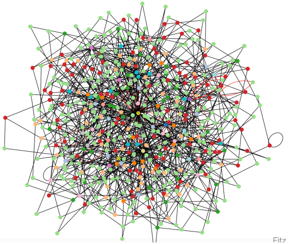
Evolution of distances in preferential attachment models

Júlia Komjáthy & Joost Jorritsma

A day on random graphs 30 June 2022



Properties of real contact networks

Most large real networks are *statistically very similar*. If you understand one, you understand others.

[Adapted from *Network Science* (2015) by Albert László Barabási]

Properties of real contact networks

Most large real networks are *statistically very similar*. If you understand one, you understand others.

- **Scale-free degree distributions**: node degree distribution has fat tail (e.g. 10% of nodes might have 90% of connections)

[Adapted from *Network Science* (2015) by Albert László Barabási]

Properties of real contact networks

Most large real networks are *statistically very similar*. If you understand one, you understand others.

- **Scale-free degree distributions**: node degree distribution has fat tail (e.g. 10% of nodes might have 90% of connections)
- **Small world phenomenon**: All nodes can be reached via a few hops through the network (e.g. six degrees of separation)

[Adapted from *Network Science* (2015) by Albert László Barabási]

Properties of real contact networks

Most large real networks are *statistically very similar*. If you understand one, you understand others.

- **Scale-free degree distributions**: node degree distribution has fat tail (e.g. 10% of nodes might have 90% of connections)
- **Small world phenomenon**: All nodes can be reached via a few hops through the network (e.g. six degrees of separation)
- **Clustering, Local communities**: nodes form strongly connected groups based on geographical locations (e.g. cities).

[Adapted from *Network Science* (2015) by Albert László Barabási]

Properties of real contact networks

Most large real networks are *statistically very similar*. If you understand one, you understand others.

- **Scale-free degree distributions**: node degree distribution has fat tail (e.g. 10% of nodes might have 90% of connections)
- **Small world phenomenon**: All nodes can be reached via a few hops through the network (e.g. six degrees of separation)
- **Clustering, Local communities**: nodes form strongly connected groups based on geographical locations (e.g. cities).
- **Hierarchical**: most nodes are connected to at least one node with more connections than they have themselves

[Adapted from *Network Science* (2015) by Albert László Barabási]

Properties of real contact networks

Most large real networks are *statistically very similar*. If you understand one, you understand others.

- **Scale-free degree distributions**: node degree distribution has fat tail (e.g. 10% of nodes might have 90% of connections)
- **Small world phenomenon**: All nodes can be reached via a few hops through the network (e.g. six degrees of separation)
- **Clustering, Local communities**: nodes form strongly connected groups based on geographical locations (e.g. cities).
- **Hierarchical**: most nodes are connected to at least one node with more connections than they have themselves
- **Geometry:** nodes have a fixed (or almost fixed) location in space

[Adapted from *Network Science* (2015) by Albert László Barabási]

Properties of real contact networks

Most large real networks are *statistically very similar*. If you understand one, you understand others.

- **Scale-free degree distributions**: node degree distribution has fat tail (e.g. 10% of nodes might have 90% of connections)
- **Small world phenomenon**: All nodes can be reached via a few hops through the network (e.g. six degrees of separation)
- **Clustering, Local communities**: nodes form strongly connected groups based on geographical locations (e.g. cities).
- **Hierarchical**: most nodes are connected to at least one node with more connections than they have themselves
- **Geometry:** nodes have a fixed (or almost fixed) location in space
- **Long-range connections**: long distance connections are relatively common (e.g. airplanes, intercontinental fiberoptic cables)

[Adapted from *Network Science* (2015) by Albert László Barabási]

Properties of real contact networks

Most large real networks are *statistically very similar*. If you understand one, you understand others.

- **Scale-free degree distributions**: node degree distribution has fat tail (e.g. 10% of nodes might have 90% of connections)
- **Small world phenomenon**: All nodes can be reached via a few hops through the network (e.g. six degrees of separation)
- **Clustering, Local communities**: nodes form strongly connected groups based on geographical locations (e.g. cities).
- **Hierarchical**: most nodes are connected to at least one node with more connections than they have themselves
- **Geometry:** nodes have a fixed (or almost fixed) location in space
- **Long-range connections**: long distance connections are relatively common (e.g. airplanes, intercontinental fiberoptic cables)

[Adapted from *Network Science* (2015) by Albert László Barabási]

Scale-free property

Many real-life networks have *power-law degrees*.

Scale-free property

Many real-life networks have *power-law degrees*.

Power-law paradigm

For some $\tau > 2$, the degree of a uniformly chosen vertex satisfies

$$\mathbb{P}(\mathsf{deg}(v) = x) \asymp \frac{C}{x^{\tau}}$$

Scale-free property

Many real-life networks have *power-law degrees*.

Power-law paradigm

For some $\tau > 2$, the degree of a uniformly chosen vertex satisfies

$$\mathbb{P}(\deg(v) = x) \asymp \frac{C}{x^{\tau}}$$

$$\log \mathbb{P}(\deg(v) = x) \asymp \log C - \tau \log x$$

log(proportion of degree x vertices) vs $\log x$ is a straight line.

Power laws

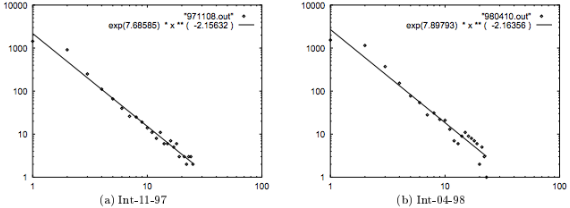
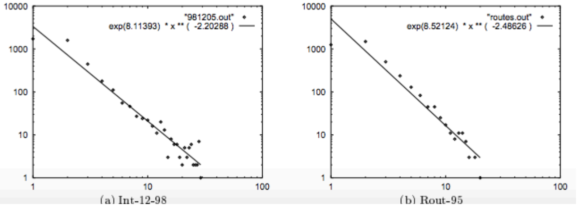


Figure 5: The outdegree plots: Log-log plot of frequency f_d versus the outdegree d.



Log-log plot of degree distribution of the router level internet network

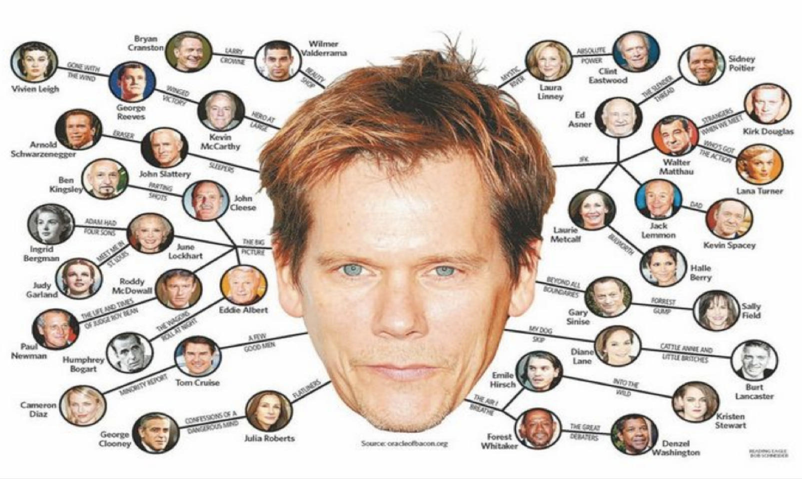
from Faloutsos, Faloutsos, Faloutsos. 1999

Growing networks

A question:

How did the network evolve around the servers of ‘99?

