The push&pull protocol for rumour spreading

Abbas Mehrabian

Simons Institute for Theory of Computing, UC Berkeley

University of Washington
28 February 2017

Co-authors

Omer Angel

Hüseyin Acan

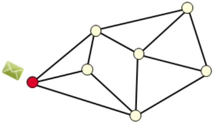
Andrea Collevecchio

Yuval Peres

Nick Wormald

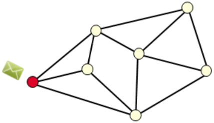
Example

ROUND 0



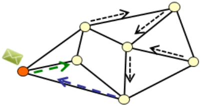
Example

ROUND 0

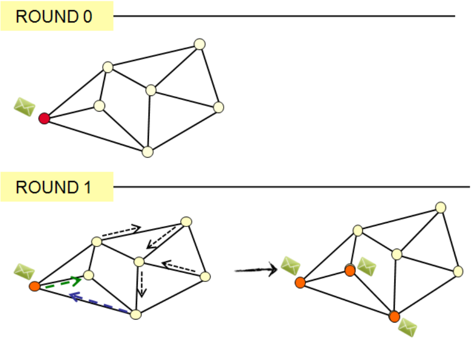


In each round, every vertex calls a random neighbour

ROUND 1



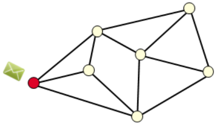
Example



In each round, every vertex calls a random neighbour and they exchange their information

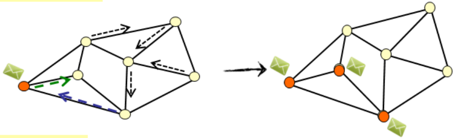
Example

ROUND 0

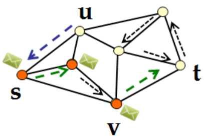


In each round, every vertex calls a random neighbour and they exchange their information

ROUND 1



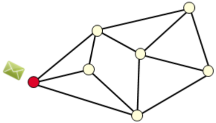
ROUND 2



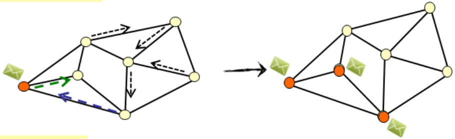
u pulls from s
v pushes to t

Example

ROUND 0

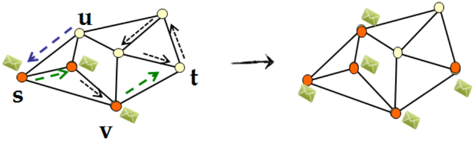


ROUND 1



In each round, every vertex calls a random neighbour and they exchange their information

ROUND 2



u pulls from s
v pushes to t

The push&pull rumour spreading protocol

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

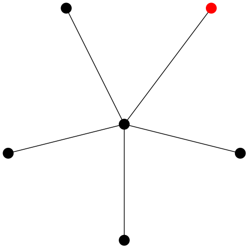
1. Consider a simple connected graph.
2. At time 0, one vertex knows a rumour.
3. At each time-step $1, 2, \ldots,$
   every informed vertex sends the rumour to a random neighbour (PUSH);
   and every uninformed vertex queries a random neighbour about the rumour (PULL).

We are interested in the spread time.

Applications

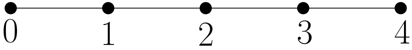
1. Replicated databases
2. Broadcasting algorithms
3. News propagation in social networks
4. Spread of viruses on the Internet.

Example: a star



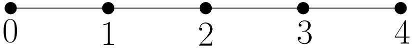
2 rounds

Example: path graph



vertex 0 knows rumour at round 0

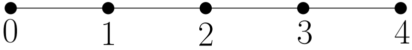
Example: path graph



vertex 0 knows rumour at round 0

vertex 1 is informed at round 1

Example: path graph

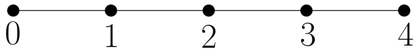


vertex 0 knows rumour at round 0

vertex 1 is informed at round 1

vertex 2 is informed at round
$1 + \min\{\text{Geo}(1/2), \text{Geo}(1/2)\} = 1 + \text{Geo}(3/4)$

