# Measuring Graph Centrality - PageRank

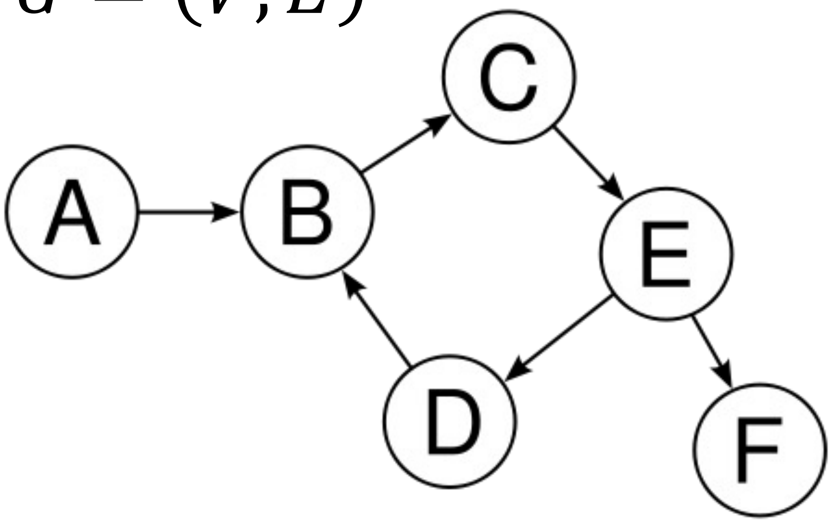
**PPT by Brandon Fain**

# Outline

- Measuring Graph Centrality: Motivation
- Random Walks, Markov Chains, and Stationarity Distributions
- Google’s PageRank Algorithm

Directed Graph

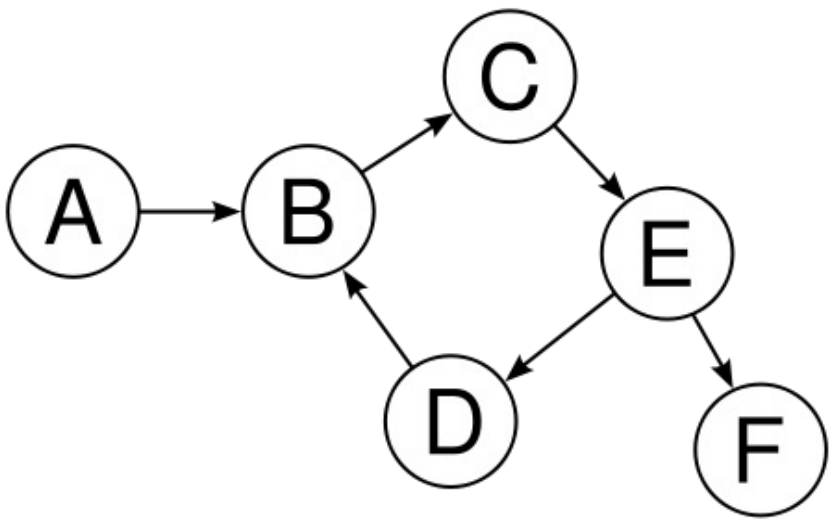
$G = (V, E)$




Adjacency Matrix

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| **A** | 0 | 1 | 0 | 0 | 0 | 0 |
| **B** | 0 | 0 | 1 | 0 | 0 | 0 |
| **C** | 0 | 0 | 0 | 0 | 1 | 0 |
| **D** | 0 | 1 | 0 | 0 | 0 | 0 |
| **E** | 0 | 0 | 0 | 1 | 0 | 1 |
| **F** | 0 | 0 | 0 | 0 | 0 | 0 |

# Graph Centrality




- Which vertex is “the most important” in this graph?
- What do we even mean by important?
- In this class, we will focus on importance as *centrality* as measured by a random walk.

