The push&pull protocol for rumour spreading

Abbas Mehrabian

Simons Institute for Theory of Computing, UC Berkeley

University of Waterloo

10 February 2017

Co-authors

Nick Wormald

Huseyin Acan

Omer Angel

Petra Berenbrink

Andrea Collvecchio

Nick Harvey

Chris Liaw

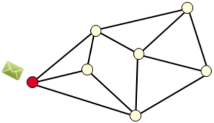
Peter Kling

Yuval Peres

Part I: Rumour spreading

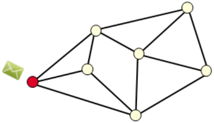
Example

ROUND 0



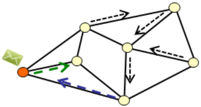
Example

ROUND 0



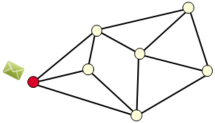
In each round, every vertex calls a random neighbour

ROUND 1

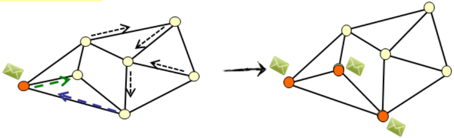


Example

ROUND 0



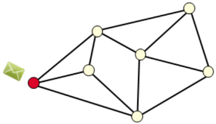
ROUND 1



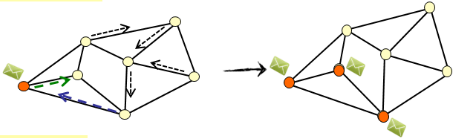
In each round, every vertex calls a random neighbour and they exchange their information

Example

ROUND 0

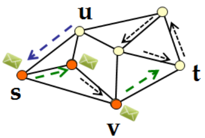


ROUND 1



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

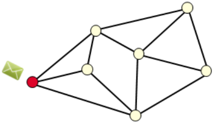
ROUND 2



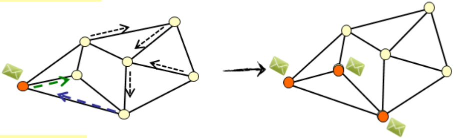
u pulls from s
v pushes to t

Example

ROUND 0

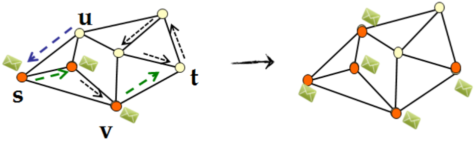


ROUND 1



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 rumour spreading protocol

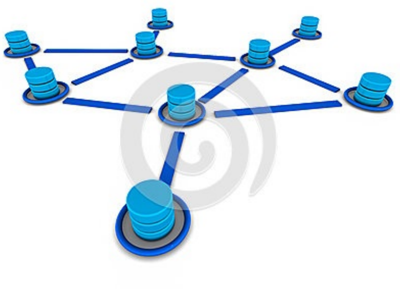
[Demers, Gealy, Greene, Hauser, Irish, Larson, Manning, Shenker, Sturgis, Swinehart, Terry, Woods'87]

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).

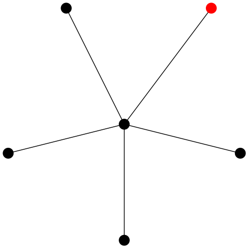
We are interested in the spread time.

Applications

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

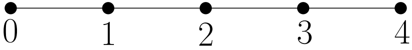


Example: a star



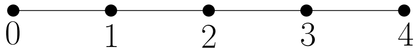
2 rounds

Example: path graph



vertex 0 knows rumour at round 0

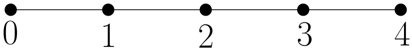
Example: path graph



vertex 0 knows rumour at round 0

vertex 1 is informed at round 1

Example: path graph



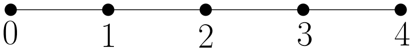
vertex 0 knows rumour at round 0

vertex 1 is informed at round 1

vertex 2 is informed at round

$1 + \min\{\mathrm{Geo}(1/2), \mathrm{Geo}(1/2)\} = 1 + \mathrm{Geo}(3/4)$

Example: path graph



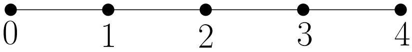
vertex 0 knows rumour at round 0

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 $1 + \mathrm{Geo}(3/4) + \mathrm{Geo}(3/4)$

