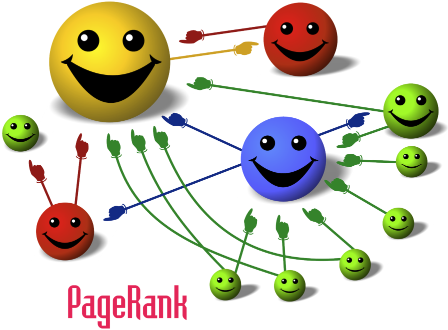
4. Spectral methods

- Linear algebra review
- Markov chain
- Perron-Frobenius theorem
- Random walk on graphs
- PageRank axioms
- Graph Laplacian matrix

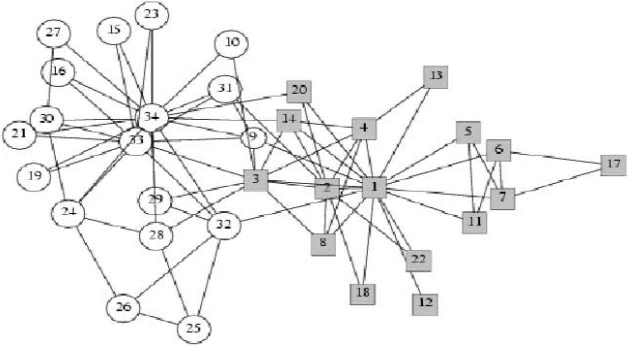
Spectral methods for network problems

- Motivating example:
- PageRank by GOOGLE



- Problem: given a search query, rank web pages according to how relevant they are
- Idea: random walk on graphs

- Motivating example:
- Spectral Graph Partitioning



- Problem: given a graph of interactions, cluster the nodes as to group connected components together
- Idea: minimize conductance $\frac{c(A,B)}{\min\{e(A),e(B)\}}$
  where $e(A) = \sum_{i \in A} \sum_{j \in V} e_{ij}$

- Motivating example:
- Spectral Clustering [Tamuz et al. 2011]

- Problem: cluster N items in high-dimensional spaces
- Idea: use pair-wise similarity graph

Linear algebra review

- Vector space $\mathbb{R}^n$
  - closed under addition:
    for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, $x + y \in \mathbb{R}^n$
  - closed under scalar multiplication:
    for all $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$, $cx \in \mathbb{R}^n$
- Inner product

$$\langle u, v \rangle \triangleq \sum_i u_i v_i = u^T v$$

- Euclidean norm

$$\|u\| \triangleq \sqrt{\sum_i u_i^2}$$

- Cauchy-Schwarz inequality

$$|\langle u, v \rangle| \leq \|u\| \, \|v\|$$

$$\cos\theta = \frac{\langle u, v \rangle}{\|u\| \, \|v\|}$$

- **Subspace** is a subset of a vector space which is itself a vector space
- Matrix $A \in \mathbb{R}^{n \times m}$
- **Range** of a matrix is a subspace defined as

$$\{Au \mid u \in \mathbb{R}^m\}$$

  It is a subspace spanned by columns of A
- **Rank** of a matrix A is the dimension of the range of A
- a set of vectors $\{v_1, \ldots, v_k\}$ is **independent** if and only if any v_i cannot be represented as a linear combination of other vectors, i.e.

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k \Rightarrow a_1 = a_2 = \cdots = a_k = 0$$

- a set of vectors $\{v_1, \ldots, v_k\}$ is a basis for a vector space V if and only if
  - ★ $\{v_1, \ldots, v_k\}$ span V, i.e. $v = \mathrm{span}(\{v_1, \ldots, v_k\})$; and
  - ★ $\{v_1, \ldots, v_k\}$ is independent

  for any vector space, the number of vectors in the basis is the same as the rank the **nullspace** of a matrix $A \in \mathbb{R}^{n \times m}$ is defined as

$$\text{null}(A) = \{x \in \mathbb{R}^m \mid Ax = 0\}$$

which is the set of vectors orthogonal to all rows in A

fact 1. $\text{rank}(A) = \text{rank}(A^T)$

fact 2. $\text{rank}(A)$ is the maximum number of independent columns (or rows) of A

fact 3. $\text{rank}(A) \leq \min(m, n)$

fact 4. $\text{rank}(A) + \dim(\text{null}(A)) = m$

interpretation: consider $y = Ax$ where we apply matrix A to an input vector x to get an output vector y

- m is the degrees of freedom in x
- $\dim(\text{null}(A))$ is the number of degrees of freedom crushed to zero by applying A
- $\text{rank}(A)$ is the number of degrees of freedom in the output y

