APPENDIX

Proof of Lemma 3.1. Let $p(s,v)$ denote the probability that v is activated by s. Suppose the blocker set $B=\emptyset$. For any $v\in(V\backslash S)$, $D_s^L(B\cup\{v\})-D_s^L(B)\geq p(s,v)\geq 0$, thus $D_s^L(\cdot)$ is monotone nondecreasing. Let $f(B)=|\cup_{v\in B}N_\phi(v)|$. Since a nonnegative linear combination of submodular functions is also submodular, we only need to prove $f(\cdot)$ is submodular for any ϕ. Due to the lower bound ignoring the combination effect, we have $f(S\cup\{x\})-f(S)=|N_\phi(x)\backslash N_\phi(S)|$ and $f(T\cup\{x\})-f(T)=|N_\phi(x)\backslash N_\phi(T)|$. Since $S\subseteq T$, we have $N_\phi(S)\subseteq N_\phi(T)$. Thus $N_\phi(x)\backslash N_\phi(S)\supseteq N_\phi(x)\backslash N_\phi(T)$. Accordingly, $f(S\cup\{x\})-f(S)\geq f(T\cup\{x\})-f(T)$, which shows that $D_s^L(\cdot)$ is submodular. The lemma holds. □

Proof of Lemma 4.2.

$$\begin{aligned}D_s^L(B)&=\sum_{v\in(V\backslash S)}\sum_{\phi\in\Phi(v,B)}p(\phi)\\&=\sum_{v\in(V\backslash S)}\sum_{\phi\in\Phi_s(v)}p(\phi)\cdot\min\{|B\cap C_\phi(s,v)|,1\}\\&=\sum_{\phi\in\Omega}p(\phi)\sum_{v\in(R_\phi(s)\backslash S)}\min\{|B\cap C_\phi(s,v)|,1\}\\&=\sum_{\phi\in\Omega}p(\phi)\cdot Cov_{C_\phi^s}(B)\\&=\mathbb{E}_{\Phi\sim\Omega}[Cov_{C_\Phi^s}(B)]=\mathbb{E}[\frac{Cov_{\mathbb{C}^s}(B)}{|\mathbb{C}^s|}],\end{aligned}$$

where $\Phi(v,B)=\{\phi\in\Omega:\exists\, u\in B \text{ s.t. } u\in\bigcap_{i=1}^{j}P_i(s,v)\}$, j is the number of paths from s to v and $\Phi_s(v)=\{\phi\in\Omega:v\in R_\phi(s)\}$. □

Proof of Lemma 4.4.

$$\begin{aligned}D_s^U(B)&=\sum_{v\in V_s'}\Pr_{\Phi\sim\Omega}[B\cap L_\Phi(v)\neq\emptyset]\\&=|V_s'|\cdot\sum_{v\in V_s'}\Pr_{\Phi\sim\Omega}[B\cap L_\Phi(v)\neq\emptyset]\cdot\frac{1}{|V_s'|}\\&=|V_s'|\cdot\Pr_{\Phi\sim\Omega,v\sim V_s'}[B\cap L_\Phi(v)\neq\emptyset]\\&=|V_s'|\cdot\mathbb{E}[\frac{Cov_{\mathbb{L}}(B)}{|\mathbb{L}|}].\end{aligned}$$

Thus, the lemma holds. □

Proof of Lemma 5.1. Martingale is a sequence of random variables $Y_1,Y_2,\ldots,Y_i$ if and only if $E[Y_i\mid Y_1,Y_2,\ldots,Y_{i-1}]=Y_{i-1}$ and $E[|Y_i|]<+\infty$ for any i. The following lemma shows a concentration result for martingales.

Lemma 8.1 ([14]). *Let $Y_1,Y_2,\ldots,Y_i$ be a martingale, such that $Y_1\leq a$, $|Y_j-Y_{j-1}|\leq a$ for any $j\in[2,i]$, and*

$$Var[Y_1]+\sum_{j=2}^{i}Var[Y_j\mid Y_1,Y_2,\ldots,Y_{j-1}]\leq b,$$

where $Var[\cdot]$ is the variance of a random variable. For any $\lambda>0$,

$$\Pr[Y_i-\mathbb{E}[Y_i]\geq\lambda]\leq\exp(-\frac{\lambda^2}{\frac{2}{3}a\lambda+2b}).$$

Let $C_1^s,C_2^s,\ldots,C_\theta^s$ denote θ random CP sequences in $\mathbb{C}^s$, $\frac{Cov_{C_i^s}(B)}{\mathbb{E}[I_G(s)]}=x_i$ and $\frac{D_s^L(B)}{\mathbb{E}[I_G(s)]}=p$. Since each C_i^s is generated from a random realization, for any $i\in[1,\theta]$, we have

$$\mathbb{E}[x_i\mid x_1,x_2,\ldots,x_{i-1}]=\mathbb{E}[x_i]=p.$$

Let $M_i=\sum_{j=1}^{i}(x_j-p)$, thus $\mathbb{E}[M_i]=\sum_{j=1}^{i}\mathbb{E}[(x_j-p)]=0$, and

$$\begin{aligned}\mathbb{E}[M_i\mid M_1,M_2,\ldots,M_{i-1}]&=\mathbb{E}[M_{i-1}+(x_i-p)\mid M_1,M_2,\ldots,M_{i-1}]\\&=M_{i-1}+\mathbb{E}[x_i]-p=M_{i-1}.\end{aligned}$$