Example: path graph



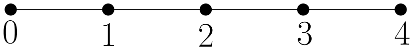
vertex 0 knows rumour at round 0

vertex 1 is informed at round 1

vertex 2 is informed at round $1 + \min\{\mathrm{Geo}(1/2), \mathrm{Geo}(1/2)\} = 1 + \mathrm{Geo}(3/4)$

vertex 3 is informed at round $1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4)$

Example: path graph



vertex 0 knows rumour at round 0

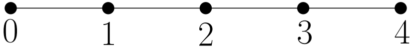
vertex 1 is informed at round 1

vertex 2 is informed at round
$1 + \min\{\mathrm{Geo}(1/2), \mathrm{Geo}(1/2)\} = 1 + \mathrm{Geo}(3/4)$

vertex 3 is informed at round $1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4)$

vertex 4 is informed at round $1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4) + 1$

Example: path graph



vertex 0 knows rumour at round 0

vertex 1 is informed at round 1

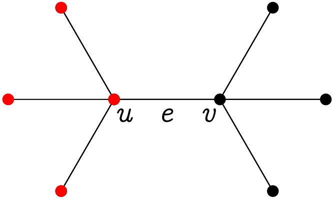
vertex 2 is informed at round
$1 + \min\{\mathrm{Geo}(1/2), \mathrm{Geo}(1/2)\} = 1 + \mathrm{Geo}(3/4)$

vertex 3 is informed at round $1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4)$

vertex 4 is informed at round $1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4) + 1$

$$\mathbb{E}[\text{Spread Time}] = \frac{4}{3}n - 2$$

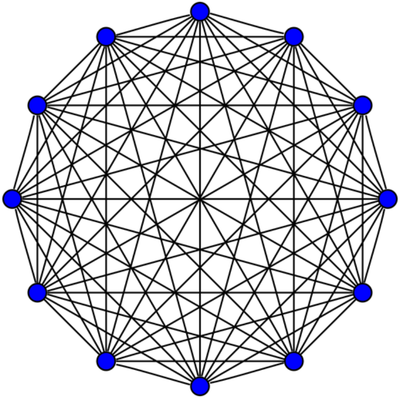
An example: double star



$$\text{Time to pass edge } e = \min\{\text{Geo}(1/4), \text{Geo}(1/4)\}$$
$$= \min\{\text{Geo}(\frac{1}{n/2}), \text{Geo}(\frac{1}{n/2})\} = \text{Geo}(\frac{4}{n} - \frac{4}{n^2})$$

Expected spread time $\sim n/4$

Example: a complete graph



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

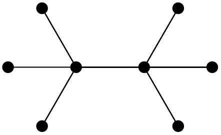
Known results

$s(G)$ expected value of spread time (for worst starting vertex)

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]

An extremal question

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



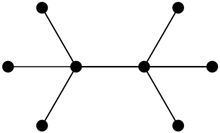
$n/4$

$4n/3$

$O(n \log n)$ upper bound by [Feige, Peleg, Raghavan, Upfal'90]

An extremal question

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



$n/4$

$4n/3$

$O(n \log n)$ upper bound by [Feige, Peleg, Raghavan, Upfal'90]

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

For any connected G on n vertices

$$s(G) < 5n$$

Only pull operations are needed!

An asynchronous variant

A (more realistic) variant

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

In each step, one random vertex performs one action (PUSH or PULL).
Each step takes time $1/n$.

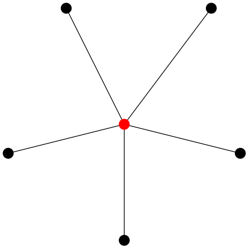
A (more realistic) variant

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

In each step, one random vertex performs one action (PUSH or PULL).
Each step takes time $1/n$.

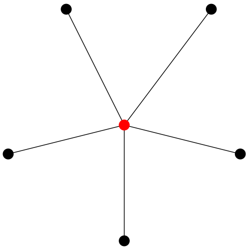
Almost equivalent definition:
every vertex has an exponential clock with rate 1,
at each clock ring, performs one action.

Example: a star



synchronous protocol: 1 round

Example: a star



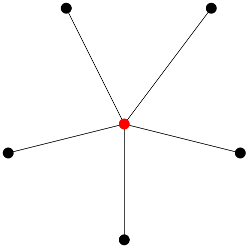
synchronous protocol: 1 round

Coupon collector: 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?

Example: a star



synchronous protocol: 1 round

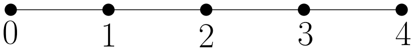
Coupon collector: Consider a bag containing n different balls.

