In today’s lecture, we were presented with **Graph Topologies** 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 vertices represent pages and edges represent Hyperlinks
- **Router Level** where vertices 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 $Vi \rightarrow Vj$.

$$TP_{i,j} = Pr(X_t = j|X_{t-1} = i)$$
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 \( \vec{P} = (P_1, P_2, \ldots, P_n) \) 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 \( \vec{p} \) such that \( \vec{p} = \vec{p}M \) where \( P_i = \lim_{T \to \infty} [\text{Prob. of walker being in state } i \text{ at time } T] \).

**Definition 3:** Cover time is the expected time to visit all of the nodes in the graph by a random walk starting 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 \( \delta S \), as \( \delta S = E(S, \overline{S}) \). 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 \( G_i \) where \( i \in \mathbb{N} \) is a collection of graphs with the following properties:

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

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 → 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

$R(u) = c \sum_{v \in B_u} \frac{R(v)}{|F_u|}$ where $c < 1$

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

