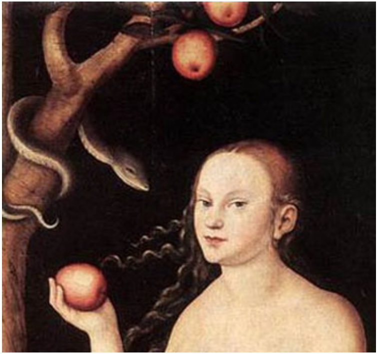
Distance in the Forest Fire Model
How far are you from Eve?



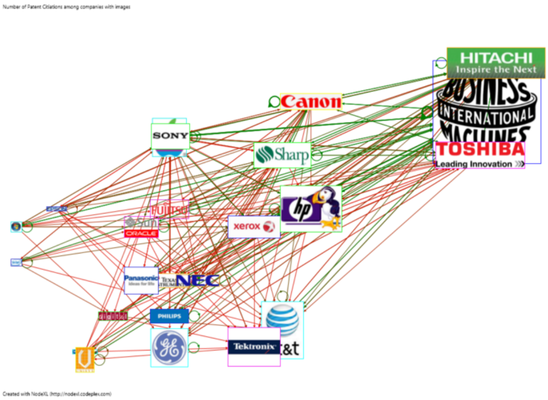
Varun Kanade, Reut Levi, Zvi Lotker,
Frederik Mallmann-Trenn, and Claire Mathieu

Motivation

- How many degrees of separation are there between you and François Hollande?

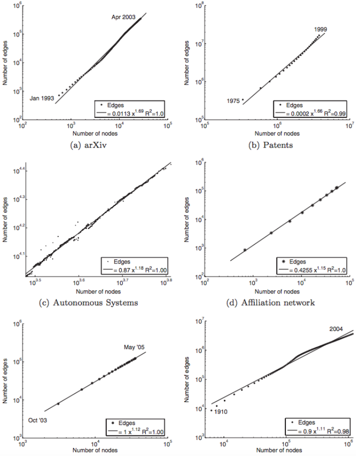
- Why bother?
  - Social networks are everywhere in daily life.
    - How do we meet new friends?
  - Has impact in economics, social science, marketing, disease propagation, ...

Motivation



- We study the distance in the Forest Fire model [Leskovec, Kleinberg and Faloutsos (2005)] ($>$ 1000 citations)
  - Random graph
  - Only simulations

Densification

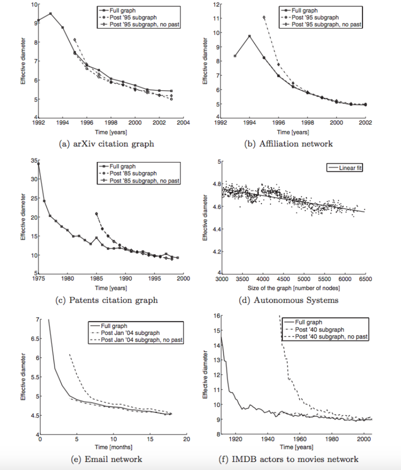


(a) arXiv (b) Patents

(c) Autonomous Systems (d) Affiliation network

(e) Email network (f) IMDB actors to movies network

Distance (between two random vertices)



(a) arXiv citation graph (b) Affiliation network

(c) Patents citation graph (d) Autonomous Systems

(e) Email network (f) IMDB actors to movies network

Distance

Empirical Result[KLF05]

Many social networks have constant expected distance.

Simulation Result[KLF05]

The Forest Fire Model has expected constant expected distance.