Therefore, $M_1,M_2,\ldots,M_\theta$ is a martingale. Since $x_i,p\in[0,1]$, we have $M_1\leq 1$ and $M_k-M_{k-1}\leq 1$ for any $k\in[2,\theta]$. In addition,

$$Var[M_1]+\sum_{k=2}^{\theta}Var[M_k\mid M_1,M_2,\ldots,M_{k-1}]=\sum_{k=1}^{\theta}Var[x_k].$$

We know that $Var[x_i]=\mathbb{E}[x_i^2]-(\mathbb{E}[x_i])^2$ and $x_i\in[0,1]$, thus we have $Var[x_i]\leq\mathbb{E}[x_i]-(\mathbb{E}[x_i])^2=p(1-p)$. Therefore,

$$Var[M_1]+\sum_{k=2}^{\theta}Var[M_k\mid M_1,M_2,\ldots,M_{k-1}]\leq\theta p(1-p).$$

Then by Lemma 8.1 and $M_\theta=\sum_{i=1}^{\theta}(x_i-p)$, we can get that Eq. (9) holds. In addition, by applying Lemma 8.1 on the martingale $-M_1,-M_2,\ldots,-M_\theta$, we can deduce that Eq. (10) holds. □

Proof of Lemma 5.2. Let $\theta_1=\frac{2\mathbb{E}[I_G(s)]\ln(12/\delta)}{\epsilon_1^2\cdot D_s^L(B_L^o)}$ and when $\theta\geq\theta_1$, we have:

$$\begin{aligned}&\Pr[Cov_{\mathbb{C}_1^s}(B_L^o)/\theta\leq(1-\epsilon_1)D_s^L(B_L^o)]\\&=\Pr[\frac{Cov_{\mathbb{C}_1^s}(B_L^o)}{\mathbb{E}[I_G(s)]}-\frac{D_s^L(B_L^o)\cdot\theta}{\mathbb{E}[I_G(s)]}\leq-\frac{\epsilon_1\cdot\theta\cdot D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]}]\\&\leq\exp[-\frac{\epsilon_1^2}{2}\cdot\theta\cdot\frac{D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]}]\leq\frac{\delta}{12},\end{aligned}\tag{15}$$

where $\epsilon_1<\epsilon$ and the first inequality is due to Eq. (10). Since $D_s^L(\cdot)$ is submodular, we can also deduce that $Cov_{\mathbb{C}_1^s}(\cdot)$ is submodular, thus $Cov_{\mathbb{C}_1^s}(B_L)\geq(1-1/e)Cov_{\mathbb{C}_1^s}(B_L^o)$ and we have:

$$\Pr[\frac{Cov_{\mathbb{C}_1^s}(B_L)}{\theta}\geq(1-1/e)(1-\epsilon_1)D_s^L(B_L^o)]\geq 1-\frac{\delta}{12}.\tag{16}$$

Let $\theta_2=\frac{(2-2/e)\mathbb{E}[I_G(s)]\ln(\binom{n-|S|}{k}\cdot 12/\delta)}{D_s^L(B_L^o)(\epsilon-(1-1/e)\epsilon_1)^2}$ and $\epsilon_2=\epsilon-(1-1/e)\cdot\epsilon_1$, when $\theta\geq\theta_2$, we assume $D_s^L(B_L)<(1-1/e-\epsilon)D_s^L(B_L^o)$, thus:

$$\begin{aligned}&\Pr[\frac{Cov_{\mathbb{C}_1^s}(B_L)}{\theta}-D_s^L(B_L)\geq\epsilon_2 D_s^L(B_L^o)]\\&=\Pr[\frac{Cov_{\mathbb{C}_1^s}(B_L)}{\mathbb{E}[I_G(s)]}-\frac{\theta\cdot D_s^L(B_L)}{\mathbb{E}[I_G(s)]}\geq\frac{\theta\cdot\epsilon_2\cdot D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]}]\\&\leq\exp(-\frac{\epsilon_2^2\cdot D_s^L(B_L^o)^2\cdot\theta}{(2D_s^L(B_L)+\frac{2}{3}D_s^L(B_L^o)\cdot\epsilon_2)\cdot\mathbb{E}[I_G(s)]})\\&\leq\exp(-\frac{\epsilon_2^2\cdot D_s^L(B_L^o)\cdot\theta}{(2(1-1/e-\epsilon)+\frac{2}{3}\epsilon_2)\cdot\mathbb{E}[I_G(s)]})\leq\frac{\delta}{12\cdot\binom{n-|S|}{k}},\end{aligned}\tag{17}$$

where the first inequality is due to Eq. (9). According to Eq. (16), Eq. (17) and there exists at most $\binom{n-|S|}{k}$ blocker sets, when $\theta \geq \max\{\theta_1, \theta_2\}$, we can get:

$$\Pr[D_s^L(B_L) \geq (1 - 1/e - \epsilon) D_s^L(B_L^o)] \geq 1 - \frac{\delta}{6}, \tag{18}$$

which contradicts the assumption. Based on Line 4 of Algorithm 4, we have $\Pr\left[(1-\beta)\mathbb{E}[I_G(s)] \leq \hat{I}_G(s) \leq (1+\beta)\mathbb{E}[I_G(s)]\right] \geq 1 - \delta/6$. In addition, OPT^L is the lower bound of $D_s^L(B_L^o)$, thus we have $\theta_{\max} \geq \max\{\theta_1, \theta_2\}$ holds at least $1 - \delta/6$ probability. Accordingly, by the union bound, the proof of this lemma is done. □

