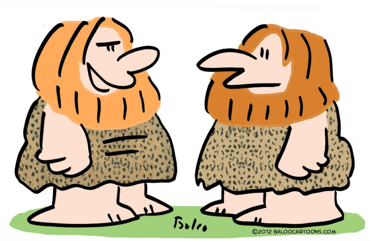
Gossip Algorithms



"Now that we've learned to talk, let's start some *rumors!*"

He Sun
School of Informatics
University of Edinburgh

THE UNIVERSITY *of* EDINBURGH

What is Gossip?

Gossip algorithms

- In a gossip algorithm, each node in the network ***periodically*** exchanges information with a subset of nodes
- This subset is usually ***the set of neighbors of each node***

Every node only has a local view of the network

- **Objective:** each node receives some desired ***global information***, through a certain number of periodically update of the nodes.

the same msg,
value of a function,
……

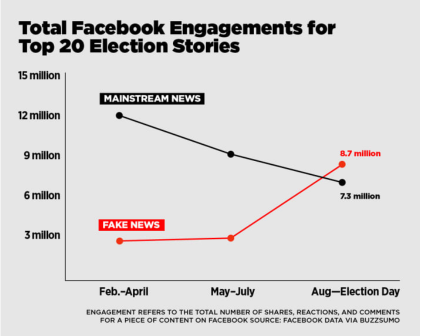
Background

Technological: Gossip algorithms are widely used in communication networks which, more and more, are likely to exhibit a social dimension. This knowledge might be exploited for more efficient communication protocols.

Application: *analysis of community structure/computer virus, help us to build better networks, ……*

Sociological: Gossip is a basic, simple form of a contagion dynamics. By studying it we hope to gain some insight into more complex diffusion phenomena.

Application: *analysis of spread of virus/fake news in an election, … …*



THE UNIVERSITY of EDINBURGH

Rumor spreading

Problem

Design an algorithm so that all the nodes receive the rumor as fast as possible.

Solution 1 Initial node sends the rumor to one of its neighbours, and every informed node forwards it to all its neighbours.

- Downside 1: every node needs to interact with all its neighbours.
- Downside 2: every node receives its degree copies of the rumor.

High communication cost

Solution 2 Construct a spanning tree, and transfer the rumor only along the edges of the tree.

- Downside: Failure of links in the tree breaks rumor spreading process.

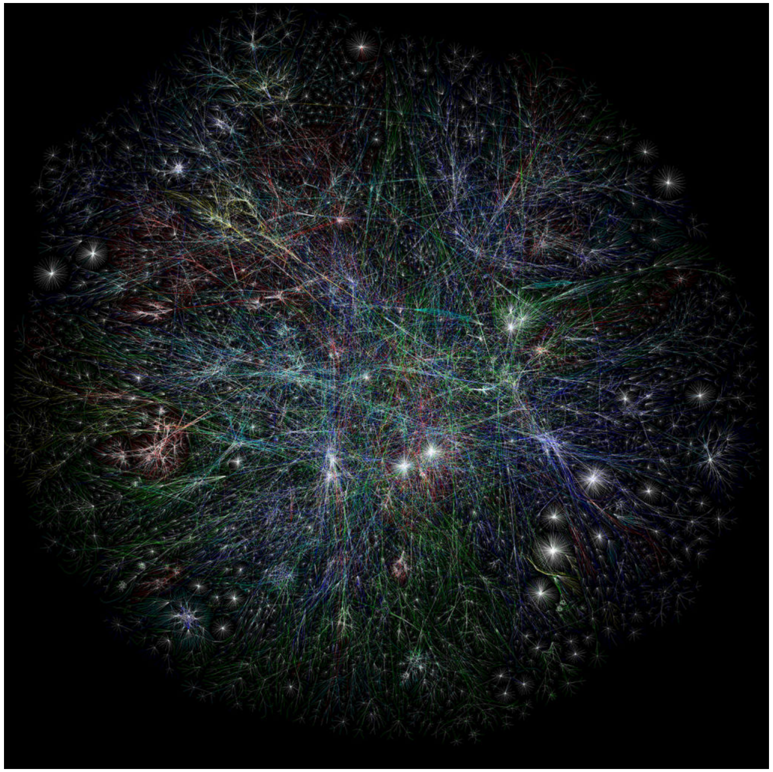
nodes only talk to their neighbours

without coordination among nodes

We need a simple, *local*, *distributed*, *fast*, and *robust* algorithm for information spreading.

tolerate certain link failures

Push protocol of rumor spreading



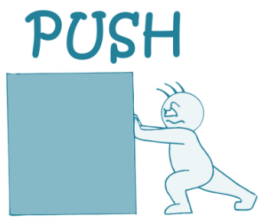
Protocol (Synchronous model)

- There is a rumor *initially* located at a node of a network;
- The protocol proceeds by **rounds**, in which each node only **contacts one of its neighbours**.