In each step we draw a random ball and put it back.

How many draws to see each ball at least once? About $n \ln n$.

asynchronous protocol: $n \ln n$ steps $= \ln n$ amount of time

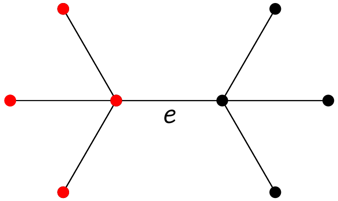
Example: a path



Spread time $\sim$ sum of $n-1$ independent exponentials

$$\mathbb{E}[\text{Spread Time}] = n - 5/3 \qquad (\text{versus } \frac{4}{3}n - 2 \text{ for synchronous})$$

An example: double star



Time to pass edge $e = \min\{\mathrm{Exp}(\frac{1}{n/2}), \mathrm{Exp}(\frac{1}{n/2})\} = \mathrm{Exp}(4/n)$

Expected spread time $\sim n/4$

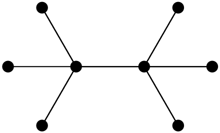
Some known results

$a(G)$ 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]

The extremal question

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

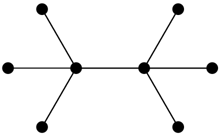


$\Omega(n)$

$\Omega(n)$

The extremal question

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



$\Omega(n)$

$\Omega(n)$

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

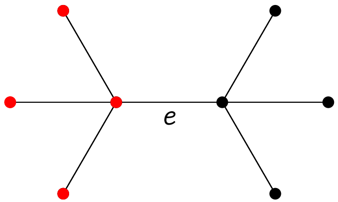
For any connected G on n vertices

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

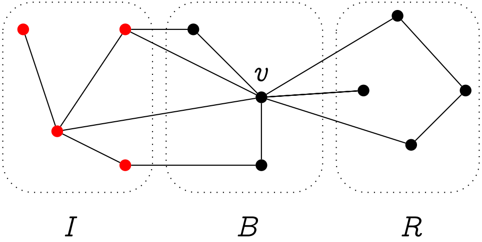
Only pull operations are needed!

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

Induction?

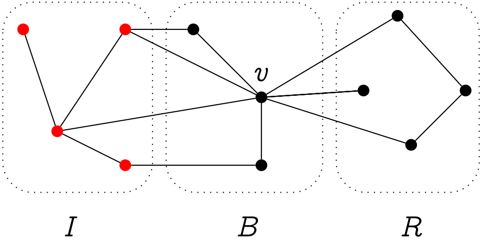


Proof idea for linear upper bound $a(G) < 4n$



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

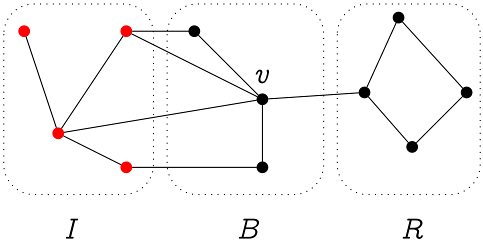
Proof idea for linear upper bound $a(G) < 4n$



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 $a(G) < 4n$

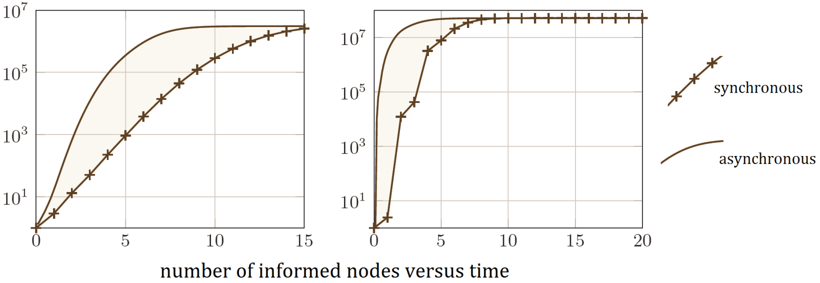


We show inductively the expected remaining time $\leq 2|B| + 4|R|$

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.

Comparison of the two variants

Comparison of the two protocols on the same graph: experiments

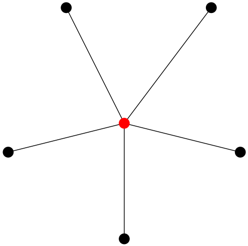


(a) Orkut network **(b)** Twitter network

Figures from: Doerr, Fouz, and Friedrich'12.