Proof of Lemma 5.3.

$$\Pr\left[\frac{D_s^L(B_L)}{\mathbb{E}[I_G(s)]} < \left((\sqrt{\frac{Cov_{\mathbb{C}_2^s}(B_L)}{\mathbb{E}[I_G(s)]} + \frac{2a_1}{9}} - \sqrt{\frac{a_1}{2}})^2 - \frac{a_1}{18} \right) \cdot \frac{1}{|\mathbb{C}_2^s|}\right]$$
$$= \Pr\left[\frac{Cov_{\mathbb{C}_2^s}(B_L)}{\mathbb{E}[I_G(s)]} - \frac{D_s^L(B_L) \cdot |\mathbb{C}_2^s|}{\mathbb{E}[I_G(s)]} > \sqrt{2a_1 \frac{D_s^L(B_L) \cdot |\mathbb{C}_2^s|}{\mathbb{E}[I_G(s)]} + \frac{a_1^2}{9}} + \frac{a_1}{3}\right]$$
$$\leq \exp(-a_1) = \frac{\delta}{3i_{\max}}.$$

Let $x = Cov_{\mathbb{C}_2^s}(B_L)/\mathbb{E}[I_G(s)]$ and $f(x) = (\sqrt{x + \frac{2a_1}{9}} - \sqrt{\frac{a_1}{2}})^2 - \frac{a_1}{18}) \cdot \frac{1}{|\mathbb{C}_2^s|}$. When $x \geq 5a_1/18$, $f(x)$ monotonically increasing, otherwise $f(x)$ decreases monotonically. Thus, when $x \geq Cov_{\mathbb{C}_2^s}(B_L) \cdot (1-\beta)/\hat{I}_G(s) \geq 5a_1/18$, $f(x)$ monotonically increasing and we have:

$$\frac{D_s^L(B_L)}{\mathbb{E}[I_G(s)]} \geq f(x) \geq f(\frac{Cov_{\mathbb{C}_2^s}(B_L) \cdot (1-\beta)}{\hat{I}_G(s)}).$$

Similarly, when $x \leq Cov_{\mathbb{C}_2^s}(B_L) \cdot (1+\beta)/\hat{I}_G(s) \leq 5a_1/18$, $f(x)$ decreases monotonically and we have:

$$\frac{D_s^L(B_L)}{\mathbb{E}[I_G(s)]} \geq f(x) \geq f(\frac{Cov_{\mathbb{C}_2^s}(B_L) \cdot (1+\beta)}{\hat{I}_G(s)}).$$

Therefore, $\sigma^L(B_L)$ is the lower bound of $\frac{D_s^L(B_L)}{\mathbb{E}[I_G(s)]}$ with at least $1 - \frac{\delta}{3i_{\max}}$ probability. Similar to the proof of Lemma 5.2 in [34], we can get $Cov^u_{\mathbb{C}_1^s}(B_L^o)$ is the upper bound of $Cov_{\mathbb{C}_1^s}(B_L^o)$, where B_i be a set containing the i nodes that are selected in the first i iterations of the Procedure Max-Coverage and $maxMC(B_i, k)$ denotes the set of k nodes with the k largest marginal coverage in $\mathbb{C}_1^s$ with respect to B_i. Thus,

$$\Pr\left[\frac{D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]} > \left(\sqrt{\frac{Cov^u_{\mathbb{C}_1^s}(B_L^o)}{\mathbb{E}[I_G(s)]} + \frac{a_2}{2}} + \sqrt{\frac{a_2}{2}}\right)^2 \cdot \frac{1}{|\mathbb{C}_1^s|}\right]$$
$$\leq \Pr\left[\frac{D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]} > \left(\sqrt{\frac{Cov_{\mathbb{C}_1^s}(B_L^o)}{\mathbb{E}[I_G(s)]} + \frac{a_2}{2}} + \sqrt{\frac{a_2}{2}}\right)^2 \cdot \frac{1}{|\mathbb{C}_1^s|}\right]$$
$$\leq \Pr\left[\frac{Cov_{\mathbb{C}_1^s}(B_L^o)}{\mathbb{E}[I_G(s)]} - \frac{|\mathbb{C}_1^s| \cdot D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]} < -\sqrt{2a_2 \cdot \frac{|\mathbb{C}_1^s| \cdot D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]}}\right]$$
$$\leq \exp(-a_2) = \frac{\delta}{3i_{\max}}.$$

Let $y = \frac{Cov^u_{\mathbb{C}_1^s}(B_L^o)}{\mathbb{E}[I_G(s)]}$ and $g(y) = \left(\sqrt{y + \frac{a_2}{2}} + \sqrt{\frac{a_2}{2}}\right)^2 \cdot \frac{1}{|\mathbb{C}_1^s|}$. $g(y)$ always increases monotonically and we have:

$$\frac{D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]} \leq g(y) \leq g(\frac{Cov^u_{\mathbb{C}_1^s}(B_L^o) \cdot (1+\beta)}{\hat{I}_G(s)}).$$

Therefore, $\sigma^U(B_L^o)$ is the upper bound of $\frac{D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]}$ with at least $1 - \frac{\delta}{3i_{\max}}$ probability. □

Proof of Theorem 5.5. The primary time overhead of LSBM lies in (i) estimating the influence of seeds of misinformation; (ii) generating the CP sequences; (iii) executing the Max-Coverage algorithm and computing $\sigma^L(B_L)$ and $\sigma^U(B_L^o)$ in all iterations.

