The push&pull protocol for rumour spreading

Abbas Mehrabian

UBC and SFU

UBC Discrete Math seminar
6 October 2015

Co-authors

Hüseyin Acan

Andrea Collevecchio

Nick Wormald

Monash University

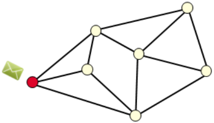
The push&pull rumour spreading protocol

[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

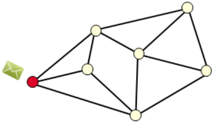


Push-Pull Protocol

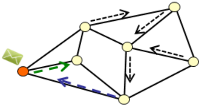
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

ROUND 0



ROUND 1

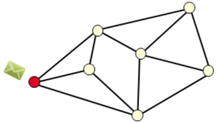


Push-Pull Protocol

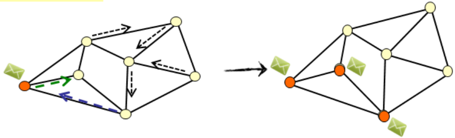
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

ROUND 0



ROUND 1

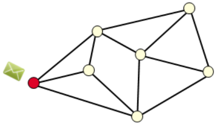


Push-Pull Protocol

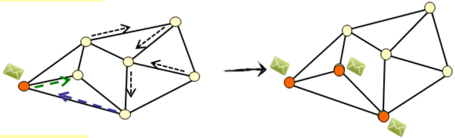
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

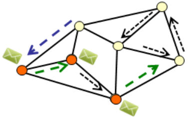
ROUND 0



ROUND 1



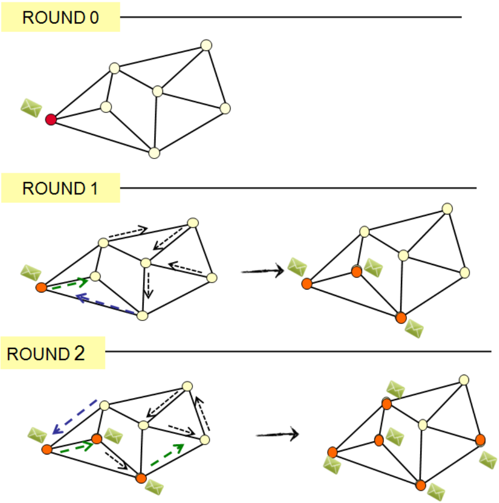
ROUND 2



Push-Pull Protocol

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

Example

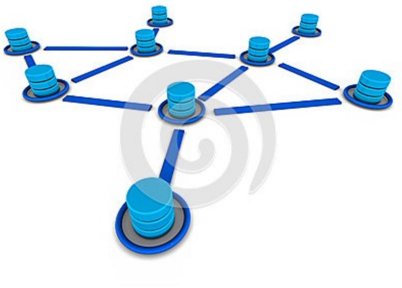


Push-Pull Protocol

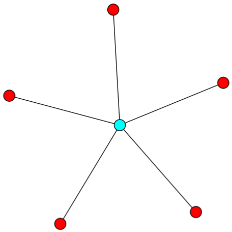
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Applications

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

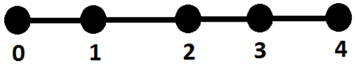


Example: a star

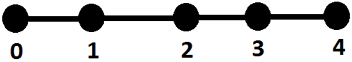


1 or 2 rounds

Example: a path

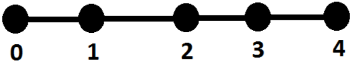


Example: a path



$$\text{inform} - \text{time}(0) = 0$$

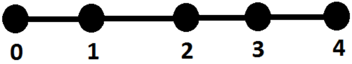
Example: a path



$$\text{inform} - \text{time}(0) = 0$$

$$\text{inform} - \text{time}(1) = 1$$

Example: a path

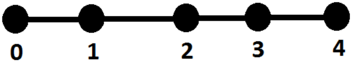


$$\mathrm{inform}-\mathrm{time}(0) = 0$$

$$\mathrm{inform}-\mathrm{time}(1) = 1$$

$$\begin{aligned}\mathrm{inform}-\mathrm{time}(2) &= 1+\min\{\mathrm{Geo}(1/2),\mathrm{Geo}(1/2)\}\\ &= 1+\mathrm{Geo}(3/4)\end{aligned}$$