# Motivation – Social Media

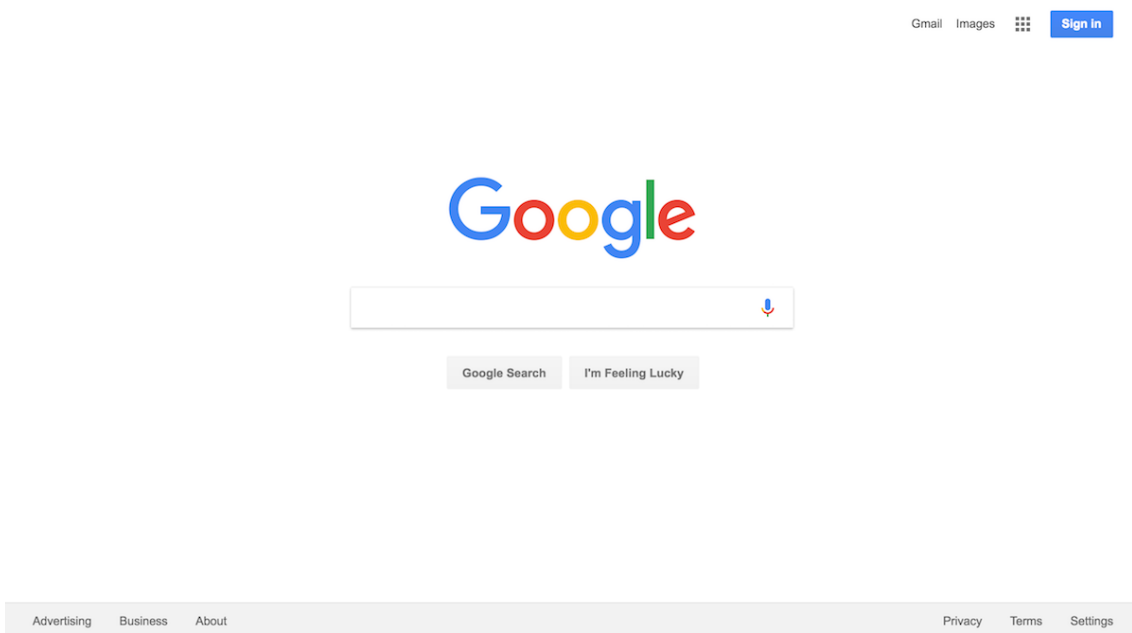
- Who is “important” in the Twitter network?

# Motivation – Academic Publishing

- How impactful is a scientific publication?

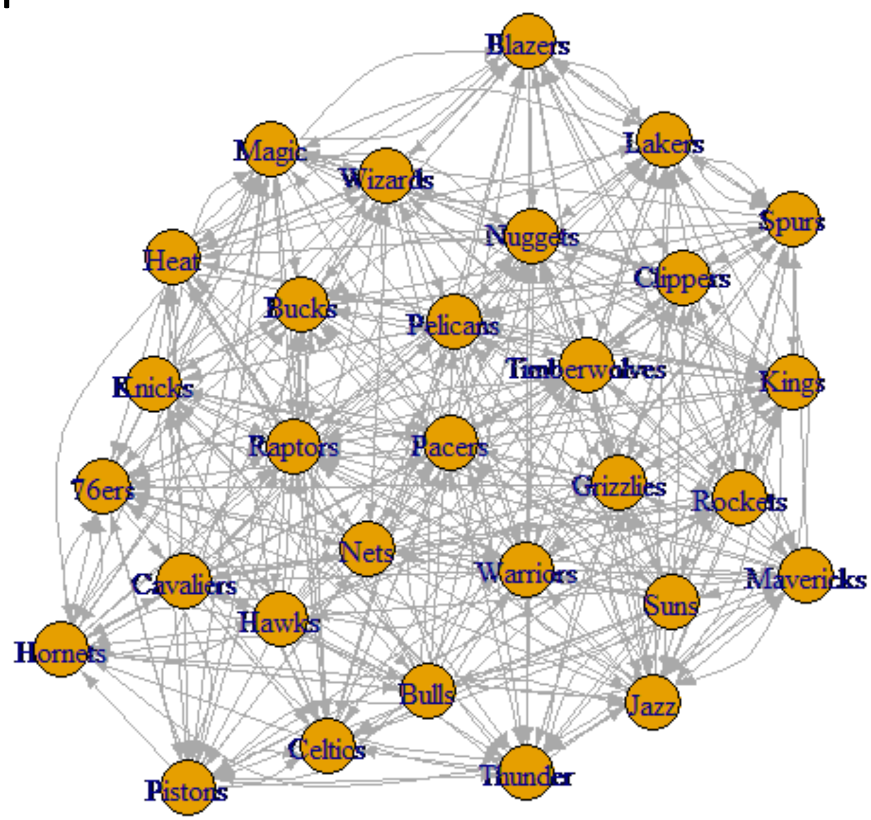
# Motivation – Web Search

- Which webpages are most important for displaying after a search query? (The original motivation).



