Rumour spreading in the spatial preferential attachment model

Abbas Mehrabian

University of British Columbia

Banff, 2016

joint work with Jeannette Janssen

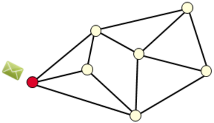
The push&pull rumour spreading protocol

[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 $1, 2, \dots,$
   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

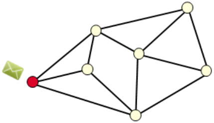


Push-Pull Protocol

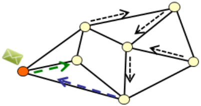
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

ROUND 0



ROUND 1

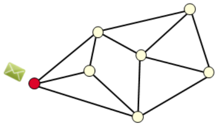


Push-Pull Protocol

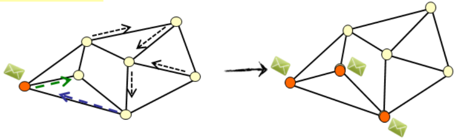
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

ROUND 0



ROUND 1

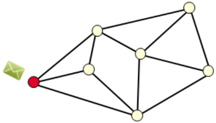


Push-Pull Protocol

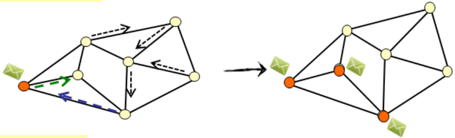
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

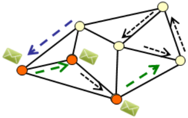
ROUND 0



ROUND 1



ROUND 2

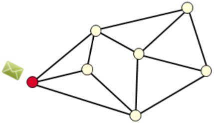


Push-Pull Protocol

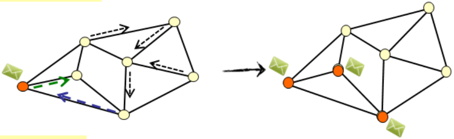
Each node contacts a random neighbor:
Node pushes the rumor (if knows);
and pulls otherwise

Example

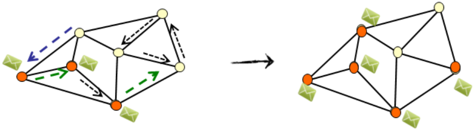
ROUND 0



ROUND 1



ROUND 2



Push-Pull Protocol

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

Some known results

Has lots of applications in distributed computing.

Some known results

Has lots of applications in distributed computing.

Spread Time: the first time everyone knows the rumour, assuming the worst starting vertex.

Some examples

- ✓ A star graph has spread time 2.

Some known results

Has lots of applications in distributed computing.

Spread Time: the first time everyone knows the rumour, assuming the worst starting vertex.

Some examples

- ✓ A star graph has spread time 2.
- ✓ A.a.s. an n-vertex path has spread time $\Theta(n)$.

Some known results

Has lots of applications in distributed computing.

Spread Time: the first time everyone knows the rumour, assuming the worst starting vertex.

Some examples

- ✓ A star graph has spread time 2.
- ✓ A.a.s. an n-vertex path has spread time $\Theta(n)$.
- ✓ A.a.s. an n-vertex complete graph has spread time $(1 + o(1)) \log_3(n)$ [Karp,Schindelhauer,Shenker,Vöcking'00]

Some known results

Has lots of applications in distributed computing.
Spread Time: the first time everyone knows the rumour, assuming the worst starting vertex.

Some examples

- ✓ A star graph has spread time 2.
- ✓ A.a.s. an n-vertex path has spread time $\Theta(n)$.
- ✓ A.a.s. an n-vertex complete graph has spread time $(1 + o(1)) \log_3(n)$ [Karp,Schindelhauer,Shenker,Vöcking'00]
- ✓ Many random graph models have been analyzed, including Erdős-Rényi model, random regular graphs, random geometric graphs, Barabási-Albert model, Chung-Lu model. In all of them spread time is Θ(diameter).

Some known results

Has lots of applications in distributed computing.
Spread Time: the first time everyone knows the rumour, assuming the worst starting vertex.