Example: a path



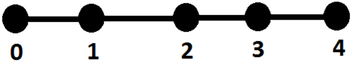
$$\mathrm{inform} - \mathrm{time}(0) = 0$$

$$\mathrm{inform} - \mathrm{time}(1) = 1$$

$$\mathrm{inform} - \mathrm{time}(2) = 1 + \mathrm{Geo}(3/4)$$

$$\mathrm{inform} - \mathrm{time}(3) = 1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4)$$

Example: a path



$$\mathrm{inform} - \mathrm{time}(0) = 0$$

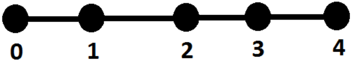
$$\mathrm{inform} - \mathrm{time}(1) = 1$$

$$\mathrm{inform} - \mathrm{time}(2) = 1 + \mathrm{Geo}(3/4)$$

$$\mathrm{inform} - \mathrm{time}(3) = 1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4)$$

$$\mathrm{inform} - \mathrm{time}(4) = 1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4) + 1$$

Example: a path



$$\text{inform} - \text{time}(0) = 0$$

$$\text{inform} - \text{time}(1) = 1$$

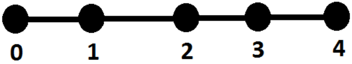
$$\text{inform} - \text{time}(2) = 1 + \text{Geo}(3/4)$$

$$\text{inform} - \text{time}(3) = 1 + \text{Geo}(3/4) + \text{Geo}(3/4)$$

$$\text{inform} - \text{time}(4) = 1 + \text{Geo}(3/4) + \text{Geo}(3/4) + 1$$

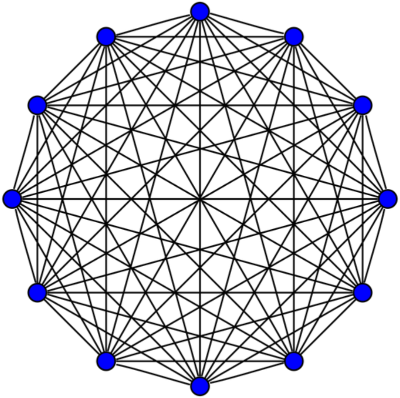
$$\mathbb{E}[\text{Spread Time}] = 2 + 2 \times 4/3 = 14/3$$

Example: a path



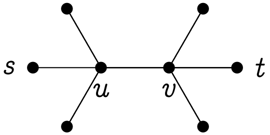
$$
\begin{aligned}
\mathrm{inform}-\mathrm{time}(0) &= 0\\
\mathrm{inform}-\mathrm{time}(1) &= 1\\
\mathrm{inform}-\mathrm{time}(2) &= 1+\mathrm{Geo}(3/4)\\
\mathrm{inform}-\mathrm{time}(3) &= 1+\mathrm{Geo}(3/4)+\mathrm{Geo}(3/4)\\
\mathrm{inform}-\mathrm{time}(4) &= 1+\mathrm{Geo}(3/4)+\mathrm{Geo}(3/4)+1\\
\mathbb{E}[\text{Spread Time}] &= 2+2\times 4/3 = 14/3\\
&= \frac{4}{3}n-2
\end{aligned}
$$

Example: a complete graph



$\log_3 n$ rounds [Karp, Schindelhauer, Shenker, Vöcking'00]

An example: double star



$$
\begin{aligned}
\mathrm{inform}-\mathrm{time}(s) &= 0 \\
\mathrm{inform}-\mathrm{time}(u) &= 1 \\
\mathrm{inform}-\mathrm{time}(v) &= 1 + \min\{\mathrm{Geo}(1/4), \mathrm{Geo}(1/4)\} = 1 + \mathrm{Geo}(7/16) \\
\mathrm{inform}-\mathrm{time}(t) &= 1 + \mathrm{Geo}(7/16) + 1 \\
\mathbb{E}[\text{Spread Time}] &= 2 + 7/16 \approx 2.44 \\
&\sim n/4
\end{aligned}
$$

