Explore of the fuzzy community structure integrating the directed line graph and likelihood optimization (Supplemental Materials)

Hui-Jia Li[1] and Xiang Ju[2]

[1]*School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100080, China.*

[2]*Neuroscience Research Center, Changsha Medical University, Changsha 410219, Hunan, China.*

(Dated: May 16, 2016)

I. THE PROOF OF EQ.(3) IN THE MAIN TEXT

In terms of the parameter of X, Θ, Δ, Z and Ω, we have:

$$\Phi = \Theta B_g Z D^{-1}, \Psi = \Delta B_g Z D^{-1} \tag{1}$$

where $D = diag(n\Omega)$.

Proof. We have

$$\phi_{pq} = \sum_{i \in C_q} \frac{1}{N_q} \theta_{pi} \tag{2}$$

where $i \in C_q$ denotes node i is in the cluster q with a size N_q, and $\frac{1}{N_q}$ is the probability of selecting node i from cluster q. Further, we have:

$$\phi_{pq} = \frac{1}{n\omega_q} \sum_{i=1}^{n} \theta_{pi} (B_g Z)_{iq}. \tag{3}$$

Similarly, we have:

$$\psi_{pq} = \frac{1}{n\omega_q} \sum_{i=1}^{n} \delta_{pi} (B_g Z)_{iq}. \tag{4}$$

Thus, we have

$$\Phi = \Theta B_g Z D^{-1}, \Psi = \Delta B_g Z D^{-1} \tag{5}$$

The Proof is end.

II. THE PROOF OF EQ.(5) IN THE MAIN TEXT

$$L(N|X, B_g) = \sum_{l=1}^{L} \sum_{b_{il} \neq 0} \sum_{k=j}^{K} \prod_{j=1}^{n} f(\theta_{kj}, a_{ij}) f(\delta_{kj}, a_{ji}) \omega_k \tag{6}$$

where $f(x, y) = x^y (1-x)^{1-y}$.

Proof. Let $v = i$ denote the event that a node with linkage structure $< a_{i1}, ..., a_{in}, a_{1i}, ...a_{ni} >$ will be observed in network N. Let $y = k$ denote the event that the cluster label assigned to a node is equal to k. Let $i \rightarrow_{a_{ij}} j$ denote the event that node v_i may or may not link to node v_j, depending on a_{ij}. Let $i \leftarrow_{a_{ji}} j$ denote the event that node v_i may or may not be linked by node v_j depending on a_{ji}. We have:

$$
\begin{aligned}
L(N|X,B_g) &= \ln \Pi_{i=1}^{n} P(v=i) = \sum_{i=1}^{n} \ln P(v=i) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} P(v=i, y=k) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} P(v=i|y=k)P(y=k) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} (P(< a_{i1}, ..., a_{in}, a_{1i}, ...a_{ni} > |y=k)P(y=k)) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} (\prod_{j=1}^{n} P(i \rightarrow_{a_{ij}} j|y=k)P(i \leftarrow_{a_{ji}} j|y=k)P(y=k)) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} (\prod_{j=1}^{n} (\theta_{kj}^{a_{ij}}(1-\theta_{kj})^{1-a_{ij}})(\delta_{kj}^{a_{ji}}(1-\delta_{kj})^{1-a_{ji}})\omega_k) \\
&= \sum_{i=1}^{n} \sum_{k=j}^{K} \prod_{j=1}^{n} f(\theta_{kj}, a_{ij}) f(\delta_{kj}, a_{ji}) \omega_k \\
&= \sum_{l=1}^{L} \sum_{b_{il}\neq 0} \sum_{k=j}^{K} \prod_{j=1}^{n} f(\theta_{kj}, a_{ij}) f(\delta_{kj}, a_{ji}) \omega_k
\end{aligned}
\tag{7}
$$

The proof is end.