Some examples

- ✓ A star graph has spread time 2.
- ✓ A.a.s. an n-vertex path has spread time $\Theta(n)$.
- ✓ A.a.s. an n-vertex complete graph has spread time $(1 + o(1)) \log_3(n)$ [Karp,Schindelhauer,Shenker,Vöcking'00]
- ✓ Many random graph models have been analyzed, including Erdős-Rényi model, random regular graphs, random geometric graphs, Barabási-Albert model, Chung-Lu model. In all of them spread time is $\Theta(\text{diameter})$.
- ✓ For a survey, see "On the push&pull rumour spreading protocol," [Acan, Collevecchio, Mehrabian, Wormald'15]

Connections with the diameter

- ✓ Clearly, spread time $\geq$ diameter

Connections with the diameter

- ✓ Clearly, spread time $\geq$ diameter
- ✓ Easy to see a.a.s. spread time $\leq 6\Delta$ (diameter $+ \log n$)

  [Feige,Peleg,Raghavan,Upfal'90]

Connections with the diameter

- ✓ Clearly, spread time $\geq$ diameter
- ✓ Easy to see a.a.s. spread time $\leq 6\Delta$ (diameter $+ \log n$)
  [Feige,Peleg,Raghavan,Upfal'90]
- ✓ In a (vertex-wise) expander graph, a.a.s.
  spread time $\leq C \log^2 n \leq C \log^2 n \times$ diameter [Giakkoupis'14]

Connections with the diameter

- ✓ Clearly, spread time $\geq$ diameter
- ✓ Easy to see a.a.s. spread time $\leq 6\Delta\,(\text{diameter} + \log n)$

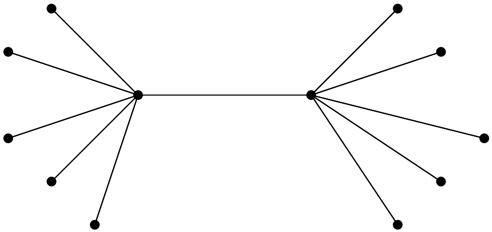
  [Feige,Peleg,Raghavan,Upfal'90]
- ✓ In a (vertex-wise) expander graph, a.a.s.
  spread time $\leq C \log^2 n \leq C \log^2 n \times$ diameter [Giakkoupis'14]

Question: is there a graph with tiny diameter and huge spread time?

Connections with the diameter

- ✓ Clearly, spread time $\geq$ diameter
- ✓ Easy to see a.a.s. spread time $\leq 6\Delta$ (diameter + $\log n$)
  [Feige,Peleg,Raghavan,Upfal'90]
- ✓ In a (vertex-wise) expander graph, a.a.s.
  spread time $\leq C \log^2 n \leq C \log^2 n \times$ diameter [Giakkoupis'14]

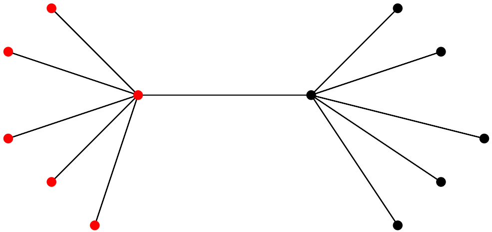
Question: is there a graph with tiny diameter and huge spread time?



Connections with the diameter

- ✓ Clearly, spread time $\geq$ diameter
- ✓ Easy to see a.a.s. spread time $\leq 6\Delta$ (diameter + $\log n$) [Feige-Peleg-Raghavan-Upfal'90]
- ✓ In a (vertex-wise) expander graph, a.a.s. spread time $\leq C \log^2 n \leq C \log^2 n \times$ diameter [Giakkoupis'14]

Question: is there a graph with tiny diameter and huge spread time?

