A Variational Bayes Generalized EM Algorithm

Given the mean-field approximation to the conditional distribution $f(\boldsymbol{Z}, \boldsymbol{\pi}|\mathcal{E}, \tilde{\boldsymbol{\theta}})$, the evidence lower bound (ELBO) of the log marginal likelihood can be derived as follow.

$$
\begin{aligned}
&ELBO := E_{q(\boldsymbol{Z},\boldsymbol{\pi})}(\log f(\boldsymbol{\theta}, \boldsymbol{Z}, \boldsymbol{\pi}|\mathcal{E}) - E_{q(\boldsymbol{Z},\boldsymbol{\pi})}(\log q(\boldsymbol{Z}, \boldsymbol{\pi}) \\
&E_{q(\boldsymbol{Z},\boldsymbol{\pi})}(\log f(\boldsymbol{\theta}, \boldsymbol{Z}, \boldsymbol{\pi}|\mathcal{E}) \\
&\quad\propto \sum_{m=1}^{M}\sum_{k=1}^{K} E_{\boldsymbol{Z}}(Z_{mk})\big[S_{e_{m1}} + R_{e_{m2}} + (\boldsymbol{U}_{e_{m1}} + \boldsymbol{V}_{e_{m2}})\boldsymbol{W}'_k - \log(f_{uk}) - \log(f_{vk} - e^{R_{e_{m1}} + V_{e_{m1}}\boldsymbol{W}'_k})\big] \\
&\quad - \frac{||\boldsymbol{S}||^2}{2\sigma_S^2} - \frac{||\boldsymbol{R}||^2}{2\sigma_R^2} - \frac{\sum_{i=1}^{n}\boldsymbol{U}_i\Sigma^{-1}\boldsymbol{U}'_i}{2\sigma_U^2} - \frac{\sum_{i=1}^{n}\boldsymbol{V}_i\Sigma^{-1}\boldsymbol{V}'_i}{2\sigma_V^2} - \frac{\sum_{k=1}^{K}\boldsymbol{W}_k\Sigma^{-1}\boldsymbol{W}'_k}{2} \\
&\quad + \Big(-\frac{a_S + n}{2} - 1\Big)\log\sigma_S^2 - \frac{b_S}{2\sigma_S^2} + \Big(-\frac{a_R + n}{2} - 1\Big)\log\sigma_R^2 - \frac{b_R}{2\sigma_R^2} \\
&\quad + \Big(-\frac{a_U + nP}{2} - 1\Big)\log\sigma_U^2 - \frac{b_U}{2\sigma_U^2} + \Big(-\frac{a_V + nP}{2} - 1\Big)\log\sigma_V^2 - \frac{b_V}{2\sigma_V^2} \\
&\quad - n\log\det(\Sigma) - \frac{K}{2}\log\det(\Sigma) - (a_d + 1)\log d - \frac{b_d}{d} \\
&\quad + \log\Gamma(K\alpha) - K\log\Gamma(\alpha) + \sum_{k=1}^{K}(\alpha + \tilde{\tilde{p}}_{.k} - 1)E_{\boldsymbol{\pi}}(\log\pi_k) + (a_a - 1)\log\alpha - b_a\alpha \\
&E_{q(\boldsymbol{Z},\boldsymbol{\pi})}(\log q(\boldsymbol{Z}, \boldsymbol{\pi})) \\
&\quad = \Big[\sum_{m=1}^{M} E_{q(\boldsymbol{Z})}(\log q(\boldsymbol{Z}_m))\Big] + E_{q(\boldsymbol{\pi})}(\log q(\boldsymbol{\pi}))
\end{aligned}
$$

Maximizing the ELBO is equivalent to minimizing the KL divergence between $q(\boldsymbol{Z}, \boldsymbol{\pi})$ and $f(\boldsymbol{\theta}, \boldsymbol{Z}, \boldsymbol{\pi}|\mathcal{E})$.The optimal optimal solutions are as follow (Wang & Blei, 2013; Beal, 2003).

