Appendix A. Proof of Lemma 1

We first use the standard quadratic bound (Jaakkola and Jordan, 1997; Bishop, 2006): for any $y \in \mathbb{R}$ and $\xi \in \mathbb{R}$,

$$\log(1+e^{-y}) \leqslant \lambda(\xi)(y^2-\xi^2) + (-y+\xi)/2 + \log(1+e^{-\xi}), \tag{29}$$

which implies:

$$\sigma(y) \geqslant \sigma(\xi) e^{(y-\xi)/2-\lambda(\xi)(y^2-\xi^2)}. \tag{30}$$

When $A = \{c_l, \bar{c}_l\}$, we have:

$$\begin{aligned} p(c=c_l|A) &\overset{(3)}{=} \frac{s_{c_l}}{s_{c_l}+s_{\bar{c}_l}} \overset{(9)}{=} \frac{1}{1+e^{-\boldsymbol{\theta}^T(\boldsymbol{x}_{c_l}-\boldsymbol{x}_{\bar{c}_l})}} = \sigma(\boldsymbol{x}_{c_l,\bar{c}_l}^T\boldsymbol{\theta}) \\ &\overset{(30)}{\geqslant} \sigma(\xi)\exp\Big((\boldsymbol{x}_{c_l,\bar{c}_l}^T\boldsymbol{\theta}-\xi)/2 - \lambda(\xi)((\boldsymbol{x}_{c_l,\bar{c}_l}^T\boldsymbol{\theta})^2-\xi^2)\Big). \end{aligned} \tag{31}$$

To prove Eq. (13), we use Bouchard's inequality (Bouchard, 2007), which states that for all $\boldsymbol{y} = [y_i]_i \in \mathbb{R}^K$ and all $\alpha \in \mathbb{R}$:

$$\log(\sum_{k=1}^K e^{y_k}) \leqslant \alpha + \sum_{k=1}^K \log(1+e^{y_k-\alpha}). \tag{32}$$

Combining Eq. (32) with Eq. (29), for every $\boldsymbol{\xi} = [\xi_i]_{i=1,\cdots,K} \in \mathbb{R}_+^K$ we get

$$\sum_{k=1}^K e^{y_k} \leqslant e^{\alpha} \prod_{k=1}^K \Big((1+e^{-\xi_k}) e^{(y_k-\alpha+\xi_k)/2+\lambda(\xi_k)((y_k-\alpha)^2-\xi_k^2)}\Big). \tag{33}$$

Hence, the top query probability under the Plackett Luce model satisfies:

$$\begin{aligned} p(c=c_l|A) &= \frac{\exp(\boldsymbol{x}_{c_l}^T\boldsymbol{\theta})}{\sum_{j\in A_l}\exp(\boldsymbol{x}_j^T\boldsymbol{\theta})} \\ &\overset{(3)(33)}{\geqslant} \frac{\exp(\boldsymbol{x}_{c_l}^T\boldsymbol{\theta})}{e^{\alpha}\prod_{j\in A}\Big((1+e^{-\xi_j})e^{(\boldsymbol{x}_j^T\boldsymbol{\theta}-\alpha+\xi_j)/2+\lambda(\xi_j)((\boldsymbol{x}_j^T\boldsymbol{\theta}-\alpha)^2-\xi_j^2)}\Big)} \\ &= \exp(\boldsymbol{x}_{c_l}^T\boldsymbol{\theta}-\alpha)\prod_{j\in A}\Big(\sigma(\xi_j)e^{(-\boldsymbol{x}_j^T\boldsymbol{\theta}+\alpha-\xi_j)/2-\lambda(\xi_j)((\boldsymbol{x}_j^T\boldsymbol{\theta}-\alpha)^2-\xi_j^2)}\Big) \end{aligned} \tag{34}$$

