1 The Model

Let D_{nj} be the data corresponding to n-th sample and j-th feature, where n runs from 1 to N and j runs from 1 to G. We assume as a first pass that the data is Gaussian in its distribution.

Let us consider a graph model with K nodes (factors) and E edges (in our model right now, $E = K(K-1)/2$).

Let us define latent variables Z and Λ such that Z_n is a $E \times 1$ vector which is one-hot, where E is the total number of edges in the graph. Again Λ_n is a $Q \times 1$ one-hot vector, where Q is the cardinality of the set of values that q can take.

$$Pr(Z|\pi) = \prod_{n=1}^{N} \prod_{k_1<k_2} \pi_{k1,k2}^{z_{n,k1,k2}}$$

We assume that q can take a finite set of values between 0 and 1, say $1/100, 2/100, \cdots, 90/100, 1$. Also

$$Pr(\Lambda|\delta) = \prod_{n=1}^{N} \prod_{q=1}^{Q} \delta_q^{\lambda_{nq}}$$

here Q is 100, if there are 100 values that q can take.

We assume the model

$$E\left[D_{nj}|Z_{n,k_1,k_2} = 1, \lambda_{n,q} = 1, F\right] = qF_{k1,j} + (1-q)F_{k2,j} \tag{1}$$

The prior we define over Λ is equivalent to writing

$$Pr\left[\Lambda_{n,q} = 1\right] = \delta_q$$

Then we can write

$$Pr\left[D_n|Z_{n,k_1,k_2} = 1, F, s^2_{j=1,2,\cdots,J}\right] = \sum_q \delta_q Pr\left[D_n|Z_{n,k_1,k_2} = 1, \Lambda_{n,q} = 1, F, s^2_{j=1,2,\cdots,J}\right]$$

where s_j^2 is the variance of the jth feature.

We also get from the prior on $Z_{n,k1,k2}$,

$$Pr\left[Z_{n,k_1,k_2} = 1\right] = \pi_{k_1,k_2} \qquad k_1 < k_2$$

Then we can write

$$Pr\left[D_n|\pi, F\right] = \sum_{k_1<k_2} \pi_{k_1,k_2} Pr\left[D_n|Z_{n,k_1,k_2} = 1, F, s^2_{j=1,2,\cdots,J}\right]$$

We define the joint prior over the edges and the fraction of the edge represented as

$$\pi_{k_1,k_2,q} = \pi_{k_1,k_2}\delta_q \qquad k_1 < k_2$$

The overall likelihood

$$L(\pi, F) = \prod_{n=1}^{N} Pr\left[D_n|\pi, F, s^2_{j=1,2,\cdots,J}\right]$$

or we can write it as

$$L(\pi, F) = \prod_{n=1}^{N} \sum_{k_1<k_2} \sum_{q} \left[\pi_{k_1,k_2,q} \times \prod_{j=1}^{G} N\left(D_{nj}; qF_{k_1,g} + (1-q)F_{k_2,g}, s_j^2\right)\right]$$

And the log likelihood

$$\ln L(\pi, F) = \sum_{n=1}^{N} \ln\left(\sum_{k_1<k_2} \sum_{q} \left[\pi_{k_1,k_2,q} \times \prod_{j=1}^{G} N\left(D_{nj}; qF_{k_1,g} + (1-q)F_{k_2,g}, s_j^2\right)\right]\right) \qquad (2)$$

This is the quantity we want to maximize.

2 EM algorithm

2.1 E step

We assume that q can take a finite set of values between 0 and 1, say $1/100, 2/100, \cdots, 90/100, 1$.

