The push&pull protocol for rumour spreading

Abbas Mehrabian

PIMS, UBC and SFU

PODC 2015
Donostia-San Sebastián

Co-authors

Hüseyin Acan

Andrea Collevecchio

Nick Wormald

Monash University

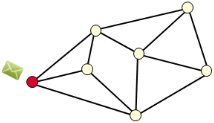
The push&pull rumour spreading protocol

[Demers, Gealy, Greene, Hauser, Irish, Larson, Manning, Shenker, Sturgis, Swinehart, Terry, Woods'87]

1. The ground is a simple connected graph.
2. At time 0, one vertex knows a rumour.
3. At each time-step $1, 2, \ldots,$
   every informed vertex sends the rumour to a random neighbour (PUSH);
   and every uninformed vertex queries a random neighbour about the rumour (PULL).

Example

ROUND 0

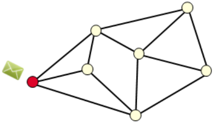


Push-Pull Protocol

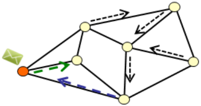
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

ROUND 0



ROUND 1

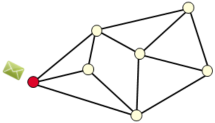


Push-Pull Protocol

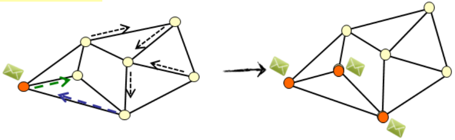
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

ROUND 0



ROUND 1

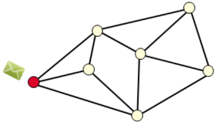


Push-Pull Protocol

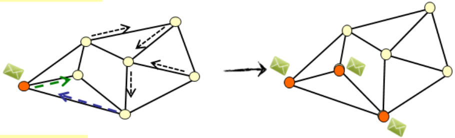
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

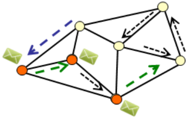
ROUND 0



ROUND 1



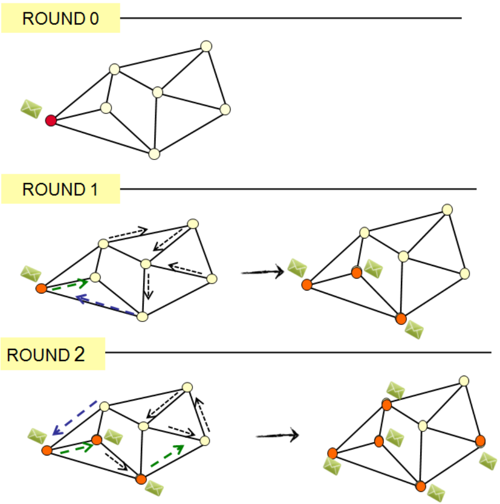
ROUND 2



Push-Pull Protocol

Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example



Push-Pull Protocol

Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Applications

1. Replicated databases
2. Broadcasting algorithms
3. News propagation in social networks and
4. Spread of viruses on the Internet.

A (more realistic) variant

Definition (The asynchronous variant: Boyd, Ghosh, Prabhakar, Shah'06)

Every vertex has an independent rate-1 Poisson process, and at times of process performs an operation (PUSH or PULL)

A (more realistic) variant

Definition (The asynchronous variant: Boyd, Ghosh, Prabhakar, Shah'06)

Every vertex has an independent rate-1 Poisson process,
and at times of process performs an operation (PUSH or PULL)
Or, equivalently,
In each step, one random vertex performs an operation (PUSH or PULL); but each step takes $1/n$ time units.

A (more realistic) variant

Definition (The asynchronous variant: Boyd, Ghosh, Prabhakar, Shah'06)

Every vertex has an independent rate-1 Poisson process,
and at times of process performs an operation (PUSH or PULL)
Or, equivalently,
In each step, one random vertex performs an operation (PUSH or PULL); but each step takes $1/n$ time units.

$s(G)$ and $a(G)$: average time it takes to broadcast the rumour.

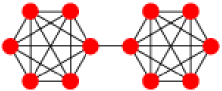
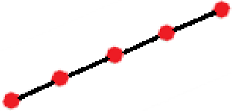
Known results

Graph G	$s(G)$	$a(G)$
Star	2	$\log n + O(1)$
Path	$(4/3)n + O(1)$	$n + O(1)$
Complete	$(1 + o(1)) \log_3 n$ [Karp,Schindelhauer,Shenker,Vöcking'00]	$\log n + o(1)$
$\mathcal{G}(n, p)$ (connected)	$\Theta(\log n)$ [Feige-Peleg-Raghavan-Upfal'90]	$(1 + o(1)) \log n$ [Panagiotou,Speidel'13]

- ✓ Many graph classes have been analyzed, including Erdős-Rényi graphs, random regular graphs, expander graphs, Barabási-Albert graphs, Chung-Lu graphs. In all of them $s(G) \asymp a(G) \asymp \log n$.
- ✓ Tight upper bounds have been found for $s(G)$ in terms of expansion profile by [Giakkoupis'11,'14].