Push protocol of rumor spreading

PUSH

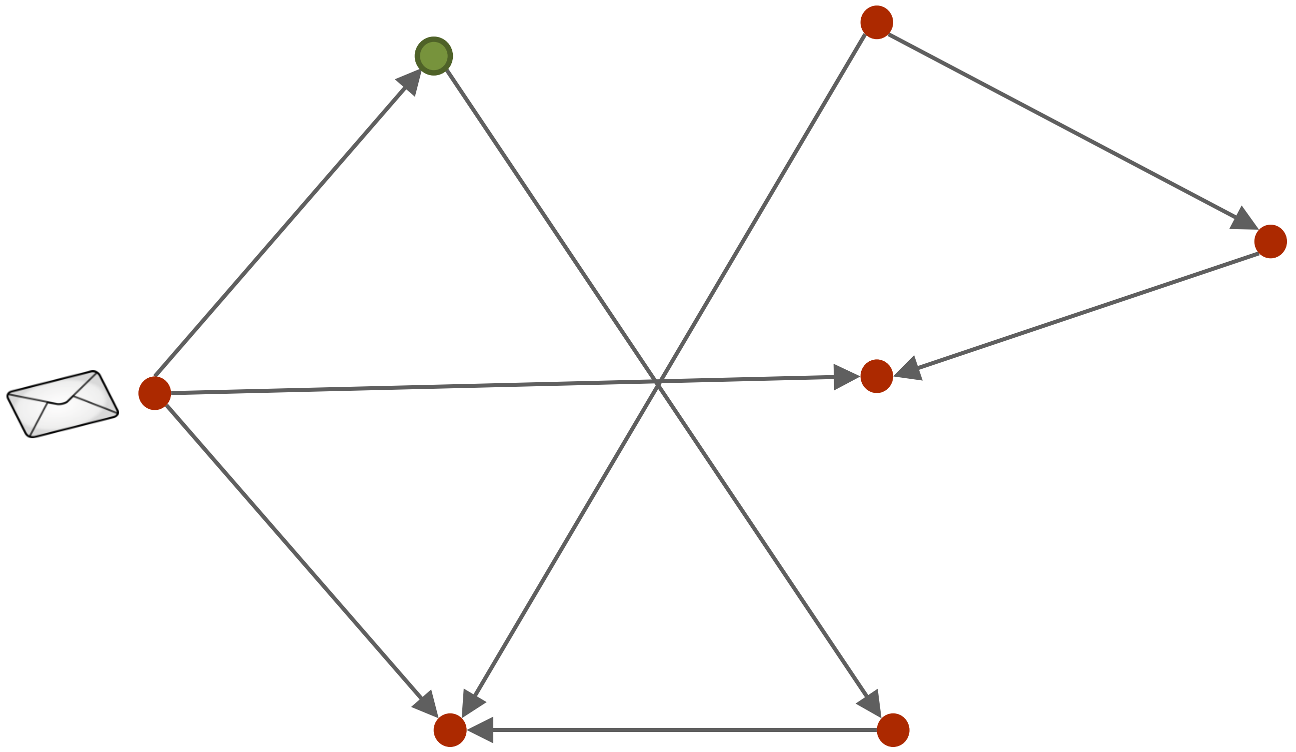
Nodes with rumor sends it to a random neighbour



