The String of Diamonds is tight for Rumor Spreading

Abbas Mehrabian

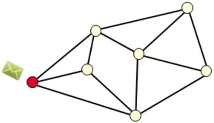
McGill University, Canada

RANDOM 2017, UC Berkeley
17 August 2017

Joint work with Omer Angel and Yuval Peres

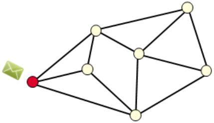
Example

ROUND 0

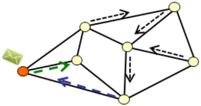


Example

ROUND 0



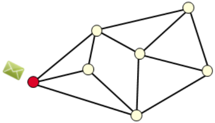
ROUND 1



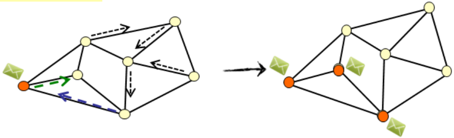
In each round, every vertex calls a random neighbour

Example

ROUND 0



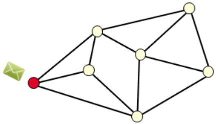
ROUND 1



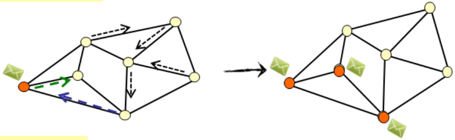
In each round, every vertex calls a random neighbour and they exchange their information

Example

ROUND 0

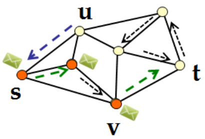


ROUND 1



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

ROUND 2

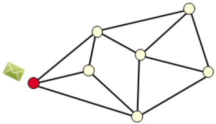


u pulls from s

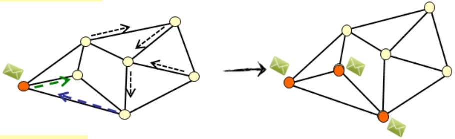
v pushes to t

Example

ROUND 0

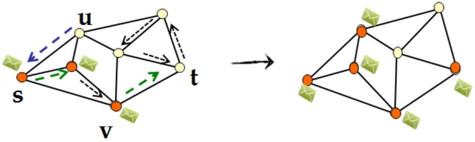


ROUND 1



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 rumor spreading protocol

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

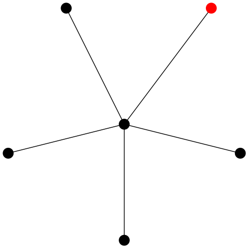
1. Consider a simple connected graph.
2. At time 0, one vertex knows a rumor.
3. At each time-step $1, 2, \ldots,$
   every informed vertex sends the rumor to a random neighbour (PUSH);
   and every uninformed vertex queries a random neighbour about the rumor (PULL).

We are interested in the spread time.

Applications

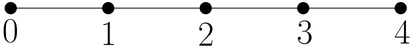
1. Replicated databases
2. Broadcasting algorithms
3. News propagation in social networks
4. Spread of viruses on the Internet.

Example: a star



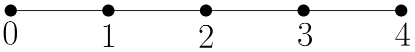
2 rounds

Example: path graph



vertex 0 knows rumor at round 0

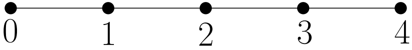
Example: path graph



vertex 0 knows rumor at round 0

vertex 1 is informed at round 1

Example: path graph

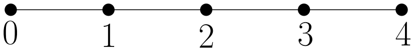


vertex 0 knows rumor at round 0

vertex 1 is informed at round 1

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

Example: path graph



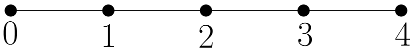
vertex 0 knows rumor at round 0

vertex 1 is informed at round 1

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

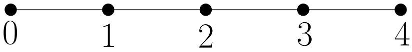
vertex 3 is informed at round $1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4)$

Example: path graph



vertex 0	knows rumor 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 rumor at round 0

vertex 1 is informed at round 1

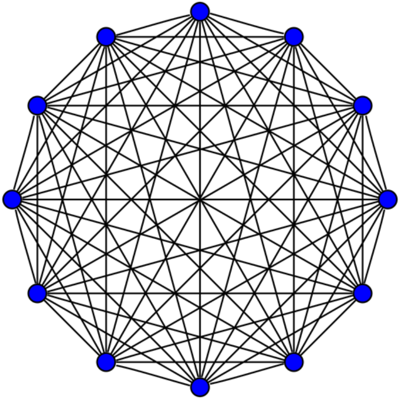
vertex 2 is informed at round
$1 + \min\{\mathrm{Geo}(1/2), \mathrm{Geo}(1/2)\} = 1 + \mathrm{Geo}(3/4)$