□

Appendix B. Proof of Lemma 2

The posterior $q(\boldsymbol{\theta})$ has the Gaussian form, i.e.:

$$q(\boldsymbol{\theta}) = \frac{1}{B_q} e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T\boldsymbol{S}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})} \tag{35}$$

where $B_q = (2\pi)^{d/2}|\boldsymbol{S}|^{1/2}$, $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{S} \in \mathbb{R}^{d\times d}$ and the ELBO $\mathbf{L}(q)$ satisfies:

$$\begin{aligned}\mathbf{L}(q) \overset{(11)}{=} &\mathbb{E}_{q(\boldsymbol{\theta})}\left[\log \frac{p_0(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right] + \mathbb{E}_{q(\boldsymbol{\theta})}\left[\sum_{l\in\mathcal{L}} \log \frac{\exp(\boldsymbol{\theta}^T \boldsymbol{x}_{c_l})}{\sum_{j\in A_l}\exp(\boldsymbol{\theta}^T\boldsymbol{x}_j)}\right]\\ =&\mathbb{E}_{q(\boldsymbol{\theta})}\left[\frac{(\boldsymbol{\theta}-\boldsymbol{\mu})^T\boldsymbol{S}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})}{2} - \frac{(\boldsymbol{\theta}-\boldsymbol{\mu}_0)^T\boldsymbol{S}_0^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_0)}{2}\right] + \frac{1}{2}\log\frac{|\boldsymbol{S}|}{|\boldsymbol{S}_0|}\\ &+ \mathbb{E}_{q(\boldsymbol{\theta})}\left[\sum_{l\in\mathcal{L}} \log \frac{\exp(\boldsymbol{\theta}^T \boldsymbol{x}_{c_l})}{\sum_{j\in A_l}\exp(\boldsymbol{\theta}^T\boldsymbol{x}_j)}\right].\end{aligned} \tag{36}$$

By Lemma 1, $\exp(\boldsymbol{\theta}^T\boldsymbol{x}_{c_l})/\sum_{j\in A_l}\exp(\boldsymbol{\theta}^T\boldsymbol{x}_j)$ is bounded by Q_l, and Lemma 2 follows. □

Appendix C. Proof of Lemma 3

The variational lower bound (14) is an expectation of a quadratic function of $\boldsymbol{\theta}$, and we can obtain the corresponding variational parameters $\boldsymbol{S}, \boldsymbol{\mu}$ by identifying the linear and quadratic terms in $\boldsymbol{\theta}$. To maximize the lower bound in Eq. (14), the quadratic term satisfies:

$$\boldsymbol{\theta}^T\boldsymbol{S}^{-1}\boldsymbol{\theta} = \boldsymbol{\theta}^T\boldsymbol{S}_0^{-1}\boldsymbol{\theta} + \boldsymbol{\theta}^T\Big(2\sum_{l\in\mathcal{L}_2}\lambda(\zeta_l^{(k)})\boldsymbol{x}_{c_l,\bar{c}_l}\boldsymbol{x}_{c_l,\bar{c}_l}^T + 2\sum_{l\in\mathcal{L}_{>2}}\sum_{j\in A_l}\lambda(\xi_{lj}^{(k)})\boldsymbol{x}_j\boldsymbol{x}_j^T\Big)\boldsymbol{\theta}, \tag{37}$$

and the linear term satisfies:

$$\boldsymbol{\theta}^T\boldsymbol{S}^{-1}\boldsymbol{\mu} = \boldsymbol{\theta}^T\boldsymbol{S}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\theta}^T\Bigg(\sum_{l\in\mathcal{L}_2}\frac{\boldsymbol{x}_{c_l,\bar{c}_l}}{2} + \sum_{l\in\mathcal{L}_{>2}}\Big(\boldsymbol{x}_{c_l} + \sum_{j\in A_l}(2\lambda(\xi_{lj}^{(k)})\alpha_l^{(k)}\boldsymbol{x}_j - \frac{\boldsymbol{x}_j}{2})\Big)\Bigg). \tag{38}$$

As the Eq. (37) and Eq. (38) hold for any $\boldsymbol{\theta}$, Lemma 3 follows.