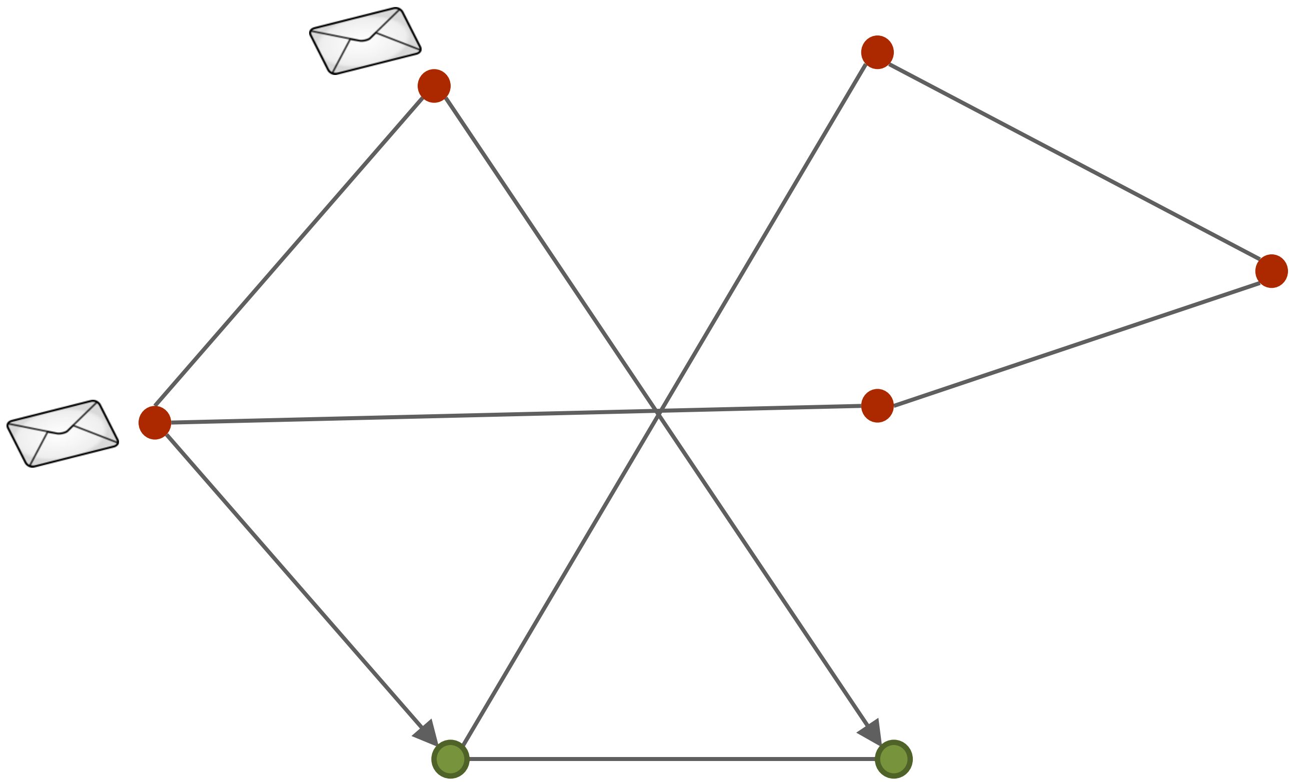
Algorithm Description

1. t=0
2. while **t<T** do
3. every informed node sends the rumor to its random neighbour.
4. t=t+1