Example: path graph



vertex 0 knows rumour at round 0

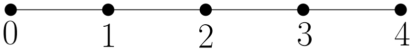
vertex 1 is informed at round 1

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

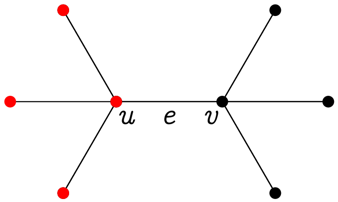
vertex 2 is informed at round
$1 + \min\{\mathrm{Geo}(1/2), \mathrm{Geo}(1/2)\} = 1 + \mathrm{Geo}(3/4)$

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

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

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

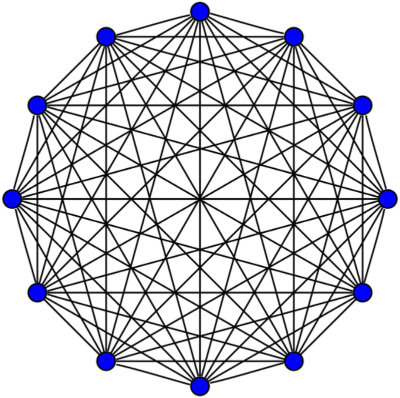
An example: double star



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

Expected spread time $\sim n/4$

Example: a complete graph



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

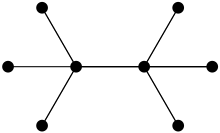
Known results

$s(G)$ 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?



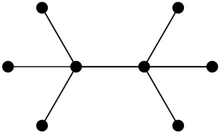
$n/4$

$4n/3$

$O(n \log n)$ 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 $1/n$.

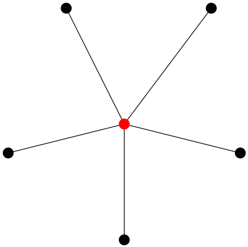
A (more realistic) variant

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 $1/n$.

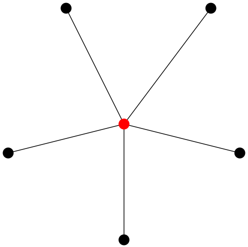
Almost equivalent definition:
every vertex has an exponential clock with rate 1,
at each clock ring, performs one action.

Example: a star



synchronous protocol: 1 round

Example: a star



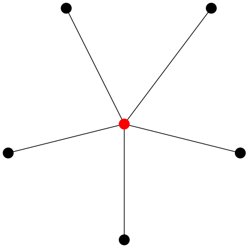
synchronous protocol: 1 round

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

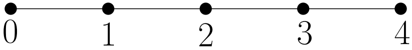
Coupon collector: Consider a bag containing n different balls.

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

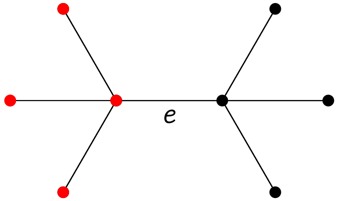
Example: a path



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

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

An example: double star



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

Expected spread time $\sim n/4$

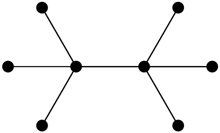
Some known results

$a(G)$ 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?