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

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

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

Example: a complete graph



$(1 + o(1)) \log_3 n$ rounds in expectation [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	$(4/3)n + O(1)$
Hypercube, $\mathcal{G}(n,p)$ (connected)	$\Theta(\ln n)$ [Feige, Peleg, Raghavan, Upfal'90]
Complete	$(1+o(1))\log_3 n$ [Karp,Schindelhauer,Shenker,Vöcking'00]
General	$O(n)$ [Acan, Collevecchio, M., Wormald'15]

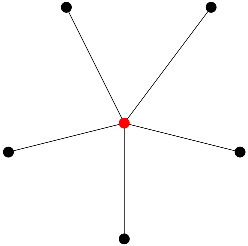
An asynchronous variant

A (more realistic) variant

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

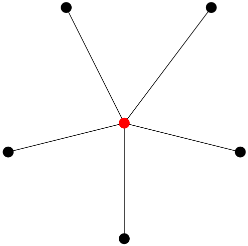
Every vertex has an exponential clock with rate 1,
at each clock ring, performs one action. (PUSH or PULL).

Example: a star



synchronous protocol: 1 time-step

Example: a star

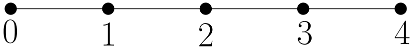


synchronous protocol: 1 time-step

asynchronous protocol:

Coupon collector: $n \ln n$ actions $= \ln n$ amount of time

Example: a path



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

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

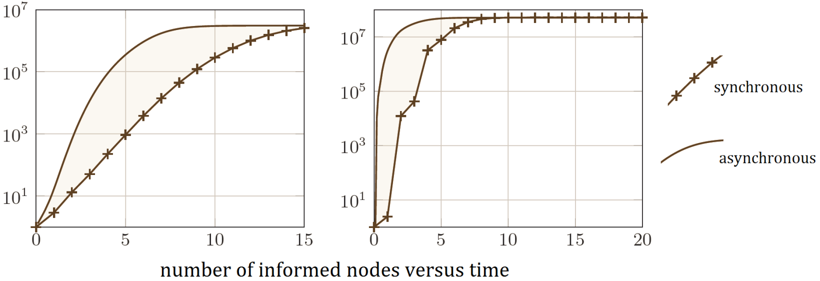
Some known results

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

Graph G	$s(G)$	$a(G)$
Star	2	$\ln n + O(1)$
Path	$(4/3)n + O(1)$	$n + O(1)$
Complete	$(1 + o(1)) \log_3 n$ [Karp,Schindelhauer,Shenker,Vöcking'00]	$\ln n + o(1)$ [Janson'99]
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]
General	$O(n)$	$\Omega(\ln n),\ O(n)$
	[Acan, Collevecchio, M., Wormald'15]	

Comparison of the two variants

Comparison of the two protocols: experiments

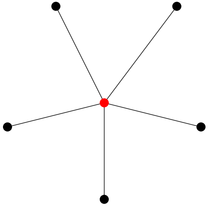


(a) Orkut network **(b)** Twitter network

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

The star

In which graph synchronous is quicker than asynchronous?

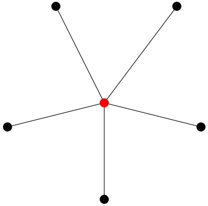


synchronous protocol: 1

asynchronous protocol: $\ln n$

The star

In which graph synchronous is quicker than asynchronous?



synchronous protocol: 1

asynchronous protocol: $\ln n$

For any G,

$a(G) \leq O(s(G) \times \ln n)$ [Acan, Collevecchio, M., Wormald'15]

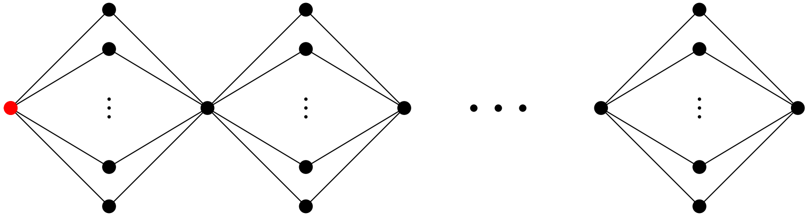
$a(G) \leq O(s(G) + \ln n)$ [Giakkoupis, Nazari, and Woelfel'16]

The string of diamonds

In which graph asynchronous is much quicker than synchronous?