The star

In which graph synchronous is quicker than asynchronous?



synchronous protocol: 1 round

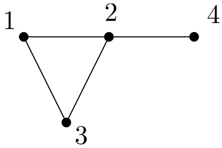
asynchronous protocol: $\ln n$ time

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

$$a(G) \leq O(s(G) \times \ln n).$$

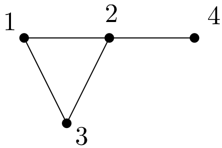
Proof idea for $a(G) \leq s(G) \times \ln n$

Consider an arbitrary calling sequence:



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

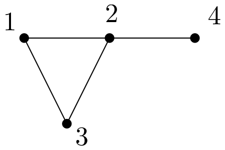
Proof idea for $a(G) \leq s(G) \times \ln n$



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

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

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

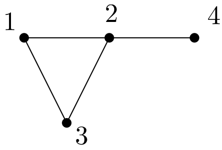


$1 \to 2 \quad 2 \to 3 \quad 1 \to 2 \quad 3 \to 2 \quad 4 \to 2$

$4 \to 2 \quad 2 \to 3 \quad 1 \to 3 \quad 4 \to 2 \quad 3 \to 1$

$1 \to 2 \quad 2 \to 3 \quad 3 \to 2 \quad 4 \to 2$

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

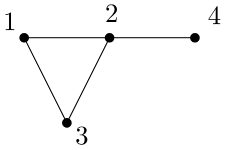


$1 \to 2 \quad 2 \to 3 \quad 1 \to 2 \quad 3 \to 2 \quad 4 \to 2$

$4 \to 2 \quad 2 \to 3 \quad 1 \to 3 \quad 4 \to 2 \quad 3 \to 1$

$1 \to 2 \quad 2 \to 3 \quad 3 \to 2 \quad 4 \to 2$

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



$$1 \to 2 \quad 2 \to 3 \quad 1 \to 2 \quad 3 \to 2 \quad 4 \to 2$$

$$4 \to 2 \quad 2 \to 3 \quad 1 \to 3 \quad 4 \to 2 \quad 3 \to 1$$

$$1 \to 2 \quad 2 \to 3 \quad 3 \to 2 \quad 4 \to 2$$

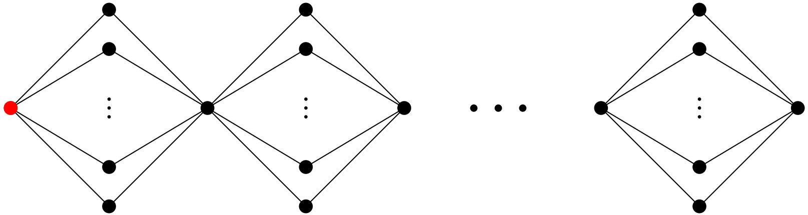
$$4 \to 2 \quad 2 \to 3 \quad 1 \to 3 \quad 3 \to 1$$

The string of diamonds

In which graph asynchronous is much quicker than synchronous?

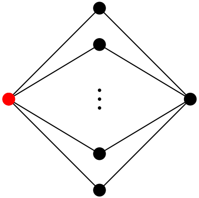
The string of diamonds

In which graph asynchronous is much quicker than synchronous?



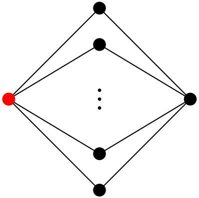
logarithmic $\ll$ polynomial

Time taken to pass through a diamond



k paths of length 2

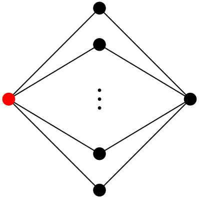
Time taken to pass through a diamond



k paths of length 2

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

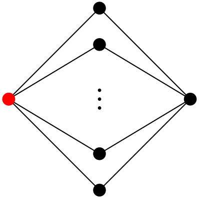
Time taken to pass through a diamond



k paths of length 2

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

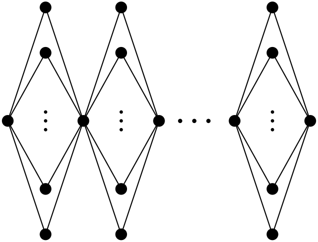
Time taken to pass through a diamond



k paths of length 2

Birthday paradox: Consider a bag containing k different balls.
In each step we draw a random ball and put it back.
How many draws to see some ball twice? $\sqrt{\pi k/2} \approx 1.25\sqrt{k}$
Time to pass the rumour
Asynchronous: $\leq 4 \times 1.25/\sqrt{k}$
Synchronous: ≥ 2