III. THE PROOF OF THEOREM 1 IN THE MAIN TEXT

Theorem 1. A local optimum of maximizing Eq.(6) will be guaranteed by recursively calculating Eq.(8) and (9):

$$
\begin{cases}
\theta_{kj} = \dfrac{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} a_{ij}\gamma_{lk}}{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} \gamma_{lk}} \\
\delta_{kj} = \dfrac{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} a_{ji}\gamma_{lk}}{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} \gamma_{lk}} \\
\omega_k = \dfrac{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} \gamma_{lk}}{n}
\end{cases}
\tag{8}
$$

$$
\begin{aligned}
\gamma_{lk} =& \frac{1}{\sum_{i=1}^{n} b_{il}} \times \\
& \sum_{b_{il}\neq 0} \frac{\prod_{j=1}^{n} f(\theta_{kj}, a_{ij}) f(\delta_{kj}, a_{ji}) w_k}{\sum_{k=1}^{K} \prod_{j=1}^{n} f(\theta_{kj}, a_{ij}) f(\delta_{kj}, a_{ji}) w_k},
\end{aligned}
\tag{9}
$$

where $\gamma_{lk} = E[z_{lk}] = P(y = k|b = l, X, B_g)$,i.e., the probability that module l is labeled with community k given X and B_g.

Proof. Let $L(G, Z|X, B_g)$ be the log-likelihood of the joint distribution of line graph G and Z given X and B_g, we have

$$L(G, Z|X, B_g) = \sum_{l=1}^{L} \sum_{b_{il}\neq 0} \sum_{k=1}^{K} z_{lk} (\sum_{j=1}^{n} (\ln f(\theta_{kj}, a_{ij}) + \ln f(\delta_{kj}, a_{ji})) + \ln \omega_k) \tag{10}$$

The detailed proof of Eq.(10) are shown in Eq.(21). Considering the expectation of $L(G, Z|X, B_g)$ on Z, we have:

$$E[L(G, Z|X, B_g)] = \sum_{l=1}^{L} \sum_{b_{il}\neq 0} \sum_{k=1}^{K} \gamma_{lk} (\sum_{j=1}^{n} (\ln f(\theta_{kj}, a_{ij}) + \ln f(\delta_{kj}, a_{ji})) + \ln \omega_k) \tag{11}$$

where $E[z_{lk}] = \gamma_{lk} = P(y = k|b = l, X, B_g)$,i.e. the probability that block l will be labeled as cluster k given X and B_g. Let $J = E[L(G, Z|X, B_g)] + \lambda(\sum_{k=1}^{K} w_k = 1)$, we have:

$$\begin{cases} \frac{\partial J}{\partial \theta_{kj}} = 0 \\ \frac{\partial J}{\partial \delta_{kj}} = 0 \\ \frac{\partial J}{\partial \omega_k} = 0 \\ \frac{\partial J}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} \theta_{kj} = \frac{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} a_{ij}\gamma_{lk}}{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} \gamma_{lk}} \\ \delta_{kj} = \frac{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} a_{ji}\gamma_{lk}}{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} \gamma_{lk}} \\ \omega_k = \frac{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} \gamma_{lk}}{\sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{b_{il}\neq 0} \gamma_{lk}} = \frac{\sum_{l=1}^{L} \sum_{b_{il}\neq 0} \gamma_{lk}}{n} \end{cases}$$

let $P(y = k|v = i)$ be the probability that node i belongs to cluster k given X and B_g, We have:

$$\gamma_{lk} = P(y = k|b = l, X, B_g) = \sum_{b_{il}\neq 0} \frac{1}{\sum_{i=1}^{n} b_{il}} P(y = k|v = i) \tag{12}$$

where $\frac{1}{\sum_{i=1}^{n} b_{il}}$ is the probability of selecting node i from block l. According to the Bayesian theorem, we have:

$$P(y = k|v = i) = \frac{P(y = k)P(v = i|y = k)}{\sum_{k=1}^{K} P(y = k)P(v = i|y = k)}. \tag{13}$$

Furthermore, we have:

$$P(y = k)P(v = i|y = k) = \prod_{j=1}^{n} f(\theta_{kj}, a_{ij}) f(\delta_{kj}, a_{ji}) \omega_k \tag{14}$$

Thus, we have:

$$\gamma_{lk} = \frac{1}{\sum_{i=1}^{n} b_{il}} \sum_{b_{il}\neq 0} \frac{\prod_{j=1}^{n} f(\theta_{kj}, a_{ij}) f(\delta_{kj}, a_{ji}) w_k}{\sum_{k=1}^{K} \prod_{j=1}^{n} f(\theta_{kj}, a_{ij}) f(\delta_{kj}, a_{ji}) \omega_k} \tag{15}$$

