Solution-1:

Since the graph $G$ is unweighted and undirected, then the BFS tree is also the shortest path from the starting node of the BFS. Thus, we can use $BFS(u)$ to get the shortest distance from the source node $u$ to each node in $G$. Let's denote the set of vertices of $G$ as $V$, and the set of edges as $E$, and $adj(u)$ denote all the vertices that are directly connected to $u$.

For the sake of notation lets denote the shortest distance between any two nodes $i$ and $j$ as $\delta(i,j)$. And the shortest path between $i$ and $j$ as $P_{ij}$. So $BFS(u)$ returns $\delta(u,v)\ \forall v \in V$. Let's denote the output of $BFS(u)$ as $\delta_u$; $\delta_u[i] = \delta(u,i)$.

Now, suppose we randomly pick a node $u$ and do the $BFS(u)$. This gives us the distance to each node in $V$ from $u$. For any two arbitrary nodes $i$ and $j$, we now have a path passing through $u$. We can say,

$$\delta(i,j) \leq \delta(i,u) + \delta(u,j) \qquad -eq(1)$$

We can repeatedly sample the vertex $u$ and if $u$ lies on $P_{ij}$ then we will have equality in $eq(1)$. Since $\delta(i,j) \leq n-1$ (the maximum path-length can be $n-1$), we have $P\left(u\ lies\ on\ P_{ij}\right) \geq \frac{1}{n-1}$.

Consider the following algorithm that does this:

*Algorithm-1:*

$BFS(u)$:

$\delta_u \leftarrow$ allocate an array of size $n$, initialize with all $\infty$.
$q \leftarrow$ empty queue
$\delta_u[u] = 0$
$q.put(u)$
while not $q.empty()$ :
    $v \leftarrow q.pop()$
    for $x \in adj(v)$:
        if $\delta_u[x] == \infty$:
            $\delta_u[x] = \delta_u[v] + 1$
            $q.put(x)$
Return $\delta_u$

$Distances(G, M)$:

$\Delta \leftarrow$ allocate an $n \times n$ distance-matrix, initialize with all $\infty$
For $i = 0$ to $i = n-1$:
    For $j = 0$ to $j = n-1$:
        If $M[i,j] == \#$: continue
        $\Delta[i,j] = M[i,j]$

Repeat $k = 276 \log n$ times:
    select uniformly-randomly a $u \in V$
    $\delta_u \leftarrow BFS(u)$
    For $i = 0$ to $i = n-1$:
        For $j = 0$ to $j = n-1$:
            $\Delta[i,j] = \min(\delta_u[i] + \delta_u[j], \Delta[i,j])$

Return $\Delta$ // $\Delta[i,j]$ *is the estimated shortest distance from i to j*

Error-Analysis:

An error can occur for any $i, j$ for which $M[i, j] = \#$. An error for $i, j$ occurs if $\Delta[i, j] \neq \delta(i, j)$. For all the $i, j$ for which $M[i, j] \neq \#$, we already have the correct value in $\Delta$. Meaning that $\Delta[i, j]$ is least $\forall i, j\ s.t\ M[i, j] \neq \#$. Thus, for these, $\delta$ retains the true values.

Define the indicator random variable $E_{ij} = \begin{cases} 1, & \Delta[i, j] \neq \delta(i, j) \\ 0, & \Delta[i, j] = \delta(i, j) \end{cases}$

Thus, an error occurs of any of $E_{ij}$ is 1:

$$P_{error} = P\left(\bigcup_{i,j} P\left(E_{ij} = 1\right)\right)$$

$$\Rightarrow P_{error} \leq \sum_{i,j} P\left(E_{ij} = 1\right) = n^2 \cdot P\left(E_{ij} = 1\right)$$

Now, to find $P\left(E_{ij} = 1\right)$:

At every reptation we choose a vertex $u$ uniformly randomly from $v$, then $\delta(u, i) + \delta(u, j) = \delta(i, j)$ occurs when $u$ lies on the shortest path from $i$ to $j$. But we also know that if $\delta(i, j) \leq \frac{n}{100}$, we already are given the exact value in $M[i, j]$. We can make an error for only those $i, j$ for which $\delta(i, j) > \frac{n}{100}$. Therefore, the minimum number of nodes along $P_{ij}$ for which $M[i, j] = \#$ has to be $n/100$.