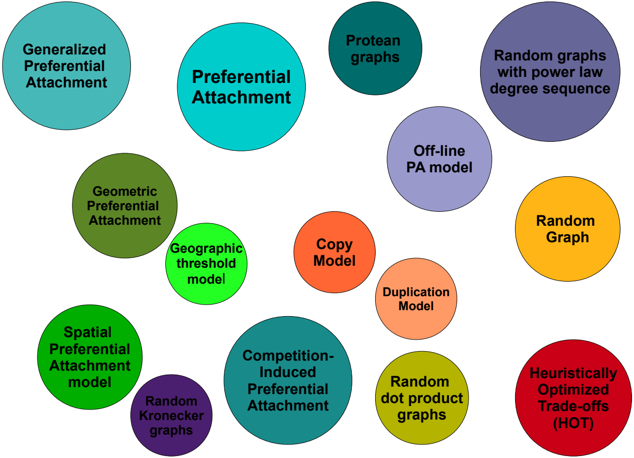


Our main contribution

is to show that a certain model for complex networks
has polylogarithmic diameter but polynomial spread time

Our main contribution

is to show that a certain model for complex networks has polylogarithmic diameter but polynomial spread time



Spatial Preferential Attachment (SPA) model

Evolving model for generating random directed graphs
Introduced by Aiello, Bonato, Cooper, Janssen, and Prałat'08.

Spatial Preferential Attachment (SPA) model

Evolving model for generating random directed graphs
Introduced by Aiello, Bonato, Cooper, Janssen, and Prałat'08.

- ✓ Parameters $A_1, A_2 > 0$.
- ✓ Vertices are embedded uniformly at random in $[0,1]^d$.

Spatial Preferential Attachment (SPA) model

Evolving model for generating random directed graphs
Introduced by Aiello, Bonato, Cooper, Janssen, and Prałat'08.

- ✓ Parameters $A_1, A_2 > 0$.
- ✓ Vertices are embedded uniformly at random in $[0,1]^d$.
- ✓ At a given time t, each vertex v has a sphere of influence $S(v,t)$ centered at v with volume

$$\operatorname{vol} S(v,t) = \frac{A_1 \times \text{in-deg}(v,t) + A_2}{t}$$

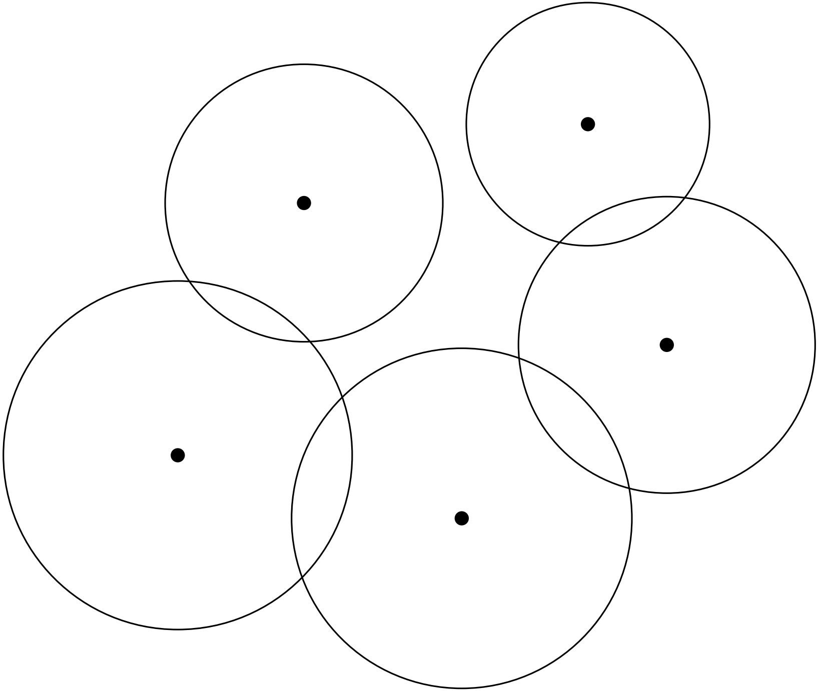
Spatial Preferential Attachment (SPA) model