Moreover, we have:

$$
\begin{aligned}
L(G|X, B_g) &= \sum_{i=1}^{n} \ln P(v=i|X, B_g) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} P(v=i, y=k|X, B_g) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} P(y=k|v=i, X^{(s)}, B_g) \frac{P(v=i, y=k|X, B_g)}{P(y=k|v=i, X^{(s)}, B_g)} \\
& (by\ Jensen's\ inequality) \\
&\geq \sum_{i=1}^{n} \sum_{k=1}^{K} P(y=k|v=i, X^{(s)}, B_g) \ln \frac{P(v=i, y=k|X, B_g)}{P(y=k|v=i, X^{(s)}, B_g)} \\
&\equiv G(X, X^{(s)})
\end{aligned}
\tag{16}
$$

Furthermore, we have:

$$
\begin{aligned}
G(X^{(s)}, X^{(s)}) &= \sum_{i=1}^{n} \sum_{k=1}^{K} P(y=k|v=i, X^{(s)}, B_g) \ln \frac{P(v=i, y=k|X, B_g)}{P(y=k|v=i, X^{s}, B_g)} \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} P(y=k|v=i, X^{(s)}, B_g) \ln P(v=i|X^{(s)}, B_g) \\
&= \sum_{i=1}^{n} \ln P(v=i|X^{(s)}, B_g) \sum_{k=1}^{K} P(y=k|v=i, X^{(s)}, B_g) \\
&= \sum_{i=1}^{n} \ln P(v=i|X^{(s)}, B_g) L(N|X^{(s)}, B_g).
\end{aligned}
\tag{17}
$$

Let $P(y=k|b=l, X^{(s)}, B_g) = \gamma_{ik}^{(s)}$, we have:

$$
\begin{aligned}
G(X, X^{(s)}) =& \sum_{l=1}^{L} \sum_{b_{il} \neq 0} \sum_{k=1}^{K} \gamma_{ik}^{(s)} \ln P(v=i, y=k|X, B_g) \\
& - \sum_{l=1}^{L} \sum_{b_{il} \neq 0} \sum_{k=1}^{K} \gamma_{ik}^{(s)} \ln P(y=k|v=i, X^{(s)}, B_g).
\end{aligned}
\tag{18}
$$

Thus, we have:

$$
\begin{aligned}
argmaxG(X, X^{(s)}) = argmax(&\sum_{l=1}^{L}\sum_{b_{il}\neq 0}\sum_{k=1}^{K}\gamma_{ik}^{(s)}\ln P(v=i, y=k|X, B_g) \\
&-\sum_{l=1}^{L}\sum_{b_{il}\neq 0}\sum_{k=1}^{K}\gamma_{ik}^{(s)}\ln P(y=k|v=i, X^{(s)}, B_g)) \\
&= argmax(\sum_{l=1}^{L}\sum_{b_{il}\neq 0}\sum_{k=1}^{K}(\gamma_{ik}^{(s)}\ln P(v=i, y=k|X, B_g))) \\
&= argmaxE[L(G, Z^{(s)}|X, B_g)] \\
&= X^{(s+1)}.
\end{aligned}
\tag{19}
$$

Recall that, the $\Theta^{(s+1)}, \Delta^{(s+1)}$, and $\Omega^{(s+1)}$ of $X^{(s+1)}$ can be computed in terms of $\gamma_{lk}^{(s)}$. Thus, we have:

$$
G(X^{(s+1)}, X^{(s)}) \geq G(X^{(s)}, X^{(s)}) = L(G|X^{(s)}, B_g). \tag{20}
$$

That is to say, the $X^{(s+1)}$ obtained in the current iteration will be not worse than $X^{(s)}$ obtained in last iteration. Thus, we have the theorem and the proof is end.

IV. THE CALCULATION OF EQ.(10)

$$
L(N, Z|X, B_g) = \sum_{l=1}^{L}\sum_{b_{il}\neq 0}\sum_{k=1}^{K} z_{lk}(\sum_{j=1}^{n}(\ln f(\theta_{kj}, a_{ij}) + \ln f(\delta_{kj}, a_{ji})) + \ln \omega_k) \tag{21}
$$

