Outline for Mean Field Variational Inference in Network Only model

In [?], they only mention they use the simplification for the indicator variational parameter for links and non links but there is no indication whether they use also stochastic variational inference.Their updates make no mention of the natural gradients, so we might as well be concerned with the whole data for now, and later come up with exact natural gradient updates. But from the updates that we derived because of the exponential family criteria the updates using natural gradients would not be different at least in most cases if not all.

The network generaive process is as follows.

Algorithm 1 Network Generation

$\forall k \in \{1 \dots K\}$
    $\beta_k \sim Beta(\eta_0, \eta_1)$
$\forall a \in \{1 \dots N\}$
    $\theta_a \sim Dir(\alpha)$
    $\forall b > a, b \in \{1 \dots N\}$
        $z_{a \to b} \sim Mult(\theta_a)$
        $z_{a \leftarrow b} \sim Mult(\theta_b)$
        $y_{ab} \sim Bern(z_{a \to b}^T B \, z_{a \leftarrow b})$

Accordingly I created networks, which graphs below explain some of their properties.

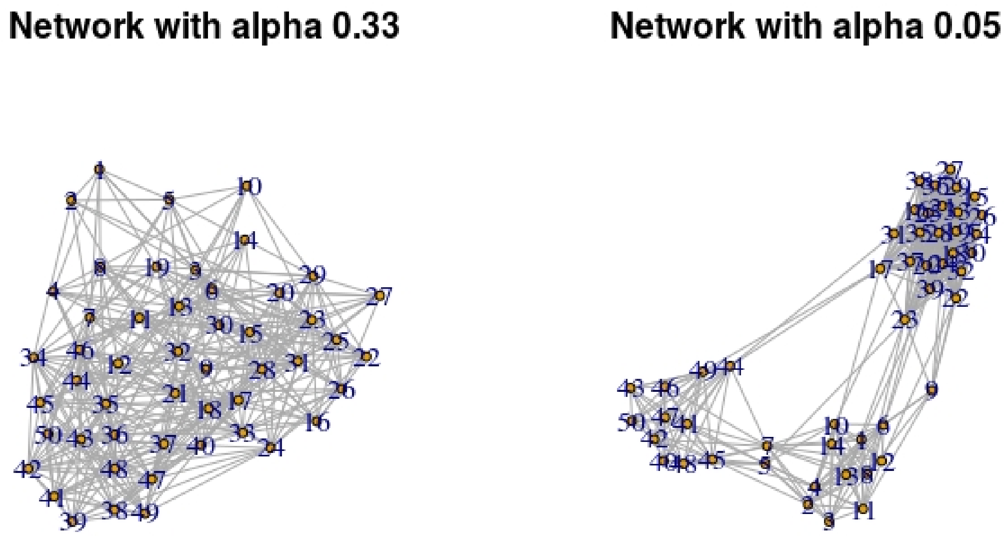

Figure 1: Networks with $\alpha = 0.33$ and $\alpha = 0.05$ from left to right

In the figure the network consists 50 nodes, and the η_0, and η_1 are set to 10, and 1 respectively to assure high chance of connectivity if two nodes share the same community when interacting with each other. To ensure different distinction distinction between roles, α can be manipulated. To observe more distinct roles we can set α to smaller values, as in the right panel of 1.

To show the values of θ's for each individual, I sort the the matrix including all individual θ's by their argmax position, so that those with higher value in the first community appear in the top-left, those with higher values for the second community appear in the middle, and those with higher values for the 3rd community appear in the bottom-right corner of the plot. The values are represented by the shade of darkness, where the darker it is the higher the value of θ for that person is.

Similarly we can show the network in this way as a binary matrix, and sorted

Figure 2: visualized θ matrix for $\alpha = 0.1$ and $K = 5$

by their θ positions as above.

