$ (see Section 3.2.4 in (Boyd and Vandenberghe, 2004)); equivalently, $-\psi(\max\{\mathbf{u} + \mathbf{A} \cdot \mathbf{1} - \mathbf{d}, \mathbf{0}\}\})$ is concave in $\mathbf{u}$. Thus this term $-\psi(\max\{\mathbf{u}+\mathbf{A}\cdot\mathbf{1}-\mathbf{d},\mathbf{0}\}\})$ is submodular in $\boldsymbol{\mu}$ if we view $\boldsymbol{\mu}$ as a Boolean vector.

Recall the definitions of β and δ:

$$\begin{aligned} \beta &= \log(\mathbf{B}'\mathbf{u} + \mathrm{LowerTri}(\mathbf{AB})' \cdot \mathbf{1}), \\ \delta &= \mathbf{D}'\mathbf{u} + \mathrm{LowerTri}(\mathbf{AD})' \cdot \mathbf{1}. \end{aligned}$$

The function $\beta(\mathbf{u}, \mathbf{A})$ is concave in $\mathbf{u}$ and $\mathbf{A}$ since the inner part

$$\mathbf{B}'\mathbf{u} + \mathrm{LowerTri}(\mathbf{AB})' \cdot \mathbf{1}$$

is linear in $\mathbf{u}$ and $\mathbf{A}$, the logarithm function is concave, and β is the composition of the linear inner part and the concave logarithm function. The function $\delta(\mathbf{u}, \mathbf{A})$ is linear in $\mathbf{u}$ and $\mathbf{A}$. Recall that $\mathbf{u}$ and $\mathbf{A}$ are linear in $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$, respectively. Thus β is concave in $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ and δ is linear in $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$. Therefore β is concave in $\boldsymbol{\gamma}$ and δ is linear in $\boldsymbol{\gamma}$, where

$$\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\mu}).$$

Thus $\mathbf{m}'\beta + \mathbf{1}'\delta$ is submodular in $\boldsymbol{\gamma}$ if $\boldsymbol{\gamma}$ is viewed as a binary vector.

Hence in light of the fact that the sum of submodular functions is submodular, the whole expression is submodular in $\boldsymbol{\gamma}$. In other words, $\log L(\boldsymbol{\gamma})$ is submodular in $\boldsymbol{\gamma}$.

D PROOF OF PROPOSITION 1

We prove it by induction.

Suppose that $\boldsymbol{\gamma} = \{v_{j_1}, v_{j_2}, \dots, v_{j_q}\}$, where $j_1 < j_2 < \dots < j_q$ and $q \le N$.

If $q = 1$, then

$$\begin{aligned} &F(\boldsymbol{\gamma}) - F(\varnothing) = F(\{v_{j_1}\}) - F(\varnothing) \\ &\ge F(V_{j_1-1} \cup v_{j_1}) - F(V_{j_1-1}) = s^g(v_{j_1}) = s^g(\boldsymbol{\gamma}), \end{aligned}$$

since $\varnothing$ must be a subset of V_{j_1-1}. Therefore,

$$F(\boldsymbol{\gamma}) \ge s^g(\boldsymbol{\gamma}) + F(\varnothing) = s^g(\boldsymbol{\gamma}),$$

since F is normalized.

Suppose that the proposition holds for all $q < r$. When $q = r$, we have

$$\begin{aligned} &F(\{v_{j_1}, v_{j_2}, \dots, v_{j_r}\}) - F(\{v_{j_1}, v_{j_2}, \dots, v_{j_{r-1}}\}) \\ &\quad \ge F(V_{j_r-1} \cup \{v_{j_r}\}) - F(V_{j_r-1}) = s^g(\{v_{j_r}\}), \end{aligned}$$

since $\{v_{j_1}, v_{j_2}, \dots, v_{j_{r-1}}\}$ is a subset of V_{j_r-1}. Therefore, we obtain

$$\begin{aligned} F(\boldsymbol{\gamma}) &= F(\{v_{j_1}, v_{j_2}, \dots, v_{j_r}\}) \\ &\ge s^g(\{v_{j_r}\}) + F(\{v_{j_1}, v_{j_2}, \dots, v_{j_{r-1}}\}) \\ &\ge s^g(\{v_{j_r}\}) + s^g(\{v_{j_1}, v_{j_2}, \dots, v_{j_{r-1}}\}) \\ &= s^g(\{v_{j_1}, v_{j_2}, \dots, v_{j_r}\}) \end{aligned}$$

by the induction assumption. This completes the proof.

E PROOF OF PROPOSITION 2

In order to show that a modular function s is a supergradient of the submodular function F at $\mathbf{x}$, we have to show that

$$\forall \mathbf{y} \in \{0,1\}^N, F(\mathbf{y}) \le F(\mathbf{x}) + s(\mathbf{y}) - s(\mathbf{x}).$$

Equivalent, if viewed as a set function, we have to show that

$$\forall Y \subseteq [N], F(Y) \le F(X) + s(Y) - s(X),$$

where X is the corresponding subset for $\mathbf{x}$, i.e.,

$$X = \{i \in [N] : \mathbf{x}_i = 1\}.$$

Since s is a modular function, it is equivalent to show

$$F(Y) + \sum_{i \in X \setminus Y} s(\{i\}) \le F(X) + \sum_{i \in Y \setminus X} s(\{i\}).$$

Grow supergradient We have to show that

$$F(Y) + \sum_{i \in X \setminus Y} \hat{s}(\{i\}) \le F(X) + \sum_{i \in Y \setminus X} \hat{s}(\{i\});$$

equivalently,

$$F(Y) + \sum_{i \in X \setminus Y} \Delta_i F(V - \{i\}) \leq F(X) + \sum_{i \in Y \setminus X} \Delta_i F(X).$$

We will show that the left-hand side is less than or equal to $F(X \cup Y)$ while the right-hand side is greater than or equal to $F(X \cup Y)$.