Theorem

Nothing proved before this work

Other social network models

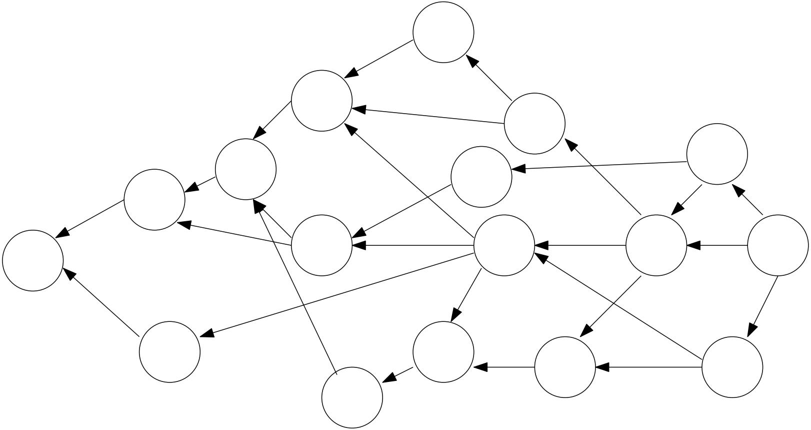
- Static Models
  - Small world; grid+random edges [Kleinberg *et al.*, 01]
  - Small world; grid+random walk [Chaintreau *et al.*, 08]
  - Kronecker product [Leskovec *et al.*, 05]
  - Chung Lu graphs [Chung and Lu, 03]
  - Block Two-Level Erdős-Rényi [Seshadhri *et al.*, 12]
- Dynamic Models
  - Preferential attachment [Barabási and R. Albert, 99]
  - Edge copying model [Kumar *et al.*, 00] [Kleinberg *et al.*, 99]
  - Community guided attachment [Leskovec *et al.*, 07]
  - Recursive search model [Vazquez, 01]
  - Random Surfer model [Blum *et al.*, 06]

Outline

Forest Fire Process

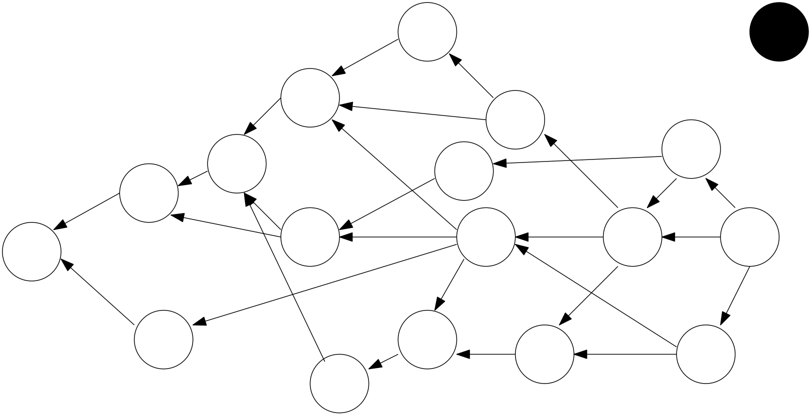
- Question: How do we meet new friends?
- Answer: We are introduced to them by our friends and the friends of our friends etc.