So, the probability that $\delta_u[i] + \delta_u[j] = \delta(i, j)$ is:

$$P\left(\delta_u[i] + \delta_u[j] = \delta(i, j)\right) = P\left(u \text{ lies on } P_{ij}\right) \geq \frac{\frac{n}{100}}{n} = \frac{1}{100}$$

$$\Rightarrow P\left(\delta_u[i] + \delta_u[j] \neq \delta(i, j)\right) \leq 1 - \frac{1}{100} = \frac{99}{100}$$

This is the probability of making an error at any iteration.

Now, since there is a repetition of $k$ times, $E_{ij} = 1$ only if for each of the $k$ iterations $\delta_u[i] + \delta_u[j] \neq \delta(i, j)$. So,

$$P\left(E_{ij} = 1\right) \leq \left(\frac{99}{100}\right)^k$$

Thus, we now have:

$$P_{error} \leq n^2 P\left(E_{ij} = 1\right) = n^2 (0.99)^k$$

We require that: $P(\Delta[\mathrm{i}, \mathrm{j}] = \delta(i, j) \forall i, j) = 1 - P_{error} \geq 1 - \frac{1}{n^2} \Rightarrow P_{error} < \frac{1}{n^2}$.

$$\Rightarrow n^2 \cdot 0.99^k < \frac{1}{n^2}$$

$$\Rightarrow k > \log_{0.99} n^{-4} = 4\log_{100/99} n = \frac{4}{\log_2 100 - \log_2 99}\log_2 n$$

We choose $k = 276\log_2 n$ since $276 > \frac{4}{\log_2 100 - \log_2 99}$

Therefore, for this $k$, we have $P_{error} < 1/n^2$. And the algorithm outputs correct with probability exceeding $1 - 1/n^2$.

Time-Complexity:

Initializing the matrix $\Delta$ takes $O(n^2)$ time. There is a repeat for $k = 276\log_2 n = O(\log_2 n)$ times. And for each of the repeats, $BFS(u)$ takes $O(n + |E|)$. Updating the matrix $\Delta$ takes $O(n^2)$ time which exceeds $O(n + |E|)$ since $|E| < n^2$. So the time complexity of loop at line 7 is $O(n^2 \log n)$ which is greater that the time complexity of initializing matrix $\Delta$.

Thus, the overall time complexity is $O(n^2 \log n)$

Solution-2:

We are concerned with the triangles in $G\left(n, \frac{1}{2}\right)$. Currently, we will do the analysis for $G(n,p)$ then substitute $p = 1/2$.

Let us denote the set of vertices of $G(n,p)$ as $V$.

Define,

$$X = \#\, triangles\ in\ G(n,p)$$

$$S = \{\{i,j,k\} \mid i,j,k \in V\ \&\ i \neq j \neq k\}$$

Let's define an indicator random variable $X_{ijk} = \begin{cases} 1, & if\ vertices\ i,j,k\ form\ a\ triangle \\ 0, & otherwise \end{cases}$

Thus,

$$X = \sum_{\{i,j,k\} \in S} X_{ijk}$$

Part-1: We are interested in the variance of $X$. We attempt compute $E[X]$ and $E[X^2]$.

For a $G(n,p)$ graph (in our case $p = \frac{1}{2}$) each edge occurs with probability $p$. Thus, the probability that vertices $i,j$ &$k$ form a triangle is that each of $i,j,k$ have an edge between them, which is $p^3$.

Therefore,

$$P(X_{ijk} = 1) = p^3 = 1/8$$

By linearity of expectation,

$$E[X] = \sum_{\{i,j,k\} \in S} E[X_{ijk}] = \binom{n}{3} p^3$$

Now, we need to compute $E[X^2]$.

