Solutions for Statistical analysis of network data – Sheet 4

1. We write F as the path on three nodes and two edges. Then

$$\begin{aligned}\mathrm{E}X_F(G) &= \sum_{F'\subset K_n} \mathrm{I}\left(F'\subset G\right)\mathrm{I}\left(F'\equiv F\right)\\ &= \Pr\{F\subset G\}\sum_{F'\subset K_n} \mathrm{I}\left(F'\equiv F\right)\\ &= \Pr\{F\subset G\}(n)_3/2,\end{aligned}$$

and also note that

$$\begin{aligned}\Pr\{F\subset G\} &= \mathrm{E}\prod_{ij\in e(F)} A_{ij}\\ &= \mathrm{Eg}\, g\left(\xi_1\right) g\left(\xi_2\right) g\left(\xi_2\right) g\left(\xi_3\right) = \|g\|_2^2.\end{aligned}$$

Therefore it follows that

$$\mathrm{E}\,X_F(G) = \|g\|_2^2(n)_3/2.$$

2. We note that

$$d_i = \sum_{j\neq i} A_{ij}.$$

Note that

$$\begin{aligned}\mathrm{E}d_i &= \sum_{j\neq i}\mathrm{E}\left\{A_{ij}\right\}\\ &= \sum_{j\neq i}\mathrm{E}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\xi}_j\right\}\\ &= (n-1)\boldsymbol{\mu}^T\boldsymbol{\mu}.\end{aligned}$$

Also note that

$$\begin{aligned}\mathrm{E}\left(d_i^2\right) &= \sum_{l\neq i}\sum_{k\neq i}\mathrm{E}A_{il}A_{ik}\\ &= \sum_{l\neq i}\sum_{k\neq i}\left(\mathrm{E}A_{il}\mathrm{E}A_{ik}\left(1-\delta_{lk}\right)+\mathrm{E}A_{il}\delta_{lk}\right).\end{aligned}$$

Then note that

$$\begin{aligned}\mathrm{Var}\left\{d_i\right\} &= \mathrm{E}\left(d_i^2\right)-\mathrm{E}^2\left(d_i\right)\\ &= \sum_{l\neq i}\sum_{k\neq i}\left(\mathrm{E}A_{il}\mathrm{E}A_{ik}\left(1-\delta_{lk}\right)+\mathrm{E}A_{il}\delta_{lk}\right)-\sum_{j\neq i}\mathrm{E}\left\{A_{ij}\right\}\sum_{l\neq i}\mathrm{E}\left\{A_{il}\right\}\\ &= \sum_{l\neq i}\mathrm{E}A_{il}-\sum_{l\neq i}\mathrm{E}^2A_{il} = \sum_{l\neq i}\mathrm{E}A_{il}\left\{1-\mathrm{E}A_{il}\right\},\end{aligned}$$

which can be reformulated in terms of $\boldsymbol{\mu}$.

3. Going back to $X_F(G)$ we still have

$$\mathrm{E}X_F(G) = \Pr\{F\subset G\}(n)_3/2.$$

In the sparse model instead

$$\begin{aligned}\Pr\{F\subset G\} &= \mathrm{E}\prod_{ij\in c(F)} A_{ij}\\ &= \rho_n^2\,\mathrm{E}\,g\left(\xi_1\right) g\left(\xi_2\right) g\left(\xi_2\right) g\left(\xi_3\right) = \rho_n^2\|g\|_2^2.\end{aligned}$$

If we go back to the definition of

$$\hat{t}_F(G) = \frac{X_F(G)}{X_F(K_n)}.$$

This is a straightforward method of moments estimator of $t_F(G) = \Pr(F \subset G)$. However with the introduction of ρ we expect all of these to converge to zero. Thus we take

$$\hat{s}_F(G) = \frac{X_F(G)}{\hat{\rho}^{|e(F)|} X_F(K_n)}.$$

This allows us to account for sparse graphs.

4. We noted the distribution of $\hat{\rho}$ for the Erdos-Renyi model. Construct a confidence interval for ρ based on this distribution. We have $n(\widehat{\rho} - \rho) \xrightarrow{L} (0, \sigma^2)$, for some specified σ^2. Thus we have

$$\Pr\left(-1.95 \leq \frac{n(\widehat{\rho} - \rho)}{\sigma} \leq 1.96\right) = 0.95,$$

and so the confidence interval becomes $\hat{\rho} \pm 1.96\frac{\sigma}{n}$. To determine σ^2 we note that

$$\begin{aligned}
\operatorname{Var}\{n(\hat{\rho} - \rho)\} &= \sigma^2 \\
n^2 \operatorname{Var}\{\hat{\rho}\} &= \sigma^2 \\
n^2 \operatorname{Var}\left\{\frac{\sum_{i<j} A_{il}}{\binom{n}{2}}\right\} &= n^2 \frac{1}{\binom{n}{2}^2} \sum_{i<j} \rho(1-\rho) \\
&= \frac{4}{(n-1)^2} \frac{n(n-1)}{2} \rho(1-\rho) \\
&= \frac{2}{\left(1 - \frac{1}{n}\right)} \rho(1-\rho) \to 2\rho(1-\rho).
\end{aligned}$$

To calculate the CI we substitute in the estimate for ρ.