$$
\begin{aligned}
q(\boldsymbol{Z}_m) &\propto e^{E_{\boldsymbol{\pi}}(\log\pi(\boldsymbol{Z}_m, \mathcal{E}|\boldsymbol{\pi}, \boldsymbol{Z}_{-m}\tilde{\boldsymbol{\theta}}))} \\
&= \exp\Big\{\sum_{k=1}^{K}\big(Z_{mk}\log(\frac{e^{S_{\boldsymbol{e}_{m1}} + \boldsymbol{U}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}{f_{uk}}\frac{e^{R_{\boldsymbol{e}_{m2}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}{f_{vk} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}) + Z_{mk}E_{\boldsymbol{\pi}}(\log\pi_k)\big)\Big\} \\
&= \prod_{k=1}^{K}\Big(\frac{e^{S_{\boldsymbol{e}_{m1}} + \boldsymbol{U}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}{f_{uk}}\frac{e^{R_{\boldsymbol{e}_{m2}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}{f_{vk} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}e^{E_{\boldsymbol{\pi}}(\log\pi_k)}\Big)^{Z_{mk}} \\
q(\boldsymbol{\pi}) &\propto e^{E_{\boldsymbol{Z}}(\log\pi(\boldsymbol{\pi}, \mathcal{E}|\boldsymbol{Z}, \tilde{\boldsymbol{\theta}})} \\
&= exp\Big\{\sum_{k=1}^{K}(\alpha + \sum_{m=1}^{M} E_{\boldsymbol{Z}}(Z_{mk}) - 1)\log\pi_k\Big\} \\
&= \prod_{k=1}^{K}\pi_k^{\alpha + \sum_{m=1}^{M} E_{\boldsymbol{Z}}(Z_{mk}) - 1}
\end{aligned}
$$

Notice that the optimal variational distribution of $\boldsymbol{Z}_m$ is a multinomial distribution with even probabilities $\tilde{\tilde{p}}_{mk} \propto \frac{e^{S_{\boldsymbol{e}_{m1}} + \boldsymbol{U}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}{f_{uk}}\frac{e^{R_{\boldsymbol{e}_{m2}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}{f_{vk} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}e^{E_{\boldsymbol{\pi}}(\log\pi_k)}$ and the optimal variational distribution of π is a dirichlet distribution with concentration parameter $\tilde{\tilde{\alpha}}_k = \alpha + \sum_{m=1}^{M} E_{\boldsymbol{Z}}(Z_{mk})$. We carried out the E-step as follow.

1. Initialize $\tilde{\tilde{p}}_{mk} \propto \frac{e^{S_{\boldsymbol{e}_{m1}} + \boldsymbol{U}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}{f_{uk}}\frac{e^{R_{\boldsymbol{e}_{m2}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}{f_{vk} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}}\boldsymbol{W}'_k}}$ for all m, k
2. Iterate over the following steps until convergence
   (a) $\tilde{\tilde{\alpha}}_k = \alpha + \sum_{m=1}^{M}\tilde{\tilde{p}}_{mk}$ for all k

(b) Calculate $E_{\boldsymbol{\pi}}(\log \pi_k) = \psi(\tilde{\tilde{\alpha}}_k) - \psi(\sum_{k=1}^K \tilde{\tilde{\alpha}}_k) = \psi(\tilde{\tilde{\alpha}}_k) - \psi(K\alpha + M)$, ψ is the Digamma function.

(c) $\tilde{\tilde{p}}_{mk} \propto \frac{e^{S_{\boldsymbol{e}_{m1}} + \boldsymbol{U}_{\boldsymbol{e}_{m1}} \boldsymbol{W}'_k}}{f_{uk}} \frac{e^{R_{\boldsymbol{e}_{m2}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}} \boldsymbol{W}'_k}}{f_{vk} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}} \boldsymbol{W}'_k}} e^{E_{\boldsymbol{\pi}}(\log \pi_k)}$ for all m, k

3. Upon convergence, we can evaluate $E_{\boldsymbol{Z}}(Z_{mk}) = \tilde{\tilde{p}}_{mk}$ and $E_{\boldsymbol{\pi}}(\log \pi_k) = \psi(\tilde{\tilde{\alpha}}_k) - \psi(K\alpha + M)$

B Gradients in Hamiltonian Monte Carlo-within Gibbs Algorithm

Let $z_m \in \{1, ..., K\}$ be the cluster assignment of the m-th edge. Let $\mathcal{M}_{i1} := \{m : \boldsymbol{e}_{m1} = i\}$ and $\mathcal{M}_{i2} := \{m : \boldsymbol{e}_{m2} = i\}$. The gradients with respect to $(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{R}, \boldsymbol{S}, \boldsymbol{W})$ are the same as in the LSEC paper (Sewell, 2021). The partial gradient of the log reduced conditional with respect to $(S_i, \boldsymbol{U}_i)$ is given by