Known results

$s(G)$: expected value of the spread time

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

Known results

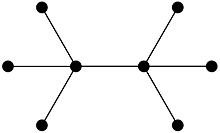
$s(G)$: expected value of the spread time

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

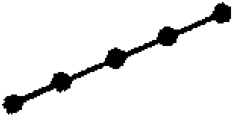
- ✓ Many graph classes have been analyzed, including hypercube graphs, random regular graphs, expander graphs, Barabási-Albert graphs, Chung-Lu graphs. In all of them $s(G) = \Theta(\ln n)$.
- ✓ Tight upper bounds have been found for $s(G)$ in terms of expansion profile by [Giakkoupis'11,'14].

An extremal question

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



$n/4$



$4n/3$

An upper bound of $13n \log_2 n$ is proved by
[Feige-Peleg-Raghavan-Upfal'90]

An asynchronous variant

A (more realistic) variant

In above protocol, all vertices act at the same time!

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

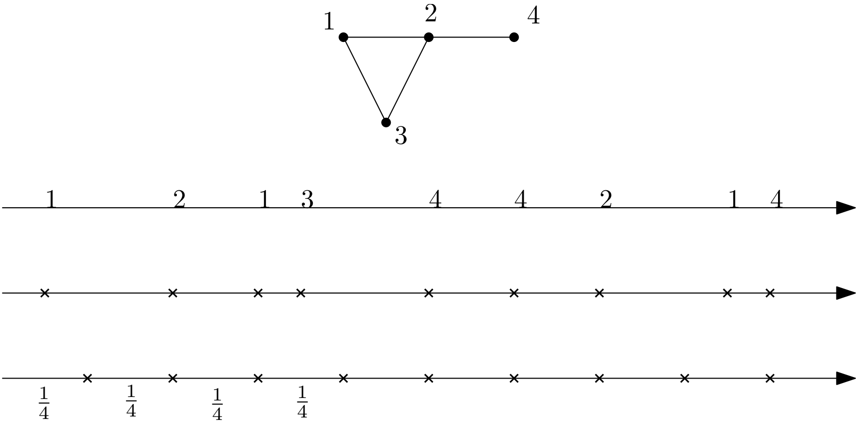
In above protocol, all vertices act at the same time!

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

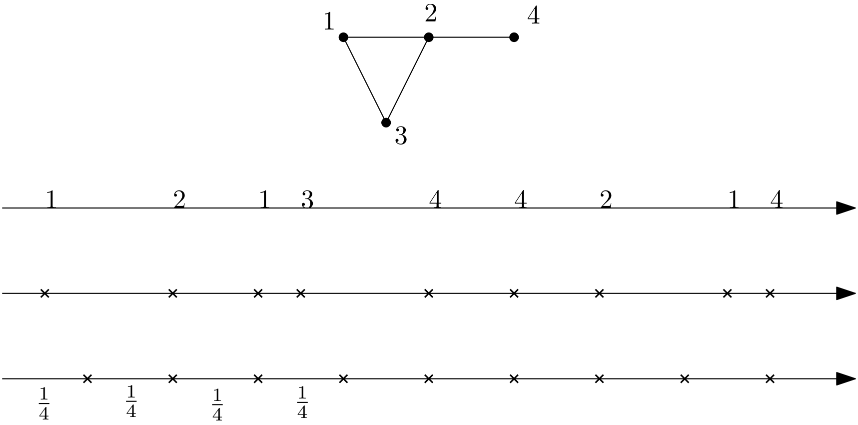
Every vertex has an independent rate-1 Poisson process,
and at times of process performs an operation (PUSH or PULL)

- ✓ Related to first-passage-percolation and Richardson's model for disease spread
- ✓ Vertices have no memory!

