1 | PURPOSE

This paper contains the proofs to accompany the theorems presented in the paper "Improving computational performance in likelihood-based network clustering" by Alan Ballard and Marcus Perry.

2 | THEOREMS

Theorem 1. Assign each vertex v $\in V$ from $G = (V, E)$ into one of k clusters. Assume the edges within cluster h $\in \{1, 2..., k\}$ are distributed according to density $f_h(y_h|\theta_h, \mathbf{z}_{(k)})$ and the edges between clusters are distributed according to density $f_b(y_b|\theta_b, \mathbf{z}_{(k)})$. The likelihood of the clustering $\mathbf{z}_{(k)}$ is then written as

$$L(\theta_1, \theta_2, ..., \theta_k, \theta_b|\mathbf{y}, \mathbf{z}_{(k)}) = \{\prod_{h=1}^{k} f_h(y_h|\theta_h, \mathbf{z}_{(k)})\} f_b(y_b|\theta_b, \mathbf{z}_{(k)}), \tag{1}$$

where each θ_h and θ_b are estimated by maximum likelihood estimators $\hat{\theta}_h$ and $\hat{\theta}_b$, respectively.

Then, consider a reclustering by reassigning vertex l^* from cluster ℓ to cluster j. If the parameter estimates for the within-cluster edge densities other than f_ℓ and f_j are unaffected by the reclustering, i.e. $\hat{\theta}^1_h = \hat{\theta}^0_h$, $h \notin \{\ell, j\}$, then the change-in-likelihood due to the reclustering can be written:

$$\Delta L = \frac{f_\ell(y_\ell|\hat{\theta}^1_\ell, \mathbf{z}^1_{(k)}) f_j(y_j|\hat{\theta}^1_j, \mathbf{z}^1_{(k)}) f_b(y_b|\hat{\theta}^1_b, \mathbf{z}^1_{(k)})}{f_\ell(y_\ell|\hat{\theta}^0_\ell, \mathbf{z}^0_{(k)}) f_j(y_j|\hat{\theta}^0_j, \mathbf{z}^0_{(k)}) f_b(y_b|\hat{\theta}^0_b, \mathbf{z}^0_{(k)})}, \tag{2}$$

where superscripts 0 and 1 indicate pre- and post-reclustering states, respectively.

Theorem 2. Assign each vertex $v \in V$ from $G = (V, E)$ into one of k clusters. Then, consider a reclustering by reassigning vertex l^* from cluster ℓ to cluster j. The elements of $\mathbf{O}$ post-reclustering can be written completely in terms of pre-reclustering information. Specifically, the elements are:

1. $o^1_{s,t} = o^0_{s,t}$
2. $o^1_{s,s} = o^0_{s,s}$
3. $o^1_{\ell,s} = o^0_{\ell,s} - Z^0_s$
4. $o^1_{j,s} = o^0_{j,s} + Z^0_s$
5. $o^1_{\ell,j} = o^0_{\ell,j} - Z^0_j + Z^0_\ell$
6. $o^1_{\ell,\ell} = o^0_{\ell,\ell} - Z^0_\ell$
7. $o^1_{j,j} = o^0_{j,j} + Z^0_j$,

where $s, t \notin \{\ell, j\}$, superscripts 0 and 1 indicate pre- and post-reclustering states, respectively, and Z^0_h is the sum of edge weights between vertex l^* and cluster h pre-reclustering, h $\in \{1, 2, ..., k\}$.

Theorem 3. Assign each vertex $v \in V$ from $G = (V, E)$ into one of k clusters. Then, consider a reclustering by reassigning vertex l^* from cluster ℓ to cluster j. The sum of the distinct off-diagonal elements of $\mathbf{O}$ post-reclustering can be written completely in terms of pre-reclustering information. Specifically:

$$o^1_b = \sum o^1_{s,t} = o^0_b + Z^0_\ell - Z^0_j, \tag{3}$$

where $s, t \in \{1, 2, ...k\}$ and $s \neq t$, superscripts 0 and 1 indicate pre- and post-reclustering states, respectively, and Z^0_h is the sum of edge weights between vertex l^* and cluster h pre-reclustering, h $\in \{1, 2, ..., k\}$.