Generative model



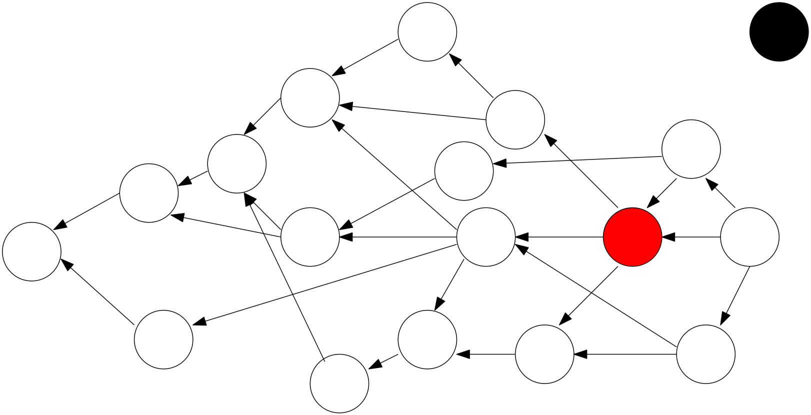
Current network

Example: arrival



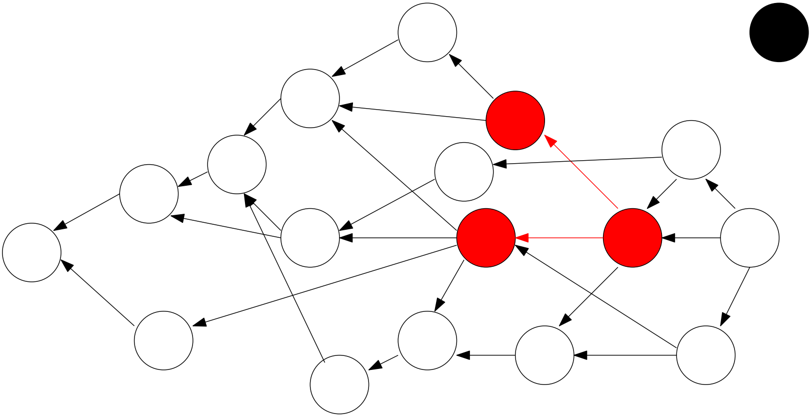
Next step: new node **u** arrives (black node).

Example



u already has one friend, $\mathrm{amb}(u)$, a uniform random node.

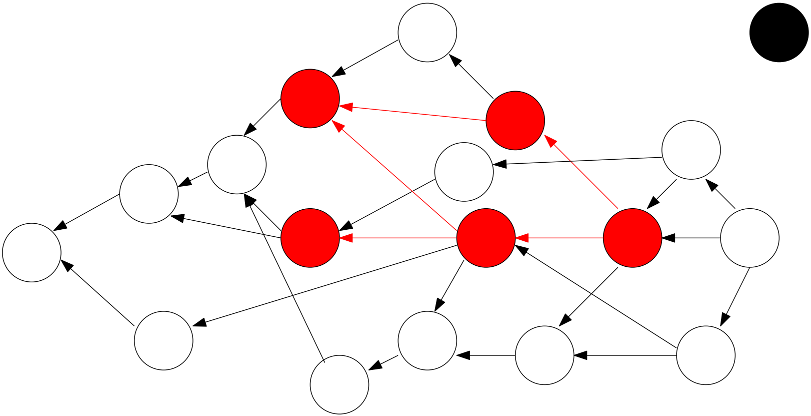
Example



Introduction to friends: start the 'forest fire'.

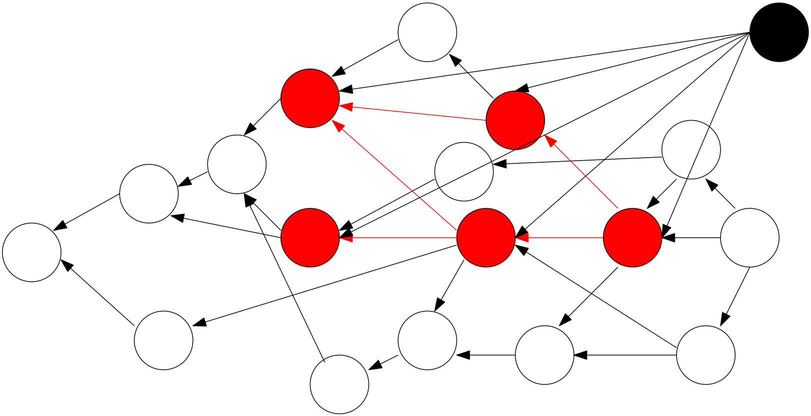
Every 'burnt' node v activates edges w.p. $\alpha/\text{outdeg}(v)$.

Example



Continue burning...

Example



Friends of u: vertex u connects to all 'burnt' nodes.