As shown in [47], the time complexity of influence estimation is $O(\frac{m \cdot \ln 1/\delta}{\beta^2})$. Then. we analyze the number of CP sequences generated by LSBM.

Let $\epsilon_1 = \epsilon$, $\tilde{\epsilon}_1 = \epsilon/e$, $\hat{\epsilon}_1 = \sqrt{\frac{2a_3\mathbb{E}[I_G(s)]}{D_s^L(B_L^o)\theta_1}}$, $\epsilon_2 = \sqrt{\frac{2a_3\mathbb{E}[I_G(s)]}{D_s^L(B_L)\theta_2}}$, $\tilde{\epsilon}_2 = (\sqrt{\frac{2a_3 D_s^L(B_L)\theta_2}{\mathbb{E}[I_G(s)]} + \frac{a_3^2}{9}} + \frac{a_3}{3}) \cdot \frac{\mathbb{E}[I_G(s)]}{D_s^L(B_L)\theta_2}$, $a_3 = c\ln(\frac{3i_{\max}}{\delta})$ for any $c \geq 1$. In addition, let $\theta_a = \frac{2\mathbb{E}[I_G(s)]\ln\frac{6}{\delta}}{(1-1/e-\epsilon)\epsilon_1^2 D_s^L(B_L^o)}$, $\theta_b = \frac{(2+2\tilde{\epsilon}_1/3)\mathbb{E}[I_G(s)]\ln\frac{6\binom{n-|S|}{k}}{\delta}}{\tilde{\epsilon}_1^2 D_s^L(B_L^o)}$, $\theta_c = \frac{27\mathbb{E}[I_G(s)]\ln\frac{3i_{\max}}{\delta}(1+\beta)^2}{(1-1/e-\epsilon)(\epsilon_1+\epsilon_1\beta-2\beta)^2 D_s^L(B_L^o)}$, $\theta_d = \frac{5\ln\frac{3i_{\max}}{\delta}\mathbb{E}[I_G(s)]}{18(1-\epsilon_1)(1-1/e-\epsilon)D_s^L(B_L^o)}$ and $\theta' = \max\{\theta_a, \theta_b, \theta_c, \theta_d\}$. It is easy to verify that

$$\theta' = O\left(\frac{(k\ln(n-|S|) + \ln 1/\delta)\mathbb{E}[I_G(s)]}{(\epsilon + \epsilon\beta - 2\beta)^2 D_s^L(B_L^o)}\right). \tag{19}$$

When $\theta_1 = \theta_2 = c\theta'$, based on Eq. (9) and Eq. (10), we have:

$$\Pr[\frac{Cov_{\mathbb{C}_1^s}(B_L^o)}{\theta_1} < (1-\epsilon_1) \cdot D_s^L(B_L^o)] \leq \left(\frac{\delta}{6}\right)^c, \tag{20}$$

$$\Pr[\frac{Cov_{\mathbb{C}_2^s}(B_L)}{\theta_2} < (1-\epsilon_1) \cdot D_s^L(B_L)] \leq \left(\frac{\delta}{6}\right)^c, \tag{21}$$

$$\Pr[\frac{Cov_{\mathbb{C}_1^s}(B_L)}{\theta_1} > D_s^L(B_L) + \tilde{\epsilon}_1 \cdot D_s^L(B_L^o)] \leq \left(\frac{\delta}{6\binom{n-|S|}{k}}\right)^c, \tag{22}$$

$$\Pr[\frac{Cov_{\mathbb{C}_1^s}(B_L^o)}{\theta_1} < (1-\hat{\epsilon}_1) \cdot D_s^L(B_L^o)] \leq \left(\frac{\delta}{3i_{\max}}\right)^c, \tag{23}$$

$$\Pr[\frac{Cov_{\mathbb{C}_2^s}(B_L)}{\theta_2} < (1-\epsilon_2) \cdot D_s^L(B_L)] \leq \left(\frac{\delta}{3i_{\max}}\right)^c, \tag{24}$$

$$\Pr[\frac{Cov_{\mathbb{C}_2^s}(B_L)}{\theta_2} > (1+\tilde{\epsilon}_2) \cdot D_s^L(B_L)] \leq \left(\frac{\delta}{3i_{\max}}\right)^c. \tag{25}$$

Specially, when $\theta_1 \geq c\theta_a$, Eq. (20) holds; when $\theta_1 \geq c\theta_b$, Eq. (22) holds. Eq. (23)-Eq. (25) are obtained based on the definition of $\hat{\epsilon}_1, \epsilon_2$ and $\tilde{\epsilon}_2$. When the event in Eq. (20) and Eq. (22) not happen, we get:

$$D_s^L(B_L) \geq (1 - 1/e - \epsilon) D_s^L(B_L^o). \tag{26}$$

Based on Eq. (26), when when $\theta_1 \geq c\theta_a$, Eq. (21) holds. Since B_L is not independent of $\mathbb{C}_1^s$ and there are at most $\binom{n-|S|}{k}$ blocker sets, based on the union bound, the probability that none of the events in Eq. (20)-Eq. (25) happens is at least:

$$1-\left((\frac{\delta}{6})^c\cdot 2+(\frac{\delta}{6\binom{n-|S|}{k}})^c\cdot\binom{n-|S|}{k}+i_{\max}\cdot(\frac{\delta}{3i_{\max}})^c\right)\geq 1-\delta^c.$$

