1 Introduction

In this lecture, we analyze the Push-Pull protocol for rumor spreading. In the rumor spreading problem, we have a rumor originating at a source node in a given graph, and the goal is to broadcast the rumor to every node in the graph. Last session, we introduced two well-studied algorithms for rumor spreading, push and pull.

- **push**: In each round, each informed node informs a random neighbor.
- **pull**: In each round, each uninformed node seeks from a random neighbor.

We have also seen that the Push protocol completes rumor spreading in $O(n \log n)$ steps with high probability. Today, we will derive a tight bound on the Push-Pull protocol, as a function of the conductance of the underlying graph. This lecture is entirely based on the elegant analysis due to Giakkoupis [1].

1.1 Two examples on push and pull

- **push on star**: Assume that on a star graph with $n$ nodes, the central node $r$ is informed. It is easy to verify that the probability that a specific node $v$, a neighbor of $r$, becomes informed in one round is $\frac{1}{n-1}$, and the expected number of rounds which $v$ becomes informed is $n - 1$. Using a standard Coupon Collector argument, we can show that the expected number of rounds to broadcast the gossip to every node in this graph is $n \ln n$.

- **pull on star**: In this example, we consider two cases; One is the case that the informed node $r$ is the central node in the star graph. Then it takes only one round for every uniformed node to get the information. Second is the case that the informed node $r$ is not the central node. Then, it takes $\Theta(n)$ rounds to inform every node in this graph.

We thus find that push works fast when the rumor is at a node with high degree, while pull is effective at moving the rumor from a node with low degree to a node with high degree. Combining the complementary strengths of these two protocols enables the desired bound on the push-pull protocol.

- **push-pull**: In each round, each node $v$ selects a neighbor uniformly at random; if $v$ already has the neighbor, it pushes the rumor the selected neighbor; otherwise, if the neighbor has the rumor, then $v$ pulls from the neighbor. If neither $v$ nor its neighbor has the rumor, then nothing is transferred.
2 Analysis of push-pull algorithm

Theorem 1. In graph $G(V, E)$ with $n$ vertices and conductance $\phi$, the push-pull algorithm completes the task in $O\left(\frac{\log n}{\phi}\right)$ rounds whp.

Definition 1. We recall the definition of conductance $\phi$ in a graph $G(V, E)$.

$$\phi = \min_{\emptyset \neq S \subseteq V} \frac{|E(S, \overline{S})|}{\min\{\text{Vol}(S), \text{Vol}(\overline{S})\}}$$

where $\text{Vol}(S) = \sum_{v \in S} d_v$ and $E(S, \overline{S})$ is the set of edges of the graph that crosses the $S, \overline{S}$ cut.

Before proving Theorem 1, we present two lemmas which help us to complete the proof of the theorem.

Lemma 1. Suppose in a graph $G(V, E)$, the maximum degree vertex is in the set of informed nodes $S$. Then, in $O\left(\frac{\log n}{\phi}\right)$ rounds of pull algorithm, $S = V$ whp.

Lemma 2. In $O\left(\frac{\log n}{\phi}\right)$ rounds of push algorithm, max degree vertex is informed and is in $S$ whp.

Remark 1. By high probability we mean probability of at least $1 - \frac{1}{n^\beta}$ for some $\beta > 0$.

Proof of Theorem 1: Lemma 1 and 2 yield the proof of Theorem 1. \qed

We now argue that Lemma 1 and 2 are, in fact, equivalent and proving one of the lemmas implies the correctness of the other one. In other word, lemma 1 $\implies$ lemma 2 and lemma 2 $\implies$ lemma 1. In the following, we establish this claim.

Let $E_{push}(u, v, t)$ be the event that if node $u$ is informed at time 0, then in at most $t$ rounds of push, $v$ is informed, and $E_{pop}(v, u, t)$ be the event that if node $v$ is informed at time 0, then in at most $t$ rounds of pull, $u$ is informed.

In the following, we show that:

$$\Pr[E_{push}(u, v, t)] = \Pr[E_{pull}(v, u, t)]$$

Consider that at each round (from round 1 to $t$), each node (independently from it has or it doesn’t have the rumor), it picks on random edge or one random neighbor.