□

Appendix D. Proof of Lemma 4

From (14), objective of (16b) is separable w.r.t. $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}, \boldsymbol{\alpha}$, and

$$\zeta_l^{(k+1)} = \underset{\zeta_l}{\mathbf{argmax}}\ \mathbb{E}_q[Q_l], \quad l \in \mathcal{L}_2, \tag{39}$$

where Q_l is given by (15). Hence, the optimal ζ_l is a stationary point:

$$\frac{\partial\mathbf{L}(\boldsymbol{\zeta},\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\mu}^{(k)},\boldsymbol{S}^{(k)})}{\partial\zeta_l} = -\lambda'(\zeta_l)\Big(\mathbb{E}_{q(\boldsymbol{\theta})}\big[(\boldsymbol{\theta}^T\boldsymbol{x}_{c_l,\bar{c}_l})^2\big] - \zeta_l^2\Big) = 0,\ l \in \mathcal{L}_2, \tag{40}$$

As $\lambda'(\zeta_l) > 0$, stationary points satisfy $\mathbb{E}_{q(\boldsymbol{\theta})}\big[(\boldsymbol{\theta}^T\boldsymbol{x}_{c_l,\bar{c}_l})^2\big] = \zeta_l^2$, which yields (19) for $\zeta_l \geq 0$. □

Appendix E. Proof of Lemma 5

By (15), we have that:

$$f_l(\boldsymbol{\xi}_l,\alpha_l) = -\alpha_l + \sum_{j\in A_l}\Big(\log\sigma(\xi_{lj}) + \frac{\alpha_l - \xi_{lj}}{2} - \lambda(\xi_{lj})\big(\mathbb{E}_{q(\boldsymbol{\theta})}[(\boldsymbol{x}_j^T\boldsymbol{\theta}-\alpha_l)^2] - \xi_{lj}^2\big)\Big). \tag{41}$$

This is a quadratic function with respect to α_l. The solution for Eq. (21a) is a stationary point, so we have:

$$\frac{\partial f_l}{\partial \alpha_l} = \frac{(m_l-2)}{2} - 2\sum_{j\in A_l}\lambda(\xi_{lj}^{(n)})\alpha_l + 2\sum_{j\in A_l}\lambda(\xi_{lj})\mathbb{E}_{q(\boldsymbol{\theta})}[\boldsymbol{x}_j^T\boldsymbol{\theta}] = 0,\ l\in\mathcal{L}_{>2}, \tag{42}$$

which implies Eq. (22). For $\boldsymbol{\xi}_l$, the solution for Eq. (21b) should also be a stationary point:

$$\frac{\partial f_l}{\partial \xi_{lj}} = -\lambda^{'}(\xi_{lj})\Big(\mathbb{E}_{q(\boldsymbol{\theta})}\big[(\boldsymbol{\theta}^T\boldsymbol{x}_j - \alpha_l^{(n+1)})^2\big] - \xi_{lj}^2\Big) = 0,\ l\in\mathcal{L}_{>2}, j\in A_l, \tag{43}$$

As $\lambda'(\xi_{lj}) > 0$, stationary points satisfy $\mathbb{E}_{q(\boldsymbol{\theta})}\big[(\boldsymbol{\theta}^T\boldsymbol{x}_{lj} - \alpha_l^{(n+1)})^2\big] = \xi_{lj}^2$, which implies (23) for $\xi_{lj}\geq 0$. □

Appendix F. Proof of Lemma 6

The derivation follows the same argument as in Jaakkola and Jordan (1997). Briefly, to see why Eq. (25) holds, note that in step (16a), covariance $\boldsymbol{S}$ and mean $\boldsymbol{\mu}$ are updated. Subsequently, $\mathbf{L}$ defined as in Eq. (14) is the sum of two terms: the KL divergence between two Gaussian distributions and a normalization factor. The optimum in step (16a) therefore occurs when the two Gaussian distributions are identical. Because of this, we can omit all quadratic and linear terms from (14) when computing (25): we only need to calculate the normalization factor by adding the constant items in Eq. (14).