And we have:

$$\hat{\epsilon}_1\leq\sqrt{\frac{2(1-1/e-\epsilon)(\epsilon_1+\epsilon_1\beta-2\beta)^2}{27(1+\beta)^2}}\leq\frac{\epsilon_1+\epsilon_1\beta-2\beta}{3(1+\beta)},$$
$$\epsilon_2\leq\sqrt{\frac{2(1-1/e-\epsilon)D_s^L(B_L^o)(\epsilon_1+\epsilon_1\beta-2\beta)^2}{27(1+\beta)^2D_s^L(B_L)}}\leq\frac{\epsilon_1+\epsilon_1\beta-2\beta}{3(1+\beta)},$$
$$\tilde{\epsilon}_2\leq\sqrt{\frac{(\epsilon_1+\epsilon_1\beta-2\beta)^2(2+2\tilde{\epsilon}_2/3)}{27(1+\beta)^2}}\leq\frac{\epsilon_1+\epsilon_1\beta-2\beta}{3(1+\beta)}.$$

In addition, when the event in Eq. (23) not happen, we have:

$$\left(\sqrt{\frac{Cov_{\mathbb{C}_1^s}(B_L^o)\cdot(1+\beta)}{\hat{I}_G(s)}+\frac{a_3}{2}}+\sqrt{\frac{a_3}{2}}\right)^2\cdot\frac{1}{\theta_1}\geq\frac{D_s^L(B_L^o)}{\mathbb{E}[I_G(s)]}.$$

Thus, it holds that:

$$\begin{aligned}1-\hat{\epsilon}_1&=1-\sqrt{\frac{2a_3\mathbb{E}[I_G(s)]}{D_s^L(B_L^o)\theta_1}}\\&\leq 1-\frac{\sqrt{2a_3}}{\sqrt{Cov_{\mathbb{C}_1^s}(B_L^o)\cdot(1+\beta)/\hat{I}_G(s)+\frac{a_3}{2}}+\sqrt{\frac{a_3}{2}}}\\&\leq\frac{Cov^u_{\mathbb{C}_1^s}(B_L^o)\cdot(1+\beta)/\hat{I}_G(s)}{(\sqrt{Cov^u_{\mathbb{C}_1^s}(B_L^o)\cdot(1+\beta)/\hat{I}_G(s)+\frac{a_3}{2}}+\sqrt{\frac{a_3}{2}})^2}.\end{aligned}$$

Since $a_2=\ln(\frac{3i_{\max}}{\delta})\leq a_3$, based on Line 19 of Algorithm 4, thus

$$\begin{aligned}\sigma^U(B_L^o)&\leq\left(\sqrt{\frac{Cov_{\mathbb{C}_1^s}(B_L^o)\cdot(1+\beta)}{\hat{I}_G(s)}+\frac{a_3}{2}}+\sqrt{\frac{a_3}{2}}\right)^2\cdot\frac{1}{\theta_1}\\&\leq\frac{Cov^u_{\mathbb{C}_1^s}(B_L^o)\cdot(1+\beta)/\hat{I}_G(s)}{1-\hat{\epsilon}_1}\cdot\frac{1}{\theta_1}.\end{aligned}\tag{27}$$

When $\theta_2\geq\theta_d$ and according to Eq. (21), we have:

$$\frac{Cov_{\mathbb{C}_2^s}(B_L)}{\mathbb{E}[I_G(s)]}\geq\frac{\theta_2\cdot(1-\epsilon_1)D_s^L(B_L)}{\mathbb{E}[I_G(s)]}\geq\frac{5a_1}{18}.$$

Thus, $f(x)=(\sqrt{x+\frac{2a_1}{9}}-\sqrt{\frac{a_1}{2}})^2-\frac{a_1}{18})$ monotonically increasing. In addition, when the event in Eq. (25) does not happen, we have:

$$\left((\sqrt{\frac{Cov_{\mathbb{C}_2^s}(B_L)\cdot(1-\beta)}{\hat{I}_G(s)}+\frac{2a_3}{9}}-\sqrt{\frac{a_3}{2}})^2-\frac{a_3}{18}\right)\cdot\frac{1}{\theta_2}\leq\frac{D_s^L(B_L)}{\mathbb{E}[I_G(s)]}.$$

Thus, it holds that:

$$\begin{aligned}&\frac{Cov_{\mathbb{C}_2^s}(B_L)\cdot(1-\beta)}{\hat{I}_G(s)}-\frac{\tilde{\epsilon}_2D_s^L(B_L)\cdot\theta_2}{\mathbb{E}[I_G(s)]}\\=&\frac{Cov_{\mathbb{C}_2^s}(B_L)\cdot(1-\beta)}{\hat{I}_G(s)}-(\sqrt{2a_3\frac{D_s^L(B_L)\cdot\theta_2}{\mathbb{E}[I_G(s)]}+\frac{a_3^2}{9}}+\frac{a_3}{3})\\\leq&\frac{Cov_{\mathbb{C}_2^s}(B_L)\cdot(1-\beta)}{\hat{I}_G(s)}-(\sqrt{\frac{2Cov_{\mathbb{C}_2^s}(B_L)(1-\beta)a_3}{\hat{I}_G(s)}+\frac{4a_3^2}{9}}-\frac{2a_3}{3})\\=&\left(\sqrt{\frac{Cov_{\mathbb{C}_2^s}(B_L)\cdot(1-\beta)}{\hat{I}_G(s)}+\frac{2a_3}{9}}-\sqrt{\frac{a_3}{2}}\right)^2-\frac{a_3}{18}.\end{aligned}$$

