Part D

(i)

The probability that an edge of graph $G = (V, E)$ is present is $p_e = \frac{M}{C_n^2} = \frac{2M}{N(N-1)}$.

Each vertex $v \in V$, could potentially connect $(N-1)$ other vertex, pair-wisely. Then the expected number of neighbors of a vertex of the graph is $(N-1)\frac{2M}{N(N-1)} = \frac{2M}{N}$.

Since each vertex is sampled identically and independently, their sum of expectation equals to their expectation of sum.

The expected number of vertex being sampled is $E(|V_p|) = N\frac{60}{\sqrt{M}}$, and the expected storage requirement is $E(|V_p| + \sum_{u \in V_p} |N(u)|) = N\frac{60}{\sqrt{M}} + N\frac{60}{\sqrt{M}}\frac{2M}{N} = \frac{60N}{\sqrt{M}} + 120\sqrt{M}$.

(ii)

The probability that an edge of graph $G = (V, E)$ is present is between 0 and 1. i.e.

$0 \le \frac{2M}{N(N-1)} \le 1$. So, $2M \le N(N-1) \le N^2$. $\sqrt{2M} \le N$.

Thus, $\frac{60N}{\sqrt{M}} + 120\sqrt{M} \ge \frac{60\sqrt{2M}}{\sqrt{M}} + 120\sqrt{M} = 60\sqrt{2} + 120\sqrt{M}$ (lower bound).

A connected graph with N vertex, has at least $(N-1)$ edges. It has $(N-1)$ when all the vertices are connected one by one and there is no circle in that graph. Reducing edge from the graph above will make the graph no longer connected. So, $M \ge N-1$. $M+1 \ge N$.

Thus, $\frac{60N}{\sqrt{M}} + 120\sqrt{M} \le \frac{60(M+1)}{\sqrt{M}} + 120\sqrt{M} = \frac{60}{\sqrt{M}} + 180\sqrt{M}$ (upper bound).

The upper bound is $\frac{60}{\sqrt{M}} + 180\sqrt{M}$ and the lower bound is $60\sqrt{2} + 120\sqrt{M}$.

(iii)

(A)

Assume that the number of neighbors of a vertex $v \in V$ of the graph $G = (V, E)$ is k. As calculated above, the expected number of neighbors of a vertex of the graph $G = (V, E)$ is $E(k) = \frac{2M}{N}$.

Since a vertex $u \in V_p$ in graph $G_p = (V, E)$ is sampled from graph $G = (V, E)$, and for every $u \in V_p$, its entire neighbors are stored, the expected number of neighbors of a vertex $u \in V_p$ of the graph $G_p = (V, E_p)$ is also $E[deg_{G_p}(u)] = (N-1)\frac{2M}{N(N-1)} = \frac{2M}{N}$.

$$E[\hat{D}] = E[\frac{1}{p}\frac{1}{|V|}\sum_{u\in V_p} deg_{G_p}(u)] = \frac{1}{p}\frac{1}{|V|}E[\sum_{u\in V_p} deg_{G_p}(u)] = \frac{1}{p}\frac{1}{|V|}\sum_{u\in V_p} E[deg_{G_p}(u)] = \frac{1}{p}\frac{1}{|V|}\frac{60N}{\sqrt{M}}\frac{2M}{N} = \frac{2M}{N}$$

(B)

For each vertex $u \in V_p$ in graph $G_p = (V, E)$, there are potentially $(N-1)$ neighbors and the probability that a neighbor is connected with the vertex is $p_e = \frac{M}{C_n^2} = \frac{2M}{N(N-1)}$. Since presence of edge between the vertex and a potential neighbor is independent and the degree of $u \in V_p$ in graph $G_p = (V, E_p)$ is sum of the presence of edges (1 indicates presence; 0 indicates absence), the distribution of degree of u, $deg_{G_p}(u)$, is binomial.

$$Prob[deg_{G_p}(u) = k] = C_{N-1}^k p_e^k (1-p_e)^{N-1-k}$$

Then the variance of degree of u is

$$Var[deg_{G_p}(u)] = (N-1)p_e(1-p_e) = (N-1)\frac{2M}{N(N-1)}[1-\frac{2M}{N(N-1)}]$$

$$Var[\hat{D}] = Var[\frac{1}{p}\frac{1}{|V|}\sum_{u\in V_p} deg_{G_p}(u)] = \frac{1}{p^2}\frac{1}{|V|^2}\sum_{u\in V_p} Var[deg_{G_p}(u)]$$

$$= \frac{1}{p^2}\frac{1}{N^2}\frac{60N}{\sqrt{M}}(N-1)\frac{2M}{N(N-1)}[1-\frac{2M}{N(N-1)}] = \frac{M\sqrt{M}}{30N^2}[1-\frac{2M}{N(N-1)}]$$

(C)

$$D = \frac{1}{|V|}\sum_{v\in V} deg_G(v);\ E(D) = \frac{1}{N}E[\sum_{v\in V} deg_G(v)] = \frac{1}{N}\sum_{v\in V} E[deg_G(v)] = \frac{1}{N}N\frac{2M}{N} = \frac{2M}{N}.$$

$$\hat{D} = \frac{1}{p|V|}\sum_{u\in V_p} deg_{G_p}(u);\ E[\hat{D}] = \frac{2M}{N}.$$

Thus, $E[D] = E[\hat{D}]$, $E[\sum_{v\in V} deg_G(v)] = \frac{1}{p}E[\sum_{u\in V_p} deg_{G_p}(u)]$.

$Prob(|D-\hat{D}| \geq \frac{D}{2}) = Prob[|\frac{1}{|V|}\sum_{v\in V} deg_G(v) - \frac{1}{p|V|}\sum_{u\in V_p} deg_{G_p}(u)| \geq \frac{1}{2}\frac{1}{|V|}\sum_{v\in V} deg_G(v)]$ can be rewritten as $Prob(|S-\mu| \geq \frac{S}{2})$, where $S = \sum_{v\in V} deg_G(v)$ and the mean of S is

$$\mu = E[\sum_{v\in V} deg_G(v)] = \frac{1}{p}E[\sum_{u\in V_p} deg_{G_p}(u)] = \frac{1}{p}\sum_{u\in V_p} deg_{G_p}(u) = 2M.$$

$(\frac{1}{p}E[\sum_{u\in V_p} deg_{G_p}(u)] = \frac{1}{p}\sum_{u\in V_p} deg_{G_p}(u)$ because it is estimation)

From Chernoff Bound,

$$Prob(S-\mu \geq \frac{S}{2}) = Prob(S \geq 2\mu) \leq e^{-\mu/3} = e^{-2M/3}$$

$$Prob(S-\mu \leq -\frac{S}{2}) = Prob(S \leq \frac{2\mu}{3}) \leq e^{-\mu(\frac{1}{3})^2/2} = e^{-M/9}$$

So,

$$Prob(|D-\hat{D}| \geq \frac{D}{2}) = Prob(|S-\mu| \geq \frac{S}{2}) = Prob(S-\mu \geq \frac{S}{2}) + Prob(S-\mu \leq -\frac{S}{2})$$

$$Prob(|D-\hat{D}| \geq \frac{D}{2}) \leq e^{-2M/3} + e^{-M/9}$$

(iv)