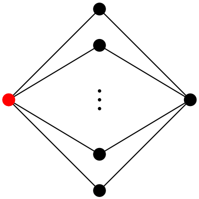
The string of diamonds

In which graph asynchronous is much quicker than synchronous?



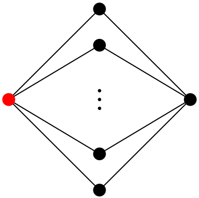
logarithmic $\ll$ polynomial

Time taken to pass through a diamond



k paths of length 2

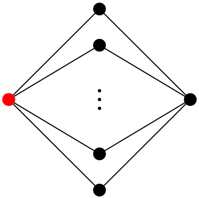
Time taken to pass through a diamond



k paths of length 2

Birthday paradox: $O(\sqrt{k})$ actions needed to have a vertex do two actions.

Time taken to pass through a diamond



k paths of length 2

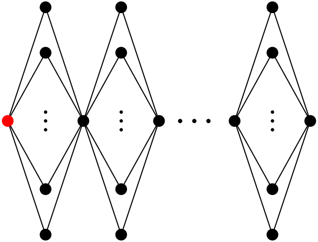
Birthday paradox: $O(\sqrt{k})$ actions needed to have a vertex do two actions.

Time to pass the rumor

Asynchronous: $\leq O(\sqrt{k}/k) = O(1/\sqrt{k})$

Synchronous: ≥ 2

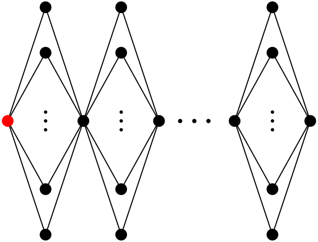
The string of diamonds, continued



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

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

The string of diamonds, continued



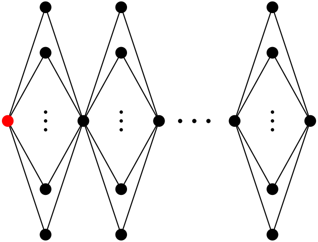
$n^{1/3}$ diamonds, each consisting of $n^{2/3}$ paths of length 2

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

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

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

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 protocols: our results

For any G,

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

[Acan, Collevecchio, M., Wormald'15]

Comparison of the protocols: our results

For any G,

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

[Acan, Collevecchio, M., Wormald'15]

$$\frac{s(G)}{a(G)} = O\left(n^{1/2}\right)$$

[Giakkoupis, Nazari, and Woelfel'16]

Comparison of the protocols: our results

For any G,

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

[Acan, Collevecchio, M., Wormald'15]

$$\frac{s(G)}{a(G)} = O\left(n^{1/2}\right)$$

[Giakkoupis, Nazari, and Woelfel'16]

Theorem (Angel, M., Peres'17)

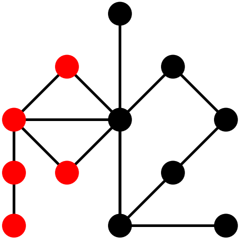
We have

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

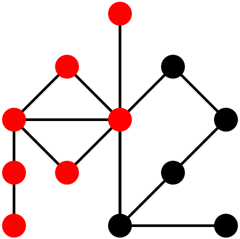
which is tight (up to a logarithmic factor).