Since $a_1\leq a_3$, based on the Line 14 of Algorithm 4, thus

$$\sigma^L(B_L)\geq\frac{Cov_{\mathbb{C}_2^s}(B_L)\cdot(1-\beta)}{\hat{I}_G(s)\cdot\theta_2}-\frac{\tilde{\epsilon}_2D_s^L(B_L)}{\mathbb{E}[I_G(s)]}.\tag{28}$$

Putting Eq. (27) and Eq. (28) together, when none of the events in Eq. (20)-Eq. (25) happens, we have:

$$\begin{aligned}\frac{\sigma^L(B_L)}{\sigma^U(B_L^o)}&\geq\frac{\frac{Cov_{\mathbb{C}_2^s}(B_L)\cdot(1-\beta)}{\hat{I}_G(s)\cdot\theta_2}-\frac{\tilde{\epsilon}_2D_s^L(B_L)}{\mathbb{E}[I_G(s)]}}{\frac{Cov^u_{\mathbb{C}_1^s}(B_L^o)\cdot(1+\beta)/\hat{I}_G(s)}{1-\hat{\epsilon}_1}\cdot\frac{1}{\theta_1}}\\&\geq\frac{\theta_1\left(\frac{Cov_{\mathbb{C}_2^s}(B_L)\cdot(1-\beta)}{(1+\beta)\cdot\theta_2}-\tilde{\epsilon}_2\cdot D_s^L(B_L)\right)(1-\hat{\epsilon}_1)}{Cov^u_{\mathbb{C}_1^s}(B_L^o)}\\&\geq\frac{\theta_1\left(\frac{1-\beta}{1+\beta}\cdot(1-\epsilon_2)-\tilde{\epsilon}_2\right)\cdot D_s^L(B_L)(1-\hat{\epsilon}_1)}{Cov^u_{\mathbb{C}_1^s}(B_L^o)}\\&\geq\frac{\theta_1\left(\frac{1-\beta}{1+\beta}\cdot(1-\epsilon_2)-\tilde{\epsilon}_2\right)\cdot D_s^L(B_L)(1-\hat{\epsilon}_1)}{Cov_{\mathbb{C}_1^s}(B_L)}(1-1/e)\\&\geq\frac{\theta_1\left(1-\epsilon_2-\frac{2\beta}{1+\beta}-\tilde{\epsilon}_2-\hat{\epsilon}_1\right)\cdot D_s^L(B_L)}{Cov_{\mathbb{C}_1^s}(B_L)}(1-1/e)\\&\geq\frac{\theta_1\,(1-\epsilon_1)\cdot D_s^L(B_L)}{Cov_{\mathbb{C}_1^s}(B_L)}(1-1/e)\\&\geq\frac{\theta_1\,(1-\epsilon_1)\cdot\left(Cov_{\mathbb{C}_1^s}(B_L)/\theta_1-\tilde{\epsilon}_1\cdot D_s^L(B_L^o)\right)}{Cov_{\mathbb{C}_1^s}(B_L)}(1-1/e)\\&\geq\frac{(1-\epsilon_1)\cdot\left(Cov_{\mathbb{C}_1^s}(B_L)-\tilde{\epsilon}_1\cdot\frac{Cov_{\mathbb{C}_1^s}(B_L^o)}{(1-\epsilon_1)}\right)}{Cov_{\mathbb{C}_1^s}(B_L)}(1-1/e)\\&\geq(1-\epsilon_1)\left(1-\frac{\tilde{\epsilon}_1}{(1-\epsilon_1)(1-1/e)}\right)(1-1/e)\\&=1-1/e-\epsilon.\end{aligned}$$

Therefore, when $\theta_1=\theta_2=c\theta'$ CP sequences are generated, LSBM does not stop only if at least one of the events in Eq. (20)-Eq. (25) happens. The probability is at most δ^c.

Let j be the first iteration in which the number of CP sequences generated by LSBM reaches θ'. From this iteration onward, the expected number of CP sequences further generated is at most

$$\begin{aligned} 2 \cdot \sum_{z \geq j} \theta_0 \cdot 2^z \cdot \delta^{2^{z-j}} &= 2 \cdot 2^j \cdot \theta_0 \sum_{z=0} 2^z \cdot \delta^{2^z} \\ &\leq 4\theta' \sum_{z=0} 2^{-2^z+z} \\ &\leq 4\theta' \sum_{z=0} 2^{-z} \leq 8\theta'. \end{aligned}$$

If the algorithm stops before this iteration, there are at most $2\theta'$ CP sequences generated. Therefore, the expected number of CP sequences generated is less than $10\theta'$, which is

$$O(\frac{(k \ln (n-|S|) + \ln 1/\delta)\mathbb{E}[I_G(s)]}{(\epsilon + \epsilon\beta - 2\beta)^2 D_s^L(B_L^o)}). \tag{29}$$

