Lecture 16 — Nov 5

Lecturer: John Byers *BOSTON UNIVERSITY* *Scribe: Vatche Ishakian*

In today's lecture, we were presented with **Graph Toplogies** and **Markov Chain Analysis** as part of our study of the PageRank algorithm.

16.1 Graph Topologies

This section will present the graphs of particular interest in our study.

Web level where verticies represent pages and edges represent Hyperlinks

Router Level where verticies represent routers, end hosts (IP-address) and edges represent links (physical connectivity)

AS-Level where vertices represent Autonomous Systems and edges represent peer relationships (customer/provider relationship)

There are other types of graphs some from the domain of Biology. Like the protein-protein interaction networks. Other interesting graphs are like Facebook graphs (social networks). Where vertices are people and edges are the relationships.

16.2 Graph Qualities

These graphs have some interesting qualities

- Hidden:
  - Edges and nodes are not known in advance
  - Learning the graph requires measuring it (can get pretty complicated)
- Dynamic:
  - Edges and nodes are changing.
- Growing:
  - These networks are growing

16.3 Markov chain Analysis

Definition 1: Given a directed graph G = G(V,E), a Transition probability TP is is the pairwise probability of moving from node Vi → Vj.

$TP_{i,j} = Pr(X_t = j | X_{t-1} = i)$

Markov property implies that the markov chain is uniquely defined by a one step probability matrix.

$$M = \begin{bmatrix} TP_{0,0} & \cdots & TP_{0,j} & \cdots \\ \vdots & \ddots & \vdots & \cdots \\ TP_{i,0} & \cdots & TP_{i,j} & \cdots \end{bmatrix}$$

Let $\overline{P} = (P_1, P_2, ..., P_n)$ be a vector giving the distribution of states, we have the following property $P_{(t)} = P_{(t-1)}$M.

Definition 2: A Steady state (stationary distribution) of a Markov chain is a probability distribution $\overline{p}$ such that $\overline{p}$=$\overline{p}$M where $P_i = \lim_{T\to\infty}$ [Prob. of walker being in state i at time T].

Definition 3: Cover time is the expected time to visit all of the nodes in the graph by a random walk stating from vertex v

Definition 4: Mixing time is the expected number of steps needed to achieve probability distribution over state space that is "close" ($< \delta$) to a steady state.

16.3.1 Expander graphs [3]

Simply put, an expander graph is a sparse graph which has high connectivity properties.

Denote a graph by G = (V,E) and $|V|$ = n. We allow self loops and multiple edges in the graph.

For S,T $\in$ V denote the set of all edges between S and T by E(S, T) = (u, v)—u $\in$ S, v $\in$ T, (u, v) $\in$ E

define the Edge Boundary of a set S, denoted δ S, as δS = E(S,$\overline{S}$). This is actually the set of outgoing edges from S.

define the Expansion Parameter of G, denoted h(G) as: h(G) = $\min_{S||S|\leq\frac{n}{2}} \frac{|\delta(S)|}{|S|}$

A family of Expander graphs Gi where i $\in$ N is a collection of graphs with the following properties:

- The graph Gi is a d-regular graph of size n_i.
- For all i, h(Gi)>= $\epsilon > 0$.

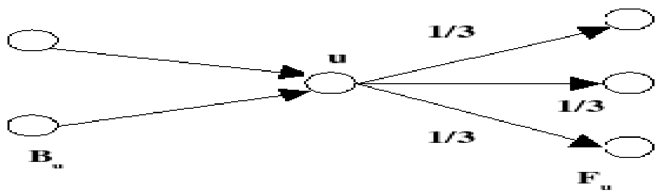
One of the interesting properties based on which the page rank algorithm converges is due to the fact that the web is an expander-like graph.[1]

16.4 Basic Page rank Algorithm:

The basic Page Rank algorithm is based on the following
1) What would a "Random Surfer" see?
2) Most frequently hit pages $\rightarrow$ most highly ranked.

Using the markov chain Analysis we compute a steady state probability $\vec{P}$, then sort the pages according to their probability. Now the "Random Surfer" will see is:

- Build a Markov chain
- Compute steady state probabilities



page u
F_u: forward lengths of u
B_u: backward lengths of u
A slightly simplified version of page rank assumes that $R(u) = c \sum_{v \in B_u} \frac{R(v)}{|F_u|}$ where c < 1

To compute the ranking of pages, iteratively run a step of the chain on $\vec{R} = [1/n.....1/n]$ until δ (where you were and where are you now) is below a tolerance $\delta = ||R_i - R_{i+1}||$

16.5 Bibliography

[1] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The Pagerank Citation Ranking: Bringing Order to theWeb. Technical report, Stanford Digital Library Technologies Project, 1998.

[2]M. Mitzenmacher and E. Upfal. Probability and Computing Cambridge University Press, 2004

[3]Lecture notes for a course on expanders given by Nati Linial and Avi Wigderson at the Hebrew University, Israel. http://www.math.ias.edu/ boaz/ExpanderCourse/