Push protocol of rumor spreading

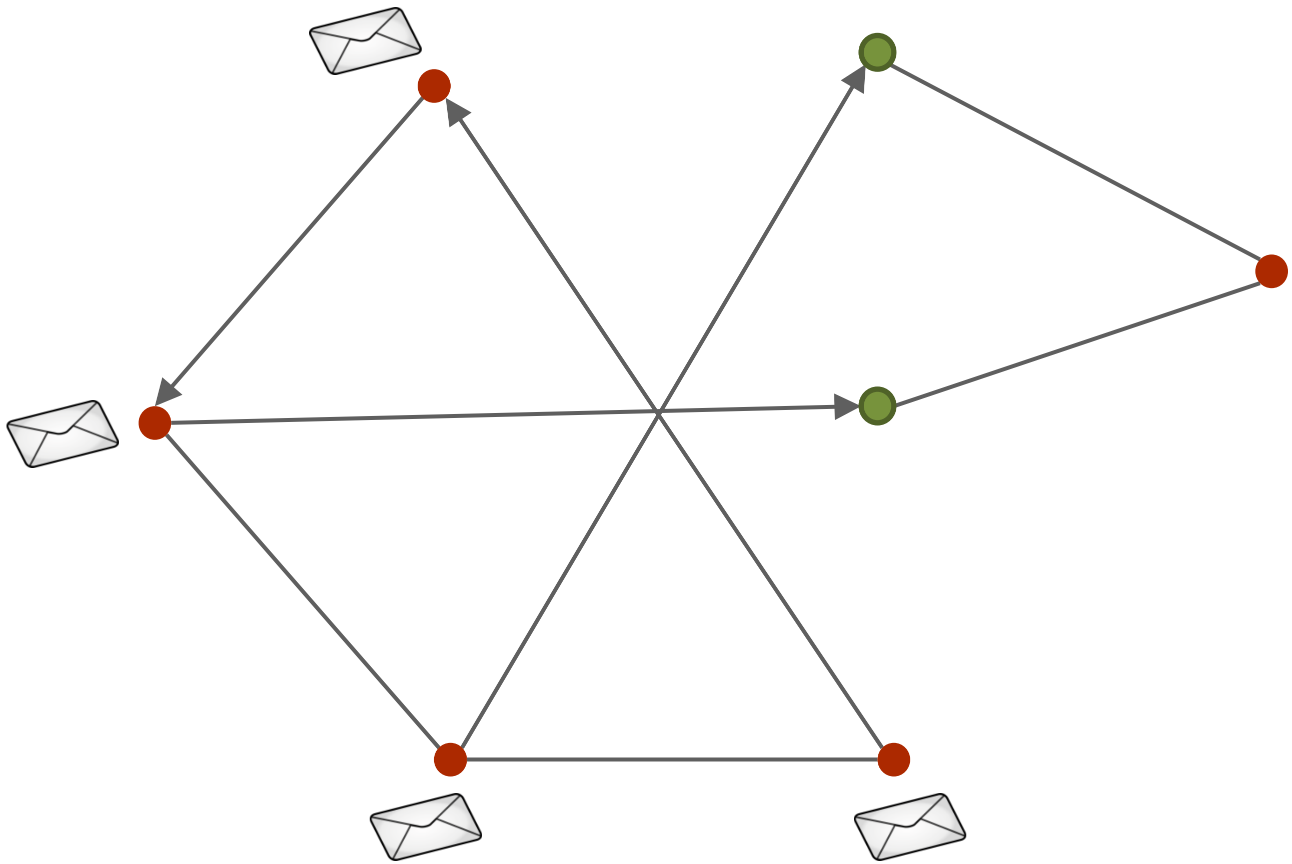


Push protocol of rumor spreading

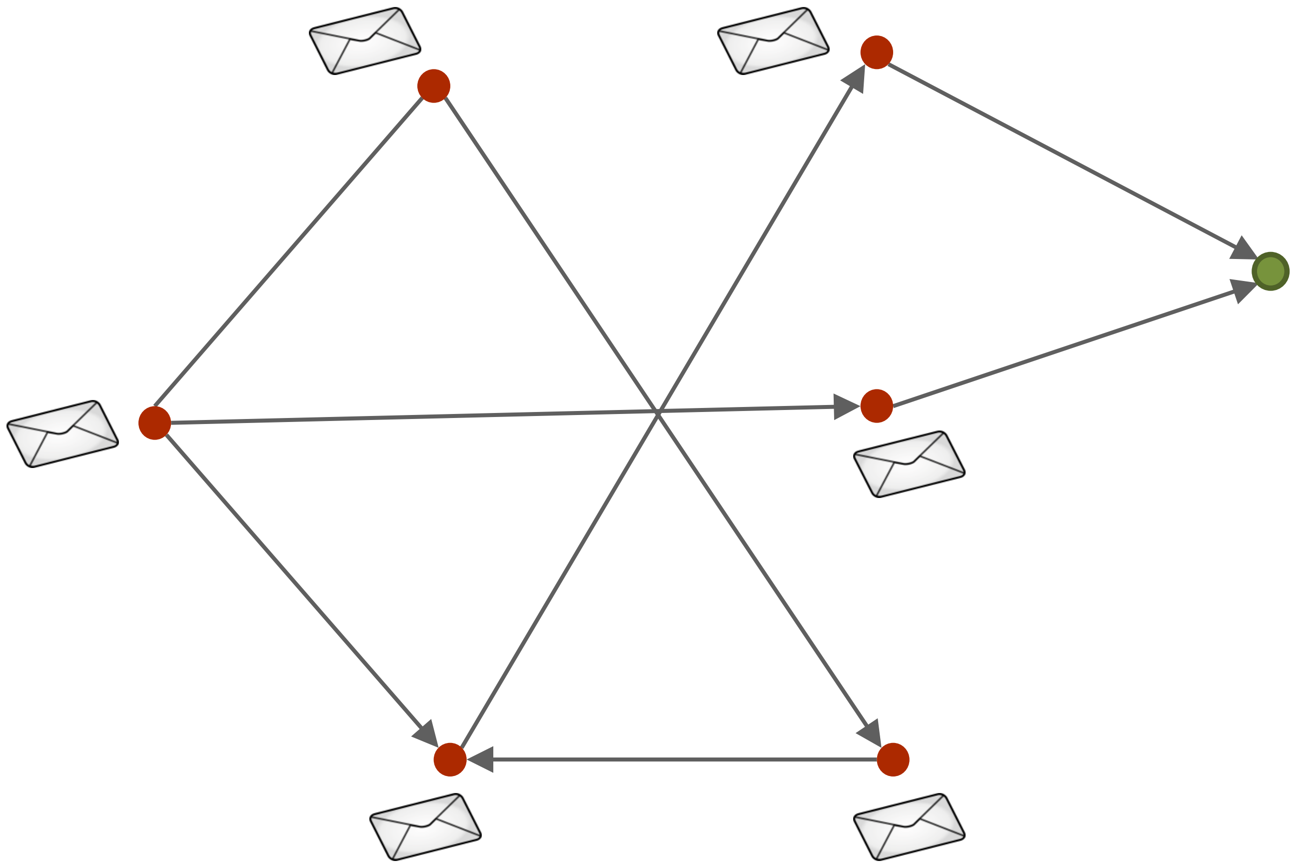


THE UNIVERSITY *of* EDINBURGH

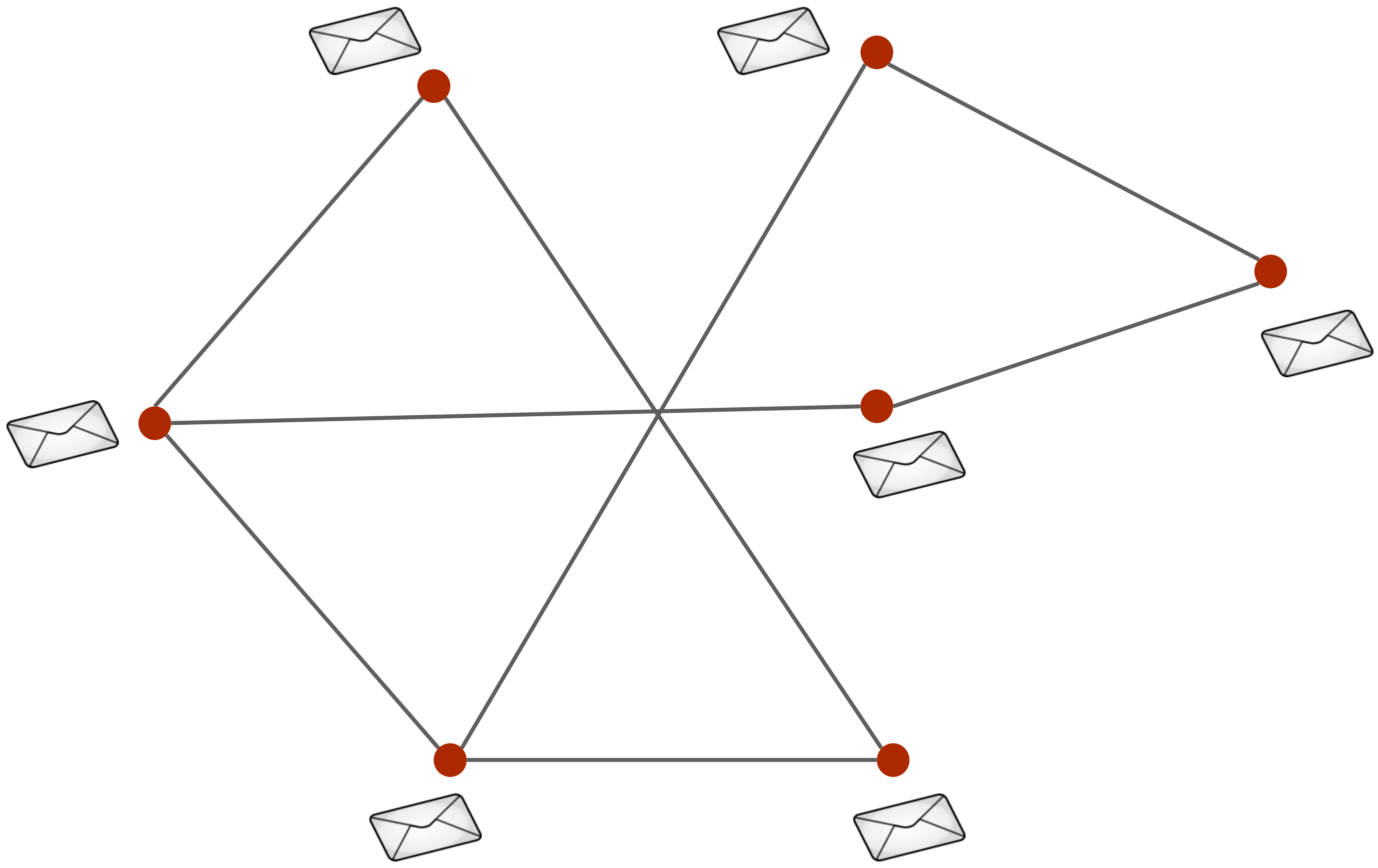
Push protocol of rumor spreading



Push protocol of rumor spreading



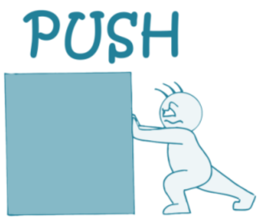
Push protocol of rumor spreading



Push protocol of rumor spreading

PUSH

Nodes with rumor sends to a random neighbour



Algorithm Description

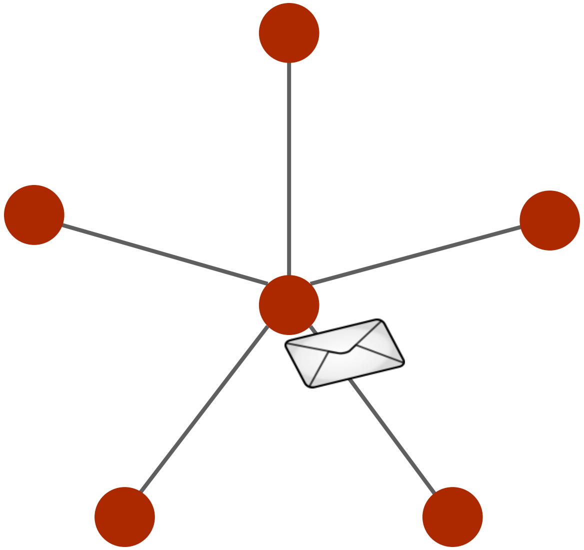
1. t=0
2. while **t<T** do
3. every informed node sends the rumor to its random neighbour.
4. t=t+1

Properties:

- Nodes only contact with their neighbours; network's global structure is unknown to each node.
- **Robust**: Failure of transmission among a few nodes won't affect the algorithm's performance.
- The algorithm **efficiently** sends a rumor to all nodes in the network.

Randomisation is the key to ensure robustness and efficiency!

Bad instance for the Push protocol

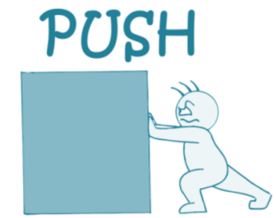


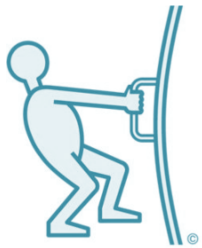
Homework: It takes $O(n \cdot \log n)$ rounds for all nodes to receive the rumor w.h.p.

