# ChoiceRank: Identifying Preferences from Node Traffic in Networks Supplementary Material

**Lucas Maystre** [1] **Matthias Grossglauser** [1]

The supplementary material consists of three parts. In Section A, we present a generalization of the network choice model, and we prove *a*) the minimal sufficiency of per-node traffic statistics and *b*) the well-posedness of MAP inference. In Section B, we deepen our discussion of the ML estimator. Finally, in Section C, we revisit the ChoiceRank algorithm. We give the modifications needed for the generalized network choice model, prove the convergence of the algorithm, and give an alternative derivation based on the expectation maximization perspective.

**Notations.** The notation follows the conventions adopted in the main text. To simplify some expressions, we use $\kappa_1, \kappa_2, \ldots$ to denote constants that do not depend on the parameter vector $\boldsymbol{\lambda}$.

## A Extensions and Proofs

In this section, we start by generalizing the network choice model to account for edge weights. Then, we present formal proofs for *a*) the (minimal) sufficiency of marginal counts and *b*) the well-posedness of MAP inference in the generalized weighted network choice model.

### A.1 Generalization of the Model

Let $G = (V, E)$ be a weighted, directed graph with edge weights $w_{ij} > 0$ for all $(i, j) \in E$. Kumar et al. (2015) propose the following generalization of Luce's choice model. Given a parameter vector $\boldsymbol{\lambda} \in \mathbf{R}^n_{>0}$, they define the choice probabilities as

$$p_{ij} = \frac{w_{ij}\lambda_j}{\sum_{k \in N_i^+} w_{ik}\lambda_k}, \quad j \in N_i^+. \tag{1}$$

We refer to this model as the *weighted network choice model*. Intuitively, the strength of each alternative is weighted by the corresponding edge's weight; Luce's original choice model is obtained by setting $w_{ij} =$ constant. In this general model, the log-likelihood becomes

$$\begin{aligned}
\ell(\boldsymbol{\lambda}; \mathcal{D}) &= \sum_{(i,j) \in E} c_{ij} \left[ \log w_{ij}\lambda_j - \log \sum_{k \in N_i^+} w_{ik}\lambda_k \right] \\
&= \sum_{(i,j) \in E} c_{ij} \left[ \log \lambda_j - \log \sum_{k \in N_i^+} w_{ik}\lambda_k \right] \\
&\quad + \sum_{(i,j) \in E} c_{ij} \log w_{ij}, \\
&= \sum_{i=1}^{n} \left[ c_i^- \log \lambda_i - c_i^+ \log \sum_{k \in N_i^+} w_{ik}\lambda_k \right] + \kappa_1,
\end{aligned} \tag{2}$$

where $c_i^- = \sum_{j \in N_i^-} c_{ji}$ and $c_i^+ = \sum_{j \in N_i^+} c_{ij}$ is the aggregate number of transitions arriving in and originating from $i$, respectively. Note that for every $i$, the weights $\{w_{ij} \mid j \in N_i^+\}$ are equivalent up to rescaling.

This generalization is relevant in situations where the current context modulates the alternatives' strength. For example, this could be used to take into account the position or prominence of a link on a page in a hyperlink graph, or the distance between two locations in a mobility network.

### A.2 Minimal Sufficiency of Marginal Counts

Recall that $c_{ij}$ denotes the number of times we observe a transition from $i$ to $j$. We set out to prove the following theorem for the weighted network choice model.

**Theorem 4.** *Let $c_i^- = \sum_{j \in N_i^-} c_{ji}$ and $c_i^+ = \sum_{j \in N_i^+} c_{ij}$ be the aggregate number of transitions arriving in and originating from i, respectively. Then, $\{(c_i^-, c_i^+) \mid i \in V\}$ is a minimally sufficient statistic for the parameter $\boldsymbol{\lambda}$ in the weighted network choice model.*

*Proof.* Let $f(\{c_{ij}\} \mid \boldsymbol{\lambda})$ be the discrete probability density function of the data under the model with parameters $\boldsymbol{\lambda}$. Theorem 6.2.13 in Casella & Berger (2002) states that $\{(c_i^-, c_i^+)\}$ is a minimally sufficient statistic for $\boldsymbol{\lambda}$ if and

[1] School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland. Correspondence to: Lucas Maystre <lucas.maystre@epfl.ch>.

