GOSSIP model

In each round, contact a single neighbor to exchange information with

Why?

• Initiating communication may be expensive
• Reduce network traffic
Information Spreading

• Each node needs to learn the inputs of all other nodes

• Also called:
  – All-to-all Dissemination
  – Rumor Spreading
  – ...

Information Spreading

In LOCAL model takes $D$ rounds

In GOSSIP model:

- Deterministic round-robin algorithm: $\Delta \cdot D$
  
  May need to contact all neighbors before reaching bottleneck

- Randomized algorithm?

$3$ rounds in LOCAL model
$O(n)$ rounds in GOSSIP model
Randomized Information Spreading

Randomized:

• Complete graph – $O(\log n)$ rounds
Information spreading

– The problem with the barbell graph is that it has
  • Large degrees, and
  • Bad connectivity

– Large degree alone is not a problem

– Bad connectivity alone is not a problem

\( O(\log n) \) rounds

\( \Theta(n) \), but we cannot hope for anything better
Randomized information spreading

- General graphs analyzed in terms of their **conductance** $\Phi$

$0 \leq \Phi \leq 1$, measure of connectivity

$$\Phi = \min_S \varphi(S)$$

- $\varphi(S) = \text{number of edges from } S \text{ to } V \setminus S$

- $\Phi \approx 1/n^2$
Conductance - Examples

• Clique

\[ \Phi(\text{clique}) = \left( \frac{n \cdot n}{2 \cdot 2} \right) / \left( \frac{n \cdot n}{2} \right) = O(1) \]

• Path

\[ \Phi(\text{path}) = O\left( \frac{1}{n} \right) \]
Randomized information spreading

$O\left(\frac{\log(n)}{\Phi_G}\right)$ rounds for all graphs $G$

$O(n^2\log(n))$
Information Spreading - Examples

• Clique

\[ \Phi(\text{clique}) = O(1) \]

Information spreading in \( O(\log n) \) rounds

• Path

\[ \Phi(\text{path}) = O\left(\frac{1}{n}\right) \]

Information spreading in \( O(n \log n) \) rounds
Faster information spreading

• \( O\left( \frac{\log(n)}{\Phi_c} \right) \) rounds for **partial** information spreading
  
  – A weakened version of the problem

\( \Phi \leq \Phi_c : \) weak conductance, conductance of subsets

\( \Phi_c = \text{constant, } \Phi = \frac{1}{n^2} \)

improves from \( O(n^2 \log(n)) \) to \( O(\log(n)) \)
Faster information spreading

• $O\left( \frac{\log(n)}{\Phi_c} \right)$ rounds for partial information spreading
  – A weakened version of the problem

• $\min_c\{O\left( \frac{c(\log(n)+c)}{\Phi_c} \right)\}$
  rounds for full information spreading

$\Phi_c = \text{constant, } \Phi = \frac{1}{n^2}$

improves from $O(n^2\log(n))$ to $O(\log(n))$
Faster information spreading

- $O\left(\frac{\log(n)}{\Phi_c}\right)$ rounds for \textit{partial} information spreading
  - A weakened version of the problem

- $\min_c\{O\left(\frac{c(\log(n)+c)}{\Phi_c}\right)\}$ rounds for \textit{full} information spreading

- $O(D+\text{polylog}(n))$ rounds
  For any graph, regardless of degrees or connectivity

compare with $D$ for the LOCAL model
Partial Information Spreading

• Relaxed spreading requirement
  
  Given a parameter $c$:
  
  – Every node needs to receive $\frac{n}{c}$ messages
  – Every message needs to be received by $\frac{n}{c}$ nodes

• For $c=1$ we get full information spreading
Weak Conductance

- Weak conductance: measure connectivity of subsets
  - Given a parameter $c \geq 1$

$$\Phi_c(G) = \min_{i \in V} \left\{ \max_{V_i \subseteq V, i \in V_i, |V_i| \geq \frac{n}{c}} \min_{S \subseteq V_i, |S| \leq \frac{|V_i|}{2}} \phi(S, V_i) \right\}$$

Over all nodes
Over all subsets of size at least $n/c$ which include this node
Conductance of the subset

Not smaller than the conductance
Partial Information Spreading

• Analyzed according to weak conductance
• Partial information spreading is obtained in

\[ O\left(\frac{\log n + \log \delta^{-1}}{\Phi_c(G)} \right) \]

rounds, with probability \(1-\delta\)
Partial Information Spreading - Examples

• Clique
  \[ \Phi_c (\text{clique}) = O(1) \]
  Partial information spreading in \( O\left(\log n + \log \delta^{-1}\right) \) rounds

