Notes on Gossip Dissemination and Averaging

Prepared by Dr. Kalpakis for CMSC 621, Fall 2012. The notes are based on Devavrat Shah’s “Gossip Algorithms” monograph.

0.1 Conductance

Let \( P = [p_{ij}] \) be the probability transition matrix of Markov chain, so that \( p_{ij} \) is the probability that node \( i \) contacts node \( j \). We assume that \( P \) is aperiodic, irreducible. It may or may not be symmetric.

\( P \) is double–stochastic matrix, i.e. the sum of elements in each row or column is equal to 1.

By the Perron–Frobenius theorem, \( P \) has a stationary distribution \( \pi = [\pi_i] \)

\[
\pi^T = \pi^T P
\]

where each \( \pi_i \) positive, and \( \sum_{i=1}^{n} \pi_i = 1 \). Moreover, the rows of \( P^k \) all tend to \( \pi^T \) as \( k \to \infty \). Moreover, \( P \) has \( n \) real eigenvalues \( \lambda_1 = \lambda > \lambda_2 \geq \ldots \geq \lambda_n \geq -1 \).

The conductance of \( P \) is defined as

\[
\Phi(P) = \min_{S \subset V} \frac{\sum_{i \in S,j \in \overline{S}} p_{ij}}{\pi(S)\pi(\overline{S})}
\]

where \( \pi(S) = \sum_{i \in S} \pi_i \). When \( P \) is symmetric, its conductance is defined as

\[
\Phi(P) = \min_{S \subset V, |S| \leq n/2} \frac{\sum_{i \in S,j \in \overline{S}} p_{ij}}{|S|}
\]

\((\pi_i = 1/n)\).

0.2 Gossip dissemination of a secret

We indicate a set of nodes by \( S \); we also use \( S \) to indicate the cardinality of \( S \). The complement of \( S \) is denoted by \( \overline{S} \).

Let \( S_t(S_t) \) be the number (set) of infected nodes at time \( t \).

Let \( P = [p_{ij}] \) be the probability transition matrix of an irreducible aperiodic and symmetric Markov chain defined on the network, so that \( p_{ij} \) is the probability that node \( i \) contacts node \( j \). We assume that \( p_{ij} = 0 \) iff \( i \) and \( j \) are not connected.

Push phase

In the push phase an “infected” node \( i \) conducts a node \( j \) with probability \( p_{ij} \), and it infects \( j \) if it has not been infected already.

In the push phase we assume that \( S_t \leq n/2 \).

Let \( X_j(t) \) be an 0–1 indicator random variable so that \( 1 \) indicates that node \( j \) got infected at time \( t \). Note that for any 0–1 random variable \( X \), we always have \( E[X] = Pr[X = 1] \).
We compute the expected value of $X_j(t)$ conditioned on the fact that the set of infected nodes is $S_t$.

$$E[X_j(t) | S_t] = 1 - Pr[X_j(t) = 0 | S_t]$$

(4)

Node $j$ remains non-infected at time $t$ if none of the infected nodes contact it, which happens with probability

$$\Pi_{i \in S_t} (1 - p_{ij}).$$

(5)

Therefore,

$$E[X_j(t) | S_t] = 1 - \Pi_{i \in S_t} (1 - p_{ij}).$$

(6)

Since $e^{-x} \geq 1 - x$ for $x \geq 0$, it follows that

$$E[X_j(t) | S_t] \geq 1 - \Pi_{i \in S_t} e^{-p_{ij}} = 1 - e^{-\sum_{i \in S_t} p_{ij}}.$$  

(7)

Since $e^{-x} \leq 1 - x/2$ for all $0 \leq x \leq 1$, and $\sum_{i \in S_t} p_{ij}$ is between 0 and 1, it follows that

$$E[X_j(t) | S_t] \geq \frac{1}{2} \sum_{i \in S_t} p_{ij}.$$  

(8)

In other words, node $j$ is becomes infected at time $t$ with probability at least $1/2$ the total probability that can flow into it from all the already infected nodes.