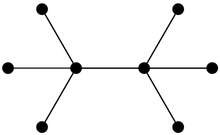


$\Omega(n)$

$\Omega(n)$

The extremal question

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



$\Omega(n)$

$\Omega(n)$

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

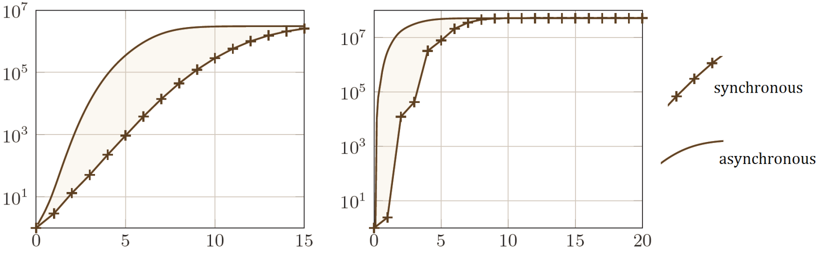
For any connected G on n vertices

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

Only pull operations are needed!

Comparison of the two variants

Comparison of the two protocols on the same graph: experiments



number of informed nodes versus time

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

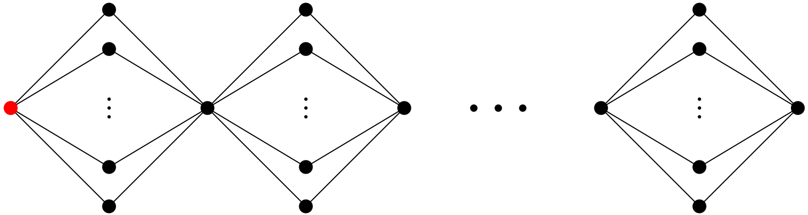
Figures from: Doerr, Fouz, and Friedrich'12.

The string of diamonds

In which graph asynchronous is much quicker than synchronous?

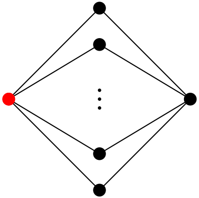
The string of diamonds

In which graph asynchronous is much quicker than synchronous?



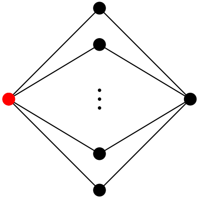
logarithmic $\ll$ polynomial

Time taken to pass through a diamond



k paths of length 2

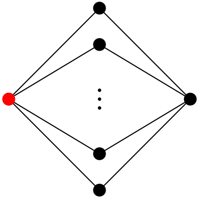
Time taken to pass through a diamond



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?

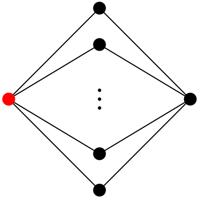
Time taken to pass through a diamond



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? $\sqrt{\pi k/2} \approx 1.25\sqrt{k}$

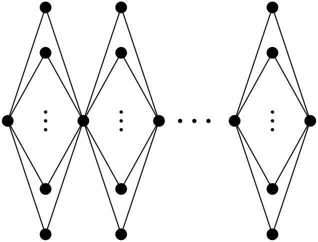
Time taken to pass through a diamond



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

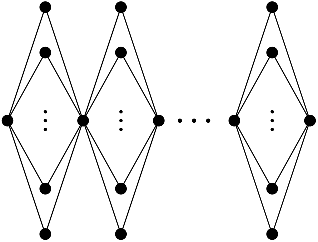
The string of diamonds, continued



$n^{1/3}$ 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



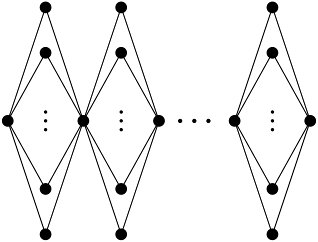
$n^{1/3}$ diamonds, each consisting of $n^{2/3}$ paths of length 2

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

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+)

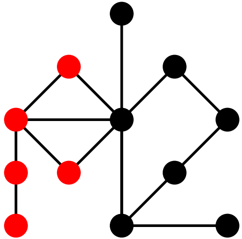
We have

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

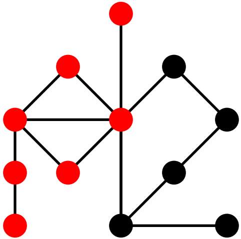
which is tight.