*Proceedings of the $34^{th}$ International Conference on Machine Learning*, Sydney, Australia, PMLR 70, 2017. Copyright 2017 by the author(s).

only if, for any $\{c_{ij}\}$ and $\{d_{ij}\}$ in the support of $f$,

$$\frac{f(\{c_{ij}\} \mid \boldsymbol{\lambda})}{f(\{d_{ij}\} \mid \boldsymbol{\lambda})} \text{ is independent of } \boldsymbol{\lambda} \iff (c_i^-, c_i^+) = (d_i^-, d_i^+) \quad \forall i. \tag{3}$$

Taking the log of the ratio on the left-hand side and using (2), we find that

$$\log \frac{f(\{c_{ij}\} \mid \boldsymbol{\lambda})}{f(\{d_{ij}\} \mid \boldsymbol{\lambda})} = \sum_{i=1}^{n} \Bigg[ (c_i^- - d_i^-) \log \lambda_i - (c_i^+ - d_i^+) \log \sum_{k \in N_i^+} w_{ik} \lambda_k \Bigg] + \kappa_2.$$

From this, it is easy to see that the ratio of densities is independent of $\boldsymbol{\lambda}$ if and only if $c_i^- = d_i^-$ and $c_i^+ = d_i^+$, which verifies (3). □

### A.3 Well-Posedness of MAP Inference

Using a Gamma$(\alpha, \beta)$ prior for each parameter, the log-posterior of the weighted network choice model can be written as

$$\log p(\boldsymbol{\lambda} \mid \mathcal{D}) = \sum_{i=1}^{n} \Bigg[ (c_i^- + \alpha - 1) \log \lambda_i - c_i^+ \log \Bigg( \sum_{k \in N_i^+} w_{ik} \lambda_k \Bigg) - \beta \lambda_i \Bigg] + \kappa_3. \tag{4}$$

We prove a theorem that guarantees that MAP estimation is well-posed in this generalized model; the proof of Theorem 2 follows trivially.

**Theorem 5.** *If i.i.d.* $\lambda_1, \ldots, \lambda_n \sim$ *Gamma*$(\alpha, \beta)$ *with* $\alpha > 1$*, then there exists a unique maximizer* $\boldsymbol{\lambda}^\star \in \mathbf{R}^n_{>0}$ *of the weighted network choice model's log-posterior* (4).

*Proof.* The log-posterior (4) is not concave in $\boldsymbol{\lambda}$, but it can be made concave using the simple reparametrization $\lambda_i = e^{\theta_i}$. Under this reparametrization, the log-prior and the log-likelihood become

$$\log p(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[ (\alpha - 1)\theta_i - \beta e^{\theta_i} \right] + \kappa_4,$$

$$\ell(\boldsymbol{\theta}; \mathcal{D}) = \sum_{i=1}^{n} \Bigg[ c_i^- \theta_i - c_i^+ \log \sum_{k \in N_i^+} w_{ik} e^{\theta_k} \Bigg] + \kappa_5.$$

It is easy to see that the log-likelihood is concave and the log-prior strictly concave in $\boldsymbol{\theta}$. As a result, the log-posterior is strictly concave in $\boldsymbol{\theta}$, which ensures that there exists at most one maximizer.

Now consider any transition counts $\{c_{ij}\}$ that satisfy $c_i^- = \sum_{j \in N_i^-} c_{ji}$ and $c_i^+ = \sum_{j \in N_i^+} c_{ij}$. The log-posterior can be written as

$$\begin{aligned} \log p(\boldsymbol{\theta} \mid \mathcal{D}) &= \sum_{i=1}^{n} \sum_{j \in N_i^+} c_{ij} \Bigg[ \theta_j - \log \sum_{k \in N_i^+} w_{ik} e^{\theta_k} \Bigg] \\ &\quad + \sum_{i=1}^{n} \left[ (\alpha - 1)\theta_i - \beta e^{\theta_i} \right] + \kappa_3 \\ &\leq -n^2 \cdot \max_{i,j} \log w_{ij} \\ &\quad + \sum_{i}^{n} \left[ (\alpha - 1)\theta_i - \beta e^{\theta_i} \right] + \kappa_3. \end{aligned}$$

For $\alpha > 1$, it follows that $\lim_{\|\boldsymbol{\theta}\| \to \infty} \log p(\boldsymbol{\theta} \mid \mathcal{D}) = -\infty$, which ensures that there is at least one maximizer. □

Note that Theorem 5 can easily be extended to independent but non-identical Gamma priors, where $\lambda_i \sim$ Gamma$(\alpha_i, \beta_i)$ and $\alpha_i \neq \alpha_j$, $\beta_i \neq \beta_j$ in general.

