1. Recall that selecting a node corresponds to a substitution being fixed at a gene. If difference nodes have different probabilities of being selected, it means that different genes have different mutation and fixation rates.

2(a) We have $K = 20, p = 0.01$, and $\alpha = 0.0087$ which gives speciation probability

$$E[S] \approx 1 - (1 - 0.01)^{0.0087 \cdot \binom{20}{2}} = 0.0165$$

Consider the yeast network with $K = 20$ and $p = 0.01$.

(b) Doubling $K$ to $K = 40$ gives

$$E[S] \approx 1 - (1 - 0.01)^{0.0087 \cdot \binom{40}{2}} = 0.0659$$

which is a 300% increase.

(c) To double the speciation probability to 0.033, trial and error shows that $K$ must be about 28 which is a 40% increase.

(d) Here $\alpha = 1$ which gives speciation probability

$$S = 1 - (1 - 0.01)^{\binom{20}{2}} = 0.85$$

(e) We must have

$$\alpha \binom{K}{2} = \binom{20}{2}$$

where $\alpha = 0.0087$ which gives $K \approx 209$, that is, 10 times as many substitutions are required.

3. We have $\mu = E[X] = 1$ and

$$g(x) = 1 - (1 - p)^x = 1 - e^{x \cdot \ln(1-p)}$$

which gives

$$g''(x) = -(\ln(1 - p))^2 e^{x \cdot \ln(1-p)} = -(\ln(1 - p))^2 (1 - p)^x$$
and with \( p = 0.1 \) and \( \mu = 1 \):

\[
g''(\mu) = -0.01
\]

As the variances are 1, 0.8, and 0.4, respectively, the second-order terms \( g''(\mu) \sigma^2 \) are -0.005, -0.004, and -0.002, respectively.

4(a) We have \( N_E = 4 \) and \( N = 5 \) which gives

\[
\alpha = \frac{N_E}{N} = \frac{4}{5} = 0.4
\]

(b) The expected value is

\[
E[X] = \alpha \binom{K}{2} = 0.4 \binom{3}{2} = 1.2
\]

and the variance equals

\[
\text{Var}[X] = N_E P_2 (1 - P_2) + 2 (N_S (P_3 - P_2^2) + N_D (P_4 - P_2^2))
\]

The general formulas for \( N_S \) and \( N_D \) are \( N_S = \sum_{j=2}^{N-1} n_j \binom{j}{2} \) where \( n_j \) is the number of nodes of degree \( j \), and \( N_D = \binom{N_E}{2} - N_S \). In this problem, direct counting is easy and gives \( N_S = 5 \) and \( N_D = 1 \). Further, \( P_i \) is the probability to get \( i \) specific nodes when \( K \) nodes are chosen at random and equals

\[
P_i = \frac{\binom{K}{i}}{\binom{N}{i}} = \frac{\binom{3}{i}}{\binom{5}{i}}
\]

for \( i = 2, 3, \) and 4. Obviously \( P_4 = 0 \) (can’t get 4 nodes if we only choose 3) and

\[
P_2 = \frac{\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10} = 0.3
\]

and

\[
P_3 = \frac{\binom{3}{3}}{\binom{5}{3}} = \frac{1}{10} = 0.1
\]
and we get

$$\text{Var}[X] = 4 \cdot 0.3 \cdot 0.7 + 2(5(0.1 - 0.3^2) + 1(0 - 0.3^2)) = 0.76$$

(c) With $\alpha = 0.4$ and $K = 3$, we get

$$P(\text{speciation}) \approx 1 - (1 - p)^{\alpha(\frac{K}{2})} = 1 - 0.9^{1.2} = 0.1188$$

5. We have $p = 0.5$ and estimate the speciation probability to be $14/30 \approx 0.47$. Over 100 generations we expect $K = 5$ substitutions and estimate $\alpha$ by solving

$$1 - (1 - p)^{\alpha(\frac{5}{2})} = 1 - 0.5^{10\alpha} = \frac{14}{30}$$

which gives $\alpha = 0.09$.

6(a) Note that each choice of 3 nodes corresponds to precisely 3 edge pairs that share a node, and there are thus 3 times as many such edge pairs as there are node triplets. As there are $\binom{N}{3}$ ways to choose 3 nodes we get

$$N_S = 3\binom{N}{3}$$

Alternatively, use the formula

$$N_S = \sum_{j=2}^{N-1} n_j \binom{j}{2}$$

where $n_j$ is the number of nodes of degree $j$. Since all $N$ nodes have degree $d_j = N - 1$ we get

$$N_S = N \binom{N-1}{2} = \frac{N(N-1)(N-2)}{2}$$

which also equals $3\binom{N}{3}$.

For $N_D$, in a similar way note that each choice of 4 nodes corresponds to precisely 3 edge pairs that do not share a node and hence

$$N_D = 3\binom{N}{4}$$
Alternatively, note that $N_D = \binom{N_E}{2} - N_S$ where $N_E = \binom{N}{2}$.

(b) No edge pairs share a node so $N_S = 0$ and $N_D = \binom{N/2}{2} = \frac{N(N-2)}{8}$.

(c) Here $N = 12$ and $N_E = 21$ which gives

$$N_S = \sum_{j=2}^{N-1} n_j \binom{j}{2} = 2 \binom{6}{2} + 10 \binom{3}{2} = 60$$

and

$$N_D = \binom{N_E}{2} - N_S = \binom{21}{2} - 60 = 150$$