Suppose we have run upto m iterations. For the $(m+1)$th iteration, we have

$$\begin{aligned}
\delta^{(m+1)}_{n,k_1,k_2,q} &= Pr\left[Z_{n,k1,k2} = 1, \Lambda_{n,q} = 1|\pi^{(m)}, F^{(m)}, s^{(m)}_{j=1,2,\cdots,J}, D_n\right] \\
&\propto Pr\left[Z_{n,k1,k2} = 1\right] Pr\left[\lambda_{n,q} = 1\right] Pr\left[D_n|\pi^{(m)}, F^{(m)}, s^{(m)}_{j=1,2,\cdots,J}, Z_{n,k1,k2} = 1, \lambda_{n,q} = 1\right] \\
&\propto \pi^{(m)}_{k_1,k_2,q} \prod_{j} N\left(D_{nj}|qF^{(m)}_{k_1,j} + (1-q)F^{(m)}_{k_2,j}, {s^{(m)}_j}^2\right)
\end{aligned}$$

where ${s^{(m)}_j}^2$ is the residual variance of feature j.

We normalize δ so that

$$\sum_{k_1<k_2} \sum_{q} \delta^{(m+1)}_{n,k_1,k_2,q} = 1 \qquad \forall n$$

We define

$$\pi^{(m+1)}_{k_1,k_2,q} = \frac{1}{N} \sum_{n=1}^{N} \delta^{(m+1)}_{n,k_1,k_2,q}$$

We have therefore updated $\pi^{(m)}_{k_1,k_2,q}$ to $\pi^{(m+1)}_{k_1,k_2,q}$.

2.1.1 Variational EM - Model 1

In this set up, we assume prior distributions of π and δ as follows

$$Pr(\pi|\alpha_0) = C(\alpha_0) \prod_{k_1<k_2} \pi_{k1,k2}^{\alpha_0 - 1}$$

Similarly the prior distribution for δ is

$$Pr(\delta|\beta_0) = C(\beta_0) \prod_{q=1}^{Q} \delta_q^{\beta_0 - 1}$$

The likelihood above can be written as

$$p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J}) = \prod_{n=1}^{N} \prod_{k_1<k_2} \prod_{q=1}^{Q} [\prod_{j=1}^{G} N(D_{ng}|qF_{k1,g} + (1-q)F_{k2,g}, s_g^2)]^{\Lambda_{nq} Z_{n,k1,k2}}$$

The joint probablity distribution distribution is given by

$$p(D,Z,\Lambda,\pi,\delta|F,s_{j=1,2,\cdots,J},\alpha_0,\beta_0) = p(\pi|\alpha_0)p(\delta|\beta_0)p(\Lambda|\delta)p(Z|\pi)p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})$$

We assume the following mean field variational distribution. In the first model, we assume the two latent variables Z and Λ are independent.

$$q(Z,\Lambda,\pi,\delta) = q(Z)q(\Lambda)q(\pi)q(\delta)$$

The variational distribution for Z

$$\begin{aligned}
\ln q^\star(Z) &= E_{\pi,\delta,\Lambda}\left[\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] \\
&= E_{\pi,\delta,\Lambda}\left[\ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] + constant \\
&= \sum_{n=1}^{N}\sum_{k_1<k_2} z_{n,k1,k2} E_\pi\left[\ln(\pi_{k1,k2})\right] + \sum_{n=1}^{N}\sum_{k_1<k_2} z_{n,k1,k2} \sum_{q} E_\Lambda(\lambda_{nq}) \left[-\sum_{j=1}^{G}\ln(s_j)\right. \\
&\qquad \left. -\frac{G}{2}\ln(2\pi) - \sum_{j=1}^{G}\frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2}\right]
\end{aligned}$$

$$\begin{aligned}
\ln q^\star(\Lambda) &= E_{\pi,\delta,Z}\left[\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] \\
&= E_{\pi,\delta,Z}\left[\ln p(\Lambda|\delta) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] + constant \\
&= \sum_{n=1}^{N}\sum_{q=1}^{Q} \lambda_{n,q} E_\delta\left[\ln(\delta_q)\right] + \sum_{n=1}^{N}\sum_{q} \lambda_{nq} \sum_{k_1<k_2} E_Z(z_{n,k1,k2}) \left[-\sum_{j=1}^{G}\ln(s_j)\right. \\
&\qquad \left. -\frac{G}{2}\ln(2\pi) - \sum_{j=1}^{G}\frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2}\right]
\end{aligned}$$