Small worlds and ultra-small worlds



Kevin Bacon game: Movie networks are small worlds

from Mark Robinson Writes

Small worlds and ultra-small worlds

Definition

A network G on N vertices is a *small world* if the average distance

$$\overline{\mathsf{dist}}(G) = \frac{1}{\binom{N}{2}} \sum_{u,v \in G} d_G(u,v) = \Theta(\log N).$$

Small worlds and ultra-small worlds

Definition
A network G on N vertices is a *small world* if the average distance

$$\overline{\text{dist}}(G) = \frac{1}{\binom{N}{2}} \sum_{u,v \in G} d_G(u,v) = \Theta(\log N).$$

G is an *ultrasmall world* if $\overline{\text{dist}(G)} = o(\log n)$.

Small worlds and ultra-small worlds

Definition
A network G on N vertices is a *small world* if the average distance

$$\overline{\mathsf{dist}}(G) = \frac{1}{\binom{N}{2}} \sum_{u,v \in G} d_G(u,v) = \Theta(\log N).$$

G is an *ultrasmall world* if $\overline{\mathsf{dist}(G)} = o(\log n)$.

Alternative def of ultrasmall:

$$\overline{\mathsf{dist}(G)} = \Theta(\log \log n)$$

.

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Dynamically growing models

Try to intrinsically explain a property of real life networks

- Original Barabási-Albert model aka Preferential attachment model
- Variable degree Preferential attachment model
- Bianconi-Barabási model aka PA with multiplicative fitness
- PA with additive fitness
- PA with power of choice
- …

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Dynamically growing models

Try to intrinsically explain a property of real life networks

- Original Barabási-Albert model aka Preferential attachment model
- Variable degree Preferential attachment model
- Bianconi-Barabási model aka PA with multiplicative fitness
- PA with additive fitness
- PA with power of choice
- ...

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Dynamically growing models

Try to intrinsically explain a property of real life networks

- Original Barabási-Albert model aka Preferential attachment model
- Variable degree Preferential attachment model
- Bianconi-Barabási model aka PA with multiplicative fitness
- PA with additive fitness
- PA with power of choice
- …

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Dynamically growing models

Try to intrinsically explain a property of real life networks

- Original Barabási-Albert model aka Preferential attachment model
- Variable degree Preferential attachment model
- Bianconi-Barabási model aka PA with multiplicative fitness
- PA with additive fitness
- PA with power of choice
- ...

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Dynamically growing models

Try to intrinsically explain a property of real life networks

- Original Barabási-Albert model aka Preferential attachment model
- Variable degree Preferential attachment model
- Bianconi-Barabási model aka PA with multiplicative fitness
- PA with additive fitness
- PA with power of choice
- …

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Dynamically growing models

Try to intrinsically explain a property of real life networks

- Original Barabási-Albert model aka Preferential attachment model
- Variable degree Preferential attachment model
- Bianconi-Barabási model aka PA with multiplicative fitness
- PA with additive fitness
- PA with power of choice
- …

Scale-free network models

Static network models

Degree distribution is superimposed on the network

- Configuration model
- Inhomogeneous random graphs
- Chung-Lu, Norros-Reitu models

Dynamically growing models

Try to intrinsically explain a property of real life networks

- Original Barabási-Albert model aka Preferential attachment model
- Variable degree Preferential attachment model
- Bianconi-Barabási model aka PA with multiplicative fitness
- PA with additive fitness
- PA with power of choice
- ...

The original Preferential Attachment Model

Definition

- Start with a single vertex

The original Preferential Attachment Model

Definition

- Start with a single vertex
- Vertices enter the network one-by-one at discrete steps $N = 1, 2, \ldots$

The original Preferential Attachment Model

Definition

- Start with a single vertex
- Vertices enter the network one-by-one at discrete steps $N = 1, 2, \ldots$
- New vertex connects to old vertices according to some increasing function of the degree

The original Preferential Attachment Model

Definition

- Start with a single vertex
- Vertices enter the network one-by-one at discrete steps $N = 1, 2, \ldots$
- New vertex connects to old vertices according to some increasing function of the degree
- Heuristically:

$$\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \mathsf{deg}_N(v_i)$$

The original Preferential Attachment Model

Classic variations

- Fixed outdegree: $(1, 0)$ edge/new vertex:

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i)$

The original Preferential Attachment Model

Classic variations

- Fixed outdegree: $(1, 0)$ edge/new vertex:

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i)$

- Fixed outdegree: $(1, \delta)$: 1 edge/new vertex.

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i) + \delta$

The original Preferential Attachment Model

Classic variations

- Fixed outdegree: $(1, 0)$ edge/new vertex:

  $\textbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i)$

- Fixed outdegree: $(1, \delta)$: 1 edge/new vertex.

  $\textbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i) + \delta$

- Fixed outdegree: (m, δ): m edges/new vertex.

  $\textbf{Prob}(v_{N+1} \xrightarrow{j} v_i) \propto \deg_{N,j}(v_i) + \delta/m$

The original Preferential Attachment Model

Classic variations

- Fixed outdegree: $(1, 0)$ edge/new vertex:

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i)$

- Fixed outdegree: $(1, \delta)$: 1 edge/new vertex.

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i) + \delta$

- Fixed outdegree: (m, δ): m edges/new vertex.

  $\mathbf{Prob}(v_{N+1} \xrightarrow{j} v_i) \propto \deg_{N,j}(v_i) + \delta/m$

- Variable outdegree: $\forall v_i \in \mathsf{PA}_N$: independently

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \frac{f(\deg_N(v_i))}{N}$,

  where $f(x)$ concave with $\lim_{x\to\infty} f(x)/x = \gamma_f \leq 1$.

The original Preferential Attachment Model

Classic variations

- Fixed outdegree: $(1,0)$ edge/new vertex:

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i)$

- Fixed outdegree: $(1,\delta)$: 1 edge/new vertex.

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \deg_N(v_i) + \delta$

- Fixed outdegree: (m,δ): m edges/new vertex.

  $\mathbf{Prob}(v_{N+1} \stackrel{j}{\longrightarrow} v_i) \propto \deg_{N,j}(v_i) + \delta/m$

- Variable outdegree: $\forall v_i \in \mathsf{PA}_N$: independently

  $\mathbf{Prob}(v_{N+1} \longrightarrow v_i) \propto \frac{f(\deg_N(v_i))}{N}$,

  where $f(x)$ concave with $\lim_{x\to\infty} f(x)/x = \gamma_f \leq 1$.

Animation

Degree distribution

Theorem (Degree sequence of PA(m,0))

Bollobás, Riordan, Spencer, Tusnády '01

Limiting degree distribution of Classic PA $(m, 0)$

$$\lim_{N \to \infty} \mathbf{Prob}(\deg(V_N) = k) \asymp \frac{1}{k^3}.$$

where V_N *uniformly chosen vertex in* PA_N.

Degree distribution

Theorem (Degree sequence of PA(m,0))

Bollobás, Riordan, Spencer, Tusnády '01
Limiting degree distribution of Classic PA $(m, 0)$

$$\lim_{N\to\infty} \mathbf{Prob}(\deg(V_N) = k) \asymp \frac{1}{k^3}.$$

where V_N uniformly chosen vertex in PA_N.

Generally

Limiting degree distribution is

$$\lim_{N\to\infty} \mathbf{Prob}(\deg(V_N) = k) \asymp \frac{1}{k^\tau},$$

where

$$\tau_{m,\delta} = 3 + \delta/m \qquad \tau_f = 1 + 1/\gamma_f.$$

Graph distances

Theorem (Distances in PAMs, $\tau \in (2,3)$)

Dommers, v/d Hofstad, Hoogiemstra '10 & Dereich and Mörters '13

Let U_N, V_N be two uniformly chosen vertices (within the giant component)[1]*. When $\tau \in (2,3)$,*

$$d_G^{(N)}(U_N, V_N) = (1 + o_{\mathbf{P}}(1)) \log\log N \cdot \frac{4}{|\log(\tau - 2)|}$$

Lower tightness holds.