Since the number of newly infected nodes that time $t$, is given by the random variable

$$\sum_{j \in S_t} X_j(t),$$

it follows that

$$E[S_{t+1} | S_t] = S_t + \sum_{j \not\in S_t} E[X_j(t) | S_t] \geq S_t + \frac{1}{2} \sum_{i \in S_t} \sum_{j \in S_t} p_{ij}.$$  

(10)

By multiplying and dividing the 2nd term in the right hand side by $S_t$, we get

$$E[S_{t+1} | S_t] = S_t + \sum_{j \not\in S_t} E[X_j(t) | S_t] \geq S_t + \frac{S_t}{2} \sum_{i \in S_t} \sum_{j \in S_t} p_{ij}.$$  

(11)

Since the last factor above is bounded from below by the conductance $\Phi(P)$, we have that

$$E[S_{t+1} | S_t] = S_t + \sum_{j \not\in S_t} E[X_j(t) | S_t] \geq S_t + \frac{S_t}{2} \Phi(P) = \left(1 + \frac{\Phi(P)}{2}\right) S_t.$$  

(12)

Note that by definition of submartingales, it then follows $S_t$ is a submartingale.

By iterating, it follows that

$$E[S_t | S_0] \geq \left(1 + \frac{\Phi(P)}{2}\right)^t S_0.$$  

(13)
Assuming \( S_0 = 1 \) and taking \( t = \log(n/2)/\log(1 + \Phi(P)/2) \),

\[
E[S_t|S_0] \geq n/2.
\] (14)

In other words, the expected number of infected nodes increases by at least a factor of \( 1 + \Phi(P)/2 \) at each time step, as long as \( S_t \leq n/2 \).

We are interested in the number of steps that are sufficient to ensure that \( S_t \) becomes larger than \( n/2 \) with high probability. Since Markov’s inequality is of the form \( E[X > a] \leq E[X]/a \), we now concentrate on \( 1/S_t \).

Using the convexity of the function \( f(x) = 1/x \) for \( x > 0 \), we have that \( f(x_2) \geq f(x_1) + f'(x_1)(x_2 - x_1) \), which for \( x_2 = S_t \) and \( x_1 = S_{t+1} \), implies that

\[
\frac{1}{S_t} \geq \frac{1}{S_{t+1}} - \frac{1}{S_{t+1}^2}(S_t - S_{t+1}),
\] (15)

or equivalently

\[
\frac{1}{S_{t+1}} \leq \frac{1}{S_t} - \frac{1}{S_{t+1}^2}(S_{t+1} - S_t),
\] (16)

Since \( S_{t+1} \leq 2S_t \), it follows that

\[
\frac{1}{S_{t+1}} \leq \frac{1}{S_t} - \frac{1}{4S_t^2}(S_{t+1} - S_t) = \frac{1}{S_t} \left( 1 - \frac{S_{t+1} - S_t}{4S_t} \right)
\] (17)

Using the inequality Eq.(12) and taking the expectation of \( 1/S_{t+1} \) conditioned on \( S_t \), we have

\[
E \left[ \frac{1}{S_{t+1}} \big| S_t \right] \leq \frac{1}{S_t} \left( 1 - \frac{\Phi(P)}{8} \right).
\] (18)

Therefore, starting with \( S_0 = 1 \), we have that

\[
E \left[ \frac{1}{S_t} \right] \leq \frac{1}{S_0} \left( 1 - \frac{\Phi(P)}{8} \right)^t \leq e^{-t\Phi(P)/8}.
\] (19)

Applying Markov’s inequality to \( 1/S_t \) we have

\[
Pr \left[ \frac{1}{S_t} > \frac{2}{n} \right] \leq \frac{ne^{-t\Phi(P)/8}}{2}.
\] (20)

Therefore, for \( t \geq 8 \log(n/\epsilon)/\Phi(P) \),

\[
Pr[S_t < n/2] \leq \epsilon.
\] (21)

Therefore, the number of infected nodes becomes at least \( n/2 \) with probability at least \( 1 - \epsilon \) in at most \( 8 \log(n/\epsilon)/\Phi(P) \) steps of pure push–gossip.
Pull phase

In the pull phase, a non–infected node \( i \) contacts a node \( j \) with probability \( p_{ij} \), and if \( j \) is infected then \( i \) becomes infected as well.