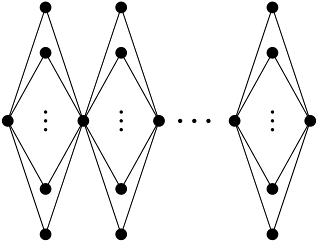
The string of diamonds, continued



$n^{1/3}$ 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



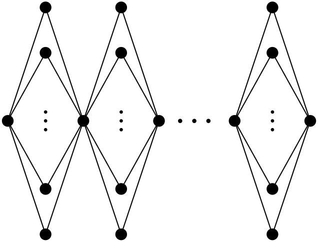
$n^{1/3}$ diamonds, each consisting of $n^{2/3}$ paths of length 2

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

while

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

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

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

$\frac{s(G)}{a(G)}$ can be as large as $\widetilde{\Omega}\left(n^{1/3}\right)$, but can it be larger?

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

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

$$\frac{s(G)}{a(G)} = \widetilde{O}\left(n^{2/3}\right)$$

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

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

$$\frac{s(G)}{a(G)} = \widetilde{O}\left(n^{2/3}\right)$$

$\frac{s(G)}{a(G)} = O\left(n^{1/2}\right)$ [Giakkoupis, Nazari, and Woelfel'16]

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

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

$$\frac{s(G)}{a(G)} = \widetilde{O}\left(n^{2/3}\right)$$

$\frac{s(G)}{a(G)} = O\left(n^{1/2}\right)$ [Giakkoupis, Nazari, and Woelfel'16]

Theorem (Angel, M, Peres'17+)

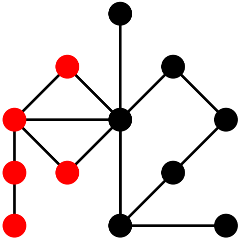
We have

$$\frac{s(G)}{a(G)} = \widetilde{O}\left(n^{1/3}\right),$$

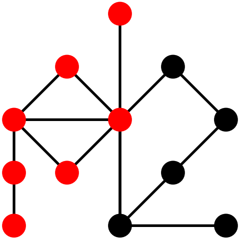
which is tight.

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Build a coupling so that



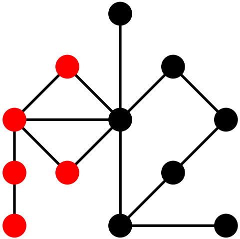
asynchronous contamination by time 1



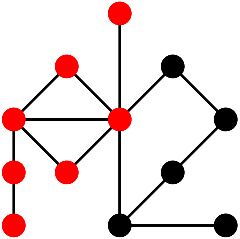
synchronous contamination by time x

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Build a coupling so that



asynchronous contamination
by time 1



synchronous contamination
by time x

If asynchronous contaminates a path of length L,
need to have $x \geq L$

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Lemma

In asynchronous, after n steps (by time 1), rumour does not pass along a path of length $> Cn^{1/3}$ (with high prob).

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Lemma

In asynchronous, after n steps (by time 1), rumour does not pass along a path of length $> Cn^{1/3}$ (with high prob).

For fixed path $v_1 v_2 \ldots v_L$, this probability is

$$\leq 2^L \times \binom{n}{L} \times n^{-L} \times \prod_{i=1}^{L-1} \max\left\{\frac{1}{\deg(v_i)}, \frac{1}{\deg(v_{i+1})}\right\}$$

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Lemma

In asynchronous, after n steps (by time 1), rumour does not pass along a path of length $> Cn^{1/3}$ (with high prob).

For fixed path $v_1 v_2 \ldots v_L$, this probability is

$$\leq 2^L \times \binom{n}{L} \times n^{-L} \times \prod_{i=1}^{L-1} \max\left\{\frac{1}{\deg(v_i)}, \frac{1}{\deg(v_{i+1})}\right\}$$

Will show

$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \leq (Cn/L)^{L/2} \qquad (1)$$

Proof sketch for $s(G) \le a(G) \times (n \ln n)^{1/3}$

Lemma

In asynchronous, after n steps (by time 1), rumour does not pass along a path of length $> Cn^{1/3}$ (with high prob).