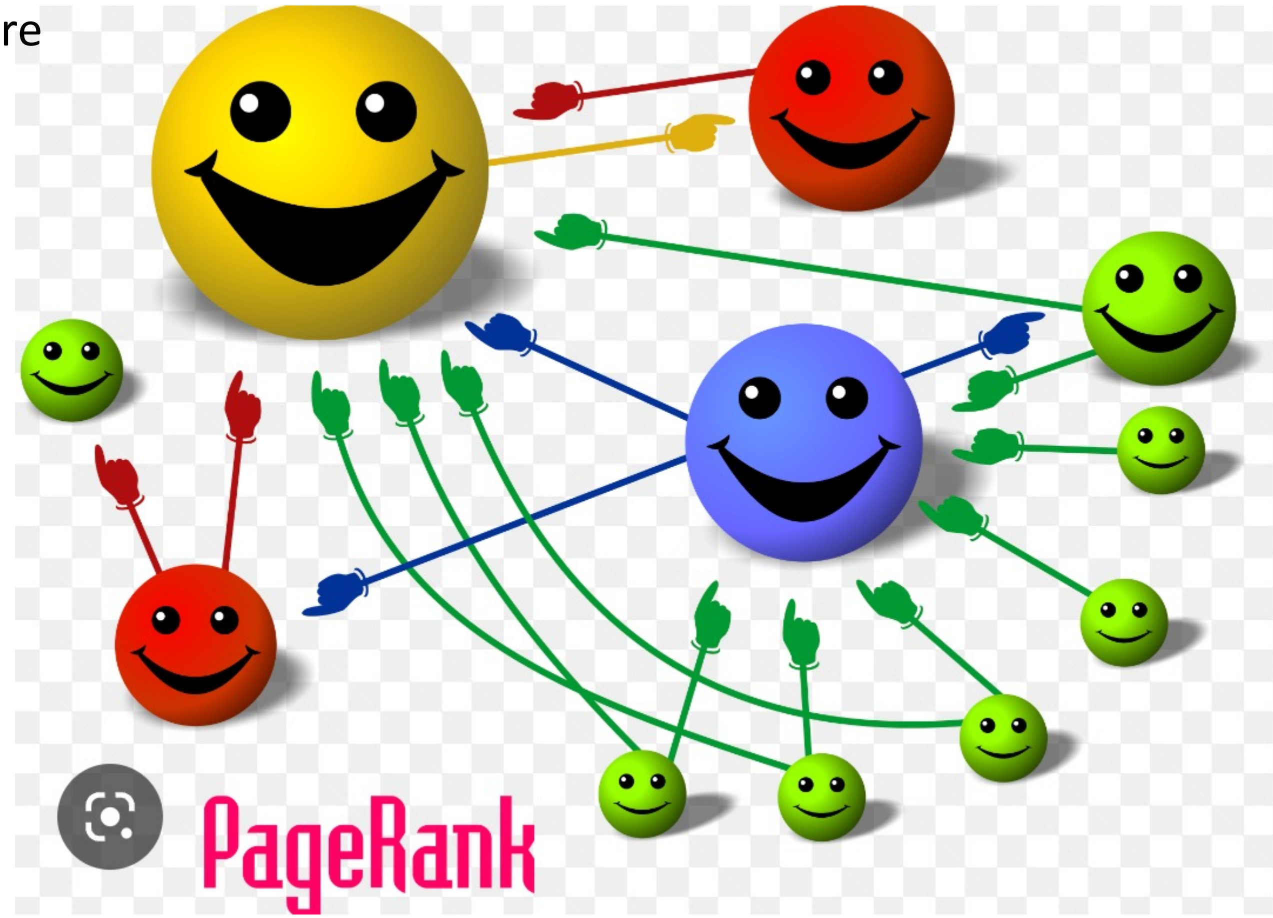
# PageRank Example

Which are more important?



# PageRank Example

Which are more popular on Facebook?