Suppose that $Y \setminus X = \{a_1, a_2, a_3, \ldots, a_r\}$, $A_i = X \cup \{a_1, a_2, a_3, \ldots, a_i\}$ and $A_0 = X$. We have

$$\begin{aligned} F(X \cup Y) &= F(X) + \sum_{i=1}^{r} (F(A_i) - F(A_{i-1})) \\ &= F(X) + \sum_{i=1}^{r} \Delta_{a_i} F(A_{i-1}) \\ &\leq F(X) + \sum_{i=1}^{r} \Delta_{a_i} F(X) \\ &= F(X) + \sum_{i \in Y \setminus X} \Delta_i F(X). \end{aligned} \tag{7}$$

Suppose that $X \setminus Y = \{b_1, b_2, \ldots, b_q\}$, $B_i = Y \cup \{b_1, b_2, \ldots, b_i\}$ and $B_0 = Y$. We have

$$\begin{aligned} F(X \cup Y) &= F(Y) + \sum_{i=1}^{q} (F(B_i) - F(B_{i-1})) \\ &= F(Y) + \sum_{i=1}^{q} \Delta_{b_i} F(B_{i-1}) \\ &\geq F(Y) + \sum_{i=1}^{q} \Delta_{b_i} F(V - \{b_i\}) \\ &= F(Y) + \sum_{i \in X \setminus Y} \Delta_i F(V - \{i\}). \end{aligned} \tag{8}$$

Shrink supergradient We have to show that

$$F(Y) + \sum_{i \in X \setminus Y} \check{s}(\{i\}) \leq F(X) + \sum_{i \in Y \setminus X} \check{s}(\{i\});$$

equivalently,

$$F(Y) + \sum_{i \in X \setminus Y} \Delta_i F(X - \{i\}) \leq F(X) + \sum_{i \in Y \setminus X} F(\{i\}).$$

We will show that the left-hand side is less than or equal to $F(X \cup Y)$ while the right-hand side is greater than or equal to $F(X \cup Y)$.

In light of Eq. (7), we have

$$\begin{aligned} F(X \cup Y) &= F(X) + \sum_{i=1}^{r} \Delta_{a_i} F(A_{i-1}) \\ &\leq F(X) + \sum_{i=1}^{r} \Delta_{a_i} F(\varnothing) \\ &= F(X) + \sum_{i \in Y \setminus X} \Delta_i F(\varnothing) \\ &= F(X) + \sum_{i \in Y \setminus X} F(\{i\}). \end{aligned} \tag{9}$$

In light of Eq. (8), we have

$$\begin{aligned} F(X \cup Y) &= F(Y) + \sum_{i=1}^{q} \Delta_{b_i} F(B_{i-1}) \\ &\geq F(Y) + \sum_{i=1}^{q} \Delta_{b_i} F(X - \{b_i\}) \\ &= F(Y) + \sum_{i \in X \setminus Y} \Delta_i F(X - \{i\}). \end{aligned}$$

Bar supergradient We have to show that

$$F(Y) + \sum_{i \in X \setminus Y} \bar{s}(\{i\}) \leq F(X) + \sum_{i \in Y \setminus X} \bar{s}(\{i\});$$

equivalently,

$$F(Y) + \sum_{i \in X \setminus Y} \Delta_i F(V - \{i\}) \leq F(X) + \sum_{i \in Y \setminus X} F(\{i\}).$$

By Eq. (8), we know that the left-hand side is less than or equal to $F(X \cup Y)$. By Eq. (9), we know that the right-hand side is greater than or equal to $F(X \cup Y)$. Therefore, the left-hand side is less than or equal to the right-hand side.

F DISCUSSION

In some ways, RDS resembles a diffusion process and in fact the continuous-time stochastic process that we formulate RDS as in this paper is a diffusion process; but RDS reveals an extra piece of information that makes reconstruction of the induced subgraph possible: the degrees of each vertex visited by the diffusion process.

For other diffusion processes, we can also derive their likelihood functions. If the optimization problem of the likelihood or the posterior is log-submodular and unconstrained, we can use the submodular variational inference method that we used in this paper. If it is a constrained problem, it is natural to relax the constraints with multiplicative log-submodular penalty factor such that the product of the original likelihood/posterior function and the penalty factor remain log-submodular. For example, if the constraint is an equality constraint of the form $q(\mathbf{A}) = \boldsymbol{c}_0$, where q is a linear function (e.g., multiplying $\mathbf{A}$ by some matrix), $\boldsymbol{c}_0$ is a fixed vector, and $\mathbf{A}$ is the adjacency matrix to be optimized over, then we can add a multiplicative factor $e^{-\|q(\mathbf{A})-\boldsymbol{c}_0\|}$ to the likelihood or posterior function, in light of the fact that every norm is a convex function, which guarantees that $e^{-\|q(\mathbf{A})-\boldsymbol{c}_0\|}$ is log-submodular with respect to $\mathbf{A}$. If the constraint is given by an inequality, we can mimic the method that we used in Section 3.1 by introducing auxiliary variables $\mathbf{u}$ and adding a multiplicative penalty term similar to Eq. (3).