A discrete viewpoint of the asynchronous protocol



A discrete viewpoint of the asynchronous protocol



Discrete viewpoint of the asynchronous variant:
In each step, one random vertex performs an operation (PUSH or PULL); but each step takes $1/n$ time units.

Coupon collector

Question: Consider a bag containing n different balls.
In each step we draw a random ball and put it back.
How many draws to see each ball at least once?

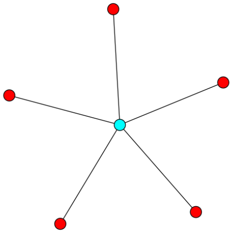
Coupon collector

Question: Consider a bag containing n different balls.
In each step we draw a random ball and put it back.
How many draws to see each ball at least once? About $n \ln n$

Corollary

The first time by which all vertices' clocks have rung at least once is about $\ln n$*.*

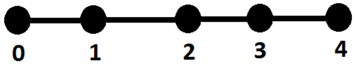
Example: a star



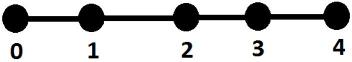
synchronous protocol: 1 or 2 rounds

asynchronous protocol: $\ln n$ amount of time

Example: a path

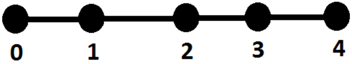


Example: a path



$$\text{inform} - \text{time}(0) = 0$$

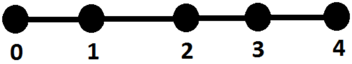
Example: a path



$$\text{inform} - \text{time}(0) = 0$$

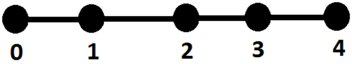
$$\text{inform} - \text{time}(1) = \min\{\text{Exp}(1), \text{Exp}(1/2)\} = \text{Exp}(3/2)$$

Example: a path



$$\begin{aligned}
\mathrm{inform}-\mathrm{time}(0) &= 0\\
\mathrm{inform}-\mathrm{time}(1) &= \mathrm{Exp}(3/2)\\
\mathrm{inform}-\mathrm{time}(2) &= \mathrm{Exp}(3/2)+\min\{\mathrm{Exp}(1/2),\mathrm{Exp}(1/2)\}\\
&= \mathrm{Exp}(3/2)+\mathrm{Exp}(1)
\end{aligned}$$

Example: a path



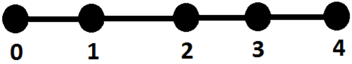
$$\text{inform} - \text{time}(0) = 0$$

$$\text{inform} - \text{time}(1) = \text{Exp}(3/2)$$

$$\text{inform} - \text{time}(2) = \text{Exp}(3/2) + \text{Exp}(1)$$

$$\text{inform} - \text{time}(3) = \text{Exp}(3/2) + \text{Exp}(1) + \text{Exp}(1)$$

Example: a path



$$\text{inform} - \text{time}(0) = 0$$

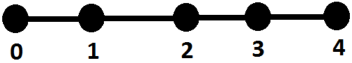
$$\text{inform} - \text{time}(1) = \text{Exp}(3/2)$$

$$\text{inform} - \text{time}(2) = \text{Exp}(3/2) + \text{Exp}(1)$$

$$\text{inform} - \text{time}(3) = \text{Exp}(3/2) + \text{Exp}(1) + \text{Exp}(1)$$

$$\text{inform} - \text{time}(4) = \text{Exp}(3/2) + \text{Exp}(1) + \text{Exp}(1) + \text{Exp}(3/2)$$

Example: a path



$$\text{inform}-\text{time}(0) = 0$$

$$\text{inform}-\text{time}(1) = \text{Exp}(3/2)$$

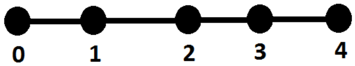
$$\text{inform}-\text{time}(2) = \text{Exp}(3/2) + \text{Exp}(1)$$

$$\text{inform}-\text{time}(3) = \text{Exp}(3/2) + \text{Exp}(1) + \text{Exp}(1)$$

$$\text{inform}-\text{time}(4) = \text{Exp}(3/2) + \text{Exp}(1) + \text{Exp}(1) + \text{Exp}(3/2)$$