Example

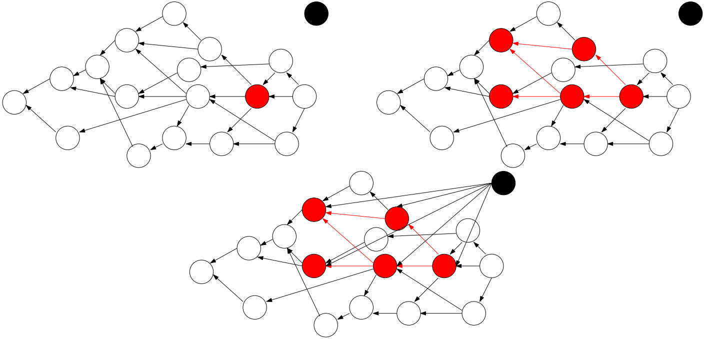


Figure: The three steps:
1) Choose $\mathrm{amb}(u)$ u.a.r. 2) Forest Fire (percolation) 3) Connect

Differences with original Forest Fire Process model

- Windy: we regard a special case in which we only burn edges in one direction (from 'younger' nodes to 'older' nodes).
- Initialization: we assume a seed-graph of constant size.

Our Theorem

Empirical Result[KLF05]

Many social networks have constant expected distance.

Simulation Result[KLF05]

The Forest Fire Model has bounded expected distance.

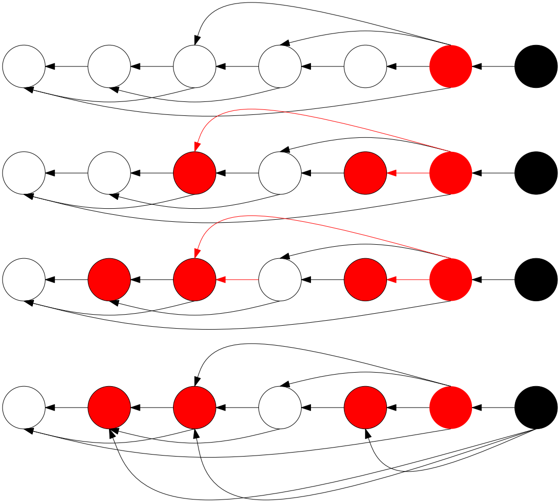
Theorem 1

Assume α, $|G_0|$ large enough. For Windy Forest Fire Process:

$$\mathbb{E}[\mathrm{dist}_{G_t}(u, G_0)] = O(1).$$

Windy Forest Line Process

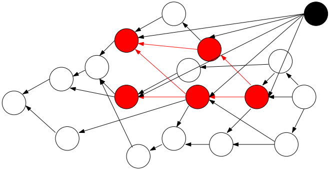
- From now on: $\mathrm{amb}(\mathbf{u_t}) = \mathbf{u_{t-1}}$.



Outline

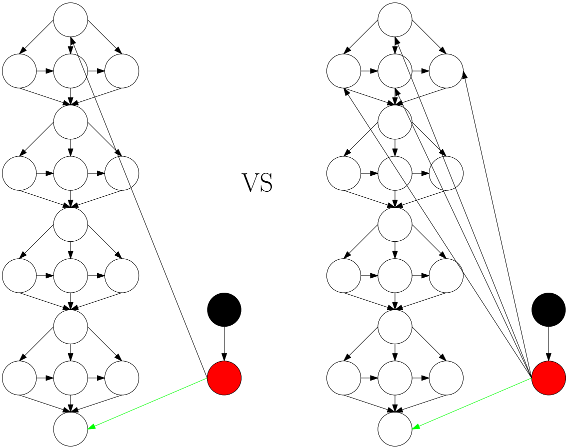
Proving bounded distance: Failed attempt #1

- Attempt: Extend analysis of Erdós-Renyi or Preferential Attachment graphs
- Problem: Independence (ER) vs Correlated edges (Forest Fire)
  - Neighborhood of u is closely related to neighborhood of $\mathrm{amb}(u)$.



