Coalescing-Branching (Cobra) Random Walks

Rajmohan Rajaraman[1]

Northeastern University, Boston, USA

27 June 2016 - ICTS, Bangalore, India

[1] Joint work with Chinmoy Dutta, Michael Mitzenmacher, Gopal Pandurangan, and Scott Roche

Introduction

There are many modes by which epidemics and information spread on networks:

- Simple random walks
- Rumor spreading mechanisms
- Random walks with speed-up techniques
- Parallel random walks
- SIR and SIS epidemics models

Introducing Cobra Walks

- Some node of a graph starts with a token/infection/piece of information
- Spreads information by sending it to k neighbors
- Any node infected at time t spreads to k neighbors
- Process continues forever

Coalescing-branching (cobra) walks defined

- Static graph $G = (V, E)$ on n nodes.
- Pick any $u' \in V$ and place a token at u'.
- For each $u \in V$, if there is more than one token at node, all tokens at that node coalesce into one.
- Each token then becomes k tokens (branching).
- For each $u \in V$, each token at u independently chooses u.a.r some $v \in N(u)$ and moves to it.
- For this talk, $k = 2$.

Illustration of a cobra walk

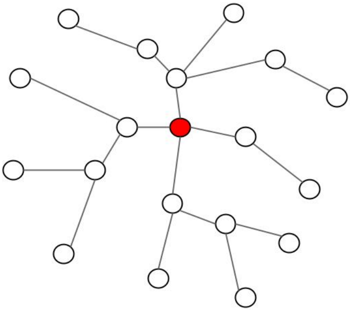


Illustration of a cobra walk

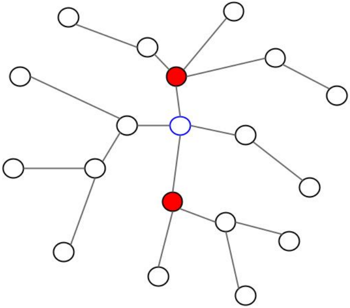


Illustration of a cobra walk

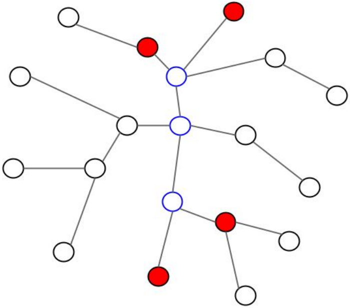


Illustration of a cobra walk

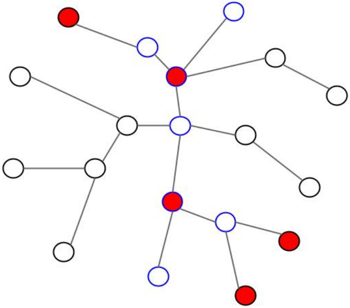


Illustration of a cobra walk

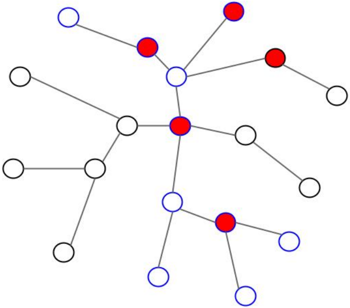


Illustration of a cobra walk

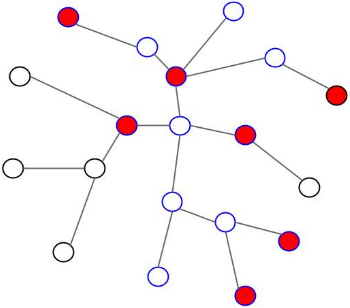


Illustration of a cobra walk

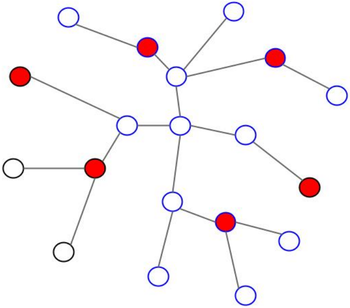
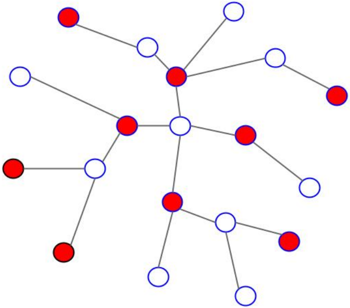


Illustration of a cobra walk



What do we want to know about cobra walks?

1. How long does it take for a cobra walk to visit every node in a graph?
   - For standard random walks $\Theta(n^3)$ [Feige 1995]

What do we want to know about cobra walks?

1. How long does it take for a cobra walk to visit every node in a graph?
   - For standard random walks $\Theta(n^3)$ [Feige 1995]
2. What are the long term-dynamics of a cobra walk?
   - What percentage of nodes of a graph are "infected" as $t \to \infty$?
   - Is there an analog to a stationary distribution or mixing time?