## B Maximum-Likelihood Estimation

In this section, we go into the analysis of the ML estimator in depth. From the definition of choice probabilities in (1), it is clear that the likelihood is invariant to a rescaling of the parameters, i.e., $\ell(\boldsymbol{\lambda}; \mathcal{D}) = \ell(s\boldsymbol{\lambda}; \mathcal{D})$ for any $s > 0$. We will therefore identify parameters up to rescaling.

### B.1 Necessary and Sufficient Conditions

In order to provide a data-dependent, necessary and sufficient condition that guarantees that the ML estimate is well-defined, we extend the definition of comparison hypergraph presented in Section 4.3.

**Definition** (Comparison graph)**.** Let $G = (V, E)$ be a directed graph and $\{a_{ij} \mid (i, j) \in E\}$ be non-negative numbers. The *comparison graph* induced by $\{a_{ij}\}$ is the directed graph $H = (V, E')$, where $(i, j) \in E'$ if and only if there is a node $k$ such that $i, j \in N_k^+$ and $a_{kj} > 0$.

The numbers $\{a_{ij}\}$ can be loosely interpreted as transition counts (although they do not need to be integer). Intuitively, there is an edge $(i, j)$ in the comparison graph whenever there is at least one instance in which $i$ and $j$ were among the alternatives and $j$ was selected. If $a_{ij} > 0$ for all edges, then the comparison graph is equivalent to its hypergraph counterpart, in that every hyperedge induces a clique in the comparison graph. As shown by the next theorem, the notion of (data-dependent) comparison graph leads to a precise characterization of whether the ML estimate is well-defined or not.

**Theorem 6.** *Let* $G = (V, E)$ *be a directed graph and* $\{(c_i^-, c_i^+)\}$ *be the aggregate number of transitions arriving in and originating from i, respectively. Let* $\{a_{ij}\}$ *be any* *set of non-negative real numbers that satisfy*

$$\sum_{j \in N_i^-} a_{ji} = c_i^-, \quad \sum_{j \in N_i^+} a_{ij} = c_i^+.$$

*Then, the maximizer of the log-likelihood* (2) *exists and is unique (up to rescaling) if and only if the comparison graph induced by* $\{a_{ij}\}$ *is strongly connected.*

The proof borrows from Hunter (2004), in particular from the proofs of Lemmas 1 and 2.

*Proof.* The log-likelihood (2) is not concave in $\boldsymbol{\lambda}$, but it can be made concave using the reparametrization $\lambda_i = e^{\theta_i}$. We can rewrite the reparametrized log-likelihood using $\{a_{ij}\}$ as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j \in N_i^+} a_{ij} \left[ \theta_j - \log \sum_{k \in N_i^+} w_{ik} e^{\theta_k} \right],$$

and, without loss of generality, we can assume that $\sum_i \theta_i = 0$ and $\min_{ij} w_{ij} = 1$.

First, we shall prove that the super-level set $\{\boldsymbol{\theta} \mid \ell(\boldsymbol{\theta}) \geq c\}$ is bounded and compact for any $c$, if and only if the comparison graph is strongly connected. The compactness of all super-level sets ensures that there is at least one maximizer. Pick any unit vector $\boldsymbol{u}$ such that $\sum_i u_i = 0$, and let $\boldsymbol{\theta} = s\boldsymbol{u}$ When $s \to \infty$, then $e^{\theta_i} > 0$ and $e^{\theta_j} \to 0$ for some $i$ and $j$. As the comparison graph is strongly connected, there is a path from $i$ to $j$, and along this path there must be two consecutive nodes $i', j'$ such that $e^{\theta_{i'}} > 0$ and $e^{\theta_{j'}} \to 0$. The existence of the edge $(i', j')$ in the comparison graph means that there is a $k$ such that $i', j' \in N_k^+$ and $a_{kj'} > 0$. Therefore, the log-likelihood can be bounded as