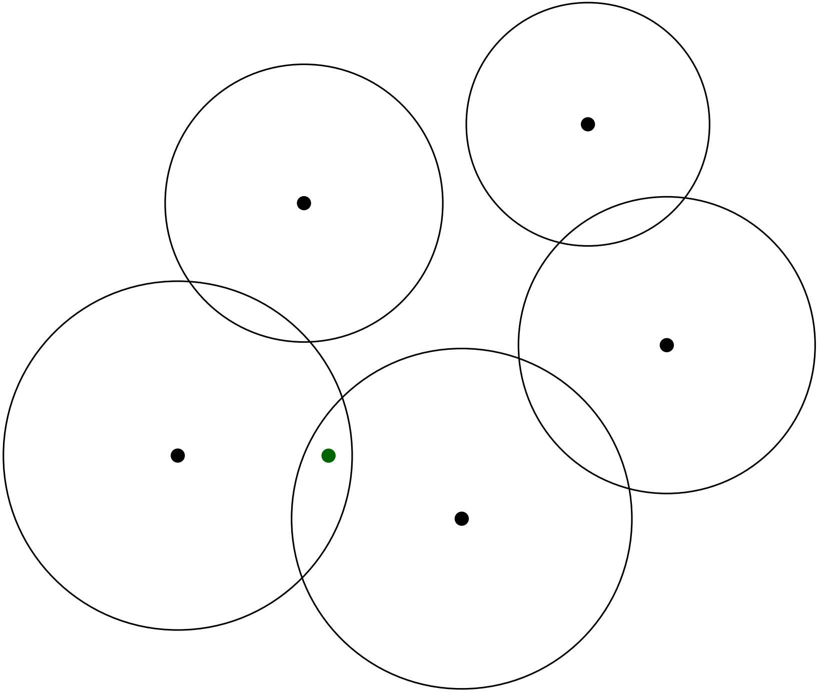
Evolving model for generating random directed graphs
Introduced by Aiello, Bonato, Cooper, Janssen, and Prałat'08.

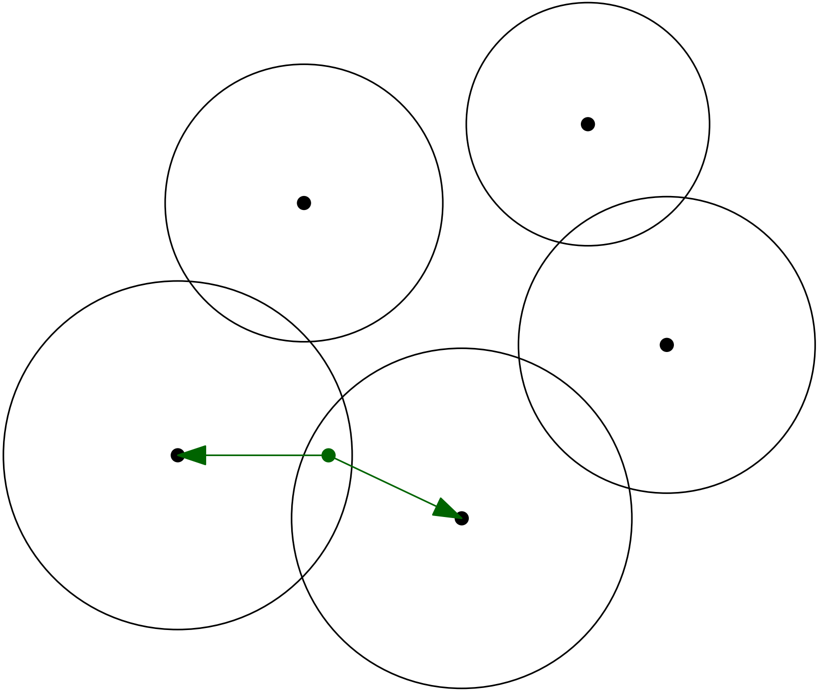
- ✓ Parameters $A_1, A_2 > 0$.
- ✓ Vertices are embedded uniformly at random in $[0,1]^d$.
- ✓ At a given time t, each vertex v has a sphere of influence $S(v,t)$ centered at v with volume