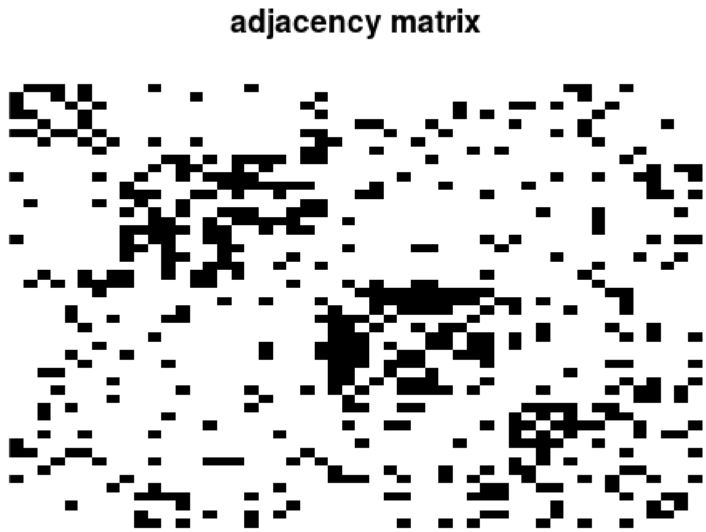

Figure 3: Adjacency matrix

So for now outline here is the for the variational inference for the whole data sampling but using the simplification impled for the indicator variational parameters in [**?**]. This means that we stick with the mean field variational inference for the time being and later extend it to the stochastic variational inference.

Algorithm 2 outline for mean field variational approximation

initialize *global variables*
repeat
    for each *local variational parameter*
        update *the local variational parameter*
    end for
    update *the global parameter*
until *ELBO converges*

1 The Variational Lower Bound

Here we explain the procedure in the **repeat** loop. The ELBO for the network only model is as follows:

$$
\begin{aligned}
\mathcal{L} \quad = \quad & \sum_{\substack{a,b \\ \in links}} \mathbb{E}_q\Big[log\, p(y_{ab}|z_{a\to b}, z_{b\to a}, B)\Big] \\
& + \sum_{\substack{a,b \\ \in nonlinks}} \mathbb{E}_q\Big[log\, p(y_{ab}|z_{a\to b}, z_{b\to a}, B)\Big] \\
& + \sum_{\substack{a,b \\ \in links}} \mathbb{E}_q\Big[log\, p(z_{a\to b}|\theta_a)\Big] + \sum_{\substack{a,b \\ \in links}} \mathbb{E}_q\Big[log\, p(z_{b\to a}|\theta_b)\Big] \\
& - \sum_{\substack{a,b \\ \in links}} \mathbb{E}_q\Big[log\, q(z_{a\to b}, z_{a\leftarrow b}|\phi_{ab})\Big] \qquad (1) \\
& + \sum_{\substack{a,b \\ \in nonlinks}} \mathbb{E}_q\Big[log\, p(z_{a\to b}|\theta_a)\Big] - \sum_{\substack{a,b \\ \in nonlinks}} \mathbb{E}_q\Big[log\, q(z_{a\to b}|\phi_{a\to b})\Big] \\
& + \sum_{\substack{a,b \\ \in nonlinks}} \mathbb{E}_q\Big[log\, p(z_{b\to a}|\theta_b)\Big] - \sum_{\substack{a,b \\ \in nonlinks}} \mathbb{E}_q\Big[log\, q(z_{b\to a}|\phi_{b\to a})\Big] \\
& + \sum_a \mathbb{E}_q\Big[log\, p(\theta_a|\alpha)\Big] - + \sum_a \mathbb{E}_q\Big[log\, q(\theta_a|\gamma_a)\Big] \\
& + \sum_k \mathbb{E}_q\Big[log\, p(\beta_k|\eta_0, \eta_1)\Big] - \sum_k \mathbb{E}_q\Big[log\, q(\beta_k|\tau_{k,0}, \tau_{k,1})\Big] \qquad (2)
\end{aligned}
$$

When dealing with the individual pairs, we distinguish between the case where there is a link between the two versus when a link is absent. This is mostly crucial in determining the complexity of the inference. We know that the communities indicated by the role indicator when there is a link is the same as the ones in the case of the absent link. As each variational element requires a variational parameter, we would need $\mathcal{O}(N^2)$ variational parameters for the the role indicators. This however complicates the estimation, and is quite wasteful of space and time. As the communities are similar in both cases of links and non-links, we replace the variational parameters of the role indicators for the case of non-links with the average of those of the links.

