A. Update equations for variational inference

A lower bound on the log likelihood of the CRF can be derived using Jensen’s inequality as shown in Eq. 1.

\[
\log p(s_v, \alpha, \beta | \Sigma_\alpha, \Sigma_\beta) \geq \mathbb{L}(q, \Psi_u, \Psi_p) = E_q [\log p(\alpha, \beta | \Sigma_\alpha, \Sigma_\beta)] + \sum_v E_q [\log p(s^v | \alpha, \beta)] + H
\]

where \( s^v \) is the complete role assignment to all people in the video \( v \), and \( H \) is the entropy of the variational distribution \( q \) shown in Eq. 2.

\[
q(\alpha, \beta, s_v | \lambda_\alpha, \lambda_\beta, \sigma_\alpha^2, \sigma_\beta^2, \phi, \psi) = \prod_j q(\alpha_j | \lambda_\alpha, \sigma_\alpha^2) \prod_k q(\beta_k | \lambda_\beta, \sigma_\beta^2) \prod_v q(s^v | \phi^v, \psi^v)
\]

Now, \( E_q [\log p(s^v | \alpha, \beta)] \) in Eq. 1 can be expanded as

\[
E_q [\log p(s^v | \alpha, \beta)] = \lambda_\alpha \cdot E_q [\Psi_u(s^v)] + \lambda_\beta \cdot E_q [\Psi_p(\phi^v, \psi^v)]
\]

where, \( Z_v = \log \left\{ \sum_{s^v} \exp \left( \alpha \cdot \Psi_u(s^v) + \beta \cdot \Psi_p(s^v) \right) \right\} \) is the log partition function. Using the fact, \( \log x \leq a^{-1}x - 1 + \log a \), we can establish a lower bound on \( E_q [\log p(s^v | \alpha, \beta)] \) as shown below

\[
E_q [\log p(s^v | \alpha, \beta)] \geq \lambda_\alpha \cdot E_q [\Psi_u(s^v)] + \lambda_\beta \cdot E_q [\Psi_p(\phi^v, \psi^v)] - \frac{h_v(q)}{\zeta_v} - \log(\zeta_v),
\]

where \( \zeta_v \) is a variational parameter and \( h_v(q) \) is defined as

\[
h_v(q) = \sum_{s^v} E_q \left[ \exp \left\{ \alpha \cdot \Psi_u(s^v) + \beta \cdot \Psi_p(s^v) \right\} \right]
\]

Given \( \Sigma_\alpha, \Sigma_\beta \), we update the parameters through a coordinate ascent method to maximize the lower bound in Eq. 1. \( \zeta_v \) is updated to \( h_v(q) \) at each iteration. The closed form update equations for \( \phi^v(p_i^v), \psi^v_j(p_j^v, s) \) at each iteration are shown in Eq. 6.

\[
\phi^v(p_i^v) \propto \exp \left\{ \lambda_{a,m} \cdot \Psi_u(p_i^v) + \sum_{j \neq i} \sum_{s \neq m} \psi^v_{ij}(p_j^v, s) \left[ \lambda_{bs} \cdot \Psi_p(p_j^v, p_i^v) \right] \right\}
\]

At each iteration, the values of \( \lambda_\alpha, \lambda_\beta, \) and \( \sigma_\alpha^2, \sigma_\beta^2 \) are updated using L-BFGS. The gradients of \( \mathbb{L} \) with respect to \( \lambda_\alpha \) and \( \lambda_\beta \) are given below

\[
\nabla_{\lambda_\alpha} \mathbb{L} = \sum \left\{ E_q[\Psi_u(p_i^v, s)] - \frac{h_v(q)}{\zeta_v} \right\}
\]

\[
\nabla_{\lambda_\beta} \mathbb{L} = \sum \left\{ E_q[\Psi_p(p_i^v, p_j^v, s)] - \frac{h_v(q)}{\zeta_v} \right\}
\]
where $\Sigma_\alpha, \Sigma_\beta$ are the components of $\Sigma_\alpha, \Sigma_\beta$ corresponding to $\alpha, \beta$ respectively. As before, $\Psi_p(p_i^v, p_j^v, s)$ is the pairwise feature when $p_i^v$ is the reference role and $p_j^v$ is assigned the role $s$. The gradients of $L$ with respect to $\sigma^2_\alpha$ and $\sigma^2_\beta$ are given below

$$\nabla_{\sigma^2_\alpha} L = -\frac{1}{2} \Sigma^{-1}_{\alpha} - \sum_v \zeta_v^{-1} \nabla_{\sigma^2_\alpha} h_v(q) + \frac{1}{2\sigma^2_\alpha},$$

$$\nabla_{\sigma^2_\beta} L = -\frac{1}{2} \Sigma^{-1}_{\beta} - \sum_v \zeta_v^{-1} \nabla_{\sigma^2_\beta} h_v(q) + \frac{1}{2\sigma^2_\beta},$$

where $\Sigma_{\alpha}^k$ and $\Sigma_{\beta}^k$ are the $k^{th}$ diagonal elements in $\Sigma_\alpha$ and $\Sigma_\beta$ respectively.

It is to be noted that the assumption of significant interaction only with the reference role, helps us exactly compute $h_v(q), \nabla \chi h_v(q), \nabla \sigma h_v(q)$ through a clique-tree message passing algorithm. The exact computation of $h_v(q)$ is intractable in a fully connected graph with interaction among all social roles.

**Implementation details** $\zeta_v$ is initialized to $1E6$ in our experiments. Also, the hyperparameters $\Sigma_\alpha$ and $\Sigma_\beta$ are diagonal matrices whose non-zero entries are all set to 0.01, 0.1, or 10 based on the variance of the unary and pairwise features. This is indicative of the amount of variance in the respective features.

**B. Optimization for role assignment from probabilities**

In every video $v$, the person $p_m^v$ with the highest value of $\phi^v(p_m^v)$ is assigned the reference role. The corresponding variational probability $\psi_{(m)}^v$ is then used to assign secondary roles to other people in the video. While assigning secondary roles, we enforce a lower $l$ and upper $u$ bound on the number of people assigned a secondary role $s$ in the event. Let $\mathbb{P}_{-mE}$ be the set of people not assigned the reference role in event $E$. Let $\psi$ be the secondary role probability matrix, where each row corresponds to a person $p_k \in \mathbb{P}_{-mE}$ and each column represents a secondary role $s$. $\psi$ can be obtained by stacking $\psi_{(m)}^v$ from all videos. $Y$ is the secondary role assignment matrix with same dimensions as $\psi$, where an entry $Y_{ks}$ is set to 1 if the person $p_k$ is assigned the role $s$. Secondary role assignment is then carried out by solving the linear integer program in Eq. 9 to maximize the probability of role assignment under given constraints.

$$\max_Y \text{Trace} \left( Y^T \psi \right),$$

subject to

$$Y 1 = 1, \quad l \leq Y^T 1 \leq u,$$

$$Y_{ks} \in \{0, 1\} \quad \forall p_k \in \mathbb{P}_{-mE}$$

We enforce the same constraints in our baseline models as well.