[1] $d_G^{(N)}$ graph distance within PA_N

Graph distances

Theorem (Distances in PAMs, $\tau \in (2,3)$)

Dommers, v/d Hofstad, Hoogiemstra '10 & Dereich and Mörters '13
Let U_N, V_N be two uniformly chosen vertices (within the giant component)[1]*. When $\tau \in (2,3)$,*

$$d_G^{(N)}(U_N, V_N) = (1 + o_{\mathbf{P}}(1)) \log\log N \cdot \frac{4}{|\log(\tau - 2)|}$$

Lower tightness holds.

$\tau > 3$

Dommers, v/d Hofstad, Hoogiemstra '10

Typical distance in $\mathrm{PA}(m, \delta)$ for $m \geq 2$, $\tau > 3$:

$$\mathsf{d}_{\mathsf{G}}^{(\mathsf{N})}(U_N, V_N) = \Theta(\log N).$$

[1] $d_G^{(N)}$ graph distance within PA_N

Universality

For static models with power law degrees, $\tau \in (2, 3)$

$$d_G(U_N, V_N) = (1 + o_{\mathbf{P}}(1)) \log \log N \cdot \frac{2}{|\log(\tau - 2)|}.$$

For PA models with power law degrees, $\tau \in (2, 3)$

$$d_G(U_N, V_N) = (1 + o_{\mathbf{P}}(1)) \log \log N \cdot \frac{4}{|\log(\tau - 2)|}.$$

Universality

For static models with power law degrees, $\tau \in (2, 3)$

$$d_G(U_N, V_N) = (1 + o_{\mathbf{P}}(1)) \log\log N \cdot \frac{2}{|\log(\tau - 2)|}.$$

For PA models with power law degrees, $\tau \in (2, 3)$

$$d_G(U_N, V_N) = (1 + o_{\mathbf{P}}(1)) \log\log N \cdot \frac{4}{|\log(\tau - 2)|}.$$

Universality

For static models with power law degrees, $\tau \in (2,3)$

$$d_G(U_N, V_N) = (1 + o_{\mathbf{P}}(1)) \log\log N \cdot \frac{2}{|\log(\tau - 2)|}.$$

For PA models with power law degrees, $\tau \in (2,3)$

$$d_G(U_N, V_N) = (1 + o_{\mathbf{P}}(1)) \log\log N \cdot \frac{4}{|\log(\tau - 2)|}.$$

Factor 2:
In PA, a high degree vertex tends to be connected via a path of 2 to a higher degree vertex.
In static network models, directly.

Back to the question:

How do distances shrink as time passes?

Evolution of graph distances

Theorem (Evolution of distances)

Jorritsma, Komjáthy, (AoAP, 22)

Let U_N, V_N be two uniformly chosen vertices in PA_N (within the giant component). When $\tau \in (2,3)$, & $m \ge 2$, for $t > N$:

Evolution of graph distances

Theorem (Evolution of distances)

Jorritsma, Komjáthy, (AoAP, 22)
Let U_N, V_N be two uniformly chosen vertices in PA_N (within the giant component). When $\tau \in (2,3)$, & $m \geq 2$, for $t > N$:

$$d_G^{(t)}(U_N, V_N)$$

Evolution of graph distances

Theorem (Evolution of distances)

Jorritsma, Komjáthy, (AoAP, 22)

Let U_N, V_N be two uniformly chosen vertices in PA_N (within the giant component). When $\tau \in (2,3)$, & $m \geq 2$, for $t > N$:

$$d_G^{(t)}(U_N, V_N) \asymp \Big(\log\log N - \log(\log(t/N) \vee 1)\Big) \cdot \frac{4}{\log(\tau-2)|} \vee 2$$

+*tightness around the main term.*

Evolution of graph distances

Theorem (Evolution of distances)

Jorritsma, Komjáthy, (AoAP, 22)

Let U_N, V_N be two uniformly chosen vertices in PA_N (within the giant component). When $\tau \in (2,3)$, & $m \geq 2$, for $t > N$:

$$X_N := \sup_{t>N} \left(d_G^{(t)}(U_N, V_N) - \left[\left(\log\log N - \log(\log(t/N) \vee 1) \right) \cdot \frac{4}{\log(\tau-2)|} \vee 2 \right] \right)$$

Evolution of graph distances

Theorem (Evolution of distances)

Jorritsma, Komjáthy, (AoAP, 22)

Let U_N, V_N be two uniformly chosen vertices in PA_N (within the giant component). When $\tau \in (2,3)$, & $m \geq 2$, for $t > N$:

$$X_N := \sup_{t>N}\left(d_G^{(t)}(U_N, V_N) - \left[\left(\log\log N - \log(\log(t/N) \vee 1)\right) \cdot \frac{4}{\log(\tau-2)|} \vee 2 \right]\right)$$

$(X_N)_{N\geq 1}$ is a tight sequence of random variables.

Evolution of graph distances

Theorem (Evolution of distances)

Jorritsma, Komjáthy, (AoAP, 22)

Let U_N, V_N be two uniformly chosen vertices in PA_N (within the giant component). When $\tau \in (2,3)$, & $m \geq 2$, for $t > N$:

$$X_N := \sup_{t>N}\left(d_G^{(t)}(U_N, V_N) - \left[\left(\log\log N - \log(\log(t/N) \vee 1) \right) \cdot \frac{4}{\log(\tau-2)|} \vee 2 \right] \right)$$

$(X_N)_{N\geq 1}$ is a tight sequence of random variables.

Message 1: Fix $a \in (0,1)$. Then at $t = N \exp\left(\varepsilon(\log N)^a\right) \ll N^{1+\varepsilon}$,

$$d_G^{(t)}(U_N, V_N) \approx (1-a) \log\log N \cdot \frac{4}{|\log(\tau-2)|}.$$

Message 2: $d_G^{(t)}(U_N, V_N)$ never leaves a (tight) strip around the main term.

Proofs

Lower bound

Lower bound in growing network

- A path counting method:
  A **possible path** π is a sequence of labelled vertices. The path's (potential) arrival time is the youngest vertex on the path. Among all possible labeled paths π_{U_N,V_N},

Lower bound in growing network

- A path counting method:
  A **possible path** π is a sequence of labelled vertices. The path's (potential) arrival time is the youngest vertex on the path. Among all possible labeled paths π_{U_N,V_N},
-

$$\mathbb{E}\Big[\sum_{t>N} \sum_{[\pi_{U_N,V_N}\text{too short at } t]} \mathbf{1}\{\pi_{U_N,V_N} \text{ present in } \mathsf{PA}_t\}\Big] \xrightarrow{?} 0.$$

Lower bound in growing network

- A path counting method:
  A **possible path** π is a sequence of labelled vertices. The path's (potential) arrival time is the youngest vertex on the path. Among all possible labeled paths π_{U_N,V_N},
- $$\mathbb{E}\Big[\sum_{t>N} \sum_{[\pi_{U_N,V_N} \text{too short at } t]} \mathbf{1}\{\pi_{U_N,V_N} \text{ present in } \mathsf{PA}_t\}\Big] \xrightarrow{?} 0.$$
- $$\mathbb{E}\Big[\sum_{\pi_{U_N,V_N} \text{too short at arrival}} \mathbf{1}\{\pi_{U_N,V_N} \text{ present in } \mathsf{PA}_t\}\Big] \xrightarrow{?} 0.$$