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

Build a coupling so that



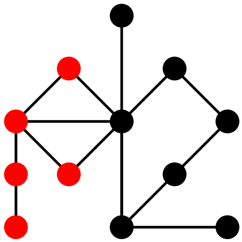
asynchronous contamination
by time 1



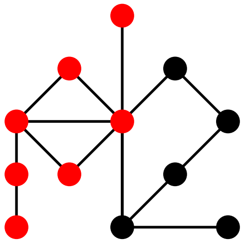
synchronous contamination
by time x

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

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

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

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

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 $v_1 v_2 \dots v_L$, 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

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

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

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 $v_1 v_2 \dots v_L$, this probability is

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

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

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

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

Want to show

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

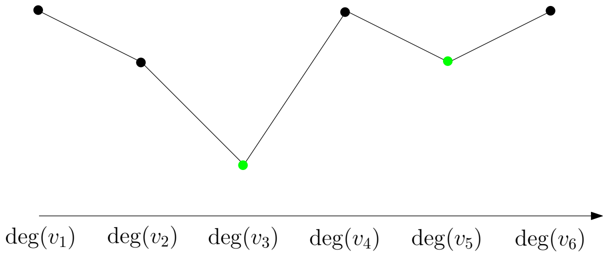
Consider the local minima vertices in the sequence $\deg(v_1), \deg(v_2), \ldots, \deg(v_L)$

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

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

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

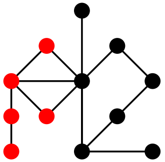
$$\sum_{L-paths} \prod_{i=1}^{L-1} \frac{1}{\min\{\deg(v_i), \deg(v_{i+1})\}} \le \sum_{s=0}^{L/2} \binom{L}{s} \cdot \binom{n}{s} \le (Cn/L)^{L/2}$$

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

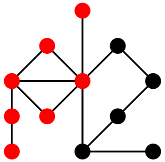
Lemma

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 $Cn^{1/3}$

$$s(G) \leq a(G) \times Cn^{1/3}$$

Summary of our results on push&pull

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

For any connected G on n vertices,

$$s(G) < 5n$$

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

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

All bounds are tight, up to constant factors.

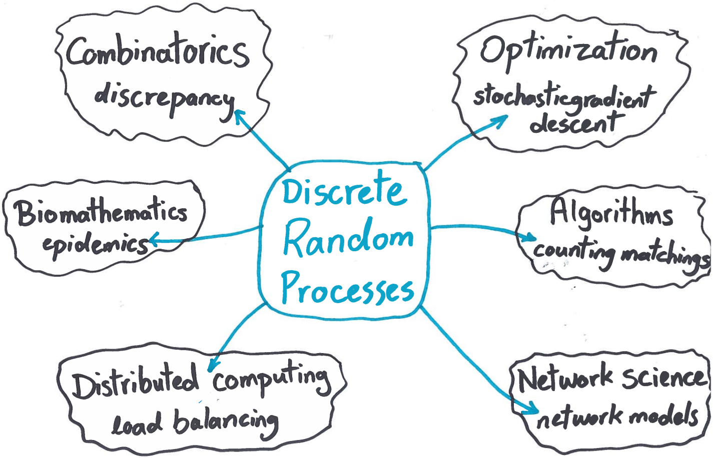
Further directions

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].