Consider the non–infected nodes \( S_t \). Assume that \( S_t \leq n/2 \). Note that if the pull phase follows the push phase this is trivially true.

As before, let \( X_j(t) \) be an 0–1 indicator random variable so that 1 indicates that node \( j \) got infected at time \( t \).

First, let’s compute the expected value for \( X_j(t) \) conditioned on \( S_t \). Node \( j \) contacts only one node at each time, and it becomes infected if it contacts an already infected node. The probability it becomes infected is equal to the sum of the probabilities it contacts any of the already infected nodes. Therefore,

\[
E[X_j(t)|S_t] = \sum_{i \in S_t} p_{ji}.
\]

(22)

Since the number of nodes infected at time \( t \) is equal to

\[
\sum_{j \in S_t} X_j(t),
\]

(23)

it follows that

\[
E[S_{t+1}|S_t] = S_t - \sum_{j \in S_t} E[X_j(t)|S_t].
\]

(24)

Substituting, we get

\[
E[S_{t+1}|S_t] = S_t - \sum_{j \in S_t} \sum_{i \in S_t} p_{ji} = S_t \left( 1 - \frac{\sum_{j \in S_t} \sum_{i \in S_t} p_{ji}}{S_t} \right).
\]

(25)

Since the 2nd term in the parenthesis is bounded from below by \( \Phi(P) \), provided that \( S_t \) is at most \( n/2 \), it follows that

\[
E[S_{t+1}|S_t] \leq (1 - \Phi(P))S_t.
\]

(26)

Note that by definition of supermartingales, it then follows \( S_t \) is a supermartingale.

In other words, the expected number of non–infected nodes decreases by at least a factor of \( (1 - \Phi(P)) \). By iterating, and assuming that \( S_0 \leq n/2 \) it follows that

\[
E[S_t|S_0] \leq (1 - \Phi(P))^t S_0 \leq e^{-t\Phi(P)}n/2.
\]

(27)

The 2nd inequality follows from \( (1 - x)^t \leq e^{-tx} \) for \( 0 \leq x \leq 1 \).

Markov’s inequality states that

\[
Pr[|X| \geq a] \leq E[X]/a
\]

(28)

for any random variable, and constant \( a > 0 \).
Applying Markov’s inequality to $S_t$ and $a = 1$, we have

$$Pr[S_t \geq 1] \leq e^{-t\Phi(P)n/2}. \quad (29)$$

For $t = (\log n + \log \epsilon^{-1})/\Phi(P)$, this implies that

$$Pr[S_t \geq 1] \leq \epsilon/2. \quad (30)$$

In other words, after $t = (\log n + \log \epsilon^{-1})/\Phi(P)$ time steps of pull the probability that there is at least 1 non–infected node is at most $\epsilon$, i.e. all nodes are infected with high probability.

**Putting both phases together.**

The separation of phases above is artificial and is done for the purposes of the analysis. There is no reason why a node can not do pull and push together when it contacts another node. The only effect would be to increase number of infected nodes.

Thus, within at most $9 \log(n/\epsilon)/\Phi(P)$ steps all nodes become infected with probability at least $(1 - \epsilon)^2$. Therefore in $O(\log n)$ steps push–pull gossip can disseminate a single piece of information to almost all the nodes of the system with high probability. The constant within the Big–Oh depends on the conductance $\Phi(P)$.

**0.3 Averaging via gossip**

Consider a connected network of $n$ nodes, where each node $i$ holds a value $x_i$. Let

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \quad (31)$$

The goal is for the nodes to each compute $\bar{x}$ by pairwise–averaging.