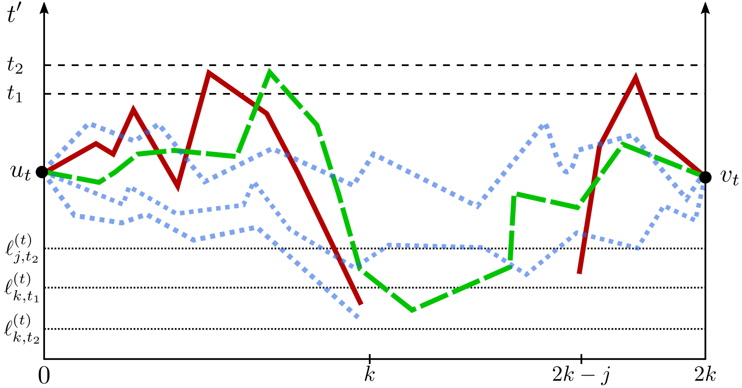
Lower bound in growing network

- A path counting method:
  A **possible path** π is a sequence of labelled vertices. The path's (potential) arrival time is the youngest vertex on the path. Among all possible labeled paths π_{U_N,V_N},
- $$\mathbb{E}\Big[\sum_{t>N} \sum_{[\pi_{U_N,V_N} \text{too short at } t]} \mathbf{1}\{\pi_{U_N,V_N} \text{ present in } \mathsf{PA}_t\}\Big] \xrightarrow{?} 0.$$
- $$\mathbb{E}\Big[\sum_{\pi_{U_N,V_N} \text{too short at arrival}} \mathbf{1}\{\pi_{U_N,V_N} \text{ present in } \mathsf{PA}_t\}\Big] \xrightarrow{?} 0.$$
- A union bound not over time, but over possible paths.

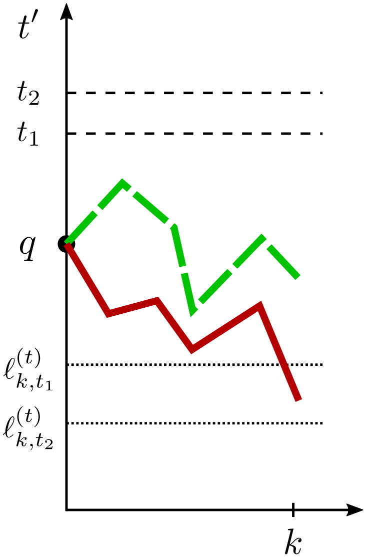
Lower bound in growing network

- A path counting method:
  A **possible path** π is a sequence of labelled vertices. The path's (potential) arrival time is the youngest vertex on the path. Among all possible labeled paths π_{U_N,V_N},
-
$$\mathbb{E}\Big[\sum_{t>N} \sum_{[\pi_{U_N,V_N}\text{too short at } t]} \mathbf{1}\{\pi_{U_N,V_N} \text{ present in } \mathsf{PA}_t\}\Big] \xrightarrow{?} 0.$$
-
$$\mathbb{E}\Big[\sum_{\pi_{U_N,V_N}\text{too short at arrival}} \mathbf{1}\{\pi_{U_N,V_N} \text{ present in } \mathsf{PA}_t\}\Big] \xrightarrow{?} 0.$$
- A union bound not over time, but over possible paths.
- This still does not tend to zero...

Bad paths

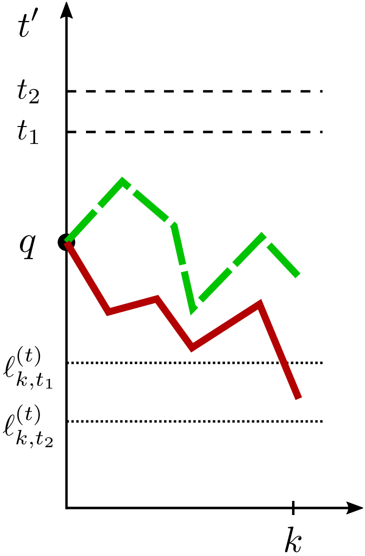


Bad paths



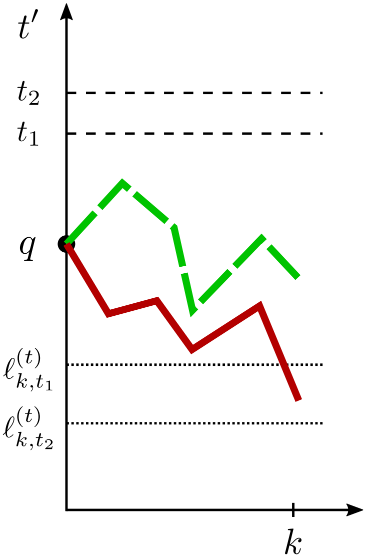
- A path is bad if it reaches a too old vertex in too few steps

Bad paths



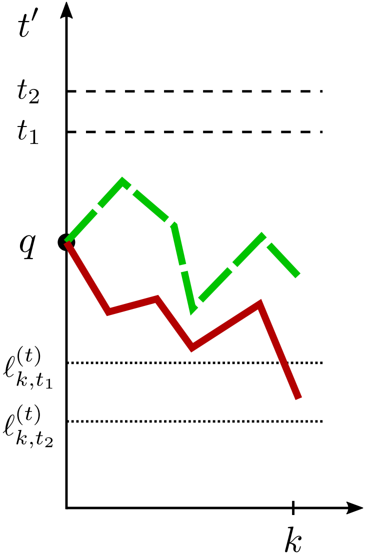
- A path is bad if it reaches a too old vertex in too few steps
- i.e., if its k vertex is older than some threshold ℓ_k

Bad paths



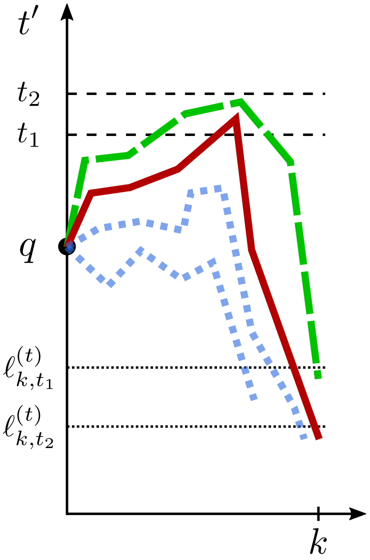
- A path is bad if it reaches a too old vertex in too few steps
- i.e., if its k vertex is older than some threshold $\ell_{k,t}^{(N)}$

Bad paths



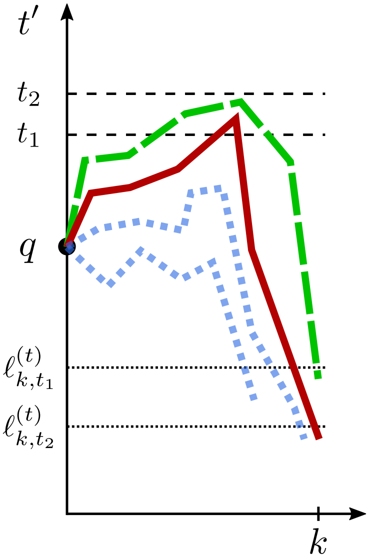
- A path is bad if it reaches a too old vertex in too few steps
- i.e., if its k vertex is older than some threshold $\ell_{k,t}^{(N)}$
- A bad path in PA_t might be already good at some later time.