# Outline

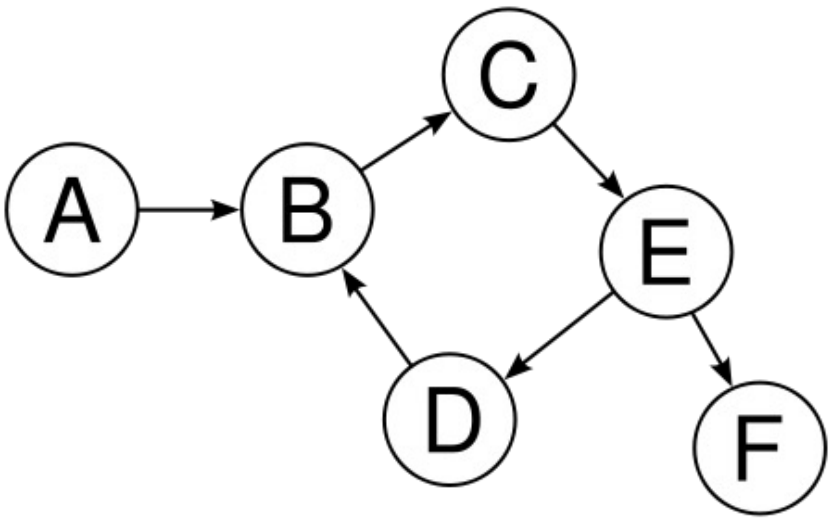
- ~~Measuring Graph Centrality: Motivation~~

- Random Walks, Markov Chains, and Stationarity Distributions

- Google's PageRank Algorithm

# Formalizing "Graph Centrality"

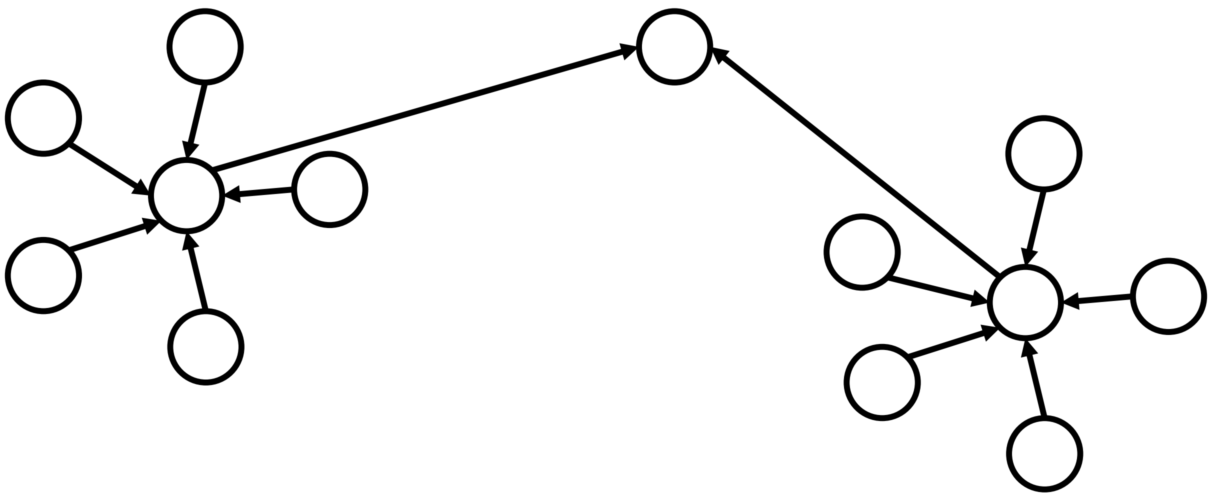
- **Attempt 1.** Measure the *in-degree* (number of incoming directed edges) of every vertex, and choose vertices with highest in-degree.




| Node | in-degree |
|---|---|
| A | 0 |
| B | 2 |
| C | 1 |
| D | 1 |
| E | 1 |
| F | 1 |