Proving bounded distance: Failed attempt #2

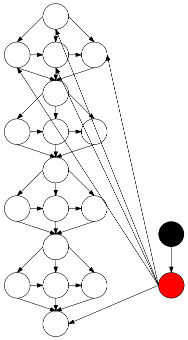
- Attempt: Simplify graph by insertion/deletion of edges
- Problem: expected distance is not monotone under insertion/deletion of edges



Proving bounded distance: Failed attempt #3

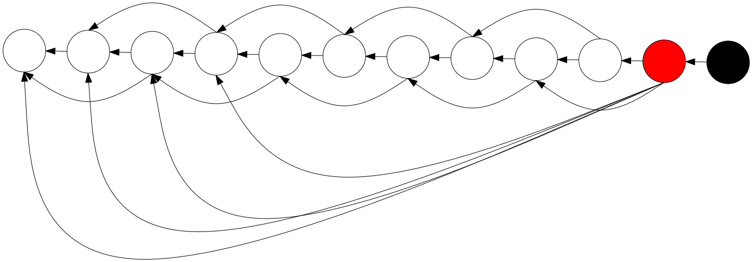
- Attempt: Show $E[\mathrm{dist}(u, root)] \leq E[\mathrm{dist}(\mathrm{amb}(u), root)]$
- Problem: false (if one doesn't assume anything about the graph structure)

For every α one can construct a graph with $E[\mathrm{dist}(u, root)] > E[\mathrm{dist}(\mathrm{amb}(u), root)]$.



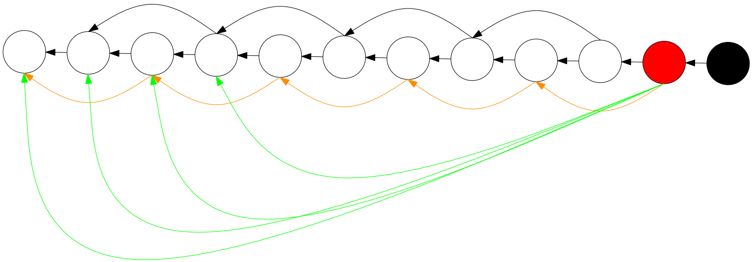
Proving bounded distance: Failed attempt #4

- Idea: Show $E[longest\_path(u, root)] \leq E[longest\_path(\mathrm{amb}(u), root)]$
- Problem: Wrong - Longest path is strictly increasing!



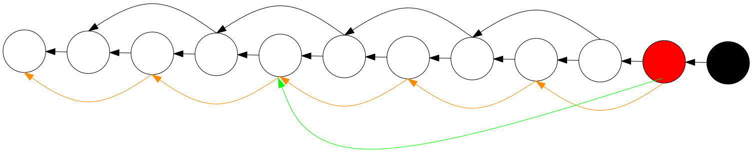
Proving bounded distance: Attempt #5

- $\phi(u)$ longest path to root without ambassador edges.
- Idea: Show $E[\phi(u)] \leq E[\phi(\mathrm{amb}(u))]$
- Problem: ?



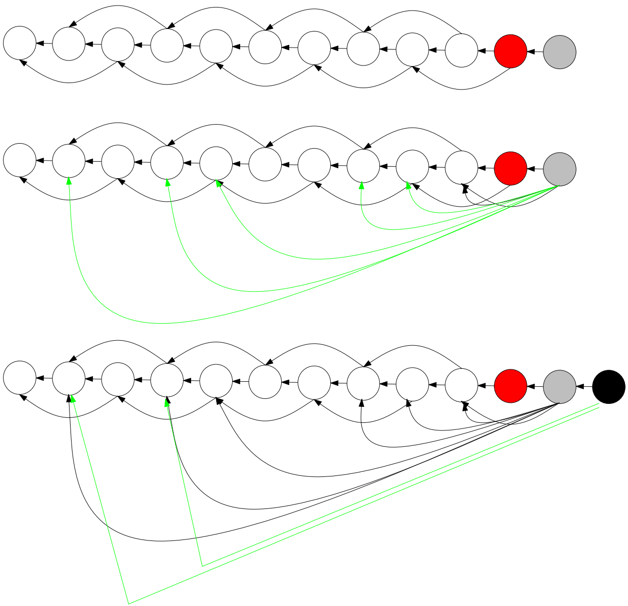
Proving bounded distance: Failed Attempt #5

- Problem: There might not be enough good edges ...