$$
\phi_{a\to m,k} \quad = \quad \frac{\sum_{b\in links(a)} \phi_{ab}^{kk}}{deg(a)} = \bar{\phi}_{a,k} \qquad (3)
$$

For short we show the variational parameter $\phi_{a\rightarrow b,k}$, and $\phi_{b\rightarrow a,k}$ in the case of links between a and b, as ϕ_{ab}^{kk}(both activating role k). Note that $\sum_{k\neq l}\phi_{ab}^{kl}=0$, and as the $\beta_{kl}=\epsilon$ in this case and $\epsilon\rightarrow 0$, then $\phi_{ab}^{kl}\propto exp\{-\infty\}$.

In the following through multiple stages we simplify the lower bound to make it easier for derivations and coding as well:

$$
\begin{aligned}
\mathcal{L} \;=\;& \sum_{a}\sum_{b\in links(a)}\sum_{k}\mathbb{E}_q\Bigg[z_{a\rightarrow b,k}z_{b\rightarrow a,k}log\,\beta_{k,0}+(1-z_{a\rightarrow b,k}z_{b\rightarrow a,k})log\,\epsilon\Bigg]\\
+&\sum_{a}\sum_{b\in nonlinks(a)}\sum_{k}\mathbb{E}_q\Bigg[z_{a\rightarrow b,k}z_{b\rightarrow a,k}log\,\beta_{k,1}+(1-z_{a\rightarrow b,k}z_{b\rightarrow a,k})log\,(1-\epsilon)\Bigg]\\
+&\sum_{a}\sum_{b\in links(a)}\sum_{k}\mathbb{E}_q\Bigg[z_{a\rightarrow b,k}log\,\theta_{a,k}+z_{b\rightarrow a,k}log\,\theta_{b,k}\Bigg]\\
-&\sum_{a}\sum_{b\in links(a)}\sum_{k}\mathbb{E}_q\Bigg[z_{a\rightarrow b,k}z_{b\rightarrow a,k}\;log\,\phi_{ab}^{kk}\Bigg]\\
+&\sum_{a}\sum_{b\in nonlinks(a)}\sum_{k}\mathbb{E}_q\Bigg[z_{a\rightarrow b,k}log\,\theta_{a,k}-z_{a\rightarrow b,k}log\,\phi_{a\rightarrow b,k}\Bigg]\\
+&\sum_{a}\sum_{b\in nonlinks(a)}\sum_{k}\mathbb{E}_q\Bigg[z_{b\rightarrow a,k}log\,\theta_{b,k}-z_{b\rightarrow a,k}log\,\phi_{b\rightarrow a,k}\Bigg]\\
+&\sum_{a}log\,\Gamma(\sum_{k}\alpha_k)-\sum_{a}\sum_{k}log\,\Gamma(\alpha_k)+\sum_{a}\sum_{k}(\alpha_k-1)\mathbb{E}_q\Bigg[log\,\theta_{a,k}\Bigg]\\
-&\sum_{a}log\,\Gamma(\sum_{k}\gamma_{a,k})+\sum_{a}\sum_{k}log\,\Gamma(\gamma_{a,k})-\sum_{a}\sum_{k}(\gamma_{a,k}-1)\mathbb{E}_q\Bigg[log\,\theta_{a,k}\Bigg]\\
+&\sum_{k}log\,\Gamma(\eta_0+\eta_1)-\sum_{k}log\,\Gamma(\eta_0)-\sum_{k}log\,\Gamma(\eta_1)+\sum_{k}(\eta_0-1)\mathbb{E}_q\Bigg[log\,\beta_{k,0}\Bigg]\\
+&\sum_{k}(\eta_1-1)\mathbb{E}_q\Bigg[log\,\beta_{k,1}\Bigg]\\
-&\sum_{k}log\,\Gamma(\tau_{k,0}+\tau_{k,1})+\sum_{k}log\,\Gamma(\tau_{k,0})+\sum_{k}log\,\Gamma(\tau_{k,1})-\sum_{k}(\tau_{k,0}-1)\mathbb{E}_q\Bigg[log\,\beta_{k,0}\Bigg]\\
-&\sum_{k}(\tau_{k,1}-1)\mathbb{E}_q\Bigg[log\,\beta_{k,1}\Bigg] \qquad (4)
\end{aligned}
$$