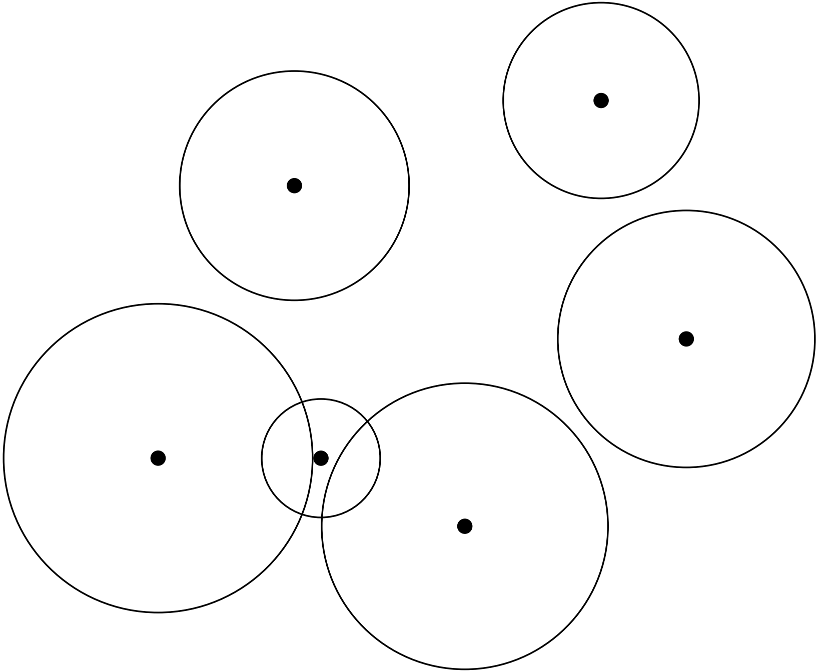
$$\mathrm{vol}\, S(v,t) = \frac{A_1 \times \text{in-deg}(v,t) + A_2}{t}$$

- ✓ In each time-step a new vertex is born.
- ✓ The new vertex links to an existing vertex v if it falls within the sphere of influence of v.









Implicit preferential attachment

The positions of the vertices are uniformly random,
the space has volume 1, so the probability that v receives a link at time t is

$$\frac{\operatorname{vol} S(v,t)}{1} = \frac{A_1 \times \text{in-deg}(v,t) + A_2}{t}$$

Implicit preferential attachment

The positions of the vertices are uniformly random, the space has volume 1, so the probability that v receives a link at time t is

$$\frac{\operatorname{vol} S(v,t)}{1} = \frac{A_1 \times \text{in-deg}(v,t) + A_2}{t}$$

- ✓ Rich get richer: with every link v receives, its sphere of influence grows and thus its link probability increases.

Implicit preferential attachment

The positions of the vertices are uniformly random,
the space has volume 1, so the probability that v receives a link at time t is

$$\frac{\mathrm{vol}\, S(v,t)}{1} = \frac{A_1 \times \text{in-deg}(v,t) + A_2}{t}$$

- ✓ Rich get richer: with every link v receives, its sphere of influence grows and thus its link probability increases.
- ✓ At time t, our graph contains a random geometric graph with t vertices and radius $c_d(A_2/t)^{1/d}$.

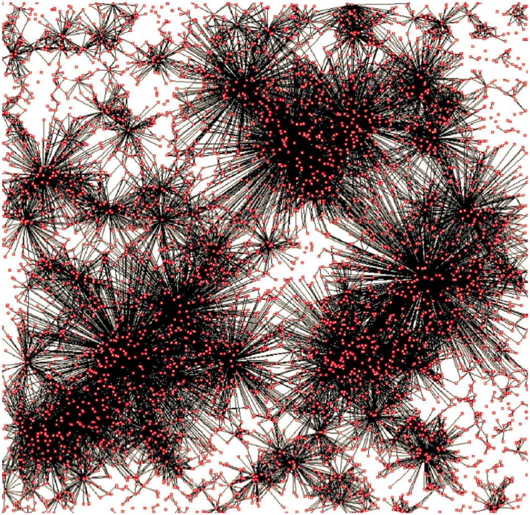
Implicit preferential attachment

The positions of the vertices are uniformly random, the space has volume 1, so the probability that v receives a link at time t is

$$\frac{\operatorname{vol} S(v,t)}{1} = \frac{A_1 \times \text{in-deg}(v,t) + A_2}{t}$$

- ✓ Rich get richer: with every link v receives, its sphere of influence grows and thus its link probability increases.
- ✓ At time t, our graph contains a random geometric graph with t vertices and radius $c_d(A_2/t)^{1/d}$.
- ✓ Original model has parameter p, we consider $p = 1$ only.