Push-Pull Protocol

PUSH

Nodes with rumor sends to a random neighbour

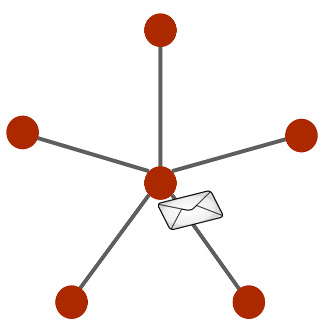




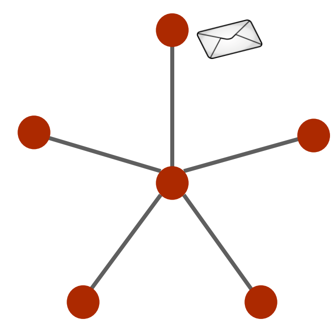
PULL

Nodes without rumor asks a random neighbour

Bad instance for PUSH



Bad instance for PULL



Push-Pull Protocol

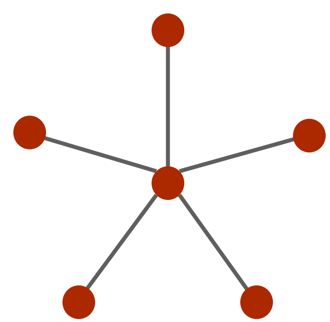
Algorithm Description

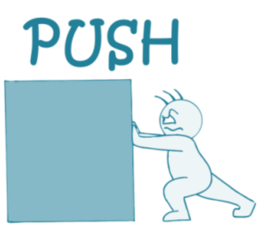
1. t=0
2. while **$t<T$** do

3-1. every informed node sends the rumor to its random neighbour.

3-2. *every uninformed node calls a random neighbour, and gets the rumor if the neighbour has one.*

4. t=t+1
5. end





Push versus Pull

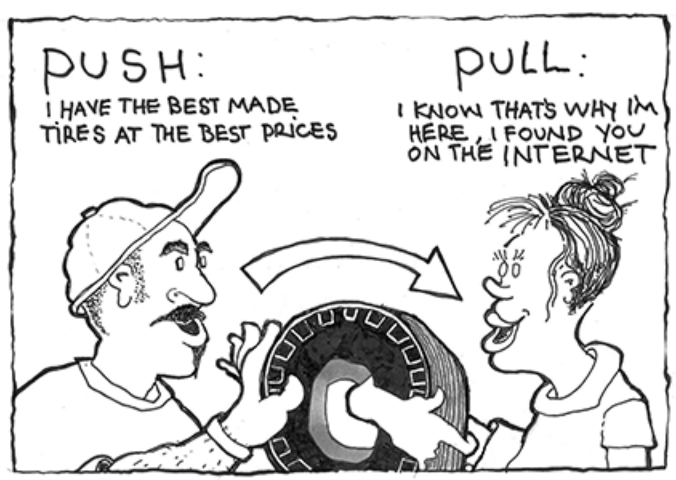
PUSH

PULL

Rumor spreads fast in social networks!

Cookie-based Advertising

Google Advertisement



Analysis of the Push protocol

Question

How many rounds are needed before every node gets the rumor w.h.p.?

Properties:

- $\Omega(\mathrm{Diam}(G))$ rounds are needed before every node gets the rumor.
- $\Omega(\log n)$ rounds are needed before every node gets the rumor.

Since the number of informed vertices at most doubles after each round.

Theorem

Let G be a complete graph with n nodes. Then, with high probability, every nodes gets the rumor after $\log n + \ln n + o(\log n)$ rounds.

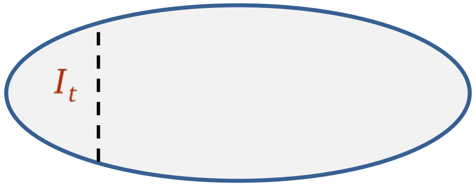
Proof sketch

Theorem

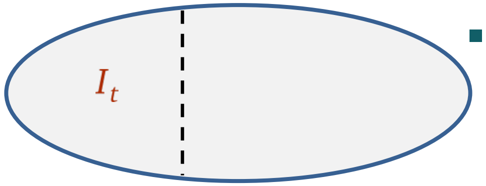
Let G be a complete graph with n nodes. Then, with high probability, every nodes gets the rumor after $\log n + \ln n + o(\log n)$ rounds.

Let I_t be the set of informed nodes in the end of round t, and U_t be the set of non-informed nodes in the end of round t. We divide the analysis into three phases:

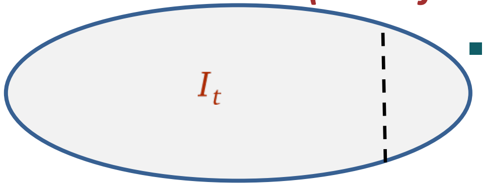
- ***(# informed nodes) is small***
  - *Since $|I_t|$ is small, there is a good chance that different informed nodes choose different non-informed nodes , in which case the number of informed nodes almost doubles after each round.*



- ***(# informed nodes)* ≈ (# uninformed nodes)**
  - *There are already a lot of informed nodes , and hence the rate of number of informed nodes becomes increasing slowly.*



- ***(#uninformed nodes) is small***
  - *There are few non-informed nodes, and the number of non-informed vertices decreases exponentially.*



Analysis on complete graphs

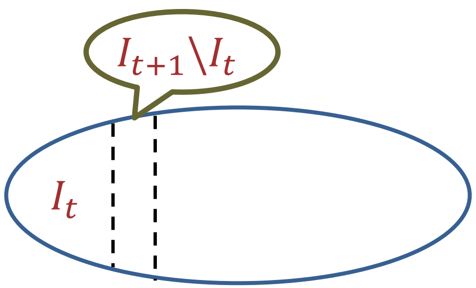
Proof: Let I_t be the set of informed nodes in the end of round t, and U_t be the set of non-informed nodes in the end of round t. We divide the analysis into three phases:

- $1 \leq |I_t| \leq (n-1)/\log n$
- $(n-1)/\log n \leq |I_t| \leq n - n/\log n$
- $n - n/\log n \leq |I_t| \leq n$