We can take the obvious expectations and the bound becomes(is the first line correct?):

$$
\begin{aligned}
\mathcal{L} \quad = \quad & \sum_{a,b \in links} \sum_{k} \left(\phi_{ab}^{kk} \mathbb{E}_q \left[log\, \beta_{k,0} \right] + (1 - \phi_{ab}^{kk}) log\, \epsilon \right) \\
+ \quad & \sum_{a,b \notin links} \sum_{k} \left(\bar{\phi}_{a,k} \bar{\phi}_{b,k} \mathbb{E}_q \left[log\, \beta_{k,1} \right] + (1 - \bar{\phi}_{a,k} \bar{\phi}_{b,k}) log\, (1 - \epsilon) \right) \\
+ \quad & \sum_{a,b \in links} \sum_{k} \left(\phi_{ab}^{kk} \mathbb{E}_q \left[log\, \theta_{a,k} \right] + \phi_{ab}^{kk} \mathbb{E}_q \left[log\, \theta_{b,k} \right] \right) \\
- \quad & \sum_{a,b \in links} \sum_{k} \left(\phi_{ab}^{kk} log\, \phi_{ab}^{kk} \right) \\
+ \quad & \sum_{a,b \notin links} \sum_{k} \left(\bar{\phi}_{a,k} \mathbb{E}_q \left[log\, \theta_{a,k} \right] - \bar{\phi}_{a,k} log\, \bar{\phi}_{a,k} \right) \\
+ \quad & \sum_{a,b \notin links} \sum_{k} \left(\bar{\phi}_{b,k} \mathbb{E}_q \left[log\, \theta_{b,k} \right] - \bar{\phi}_{b,k} log\, \bar{\phi}_{b,k} \right) \\
+ \quad & \sum_{a} log\, \Gamma(\sum_{k} \alpha_k) - \sum_{a} \sum_{k} log\, \Gamma(\alpha_k) + \sum_{a} \sum_{k} (\alpha_k - 1) \mathbb{E}_q \left[log\, \theta_{a,k} \right] \\
- \quad & \sum_{a} log\, \Gamma(\sum_{k} \gamma_{a,k}) + \sum_{a} \sum_{k} log\, \Gamma(\gamma_{a,k}) - \sum_{a} \sum_{k} (\gamma_{a,k} - 1) \mathbb{E}_q \left[log\, \theta_{a,k} \right] \\
+ \quad & \sum_{k} log\, \Gamma(\eta_0 + \eta_1) - \sum_{k} log\, \Gamma(\eta_0) - \sum_{k} log\, \Gamma(\eta_1) + \sum_{k} (\eta_0 - 1) \mathbb{E}_q \left[log\, \beta_{k,0} \right] \\
+ \quad & \sum_{k} (\eta_1 - 1) \mathbb{E}_q \left[log\, \beta_{k,1} \right] \\
- \quad & \sum_{k} log\, \Gamma(\tau_{k,0} + \tau_{k,1}) + \sum_{k} log\, \Gamma(\tau_{k,0}) + \sum_{k} log\, \Gamma(\tau_{k,1}) - \sum_{k} (\tau_{k,0} - 1) \mathbb{E}_q \left[log\, \beta_{k,0} \right] \\
- \quad & \sum_{k} (\tau_{k,1} - 1) \mathbb{E}_q \left[log\, \beta_{k,1} \right]
\end{aligned} \tag{5}
$$