SPA model: example in 2D



5000 vertices, $A_1 = A_2 = 1$ [Cooper,Frieze,Prałat'14]

Known results

- ✓ **Power law degree distribution.** If $A_1 < 1$, the in-degree distribution has a power law tail with exponent $1 + \frac{1}{A_1}$
  [Aiello,Bonato,Cooper,Janssen,Prałat'08]
- ✓ **Sparse graph.** If $A_1 < 1$, a.a.s. the average out-degree is $\frac{A_2}{1-A_1}$ [Aiello,Bonato,Cooper,Janssen,Prałat'08]
- ✓ **Small world.** A.a.s. any directed geodesic path has length $O(\log n)$ [Cooper,Frieze,Prałat'14]
- ✓ **Good clustering**

Known results

- ✓ **Power law degree distribution.** If $A_1 < 1$, the in-degree distribution has a power law tail with exponent $1 + \frac{1}{A_1}$
  [Aiello,Bonato,Cooper,Janssen,Prałat'08]
- ✓ **Sparse graph.** If $A_1 < 1$, a.a.s. the average out-degree is $\frac{A_2}{1-A_1}$ [Aiello,Bonato,Cooper,Janssen,Prałat'08]
- ✓ **Small world.** A.a.s. any directed geodesic path has length $O(\log n)$ [Cooper,Frieze,Prałat'14]
- ✓ **Good clustering**
- ✓ **Large maximum degree.** A.a.s. maximum total degree is $n^{\Omega(1)}$ [Aiello,Bonato,Cooper,Janssen,Prałat'08]
- ✓ **Not an expander.** A.a.s. the minimum bisection has size $o(n)$ [Cooper,Frieze,Prałat'14]

Our results

Consider the giant component of the undirected underlying graph generated by the SPA model;

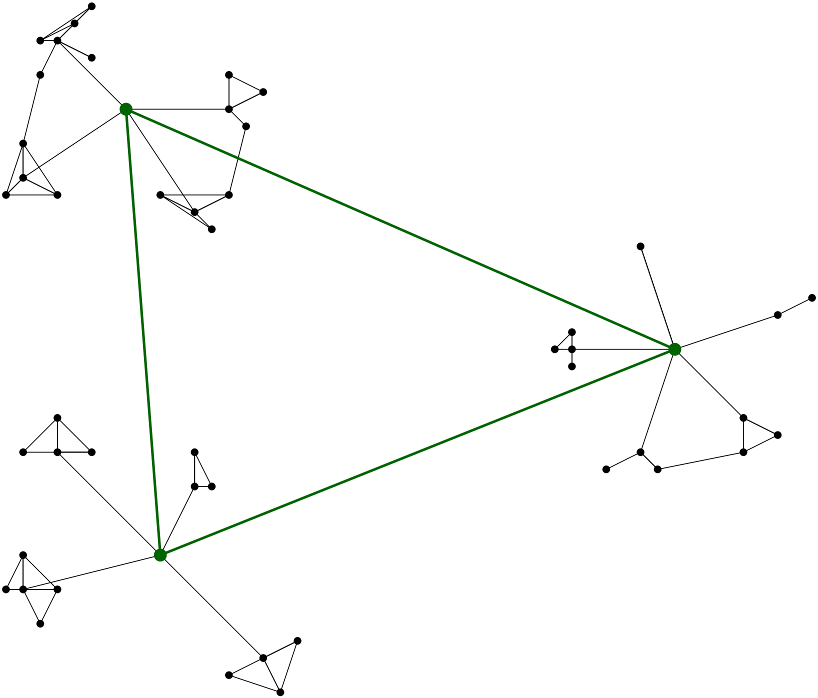
Theorem (upper bound for the diameter)

If A_2 is sufficiently large, a.a.s. the graph has diameter $O(\log^2 n)$.