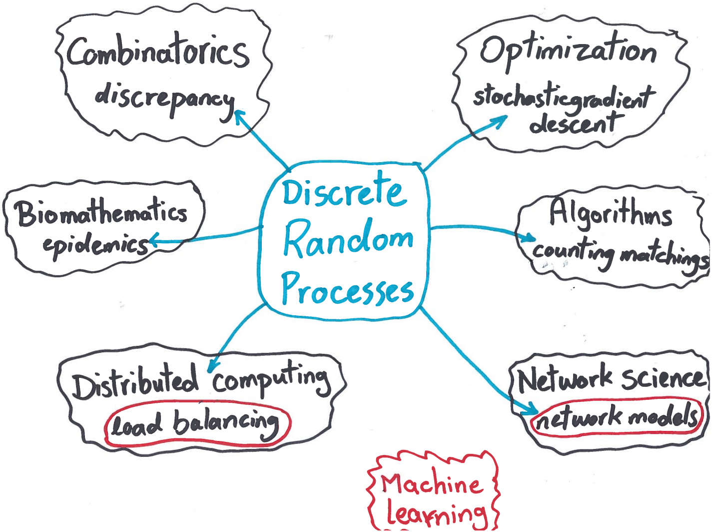
Part II: Broader overview of my research

discrete random processes

Applications of discrete random processes

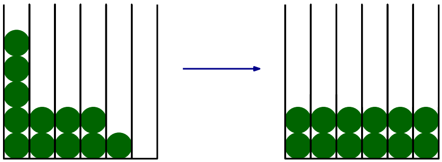


Applications of discrete random processes

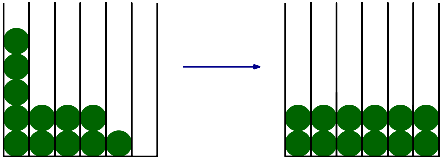


Load Balancing

Load balancing



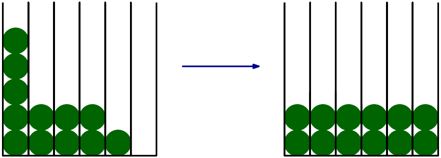
Load balancing



Definition (Randomized local search)

Each ball has an exponential clock of rate 1. When the clock rings, the ball is activated.
On activation, the ball chooses a random bin and moves there if its own load is improved by doing so.

Load balancing



Definition (Randomized local search)

Each ball has an exponential clock of rate 1. When the clock rings, the ball is activated.
On activation, the ball chooses a random bin and moves there if its own load is improved by doing so.

n = number of bins, m = number of balls

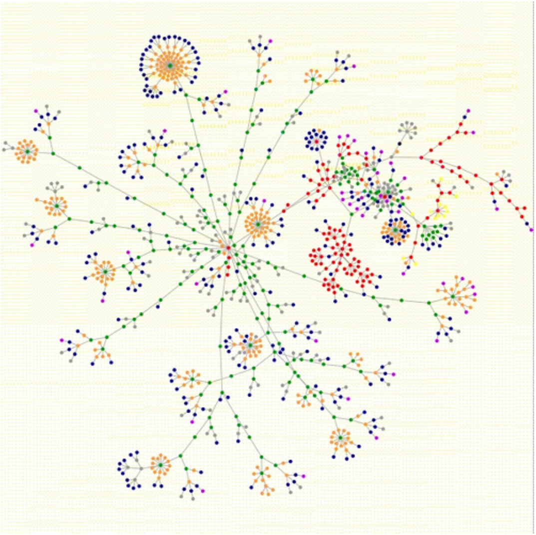
$O(n^2)$ Bound on expected time to reach perfect balance [Goldberg'04]

$O\left(\ln(n)^2 + \ln(n)n^2/m\right)$ [Ganesh et al.'12]

$O(\ln n + n^2/m)$ (tight!) [Berenbrink, Kling, Liaw, M'16]

Diameters of complex networks

Diameters of complex networks



Facebook graph in May 2011 had 700 million vertices, diameter 41

Diameters of complex networks

Our contribution: a technique for proving certain random graph models have diameter at most $O(\log n)$.

Diameters of complex networks

Our contribution: a technique for proving certain random graph models have diameter at most $O(\log n)$.