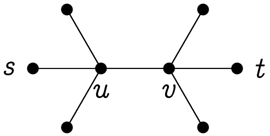
$$\mathbb{E}[\text{Spread Time}] = 2/3 + 1 + 1 + 2/3 = 10/3$$

Example: a path



$$
\begin{aligned}
\mathrm{inform-time}(0) &= 0 \\
\mathrm{inform-time}(1) &= \mathrm{Exp}(3/2) \\
\mathrm{inform-time}(2) &= \mathrm{Exp}(3/2) + \mathrm{Exp}(1) \\
\mathrm{inform-time}(3) &= \mathrm{Exp}(3/2) + \mathrm{Exp}(1) + \mathrm{Exp}(1) \\
\mathrm{inform-time}(4) &= \mathrm{Exp}(3/2) + \mathrm{Exp}(1) + \mathrm{Exp}(1) + \mathrm{Exp}(3/2) \\
\mathbb{E}[\text{Spread Time}] &= 2/3 + 2 \times 2/3 = 10/3 \\
&= n - 5/3 \qquad \left(\text{versus } \frac{4}{3}n - 2 \text{ for synchronous}\right)
\end{aligned}
$$

An example: double star



$$\mathrm{inform}-\mathrm{time}(u) = 0$$

$$\mathrm{inform}-\mathrm{time}(v) = \min\{\mathrm{Exp}(2/n), \mathrm{Exp}(2/n)\} = \mathrm{Exp}(4/n)$$

$$\mathbb{E}[\text{Spread Time}] = n/4 + \ln n$$

Known results

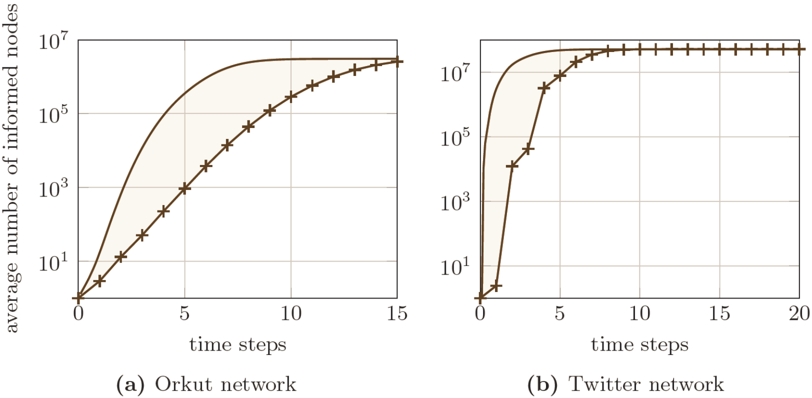
$a(G)$: expected value of spread time in asynchronous protocol

Graph G	$s(G)$	$a(G)$
Star	2	$\ln n + O(1)$
Path	$(4/3)n + O(1)$	$n + O(1)$
Double star	$(1 + o(1))n/4$	$(1 + o(1))n/4$
Complete	$(1 + o(1)) \log_3 n$ [Karp,Schindelhauer,Shenker,Vöcking'00]	$\ln n + o(1)$
Hypercube graph	$\Theta(\ln n)$ [Feige-Peleg-Raghavan-Upfal'90]	$\Theta(\ln n)$ [Fill,Pemantle'93]
$\mathcal{G}(n, p)$ (connected)	$\Theta(\ln n)$ [Feige-Peleg-Raghavan-Upfal'90]	$(1 + o(1)) \ln n$ [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 $s(G), a(G) = \Theta(\ln n)$