Proving bounded distance: Successful Attempt

- What if we look at two consecutive arrivals?
- The first node prepares the graph structure
- The second node exploits it



Outline

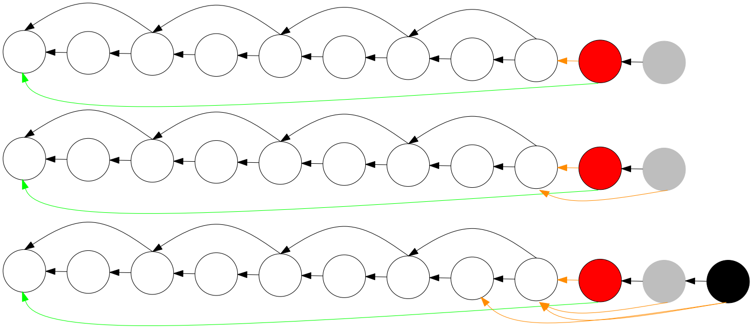
Define a potential that drops?

Define ϕ so that in expectation,

- $\phi(v) - \phi(\mathrm{amb}(\mathrm{amb}(v))) < 0$, assuming that $\phi(\mathrm{amb}(\mathrm{amb}(v)))$ is 'big'

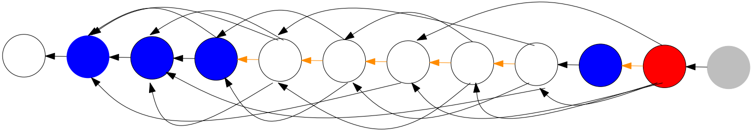
The last issue: ϕ may increase suddenly

- Problem: $\phi(\mathrm{amb}(\mathrm{amb}(u)))$ small but $\mathbf{E}[\phi(\mathbf{u})]$ huge.
- $E[\phi(v) - \phi(\mathrm{amb}(\mathrm{amb}(v)))|\phi(\mathrm{amb}(\mathrm{amb}(v)))$ is 'big'$] < 0$ fails.



The ϕ-function

$$\phi(v) = \begin{cases} 0 & \text{if } v \in seed \\ \max\begin{cases} 1 + \max\{\phi(u) : u \in \mathcal{N}(v) \setminus \{\mathrm{amb}(v)\}\} \\ 1 + \phi(\mathrm{amb}(v)) \end{cases} & \text{if } \mathrm{outdeg}(v) \\ \max\begin{cases} 1 + \max\{\phi(u) : u \in \mathcal{N}(v) \setminus \{\mathrm{amb}(v)\}\} \\ \phi(\mathrm{amb}(v)) - 2 \end{cases} & \text{otherwise} \end{cases}$$



ϕ dominates the distance to the seed

$$\mathrm{dist}_{L_t}(v, L_0) = \begin{cases} 0 \text{ if } v \in \text{seed} \\ 1 + \min_{u \in \mathcal{N}(v)} \{\mathrm{dist}_{L_t}(u, \text{seed})\} \text{ otherwise} \end{cases}$$

Using definitions, by induction $\mathrm{dist}_{L_t}(v, L_0) \leq \phi(v)$.

