

Distance in the Forest Fire Model

How far are you from Eve?



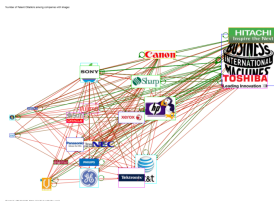
Varun Kanade, Reut Levi, Zvi Lotker,
Frederik Mallmann-Trenn, and **Claire Mathieu**

Motivation

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



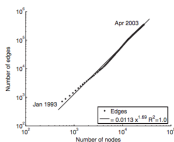
Social networks are everywhere. How do we meet new friends? Impact in economics, social science, marketing, disease propagation, ...



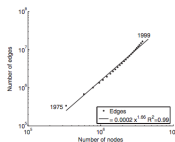
Social network feature: Densification

[Leskovec, Kleinberg and Faloutsos (2005)]

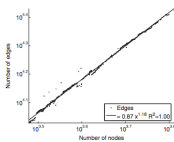
Densification on several networks: Arxiv, Patents, Autonomous systems, Email, Actors-Movies, . . .



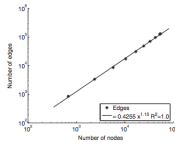
(a) arXiv



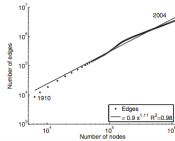
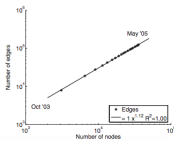
(b) Patents



(c) Autonomous Systems

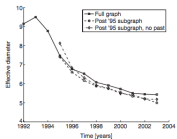


(d) Affiliation network

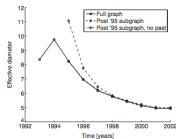


Social network feature: Distance (between two random vertices)

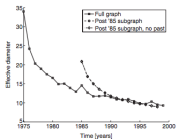
Arxiv, Patents, Autonomous systems, Email, Actors-Movies, ...



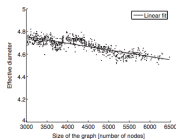
(a) arXiv citation graph



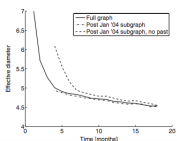
(b) Affiliation network



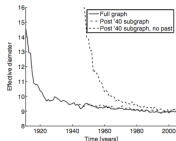
(c) Patents citation graph



(d) Autonomous Systems



(e) Email network



(f) IMDB actors to movies network

Distance: previous work

Empirical Result[KLF05]

Many social networks have bounded expected distance.

Simulation Result[KLF05]

The Forest Fire Model has expected bounded expected distance.