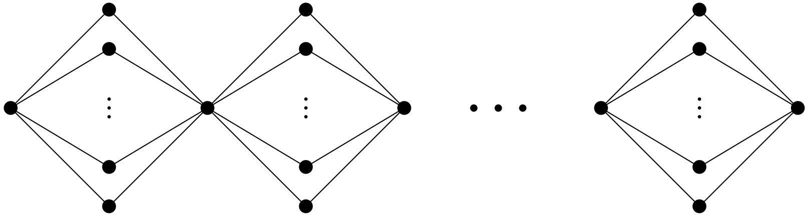
Comparison of the two variants

Comparison of the two protocols on the same graph: experiments



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

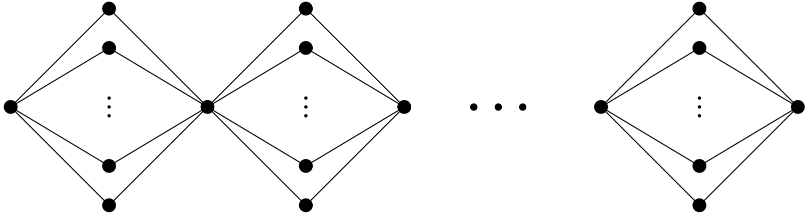
The string of diamonds



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!

Birthday paradox

Question: Consider a bag containing n different balls.
In each step we draw a random ball and put it back.
How many draws to see some ball twice?

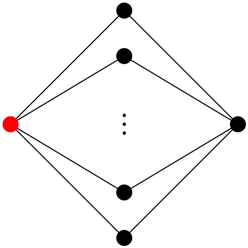
Birthday paradox

Question: Consider a bag containing n different balls.
In each step we draw a random ball and put it back.
How many draws to see some ball twice?
About $\sqrt{\pi n/2} \approx 1.25\sqrt{n}$

Corollary

The first time to have some clock ring twice is about $1.25/\sqrt{n}$.

Time taken to pass through a diamond

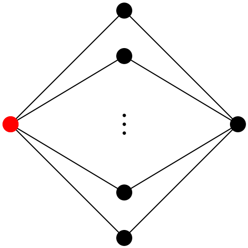


k paths of length 2

Average time to pass the rumour:

Synchronous: 2 rounds

Time taken to pass through a diamond



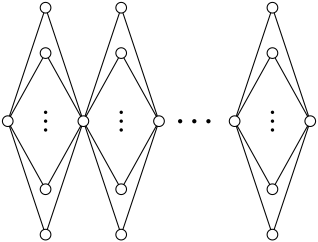
k paths of length 2

Average time to pass the rumour:

Synchronous: 2 rounds

Asynchronous: $\leq 4 \times 1.25/\sqrt{k}$

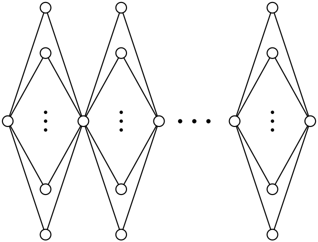
The string of diamonds, continued



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

$$a(G) \leq n^{1/3} \times \frac{5}{\sqrt{n^{2/3}}} + \ln n = 5 + \ln n$$

The string of diamonds, continued



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

$$a(G) \leq n^{1/3} \times \frac{5}{\sqrt{n^{2/3}}} + \ln n = 5 + \ln n$$

while

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

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

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

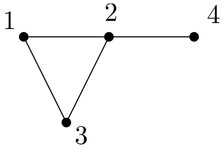
We have

$$\frac{1}{\ln n} \leq \frac{s(G)}{a(G)} \leq 200 n^{2/3} \ln n$$

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

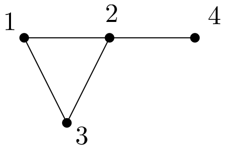
Proof idea for $a(G) \leq s(G) \times \ln n$

Consider an arbitrary calling sequence:



$1 \downarrow 2 \quad 2 \downarrow 3 \quad 1 \downarrow 2 \quad 3 \downarrow 2 \quad 4 \downarrow 2 \quad 4 \downarrow 2 \quad 2 \downarrow 3 \quad 1 \downarrow 3 \quad 4 \downarrow 2 \quad 3 \downarrow 1$

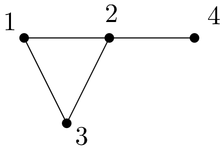
Proof idea for $a(G) \leq s(G) \times \ln n$