Bad paths



- A path is bad if it reaches a too old vertex in too few steps
- i.e., if its k vertex is older than some threshold $\ell_{k,t}^{(N)}$
- A bad path in PA_t might be already good at some later time.
- a good path can never turn bad.

Bad paths

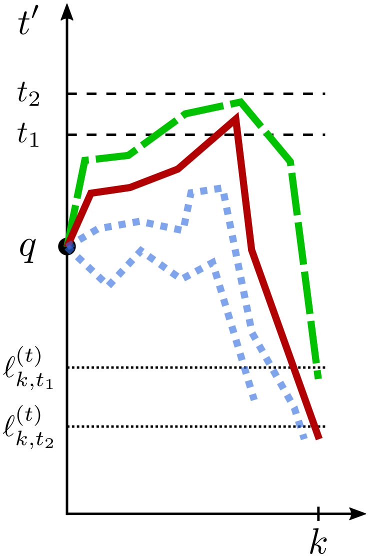


- A path is bad if it reaches a too old vertex in too few steps
- i.e., if its k vertex is older than some threshold $\ell_{k,t}^{(N)}$
- A bad path in PA_t might be already good at some later time.
- a good path can never turn bad.
- $\Rightarrow$ Enough to bound:

$$\mathbb{E}\Big[\sum_t \sum_{\pi(U_N,V_N)\ni v_t \text{ bad}} \mathbf{1}\{\pi \text{ present in } PA_t\}\Big].$$

i.e., containing the newly added v_t.

Bad paths



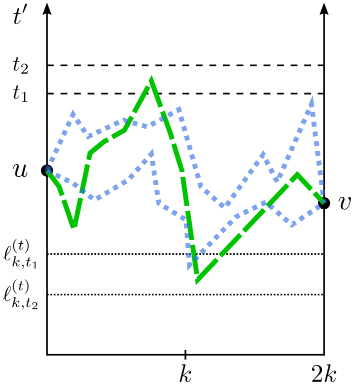
- A path is bad if it reaches a too old vertex in too few steps
- i.e., if its k vertex is older than some threshold $\ell^{(N)}_{k,t}$
- A bad path in PA_t might be already good at some later time.
- a good path can never turn bad.
- $\Rightarrow$ Enough to bound:

$$\mathbb{E}\Big[\sum_t \sum_{\pi(U_N,V_N)\ni v_t \text{ bad}} \mathbf{1}\{\pi \text{ present in } PA_t\}\Big].$$

i.e., containing the newly added v_t.

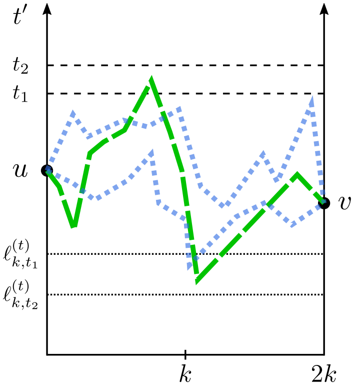
- This tends to zero $\Rightarrow$ whp **no bad paths** present **ever**.

Good paths that are too short



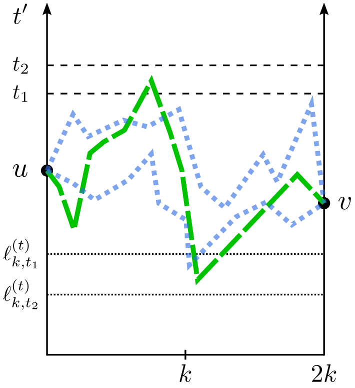
- A good path is too short at time $t > N$ if it is shorter than the [main term$-C$].

Good paths that are too short



- A good path is too short at time $t > N$ if it is shorter than the [main term$-C$].
- A too short path may become not too short as t passes.

Good paths that are too short

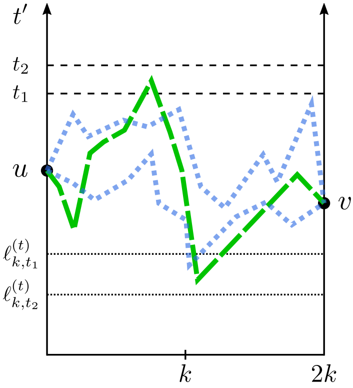


- A good path is too short at time $t > N$ if it is shorter than the [main term$-C$].
- A too short path may become not too short as t passes.
- $\Rightarrow$ Enough to compute:

$$\mathbb{E}\Big[\sum_t \sum_{\pi(U_N,V_N)\ni v_t \text{ short good}} \mathbf{1}\{\pi \text{ present in } PA_t\}\Big]$$

i.e., containing the newly added v_t.

Good paths that are too short



- A good path is too short at time $t > N$ if it is shorter than the [main term$-C$].
- A too short path may become not too short as t passes.
- $\Rightarrow$ Enough to compute:

$$\mathbb{E}\Big[\sum_t \sum_{\pi(U_N,V_N)\supset v_t \text{ short good}} \mathbf{1}\{\pi \text{ present in } PA_t\}\Big]$$

  i.e., containing the newly added v_t.
- This tends to zero $\Rightarrow$ whp **no too short good paths** present **ever**.

Upper bound

Local weak limit of PA

- First we understand how the graph looks locally around U_N and around V_N.

Local weak limit of PA

- First we understand how the graph looks locally around U_N and around V_N.
- Up to some radius R_N, these are disjoint trees (whp)

Local weak limit of PA

- First we understand how the graph looks locally around U_N and around V_N.
- Up to some radius R_N, these are disjoint trees (whp)
- Limiting distribution: exists. Dereich, Mörters '13 & Berger, Borgs, Chayes, Saberi '14

Local weak limit of PA

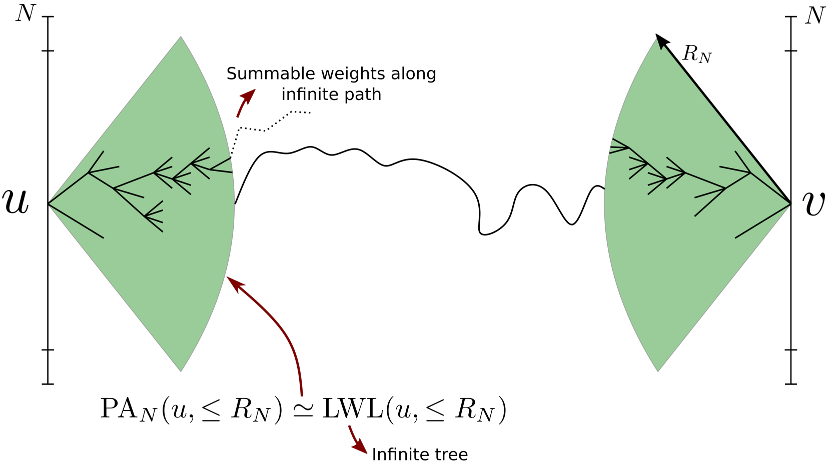
- First we understand how the graph looks locally around U_N and around V_N.
- Up to some radius R_N, these are disjoint trees (whp)
- Limiting distribution: exists. Dereich, Mörters '13 & Berger, Borgs, Chayes, Saberi '14
- They are (killed) multitype branching processes.

Local weak limit of PA

- First we understand how the graph looks locally around U_N and around V_N.
- Up to some radius R_N, these are disjoint trees (whp)
- Limiting distribution: exists. Dereich, Mörters '13 & Berger, Borgs, Chayes, Saberi '14
- They are (killed) multitype branching processes.
- We will use this info as a black box.