Theorem

Nothing proved before this work

Outline

Introduction

Related Work

1st Model: The Windy Forest Fire model

Model definition

Proof intuition

A potential function

Ambassador tree

2nd Model: Random Walk model

Conclusion/Open questions

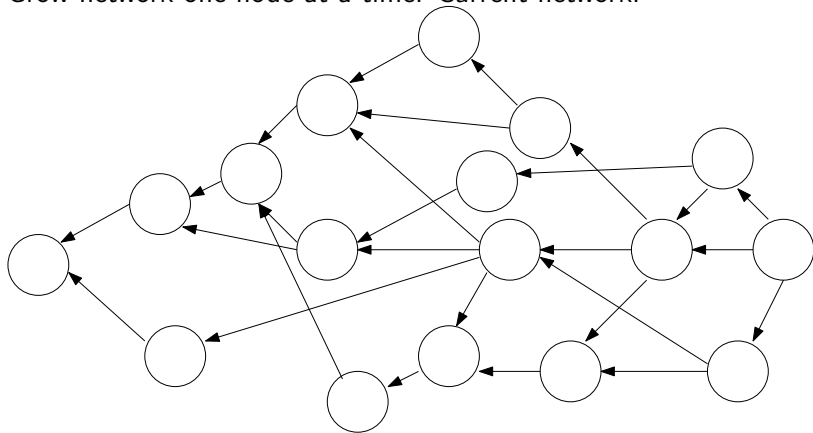
Social networks model: Forest Fire

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

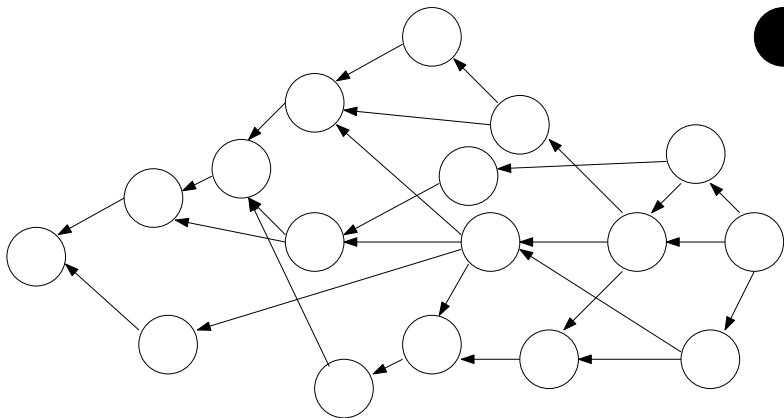


Example

Grow network one node at a time. Current network:

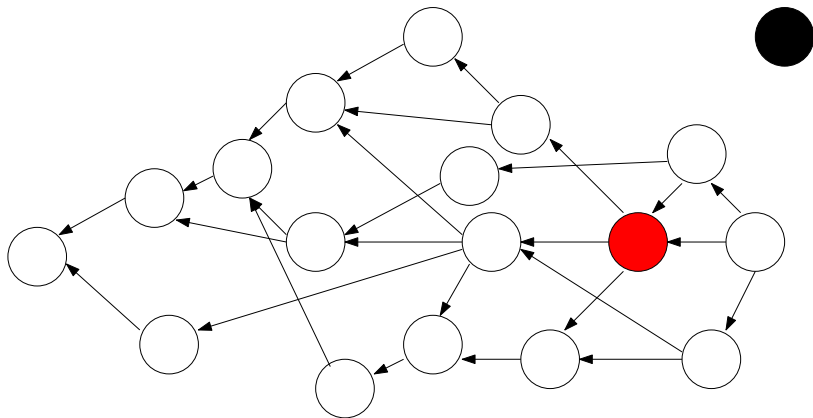


Example



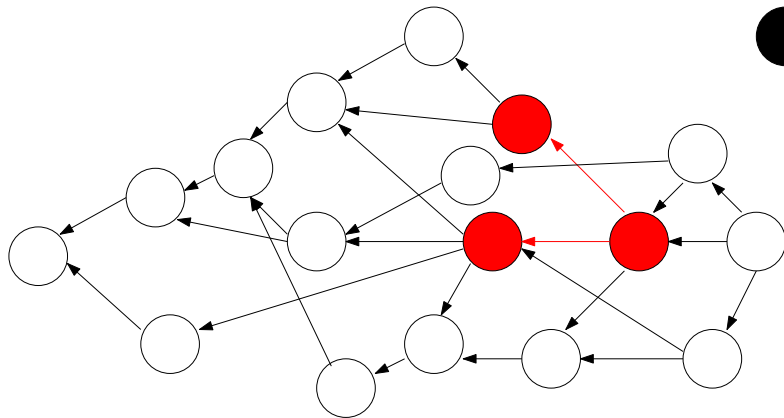
Next step: new node u arrives (black node).

Example



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

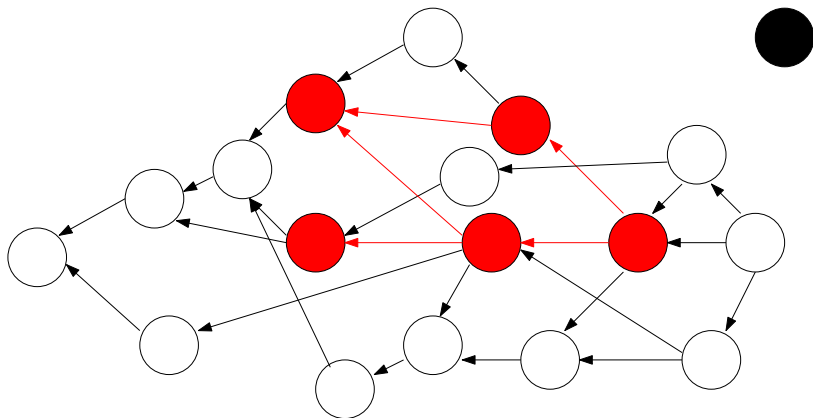
Example



Introduction to friends: start the 'forest fire'.