For fixed path $v_1 v_2 \ldots v_L$, this probability is

$$\le 2^L \times \binom{n}{L} \times n^{-L} \times \prod_{i=1}^{L-1} \max\left\{\frac{1}{\deg(v_i)}, \frac{1}{\deg(v_{i+1})}\right\}$$

Will show

$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \le (Cn/L)^{L/2} \tag{1}$$

Implies the total probability is $\le (C\sqrt{n}/L\sqrt{L})^L$.
Putting $L = Cn^{1/3}$ makes this $o(1)$.

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Want to show

$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \leq (Cn/L)^{L/2}$$

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Want to show

$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \leq (Cn/L)^{L/2}$$

Baby version: we have

$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\deg(v_i)} \leq n$$

Once we choose the first vertex, the $1/\deg$ factors cancel number of choices for next vertices!

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Want to show

$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \leq (Cn/L)^{L/2}$$

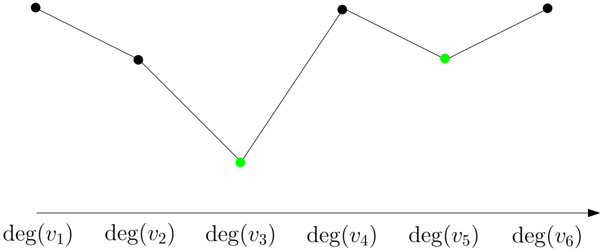
Consider the local minima vertices in the sequence $\deg(v_1), \deg(v_2), \ldots, \deg(v_L)$

Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Want to show

$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \leq (Cn/L)^{L/2}$$

Consider the local minima vertices in the sequence $\deg(v_1), \deg(v_2), \ldots, \deg(v_L)$



Proof sketch for $s(G) \leq a(G) \times (n \ln n)^{1/3}$

Want to show

$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \leq (Cn/L)^{L/2}$$

Consider the local minima vertices in the sequence $\deg(v_1), \deg(v_2), \ldots, \deg(v_L)$.
Once we choose these vertices, the $1/\min\{\deg, \deg\}$ factors cancel out number of choices for other vertices, so

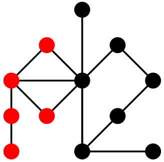
$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \leq \sum_{s=0}^{L/2} \binom{L}{s} \cdot \binom{n}{s} \leq (Cn/L)^{L/2}$$

Proof sketch for $s(G) \le a(G) \times (n \ln n)^{1/3}$

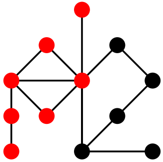
Lemma

In asynchronous, during $[0,1]$, rumour does not pass along a path of length $> Cn^{1/3}$ (with high prob).

Using careful couplings,



asynchronous contamination
by time 1



synchronous contamination
by time $C(n \ln n)^{1/3}$

$$s(G) \le a(G) \times C(n \ln n)^{1/3}$$

Summary of our results on push&pull

Theorem (Acan, Angel, Collevecchio, M, Peres, Wormald'15,'17)

For any connected G on n vertices,

$$s(G) < 5n$$

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

$$\frac{1}{\ln n} < \frac{s(G)}{a(G)} < C(n \ln n)^{1/3}$$

All bounds are tight, up to constant factors.

Future directions

1. Connect $s(G)/a(G)$ with other graph properties.
2. How to choose first vertex(es) carefully to minimize the spread time? [Kempe, J. Kleinberg, E. Tardos'03]
3. Number of passed messages? [Fraigniaud, Giakkoupis'10]
4. More than one message? [Censor-Hillel, Haeupler, Kelner, Maymounkov'12]
5. Variation: each node spreads for a bounded number of rounds [Akbarpour, Jackson'16].

Future directions

1. Connect $s(G)/a(G)$ with other graph properties.
2. How to choose first vertex(es) carefully to minimize the spread time? [Kempe, J. Kleinberg, E. Tardos'03]
3. Number of passed messages? [Fraigniaud, Giakkoupis'10]
4. More than one message? [Censor-Hillel, Haeupler, Kelner, Maymounkov'12]
5. Variation: each node spreads for a bounded number of rounds [Akbarpour, Jackson'16].