In more details, quantity Q_l defined in Eq. (15) can be written as:

$$Q_l = Q_l^{'} + Q_l^{''} + \overline{Q}_l \tag{44}$$

where $Q_l^{'}, Q_l^{''}$ and $\overline{Q}_l$ are defined as:

$$Q_l^{'} = \begin{cases} \boldsymbol{x}_{c_l}^T\boldsymbol{\theta} + \sum_{j\in A_l}\Big[-\boldsymbol{x}_j^T\boldsymbol{\theta}/2 + 2\alpha_l\lambda(\xi_{lj})\big(\boldsymbol{x}_j^T\boldsymbol{\theta}\big)\Big], & l\in\mathcal{L}_{>2},\\ \boldsymbol{x}_{c_l,\bar{c}_l}^T\boldsymbol{\theta}/2, & l\in\mathcal{L}_2,\end{cases} \tag{45a}$$

$$Q_l^{''} = \begin{cases} -\sum_{j\in A_l}\lambda(\xi_{lj})\big(\boldsymbol{x}_j^T\boldsymbol{\theta}\big)^2, & l\in\mathcal{L}_{>2},\\ -\lambda(\zeta_l)(\boldsymbol{x}_{c_l,\bar{c}_l}^T\boldsymbol{\theta})^2, & l\in\mathcal{L}_2,\end{cases} \tag{45b}$$

$$\overline{Q}_l = \begin{cases} -\alpha_l + \sum_{j\in A_l}\Big[\log\sigma(\xi_{lj}) + \frac{\alpha_l-\xi_{lj}}{2} - \lambda(\xi_{lj})\big(\alpha_l^2-\xi_{lj}^2\big)\Big], & l\in\mathcal{L}_{>2},\\ \log\sigma(\zeta_l) - \zeta_l/2 + \lambda(\zeta_l)\zeta_l^2, & l\in\mathcal{L}_2.\end{cases} \tag{45c}$$

After step (16a), the following equations hold:

$$\mathbb{E}_{q(\boldsymbol{\theta})}\Big[\sum_{l\in\mathcal{L}}Q_l^{'}\Big] + \mathbb{E}_{q(\boldsymbol{\theta})}\Big[-\boldsymbol{\theta}^T\boldsymbol{S}^{-1}\boldsymbol{\mu} + \boldsymbol{\theta}^T\boldsymbol{S}_0^{-1}\boldsymbol{\mu}_0\Big] = 0, \tag{46a}$$

$$\mathbb{E}_{q(\boldsymbol{\theta})}\Big[\sum_{l\in\mathcal{L}}Q_l^{''}\Big] + \mathbb{E}_{q(\boldsymbol{\theta})}\Big[\boldsymbol{\theta}^T\boldsymbol{S}^{-1}\boldsymbol{\theta}/2 - \boldsymbol{\theta}^T\boldsymbol{S}_0^{-1}\boldsymbol{\theta}_0/2\Big] = 0, \tag{46b}$$

which are equivalent to Eq. (38) and Eq. (37). Thus, Eq. (14) can be written as:

$$\begin{aligned}\mathbf{L}(\boldsymbol{\zeta},\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{S}) =& \mathbb{E}_{q(\boldsymbol{\theta})}\left[\sum_{l\in\mathcal{L}}\overline{Q}_l\right] + \frac{1}{2}\log\frac{|\boldsymbol{S}|}{|\boldsymbol{S}_0|} + \mathbb{E}_{q(\boldsymbol{\theta})}\left[\frac{\boldsymbol{\mu}^T\boldsymbol{S}^{-1}\boldsymbol{\mu}}{2} - \frac{\boldsymbol{\mu}_0^T\boldsymbol{S}_0^{-1}\boldsymbol{\mu}_0}{2}\right] \\ =& \sum_{l\in\mathcal{L}}\overline{Q}_l + \frac{1}{2}\log\frac{|\boldsymbol{S}|}{|\boldsymbol{S}_0|} + \frac{\boldsymbol{\mu}^T\boldsymbol{S}^{-1}\boldsymbol{\mu}}{2} - \frac{\boldsymbol{\mu}_0^T\boldsymbol{S}_0^{-1}\boldsymbol{\mu}_0}{2}.\end{aligned} \tag{47}$$

This is equivalent to Eq. (25). □