$$\begin{aligned}
\frac{\partial \log f(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W} | (\cdot \setminus \boldsymbol{\pi}))}{\partial (S_i, \boldsymbol{U}_i)} &= \sum_{m \in \mathcal{M}_{i1}} \begin{pmatrix} 1 & \boldsymbol{W}_{z_m} \end{pmatrix} - \sum_{m=1}^{M} \frac{e^{S_i + \boldsymbol{U}_i \boldsymbol{W}'_{z_m}}}{f_{uz_m}} \begin{pmatrix} 1 & \boldsymbol{W}_{z_m} \end{pmatrix} - (S_i, \boldsymbol{U}_i) \begin{pmatrix} \tau_S & \boldsymbol{0} \\ \boldsymbol{0}' & \tau_U \mathcal{I}_p \end{pmatrix} \\
&= \sum_{m \in \mathcal{M}_{i1}} \begin{pmatrix} 1 & \boldsymbol{W}_{z_m} \end{pmatrix} - \sum_{k=1}^{K} (\sum_{m=1}^{M} \boldsymbol{Z}_{mk}) \frac{e^{S_i + \boldsymbol{U}_i \boldsymbol{W}'_k}}{f_{uk}} \begin{pmatrix} 1 & \boldsymbol{W}_k \end{pmatrix} - (S_i, \boldsymbol{U}_i) \begin{pmatrix} \tau_S & \boldsymbol{0} \\ \boldsymbol{0}' & \tau_U \mathcal{I}_p \end{pmatrix}.
\end{aligned}$$

If we precompute $\tilde{H}_k := \sum_{m=1}^{M} \frac{\mathbf{1}_{\{z_m = k\}}}{f_{vk} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}} \boldsymbol{W}'_k}}$, then the partial gradient of with respect to $(R_i, \boldsymbol{V}_i)$ is given by

$$\begin{aligned}
\frac{\partial \log f(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W} | (\cdot \setminus \boldsymbol{\pi}))}{\partial (R_i, \boldsymbol{V}_i)} &= \sum_{m \in \mathcal{M}_{i2}} \begin{pmatrix} 1 & \boldsymbol{W}_{z_m} \end{pmatrix} - \sum_{m \notin \mathcal{M}_{i1}} \frac{e^{R_i + \boldsymbol{V}_i \boldsymbol{W}'_{z_m}}}{f_{vz_m} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}} \boldsymbol{W}'_{z_m}}} \begin{pmatrix} 1 & \boldsymbol{W}_{z_m} \end{pmatrix} \\
&\quad - (R_i, \boldsymbol{V}_i) \begin{pmatrix} \tau_R & \boldsymbol{0} \\ \boldsymbol{0}' & \tau_V \mathcal{I}_p \end{pmatrix} \\
&= \sum_{m \in \mathcal{M}_{i2}} \begin{pmatrix} 1 & \boldsymbol{W}_{z_m} \end{pmatrix} - \sum_{k=1}^{K} e^{R_i + \boldsymbol{V}_i \boldsymbol{W}'_k} \left(\tilde{H}_k - \frac{|\mathcal{M}_{i1} \cap \{m : z_m = k\}|}{f_{vk} - e^{R_i + \boldsymbol{V}_i \boldsymbol{W}'_k}} \right) \begin{pmatrix} 1 & \boldsymbol{W}_k \end{pmatrix} \\
&\quad - (R_i, \boldsymbol{V}_i) \begin{pmatrix} \tau_R & \boldsymbol{0} \\ \boldsymbol{0}' & \tau_V \mathcal{I}_p \end{pmatrix}.
\end{aligned}$$

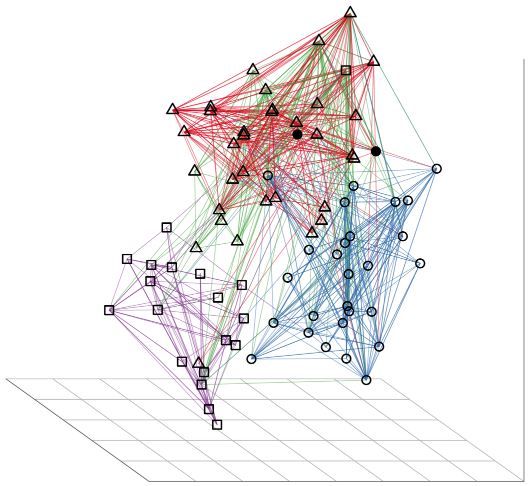
The partial gradient of with respect to $\boldsymbol{W}_k$ is given by

$$\frac{\partial \log f(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W} | (\cdot \setminus \boldsymbol{\pi}))}{\partial \boldsymbol{W}_k} = \sum_{m : z_m = k} \left(\boldsymbol{U}_{\boldsymbol{e}_{m1}} + V_{\boldsymbol{e}_{m2}} - \frac{\boldsymbol{s}_{uk}}{f_{uk}} - \frac{\boldsymbol{s}_{vk} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}} \boldsymbol{W}'_k} \boldsymbol{V}_{\boldsymbol{e}_{m1}}}{f_{vk} - e^{R_{\boldsymbol{e}_{m1}} + \boldsymbol{V}_{\boldsymbol{e}_{m1}} \boldsymbol{W}'_k}} \right) - \boldsymbol{W}_k,$$