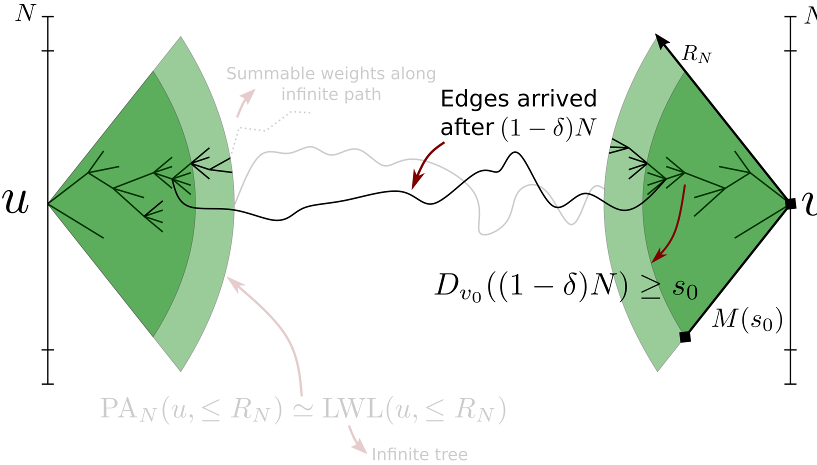
Proof outline for fixed N

y-axes=arrival time of vertices; x-axis: graph distance. LWL=limit distribution of the trees, depth R_N (large constant).



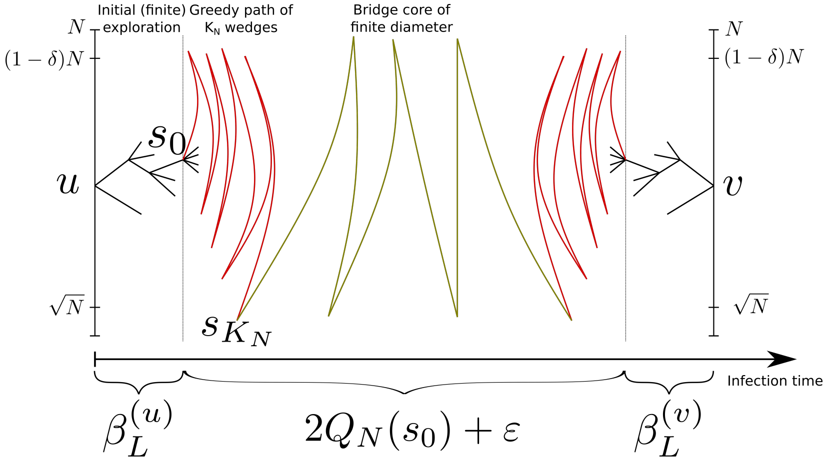
Proof outline for fixed N

y-axes=arrival time of vertices; x-axis: graph distance. LWL=limit distribution of the trees, depth R_N (large constant).



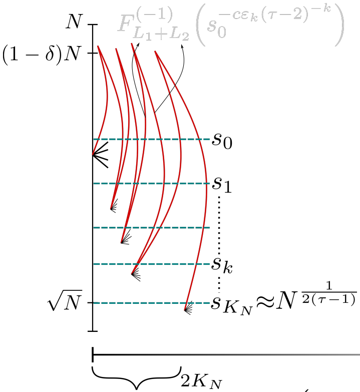
Building the connecting path

y-axes=arrival time of vertices; *x*-axis: graph distance.



Greedy path to the core

Core: vertices born before time $\sqrt{N}$. Core has bounded diameter.



Careful choice of the sequence $(s_k)_{k\geq 1}$ yields
- Short path to dense inner core

$$K_N \leq \frac{\log\log(N)}{|\log(\tau-2)|} + C$$

- From any vertex in Layer k, there are at least $s_k^{\varepsilon_k}$ young wedges to Layer $k+1$

- "Young" edges ensures independent weights w.r.t. to weighted LWL

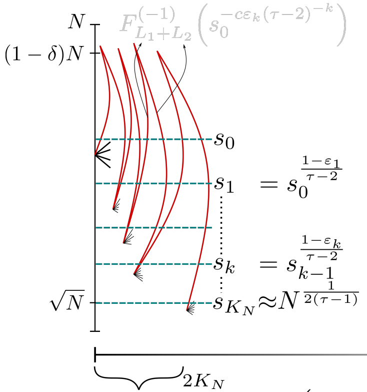
- Minimum of i.i.d. weighted wedges

$$\min_{j\leq M}\{L_1^{(j)} + L_2^{(j)}\} \approx F^{(-1)}_{L_1+L_2}(1/M)$$

$$Q_N(s_0) \approx \sum_{k=1}^{2K_N} F_L^{(-1)}\left(s_0^{-(\tau-2)^{-k/2}}\right)$$

Greedy path to the core

Core: vertices born before time $\sqrt{N}$. Core has bounded diameter.



$$Q_N(s_0) \approx \sum_{k=1}^{2K_N} F_L^{(-1)}\left(s_0^{-(\tau-2)^{-k/2}}\right)$$

Careful choice of the sequence $(s_k)_{k\geq 1}$ yields
- Short path to dense inner core

$$K_N \leq \frac{\log\log(N)}{|\log(\tau-2)|} + C$$

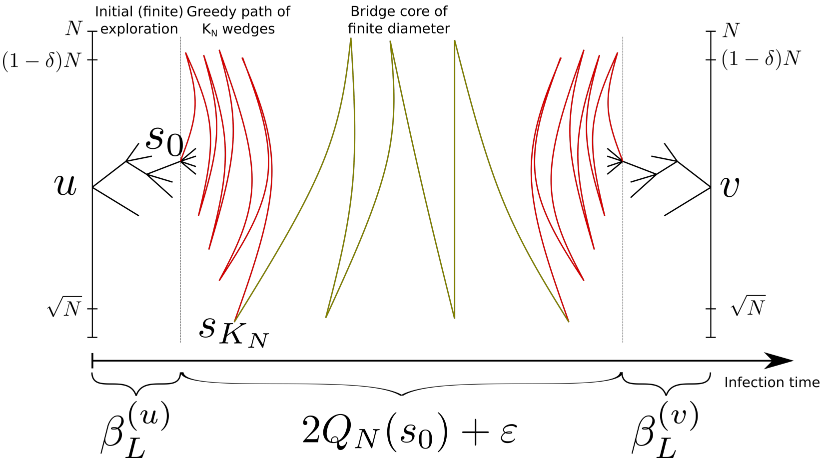
- From any vertex in Layer k, there are at least $s_k^{\varepsilon_k}$ young wedges to Layer $k+1$

- "Young" edges ensures independent weights w.r.t. to weighted LWL

- Minimum of i.i.d. weighted wedges

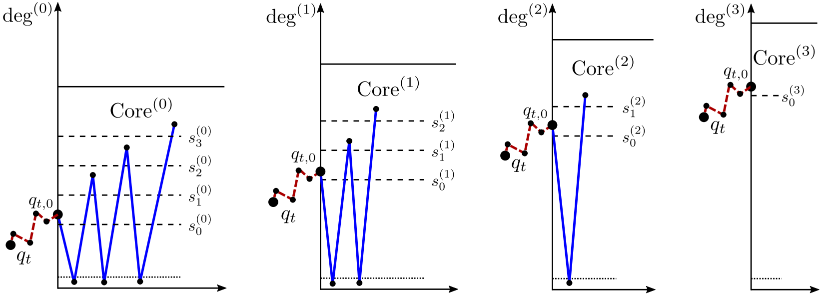
$$\min_{j\leq M}\{L_1^{(j)} + L_2^{(j)}\} \approx F^{(-1)}_{L_1+L_2}(1/M)$$

Dense core has small diameter



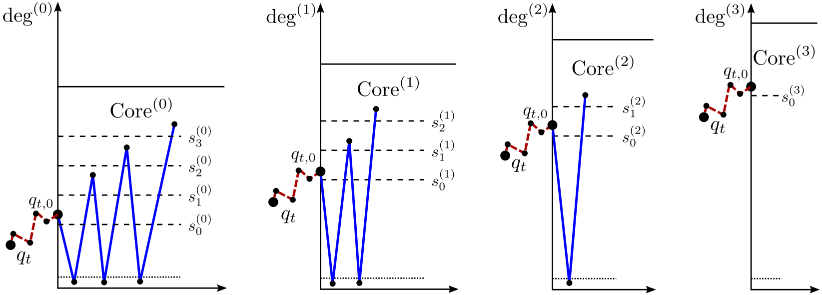
Extension to the growing network

- We need to track the degree-growth of the vertex that starts the wedging procedure (via a Móri martingale).