Proof :

$$
\begin{aligned}
&L(N, Z|X, B_g) \\
&= \ln \Pi_{i=1}^{n} P(v = i, y = y(i)) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} m_{ik} P(v = i, y = k) \\
&= \sum_{i=1}^{n} \ln \sum_{k=1}^{K} m_{ik} P(v = i|y = k) P(y = k) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} \ln(P(v = i|y = k) P(y = k))^{m_{ik}} \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} m_{ik} \ln(P(v = i|y = k) P(y = k)) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} m_{ik} \ln(\prod_{j=1}^{n} (\theta_{kj}^{a_{ij}} (1 - \theta_{kj})^{1-a_{ij}} \delta_{kj}^{a_{ji}} (1 - \delta_{kj})^{1-a_{ji}}) \omega_k \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} m_{ik} (\sum_{j=1}^{n} (\ln f(\theta_{kj}, a_{ij}) + \ln f(\delta_{kj}, a_{ji})) + \ln \omega_k) \\
&= \sum_{l=1}^{L} \sum_{b_{il} \neq 0} \sum_{k=1}^{K} z_{lk} (\sum_{j=1}^{n} (\ln f(\theta_{kj}, a_{ij}) + \ln f(\delta_{kj}, a_{ji})) + \ln \omega_k)
\end{aligned}
\tag{22}
$$

The proof is end.

V. THE CALCULATION OF EQ.(10) IN THE MAIN TEXT

Now, we will discuss how to approximately estimate the prior $P(X|B_g)$ based on the information theory. Note that $1 \leq K \leq L = n/g$,which implies, that the coarser the granularity is, the smaller if K. It will be shown in the following that a smaller K will indicate a less complexity of X. Thus, we have the following: a coarser granularity prefers simpler models,which can be mathematically written as

$$P(X|B_g) = \eta(X)^g \tag{23}$$

where the function $\eta(X)$ measures the complexity of X in terms of its parameters. In this paper, we set $\eta(X) = P(X|B_1) = P(X)$, the prior of X under $g = 1$.

According to Shannon and Weaver, $ln(1/P(X))$ is the minimum description length of X with a prior $P(X)$ in its model space. Let $\bar{X}$ denote the optimal coding schema for X, and let $C(\bar{X})$ be the coding length (or complexity) of X under this schema. We have

$$-lnP(X|B_g) = -glnP(X) = gC(\bar{X}) \tag{24}$$

Now, to estimate the prior $P(X|B_g)$ is to design a coding(or compressing) schema as close to $\bar{X}$ as possible.

As discussed before, the main purpose of the model X is to characterize the behavior of network N in terms of node couplings measured together by its five parameters. Next, we will show that all node couplings can be approximately measured by three compressed parameters instead of original five ones. In this way, one hopes to get a more compact coding schema much closer to $\bar{X}$.

First, we can compress four parameters of X,i.e.,Θ, Δ, Z, and Ω, into two

$$\begin{cases} \Phi = \Theta B_1 Z D_{-1} = \Theta Z D_{-1} \\ \Psi = \Delta B_1 Z D_{-1} = \Delta Z D_{-1} \end{cases}$$

where $D = diag(n\Omega)$.

Second, parameter Z can be compressed into a map y, where $y(i) = k$ if the entry (i, k) of $B_1 Z = Z$ is equal to one. Now, given y, Φ, and Ψ, node coupling p_{ij} and q_{ij} can be measured by

$$p_{ij} = \phi_{y(i),y(j)} q_{ij} = \psi_{y(i),y(j)} \tag{25}$$

Equation (25) says that all node coupling can be approximately characterized by y, Φ, and Ψ. Correspondingly, the compressed coding schema of X is

$$\hat{X} = (K, \Phi_{K\times K}, \Psi_{K\times K}, y_{n\times 2}). \tag{26}$$

Now, we compare the complexity of X and $\hat{X}$

$$\begin{aligned} C(\hat{X}) &= 1 \times (-\ln\frac{1}{1}) + 2K^2(-\ln\frac{1}{K^2}) + 2n(-\ln\frac{1}{2n}) \\ &= 2nK\ln K^2 + 2n\ln 2n. \end{aligned} \tag{27}$$

Note that the dimension of Z is $n \times K$ under B_1, so we have

$$\begin{aligned} C(X) =& 1 \times (-\ln\frac{1}{1}) + 3nK(-\ln\frac{1}{nK}) + K(-\ln\frac{1}{K}) \\ &= 2nK\ln nK + nK\ln nK + K\ln K. \end{aligned} \tag{28}$$

It is easy to verify $C(\hat{X}) \ll C(X)$ in the case of $K \ll n$. That is to say that the coding schema closest to $\bar{X}$ that we have found is given by 26. Thus, we have

$$C(\bar{X}) \approx C(\hat{X}) = 2K^2 \ln K^2 + 2n \ln 2n. \tag{29}$$

The proof is end.