5. Let us solve a slightly more general problem. Assume that $\rho = n^{-\alpha}$ for $0 < \alpha < 1$. We note that

$$\begin{aligned}
\mathrm{E}(\hat{\rho}) &= \rho \\
\operatorname{Var}(\hat{\rho}) &= \frac{\rho(1-\rho)}{\binom{n}{2}}
\end{aligned}$$

Now define the estimate $\widehat{\rho}_0 = 0$. This has

$$\begin{aligned}
\mathrm{E}\left(\widehat{\rho}_0\right) &= 0 \\
\operatorname{Var}\left(\widehat{\rho}_0\right) &= 0.
\end{aligned}$$

Then we may note that

$$\begin{aligned}
\operatorname{MSE}(\widehat{\rho}) &= \frac{\rho(1-\rho)}{\binom{n}{2}} \\
\operatorname{MSE}\left(\widehat{\rho}_0\right) &= \rho^2.
\end{aligned}$$

We note that

$$\begin{aligned}
\rho^2 &< \tfrac{\rho(1-\rho)}{\binom{n}{2}} \\
\rho &< \tfrac{(1-\rho)}{\binom{n}{2}}.
\end{aligned}$$

Because of our constraint on α, that is never true.

6. Now assume that

$$A_{ij} \mid \pi \sim \text{Bernoulli}\left(\pi_j \pi_j\right).$$

Determine the mean, variance and covariance of d_i and d_j if we draw realizations from that model.

This problem has partially already been solved, but is retained here to be in contrast with the first question on this sheet. We note that

$$\mathrm{E} d_i = \sum_{j \neq i} \pi_i \pi_j = \pi_i \sum \pi_j = \pi_i \|\pi\|_1.$$

Also

$$\begin{aligned} \mathrm{E}\left(d_i^2\right) &= \sum_{l \neq i} \sum_{k \neq i} \mathrm{E} A_{il} A_{ik} \\ &= \sum_{l \neq i} \sum_{k \neq i} \left(\mathrm{E} A_{il} \mathrm{E} A_{ik}\left(1-\delta_{lk}\right) + \mathrm{E} A_{il} \delta_{lk}\right). \end{aligned}$$

Then note that

$$\begin{aligned} \operatorname{Var}\left\{d_i\right\} &= \mathrm{E}\left(d_i^2\right) - \mathrm{E}^2\left(d_i\right) \\ &= \sum_{l \neq i} \sum_{k \neq i} \left(\mathrm{E} A_{il} \mathrm{E} A_{ik}\left(1-\delta_{lk}\right) + \mathrm{E} A_{il} \delta_{lk}\right) - \sum_{j \neq i} \mathrm{E}\left\{A_{ij}\right\} \sum_{l \neq i} \mathrm{E}\left\{A_{il}\right\} \\ &= \sum_{l \neq i} \mathrm{E} A_{il} - \sum_{l \neq i} \mathrm{E}^2 A_{il} = \sum_{l \neq i} \mathrm{E} A_{il}\left\{1 - \mathrm{E} A_{il}\right\} \\ &= \sum_{l \neq i} \pi_i \pi_l \left\{1 - \pi_i \pi_l\right\}. \end{aligned}$$

Finally we note that

$$\begin{aligned} \operatorname{Cov}\left\{d_i d_j\right\} &= \mathrm{E}\left(d_i d_j\right) - \mathrm{E}\left(d_i\right) \mathrm{E}\left(d_j\right) \\ &= \sum_{l \neq i} \sum_{k \neq j} \left(\mathrm{E} A_{il} \mathrm{E} A_{jk}\left(1-\delta_{lk}\right) + \mathrm{E} A_{il} \mathrm{E} A_{jk} \delta_{lk}\right) - \sum_{j \neq i} \mathrm{E}\left\{A_{ij}\right\} \sum_{l \neq i} \mathrm{E}\left\{A_{il}\right\}. \end{aligned}$$

The calculations complete like before.