$$\begin{aligned}\ell(\boldsymbol{\theta}) &\leq a_{kj'} \left[ \theta_{j'} - \log \sum_{q \in N_k^+} w_{kq} e^{\theta_q} \right] \\ &\leq a_{kj'} \left[ \theta_{j'} - \log(e^{\theta_{j'}} + e^{\theta_{i'}}) \right],\end{aligned}$$

and $\lim_{s \to \infty} \ell(\boldsymbol{\theta}) = -\infty$. Conversely, suppose that the comparison graph is not strongly connected and partition the vertices into two non-empty subsets $S$ and $T$ such that there is no edge from $S$ to $T$. Let $c > 0$ be any positive constant, and take $\tilde{\theta}_i = \theta_i + c$ if $i \in S$ and $\tilde{\theta}_i = \theta_i$ if $i \in T$ (renormalize such that $\sum_i \tilde{\theta}_i = 0$). Clearly, $\ell(\tilde{\boldsymbol{\theta}}) \geq \ell(\boldsymbol{\theta})$, and by repeating this procedure $\|\boldsymbol{\theta}\|$ may be driven to infinity without decreasing the likelihood.

Second, we shall prove that if the comparison graph is strongly connected, the log-likelihood is strictly concave (in $\boldsymbol{\theta}$). In particular, for any $p \in (0, 1)$,

$$\ell\left[p\boldsymbol{\theta} + (1-p)\boldsymbol{\eta}\right] \geq p\ell(\boldsymbol{\theta}) + (1-p)\ell(\boldsymbol{\eta}), \tag{5}$$

with equality if and only if $\boldsymbol{\theta} \equiv \boldsymbol{\eta}$ up to a constant shift. Strict concavity ensures that there is at most one maximizer of log-likelihood. We start with Hölder's inequality, which implies that, for positive $\{x_k\}$ and $\{y_k\}$, and $p \in (0, 1)$,

$$\log \sum_k x_k^p y_k^{1-p} \leq p \log \sum_k x_k + (1-p) \log \sum_k y_k.$$

with equality if and only $x_k = c y_k$ for some $c > 0$. Letting $x_k = w_{ik} e^{\theta_k}$ and $y_k = w_{ik} e^{\eta_k}$, we find that for all $i$

$$\begin{aligned}&\log \sum_{k \in N_i^+} w_{ik} e^{p\theta_k + (1-p)\eta_k} \\ &\quad \leq p \log \sum_{k \in N_i^+} w_{ik} e^{\theta_k} + (1-p) \log \sum_{k \in N_i^+} w_{ik} e^{\eta_k},\end{aligned} \tag{6}$$

with equality if and only if there exists $c \in \mathbf{R}$ such that $\theta_k = \eta_k + c$ for all $k \in N_i^+$. Multiplying by $a_{ij}$ and summing over $i$ and $j$ on both sides of (6) shows that the log-likelihood is concave in $\boldsymbol{\theta}$. Now, consider any partition of the vertices into two non-empty subsets $S$ and $T$. Because the comparison graph is strongly connected, there is always $k \in V$, $i \in S$ and $j \in T$ such that $i, j \in N_k^+$ and $a_{ki} > 0$. Therefore, the left and right side of (5) are equal if and only if $\boldsymbol{\theta} \equiv \boldsymbol{\eta}$ up to a constant shift.

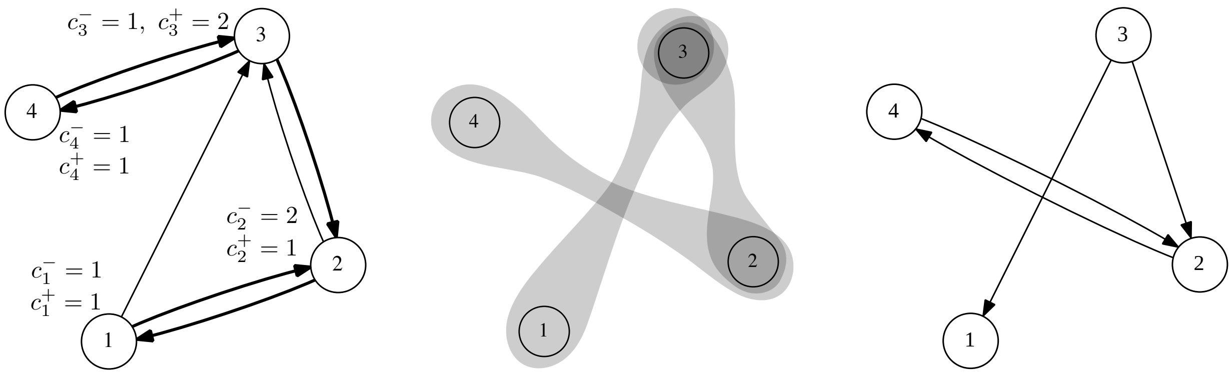
Bounded super-level sets and strict concavity form necessary and sufficient conditions for the existence and uniqueness of the maximizer. □