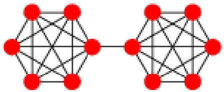
An extremal question

What's the maximum broadcast time of an n-vertex graph?

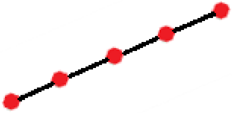
	
$n/4$	$4n/3$

An extremal question

What's the maximum broadcast time of an n-vertex graph?



$n/4$



$4n/3$

Theorem (Acan, Collevecchio, M, Wormald'15)

For any connected G on n vertices

$$s(G) < 4.6n$$

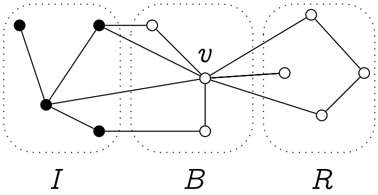
$$\log(n)/5 \leq a(G) < 4n$$

Proof idea for linear upper bound $a(G) < 4n$

Only pull operations are needed!

Proof idea for linear upper bound $a(G) < 4n$

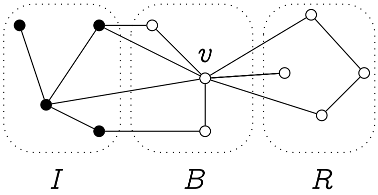
Only pull operations are needed!



We show inductively the expected remaining time $\leq 2|B| + 4|R|$

Proof idea for linear upper bound $a(G) < 4n$

Only pull operations are needed!

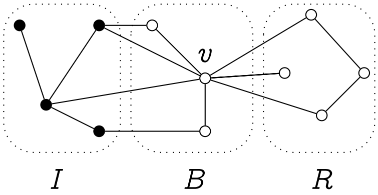


We show inductively the expected remaining time $\leq 2|B| + 4|R|$
Two cases:

1. If there is some boundary vertex v with $\deg_R(v) > \deg_B(v)$: it may take a lot of time to inform v, but once it is informed, $R \downdownarrows$ and $B \upuparrows$

Proof idea for linear upper bound $a(G) < 4n$

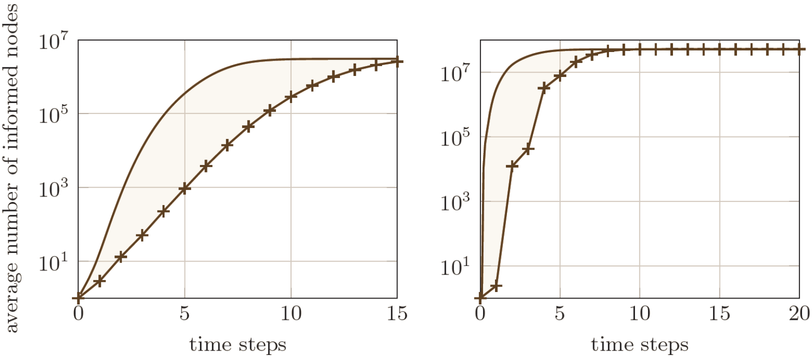
Only pull operations are needed!



We show inductively the expected remaining time $\leq 2|B| + 4|R|$
Two cases:

1. If there is some boundary vertex v with $\deg_R(v) > \deg_B(v)$: it may take a lot of time to inform v, but once it is informed, $R \downdownarrows$ and $B \upuparrows$
2. Otherwise, boundary vertices work together "in parallel" and average time for one of them to pull the rumour is 2.

Comparison of the two protocols on the same graph: experiments



(a) Orkut network **(b)** Twitter network

Figures from: Doerr, Fouz, and Friedrich. MedAlg 2012.

Comparison of the two protocols on the same graph: our results

Theorem (Acan, Collevecchio, M, Wormald'15)

We have

$$\frac{C_1}{\log n} \leq \frac{s(G)}{a(G)} \leq C_2 n^{2/3} \log n$$

Moreover, there exist infinitely many graphs for which this ratio is $\Omega\left((n/\log n)^{1/3}\right)$.