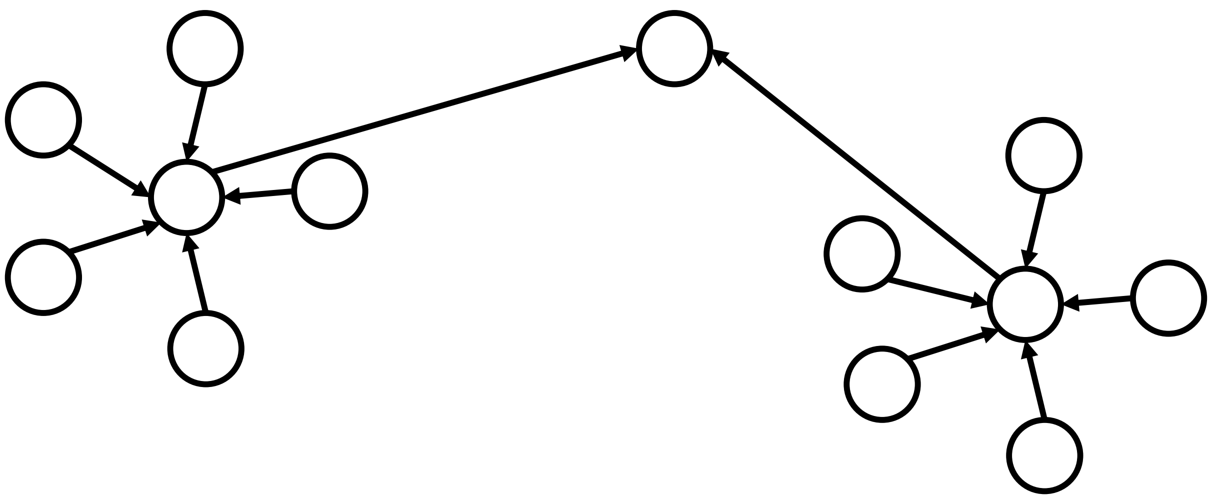
# Formalizing Graph Centrality

- **Problem.** Why do edges from unimportant and important nodes contribute equally?
- What is the most important and central vertex in this graph?



# Formalizing Graph Centrality

- **Attempt 2. Say that a vertex is "central" if we are likely to arrive at the vertex while traversing the graph.**
- For example, in this graph, *all* traversals end at the same place.



# Random Walk

- **Question.** What do we mean by "likely" in a traversal? Where is the probability coming from?
- **Answer.** We consider a ***random walk*** **on graph** $G = (V, E)$**:**

- Start at a random vertex
- For t from 1 to T steps:
  - **Choose an outgoing edge uniformly at random and follow it**

- Let $\boldsymbol{\pi_i^t}$ be the **probability that we are at node $\boldsymbol{i}$ at time $\boldsymbol{t}$**.
- Then the **centrality of node $\boldsymbol{i}$ is $\boldsymbol{\lim_{t\to\infty} \pi_i^t}$.**

# Transition Probabilities

- **We are doing a random walk on the vertices of G** with equal likelihood of moving to any adjacent vertex.
  - Recall that on each step of the random walk, we **choose an outgoing edge uniformly at random and follow it.**
- **Let $\boldsymbol{d_i}$ be the out-degree of vertex $\boldsymbol{i}$.**
- Then

$$\pi_j^{t+1} = \sum_{i:(i,j)\in E} \frac{\pi_i^t}{d_i}.$$

- Note that $\overrightarrow{\pi^{t+1}}$ only depends on $\overrightarrow{\pi^t}$.
- **Consider graph** $G = (V, E)$ **with n vertices and n x n adjacency matrix A.**
- **The n x n transition matrix $\boldsymbol{P}$** is defined below: (Assume $d_i \geq 1$ for all $i$)

$$P_{ij} = \begin{cases} \frac{1}{d_i}, & A_{ij} = 1 \\ 0, & A_{ij} = 0 \end{cases}$$

Each row of **transition matrix $\boldsymbol{P}$** represents a conditional probability distribution: we can interpret $P_{ij}$ as the probability that we move to $j$ given we are at $i$.

# Markov Chain

- Each row of **transition matrix $\boldsymbol{P}$** represents a conditional probability distribution: we can interpret $P_{ij}$ as the probability that we move to $j$ given we are at $i$.
- **The updates of $\overrightarrow{\boldsymbol{\pi}^{t}}$ to $\overrightarrow{\boldsymbol{\pi}^{t+1}}$ can be expressed a vector multiplication by the transition matrix $\boldsymbol{P}$**:

$$\overrightarrow{\pi^{t+1}} = \overrightarrow{\pi^{t}}\, P$$

- Note that $\overrightarrow{\pi^{t+1}}$ is independent the prior history, conditional on $\overrightarrow{\pi^{t}}$, i.e.,

$$\left(\overrightarrow{\pi^{t+1}}\ \middle|\ \overrightarrow{\pi^{1}}, \overrightarrow{\pi^{2}}, \dots, \overrightarrow{\pi^{t}}\right) = \left(\overrightarrow{\pi^{t+1}}\ \middle|\ \overrightarrow{\pi^{t}}\right).$$

- Thus, **this random walk is a *Markov Chain.***

# Stationary Distribution

- **Recall** $\overrightarrow{\pi^{t+1}} = \overrightarrow{\pi^{t}}\, P$
- $\lim_{t\to\infty} \overrightarrow{\boldsymbol{\pi}^{t}}$**, our measure of graph centrality,** is the ***stationary distribution*** **of the Markov chain.**