The space for such decisions $\Omega$ are points of the form $x = (\prod_{i} d_i)^t$. Then, if $x \in E_{push}(u, v, t)$, it means there are some edges at each round that let rumor starts from $u$ and reaches $v$ in at most $t$ rounds. This exactly means that using that edges $u$ can pull rumor from $v$ during these rounds. Thus $x \in E_{pull}(v, u, t)$.

$$(x_1^1, x_2^1, ..., x_n^1)$$

$$(x_1^2, x_2^2, ..., x_n^2)$$

$$\vdots$$

$$(x_1^t, x_2^t, ..., x_n^t)$$
Then we have:
\[ x \in E_{\text{push}}(u,v,t) \iff x \in E_{\text{pull}}(v,u,t) \]
This implies that
\[ \Pr[E_{\text{push}}(u,v,t)] = \Pr[E_{\text{pull}}(v,u,t)] \]
Having above argument, it suffices to prove only one of the above lemmas. Here we give a proof for lemma 1.

The proof for lemma 1 relies on the following key claim on the pull protocol. For any integer \( t \geq 0 \), let \( S_t \) denote the set of informed nodes, and \( U_t \) the set of uninformed nodes at time \( t \).

**Lemma 3.** In pull, consider two cases:

a. If \( \Delta \leq \text{Vol}(S_0) \leq |E| \), then after \( t = \lceil \frac{4}{\phi} \rceil \) rounds, we have the following:
\[ \Pr[\text{Vol}(S_t) > \min(2 \text{Vol}(S_0), |E|)] \geq \frac{1}{2}. \]

b. If \( \text{Vol}(S_0) > |E| \), then after \( t = \lceil \frac{6}{\phi} \rceil \) rounds, we have the following:
\[ \Pr[\text{Vol}(U_t) < \frac{\text{Vol}(U_0)}{2}] \geq \frac{1}{2}. \]

**Proof:** We only prove part a. The second part has a similar argument. Assume that at round \( i \), \( \text{Vol}(S_{i-1}) \leq |E| \), then from the definition of graph conductance, we have:
\[ |E(S_{i-1}, U_{i-1})| \geq \phi \cdot \text{Vol}(S_{i-1}) \geq \lceil \phi \cdot \text{Vol}(S_0) \rceil \]
Let \( M = \lceil \phi \cdot \text{Vol}(S_0) \rceil \), and let \( E_i \) be an arbitrary subset of \( |E(S_{i-1}, U_{i-1})| \) of size \( M \). Then for any node \( u \) in \( U_{i-1} \), we define \( g_i(u) \) be the number of edges \( u \) has in \( E_i \), and we define random variable \( L_i(u) \) as following:
\[ L_i(u) = \begin{cases} 1 & \text{if } u \text{ pulls using an edge in } E_i \\ 0 & \text{otherwise} \end{cases} \]
Then, we define random variable \( L_i = \sum_{u \in U_{i-1}} d_u \cdot L_i(u) \). Then, it is easy to see that:
\[ \text{Vol}(S_i) \geq \text{Vol}(S_{i-1}) + L_i \]
Now, we want to see how the \( \text{Vol}(S_i) \) grows from round to round. Assuming all random choices until now are fixed, we want to find the expected value of \( L_i \), using linearity of expectation we have:
\[ E[L_i] = \sum_{u \in U_{i-1}} d_u \cdot E[L_i(u)] = \sum_{u \in U_{i-1}} d_u \cdot \frac{g_i(u)}{d_u} \]
\[ = \sum_{u \in U_{i-1}} g_i(u) = M \]
As we see at each round the size of $S$, set of informed nodes, grows by $M$ in expectation.

Let’s define $L_i = L_1 + L_2 + \ldots + L_i$, then using linearity of expectation, we get $E[L_i] = iM$.

For $t = \frac{\text{Vol}(S_0)}{M}$ rounds, from $|S_0|$ we reach to $S_t$ with size at least $2|S_0|$ in exception, each round increases by $M$. Since $M$ is $\lceil \phi \cdot \text{Vol}(S_0) \rceil$ (from its definition), $t \approx \frac{\text{Vol}(S_0)}{\phi \cdot \text{Vol}(S_0)} \approx \frac{1}{\phi}$. Note that so far we showed size of $S_t \geq 2|S_0|$, if its size increases by $M$ at every round which is the expectation.

Now we use Chebyshev’s bound to show that for $t = \lceil \frac{4}{\phi} \rceil$, the probability that $\text{Vol}(S_t) < 2\text{Vol}(S_0)$ is less than $\frac{1}{2}$.