Comparison of the two protocols on the same graph: our results

Theorem (Acan, Collevecchio, M, Wormald'15)

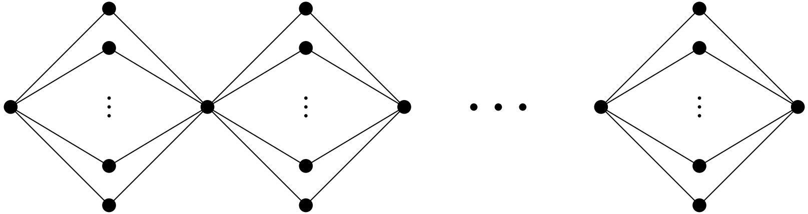
We have

$$\frac{C_1}{\log n} \leq \frac{s(G)}{a(G)} \leq C_2 n^{2/3} \log n$$

Moreover, there exist infinitely many graphs for which this ratio is $\Omega\left((n/\log n)^{1/3}\right)$.

The inequalities are proved by building careful couplings between the two variants.

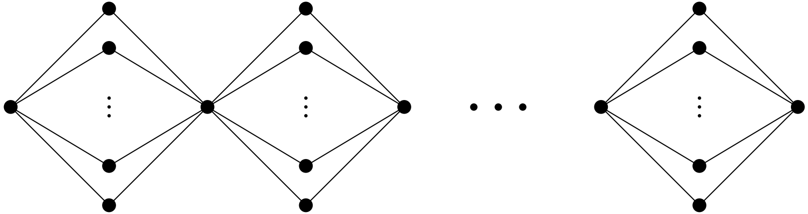
The string of diamonds



The asynchronous protocol is much quicker than its synchronous variant!

$$a(G) \ll s(G)$$

The string of diamonds



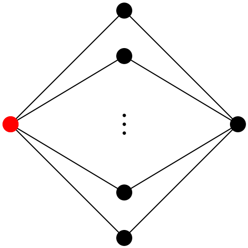
The asynchronous protocol is much quicker than its synchronous variant!

$$a(G) \ll s(G)$$

Indeed, asynchronous can be logarithmic, while synchronous is polynomial

counter-inuititive: synchrony harms!

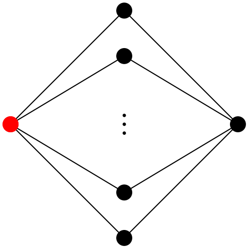
Time taken to pass through a diamond



k paths of length 2

Synchronous: needs ≥ 2 rounds

Time taken to pass through a diamond

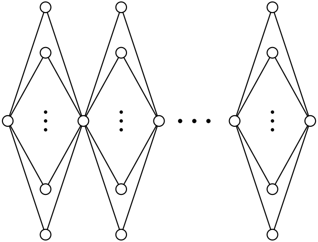


k paths of length 2

Synchronous: needs ≥ 2 rounds

Asynchronous: using a birthday-paradox type argument, the average time needed to pass the rumour is $O(1/\sqrt{k})$

The string of diamonds, continued

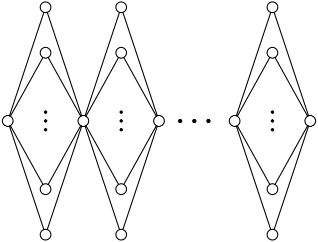


$\approx n^{1/3}$ diamonds, each consisting of $\approx n^{2/3}$ paths of length 2

Then

$$a(G) \le n^{1/3} \times \frac{C}{\sqrt{n^{2/3}}} + \log n = C + \log n$$

The string of diamonds, continued



$\approx n^{1/3}$ diamonds, each consisting of $\approx n^{2/3}$ paths of length 2

Then

$$a(G) \le n^{1/3} \times \frac{C}{\sqrt{n^{2/3}}} + \log n = C + \log n$$

while

$$s(G) \ge 2n^{1/3}$$

Final slide

Theorem (Acan, Collevecchio, M, Wormald'15)

For any connected G on n vertices

$$s(G) < 4.6n$$

$$\log(n)/5 \le a(G) < 4n$$

$$\frac{C_1}{\log n} \le \frac{s(G)}{a(G)} \le C_2 n^{2/3} \log n$$

For infinitely many graphs this ratio is $\Omega\left((n/\log n)^{1/3}\right)$.

Final slide

Theorem (Acan, Collevecchio, M, Wormald'15)

For any connected G on n vertices

$$s(G) < 4.6n$$
$$\log(n)/5 \leq a(G) < 4n$$
$$\frac{C_1}{\log n} \leq \frac{s(G)}{a(G)} \leq C_2 n^{2/3} \log n$$

For infinitely many graphs this ratio is $\Omega\left((n/\log n)^{1/3}\right)$.

Two weeks ago, Giakkoupis, Nazari, and Woelfel improved upper bound to $O\left(n^{1/2}\right)$