• Path
  \[ \Phi_c (\text{path}) = \frac{1}{n} \frac{n}{c} = \frac{c}{n} \]
  Partial information spreading in \( O\left(\frac{n}{c} \left(\log n + \log \delta^{-1}\right)\right) \) rounds
The c-Barbell

- c-Barbell

\[ \Phi(c - \text{barbell}) = O\left(\frac{1}{\left(\frac{n}{c}\right)^2} c \right) = O\left(\frac{c}{n^2}\right) \]

Information spreading in \( O\left(\frac{n^2}{c\left(\log n + \log \delta^{-1}\right)}\right) \) rounds

- Even worse than for the path
  - The bottleneck has a smaller probability
The c-Barbell

- c-Barbell

\[ \Phi_c(c \text{- barbell}) = O\left(\left(\frac{n}{2c}\right)^2 / \left(\frac{n}{2c}\right)^2\right) = O(1) \]

Partial information spreading in

\[ O\left(\log n + \log \delta^{-1}\right) \] rounds

- Improving from \( O(n^2 \log n) \) to \( O(\log n) \)
Interlude

• Partial information spreading is sufficient for certain applications
  – Distributed maximum coverage

• Still, we want to be able to cope with those bottlenecks
Problem

• Define a bottleneck
  – The edge connecting the two cliques in the barbell?
  – Any edge within the two cliques?

• Other cases?
A Hybrid Approach

• Deterministic or Random choices of a neighbor seem to be insufficient

• Use both
  – Even numbered round: contact a random neighbor
  – Odd numbered round: contact a neighbor deterministically
    • Go over list of neighbors
    • Crux: make list shrink as...
Faster spreading

• **Theorem 1:** There is an algorithm which obtains full information spreading in 

\[ O\left(c\left(\frac{\log n + \log \delta^{-1}}{\Phi_c(G)} + c\right)\right) \]

rounds, with probability at least \(1 - 3c\delta\)

• This is \(\text{polylog}(n)\) if \(c\) is \(\text{polylog}(n)\) and \(\Phi_c(G)\) is \(1/\text{polylog}(n)\)
The Algorithm in a Nutshell

• When $v$ first gets a message $m(u)$ it erases $u$ from the list of neighbors
  – Who are accessed in odd numbered rounds
• **Unless** the message was received from $u$ itself, when being contacted by $v$
The Algorithm in a Nutshell

• When v first gets a message $m(u)$ it erases u from the list of neighbors
  – Who are accessed in odd numbered rounds

• **Unless** the message was received from u itself, when being contacted by v
Grey-Black-White Graph

- Associate the lists with a directed graph (grey)
- Edges to neighbors removed from list turn white
- Edges to neighbors remaining forever turn black
Analysis

• **Lemma 1**: The graph without the *white* edges is always connected (weakly, not directed paths)
  – Proof by induction

• **Lemma 2**: The number of *black* edges leaving node \( v \) is bounded by the number of rounds
  – Since \( v \) has to contact that neighbor
Analysis

• Let $T = O\left(\frac{\log n + \log \delta^{-1}}{\Phi_c(G)}\right)$ be the number of rounds required for $v$ to get all messages in $V_v$ by random choices.

• **Claim:** After $2T$ rounds, $v$ has $m(u)$ of every $u \in V_v$
  - With probability $1-\delta$
The set of nodes $A_v$

- Nodes with intersecting sets:
  \[ I_v = \{ u \in V \mid V_v \cap V_u \neq \emptyset \} \]

- Union of all these sets:
  \[ A_v = V_v \cup \bigcup_{u \in I_v} V_u \]
Analysis of receiving $A_v$

- **Lemma 3**: After $6T$ rounds, $v$ has $m(x)$ for every $x \in A_v$
  - With probability $1-3\delta$

- **Proof**: By claim.
Analysis of receiving $N(v)$

- **Lemma 4**: After $v$ has $m(x)$ for every $x \in A_v$, it takes an additional $2c$ rounds to get $m(y)$ for every $y \in N(v)$
2c additional rounds

• **Proof:** Induction argument:

After 2i rounds (if not done) we have nodes \( u_1, \ldots, u_i \)

- For every \( 1 \leq j \neq k \leq i \), \( V u_j \cap V u_k = \emptyset \)
- For every \( 1 \leq j \leq i \), \( v \) receives \( m(x) \) for every \( x \in A u_j \)

because (if not done):

- \( v \) contacts a grey neighbor \( u \)
- If there was \( 1 \leq j \leq i-1 \) s.t. \( V u_j \cap V u \neq \emptyset \) then \( u \in A u_j \)
- But then \( u \) was not grey (Lemma 3)
2c