**Remark 2. (Chebyshev’s inequality)** For random variable $X$, with expected value $\mu$, we have the following:

$$\Pr[|X - \mu| > \delta] = \Pr[(X - \mu)^2 > \delta^2] \leq \frac{E[(X - \mu)^2]}{\delta^2} = \frac{\text{Var}(X)}{\delta^2}$$

Note that Chebyshev’s inequality is just Markov’s inequality on second moment of $X$.

Before applying Chebyshev’s inequality, we need to figure out $\text{Var}(L_i)$. We know $L_i = \sum_{j \leq i} L_j$. Although $L_j$s are not independent, it can be proved that $\text{Var}(L_i) = \sum_{j \leq i} \text{Var}(L_j)$ [1]. For $\text{Var}(L_j)$ we have:

$$\text{Var}(L_j) = E[((\sum_{u \in U_{j-1}} L_j(u) \cdot d_u) - M)^2]$$

$$= (\sum_{u \in U_{j-1}} E[L_j(u)^2 \cdot d_u^2]) - M^2$$

$$\leq \sum_{u \in U_{j-1}} E[L_j(u) \cdot d_u^2]$$

$$\leq \sum_{u \in U_{j-1}} \Delta \cdot E[L_j(u) \cdot d_u] = \Delta \cdot M$$

Thus, it implies that $\text{Var}(L_i) = i\Delta M$.

For $t = \lceil \frac{4}{\phi} \rceil$, we have:

$$\Pr[\text{Vol}(S_t) < 2\text{Vol}(S_0)] = \Pr[\text{Vol}(S_t) - \text{Vol}(S_0) < \text{Vol}(S_0)]$$

$$\leq \Pr[L_t < \text{Vol}(S_0)]$$

$$\leq \Pr[|L_t - \text{Vol}(S_0)| > tM - \text{Vol}(S_0)]$$

$$\leq \Pr[|L_t - \text{Vol}(S_0)| > 3\text{Vol}(S_0)]$$

(Apply Chebyshev’s inequality here)

$$\leq \frac{tM \Delta}{9\text{Vol}(S_0)^2} = \frac{4 \cdot \phi \cdot \text{Vol}(S_0) \Delta}{9 \cdot \phi \cdot \text{Vol}(S_0)^2}$$

(From assumption in part a of lemma 3, $\text{Vol}(S_0) \geq \Delta$)

$$\leq \frac{4}{9}.$$
This completes the proof of lemma 3 part a.

Now we are ready to prove lemma 1.

**Proof of Lemma 1:** Using lemma 3, after each $\frac{4}{\phi}$ rounds $\text{Vol}(S_t)$ with at least half probability becomes twice. The question is after how many rounds the volume of informed nodes becomes greater than $|E|$. The expected number of rounds is $\log \left( \frac{|E|}{\text{Vol}(S_0)} \right) \cdot \frac{4}{\phi}$. But this is not enough, and we need to bound the probability that it takes a longer time for volume of informed nodes to hit $|E|$.

Here we abstract our problem with coin toss problem with same parameter. Assume that we have a coin that comes head with probability at least $\frac{1}{2}$. How many times should we toss the coin to get at least $2 \log n$ heads. If we toss it $8 \log n$ times, what is the following probability:

$$\Pr[\text{# heads} \leq 2 \log n] = ?$$

We define $X_i$ be the random variable for the outcome of $i$th coin toss:

$$X_i = \begin{cases} 1 & \text{if coin turns up head} \\ 0 & \text{otherwise} \end{cases}$$

And we define $X = \sum_{i=1}^{n} X_i$. Then we have $\mu = E[X] = 4 \log n$. Using standard Chernoff bound:

$$\Pr[X \leq 2 \log n] \leq e^{-\frac{1}{2} \cdot \frac{4 \log n}{n}} = \frac{1}{\sqrt{n}}$$

This implies that $O(\log \left( \frac{|E|}{\text{Vol}(S)} \right) \cdot \frac{4}{\phi}) = O(\frac{\log n}{\phi})$ rounds suffices for Volume of $S$ to hit $|E|$ whp. Using similar argument, $O(\frac{\log n}{\phi})$ rounds suffices for volume of $U$ to get to 0 whp. This completes the proof for lemma 1.

References