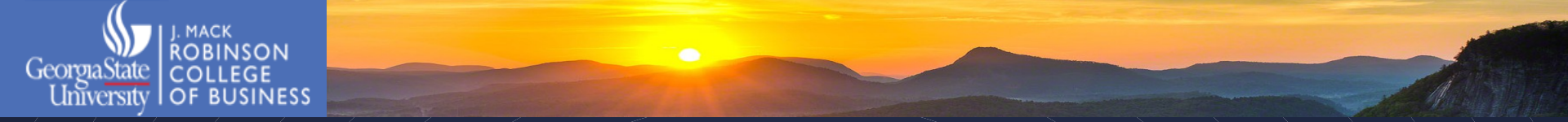

CIS 8045

Neo4j: Concepts in Graph Network Analysis

What can be done with Neo4j?

- Recommendation
- Path Tracing
- Influence/Node Importance
- Node Similarity

Influence/Node Importance

- Two types of basic guidance
  - Information spread within a social network
  - Critical nodes in a road or computer network

- Degree centrality (including incoming and outgoing degrees):

$$C_n = deg(n)$$

- PageRank (variants: ArticleRank, Personalized PageRank, etc.)

$$PR(A) = (1 - d) + d * (PR(N1)/C(N1) + ... + PR(Nn)/C(Nn))$$

- Eigenvector centrality

$$M.X = \lambda.X$$

- Closeness centrality $C_n = (N-1) / \sum d(n, m)$

- Betweenness centrality: $C_n = \sum \sigma(u, v \mid n) / \sum \sigma(u, v)$

Similarity

- **Set-based measures**: Compare the content of two sets globally. For instance, sets (A, B, C) and (C, D, B) have two common elements.
  - **Overlapping similarity** $O(A, B) = |A \cap B| / min(|A|, |B|)$
  - **Jaccard similarity** $J(A, B) = |A \cap B| / |A \cup B|$

- **Vector-based measures**: Compare vectors element-wise, meaning that the position of each element is important. Euclidean distance is an example of such measures.
  - **Euclidean distance** $d(u, v) = \sqrt{((u_1 - v_1)^2 + (u_2 - v_2)^2 + \ldots + (u_n - v_n)^2)}$
  - **Cosine similarity**