We have shown that the expected time required to generate a CP sequence is $O(m \cdot \alpha(m,n))$. Based on Wald's equation, LSBM requires $O(\frac{(k \ln (n-|S|)+\ln 1/\delta)\mathbb{E}[I_G(s)] m \cdot \alpha(m,n)}{(\epsilon+\epsilon\beta-2\beta)^2 D_s^L(B_L^o)})$ in CP sequences generation. In addition, the total expected time used for executing the Max-Coverage and computing $\sigma^L(B_L)$ and $\sigma^U(B_L^o)$ in all the iterations is

$$\begin{aligned} &O(k(n-|S|) \cdot i_{\max} + 2\mathbb{E}[|\mathbb{C}_1^s \cup \mathbb{C}_2^s|] \cdot \mathbb{E}[|C^s|]) \\ =&O(\frac{(k \ln (n-|S|) + \ln 1/\delta)\mathbb{E}[I_G(s)]}{(\epsilon + \epsilon\beta - 2\beta)^2}). \end{aligned}$$

Thus, the theorem holds. □

Lemma 8.2. *Given a blocker set B, a seed node s and a fixed number of θ random LRR sets $\mathbb{L}$. For any $\lambda > 0$,*

$$\Pr[Cov_{\mathbb{L}}(B) - \frac{D_s^U(B) \cdot \theta}{|V_s'|} \geq \lambda] \leq \exp(-\frac{\lambda^2}{\frac{2D_s^U(B)}{|V_s'|} \cdot \theta + \frac{2}{3}\lambda}), \tag{30}$$

$$\Pr[Cov_{\mathbb{L}}(B) - \frac{D_s^U(B) \cdot \theta}{|V_s'|} \leq -\lambda] \leq \exp(-\frac{\lambda^2}{\frac{2D_s^U(B)}{|V_s'|} \cdot \theta}). \tag{31}$$

Proof. Let $L_1, L_2, \ldots, L_\theta$ denote θ random LRR sets in $\mathbb{L}$, $Cov_{L_i}(B) = x_i$ and $\frac{D_s^U(B)}{|V_s'|} = p$. Since each L_i is generated from a random realization, for any $i \in [1, \theta]$, we have

$$\mathbb{E}[x_i \mid x_1, x_2, \ldots, x_{i-1}] = \mathbb{E}[x_i] = p.$$

Let $M_i = \sum_{j=1}^i (x_j - p)$, thus $\mathbb{E}[M_i] = \sum_{j=1}^i \mathbb{E}[(x_j - p)] = 0$, and

$$\begin{aligned} \mathbb{E}[M_i \mid M_1, M_2, \ldots, M_{i-1}] &= \mathbb{E}[M_{i-1} + (x_i - p) \mid M_1, M_2, \ldots, M_{i-1}] \\ &= M_{i-1} + \mathbb{E}[x_i] - p = M_{i-1}. \end{aligned}$$

Therefore, $M_1, M_2, \ldots, M_\theta$ is a martingale. Since $x_i, p \in [0,1]$, we have $M_1 \leq 1$ and $M_k - M_{k-1} \leq 1$ for any $k \in [2, \theta]$. In addition,

$$\mathrm{Var}[M_1] + \sum_{k=2}^{\theta} \mathrm{Var}[M_k \mid M_1, M_2, \ldots, M_{k-1}] = \sum_{k=1}^{\theta} \mathrm{Var}[x_k].$$

We know that $\mathrm{Var}[x_i] = \mathbb{E}[x_i^2] - (\mathbb{E}[x_i])^2$ and $x_i \in [0,1]$, thus we have $\mathrm{Var}[x_i] \leq \mathbb{E}[x_i] - (\mathbb{E}[x_i])^2 = p(1-p)$. Therefore,

$$\mathrm{Var}[M_1] + \sum_{k=2}^{\theta} \mathrm{Var}[M_k \mid M_1, M_2, \ldots, M_{k-1}] \leq \theta p(1-p).$$

Then by Lemma 8.1 and $M_\theta = \sum_{i=1}^{\theta}(x_i - p)$, we can get that Eq. (30) holds. In addition, by applying Lemma 8.1 on the martingale $-M_1, -M_2, \ldots, -M_\theta$, we can deduce that Eq. (31) holds. □

Proof of Theorem 5.7. The proof of Theorem 5.7 is similar to that of Theorem 5.5 and the differences between them are that we use Lemma 8.1 as the concentration bounds and we set:

$$\theta_1 = \frac{2|V_s'| \ln(6/\delta)}{\epsilon_1^2 \cdot D_s^U(B_U^o)},$$

$$\theta_2 = \frac{(2-2/e)|V_s'| \ln(\binom{n-|S|}{k} \cdot 6/\delta)}{D_s^U(B_U^o)(\epsilon - (1-1/e)\epsilon_1)^2}, \epsilon_2 = \epsilon - (1-1/e) \cdot \epsilon_1,$$

$$\theta_{\max} = \frac{2|V_s'| \left((1-1/e)\sqrt{\ln \frac{6}{\delta}} + \sqrt{(1-1/e)(\ln \binom{|V_s'|-|S|}{k} + \ln \frac{6}{\delta})}\right)^2}{\epsilon^2 \mathrm{OPT}^L},$$

$$\theta_0 = \theta_{\max} \cdot \epsilon^2 \mathrm{OPT}^L / |V_s'|,$$

$$\sigma^L(B_U) = \left((\sqrt{Cov_{\mathbb{L}_2}(B_U) + \frac{2a_1}{9}} - \sqrt{\frac{a_1}{2}})^2 - \frac{a_1}{18} \right) \cdot \frac{|V_s'|}{|\mathbb{L}_2|},$$

$$\sigma^U(B_U^o) \leftarrow \left(\sqrt{Cov_{\mathbb{L}_1}^u(B_U^o) + \frac{a_2}{2}} + \sqrt{\frac{a_2}{2}} \right)^2 \cdot \frac{|V_s'|}{|\mathbb{L}_1|}.$$