Assume an irreducible, aperiodic, and symmetric Markov chain $P = [p_{ij}]$ defined on the nodes, such that $p_{ij}$ is the contact probability between nodes $i$ and $j$. We assume that $P$ is conformant, i.e. $p_{ij} = 0$ iff $i$ and $j$ are not connected. Matrix $P$ is doubly stochastic (each one of its rows and columns sum up to 1). By the Birkhoff–Vo Neumann theorem, $P$ can be expressed as a weighted sum of up to $n^2$ permutation matrices $\Pi_k$

$$P = \sum_{k} a_k \Pi_k \quad (32)$$

where $\sum_k a_k = 1$ and $a_k > 0$.

Let $y_i(t)$ is the estimate of the average that node $i$ has at time step $t$. Initially, $y_i(0) = x_i$ for all nodes. Whenever two nodes gossip at time step $t + 1$, they both replace their own estimates as follows

$$y_{i,j}(t + 1) \leftarrow \frac{1}{2} (y_i(t) + y_j(t)) \quad (33)$$

Let $y(t)$ be the vector of node estimates.
We assume that at time step $t + 1$, we select a permutation matrix $\Pi_k$ with probability $a_k$, and have the nodes gossip according to that permutation matrix. This corresponds to

$$y(t + 1) = W_t y(t)$$  \hspace{1cm} (34)$$

where

$$W_t = \frac{1}{2} (I_n + \Pi_k)$$  \hspace{1cm} (35)$$

and $I_n$ is the $n \times n$ identify matrix.

We study the error vector

$$z_t = y(t) - x_1$$  \hspace{1cm} (36)$$

where $1$ is the vector of $n$ ones.

Since $W_t 1 = 1$, it follows that

$$z_{t+1} = W_t z_t.$$  \hspace{1cm} (37)$$

which implies that

$$\|z_{t+1}\|^2 = z_{t+1}^T z_{t+1} = z_t^T W_t^T W_t z_t.$$  \hspace{1cm} (38)$$

Then, the expectation of the norm of $z_t + 1$ conditioned on $z_t$ is

$$E[\|z_{t+1}\|^2 | z_t] = z_t^T E[W_t^T W_t] z_t.$$  \hspace{1cm} (39)$$

Observe that, since $\Pi^T \Pi = I$ for any permutation matrix $\Pi$, we have

$$W_t^T W_t = \frac{1}{2} (I_n + \Pi_k)$$  \hspace{1cm} (40)$$

Therefore,

$$E[W_t^T W_t] = \frac{1}{2} (I_n + P) = \overline{W}$$  \hspace{1cm} (41)$$

which implies that

$$E[\|z_{t+1}\|^2 | z_t] = z_t^T \overline{W} z_t.$$  \hspace{1cm} (42)$$

Note that due to “mass preservation”,

$$1^T z_t = 0.$$  \hspace{1cm} (43)$$

Using the Rayleigh quotient theorem for the 2nd largest eigenvalue of $\overline{W}$, and the fact that $1$ is an eigenvector for the largest eigenvalue of $\overline{W}$, we have that

$$\lambda_2(\overline{W}) = \max_{v: 1^T v = 0} \frac{v^T \overline{W} v}{v^T v},$$  \hspace{1cm} (44)$$

which implies that

$$E[\|z_{t+1}\|^2 | z_t] \leq \lambda_2(\overline{W}) \|z_t\|^2$$  \hspace{1cm} (45)$$
Therefore, by induction on \( t \), we have

\[
E[\|z_t\|^2] \leq \lambda_2(\mathbf{W})^t\|z_0\|^2 \tag{46}
\]

Note that \( \|z_0\|^2 \) is actually equal to the variance \( \sigma_x \) of the vector \( x = [x_i] \).

Applying Markov’s inequality to the random variable \( \|z_t\|^2/\sigma_x \), we have that

\[
Pr \left[ \frac{\|z_t\|^2}{\sigma_x} \geq \epsilon^2 \right] \leq \epsilon^{-2} \frac{E[\|z_t\|^2]}{\sigma_x} \leq \epsilon^{-2} \lambda_2(\mathbf{W}) \tag{47}
\]

In other words, the norm of the error decreases exponentially with the number of steps with high probability, so \( y_i(t) \) approaches \( \bar{x} \) at rate that is exponential in \( \lambda_2(\mathbf{W}) \).