Proof sketch for $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

Build a coupling so that



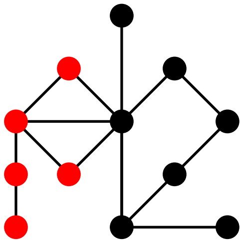
asynchronous contamination
by time 1



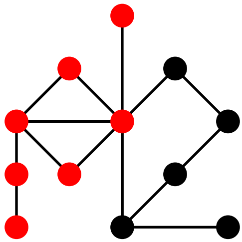
synchronous contamination
by time x

Proof sketch for $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

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 $s(G) \le a(G) \times \widetilde{O}\left(n^{1/3}\right)$

Lemma

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

Proof sketch for $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

Lemma

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

For fixed path $v_1 v_2 \dots v_L$, 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 $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

Lemma

In asynchronous, after one time unit, rumor 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 $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

Lemma

In asynchronous, after one time unit, rumor 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

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

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

Proof sketch for $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

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 $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

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 $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

Want to show

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

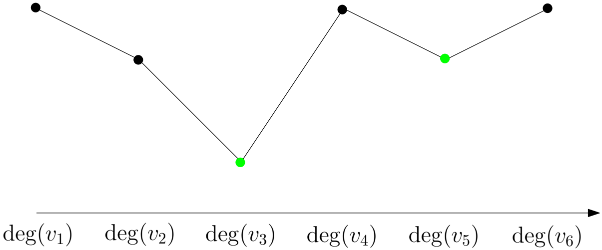
Consider the local minima vertices in the sequence $\deg(v_1), \deg(v_2), \ldots, \deg(v_L)$

Proof sketch for $s(G) \leq a(G) \times \widetilde{O}\left(n^{1/3}\right)$

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 $s(G) \le a(G) \times \widetilde{O}\left(n^{1/3}\right)$

Want to show

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

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

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

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

Lemma

In asynchronous, during $[0, t]$, rumor does not pass along a path of length $> Cn^{1/3}t^{2/3}$ (with high prob).

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

Lemma

In asynchronous, during $[0, t]$, rumor does not pass along a path of length $> Cn^{1/3}t^{2/3}$ (with high prob).

Let s be starting vertex. Observe there are independent exponential random variables $Y_{x,y}$:

$$A = \text{asynchronous spread time} = \max_{v \in V} \min_{\Gamma:(s,v)\text{-path}} \sum_{xy \in E(\Gamma)} \min\{Y_{x,y}, Y_{y,x}\}.$$

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

Lemma

In asynchronous, during $[0, t]$, rumor does not pass along a path of length $> Cn^{1/3}t^{2/3}$ (with high prob).

Let s be starting vertex. Observe there are independent exponential random variables $Y_{x,y}$:

$$A = \text{asynchronous spread time} = \max_{v \in V} \min_{\Gamma:(s,v)\text{-path}} \sum_{xy \in E(\Gamma)} \min\{Y_{x,y}, Y_{y,x}\}.$$

Similarly, there are non-independent geometric random variables $T_{x,y}$:

$$S = \text{synchronous spread time} = \max_{v \in V} \min_{\Gamma:(s,v)\text{-path}} \sum_{xy \in E(\Gamma)} \min\{T_{x,y}, T_{y,x}\}.$$

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

Lemma

In asynchronous, during $[0, t]$, rumor does not pass along a path of length $> Cn^{1/3}t^{2/3}$ (with high prob).

Let s be starting vertex. Observe there are independent exponential random variables $Y_{x,y}$:

$$A = \text{asynchronous spread time} = \max_{v \in V} \min_{\Gamma:(s,v)\text{-path}} \sum_{xy \in E(\Gamma)} \min\{Y_{x,y}, Y_{y,x}\}.$$

Similarly, there are non-independent geometric random variables $T_{x,y}$:

$$S = \text{synchronous spread time} = \max_{v \in V} \min_{\Gamma:(s,v)\text{-path}} \sum_{xy \in E(\Gamma)} \min\{T_{x,y}, T_{y,x}\}.$$

Fortunately, can couple them with independent exponentials $X_{x,y}$ s.t. $T_{x,y} \leq \ln n + X_{x,y}$, so

$$S \leq \max_{v \in V} \min_{\Gamma:(s,v)\text{-path}} \sum_{xy \in E(\Gamma)} (\ln n + \min\{X_{x,y}, X_{y,x}\}) \leq A^{2/3} n^{1/3} \times \ln n + A.$$

Conclusion

$s(G) :=$ expected spread time in G in synchronous time model
$a(G) :=$ expected spread time in G in asynchronous time model

Theorem (Angel, M., Peres'17)

For any connected G on n vertices,

$$\frac{s(G)}{a(G)} = O\left(n^{1/3} \ln^{2/3} n\right)$$

For any n there exists G for which

$$\frac{s(G)}{a(G)} = \Omega\left(n^{1/3} \ln^{-1/3} n\right)$$

THANKS!