1.022 Introduction to Network Models

Amir Ajorlou

Laboratory for Information and Decision Systems
Institute for Data, Systems, and Society
Massachusetts Institute of Technology

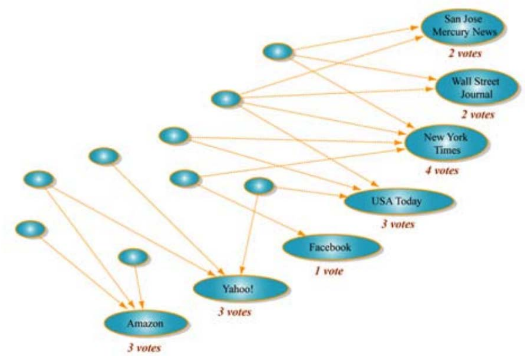
Lecture 5

Centrality and web search

- When you go to Google and type MIT
  - $\Rightarrow$ First result is the home page of MIT
  - $\Rightarrow$ How does Google know that this was the best answer?
- Problem of information retrieval
  - $\Rightarrow$ Search data repositories in response to keyword queries
  - $\Rightarrow$ Classical approach has been textual analysis $\Rightarrow$ No link structure
- Links that point to the webpage $\Rightarrow$ authority of a page on the topic
  - $\Rightarrow$ First, collect a large sample of pages relevant to a topic
  - $\Rightarrow$ Then, look at the number of in-links (score) from these pages

Link structure

- For a query like "newspaper", the most important page is less obvious



Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. *Mining of Massive Datasets.* Cambridge University Press, 2019.

- Mix of newspapers and other pages that always receive links

Page Rank

- Multi-billion query-independent idea of Google
- Each node (or page) is important if it is cited by other important pages
- Each node j has a centrality value (PageRank value) $w(j)$
  - ⇒ Function of the centralities of his (incoming) neighbors
  - ⇒ Similar to eigenvector centrality

$$w(j) = \sum_i \frac{w(i)}{d_{out}(i)} A_{ij}$$

- Dividing by degree dilutes the importance of pages linking to many nodes
- In matrix notation $\mathbf{w}^T = \mathbf{w}^T \mathbf{P}$ where $P_{ij} = A_{ij}/d_{out}(i)$
  - ⇒ Note that $\sum_j Pij = 1$

Page Rank

- This defines a random walk on the nodes of the network
  - ⇒ Walker starts from a node chosen uniformly at random
  - ⇒ Walker moves out choosing uniformly among the out-links
- PageRank is the limiting probability of the random walk
  - ⇒ But, dangling ends may cause the walk to get trapped
- We allow random walk to teleport with probability $1 - s$

$$\mathbf{w}_{k+1}^T = s\mathbf{w}_{k+1}^T\mathbf{P} + \frac{1-s}{n}\mathbf{1}$$

- This is a simpler version of PageRank ⇒ More tricks in practice

- Vector v is an **eigenvector** with **eigenvalue** λ if

$$Mv = \lambda v.$$

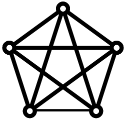
For any symmetric real $n \times n$ matrix M:

- If v and w are eigenvectors with distinct eigenvalues, then v and w are orthogonal.
- If v and w are eigenvectors corresponding to the same eigenvalue, then for any scalars a and b, $av + bw$ is an eigenvector with the same eigenvalue as v and w.
- M has a full orthonormal basis of eigenvectors $v_1, v_2, \ldots, v_n$. All eigenvalues and eigenvectors are real.
- M is diagonalizable. That is,

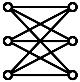
$$M = V \Lambda V^T,$$

  - V: matrix with n orthonormal eigenvectors as columns
  - Λ: diagonal matrix with eigenvalues on diagonal
  - $M = \sum_i \lambda_i v_i v_i^T$ (note that $VV^T = I$).

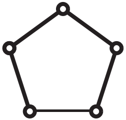
- Complete graph K_5:
  $\lambda(A) = \{4, -1, -1, -1, -1\}$



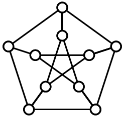
- Bipartite graph $K3, 3$:
  $\lambda(A) = \{3, 0, 0, 0, 0, -3\}$



- Ring graph P_5:
  $\lambda(A) =$
  $\{2, \frac{-1+\sqrt(5)}{2}, \frac{-1+\sqrt(5)}{2}, \frac{-1-\sqrt(5)}{2}, \frac{-1-\sqrt(5)}{2}\}$

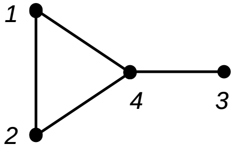


- Peterson graph:
  $\lambda(A) = \{3, 1, 1, 1, 1, 1, -2, -2, -2, -2\}$



- Vertex degrees often stored in the diagonal matrix $\mathbf{D}$, where $D_{ii} = d_i$

$$\mathbf{D} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$



- The $|V| \times |V|$ symmetric matrix $\mathbf{L} := \mathbf{D} - \mathbf{A}$ is called graph Laplacian

$$L_{ij} = \begin{cases} d_i, & \text{if } i = j \\ -1, & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases}, \quad \mathbf{L} = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}$$

- Variants of the Laplacian exist, with slightly different interpretations
  - $\Rightarrow$ Normalized Laplacian $\mathbf{L}_n = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$
  - $\Rightarrow$ Random-walk Laplacian $\mathbf{L}_{rw} = \mathbf{D}^{-1}\mathbf{L}$

Laplacian matrix properties

- Smoothness: For any vector $\mathbf{x} \in \mathbb{R}^{|V|}$ of "vertex values", one has

$$\mathbf{x}^\top \mathbf{L}\mathbf{x} = \sum_{(i,j)\in E} (x_i - x_j)^2$$

  which can be minimized to enforce smoothness of functions on G

- Incidence relation: $\mathbf{L} = \mathbf{B}\mathbf{B}^\top$ where $\mathbf{B}$ has arbitrary orientation
- Positive semi-definiteness: Follows since $\mathbf{x}^\top \mathbf{L}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^{|V|}$
- Rank deficiency: Since $\mathbf{L}\mathbf{1} = \mathbf{0}$, $\mathbf{L}$ is rank deficient