**Questions**.

1. **Does the limit even exist?**
2. **Does the limit depend on the starting state** $\overrightarrow{\boldsymbol{\pi}^{1}}$**?**
3. **Can we compute** $\lim_{t\to\infty} \overrightarrow{\boldsymbol{\pi}^{t}}$ **efficiently?**

# Existence and Uniqueness

- The transition matrix $P^T$ is a matrix with rows swapped by the columns of $P$.
- An ***eigenvector* of a matrix** is a vector that when multipled by the matrix gives the same vector.
- Note that if $\lim_{t\to\infty} \overrightarrow{\pi^t}$ exists, then it must be some $\overrightarrow{\pi^*}$ such that
$$\overrightarrow{\pi^*} = \overrightarrow{\pi^*}\, P \;\; so\; P^T\, \overrightarrow{\pi^*} = \overrightarrow{\pi^*}\, .$$
- That is, the stationary distribution $\overrightarrow{\pi^*}$ is an ***eigenvector* of the transposed transition matrix $\boldsymbol{P^T}$**, with eigenvalue 1.

- Is it the only one? We need a theorem from linear algebra. Suppose for a moment that $P$ has all strictly positive values.

# Existence and Uniqueness

- **Perron-Frobenius Theorem** (abbreviated). Let A be a square matrix with real, strictly positive entries. Then the following hold.
  1. The largest eigenvalue (call it $\lambda_1$) of A is unique.
  2. There is a *unique* eigenvector (call it $\overrightarrow{v^*}$) corresponding to $\lambda_1$, all entries of which are positive, and this is the *only* eigenvector with all positive entries.
  3. The power iteration method that repeatedly applies $\overrightarrow{v^{t+1}} = A\overrightarrow{v^t}$ beginning from an initial vector $\overrightarrow{v^1}$ not orthogonal to $\overrightarrow{v^*}$ converges to $\overrightarrow{v^*}$ as $t \to \infty$.

- Every row of $P$ is a probability distribution, so $P\,\vec{1} = \vec{1}$.
- By conditions 2 and 1, it must be that the largest eigenvalue of $P$ is 1.
- Since $P$ is square, $P$ and $P^T$ have the same eigenvalues, so 1 is the largest eigenvalue of $P^T$ too!

# Existence and Uniqueness

- Assume matrix $P$ has all positive entries.
- Since 1 is the largest eigenvalue of $P^T$, the theorem implies that:

$\overrightarrow{\boldsymbol{\pi}^*}$ **exists and is the *unique* eigenvector of** $\boldsymbol{P^T}$.

- So we have answered questions 1 and 2: **the stationary distribution exists, and it is unique.**

- **What about computation?** The theorem tells us that the power iteration method converges in the limit…but how long does that take?

# Computation

- **Recall $\overrightarrow{\boldsymbol{\pi}^{t+1}} = \overrightarrow{\boldsymbol{\pi}^{t}}\, \boldsymbol{P}$ and this process limits to $\overrightarrow{\boldsymbol{\pi}^{*}}$ is the *unique* eigenvector of $\boldsymbol{P}^{\boldsymbol{T}}$**.
- The **spectral gap is $\boldsymbol{\lambda_1} - \boldsymbol{\lambda_2}$ where**
  - Let $\lambda_1 = 1$ is the largest eigenvalue of $P^T$, and
  - let $\lambda_2$ be the second largest eigenvalue of $P^T$.

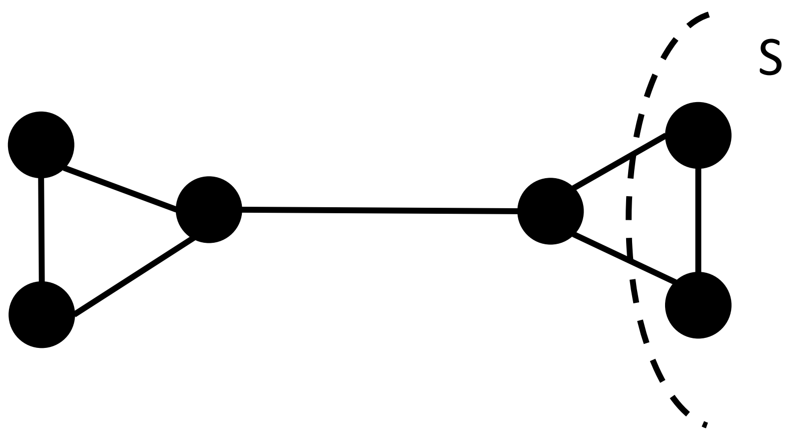
In general, **the convergence rate is determined by the *spectral gap*.**