Main idea: Embed a random recursive tree

Definition (random recursive tree)

Initially we have a single node; in every round a uniformly random node gives birth to a new child.

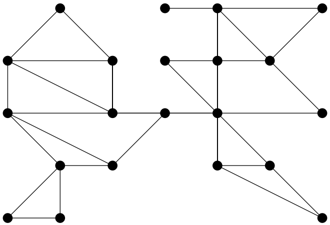
Diameters of complex networks

Our contribution: a technique for proving certain random graph models have diameter at most $O(\log n)$.

Main idea: Embed a random recursive tree

Definition (random recursive tree)

Initially we have a single node; in every round a uniformly random node gives birth to a new child.



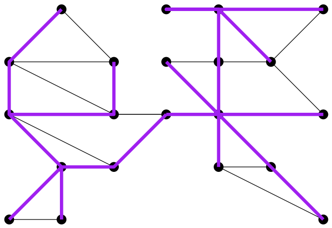
Diameters of complex networks

Our contribution: a technique for proving certain random graph models have diameter at most $O(\log n)$.

Main idea: Embed a random recursive tree

Definition (random recursive tree)

Initially we have a single node; in every round a uniformly random node gives birth to a new child.



New results using our approach

Theorem (M'14)

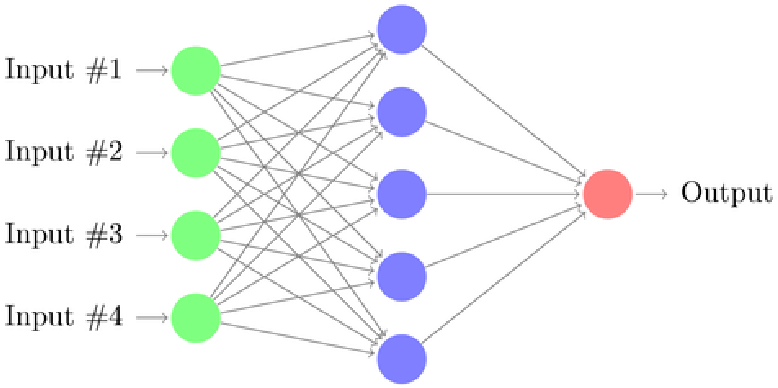
These random graph models have diameter $O(\log n)$ with high probability:

1. The (edge) copying model [Kumar,Raghavan,Rajagopalan,Sivakumar,Tomkins,Upfal'00]
2. Aiello-Chung-Lu models [Aiello,Chung,Lu'01]
3. The Cooper-Frieze model [Cooper,Frieze'01]
4. The generalized linear preference model [Bu,Towsley'02]
5. The PageRank-based selection model [Pandurangan,Raghavan,Upfal'02]
6. Directed scale-free graphs [Bollobás,Borgs,Chayes,Riordan'03]
7. The forest fire model [Leskovec,Kleinberg,Faloutsos'05]

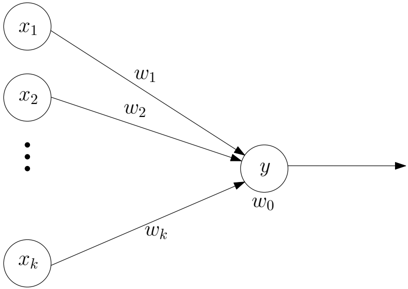
VC-dimension of neural networks

a somewhat different problem

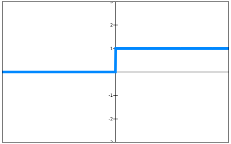
Artificial Neural networks



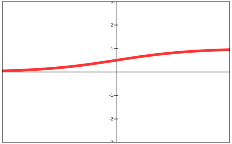
Artificial Neural networks



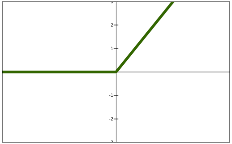
$$y = \sigma(w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_k x_k)$$



threshold



sigmoid



ReLU

How many data points is needed to learn?