Analysis of Phase I Let t be any round with $1 \leq |I_t| \leq (n-1)/\log n$. Notice that

$$\mathbb{E}[|I_{t+1} \backslash \mathrm{I}_t|] = \sum_{u \notin I_t} \mathbb{P}[u \in I_{t+1}] = \sum_{u \notin I_t} 1 - \mathbb{P}[u \in U_{t+1}] = \sum_{u \notin I_t} 1 - \left(1 - \frac{1}{n-1}\right)^{|I_t|}$$

With the inequality $(1-x)^n \leq 1 - nx + n^2x^2$, it holds that



$$\sum_{u \notin I_t} 1 - \left(1 - \frac{1}{n-1}\right)^{|I_t|} \geq \sum_{u \notin I_t} 1 - \left(1 - \frac{|I_t|}{n-1} + \frac{|I_t|^2}{(n-1)^2}\right)$$

$$\geq \sum_{u \notin I_t} \frac{|I_t|}{n-1} \cdot \left(1 - \frac{|I_t|}{n-1}\right) \geq (n - |I_t|) \frac{|I_t|}{n-1} \left(1 - \frac{1}{\ln n}\right)$$

$$\geq \left(n - \frac{n-1}{\ln n}\right) \frac{|I_t|}{n-1} \left(1 - \frac{1}{\ln n}\right) \geq \left(1 - \frac{2}{\ln n}\right) |I_t|$$

Since $|I_{t+1} \backslash \mathrm{I}_t| \leq |I_t|$, it follows by Markov inequality that

$$\mathbb{P}[|I_t| - |I_{t+1} \backslash I_t| \geq c] \leq \frac{\mathbb{E}[|I_t| - |I_{t+1} \backslash I_t|]}{c} \leq \frac{2|I_t| / \ln n}{c}$$

Analysis on complete graphs

Proof: Analysis of Phase I (Contd) $\mathbb{P}[|I_t| - |I_{t+1} \setminus I_t| \geq c] \leq \frac{\mathbb{E}[|I_t| - |I_{t+1} \setminus I_t|]}{c} \leq \frac{2|I_t|/\ln n}{c}$

Choosing $c = 2|I_t|/\sqrt{\ln n}$ yields

$$\mathbb{P}\left[|I_t| - |I_{t+1} \setminus I_t| \geq \frac{2|I_t|}{\sqrt{\ln n}}\right] \leq \frac{1}{\sqrt{\ln n}}$$

with high prob., #informed nodes almost doubles

which is equivalent to

$$\mathbb{P}\left[|I_{t+1} \setminus I_t| \geq \left(1 - \frac{2}{\sqrt{\ln n}}\right)|I_t|\right] \geq 1 - \frac{1}{\sqrt{\ln n}}$$

We call a round ***good*** if the above happens. Notice that after

$$\log_{2-2/\sqrt{\ln n}}(n/\log n) = \log_2 n + o(\log n) \coloneqq \beta$$

good rounds, we have $|I_t| \geq n/\ln n$.

If we consider $\beta + 8\ln n/\ln\ln n$ consecutive rounds, the probability for having more than $8\ln n/\ln\ln n$ bad rounds is upper bounded by

$$\binom{\beta + 8\ln n/\ln\ln n}{8\ln n/\ln\ln n} \cdot \left(\frac{1}{\sqrt{\ln n}}\right)^{\frac{8\ln n}{\ln\ln n}} \leq 2^{2\log_2 n} \cdot n^{-4} = n^{-2}$$

Hence, with probability at least $1 - n^{-2}$, there is a round $\tau \leq \log_2 n + o(\log n)$ such that $|I_\tau| \geq n/\log n$.

Analysis on complete graphs

Proof: Let I_t be the set of informed nodes in the end of round t, and U_t be the set of non-informed nodes in the end of round t. We divide the analysis into three phases:

- $1 \leq |I_t| \leq (n-1)/\log n$
- $(n-1)/\log n \leq |I_t| \leq n - n/\log n$
- $n - n/\log n \leq |I_t| \leq n$

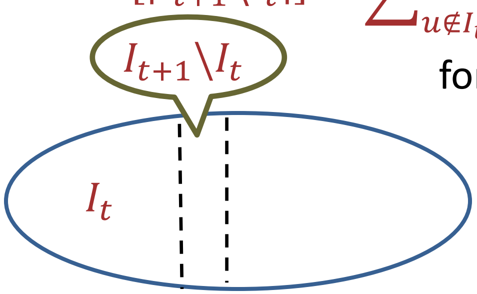
Analysis of Phase II Let t be the first round with $|I_t| \geq (n-1)/\log n$, and assume $|I_t| \leq n/2$.

As in Phase I, we have

$$\mathbb{E}[|I_{t+1} \backslash I_t|] = \sum_{u \notin I_t} \mathbb{P}[u \in I_{t+1}] = \sum_{u \notin I_t} 1 - \left(1 - \frac{1}{n-1}\right)^{|I_t|} \geq \frac{n}{2} \cdot \frac{|I_t|}{n-2} \cdot \left(1 - \frac{n/2}{n-1}\right) \geq |I_t|/3$$

for sufficiently large n.