So we get

$$q^{\star}(Z) \propto \prod_{n=1}^{N} \prod_{k_1<k_2} \rho_{n,k1,k2}^{Z_{n,k1,k2}}$$

where we define

$$\rho_{n,k1,k2} \propto exp \Bigg(E_{\pi} \left[\ln(\pi_{k1,k2}) \right] + \sum_{q} E_{\Lambda}(\lambda_{nq}) \Bigg[- \sum_{j=1}^{G} \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^{G} \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \Bigg]$$

$$\rho_{n,k1,k2} \propto exp \left(E_{\pi} \left[\ln(\pi_{k1,k2}) \right] + \sum_{q} \nu_{nq} \left[- \sum_{j=1}^{G} \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^{G} \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

$$\rho_{n,k1,k2} \propto exp \left(\psi_{a_{k1,k2}} - \psi(\sum_{k1<k2} a_{k1,k2}) + \left[- \sum_{j=1}^{G} \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^{G} \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

$\rho_{n,k1,k2}$ is normalized to sum to 1 for each n over $k1$ and $k2$.
We also get

$$q^{\star}(\Lambda) \propto \prod_{n=1}^{N} \prod_{q=1}^{Q} \nu_{nq}^{\Lambda_{nq}}$$

where

$$\nu_{nq} \propto exp \left(E_{\delta} \left[\ln(\delta_q) \right] + \sum_{k_1<k_2} E_Z(z_{n,k1,k2}) \left[- \sum_{j=1}^{G} \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^{G} \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

$$\nu_{nq} \propto exp \left(E_{\delta} \left[\ln(\delta_q) \right] + \sum_{k_1<k_2} \rho_{n,k1,k2} \left[- \sum_{j=1}^{G} \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^{G} \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

$$\nu_{nq} \propto exp \left(\psi(b_q) - \psi(\sum_{q=1}^{Q} b_q) + \left[- \sum_{j=1}^{G} \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^{G} \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

ν_{nq} are normalized to sum to 1.
We can also derive variational distributions similarly for π and δ.

$$\begin{aligned}
\ln q^\star(\pi) &= E_{\Lambda,Z,\delta}\left[\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] \\
&= E_Z\left[\ln p(Z|\pi)\right] + \ln p(\pi|\alpha_0) + constant \\
&= \sum_{n=1}^{N}\sum_{k1<k2} E(z_{n,k1,k2}) \ln \pi_{k1,k2} + (\alpha_0 - 1)\sum_{k1<k2} \ln \pi_{k1,k2} \\
&= \sum_{k1<k2}\left[\sum_{n=1}^{N} \rho_{n,k1,k2} + (\alpha_0 - 1)\right] \ln \pi_k
\end{aligned}$$

We define

$$a_{k1,k2} = \alpha_0 + \sum_{n=1}^{N} \rho_{n,k1,k2}$$

$$q^\star(\pi) = Dir(\pi|a)$$

$$\begin{aligned}
\ln q^\star(\delta) &= E_{\Lambda,Z,\pi}\left[\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] \\
&= E_\Lambda\left[\ln p(\Lambda|\delta)\right] + \ln p(\delta|\beta_0) + constant \\
&= \sum_{n=1}^{N}\sum_{q=1}^{Q} E(\lambda_{n,q}) \ln \delta_q + (\beta_0 - 1)\sum_{q=1}^{Q} \ln \delta_q \\
&= \sum_{q=1}^{Q}\left[\sum_{n=1}^{N} \nu_{n,q} + (\beta_0 - 1)\right] \ln \delta_q
\end{aligned}$$

We define

$$b_q = \beta_0 + \sum_{n=1}^{N} \nu_{n,q}$$

$$q^\star(\delta) = Dir(\delta|b)$$

We alternate between the Variational E and M steps, E steps being the ones where we compute the responsibilities $\rho_{n,k1,k2}$ and $\nu_{n,q}$ and the M step is where we update the variational distribution of the parameters π and δ.