- The spectral gap is in turn related to the ***conductance* of the underlying graph**:

- Let $S \subseteq V$ be a cut in graph $G = (V, E)$ that disconnects vertices of $V$ - $S$ from $S$.
- **The *conductance* of the cut is $\boldsymbol{\phi}(\boldsymbol{S}) = \frac{|\{(\boldsymbol{i},\boldsymbol{j}) \in \boldsymbol{E} : \boldsymbol{i} \in \boldsymbol{S}, \boldsymbol{j} \notin \boldsymbol{S}\}|}{\min(\sum_{\boldsymbol{i} \notin \boldsymbol{S}} \boldsymbol{d_i},\, \sum_{\boldsymbol{i} \in \boldsymbol{S}} \boldsymbol{d_i})}$ .**

# Computation

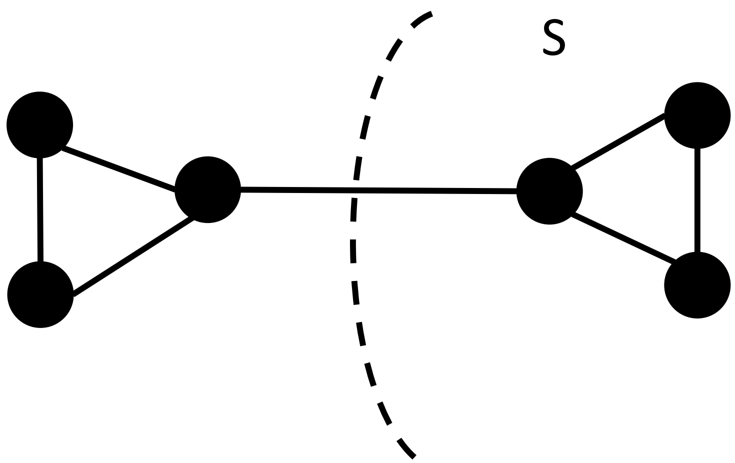
- The conductance of a graph is the minimum conductance of any cut.
- Example: A cut separating the 4 vertices of $V$ - $S$ (on left) from 2 vertices of $S$ (on right), with
  - $|\{(i,j) \in E: i \in S, j \notin S\}|$ =2,
  - $\sum_{i \notin S} d_i$ =10 and
  - $\sum_{i \in S} d_i$ =4.




$$\phi(S) = \frac{2}{\min(10,4)} = \frac{1}{2}$$

# Computation

- Example 2: A cut separating the 3 vertices of $V$ - $S$ (on left) from 3 vertices of $S$ (on right), with
  - $|\{(i,j) \in E\colon i \in S, j \notin S\}|$ =1,
  - $\sum_{i \notin S} d_i$ =7 and
  - $\sum_{i \in S} d_i$ =7.




$$\phi(S) = \frac{1}{\min(7,7)} = \frac{1}{7}$$

# Computation

- **Recall $\overrightarrow{\boldsymbol{\pi^{t+1}}} = \overrightarrow{\boldsymbol{\pi^{t}}}\, \boldsymbol{P}$ and this process limits to $\overrightarrow{\boldsymbol{\pi}^{*}}$ is the *unique* eigenvector of $\boldsymbol{P^{T}}$ .**
- **How is conductance related to convergence time of iterations to convergence?**
  - **Intuitively, lower conductance graphs have bottlenecks, and it may take a longer time for the random walk to traverse the cut.**
  - **By contrast, power iteration converges rapidly on graphs with high conductance (e.g., complete graphs).**
- **To converge (to within some constant error term), one needs $\boldsymbol{O}\left(\frac{\boldsymbol{log(n)}}{\boldsymbol{\phi^{2}}}\right)$ iterations.** What does that look like in practice?

# Outline

- ~~Measuring Graph Centrality: Motivation~~
- ~~Random Walks, Markov Chains, and Stationarity Distributions~~
- Google's PageRank Algorithm

# PageRank

- **Page rank is named after Larry Page.**

- He was doing a PhD at Stanford when he started working on the project of building a search engine.

