Analysis of Network Diffusion and Distributed Network Algorithms

Rajmohan Rajaraman
Northeastern University, Boston

May 2012

Overview of the 4 Sessions

- Random walks
- Percolation processes
  - Branching processes, random graphs, and percolation phenomena
- Rumors & routes
  - Rumor spreading, small-world model, network navigability
- Distributed algorithms
  - Maximal independent set, dominating set, local balancing algorithms

Random Walks

Rajmohan Rajaraman
Northeastern University, Boston

May 2012

Outline

- Basic definitions and notation
- Applications
- Two results:
  - Mixing time and convergence of random walks
  - Cover time of random walks
- Applications to clustering
- Techniques:
  - Probability theory
  - Spectral graph theory

What is a Random Walk?

- Let G be an arbitrary undirected graph
- A walk starts at an arbitrary vertex v_0
- At the start of step t, the walker moves from v_{t-1} to vertex v_t chosen uniformly at random from neighbors of v_{t-1} in G
- For all $t > 0$, v_t is a random variable

Notation

- Let G be an arbitrary undirected graph and A be its adjacency matrix
  - A_{ij} is 1 whenever there is an edge (i,j)
- Define the random walk matrix M
  - M_{ij} is A_{ij}/degree(i)
- Let x denote the initial probability distribution (row) vector
- After t steps, the probability distribution vector equals xM^t

Definitions

- Stationary distribution
  - Probability vector π such that $\pi M = \pi$
- Hitting time h_{ij}
  - Expected time for random walk starting from i to visit j
- Cover time C
  - Expected time for random walk starting from an arbitrary vertex to visit all nodes of G
- Mixing time
  - Time it takes for the random walk to converge to a stationary distribution

Questions of Interest

- Stationary distribution:
  - Do they always exist?
  - Is the stationary distribution unique?
  - Does a random walk always converge to a stationary distribution? If it does, what is the mixing time?
- Hitting time:
  - For a given graph G and vertices i,j, what is the hitting time h_{ij}
- Cover time:
  - For a given graph G, what is its cover time C?

Applications

- Probabilistic process whose variants capture social and physical phenomena
  - Brownian motion in physics
  - Spread of epidemics in contact networks
  - Spread of innovation and influence in social networks
  - Connections to electrical networks
  - Markov chains arise in numerous scenarios
- Pseudo-random number generators
  - Random walk in an expander graph is an efficient way to generate pseudo-random bits from a small random seed
- Use in randomized algorithms
- Google's PageRank
  - PageRank is the probability vector of the stationary distribution of an appropriately defined random walk

Stationary Distribution and Mixing Time

- Lemma 1: A stationary distribution always exists and is unique
- For d-regular undirected graphs G, let $\lambda(G)$ denote the second largest eigenvalue of M
- Theorem 1: The random walk is within ε of the stationary distribution in

$$O\left(\frac{\ln(n/\varepsilon)}{1-\lambda(G)}\right) \text{ steps}$$

  – For non-regular graphs, replace M by a normalized version

Cover Time

- Matthews Bound:
  - Let h_{max} be the maximum hitting time
  - The cover time is at most $h_{max}\ln(n)$
- Exercise: Prove that time for a random walk to cover every vertex is $O(h_{max} \log(n))$ whp
- Theorem 2: For any m-edge n-vertex undirected graph G, the cover time is O(mn)
- [Mitzenmacher-Upfal 04]

Lovasz-Simonovits Theorem

- Lazy random walk:
  - With probability ½, walk stays at current node;
  - With probability ½, does regular random walk
- Theorem 3: [LS 93] For any initial probability distribution and every t, we have

$$I_t(x) \le \min(\sqrt{x}, \sqrt{2m - x})\left(1 - \frac{\phi^2}{2}\right)^t + \frac{x}{2m}$$

$$p_t(e_1) \ge p_t(e_1) \ge \cdots \ge p_t(e_{2m})$$

Define $I_t(k) = \sum_{i=1}^{k} p_t(e_i)$ and extend to interval $[0, 2m]$ by linear interpolation

Sparsest Cut

- Conductance: Φ measures the "expansion" of a graph

$$\phi = \min_{S \subset V, S \neq \emptyset} \frac{e(S, V-S)}{\min(e(S), e(V-S))}$$

- Finding the cut (S,V-S) that yields the above minimum ratio is the sparsest cut problem
- LP rounding: Yields O(log(n))-approximation [Leighton-Rao 88, Linial-London-Rabinovich 94]
- SDP rounding: Yields O(√log(n)) approximation [Arora-Rao-Vazirani 05]

Local Clustering

- Suppose you are given a massive graph and want to find a "good" cluster containing a given vertex v
  - Good means low conductance
- Approach: Solve the sparsest cut problem and return the cluster containing v
  - Too expensive
- Local clustering [Spielman-Teng 08, Andersen-Chung-Lang 08]:
  - Start a random walk from v, maintaining the probability vector for each vertex
  - Keep zeroing out vertices that have very low probability
  - LS Theorem helps in showing that in time nearly proportional to the size of the cluster, can achieve close to desired conductance

Take Away Messages

- Random walk and related processes
  – Arise in several scenarios
  – Are useful primitives for designing fast algorithms
  – Yield effective and practical pseudo-random sources
- Analysis tools for random walks
  – Basic probability (Markov's inequality, Chernoff-type bounds, Martingales)
  – Spectral graph theory