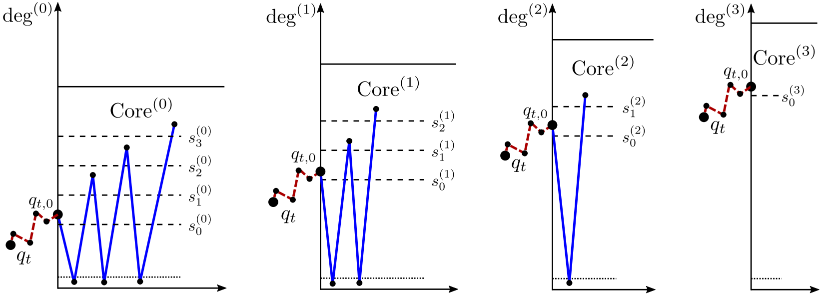
Extension to the growing network

- We need to track the degree-growth of the vertex that starts the wedging procedure (via a Móri martingale).
- We redefine the layers for each time $t > N$



Extension to the growing network

- We need to track the degree-growth of the vertex that starts the wedging procedure (via a Móri martingale).
- We redefine the layers for each time $t > N$
- Careful union bound: summing error probs only where the event changes



Thank you for the attention!

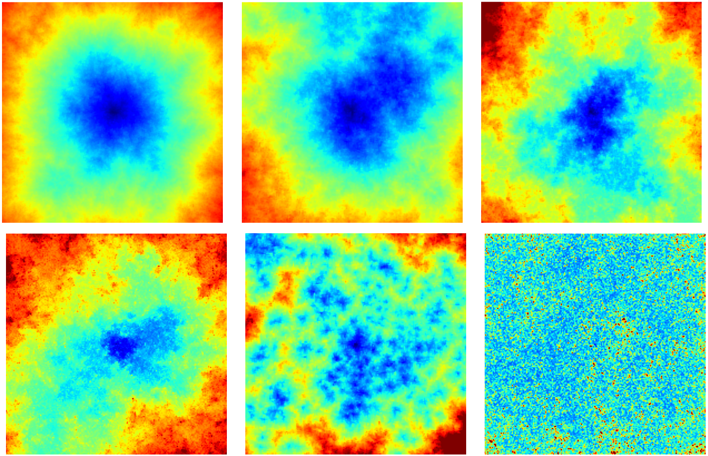
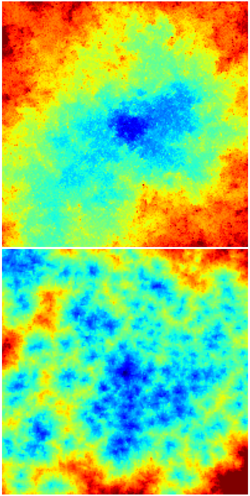


Figure: Six instances of an infection spreading on a two-dimensional spatial scale free networks with different parameters.

Spreading processes on networks



Susceptible-Infected model:

- At time $t = 0$ the source node is infected, all other nodes susceptible.
- if, on an edge $\{u, v\}$, u is infected and v is not, then v becomes infected after a random iid transmission delay $L_{(u,v)}$.

The epidemic curve

The function that counts the total number of infected nodes before time t:

$$I(t) = \#\{\ \textit{infected nodes before time } t\}$$

Weighted-network point of view

Pre-sampling all randomness

Add *iid* weights from distribution L to existing edges.

Weighted-network point of view

Pre-sampling all randomness

Add *iid* weights from distribution L to existing edges.

Spreading time = weighted distance

Edge weight = transmission delay through the edge

Weighted-network point of view

Pre-sampling all randomness

Add *iid* weights from distribution L to existing edges.

Spreading time = weighted distance

Edge weight = transmission delay through the edge
The spreading time between two vertices u, v
= the weighted distance:

Weighted-network point of view

Pre-sampling all randomness

Add *iid* weights from distribution L to existing edges.

Spreading time = weighted distance

Edge weight = transmission delay through the edge
The spreading time between two vertices u, v
= the weighted distance:

$$d_L(u,v) = \min_{\pi:\text{path } u \leftrightarrow v} \text{sum of } L_e \text{ on edges on } \pi$$

Weighted-network point of view

Pre-sampling all randomness

Add *iid* weights from distribution L to existing edges.

Spreading time = weighted distance

Edge weight = transmission delay through the edge
The spreading time between two vertices u, v
= the weighted distance:

$$d_L(u,v) = \min_{\pi:\text{path } u\leftrightarrow v} \text{sum of } L_e \text{ on edges on } \pi$$

Epidemic curve:

$$I(t) = \#\{v : d_L(u,v) \le t\}.$$

Weighted distances in PAMs

$L \equiv 1$: graph distance

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$

Weighted distances in PAMs

$L \equiv 1$: graph distance

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$

$L \equiv 2$:

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{8}{|\log(\tau-2)|}$

Weighted distances in PAMs

$L \equiv 1$: graph distance

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$

$L \equiv 2$:

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{8}{|\log(\tau-2)|}$

$L \sim U[1,2]$:

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$

Weighted distances in PAMs

$L \equiv 1$: graph distance

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$

$L \equiv 2$:

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{8}{|\log(\tau-2)|}$

$L \sim U[1, 2]$:

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$

$L \sim U[0, 1]$:

Weighted distances in PAMs

$L \equiv 1$: graph distance

$$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$$

$L \equiv 2$:

$$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{8}{|\log(\tau-2)|}$$

$L \sim U[1,2]$:

$$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$$

$L \sim U[0,1]$:

$$d_G^{(N)}(U_N, V_N) \asymp \Theta(1)$$

Weighted distances in PAMs

$L \equiv 1$: graph distance

$$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$$

$L \equiv 2$:

$$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{8}{|\log(\tau-2)|}$$

$L \sim U[1,2]$:

$$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$$

$L \sim U[0,1]$:

$$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L \overset{a.s.}{<} \infty$$

Weighted distances in PAMs

$L \equiv 1$: **graph distance**

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$

$L \equiv 2$:

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{8}{|\log(\tau-2)|}$

$L \sim U[1,2]$:

$d_G^{(N)}(U_N, V_N) \asymp \log\log N \cdot \frac{4}{|\log(\tau-2)|}$

$L \sim U[0,1]$,

$L \sim \mathrm{Exp}(\lambda), \mathrm{Gamma}(\lambda, k), \dots$:

$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L \overset{a.s.}{<} \infty$

Weighted distances on PA

Theorem (Explosion in PA)

Jorritsma, K, '20

Consider PA with $\tau \in (2,3)$, and U_N, V_N two typical vertices in the giant component of PA_N. Then if

$$E(L) := \sum_{k=1}^{\infty} F_L^{(-1)}\left(1/\mathrm{e}^{\mathrm{e}^k}\right) < \infty,$$

then

$$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L \overset{a.s.}{<} \infty.$$