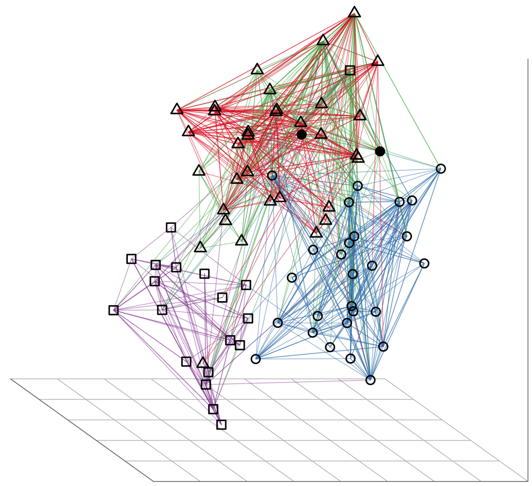
where $\boldsymbol{s}_{uk} := \sum_{i=1}^{n} e^{S_i + \boldsymbol{U}_i \boldsymbol{W}'_k} \boldsymbol{U}_i$ and $\boldsymbol{s}_{vk} := \sum_{i=1}^{n} e^{R_i + \boldsymbol{V}_i \boldsymbol{W}'_k} \boldsymbol{V}_i$.

C Sensitivity Results of Applying the GEM Algorithm to UK Faculty Network

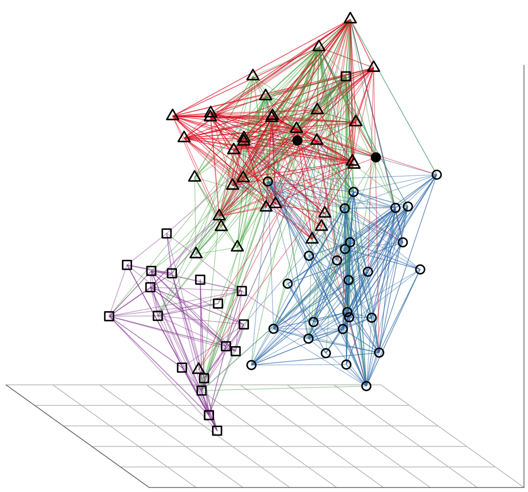
We carried out a sensitivity analysis for the GEM algorithm varying the number of cluster $K \in \{10, 20, 30\}$, number of dimension $p \in \{2, 3, 4\}$, the initial value of α (random draw from prior vs. set as prior mode), and the hyperparameters of the priors on the τ's $\{0.1, 1\}$. Higher values of K implies more superfluous clusters to empty out, and thus, requires a larger value of b_α. We set b_α to be 200, 400, and 600 for the value of K being 10, 20, and 30, respectively. We found that changing the initialization of α and the hyperparameters values did not have any major impact on the clustering results (Figure 1). Figure 2 showed the clustering results of varying K and p, while assuming that α was initialized at the prior mode and $a_U = b_U = a_S = b_S = a_V = b_V = a_R = b_R = 1$. There were minimal changes to the clustering results, except for when $p = 4$ and $K = 30$, the algorithm yielded an extra sub-cluster (yellow).



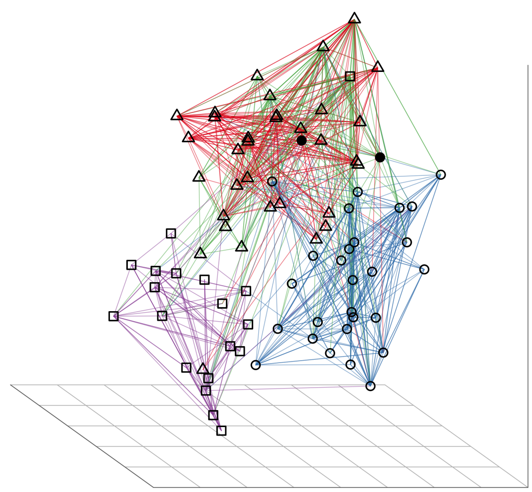
Hyperparameter = 0.1, random α initialization



Hyperparameter = 0.1, initialize $\alpha = a_\alpha / b_\alpha$



Hyperparameter = 1, random α initialization



Hyperparameter = 1, initialize $\alpha = a_\alpha / b_\alpha$

Figure 1: Sensitivity results of applying aLSEC to UK Faculty data set assuming different hyperparameters of the precision parameters and initialization mechanism for α, while assuming $K = 10, p = 3$. Edge colors correspond to different edge clusters. The three hollow shapes represent the three schools. The two solid circles represent the two individuals who did not mention their schools.