We can start with $a = \alpha_0$ and $b = \beta_0$. We can then estimate $\rho_{n,k1,k2}$ and also ν_{nq} and then then product of these two terms to get new responsibility

$$\delta_{n,k1,k2,q} = \rho_{n,k1,k2}\nu_{nq}$$

and we use this $\delta_{n,k1,k2,q}$ as the responsibility for the M-step of the original EM updates.

2.1.2 Variational EM - Model 2

In model 2, we do not assume independence of the latent variables Z and Λ and instead estimate their joint variational distribution.

$$q(Z,\Lambda,\pi,\delta) = q(Z,\Lambda)q(\pi)q(\delta)$$

$$\begin{aligned}
\ln q^{\star}(Z,\Lambda) &= E_{\pi,\delta}\left[\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] \\
&= E_{\pi,\delta,\Lambda}\left[\ln p(Z|\pi) + \ln p(\Lambda|\delta) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] + constant \\
&= \sum_{n=1}^{N}\sum_{k_1<k_2}\sum_{q=1}^{Q} z_{n,k1,k2}\lambda_{nq}E_{\pi}\left[\ln(\pi_{k1,k2})\right] + \sum_{n=1}^{N}\sum_{q=1}^{Q}\sum_{k_1<k_2}\lambda_{n,q}z_{n,k1,k2}E_{\delta}\left[\ln(\delta_q)\right] \\
&\quad + \sum_{n=1}^{N}\sum_{k_1<k_2}\sum_{q} z_{n,k1,k2}\lambda_{nq}\left[-\sum_{j=1}^{G}\ln(s_j) - \frac{G}{2}\ln(2\pi) - \sum_{j=1}^{G}\frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2}\right]
\end{aligned} \tag{3}$$

From here one can get

$$q^{\star}(Z,\Lambda) \propto \prod_{n=1}^{N}\prod_{k_1<k_2}\prod_{q=1}^{Q}\delta_{n,k1,k2,q}^{Z_{n,k1,k2}\Lambda_{n,q}}$$

then

$$\begin{aligned}
\delta_{n,k1,k2,q} \propto exp\Big(&E_{\pi}\left[\ln(\pi_{k1,k2})\right] + E_{\delta}\left[\ln(\delta_q)\right] \\
&+ \left[-\sum_{j=1}^{G}\ln(s_j) - \frac{G}{2}\ln(2\pi) - \sum_{j=1}^{G}\frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2}\right]\Big)
\end{aligned} \tag{4}$$

$$\begin{aligned}
\delta_{n,k1,k2,q} \propto exp\Bigg(&\psi_{a_{k1,k2}} - \psi(\sum_{k1<k2} a_{k1,k2}) + \psi(b_q) - \psi(\sum_{q=1}^{Q} b_q) \\
&\left[-\sum_{j=1}^{G}\ln(s_j) - \frac{G}{2}\ln(2\pi) - \sum_{j=1}^{G}\frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2}\right]\Bigg)
\end{aligned} \tag{5}$$

The updates for π and δ are same as before. Here also we have the same way of initializing π and δ first, then use $a_{k1,k2} = \alpha_0$ and $b_q = \beta_0$ to begin with and estimate $\delta_{n,k1,k2,q}$. Then use the $\delta_{n,k1,k2,q}$ to update $a_{k1,k2}$ and b_q and proceed in this way. In this case, we do not assume independence of the Λ and Z variational distributions, so this model is more generalized.