$$X^2 = \sum_{\{i,j,k\} \in S} \sum_{\{i',j',k'\} \in S} X_{ijk} X_{i'j'k'}$$

Let $T = \{(\{i,j,k\},\{i',j',k'\}) \mid \{i,j,k\},\{i',j',k'\} \in S\}$

We can partition the set $T$ into:

$$T_1 = \{(\{i,j,k\},\{i',j',k'\}) \mid \{i,j,k\},\{i',j',k'\} \in S\ and\ |\{i,j,k\} \cap \{i',j',k'\}| = 0\}$$

$$T_2 = \{(\{i,j,k\},\{i',j',k'\}) \mid \{i,j,k\},\{i',j',k'\} \in S\ and\ |\{i,j,k\} \cap \{i',j',k'\}| = 1\}$$

$$T_3 = \{(\{i,j,k\},\{i',j',k'\}) \mid \{i,j,k\},\{i',j',k'\} \in S\ and\ |\{i,j,k\} \cap \{i',j',k'\}| = 2\}$$

$$T_4 = \{(\{i,j,k\},\{i',j',k'\}) \mid \{i,j,k\},\{i',j',k'\} \in S\ and\ |\{i,j,k\} \cap \{i',j',k'\}| = 3\}$$

We have $T = T_1 \cup T_2 \cup T_3 \cup T_4$

We can rewrite $X^2$ as:

$$X^2 = \sum_{(\{i,j,k\},\{i',j',k'\})\in T} X_{ijk}X_{i'j'k'}$$
$$= \sum_{T_l\in\{T_!,T_2,T_3,T_4\}} \sum_{(\{i,j,k\},\{i',j',k'\})\in T_l} X_{ijk}X_{i'j'k'}$$

We can write $E[X^2]$ as:

$$E[X^2] = \sum_{T_l\in\{T_!,T_2,T_3,T_4\}} \sum_{(\{i,j,k\},\{i',j',k'\})\in T_l} E\left[X_{ijk}X_{i'j'k'}\right]$$

There are the following cases of interactions of the triangles formed by the sets of vertices $\{i, j, k\}$ and $\{i', j', k'\}$:

a) Both the sets are disjoint (the triangles have no vertices in common) denote as $T_1$:
   Select 6 distinct vertices, among them form two triangles by selecting 3 vertices: $= |T_1| = \binom{n}{6}\cdot\binom{6}{3} = 20\binom{n}{6}$
   There should be 3 edges in each $\{i, j, k\}$ and $\{i', j', k'\}$, $p^6$ is the probability that all 6 edges exist.
   $$\Rightarrow E_{\{\{i,j,k\},\{i',j',k'\}\}\in T_1}\left[X_{ijk}X_{i'j'k'}\right] = p^6$$

b) WLOG, $i = i'$. One vertex in common, denoted as $T_2$:
   Select 5 distinct vertices, among these make $1^{st}$ triangle by selecting 3. Then among the vertices in $1^{st}$ triangle select one vertex to be the common vertex with $2^{nd}$ triangle. $= |T_2| = \binom{n}{5}\binom{5}{3}\binom{3}{1} = 30\binom{n}{5}$ ways.
   There should be 6 edges between the appropriate vertices, $p^6$ probability that all 6 edges exist.
   $$\Rightarrow E_{\{\{i,j,k\},\{i',j',k'\}\}\in T_2}\left[X_{ijk}X_{i'j'k'}\right] = p^6$$

c) WLOG, $i = i'\,\&\, j = j'$. Two vertices in common, denoted as $T_3$:
   Select 4 distinct vertices in $\binom{n}{4}$ ways, among these select 3 to make the 1st triangle and among the vertices in the 1st triangle select two to serve as the common edge. Total ways $= |T_3| = \binom{n}{4}\binom{4}{3}\binom{3}{2} = 12\binom{n}{4}$.
   And 5 edges must exist, so $p^5$ probability that these edges exist.
   $$\Rightarrow E_{\{\{i,j,k\},\{i',j',k'\}\}\in T_3}\left[X_{ijk}X_{i'j'k'}\right] = p^5$$

d) All vertices are the same, denoted as $T_4$:
   There can only be $\binom{n}{3}$ such pairs of $\{i, j, k\}$ and $\{i', j', k'\}$. And since there are only 3 edges, probability is $p^3$ that the 3 edges exist.
   $$\Rightarrow E_{\{\{i,j,k\},\{i',j',k'\}\}\in T_4}\left[X_{ijk}X_{i'j'k'}\right] = p^3$$
   And, $|T_4| = \binom{n}{3}$