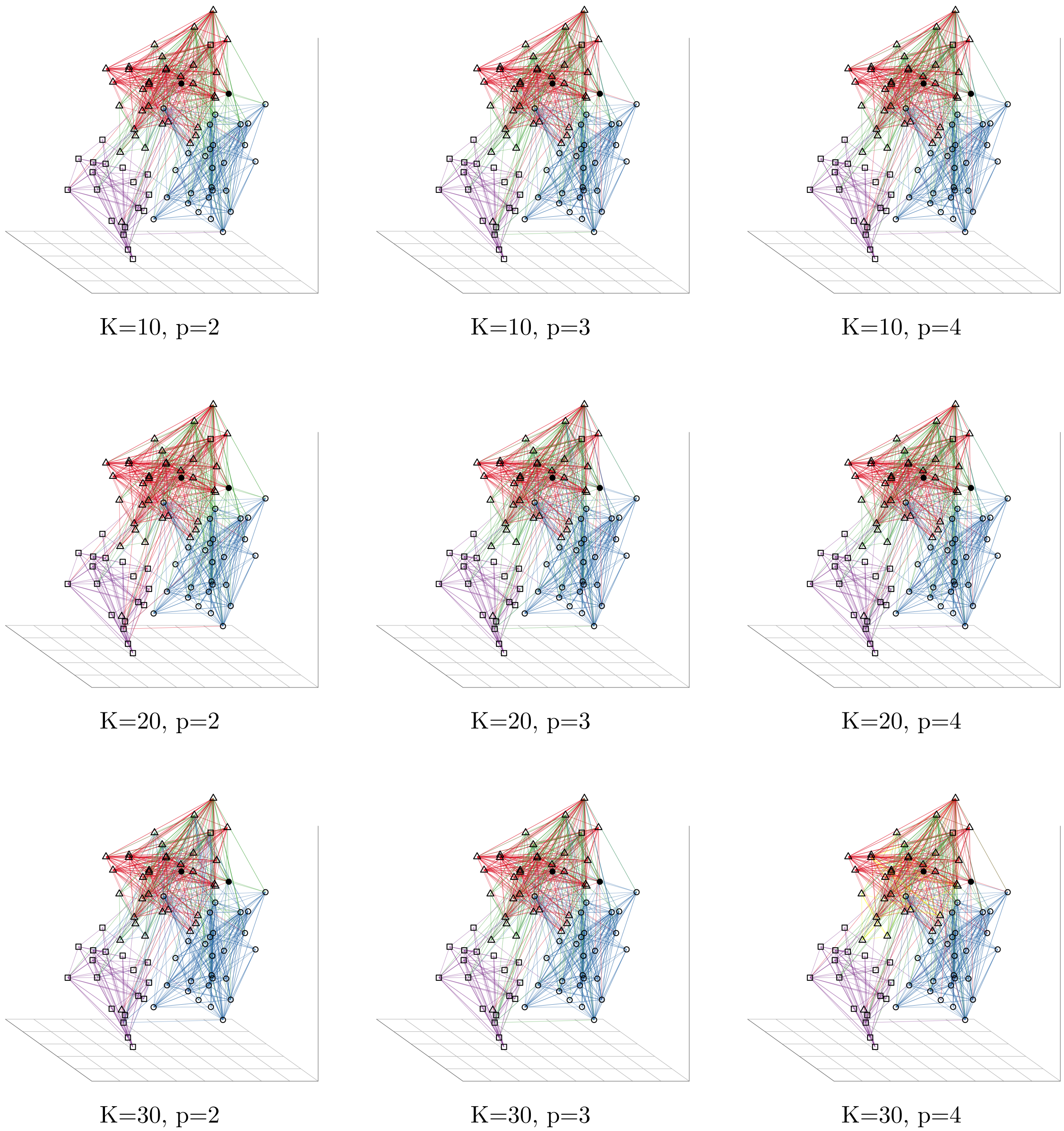


Figure 2: Sensitivity results of applying aLSEC to UK Faculty data set assuming different value of K, p, while setting hyperparameters of the precision parameters to be 0.1 and initializing $\alpha = a_\alpha/b_\alpha$. Edge colors correspond to different edge clusters. The three hollow shapes represent the three schools. The two solid circles represent the two individuals who did not mention their schools.