Theorem 4. Assign each vertex $v \in V$ from $G = (V, E)$ into one of k clusters. Then, consider a reclustering by reassigning vertex l^* from cluster ℓ to cluster j. The elements of $\mathbf{P}$ post-reclustering can be written completely in terms of pre-reclustering information. Specifically, the elements are:

1. $p^1_{s,t} = p^0_{s,t}$
2. $p^1_{s,s} = p^0_{s,s}$
3. $p^1_{\ell,s} = n^0_s(n^0_\ell - 1)$
4. $p^1_{j,s} = n^0_s(n^0_j + 1)$
5. $p^1_{\ell,j} = (n^0_\ell - 1)(n^0_j + 1)$
6. $p^1_{\ell,\ell} = \frac{(n^0_\ell - 1)(n^0_\ell - 2)}{2}$

7. $p^1_{j,j} = \frac{(n^0_j+1)n^0_j}{2}$,

where $s, t \notin \{\ell, j\}$, superscripts 0 and 1 indicate pre- and post-reclustering states, respectively, and n^0_h is the number of vertices in cluster h pre-reclustering.

Theorem 5. Assign each vertex $v \in V$ from $G = (V, E)$ into one of k clusters. Then, consider a reclustering by reassigning vertex l^* from cluster ℓ to cluster j. The sum of the distinct off-diagonal elements of $\mathbf{P}$ post-reclustering can be written completely in terms of pre-reclustering information. Specifically:

$$p^1_b = \sum p^1_{s,t} = p^0_b + n^0_\ell - n^0_j - 1, \tag{4}$$

where $s, t \in \{1, 2, ...k\}$ and $s \neq t$, superscripts 0 and 1 indicate pre- and post-reclustering states, respectively, and n^0_h is the number of vertices in cluster h pre-reclustering, h $\in \{1, 2, ..., k\}$.

3 | PROOFS

Proof. Theorem 1: The change-in-likelihood can be written in an abbreviated form given some restrictions on the parameters. Let $\hat\theta^1_s = \hat\theta^0_s, s \neq \{\ell, j\}$. Then, $f_s(y_s|\hat\theta^1_s, \mathbf{z}^1_{(k)}) = f_s(y_s|\hat\theta^0_s, \mathbf{z}^0_{(k)})$ Finally, for space-conservation purposes, let $\Lambda^0_t = y_t|\hat\theta^0_t, \mathbf{z}^0_{(k)}$ and $\Lambda^1_t = y_t|\hat\theta^1_t, \mathbf{z}^1_{(k)}, t \in \{1, 2, ...k, b\}$. Then, $f_s(\Lambda^1_s) = f_s(\Lambda^0_s), s \neq \{\ell, j\}$ and,

$$\Delta L = \frac{\{\prod_{s\neq\{\ell,j\}} f_s(\Lambda^1_s)\} f_\ell(\Lambda^1_\ell) f_j(\Lambda^1_j) f_b(\Lambda^1_b)}{\{\prod_{s\neq\{\ell,j\}} f_s(\Lambda^0_s)\} f_\ell(\Lambda^0_\ell) f_j(\Lambda^0_j) f_b(\Lambda^0_b)} = \frac{\{\prod_{s\neq\{\ell,j\}} f_s(\Lambda^0_s)\} f_\ell(\Lambda^1_\ell) f_j(\Lambda^1_j) f_b(\Lambda^1_b)}{\{\prod_{s\neq\{\ell,j\}} f_s(\Lambda^0_s)\} f_\ell(\Lambda^0_\ell) f_j(\Lambda^0_j) f_b(\Lambda^0_b)} = \frac{f_\ell(\Lambda^1_\ell) f_j(\Lambda^1_j) f_b(\Lambda^1_b)}{f_\ell(\Lambda^0_\ell) f_j(\Lambda^0_j) f_b(\Lambda^0_b)} \quad \square$$

$\square$

Proof. Theorem 2: Post-reclustering edge weight sums can be written in terms of pre-reclustering edge weight sums.
Let $\mathbf{1_i}$ be a N x 1 vector of 0's and a single 1 in the i^{th} position. Then, if vertex l^* is removed from cluster ℓ and placed in cluster j, $\boldsymbol{\omega}^1_\ell = \boldsymbol{\omega}^0_\ell - \mathbf{1}_{l^*}$ and $\boldsymbol{\omega}^1_j = \boldsymbol{\omega}^0_j + \mathbf{1}_{l^*}$, where superscripts 0 and 1 indicate pre- and post-reclustering states, respectively. Then,

1. $$\begin{aligned} o^1_{s,t}, s, t \notin \{\ell, j\} &= {\boldsymbol{\omega}^1_s}^T \mathbf{A} \boldsymbol{\omega}^1_t \\ &= {\boldsymbol{\omega}^0_s}^T \mathbf{A} \boldsymbol{\omega}^0_t \\ &= obs^0_{s,t} \quad \square \end{aligned}$$

2. $$\begin{aligned} o^1_{s,s}, s \notin \{\ell, j\} &= \frac{1}{2}{\boldsymbol{\omega}^1_s}^T \mathbf{A} \boldsymbol{\omega}^1_s \\ &= \frac{1}{2}{\boldsymbol{\omega}^0_s}^T \mathbf{A} \boldsymbol{\omega}^0_s \\ &= obs^0_{s,s} \quad \square \end{aligned}$$