Hajek's theorem

We have defined ϕ so that

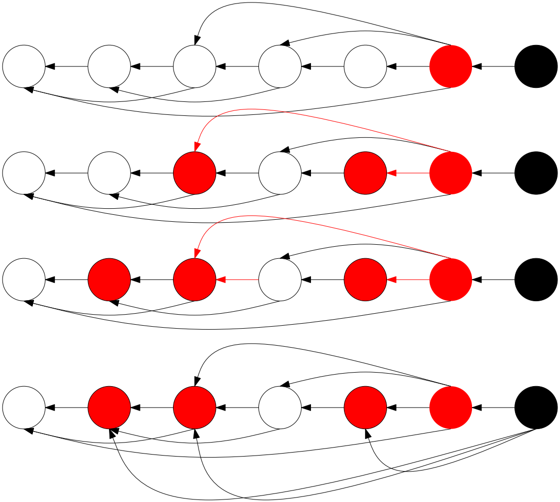
- ϕ dominates distances to seed: $\mathrm{dist}_{L_t}(v, L_0) \leq \phi(v)$
- Expected negative bias:
  $E[\phi(v) - \phi(\mathrm{amb}(\mathrm{amb}(v)))|\phi(\mathrm{amb}(\mathrm{amb}(v)))$ is 'big'$] < 0$
- $\phi(v)$ has an exponential tail

Hajek's theorem then implies:

$$E[\mathrm{dist}_{L_t}(v, L_0)] \leq E[\phi(v)] = O(1).$$

Windy Forest Line Process

- $\text{amb}(\mathbf{u_t}) = \mathbf{u_{t-1}}$.



Windy Forest Fire Process

- $\mathrm{amb}(\mathbf{u_t})$ =uniform random node.

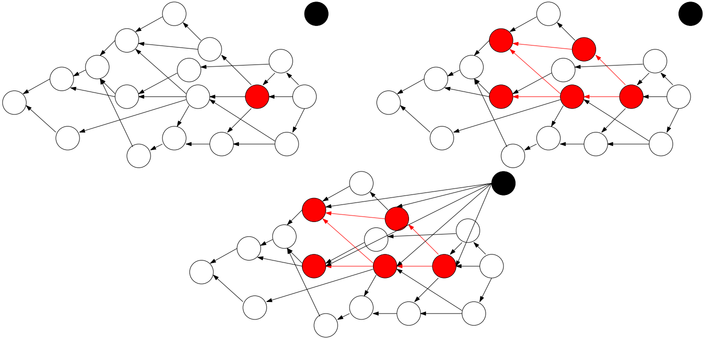
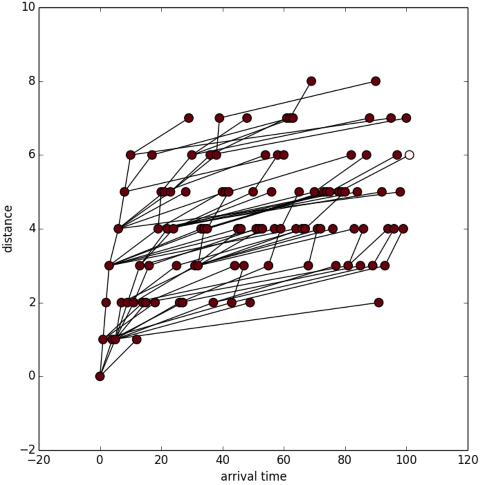


Figure: DAG
Every u has a special outgoing edge, to its ambassador

Relating the two processes: Ambassador tree

- A_t: Take G_t and consider only the edges $\{(u, \mathrm{amb}(u))\}$.



Ambassador tree

Line Fire result: $E[\mathrm{dist}_{L_t}(v, L_0)] = O(1)$,

$+$ Ambassador tree properties

$\Rightarrow$ Forest Fire result:

$$E[\mathrm{dist}_{G_t}(v, G_0)] = O(1).$$

Outline

Model

Random walk model with parameter $0 < p < 1$.
At arrival of u

- Connect to a node $\mathrm{amb}(u)$ chosen u.a.r.
- Perform random walk starting from $\mathrm{amb}(u)$.
- At every step, w.p. p, stop the random walk.
- Connect u to all nodes visited by the random walk.