2.2 M step

We define the parameter

$$\theta := (\pi_{k_1,k_2,q}, F, s_{j=1,2,\cdots,J})$$

We define the complete loglikelihood

$$logL_c(\theta; D, Z, \lambda) = log\pi_{k_1,k_2,q} + logL(D|Z, \lambda, q, F)$$

We take the expectation of this quantity with respect to $\left[Z, \lambda | D, \theta^{(m)}\right]$.

$$Q(\theta|\theta^{(m)}) \propto -\sum_{n=1}^{N} \sum_{k_1<k_2} \sum_{q} \delta^{(m+1)}_{n,k_1,k_2,q} \sum_{j} \left[logs_j^{(m+1)} + \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^{(m+1)2}} \right] \quad (6)$$

We try to maximize this quantity with respect to F, So, we can take derivative with respect to F and try to solve the resulting normal equation.

This equation, conditional on $\left[Z, \lambda | D, \theta^{(m)}\right]$, can be written as

$$D_{N\times J} = L_{N\times K} F_{K\times J} + E_{N\times J} \quad (7)$$

where

$$e_{nj} \sim N(0, s_j^2)$$

We define

$$D'_{nj} := \frac{D_{nj}}{s_j}$$

If we consider finding the factors on a feature-by-feature basis, we do not need to worry about s_j.

$$L_{nk} = \begin{cases} q \text{ or } (1-q) & \lambda_n = q \\ 0 & \text{o.w.} \end{cases}$$

We have

$$E_{Z,\lambda|D,\theta^{(m)}}\left[L_{nk}\right] = \sum_{q} \sum_{k_2>k} q\delta^{(m+1)}_{n,k,k_2,q} + \sum_{q} \sum_{k_1<k} (1-q)\delta^{(m+1)}_{n,k1,k,q}$$

$$E_{Z,\lambda|D,\theta^{(m)}}\left[L^2_{nk}\right] = \sum_{q} \sum_{k_2>k} q^2\delta^{(m+1)}_{n,k,k_2,q} + \sum_{q} \sum_{k_1<k} (1-q)^2\delta^{(m+1)}_{n,k1,k,q}$$

Also for any $k \neq l$,

$$E_{Z,\lambda|D,\theta^{(m)}}\left[L_{nk}L_{nl}\right] = \sum_{q} q(1-q)\delta_{n,k,l,q}^{(m+1)}$$

We use these to solve for the equation

$$\left[E_{Z,\lambda|D,\theta^{(m)}}\left(L^T L\right)\right] F \approx \left[E_{Z,\lambda|D,\theta^{(m)}}(L)\right]^T D$$

The solution therefore is

$$F \approx \left[E_{Z,\lambda|D,\theta^{(m)}}\left(L^T L\right)\right]^{-1} \left[E_{Z,\lambda|D,\theta^{(m)}}(L)\right]^T D$$

For $W = L^T L$

$$W_{kl} = \sum_{n} L_{kn}L_{nl}$$

$$E_{Z,\lambda|D,\theta^{(m)}}\left(W_{kl}\right) = \sum_{n} E_{Z,\lambda|D,\theta^{(m)}}\left(L_{nk}L_{nl}\right)$$

We use the definition of $E_{Z,\lambda|D,\theta^{(m)}}\left[L_{nk}^2\right]$ and $E_{Z,\lambda|D,\theta^{(m)}}\left[L_{nk}L_{nl}\right]$ from above to solve F.

In the same way as we computed F by solving for the normal equation obtained from taking derivative of the function $Q(\theta|\theta^{(m)})$, we take derivative of the latter with respect to s_j^2 to obtain EM updates of the residual variance terms. Taking the derivative, we obtain the estimate as

$$\widehat{s_j^{(m+1)}}^2 = \frac{1}{N}\sum_{n=1}^{N}\sum_{k_1<k_2}\sum_{q} \delta_{n,k_1,k_2,q}^{(m+1)}\left(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j}\right)^2 \tag{8}$$

where the F are the estimated values of the factors from the previous step.

We then continue this procedure described above for multiple iterations.