What do we want to know about cobra walks?

1. How long does it take for a cobra walk to visit every node in a graph?
   - For standard random walks $\Theta(n^3)$ [Feige 1995]
2. What are the long term-dynamics of a cobra walk?
   - What percentage of nodes of a graph are "infected" as $t \to \infty$?
   - Is there an analog to a stationary distribution or mixing time?
3. What are possible applications of cobra walks?

What do we want to know about cobra walks?

1. **How long does it take for a cobra walk to visit every node in a graph?**
   - For standard random walks $\Theta(n^3)$ [Feige 1995]
2. What are the long term-dynamics of a cobra walk?
   - What percentage of nodes of a graph are "infected" as $t \to \infty$?
   - Is there an analog to a stationary distribution or mixing time?
3. What are possible applications of cobra walks?

Challenges in analyzing cobra walks

- Size of active set is non-monotonic.
- Lack of independence in number and distribution of pebbles
- Dynamics of cobra walks governed by graph topology (e.g. number of common neighbors of nodes)...
- ... and by current distribution of pebbles on the walks.

Cobra Walk as a Model for Epidemics

- Cobra walks are a discrete-time version of the SIS (susceptible-infected-susceptible) epidemic mode with probability 0 of extinction:
  - Branching simulates probability of infecting neighbors
  - Coalescing simulates receiving infection from multiple neighbors
  - Recovery time (return to susceptible state) is one time step
- Extensive work in related contact processes [Harris 1974; Durrett 1980; Ganesh-Massoulié-Towsley 2005; Berger-Borgs-Chayes-Saberi 2005]
- Most work focuses on questions of extinction time and persistence of the epidemic [Durret 2010, Kessler 2007], [Draief and Ganesh 2011]
- Some models involve a mean-field approximation of some part of the epidemic. [Van Mieghem 2011].

Cobra Walks vs Rumor Spreading

- Rumor spreading is well-studied [Chierichetti and Panconsesi 2010, Chierichetti et al 2011, Giakkoupis and Sauerwald 2012, Fountoulakis and Panagiotou 2010, and many others..]
- Fast coverage of any graph $O(n \log n)$ [Feige et al 1990]
- Set of nodes with rumor monotonically non-decreasing
- Message complexity can be high

Cobra Walks vs Parallel Random Walks

Parallel (independent) random walks [Alon et al 2008, Elsaesser and Sauerwald 2009, Efremenko and Reingold 2009]

- Can provide significant speed up for many classes of graphs (e.g. expanders)
- Cover time not necessarily equivalent to cobra walks
- Independence of walks a powerful tool in proofs

Bounds on Cover Time of Cobra Walks

Let n be the number of nodes in the given graph.

- Lower bound: $\Omega(n \log n)$.

Bounds on Cover Time of Cobra Walks

Let n be the number of nodes in the given graph.

- Lower bound: $\Omega(n \log n)$.
- Trees: $O(n \log n)$ [Dutta-Pandurangan-R-Roche 2013].

Bounds on Cover Time of Cobra Walks

Let n be the number of nodes in the given graph.

- Lower bound: $\Omega(n \log n)$.
- Trees: $O(n \log n)$ [Dutta-Pandurangan-R-Roche 2013].
- d-Dimensional grid: $O(n^{\frac{1}{d}})$ [Dutta-Pandurangan-R-Roche 2013].

Bounds on Cover Time of Cobra Walks

Let n be the number of nodes in the given graph.

- Lower bound: $\Omega(n \log n)$.
- Trees: $O(n \log n)$ [Dutta-Pandurangan-R-Roche 2013].
- d-Dimensional grid: $O(n^{\frac{1}{d}})$ [Dutta-Pandurangan-R-Roche 2013].
- Conductance Φ and degree d: $O\left(\frac{d^4 \log^2 n}{\Phi^2}\right)$ [Mitzenmacher-R-Roche 2016].
  - $O(\log n)$ bound for expanders [Cooper-Radzik-Rivera 2016]

Bounds on Cover Time of Cobra Walks

Let n be the number of nodes in the given graph.

- Lower bound: $\Omega(n \log n)$.
- Trees: $O(n \log n)$ [Dutta-Pandurangan-R-Roche 2013].
- d-Dimensional grid: $O(n^{\frac{1}{d}})$ [Dutta-Pandurangan-R-Roche 2013].
- Conductance Φ and degree d: $O\left(\frac{d^4 \log^2 n}{\Phi^2}\right)$ [Mitzenmacher-R-Roche 2016].
  - $O(\log n)$ bound for expanders [Cooper-Radzik-Rivera 2016]
- Cover time for an arbitrary graph is $O(n^{11/4} \log n)$ [Mitzenmacher-R-Roche 2016].