Note that, $\sigma^L(B_U)$ is the lower bound of $D_s^U(B_U)$ with at least $1 - \frac{\delta}{3i_{\max}}$ probability, and $\sigma^U(B_U^o)$ is the upper bound of $D_s^U(B_U^o)$ with at least $1 - \frac{\delta}{3i_{\max}}$ probability. □

Proof of Theorem 5.8. The proof of Theorem 5.8 is similar to that of Theorem 5.6 and the differences between them are that we use Lemma 8.1 as the concentration bounds and we set: $\epsilon_1 = \epsilon$, $\tilde{\epsilon}_1 = \epsilon/e$, $\hat{\epsilon}_1 = \sqrt{\frac{2a_3|V_s'|}{D_s^U(B_U^o)\theta_1}}$, $\epsilon_2 = \sqrt{\frac{2a_3|V_s'|}{D_s^U(B_U)\theta_2}}$, $\tilde{\epsilon}_2 = (\sqrt{\frac{2a_3 D_s^U(B_U)\theta_2}{|V_s'|} + \frac{a_3^2}{9}} + \frac{a_3}{3}) \cdot \frac{|V_s'|}{D_s^U(B_U)\theta_2}$, $a_3 = c \ln(\frac{3i_{\max}}{\delta})$ for any $c \geq 1$. In addition, let $\theta_a = \frac{2|V_s'| \ln \frac{6}{\delta}}{\epsilon_1^2 D_s^U(B_U^o)}$, $\theta_b = \frac{(2+2\tilde{\epsilon}_1/3)|V_s'| \ln \frac{6\binom{n-|S|}{k}}{\delta}}{\tilde{\epsilon}_1^2 D_s^U(B_U^o)}$, $\theta_c = \frac{27|V_s'| \ln \frac{3i_{\max}}{\delta}}{(1-1/e-\epsilon)\epsilon_1^2 D_s^U(B_U^o)}$, and $\theta' = \max\{\theta_a, \theta_b, \theta_c\}$.

Based on the above setting, the expected number of LRR sets generated by GSBM is proved to be $O(\frac{(k \ln(|V_s'|-|S|) + \ln(1/\delta))|V_s'|}{\epsilon^2 \cdot D_s^U(B_U^o)})$. Besides, let $\mathcal{L}$ be the expected number of edges being traversed required to generate one LRR set for a randomly selected node in $|V_s'|$. Then, $\mathcal{L} = \mathbb{E}[p_L \cdot m]$, where the expectation is taken over the random choices of L. v^* is chosen from $|V_s'|$ according to a probability that is proportional to its in-degree, and we define the random variable:

$$X(v^*, L) = \begin{cases} 1 & \text{if } v^* \in L \\ 0 & \text{otherwise} \end{cases}$$

Then, for any fixed L, we have:

$$p_L = \sum_{v^*} \left(Pr[v^*] \cdot X(v^*, L) \right).$$

Consider the fixed v^* and vary L, we have:

$$p_{v^*} = \sum_{L} \left(Pr[L] \cdot X(v^*, L)\right).$$

In addition, by Lemma 4.4, we have $\mathbb{E}[p_{v^*}] = D_s^U(v^*)/|V_s'|$. Therefore,

$$\begin{aligned}
\mathcal{L}/m = \mathbb{E}[p_L] &= \sum_{L}(Pr[L] \cdot p_L) \\
&= \sum_{L}\left(Pr[L] \cdot \sum_{v^*}\left(Pr[v^*] \cdot X(v^*, L)\right)\right) \\
&= \sum_{v^*}\left(Pr[v^*] \cdot \sum_{L}\left(Pr[L] \cdot X(v^*, L)\right)\right) \\
&= \sum_{v^*}\left(Pr[v^*] \cdot p_{v^*}\right) = \mathbb{E}[p_{v^*}] = D_s^U(v^*)/|V_s'|.
\end{aligned}$$

Thus, we have $\mathcal{L} = D_s^U(v^*) \cdot \frac{m}{|V_s'|}$. Based on Wald's equation, we can show that GSBM requires $O(\frac{(k \ln(|V_s'|-|S|)+\ln(1/\delta))m}{\epsilon^2})$ expected time in LRR set generation. Furthermore, the total expected time used for executing the Max-Coverage and computing $\sigma^L(B_U)$ and $\sigma^U(B_U^o)$ in all the iterations is

$$\begin{aligned}
&O(k(n-|S|) \cdot i_{\max} + 2\mathbb{E}[|\mathbb{L}_1 \cup \mathbb{L}_2|] \cdot \mathbb{E}[|L|]) \\
=&O(\frac{(k \ln(n-|S|) + \ln(1/\delta))|V_s'|}{\epsilon^2}).
\end{aligned}$$

Thus, the theorem holds. □