So for $i=c$ we’re done because we have $c$ disjoint sets of size $n/c$

• **Proof:** Induction argument:

After $2i$ rounds (if not done) we have nodes $u_1,...,u_i$

– For every $1 \leq j \neq k \leq i$, $V_{u_j} \cap V_{u_k} = \emptyset$

– For every $1 \leq j \leq i$, $v$ receives $m(x)$ for every $x \in A_{u_j}$

because (if not done):

– $v$ contacts a grey neighbor $u$

– If there was $1 \leq j \leq i-1$ s.t. $V_{u_j} \cap V_u \neq \emptyset$ then $u \in A_{u_j}$

– But then $u$ was not grey (Lemma 3)
Rest of nodes?

• Now $v$ needs to receive $m(x)$ for rest of the nodes $x$
  – No more grey edges
  – Number of black edges is at most $6T+2c$ (Lemma 2)
  – Black sub-graph is connected (Lemma 1)

• Same idea as proof of Lemma 4:
  Someone is connected to $x$ with a black edge, will get $m(x)$ after at most $6T+2c$ rounds
• Now $v$ needs to receive $m(x)$ for rest of the nodes $x$
  – No more grey edges
  – Number of black edges is at most $6T+2c$ (Lemma 2)
  – Black sub-graph is connected (Lemma 1)

• Same idea as proof of Lemma 4:
Someone is connected to $x$ with a black edge, will get $m(x)$ after at most $6T+2c$ rounds
Wrap up

• Hybrid algorithm obtaining full information spreading in

\[ O\left(c\left(\frac{\log n + \log \delta^{-1}}{\Phi_c(G)} + c\right)\right) \]

rounds, with probability at least \(1 - 3c\delta\)
Simulating LOCAL with GOSSIP

• Can simulate 1 round of LOCAL with $\Delta$ rounds of GOSSIP

• Suppose we have a $k$-spanner $S$?
• Can simulate 1 round of LOCAL with $\Delta_Sk$ rounds of GOSSIP
  – By going over neighbors in $S$ for $k$ iterations
  – Every neighbor in $G$ is in distance at most $k$ in $S$
Wish upon a star

• For some graphs, a spanner with \( \Delta_s < \Delta \) does not exist
  – For any k...

• Instead, we use a spanner that is “sparse enough”
  – Takes \( O(1) \) rounds to reach the next hop
Hereditary density $\delta$

- The hereditary density of $S$ is $\delta$, if the maximal density of any induced subgraph is $\delta$
  - any induced subgraph has at most $\delta |U|$ edges

\[ \text{density}(U) = 1 \]
Sparse Spanner

• Still many degrees in $S$ can be large
• But the edges can be oriented such that the out-degrees are at most $O(\delta)$

• **Theorem 2**: Let $S$ be a
  – $k$-spanner
  – with hereditary density $\delta$, 
  – whose edges are oriented with out-degrees $O(\delta)$.

Then any $T$-round algorithm in the LOCAL model can be simulated in $O(\delta k T)$ rounds in the GOSSIP model.
Sparse Spanner

- **Theorem 3**: Let $S$ be an
  - $(\alpha, \beta)$-spanner
  - with hereditary density $\delta$,
  - whose edges are oriented with out-degrees $O(\delta)$.

Then any $T$-round algorithm in the LOCAL model can be simulated in $O(\delta \alpha T + \delta \beta)$ rounds in the GOSSIP model.
Sparse Spanner

• **Facts:**
  – For every $G$ there is an $(O(1), \text{polylog}(n))$-spanner $S$ with hereditary density $\delta = O(1)$
  – $S$ can be constructed in the LOCAL model in $\text{polylog}(n)$ rounds
  – The edges of $S$ can be oriented to have out-degree $O(\delta) = O(1)$ within $O(\delta \log n) = O(\log n)$ rounds in the GOSSIP model

• **Corollary:** Any $T$-round algorithm in the LOCAL model can be simulated in $O(T + \text{polylog}(n))$ rounds in the GOSSIP model.
A simulation implies a spanner

• **Theorem 4**: A simulation with an overhead of $\alpha$ implies an $\alpha$-spanner with hereditary density $\alpha$, constructed in $\alpha$ rounds the GOSSIP model

• **Proof**: The number of rounds used to simulate a single LOCAL round is a bound for:
  – the number of outgoing edges (density)
  – and the distance to neighbors (stretch)
  – Number of messages is $O(n\alpha)$
  (compare to $O(n^2)$ in LOCAL)
Have a great SUMMER!