Open Problems

- Close the substantial gap between upper and lower bounds for general graphs.
- Tight bounds in terms of conductance and expansion of the graph.
- Analyze other measures of interest.
- Better bounds for special graph classes.
- Explore implications for real-world processes.

Conductance of a Graph

- A measure of how well the graph locally expands everywhere.
- Define the *volume* of a set $S \subseteq V$ to be the sum of the degrees of the vertices in S.
- Let the *conductance* of a set $S \subseteq V$ of vertices be the ratio of the number of edges between S and $V \setminus S$ to the volume of S.
  - Formally, define $\phi(S) = |\partial(S)|/\text{vol}(S)$, where $\partial(S) = \sum_{(u,v):u \in S, v \notin S} 1$ and $\text{vol}(S) = \sum_{u \in S} d(u)$.
- Then the conductance Φ of the graph is the minimum conductance of a set whose volume is at most half the volume of V.
  - Formally, Φ is $\min_{S:\text{vol}(S) \leq \text{vol}(V)/2} \phi(S)$.
- A constant-degree constant-conductance graph is referred to as an *expander*.

Challenges and Techniques

Challenges

- The number of active nodes is non-monotonic.
- Though the "pebbles" make independent branching moves, coalescing introduces dependencies.

Analysis plan:

- Introduce W_{alt}, a process that stochastically dominates cobra walks.
- Break the process into $O(\log n)$ epochs of length $O(\log n/\Phi^2)$.
- In each epoch, show that any node v has a constant probability of being covered.
- This gives the desired high probability bound of $O(\log^2 n/\Phi^2)$ for cover time.

Description of W_{alt}.

- There are δn pebbles distributed arbitrarily around the graph.
- No more branching or coalescing occurs.
- Pebbles are arbitrarily labeled using a total order.
- First two pebbles at v at time t continue to move independently.
- Third and higher-ranked pebbles at v at time t chose one of the destinations of the first two pebbles independently with probability $1/2$.

Illustration of W_{alt}

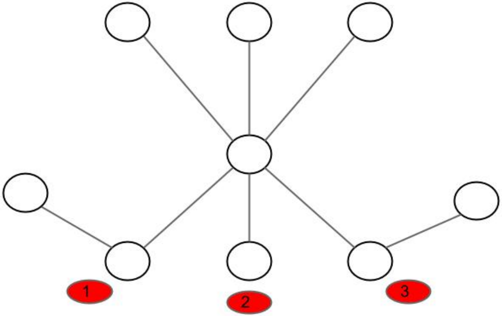


Illustration of W_{alt}

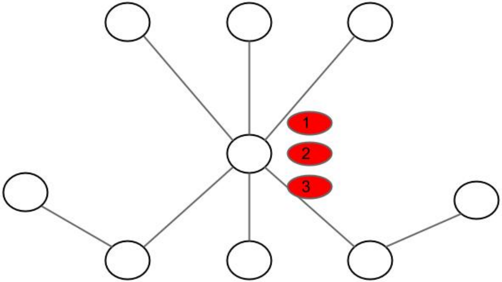


Illustration of W_{alt}

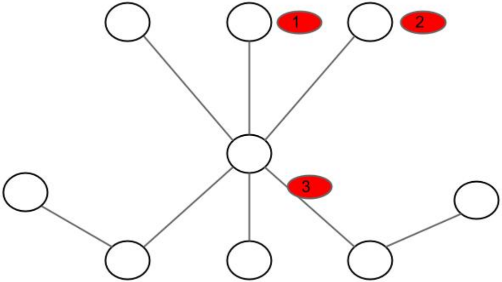
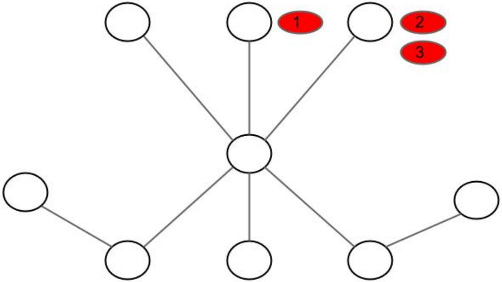


Illustration of W_{alt}



W_{alt} covers G in $O(\log^2 n/\Phi^2)$ rounds

Main idea:

- Show that each node of G will be covered by at least one pebble in $O(\log n/\Phi^2)$ steps with constant probability,
- Repeat $O(\log n)$ times to get a high probability bound on every node.

Therefore,

- Need to show that probability that any pebble i is at vertex v at time $s = \Theta(\log n/\Phi^2)$ is $\Theta(\frac{1}{n})$.
- Can't use standard methods because walks are not fully independent.