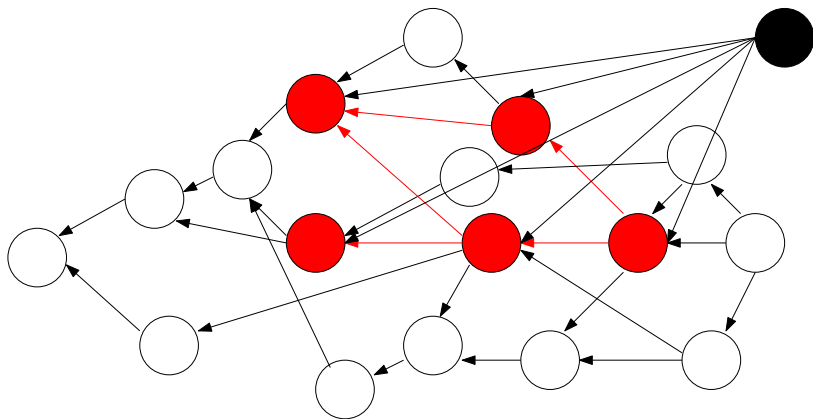
Every 'burnt' node v activates edges w.p. $\alpha/\text{outdeg}(v)$.

Example



Continue burning...

Example



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

Social network model: Forest Fire

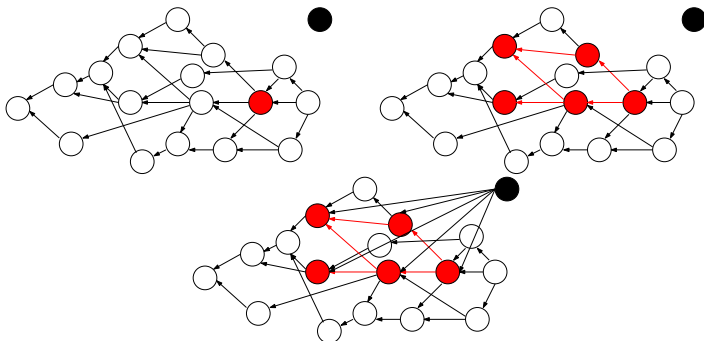


Figure: The three steps of adding a node to the network:
 1) Choose $\text{amb}(u)$ u.a.r. 2) Forest Fire (percolation) 3) Connect

Differences with original Forest Fire model

- ▶ Windy: we study 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 (> 1).



Forest Fire Expected Diameter Theorem

Empirical Result[KLF05]

Many social networks have bounded expected distance.

Simulation Result[KLF05]

The Forest Fire Model has bounded expected distance.

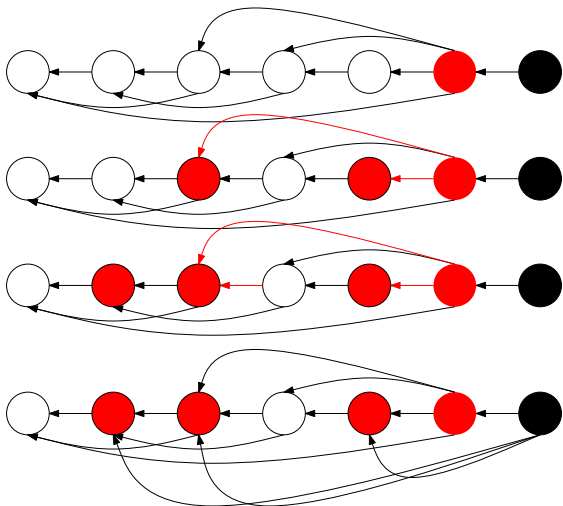
Theorem 1

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

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

A simpler model: a Line Process

- From now on: $\text{amb}(\mathbf{u}_t) = \mathbf{u}_{t-1}$.