We now give a proof for Theorem 1, presented in the main body of text.

*Proof of Theorem 1.* If the comparison hypergraph is disconnected, then for any data $\mathcal{D}$, the (data-induced) comparison graph is disconnected too. Furthermore, the connected components of the comparison graph are subsets of those of the hypergraph. Partition the vertices into two non-empty subsets $S$ and $T$ such that there is no hyperedge between $S$ to $T$ in the comparison hypergraph. Let $A = \{i \mid N_i^+ \subset S\}$ and $B = \{i \mid N_i^+ \subset T\}$. By construction of the comparison hypergraph, $A \cap B = \varnothing$ and $A \cup B = V$. The log-likelihood can be therefore be rewritten as

$$\begin{aligned}\ell(\boldsymbol{\theta}) = &\sum_{i \in A} \sum_{j \in N_i^+} a_{ij} \left[ \log \lambda_j - \log \sum_{k \in N_i^+} w_{ik} \lambda_k \right] \\ &+ \sum_{i \in B} \sum_{j \in N_i^+} a_{ij} \left[ \log \lambda_j - \log \sum_{k \in N_i^+} w_{ik} \lambda_k \right].\end{aligned}$$

The sum over $A$ involves only parameters related to nodes in $S$, while the sum over $B$ involves only parameters related to nodes in $T$. Because the likelihood is invariant to a rescaling of the parameters, it is easy to see that we can arbitrarily rescale the parameters of the vertices in either $S$ or $T$ without affecting the likelihood. □




(a) network structure (b) comparison hypergraph (c) comparison graph

*Figure 1.* An innocent-looking example where the ML estimate does not exist. The network structure, aggregate traffic data and compatible transitions are shown on the left. While the comparison hypergraph is connected, the (data-dependent) comparison graph is not strongly connected.

**Verifying the condition of Theorem 6.** In order to verify the necessary and sufficient condition given $\{(c_i^-, c_i^+)\}$, one has to find a non-negative solution $\{a_{ij}\}$ to the system of equations

$$\sum_{j \in N_i^-} a_{ji} = c_i^-,$$
$$\sum_{j \in N_i^+} a_{ij} = c_i^+.$$

Dines (1926) presents a remarkably simple algorithm to find such a non-negative solution. Alternatively, Kumar et al. (2015) suggest recasting the problem as one of maximum flow in a network. However, the computational cost of running Dines' or max-flow algorithms is significantly greater than that of running ChoiceRank.

### B.2 Example

To conclude our discussion, we provide an innocuous-looking example that highlights the difficulty of dealing with the ML estimate. Consider the network structure and traffic data depicted in Figure 1. The network is strongly connected, and its comparison hypergraph is connected as well; as such, the network satisfies the necessary condition stated in Theorem 1 in the main text. Nevertheless, the condition is not sufficient for the ML-estimate to be well-defined. In this example, the (data-dependent) comparison graph is *not* strongly connected, and it is easy to see that the likelihood can always be increased by increasing $\lambda_1$, $\lambda_2$ and $\lambda_4$. Hence, the ML estimate does not exist.

In this simple example, we indicate the edge transitions that generated the observed marginal traffic in bold. Given this information, the comparison graph is easy to find, and the necessary and sufficient conditions of Theorem 6 are easy to check. But in general, finding a set of transitions that is compatible with given marginal per-node traffic data is computationally expensive (see discussion above).

## C ChoiceRank Algorithm

In this section, we start by generalizing the ChoiceRank algorithm to the weighted network choice model. We then prove the convergence of this generalized algorithm. Finally, we show how the same algorithm can be obtained from an EM viewpoint by introducing suitable latent variables.