3. $$\begin{aligned} o^1_{\ell,s} - obs^0_{\ell,s}, s \notin \{\ell, j\} &= {\boldsymbol{\omega}^1_\ell}^T \mathbf{A} {\boldsymbol{\omega}_s}^1 - {\boldsymbol{\omega}^0_\ell}^T \mathbf{A} \boldsymbol{\omega}^0_s \\ &= {\boldsymbol{\omega}^1_\ell}^T \mathbf{A} {\boldsymbol{\omega}_s}^0 - {\boldsymbol{\omega}^0_\ell}^T \mathbf{A} {\boldsymbol{\omega}_s}^0 \\ &= -[\boldsymbol{\omega}^0_\ell - \boldsymbol{\omega}^1_\ell]^T \mathbf{A} {\boldsymbol{\omega}_s}^0 \\ &= -[\mathbf{1}_{l^*}]^T \mathbf{A} \boldsymbol{\omega}^0_s \\ &= -Z^0_s \quad \square \end{aligned}$$

4. $$\begin{aligned} o^1_{j,s} - o^0_{j,s}, s \notin \{\ell, j\} &= {\boldsymbol{\omega}^1_j}^T \mathbf{A} \boldsymbol{\omega}^1_s - {\boldsymbol{\omega}^0_j}^T \mathbf{A} \boldsymbol{\omega}^0_s \\ &= {\boldsymbol{\omega}^1_j}^T \mathbf{A} \boldsymbol{\omega}^0_s - {\boldsymbol{\omega}^0_j}^T \mathbf{A} \boldsymbol{\omega}^0_s \\ &= [\boldsymbol{\omega}^1_j - \boldsymbol{\omega}^0_j]^T \mathbf{A} {\boldsymbol{\omega}_s}^0 \\ &= [\mathbf{1}_{l^*}]^T \mathbf{A} \boldsymbol{\omega}^0_s \\ &= Z^0_s \quad \square \end{aligned}$$

5. $$\begin{aligned} o^1_{\ell,j} &= {\boldsymbol{\omega}^1_\ell}^T \mathbf{A} \boldsymbol{\omega}^1_j \\ &= [\boldsymbol{\omega}^0_\ell - \mathbf{1}_{l^*}]^T \mathbf{A} [\boldsymbol{\omega}^0_j + \mathbf{1}_{l^*}] \\ &= {\boldsymbol{\omega}^0_\ell}^T \mathbf{A} {\boldsymbol{\omega}_j}^0 - [\mathbf{1}_{l^*}]^T \mathbf{A} \boldsymbol{\omega}^0_j + {\boldsymbol{\omega}^0_\ell}^T \mathbf{A} \mathbf{1}_{l^*} \\ &= o^0_{\ell,j} - Z^0_j + Z^0_\ell \quad \square \end{aligned}$$

6. $$\begin{aligned} o^1_{\ell,\ell} &= \frac{1}{2}{\boldsymbol{\omega}^1_\ell}^T \mathbf{A} \boldsymbol{\omega}^1_\ell \\ &= \frac{1}{2}[\boldsymbol{\omega}^0_\ell - \mathbf{1}_{l^*}]^T \mathbf{A} [\boldsymbol{\omega}^0_\ell - \mathbf{1}_{l^*}] \\ &= \frac{1}{2}[{\boldsymbol{\omega}^0_\ell}^T \mathbf{A} \boldsymbol{\omega}^0_\ell - [\mathbf{1}_{l^*}]^T \mathbf{A} \boldsymbol{\omega}^0_\ell - [\mathbf{1}_{l^*}]^T \mathbf{A} \boldsymbol{\omega}^0_\ell] \\ &= \frac{1}{2}[2o^0_{\ell,\ell} - 2[\mathbf{1}_{l^*}]^T A \omega^0_\ell] \\ &= o^0_{\ell,\ell} - Z^0_\ell \quad \square \end{aligned}$$

7. $$\begin{aligned} o^1_{j,j} &= \frac{1}{2}{\boldsymbol{\omega}^1_j}^T \mathbf{A} \boldsymbol{\omega}^1_j \\ &= \frac{1}{2}[\boldsymbol{\omega}^0_j + \mathbf{1}_{l^*}]^T \mathbf{A} [\boldsymbol{\omega}^0_j + \mathbf{1}_{l^*}] \\ &= \frac{1}{2}[{\boldsymbol{\omega}^0_j}^T \mathbf{A} \boldsymbol{\omega}^0_j + [\mathbf{1}_{l^*}]^T \mathbf{A} \boldsymbol{\omega}^0_j + [\mathbf{1}_{l^*}]^T \mathbf{A} \boldsymbol{\omega}^0_j] \\ &= \frac{1}{2}[2o^0_{j,j} + 2[\mathbf{1}_{l^*}]^T \mathbf{A} \boldsymbol{\omega}^0_j] \\ &= o^0_{j,j} + Z^0_j \quad \square \end{aligned}$$