Outline

Introduction

Related Work

1st Model: The Windy Forest Fire model

Model definition

Proof intuition

A potential function

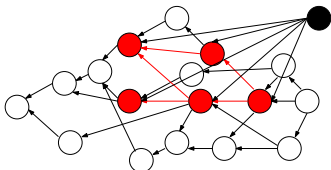
Ambassador tree

2nd Model: Random Walk model

Conclusion/Open questions

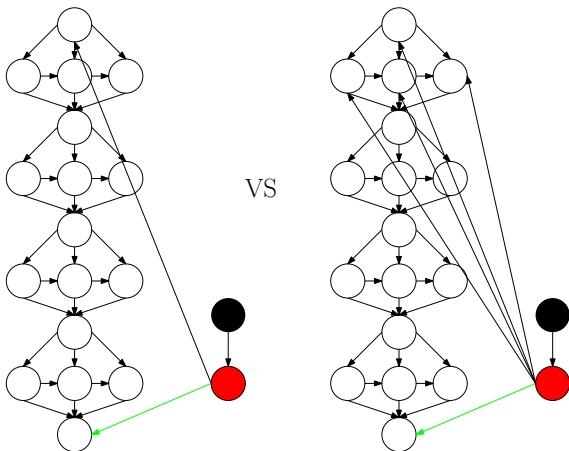
Proving bounded distance on line: Failed attempt #1

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



Proving bounded distance on line: Failed attempt #2

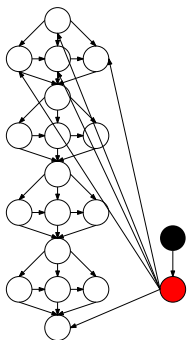
- ▶ Attempt: Simplify graph by insertion/deletion of edges
- ▶ Problem: expected distance is not monotone under insertion/deletion of edges



Proving bounded distance on line: Failed attempt #3

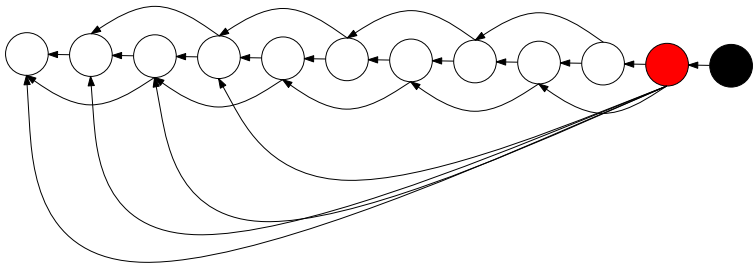
- ▶ Attempt: Show $E[\text{dist}(u, \text{root})] \leq E[\text{dist}(\text{amb}(u), \text{root})]$
- ▶ Problem: false (if one doesn't assume anything about the graph structure)

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



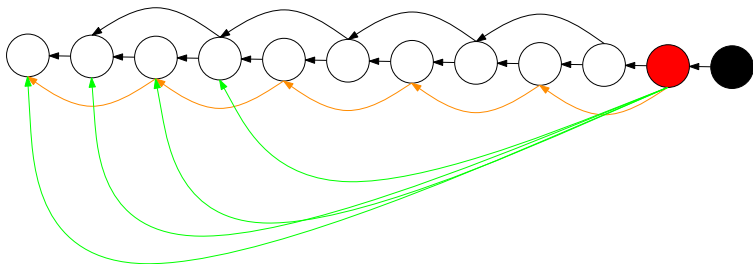
Proving bounded distance on line: Failed attempt #4

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



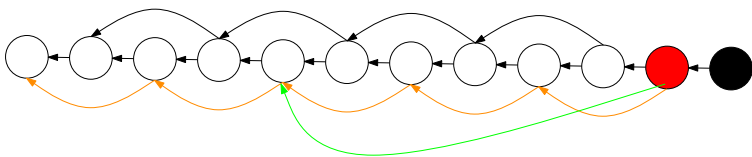
Proving bounded distance on line: Attempt #5