Taking the remaining expectations for Beta and Dirichlet distribution parameters:

$$
\begin{aligned}
\mathcal{L} \;=\;& \sum_{a,b\in links}\sum_{k}\left(\phi_{ab}^{kk}(\Psi(\tau_{k,0})-\Psi(\tau_{k,0}+\tau_{k,1}))+(1-\phi_{ab}^{kk})log\,\epsilon\right)\\
+\;& \sum_{a,b\notin links}\sum_{k}\left(\bar{\phi}_{a,k}\bar{\phi}_{b,k}(\Psi(\tau_{k,1})-\Psi(\tau_{k,0}+\tau_{k,1}))+(1-\bar{\phi}_{a,k}\bar{\phi}_{b,k})log\,(1-\epsilon)\right)\\
+\;& \sum_{a,b\in links}\sum_{k}\left(\phi_{ab}^{kk}(\Psi(\gamma_{a,k})-\Psi(\sum_{k}\gamma_{a,k}))+\phi_{ab}^{kk}(\Psi(\gamma_{b,k})-\Psi(\sum_{k}\gamma_{b,k}))\right)\\
-\;& \sum_{a,b\in links}\sum_{k}\left(\phi_{ab}^{kk}log\,\phi_{ab}^{kk}\right)\\
+\;& \sum_{a,b\notin links}\sum_{k}\left(\bar{\phi}_{a,k}(\Psi(\gamma_{a,k})-\Psi(\sum_{k}\gamma_{a,k}))-\bar{\phi}_{a,k}log\,\bar{\phi}_{a,k}\right)\\
+\;& \sum_{a,b\notin links}\sum_{k}\left(\bar{\phi}_{b,k}(\Psi(\gamma_{b,k})-\Psi(\sum_{k}\gamma_{b,k}))-\bar{\phi}_{b,k}log\,\bar{\phi}_{b,k}\right)\\
+\;& \sum_{a}log\,\Gamma(\sum_{k}\alpha_k)-\sum_{a}\sum_{k}log\,\Gamma(\alpha_k)+\sum_{a}\sum_{k}(\alpha_k-1)(\Psi(\gamma_{a,k})-\Psi(\sum_{k}\gamma_{a,k}))\\
-\;& \sum_{a}log\,\Gamma(\sum_{k}\gamma_{a,k})+\sum_{a}\sum_{k}log\,\Gamma(\gamma_{a,k})-\sum_{a}\sum_{k}(\gamma_{a,k}-1)(\Psi(\gamma_{a,k})-\Psi(\sum_{k}\gamma_{a,k}))\\
+\;& \sum_{k}log\,\Gamma(\eta_0+\eta_1)-\sum_{k}log\,\Gamma(\eta_0)-\sum_{k}log\,\Gamma(\eta_1)+\sum_{k}(\eta_0-1)(\Psi(\tau_{k,0})-\Psi(\tau_{k,0}+\tau_{k,1}))\\
+\;& \sum_{k}(\eta_1-1)(\Psi(\tau_{k,1})-\Psi(\tau_{k,0}+\tau_{k,1}))\\
-\;& \sum_{k}log\,\Gamma(\tau_{k,0}+\tau_{k,1})+\sum_{k}log\,\Gamma(\tau_{k,0})+\sum_{k}log\,\Gamma(\tau_{k,1})-\sum_{k}(\tau_{k,0}-1)(\Psi(\tau_{k,0})\\
-\;& \Psi(\tau_{k,0}+\tau_{k,1}))-\sum_{k}(\tau_{k,1}-1)(\Psi(\tau_{k,1})-\Psi(\tau_{k,0}+\tau_{k,1}))
\end{aligned}
\tag{6}
$$

This is all we have for the variational lower bound

So we need to make sure we have a good computation for each term as we go on. As we need to keep track of each variational update, we can use those inputs for the their contribution to the lower bound on the go.