Note: Result of Cooper-Frieze-Prałat does not apply, as they consider directed paths only.

Theorem (lower bound for rumour spreading)

A.a.s. it takes $n^{\Omega(1)}$ rounds to spread the rumour.



Proof sketch of the upper bound for diameter

Say n is a power of 2

1. $G_t :=$ subgraph induced by first t vertices
2. Since A_2 is large, $G_n, G_{n/2}, \ldots, G_1$ have giant components

Proof sketch of the upper bound for diameter

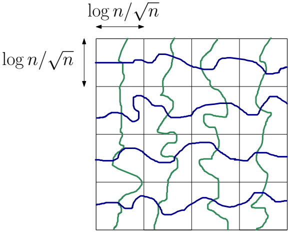
Say n is a power of 2

1. $G_t :=$ subgraph induced by first t vertices
2. Since A_2 is large, $G_n, G_{n/2}, \ldots, G_1$ have giant components
3. Idea: find a path of length $O(\log n)$ from arbitrary vertex v in giant of G_n to some vertex in giant of $G_{n/2}$, and iterate.

Proof sketch of the upper bound for diameter

Say n is a power of 2

1. $G_t :=$ subgraph induced by first t vertices
2. Since A_2 is large, $G_n, G_{n/2}, \ldots, G_1$ have giant components
3. Idea: find a path of length $O(\log n)$ from arbitrary vertex v in giant of G_n to some vertex in giant of $G_{n/2}$, and iterate.
   - $G_{n/2}$ contains a random geometric graph.
   - By known results on RGG's, any point not in the giant of $G_{n/2}$ is within $\log n/\sqrt{n}$ Euclidean distance to some vertex in the giant of $G_{n/2}$. [Ganesan'13]



Proof sketch of the upper bound for diameter

Say n is a power of 2

1. $G_t :=$ subgraph induced by first t vertices
2. Since A_2 is large, $G_n, G_{n/2}, \ldots, G_1$ have giant components
3. Idea: find a path of length $O(\log n)$ from arbitrary vertex v in giant of G_n to some vertex in giant of $G_{n/2}$, and iterate.
   - $G_{n/2}$ contains a random geometric graph.
   - By known results on RGG's, any point not in the giant of $G_{n/2}$ is within $\log n/\sqrt{n}$ Euclidean distance to some vertex in the giant of $G_{n/2}$. [Ganesan'13]
   - By known results on stretch factor of RGG's, this leads to a path of length $O(\log n)$ from v to some vertex in the giant of $G_{n/2}$. [Bradonjic, Elsasser, Friedrich, Sauerwald, Stauffer'13]

Upper bound for diameter (the catch)

$$G_n \supseteq \bigcup_{i=1}^{n} R_i,$$

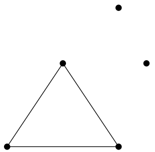
where each R_i is a RGG.

We showed diameter of giant of $\bigcup_{i=1}^{n} R_i = O(\log^2 n)$.

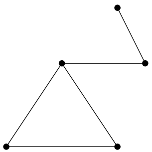
Upper bound for diameter (the catch)

$$G_n \supseteq \bigcup_{i=1}^{n} R_i,$$

where each R_i is a RGG.
We showed diameter of giant of $\bigcup_{i=1}^{n} R_i = O(\log^2 n)$.



diameter of giant$(\cup_{i=1}^{n} R_i) = 1$



diameter of giant$(G_n) = 3$

Upper bound for diameter (the catch)

$$G_n \supseteq \bigcup_{i=1}^{n} R_i,$$

where each R_i is a RGG. We showed diameter of giant of $\bigcup_{i=1}^{n} R_i = O(\log^2 n)$.
By [Penrose'03], 99% of vertices lie in the giant.

Theorem (Janssen, M'15)

In the SPA model graph with $p = 1$ and A_2 sufficiently large, a.a.s. 99% of vertices are within distance $O(\log^2 n)$ of each other, in dimension 2.