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



Proving bounded distance on line: Failed Attempt #5

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



Proving bounded distance on line: Successful Attempt

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

Define ϕ so that in expectation,

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

Hajek's theorem

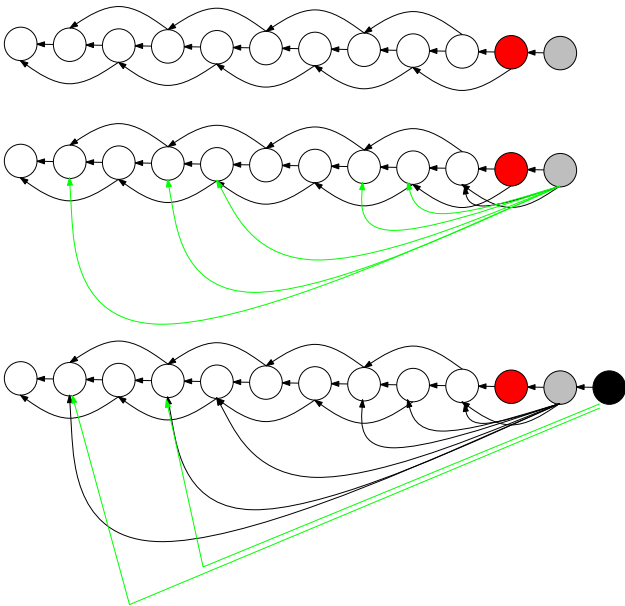
We define ϕ so that

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

Hajek's theorem then implies:

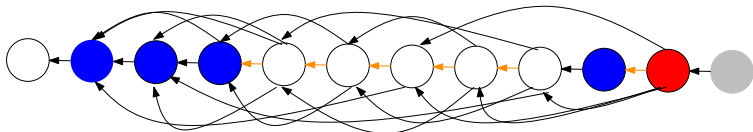
$$E[\text{dist}_{L_t}(v, L_0)] \leq E[\phi(v)] = O(1).$$

Bounded expected distance on Line process



The ϕ -function

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



Back to the Forest Fire Process

- ▶ So far: Line process, $\text{amb}(\mathbf{u}_t) = \mathbf{u}_{t-1}$.
- ▶ Forest fire process: $\text{amb}(\mathbf{u}_t) = \text{uniform random node}$.

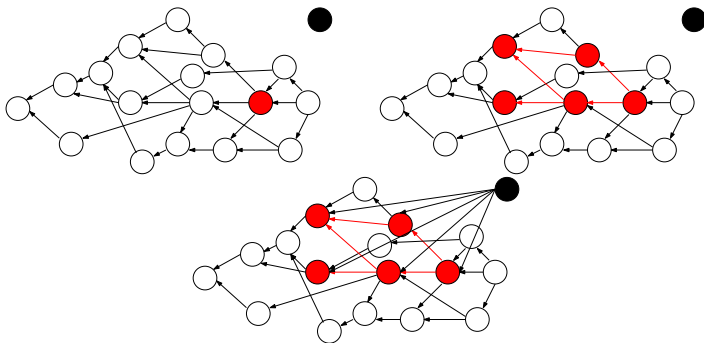
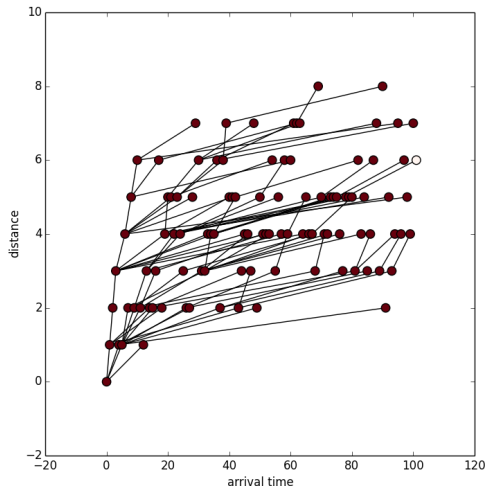


Figure: DAG

Consider only the edges $\{(u, \text{amb}(u))\}$

Relating the two processes: Ambassador tree

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



Ambassador tree

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

+ Ambassador tree properties

\Rightarrow Forest Fire result:

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

Outline

Introduction

Related Work

1st Model: The Windy Forest Fire model

Model definition

Proof intuition

A potential function

Ambassador tree

2nd Model: Random Walk model

Conclusion/Open questions

Another Social Network model: Random Walk

Random walk model with parameter $0 < p < 1$.

At arrival of u

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

Results for Random Walk Model

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 ,

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

Theorem 3

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

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

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.

Conclusion/Open questions

Main approach:

- ▶ Identify critical, surprising features of existing networks
- ▶ Propose mathematical model of network and of features
- ▶ Realistic and simple (a contradiction in terms)
- ▶ Sanity check: Prove that model has features

Better (less ad hoc) approach:

- ▶ Identify critical, surprising features of existing networks
- ▶ Take existing mathematical model of network
- ▶ Propose mathematical model of features, realistic and simple (a contradiction in terms)
- ▶ Sanity check: Prove that model has features

Then: Understanding \implies can propose improved algorithm...

Thank you!