- He didn't finish his PhD, but he is currently the Alphabet CEO and worth around 83 billion USD.
- This is largely due to his PageRank algorithm developed as a graduate student at Stanford.

# PageRank

- PageRank treats the web as a huge graph, where webpages are vertices, and hyperlinks are directed edges.
- The PageRank algorithm simply applies the **power iteration method** to **compute the stationary distribution of a random walk on the web**.
- Recall that we needed *all* entries in $P$ to be strictly positive to be guaranteed that this works.
- That means that from any vertex, there has to be nonzero probability of transitioning to *any* other vertex.

# PageRank

- To satisfy this, PageRank assumes a slightly different random walk than we described. In particular:

- **Start at a random vertex**
- **For t from 1 to T steps:**
  - **If current page has no links**
    - **Choose a page uniformly at random.**
  - **Else**
    - **With probability 0.15, choose a page uniformly at random.**
    - **With the remaining probability, choose a link from the current page uniformly at random and follow it.**

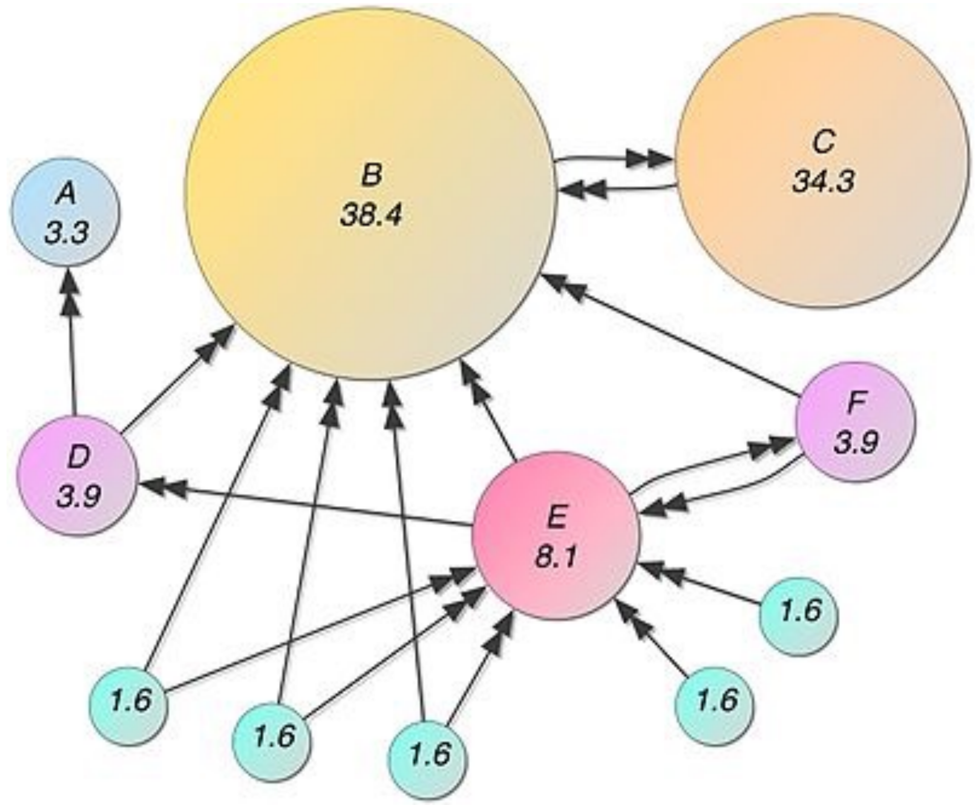
# PageRank

- Thus, if there are n web pages in total, the transition matrix for this random walk is given by

$$P_{ij} = \begin{cases} \dfrac{0.85A_{ij}}{d_i} + \dfrac{0.15}{n}, & i \text{ has links} \\ \dfrac{1}{n}, & i \text{ has no links} \end{cases}$$

- Then we just compute the stationary distribution by the power iteration method.
- What kind results does this generate?

# Resulting PageRank



# PageRank

- Note that our modification also ensures that the conductance of the graph is not too small. In practice, 50 to 100 power iterations suffice for a reasonable approximation to the stationary distribution.
- This might seem hard for large n, but note that the graph itself is extremely sparse, so matrix – vector multiplication can be implemented efficiently.
- All other things equal, google search prefers to show results with higher PageRank.
- The #1 thing that increases your PageRank?
  - Having other important pages link to you.
  - For example, develop an **important algorithm**, like Larry Page's PageRank and found a startup company based on your algorithm.