Our results

Theorem 2

Let G_0 be a strongly connected graph. The Random Walk Process with parameters $p < 1/3$ and G_0 has the property of non-increasing distance to G_0, i.e., for every t,

$$\mathbb{E}[\mathrm{dist}_{G_t}(u, G_0)] = O(1).$$

Theorem 3

For all G_0, the Random Walk Process with parameters $p > 1/3$ and G_0, has the property that

$$\mathbb{E}[\mathrm{dist}_{G_t}(u, G_0)] = \Omega(\log t).$$

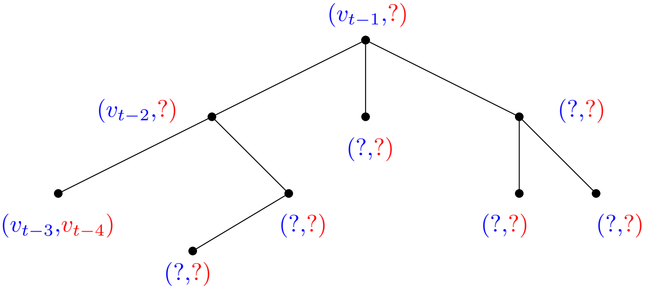
The proof

- Principle of deferred randomness. We uncover the random choices 'backwards'.
- The degree of node u distributed $deg(u) \sim \text{Geom}(p)$.
- When the random walk arrives at a node u it takes the edge with ID $X \sim \text{Uniform}(\text{Geom}(p))$.

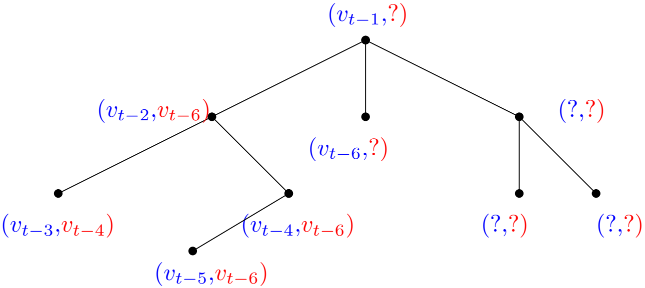
The proof

- Consider the decisions of the random walk with r steps $X_1, X_2, ..., X_r$.
- Where does us the 5'th edge of node u take? Need to know where the edges $0, 1, 2, 3, 4$ take us.
- We don't care about the other edges of u!

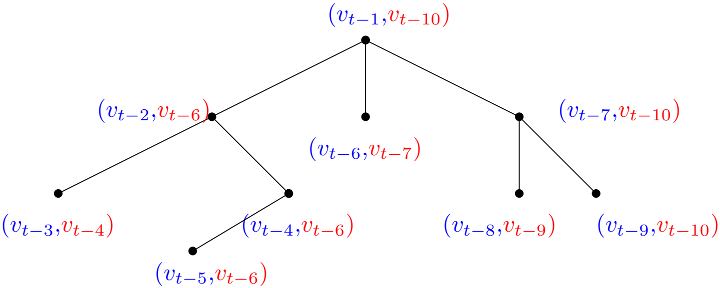
The proof



The proof



The proof


(v_{t-1}, v_{t-10})
(v_{t-2}, v_{t-6})
(v_{t-6}, v_{t-7})
(v_{t-7}, v_{t-10})
(v_{t-3}, v_{t-4})
(v_{t-4}, v_{t-6})
(v_{t-8}, v_{t-9})
(v_{t-9}, v_{t-10})
(v_{t-5}, v_{t-6})

Conclusion/Open questions

Main result:

- François Hollande is probably not too 'far' from you (if you care), i.e., Forest Fire Process and Random Walk model have constant expected distance.

Open questions for Forest Fire Process:

- We analyzed the Windy FF. What about the original one?
- Prove densification.

Thank you!