Using the above, we can now write $E[X^2]$ as:

$$E[X^2] = 20\binom{n}{6}p^6 + 30\binom{n}{5}p^6 + 12\binom{n}{4}p^5 + \binom{n}{3}p^3$$

Thus, the variance:

$$var(X) = E[X^2] - E[X]^2$$

$$\Rightarrow var(X) = 20\binom{n}{6}p^6 + 30\binom{n}{5}p^6 + 12\binom{n}{4}p^5 + \binom{n}{3}p^3 - \binom{n}{3}^2 p^6$$

Notice that since $T = T_1 \cup T_2 \cup T_3 \cup T_4$, and $|T| = \binom{n}{3}^2$ (because it is a cartesian product):

$$20\binom{n}{6} + 30\binom{n}{5} + 12\binom{n}{4} + \binom{n}{3} = \binom{n}{3}^2$$

Therefore,

$$\begin{aligned} var(X) = 20\binom{n}{6}p^6 + 30\binom{n}{5}p^6 + 12\binom{n}{4}p^5 + \binom{n}{3}p^3 \\ -20\binom{n}{6}p^6 - 30\binom{n}{5}p^6 - 12\binom{n}{4}p^6 - \binom{n}{3}p^6 \end{aligned}$$

$$\begin{aligned} \Rightarrow \boldsymbol{var(X) = 12\binom{n}{4}p^5(1-p) + \binom{n}{3}p^3(1-p^3)} \\ = \binom{n}{3}p^3\big(1 + 3p^2(n-3) + p^3(8-3n)\big) \end{aligned}$$

Substituting $p = 1/2$,

$$var(X) = \frac{12}{64}\binom{n}{4} + \frac{7}{64}\binom{n}{3} = \boldsymbol{\frac{3}{16}\binom{n}{4} + \frac{7}{64}\binom{n}{3}}$$

Or,

$$var(X) = \frac{\binom{n}{3}\left(1 + \frac{3}{4}(n-3) + \frac{1}{8}(8-3n)\right)}{8} = \frac{\binom{n}{3}(3n-2)}{64}$$

From Chebyshev's Inequality we know: $P[|\boldsymbol{X} - \boldsymbol{E}[\boldsymbol{X}]| \geq \delta] \leq \frac{\boldsymbol{\sigma^2}}{\delta^2}$. Putting $\delta = E[X]/2$:

$$P\left[|\boldsymbol{X} - \boldsymbol{E}[\boldsymbol{X}]| \geq \frac{E[X]}{2}\right] \leq \frac{4\boldsymbol{\sigma^2}}{E[X]^2} = 4\frac{\frac{\binom{n}{3}(3n-2)}{64}}{\left(\binom{n}{3}\cdot\frac{1}{8}\right)^2} = \frac{4(3n-2)}{\binom{n}{3}} = \frac{12\cdot(3n-2)}{n(n-1)(n-2)} = O\left(\frac{1}{n^2}\right)$$

Part-2:

Notice that if we have $X_{ijk} = 0$, then $X_{ijk'} = X_{ij'k} = X_{i'jk} = 0$.

So, clearly $X_{ijk}$ are dependent.

Let's define a couple of new things:

$$C = \big\{\{i,j\} |\, i,j \in V; i \neq j\big\}; |C| = \binom{n}{2}$$

$$c_i = the\ i^{th}\ element\ in\ C$$

$$e_c = edge\ between\ i\ and\ j, where\ c = (i,j) \in C$$

$$E_i = \begin{cases} 1, & if\ e_{c_i}\ exists\ in\ G \\ 0, & otherwise \end{cases}$$

$$X_{c_i k} = \begin{cases} 1, & triangle\ with\ c_i\ and\ vertex\ k\ exists \\ 0, & otherwise \end{cases}$$

Let $A_{i-1} \in \{0,1\}^{i-1}$, and $E_{i-1} = A_{i-1}$ denotes the values of the first $i-1$ $E_i s$.