$\boxed{1 \downarrow 2 \quad 2 \downarrow 3 \quad 1 \downarrow 2 \quad 3 \downarrow 2 \quad 4 \downarrow 2}$ $\boxed{4 \downarrow 2 \quad 2 \downarrow 3 \quad 1 \downarrow 3 \quad 4 \downarrow 2 \quad 3 \downarrow 1}$

$1 \downarrow 2 \quad 2 \downarrow 3 \quad 1 \downarrow 2 \quad 3 \downarrow 2 \quad 4 \downarrow 2$

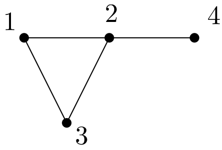
Proof idea for $a(G) \leq s(G) \times \ln n$



$$\boxed{\begin{array}{ccccc} 1 & 2 & 1 & 3 & 4 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 2 & 3 & 2 & 2 & 2 \end{array}} \quad \boxed{\begin{array}{ccccc} 4 & 2 & 1 & 4 & 3 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 2 & 3 & 3 & 2 & 1 \end{array}}$$

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 2 & 3 & 2 & 2 \end{array}$$

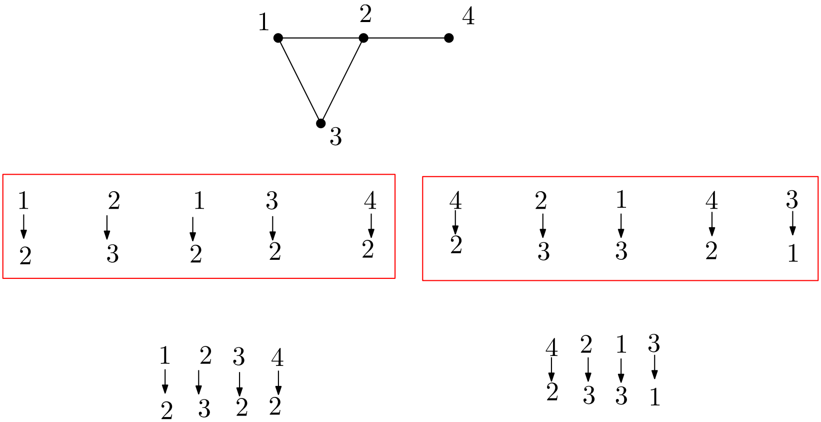
Proof idea for $a(G) \leq s(G) \times \ln n$



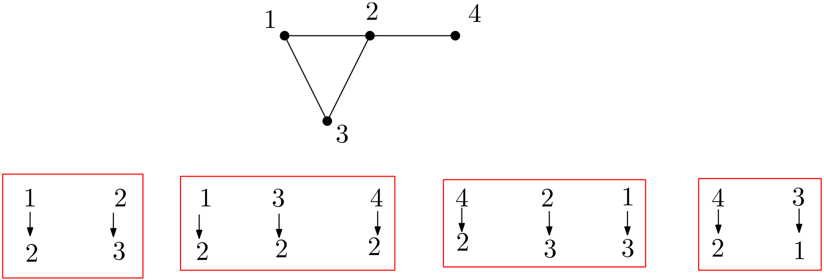
$$\begin{array}{ccccc} 1 & 2 & 1 & 3 & 4 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 2 & 3 & 2 & 2 & 2 \end{array} \qquad \begin{array}{ccccc} 4 & 2 & 1 & 4 & 3 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 2 & 3 & 3 & 2 & 1 \end{array}$$

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 2 & 3 & 2 & 2 \end{array}$$