### C.1 Algorithm for the Generalized Model

Using the same linear upper-bound on the logarithm as in Section 5 of the main text, we can lower-bound the log-posterior (4) in the weighted model by

$$f^{(t)}(\boldsymbol{\lambda}) = \kappa_2 + \sum_{i=1}^{n} \Bigg[ (c_i^- + \alpha - 1) \log \lambda_i - \beta \lambda_i - c_i^+ \left( \log \sum_{k \in N_i^+} w_{ik} \lambda_k^{(t)} + \frac{\sum_{k \in N_i^+} w_{ik} \lambda_k}{\sum_{k \in N_i^+} w_{ik} \lambda_k^{(t)}} - 1 \right) \Bigg], \tag{7}$$

with equality if and only if $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(t)}$. Starting with an arbitrary $\boldsymbol{\lambda}^{(0)} \in \mathbf{R}_{>0}^n$, we repeatedly maximize the lower-bound $f^{(t)}$. This surrogate optimization problem has a closed form solution, obtained by setting $\nabla f^{(t)}$ to 0:

$$\lambda_i^{(t+1)} = \frac{c_i^- + \alpha - 1}{\sum_{j \in N_i^-} w_{ji} \gamma_j^{(t)} + \beta}, \tag{8}$$

where

$$\gamma_j^{(t)} = \frac{c_j^+}{\sum_{k \in N_j^+} w_{jk} \lambda_k^{(t)}}.$$

The iterates provably converge to the maximizer of (4), as shown by the following theorem.

**Theorem 7.** *Let $\boldsymbol{\lambda}^\star$ be the unique maximum a-posteriori estimate. Then for any initial $\boldsymbol{\lambda}^{(0)} \in \mathbf{R}^n_{>0}$ the sequence of iterates defined by* (8) *converges to $\boldsymbol{\lambda}^\star$.*

The proof follows that of Hunter's Theorem 1 (2004).

*Proof.* Let $M : \mathbf{R}^n_{>0} \to \mathbf{R}^n_{>0}$ be the (continuous) map implicitly defined by one iteration of the algorithm. For conciseness, let $g(\boldsymbol{\lambda}) \doteq \log p(\boldsymbol{\lambda} \mid \mathcal{D})$. As $g$ has a unique maximizer and is concave using the reparametrization $\lambda_i = e^{\theta_i}$, it follows that $g$ has a single stationary point. First, observe that the minorization-maximization property guarantees that $g\left[M(\boldsymbol{\lambda})\right] \geq g(\boldsymbol{\lambda})$. Combined with the strict concavity of $g$, this ensures that $\lim_{t\to\infty} g(\boldsymbol{\lambda}^{(t)})$ exists and is unique for any $\boldsymbol{\lambda}^{(0)}$. Second, $g\left[M(\boldsymbol{\lambda})\right] = g(\boldsymbol{\lambda})$ if and only if $\boldsymbol{\lambda}$ is a stationary point of $g$, because the minorizing function is tangent to $g$ at the current iterate. It follows that $\lim_{t\to\infty} \boldsymbol{\lambda}^{(t)} = \boldsymbol{\lambda}^\star$. □

Theorem 3 of the main text follows directly by setting $w_{ij} \equiv 1$. For completeness, the edge-streaming implementation adapted to the weighted model is given in Algorithm 2. The only changes with respect to Algorithm 1 (presented in the main text) are in lines 4 and 7: Every message $\gamma_i$ or $\lambda_j$ flowing through an edge $(i, j)$ is multiplied by the edge weight $w_{ij}$.

**Algorithm 2** ChoiceRank for the weighted model

**Require:** graph $G = (V, E)$, counts $\{(c_i^-, c_i^+)\}$
1: $\boldsymbol{\lambda} \leftarrow [1, \ldots, 1]$
2: **repeat**
3: $\quad \boldsymbol{z} \leftarrow \mathbf{0}_n$ ▷ Recompute $\boldsymbol{\gamma}$
4: $\quad$ **for** $(i, j) \in E$ **do** $z_i \leftarrow z_i + w_{ij}\lambda_j$
5: $\quad$ **for** $i \in V$ **do** $\gamma_i \leftarrow c_i^+ / z_i$
6: $\quad \boldsymbol{z} \leftarrow \mathbf{0}_n$ ▷ Recompute $\boldsymbol{\lambda}$
7: $\quad$ **for** $(i, j) \in E$ **do** $z_j \leftarrow z_j + w_{ij}\gamma_i$
8: $\quad$ **for** $i \in V$ **do** $\lambda_i \leftarrow (c_i^- + \alpha - 1)/(z_i + \beta)$
9: **until** $\boldsymbol{\lambda}$ has converged