To use the method of bounded difference, we need to show that for a set of $o_1, o_2 \ldots, o_i$:

$$|E[X|E_{i-1} = A_{i-1}, E_i = 1] - E[X|E_{i-1} = A_{i-1}, E_i = 0]| \leq o_i$$

Clearly if $E_{c_i} = 0$, then $X_{c_i k} = 0\ \forall k \in V/\{i,j\}$, where $i,j$ are the vertices forming the edge $c_i$. And if $E_{c_i} = 1$, then $X_{c_i k} = 1 \forall k \in V/\{i,j\}$ with probability $p^2$.

$$\therefore |E[X|E_{i-1} = A_{i-1}, E_i = 1] - E[X|E_{i-1} = A_{i-1}, E_i = 0]| \leq (n-2)p^2$$

Thus, taking $o_i = (n-2)p^2$

$$P(|X - E[X]| > E[X]/2) < \exp\left(-\frac{E[X]^2}{4\sum_{i=1}^{|c|} o_i^2}\right) = \exp\left(-\frac{\binom{n}{3}^3 p^6}{4\binom{n}{2}(n-2)^2 p^4}\right)$$

$$= \exp\left(-\frac{\binom{n}{3}^3 p^6}{12\binom{n}{3}(n-2)p^4}\right) = \exp\left(-\frac{n(n-1)}{72}p^2\right)$$

For, $p = \frac{1}{2}$:

$$P\left(|X - E[X]| > \frac{E[X]}{2}\right) < \exp\left(-\frac{n(n-1)}{288}\right)$$

Part-3:

The method of bounded difference gives an $O\left(e^{-n^2}\right)$ bound while Chebyshev's inequality gives an inverse polynomial bound of $O(n^{-2})$. The bound given by method of bounded differences is much tighter than that given by Chebyshev's inequality

Solution-3: -- (not yet complete)

What is the probability that a path of length $O(\log n)$ to exist in a $G(n,p)$ graph? $= p^{C\log n}$

We are now given that every vertex now adds an edge to every other vertex with probability $c/n$. Since the addition of new edges in the graph can only reduce the diameter, and since we expect a diameter of $O(\log n)$ after the addition of $\Theta(n)$ new edges, we need to find an estimate of how much the diameter reduces after addition of new edges. (The idea is to show (if possible) that reducing a diameter beyond $O(\log n)$ in $\Theta(n)$ new edges is difficult)

WLOG, consider that the current graph has a diameter with $l$+1 nodes on it, the length of the diameter is $l$.

What is the probability that the next edge will be selected on the diameter?

$$P(next\ edge\ in\ diameter) = 1 - P(next\ edge\ not\ on\ diameter)$$
$$= 1 - (1-p)^{\binom{l+1}{2}} \cong 1 - (1-p)^{l^2}$$

Let $q = 1 - p$

Now given a diameter of length $l$, what is the expected size of the diameter upon addition of a new edge?

The length of the diameter decreases by $x$ if two points chosen to form the edge are on the diameter and are separated by $x+1$. There are $l - x$ such pairs of points on $l$. And an edge can form independently for any such pair with probability $p$.

If we want that the probability of the diameter on the addition of next edge reduces by at least $x$, then we are also open to pairs of points that give a reduction $> x$, there are $\sum_{i=x}^{l} l - i = \frac{(l-x)(l-x+1)}{2}$ such pairs.

$$\therefore P(length\ decreases\ atleast\ by\ x \mid edge\ on\ diameter)$$
$$= 1 - P(no\ edge\ of\ such\ a\ pair\ of\ point)$$
$$= 1 - q^{\frac{(l-x)(l-x+1)}{2}}$$

If we put $x = \frac{l}{2}$:

$$P(length\ decreases\ atleast\ by\ x \mid edge\ on\ diameter) = 1 - q^{\frac{l(l+2)}{8}}$$
$$\Rightarrow P(length\ decreases\ atleast\ by\ x\ ) = P(next\ edge\ in\ diameter)\left(1 - q^{\frac{l(l+2)}{8}}\right)$$
$$\cong \left(1 - q^{l^2}\right)\left(1 - q^{\frac{l(l+2)}{8}}\right) = p_l$$

What is the expected number of edges that we need to draw before the diameter reduces to at least half of the original?$= 1/p_l$

We can partition the evolution of the graph on addition of the next edges as:

X: The diameter reduced by at least half on addition of edge
O: The diameter did not reduce by at least half on addition of edge

OOX|OOOOX|OOOOOOOX|OOOOOOOOOOOOOOX|….OO
There will be a total of n events in the sequence