#informed nodes grows slower!!!



Applying the same analysis as in Phase I, we have there is a constant c, s.t.

$$\mathbb{P}\left[|I_{t+1}| \geq \frac{5}{4}|I_t|\right] \geq c$$

Call a round ***good*** if $|I_{t+1}| \geq 5|I_t|/4$. Starting with $|I_t| \geq (n-1)/\log n$, we only need $O(\log\log n)$ good rounds before the number of informed nodes reaches $n/2$. Since every good round happens with constant probability, if we spend $O(\sqrt{\log n})$ rounds, then the probability for having less than $O(\log\log n)$ good rounds is $o(1)$.

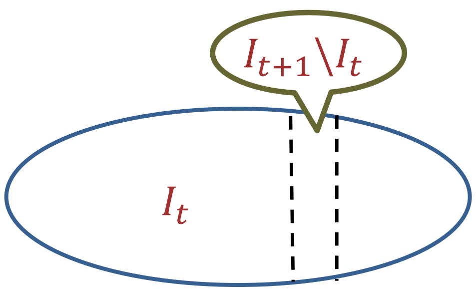
Analysis on complete graphs

Proof: Let I_t be the set of informed nodes in the end of round t, and U_t be the set of non-informed nodes in the end of round t. We divide the analysis into three phases:

- $1 \leq |I_t| \leq (n-1)/\log n$
- $(n-1)/\log n \leq |I_t| \leq n - n/\log n$
- $n - n/\log n \leq |I_t| \leq n$

Analysis of Phase II (contd.) Let t be a round with $n/2 \leq |I_t| \leq n - n/\log n$. We upper bound the expected number of ***non-informed*** nodes by

$$\mathbb{E}[|U_{t+1}|] = \sum_{u \in U_t} \mathbb{P}[u \in U_{t+1}] = \sum_{u \in U_t} \left(1 - \frac{1}{n-1}\right)^{|I_t|} \leq |U_t| \cdot \left(1 - \frac{1}{n}\right)^{\frac{n}{2}} \leq |U_t| \cdot \mathrm{e}^{-1/2}$$



Iterating this argument by for any $\tau \in \mathbb{N}$ rounds yields

$$\mathbb{E}[|U_{t+\tau}|] \leq |U_t| \cdot \mathrm{e}^{-\tau/2}$$

\# non-informed vertices shrinks by const factor!!!

Hence by choosing $\tau = 4 \ln \ln n$ gives us that

$$\mathbb{E}[|U_{t+\tau}|] \leq \frac{n}{2} \cdot \mathrm{e}^{-2 \ln \ln n} = n/2 \cdot (\ln n)^{-2}$$

By Markov's inequality, it holds that

$$\mathbb{P}\left[|U_{t+\tau}| \geq \frac{n}{2} \cdot (\ln n)^{-1}\right] \leq \mathbb{P}[|U_{t+\tau}| \geq (\ln n) \cdot \mathbb{E}[|U_{t+\tau}|]] \leq \frac{1}{\ln n}$$

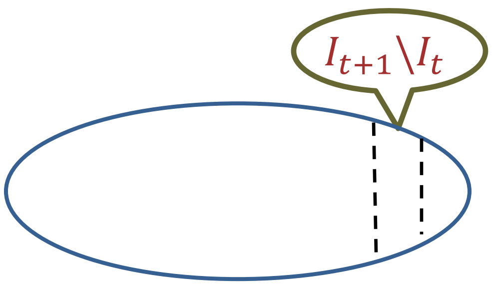
Analysis on complete graphs

Proof: Let I_t be the set of informed nodes in the end of round t, and U_t be the set of non-informed nodes in the end of round t. We divide the analysis into three phases:

- $1 \le |I_t| \le (n-1)/\log n$
- $(n-1)/\log n \le |I_t| \le n - n/\log n$
- $n - n/\log n \le |I_t| \le n$

<u>**Analysis of Phase III**</u> Let t be a round with $|I_t| \ge n - n/\log n$.

The probability that a fixed node is not informed by any vertex in I_t for $\alpha = \ln n + \ln n / \ln\ln n$ rounds is at most



$$\left(1 - \frac{1}{n-1}\right)^{(n-n/\ln n)\cdot\alpha} \le \exp\left(-\left(1-\frac{1}{\ln n}\right)\cdot\left(\ln n + \frac{\ln n}{\ln\ln n}\right)\right) = \exp\left(-\ln n - \frac{\ln n}{\ln\ln n} + 1 + \frac{1}{\ln\ln n}\right) = n^{-1} \cdot o(1)$$

Taking the union bound, with high probability every node gets informed after α rounds.

A similar proof can be applied for highly-connected graphs.

Application: ID Distribution

Algorithm Description

1. Initial node v sets $\mathrm{ID}_v = 0$.
2. t=0
3. while **t<T** do

4-1. every node *v* with ID sends (ID_v, t) to its random neighbour.

4-2. *if a node u without ID receives* (ID_v, t) *from its neigbhour, then*

$$\mathrm{ID}_u = 2^{t-1} + \mathrm{ID}_v$$

Note: if node u receives msg from multiple neighbours, u chooses a random one to perform the operation above.

4. t=t+1
5. end

Homework: Prove that every node receives a unique ID.