$\square$

Proof. Theorem 3: Post-reclustering between-cluster edge weight sums can be written in terms of pre-reclustering edge weight sums.

$$
\begin{aligned}
o_b^1 &= \sum o_{s,t}^1 + \sum o_{\ell,t}^1 + \sum o_{j,t}^1 + o_{\ell,j}^1 \\
&= \sum o_{s,t}^0 + \sum (o_{\ell,t}^0 - Z_t^0) + \sum (o_{j,t}^0 + Z_t^0) + (o_{\ell,j}^0 + Z_\ell^0 - Z_j^0), \text{ by Theorem 2} \\
&= \sum o_{s,t}^0 + \sum o_{\ell,t}^0 + \sum o_{j,t}^0 + (o_{\ell,j}^0 + Z_\ell^0 - Z_j^0) \\
&= (\sum o_{s,t}^0 + \sum o_{\ell,t}^0 + \sum o_{j,t}^0 + o_{\ell,j}^0) + Z_\ell^0 - Z_j^0 \\
&= o_b^0 + Z_\ell^0 - Z_j^0, s,t \notin \{\ell, j\}, s \neq t \quad \square
\end{aligned}
$$

$\square$

Proof. Theorem 4: Post-reclustering maximum edge counts can be written in terms of pre-reclustering maximum edge counts. For a given clustering $\mathbf{z}_{(k)}$ on an undirected network, $\mathbf{P}$ is a symmetric k x k matrix with element $p_{s,t}$ containing the maximum number of distinct edges possible between the vertices of clusters s and t. This value is $p_{s,t} = \binom{n_s+n_t}{2} - \binom{n_s}{2} - \binom{n_t}{2} = n_s n_t$ when $s \neq t$ and $p_{s,t} = \frac{n_s(n_s-1)}{2}$ when $s = t$, where n_h is the number of distinct vertices in cluster $h \in \{1, 2, ..., k\}$. Then,

1. $p_{s,t}^1, s,t \notin \{\ell, j\} = n_s^0 n_t^0 \quad = p_{s,t}^0 \quad \square$
2. $p_{s,s}^1, s \notin \{\ell, j\} = \dfrac{n_s^0(n_s^0 - 1)}{2} \quad = p_{s,s}^0 \quad \square$
3. $p_{\ell,s}^1, s \notin \{\ell, j\} = n_s^0(n_\ell^0 - 1) \quad \square$
4. $p_{j,s}^1, s \notin \{\ell, j\} = n_s^0(n_j^0 + 1) \quad \square$
5. $p_{\ell,j}^1 = (n_\ell^0 - 1)(n_j^0 + 1) \quad \square$
6. $p_{\ell,\ell}^1 = \dfrac{(n_\ell^0 - 1)((n_\ell^0 - 1) - 1)}{2} \quad = \dfrac{(n_\ell^0 - 1)(n_\ell^0 - 2)}{2} \quad \square$
7. $p_{j,j}^1 = \dfrac{(n_j^0 + 1)((n_j^0 + 1) - 1)}{2} \quad = \dfrac{(n_j^0 + 1)n_j^0}{2} \quad \square$

$\square$

Proof. Theorem 5: Post-reclustering between-cluster maximum edge counts can be written in terms of pre-reclustering maximum edge counts.

$$
\begin{aligned}
p_b^1 &= \sum p_{s,t}^1 + \sum p_{\ell,t}^1 + \sum p_{j,t}^1 + \sum p_{\ell,j}^1 \\
&= \sum p_{s,t}^0 + \sum n_t^0(n_\ell^0 - 1) + \sum n_t^0(n_j^0 + 1) + (n_\ell^0 - 1)(n_j^0 + 1), \text{ by Theorem 4} \\
&= \sum p_{s,t}^0 + \sum n_t^0(n_\ell^0) + \sum n_t^0(n_j^0) + ((n_\ell^0)(n_j^0) + n_\ell^0 - n_j^0 - 1) \\
&= (\sum p_{s,t}^0 + \sum p_{\ell,t}^0 + \sum p_{j,t}^0 + p_{\ell,j}^0) + n_\ell^0 - n_j^0 - 1 \\
&= p_b^0 + n_\ell^0 - n_j^0 - 1, s,t \notin \{\ell, j\}, s \neq t \quad \square
\end{aligned}
$$

$\square$

$\square$