Analysis of W_{alt}

- Let E_i be the event that pebble i visits (arbitrary) node v at time s. Then the probability that v gets visited by any of the pebbles is:

$$\Pr\left[\bigcup_i E_i\right] \geq \sum_i \Pr[E_i] - \frac{1}{2}\sum_{i \neq j} \Pr[E_i \cap E_j] \quad (1)$$

$$(2)$$

- We can use standard random walk theory to show that $Pr[E_i = 1] \geq \frac{1}{2n}$
- The challenge is to upper bound $Pr[E_j \cap E_i]$ suitably.
  - We show that $Pr[E_j \cap E_i]$ is at most $\frac{2}{n^2+n} + \frac{1}{n^4}$ for suitably large s.

Bound on $\Pr[E_i \cap E_j]$

- Fix the walk of pebble i, and and make the assumption that if j and i arrive at the same time at the same node, j is the third or higher priority pebble at that node.
- View the random walks of i and j as a single random walk over the tensor product graph $G \times G$.
  - Cartesian product $V(G) \times V(G)$ as vertex set, and an edge set defined as follows: vertex $(u, u') \in V(G \times G)$ has an edge to $(v, v') \in V(G \times G)$ if and only if $(u, v), (u', v') \in E(G)$.
  - Make edges directed, and attach weights to them such that the walk on the directed graph $\mathcal{D}$ is isomorphic to the movement of pebbles i, j in W_{alt} on G.
- Using directed Cheeger inequality [Chung 2005]:
  - Argue that second smallest eigenvalue of normalized Laplacian of $\mathcal{D}$ is at least $1/\Phi^2$.
  - Argue that walk converges in $O(\log n/\Phi^2)$ time to stationary probability distribution of $2/(n^2 + n)$.

Can We Beat the Trivial Random Walk Bound?

Cover time of a standard random walk:

- Upper bound of $O(mn)$, where m is number of edges, by considering an Eulerian walk over a spanning tree.
- Tight bound of $4n^3/27 + o(n)$ [Brightwell-Winkler; Feige].

Analysis using hitting time:

- Consider a cobra walk starting from a node s.
- What is the hitting time to an arbitrary node t?
  - Can try to follow the pebble nearest to t.
  - Leads to the notion of a *biased random walk*.
- We show that hitting time is $O(n^{11/4})$.
- Invoke Matthews Theorem to obtain desired cover time bound.

Biased Random Walks

ε-Biased random walks introduced by [Azar-Broder-Karlin-Linial-Phillips '92]:

- With probability $1 - \varepsilon$, make a random walk.
- With probability ε, controller selects an arbitrary neighbor to move to.

Theorem: (Azar et al) Consider a biased random walk, and arbitrary vertex x. There is an ε-biased random walk for which the stationary probability at x is at least

$$\frac{\mathrm{vol}(S)}{\mathrm{vol}(S) + \sum_{v \neq x} (1-\varepsilon)^{\Delta(v,x)-1} \cdot \delta(v)},$$

where $\Delta(v, x)$ is the length of the shortest path from v to x.

Inverse-Degree-Biased Walks

Inverse-degree-biased walk with target x:

- If the walk is at x, then it moves to a uniformly random neighbor.
- If the walk is at $v \neq x$:
  - With probability $1 - 1/\delta(v)$, it moves to a uniformly random neighbor.
  - With probability $1/\delta(v)$, controller selects neighbor to move the walk to.

Lemma: [Mitzenmacher-R-Roche 2016] Hitting time to x of cobra walk is stochastically dominated by hitting time to x of inverse-degree-biased walk with target x.

Hitting Time of an Inverse-Degree-Biased Walk

- Consider a shortest path from a given source to target x.
- Bound the hitting time as the sum of the traversal times along the edges of P.
- Bound the traversal time from u to v using the return time of another inverse-degree-biased walk.
- Derive a variant of the Azar et al theorem in terms of a revised notion of shortest path, using a beautiful theorem of Metropolis et al.
  - **Theorem:** [Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953] For every graph G and strictly positive probability distribution π over the vertices of G, there exists a Markov chain M whose stationary distribution is π.
- Further calculations using the fact that P is a shortest path yield the $O(n^{11/4})$ bound.

Cover time results for

- Grids and trees: Tight.
- Conductance-based analysis: In the same ball-park, but not tight.
- General graphs: Could be far from tight.

Many open problems remain:

- $O(n \log n)$ time for general graphs?
- Better bounds in terms of expansion, conductance, etc.?
- Analyze other measures of interest.
- Viewed as an information dissemination method, what is message complexity?
- Analysis for other classes of graphs.
- Explore connections to SIS epidemics.