Weighted distances on PA

Theorem (Explosion in PA)

Jorritsma, K, '20

Consider PA with $\tau \in (2,3)$, and U_N, V_N two typical vertices in the giant component of PA_N. Then if

$$E(L) := \sum_{k=1}^{\infty} F_L^{(-1)}\big(1/\mathrm{e}^{\mathrm{e}^k}\big) < \infty,$$

then

$$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L \overset{a.s.}{<} \infty.$$

If $E(L) = \infty$ then

$$d_G^{(N)}(U_N, V_N) = 2 \sum_{k=1}^{2\log\log N/|\log(\tau-2)|} F_L^{(-1)}\big(1/\mathrm{e}^{1/(\tau-2)^k}\big) + O_{\mathbb{P}}(1).$$

From small to mini-worlds

- $E(L) = \infty$: fluctuations are tight in most cases. Lower tightness always, upper tightness under technical condition that holds for most Ls.

$$E_2(L) = \sum_{k=1}^{\infty} \frac{1}{k} F^{(-1)}_{L-\inf \mathrm{supp} L}\left(1/\mathrm{e}^{\mathrm{e}^k}\right) < \infty$$

From small to mini-worlds

- $E(L) = \infty$: fluctuations are tight in most cases. Lower tightness always, upper tightness under technical condition that holds for most Ls.
$$E_2(L) = \sum_{k=1}^{\infty} \frac{1}{k} F^{(-1)}_{L-\inf \mathrm{supp} L}(1/e^{e^k}) < \infty$$
- Fix your favorite $1 \ll g(N) = O(\log\log N)$. Then one can construct a distribution L such that
$$d_G^{(N)}(U_N, V_N) = g(N) + O_{\mathbb{P}}(1).$$

Shrinking of weighted distances when $I(L) = \infty$

For graph distances

$$\sup_{t' \geq t} \left(d_G^{(N')}(U_N, V_N) - 2\Big(\log\log N - \log(1 \vee \log(N'/N)) \Big) \cdot \frac{2}{\log(\tau - 2)|} \vee 1 \right) = O_{\mathsf{P}}(1).$$

Shrinking of weighted distances when $I(L) = \infty$

For graph distances

$$\sup_{t' \geq t} \Big(d_G^{(N')}(U_N, V_N) - 2\Big(\log\log N - \log(1 \vee \log(N'/N))\Big) \cdot \frac{2}{\log(\tau-2)|} \vee 1 \Big) = O_{\mathrm{P}}(1).$$

Theorem

Jorritsma, K, '20

Consider PA with $\tau \in (2,3)$, and U_N, V_N two typical vertices in the giant component of PA_N, and $I(L) = \infty$. Then

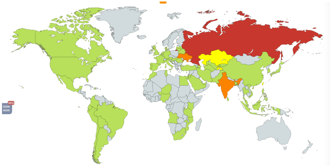
$$\sup_{N' \geq N} \Big(d_G^{(N')}(U_N, V_N) - 2 \sum_{k=1}^{K_{N,N'}} F_L^{(-1)}\big(1/\mathrm{e}^{1/(\tau-2)^k}\big)\Big) = O_{\mathbb{P}}(1)$$

with $K_{N,N'} = \Big(\log\log N - \log(1 \vee \log(N'/N))\Big) \cdot \frac{2}{\log(\tau-2)|} \vee 1$

What is explosion?

On infinite networks

A spreading process is explosive on an infinite network if $I(t) = \infty$ for some $t < \infty$.



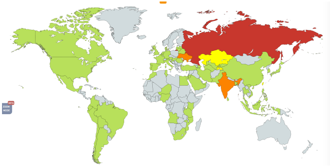
What is explosion?

On infinite networks

A spreading process is explosive on an infinite network if $I(t) = \infty$ for some $t < \infty$.

On very large but finite networks

A *positive fraction* of the nodes is infected within *constant time*, *no matter how large the network is*.



What is explosion?

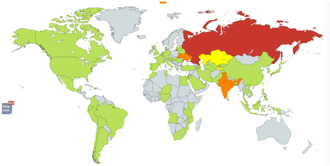
On infinite networks

A spreading process is explosive on an infinite network if $I(t) = \infty$ for some $t < \infty$.

On very large but finite networks

A *positive fraction* of the nodes is infected within *constant time*, *no matter how large the network is*.

- **1970s:** Grey, Harris, Sevastanov: explosion in Branching processes



What is explosion?

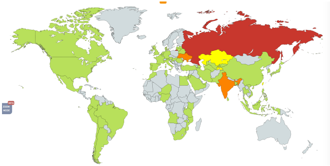
On infinite networks

A spreading process is explosive on an infinite network if $I(t) = \infty$ for some $t < \infty$.

On very large but finite networks

A *positive fraction* of the nodes is infected within *constant time*, *no matter how large the network is.*

- **1970s:** Grey, Harris, Sevastanov: explosion in Branching processes
- **2010s:** Amini, Devroye, Griffith, Olver: explosion in Branching random walks



What is explosion?

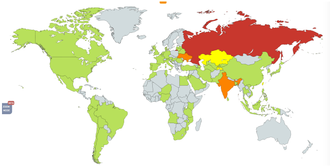
On infinite networks

A spreading process is explosive on an infinite network if $I(t) = \infty$ for some $t < \infty$.

On very large but finite networks

A *positive fraction* of the nodes is infected within *constant time*, *no matter how large the network is*.

- **1970s:** Grey, Harris, Sevastanov: explosion in Branching processes
- **2010s:** Amini, Devroye, Griffith, Olver: explosion in Branching random walks
- **2017+:** Me: explosion on networks



Explosion in PA

$$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L = Y_1 + Y_2$$

Then

A deterministic curve with a random constant shift.

Explosion in PA

$$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L \;= Y_1 + Y_2$$

Then

$$I(t) = \frac{1}{N}\#\{v : d^{(N)}(U_N, v) \le t\}$$

A deterministic curve with a random constant shift.

Explosion in PA

$$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L \;\; = Y_1 + Y_2$$

Then

$$I(t) = \frac{1}{N}\#\{v : d^{(N)}(U_N, v) \le t\}$$
$$= \frac{1}{N}\sum_{v \in \mathsf{PA}_N} \mathbb{1}_{\{d_G(U_n, v) \le t\}}$$

A deterministic curve with a random constant shift.

Explosion in PA

$$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L \;= Y_1 + Y_2$$

Then

$$\begin{aligned} I(t) &= \frac{1}{N} \#\{v : d^{(N)}(U_N, v) \le t\} \\ &= \frac{1}{N} \sum_{v \in \mathsf{PA}_N} \mathbb{1}_{\{d_G(U_n, v) \le t\}} \\ &= \mathbb{P}_{V_N}(d_G(U_n, V_n) \le t \mid U_n) \end{aligned}$$

A deterministic curve with a random constant shift.

Explosion in PA

$$d_G^{(N)}(U_N, V_N) \xrightarrow{d} \beta_L \ = Y_1 + Y_2$$

Then

$$\begin{aligned} I(t) &= \frac{1}{N}\#\{v : d^{(N)}(U_N, v) \le t\} \\ &= \frac{1}{N}\sum_{v \in \mathbf{PA}_N} \mathbb{1}_{\{d_G(U_n, v) \le t\}} \\ &= \mathbb{P}_{V_N}(d_G(U_n, V_n) \le t \mid U_n) \\ &\xrightarrow{d} \mathbb{P}(Y_2 \le t - Y_1 \mid Y_1) = g(t - Y_1) \end{aligned}$$

A deterministic curve with a random constant shift.