2 Variational updates

The updates for the network only models are:

2.1 The local update

$$\begin{aligned}\phi_{ab}^{kk}|ab &\in links \\ &\propto exp\left\{\Psi(\tau_{k,0}) - \Psi(\tau_{k,0}+\tau_{k,1}) + \Psi(\gamma_{a,k}) - \Psi(\sum_k \gamma_{a,k}) + \Psi(\gamma_{b,k}) - \Psi(\sum_k \gamma_{b,k})\right\}\end{aligned} \quad (7)$$

after this we need to normalize the ϕ_{ab}^{kk} as over k they need to sum to 1. so $\sum_k \phi_{ab}^{kk} = 1$.

After this we can also update the $\bar{\phi}_{a,k}$. As mentioned before

2.2 The global update

$$\tau_{k,0} \leftarrow \sum_{a,b \in links} \phi_{ab}^{kk} + \eta_0 \quad (8)$$

$$\tau_{k,1} \leftarrow \sum_{a,b \in nonlinks} \phi_{a\to b,k}\phi_{b\to a,k} + \eta_1$$

or simply

$$\tau_{k,1} \leftarrow \eta_1 + (\sum_a \bar{\phi}_{a,k} \sum_a \bar{\phi}_{a,k} - \sum_a (\bar{\phi}_{a,k})^2)/2 \ - \sum_{(a,b)\in links} \bar{\phi}_{a,k}\bar{\phi}_{b,k} \quad (9)$$

$$\gamma_{a,k} \leftarrow \alpha_k + \sum_{b \in links(a)} \phi_{ab}^{kk} + \sum_{b \in nonlinks(a)} \phi_{a\to b,k}$$

or simply

$$\gamma_{a,k} \leftarrow \alpha_k + \sum_{b \in links(a)} \phi_{ab}^{kk} + \big(N - 1 - deg(a)\big)\bar{\phi}_{a,k} \quad (10)$$

The complete algorithm then becomes :

Algorithm 3 Outline for network inference

---

repeat
local part
  for a **in** $\mathcal{N}$
    for b **in** $links(a)$
      for k **in** $1:K$
        $\phi_{ab}^{kk^{(t)}} = exp\Big\{\Psi(\tau_{k,0}) - \Psi(\tau_{k,0}+\tau_{k,1}) + \Psi(\gamma_{a,k}) - \Psi(\sum_k \gamma_{a,k}) + \Psi(\gamma_{b,k}) - \Psi(\sum_k \gamma_{b,k})\Big\}$
        $\bar{\phi}_{a,k}^{(t)} = \frac{\sum_{b\in links(a)} \phi_{ab}^{kk}}{deg(a)}$ modify this
      end for
    end for
  end for
global part
  for k **in** $1:K$
    for a **in** $\mathcal{N}$
      $\gamma_{a,k}^{(t)} = \alpha_k + \sum\limits_{(a,b)\in links(a)} \phi_{ab}^{kk} + (N-1-deg(a))\bar{\phi}_{a,k}$
    end for
    $\tau_{k,0}^{t} = \eta_0 + \sum\limits_{(a,b)\in links} \phi_{ab}^{kk}$
    $\tau_{k,1}^{t} = \eta_1 + \frac{\sum\limits_a \bar{\phi}_{a,k} \sum\limits_a \bar{\phi}_{a,k} - \sum\limits_a (\bar{\phi}_{a,k})^2}{2} - \sum\limits_{(a,b)\in links} \bar{\phi}_{a,k}\bar{\phi}_{b,k}$
  end for
compute $\mathcal{L}$ *on the go*
until $\mathcal{L}^{(t)} - \mathcal{L}^{(t-1)} < 1e-08$

We stop the process if the difference between the two most recent lower bounds are negligible(~1e-8) or in other words when we have convergence.

For now, I am fixing the model parameters α and η to their true values. But we need to also derive these in the M-step, where we optimize the lower bound fixing the variational parameters. For now I am skipping this part.

After the convergence I estimate them as below:

$$\hat{\beta}_{k,k} = \frac{\tau_{0,k}}{\tau_{0,k}+\tau_{1,k}}$$

$$\hat{\theta}_{a,k} = \frac{\gamma_{a,k}}{\sum_l \gamma_{a,l}}$$

References