How many data points is needed to learn?

Theorem (fundamental theorem of statistical learning (Vapnik, Chervonenkis'71))

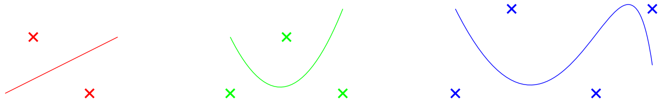
The number of samples required for PAC learning within error ε in a model with VC-dimension v is essentially $\Theta(v/\varepsilon)$.

How many data points is needed to learn?

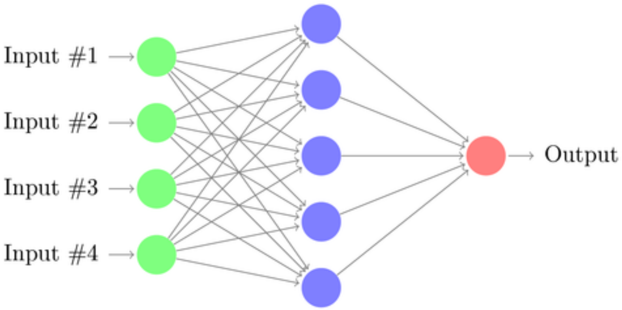
Theorem (fundamental theorem of statistical learning (Vapnik, Chervonenkis'71))

The number of samples required for PAC learning within error ε in a model with VC-dimension v is essentially $\Theta(v/\varepsilon)$.

e.g. VC-dimension of polynomials of degree d is $d+1$.



VC-dimension of neural networks



$v(e, \ell) :=$ maximum VC-dimension of a neural network

e edges

ℓ layers

VC-dimension of neural networks

If the activation function is piecewise polynomial,

$$v(e, \ell) \le Ce^2 \qquad \text{[Goldberg, Jerrum'95]}$$

$$ce\ell \le v(e, \ell) \le Ce\ell^2 \qquad \text{[Bartlett, Maiorov, Meir'98]}$$

VC-dimension of neural networks

If the activation function is piecewise polynomial,

$$v(e,\ell) \le Ce^2 \qquad \text{[Goldberg, Jerrum'95]}$$

$$ce\ell \le v(e,\ell) \le Ce\ell^2 \qquad \text{[Bartlett, Maiorov, Meir'98]}$$

Theorem (Harvey, Liaw, M'16)

If the activation function is piecewise linear,

$$ce\ell \log(e/\ell) \le v(e,\ell) \le Ce\ell \log e$$

VC-dimension of neural networks

If the activation function is piecewise polynomial,

$$v(e,\ell) \le Ce^2 \qquad \text{[Goldberg, Jerrum'95]}$$

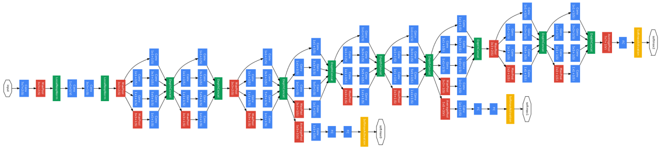
$$ce\ell \le v(e,\ell) \le Ce\ell^2 \qquad \text{[Bartlett, Maiorov, Meir'98]}$$

Theorem (Harvey, Liaw, M'16)

If the activation function is piecewise linear,

$$ce\ell\log(e/\ell) \le v(e,\ell) \le Ce\ell\log e$$

GoogleNet'14: $\ell = 41$, $e = 7$ million



Further directions

1. What is the effect of depth on representation power of neural networks?
2. Why stochastic gradient descent "works" for training neural networks in practice although the objective function is non-convex?

Further directions

1. What is the effect of depth on representation power of neural networks?
2. Why stochastic gradient descent "works" for training neural networks in practice although the objective function is non-convex?