Further questions. . .

Proof sketch of the lower bound for rumour spreading

1. Categorize the edges into *short* and *long*, and prove that no *long* edge is used during the first $n^{O(1)}$ rounds.
2. Vertices that are born late, have small spheres of influence

$$\operatorname{vol} S(v, t) = \frac{A_1 \times \text{in-deg}(v, t) + A_2}{t}$$

so no incident *long* edges anyway.

Proof sketch of the lower bound for rumour spreading

1. Categorize the edges into *short* and *long*, and prove that no *long* edge is used during the first $n^{O(1)}$ rounds.
2. Vertices that are born late, have small spheres of influence

$$\text{vol}\, S(v,t) = \frac{A_1 \times \text{in-deg}(v,t) + A_2}{t}$$

   so no incident *long* edges anyway.
3. Vertices that are born early, accumulate a lot of neighbours, so the proportion of *long* incident edges is pretty small, so they tend not to send the rumour along such edges.

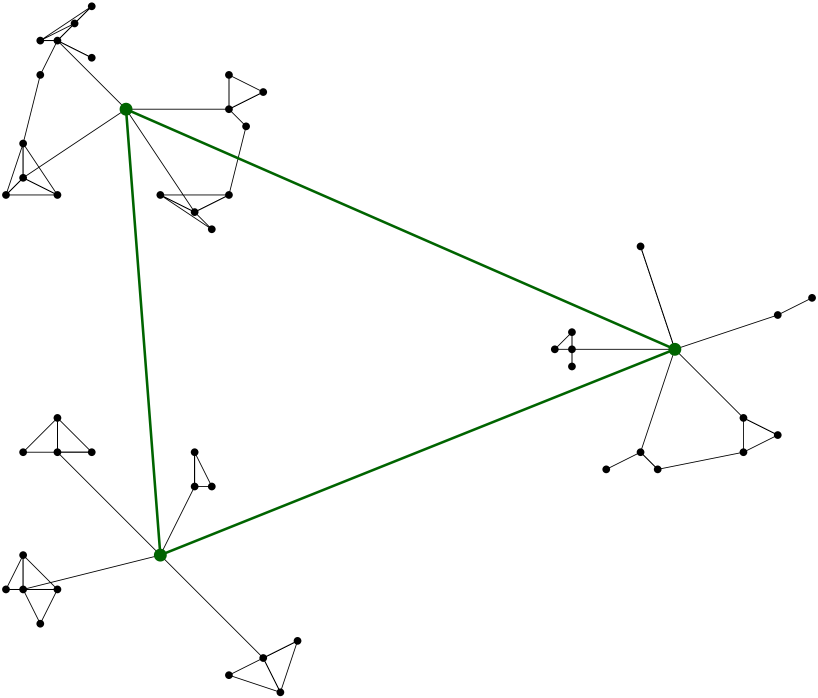
Proof sketch of the lower bound for rumour spreading

1. Categorize the edges into *short* and *long*, and prove that no *long* edge is used during the first $n^{O(1)}$ rounds.
2. Vertices that are born late, have small spheres of influence

   $$\mathrm{vol}\, S(v,t) = \frac{A_1 \times \text{in-deg}(v,t) + A_2}{t}$$

   so no incident *long* edges anyway.
3. Vertices that are born early, accumulate a lot of neighbours, so the proportion of *long* incident edges is pretty small, so they tend not to send the rumour along such edges.

Proof technique: Old and new concentration inequalities for vertices' degrees.



Wrap up

Consider the giant component of the undirected underlying graph generated by the SPA model

Theorem (Janssen, M'15)

If $p = 1$ and A_2 is sufficiently large, a.a.s. 99% of vertices are within distance $O(\log^2 n)$ of each other, in dimension 2.

Theorem (Janssen, M'15)

Suppose $pA_1 < 1$. If rumour starts from a random vertex, a.a.s. after n^α rounds, number of informed vertices is $o(n)$.

$\alpha = \frac{pA_1(1-pA_1)}{(3+pA_1)\times\text{dimension}+1-pA_1}$