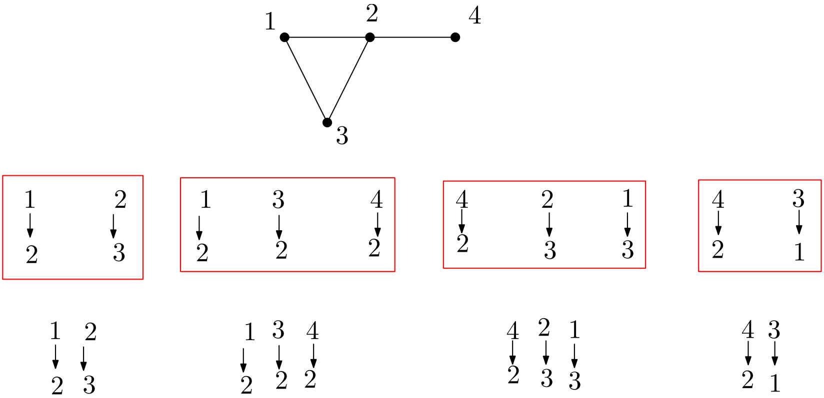
Proof idea for $a(G) \leq s(G) \times \ln n$



Proof idea for $s(G) \leq a(G) \times n^{2/3}$



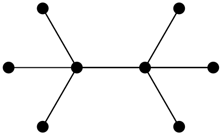
Proof idea for $s(G) \leq a(G) \times n^{2/3}$



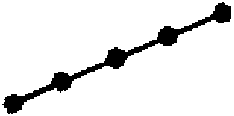
Extremal spread times

The extremal question

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



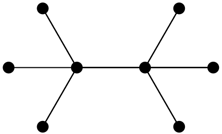
$s(G), a(G) \sim n/4$



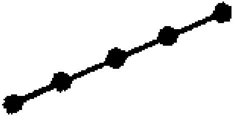
$s(G) \sim 4n/3$ and $a(G) \sim n$

The extremal question

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



$s(G), a(G) \sim n/4$



$s(G) \sim 4n/3$ and $a(G) \sim n$

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

For any connected G on n vertices

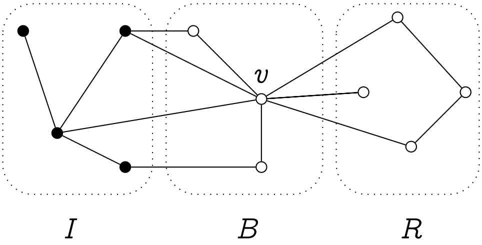
$$s(G) < 5n$$
$$\ln(n)/5 < a(G) < 4n$$

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

Only pull operations are needed!

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

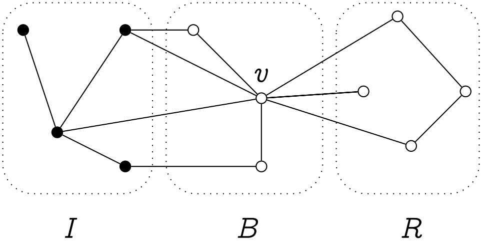
Only pull operations are needed!



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!

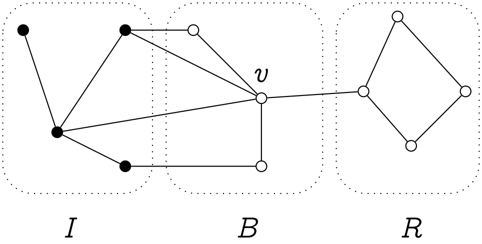


We show inductively the expected remaining time $\leq 2|B| + 4|R|$

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 $\leq 2|B| + 4|R|$

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, each boundary vertex has pulling rate $\geq 1/2|B|$, and the B boundary vertices work together "in parallel" and average time for one of them to pull the rumour is 2.

Final slide

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

For any connected G on n vertices

$$s(G) < 5n$$
$$\ln(n)/5 < a(G) < 4n$$
$$\frac{1}{\ln n} < \frac{s(G)}{a(G)} < 200n^{2/3} \ln n$$

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

Final slide

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

For any connected G on n vertices

$$s(G) < 5n$$
$$\ln(n)/5 < a(G) < 4n$$
$$\frac{1}{\ln n} < \frac{s(G)}{a(G)} < 200n^{2/3}\ln n$$

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

Giakkoupis, Nazari, and Woelfel [July 2015] improved upper bound to $O\left(n^{1/2}\right)$