### C.2 EM Viewpoint

The MM algorithm can be seen from an EM viewpoint, following the ideas of Caron & Doucet (2012). We introduce $n$ independent random variables $\mathcal{Z} = \{Z_i \mid i = 1, \ldots, n\}$, where

$$Z_i \sim \text{Gamma}\Bigg(c_i^+, \sum_{j \in N_i^+} w_{ij}\lambda_j\Bigg).$$

With the addition of these latent random variables the complete log-likelihood becomes

$$\begin{aligned}
\ell(\boldsymbol{\lambda}; \mathcal{D}, \mathcal{Z}) &= \ell(\boldsymbol{\lambda}, \mathcal{D}) + \sum_{i=1}^n \log p(z_i \mid \mathcal{D}, \boldsymbol{\lambda}) \\
&= \sum_{i=1}^n \left[ c_i^- \log \lambda_i - c_i^+ \log \sum_{k \in N_i^+} w_{ik}\lambda_k \right] \\
&\quad + \sum_{i=1}^n \left[ c_i^+ \log \sum_{k \in N_i^+} w_{ik}\lambda_k - z_i \sum_{k \in N_i^+} w_{ik}\lambda_k \right] + \kappa_6 \\
&= \sum_{i=1}^n \left[ c_i^- \log \lambda_i - z_i \sum_{k \in N_i^+} w_{ik}\lambda_k \right] + \kappa_6.
\end{aligned}$$

Using a Gamma$(\alpha, \beta)$ prior for each parameter, the expected value of the log-posterior with respect to the conditional $\mathcal{Z} \mid \mathcal{D}$ under the estimate $\boldsymbol{\lambda}^{(t)}$ is

$$\begin{aligned}
Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(t)}) &= \mathbf{E}_{\mathcal{Z} \mid \mathcal{D}, \boldsymbol{\lambda}^{(t)}} \left[\ell(\boldsymbol{\lambda}; \mathcal{D}, \mathcal{Z})\right] + \log p(\boldsymbol{\lambda}) \\
&= \sum_{i=1}^n \left[ c_i^- \log \lambda_i - c_i^+ \frac{\sum_{k \in N_i^+} w_{ik}\lambda_k}{\sum_{k \in N_i^+} w_{ik}\lambda_k^{(t)}} \right] \\
&\quad + \sum_{i=1}^n \left[ (\alpha - 1)\log \lambda_i - \beta\lambda_i \right] + \kappa_7
\end{aligned}$$

The EM algorithm starts with an initial $\boldsymbol{\lambda}^{(0)}$ and iteratively refines the estimate by solving the optimization problem $\boldsymbol{\lambda}^{(t+1)} = \arg\max_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(t)})$. It is not difficult to see that for a given $\boldsymbol{\lambda}^{(t)}$, maximizing $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(t)})$ is equivalent to maximizing the minorizing function $f^{(t)}(\boldsymbol{\lambda})$ defined in (7). Hence, the MM and the EM viewpoint lead to the exact same sequence of iterates.

The EM formulation leads to a Gibbs sampler in a relatively straightforward way (Caron & Doucet, 2012). We leave a systematic treatment of Bayesian inference in the network choice model for future work.

## References

Caron, F. and Doucet, A. Efficient Bayesian Inference for Generalized Bradley–Terry models. *Journal of Computational and Graphical Statistics*, 21(1):174–196, 2012.

Casella, G. and Berger, R. L. *Statistical Inference*. Duxbury Press, second edition, 2002.

Dines, L. L. On Positive Solutions of a System of Linear Equations. *Annals of Mathematics*, 28(1/4):386–392, 1926.

Hunter, D. R. MM algorithms for generalized Bradley–Terry models. *The Annals of Statistics*, 32(1):384–406, 2004.

Kumar, Ravi, Tomkins, Andrew, Vassilvitskii, Sergei, and Vee, Erik. Inverting a Steady-State. In *WSDM'15*, pp. 359–368. ACM, 2015.