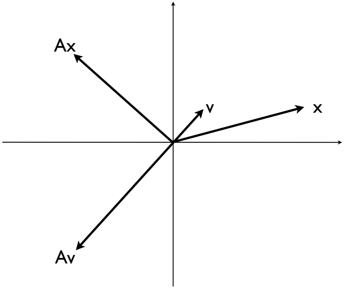
fact 5. $\text{rank}(B) - \dim(\text{null}(A)) \ \leq \ \text{rank}(AB) \ \leq \ \min(\text{rank}(A), \text{rank}(B))$

a matrix $A \in \mathbb{R}^{m \times n}$ is **full rank** if and only if $\mathrm{rank}(A) = \min(m, n)$

- ⋆ even if A and B are full rank, AB might not be full rank (e.g. low-rank factorization)
- ⋆ even if AB is full rank, either one of A and B might not be full rank
- ⋆ if A and B have empty null spaces, then AB has an empty null space

- ⋆ give a non-zero matrix A such that $A^2 = 0$ is a all-zeros matrix
- ⋆ if $\angle(Ax, x) = 0$ for all $x \in \mathbb{R}^n$, i.e. all vectors are eigenvectors, then what can we say about A?

- Eigenvectors and eigenvalues
  - $\lambda \in \mathbb{C}$ is an **eigenvalue** of $A \in \mathbb{R}^{n \times n}$ if
$$Av = \lambda v$$
    and any such v is called an **eigenvector** of A.
  - If v is an eigenvector of A, then so is av.
  - Even when A is real, eigenvalue λ and eigenvector v can be complex
  - Rank of A is the number of non-zero eigenvalues
- Scaling interpretation (assume $\lambda \in \mathbb{R}$ for now)
  - if v is an eigenvector, it is scaled by λ: $Av = \lambda v$.
  - if $x = c_1 v_1 + c_2 v_2$, then $Ax = c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2$.



$$\begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = A \begin{bmatrix} 1 \\ 1 \end{bmatrix} + A \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

$$= 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 8 \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

- scaling interpretation
  - $\lambda > 0$: Av point in the same direction as v
  - $\lambda < 0$: Av point in opposite direction as v
  - $|\lambda| < 1$: Av smaller than v
  - $|\lambda| > 1$ Av larger than v

- eigenvectors are not unique when multiple eigenvalues have same value

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

- symmetric matrices have real eigenvalues and eigenvectors

$$\lambda v = Av$$

$$\lambda^* v^* = v^* A^T$$

$$\lambda^* v^* v = v^* A^T v$$

$$\text{but we also have } \lambda v^* v = v^* A^T v$$

since $v^* v$ is a real number, we can conclude that $\lambda = \lambda^*$

- it is not immediately clear why eigenvalues and eigenvectors play important role in discrete mathematics; eigenvalues have many equivalent characterizations, and perhaps these equivalent representations shine a light on why they are significant the **Rayleigh quotient** of a non-zero vector x with respect to a matrix A is defined as the ratio

$$\frac{x^T A\, x}{x^T x}$$

theorem. let v be the one that maximizes the Rayleigh quotient of a symmetric matrix A. Then v is an eigenvector with the eigenvalue equal to the Rayleigh quotient, and this eigenvalue is the largest eigenvalue of A.

proof. we solve the unconstrained maximization by setting the gradient to zero

$$\frac{\partial \frac{x^T A x}{x^T x}}{\partial x} \quad = \quad \frac{-(x^T A x) 2x + (x^T x) 2 A x}{(x^T x)^2} \quad = \quad 0$$

this gives

$$Ax \quad = \quad \left(\frac{x^T A x}{x^T x}\right) x$$

which implies that the maximizer is a eigenvector, and that the Rayleigh quotient is equal to the largest eigenvalue

Courant-Fischer theorem. let A be a symmetric matrix with eigenvalue $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then

$$\begin{aligned}\lambda_k &= \min_{H \in \mathbb{R}^n, \dim(H)=n-k+1} \max_{x \in H} \frac{x^T A x}{x^T x} \\ &= \max_{H \in \mathbb{R}^n, \dim(H)=k} \min_{x \in H} \frac{x^T A x}{x^T x}\end{aligned}$$

proof. here we only prove the second equation.

- ⋆ A symmetric matrix has a orthogonal and normalized eigenvectors $\{v_1, \ldots, v_n\}$. Any subspace H that has dimension k has a non-empty intersection with the subspace spanned by $\{v_k, \ldots, v_n\}$. Let x be a vector in this subspace such that $x = \sum_{i=k}^n \alpha_i v_i$ for some scalars α_i's. Then,

$$\frac{x^T A x}{x^T x} = \frac{\sum_{i=k}^n \alpha_i^2 \lambda_i}{\sum_{i=k}^n \alpha_i^2} \leq \lambda_k$$

it follows that $\min_{x \in H} \frac{x^T A x}{x^T x} \leq \lambda_k$ for any k dimensional H and in particular $\min_{x \in H} \frac{x^T A x}{x^T x} \leq \lambda_k$. And we know this can be achieved with equality by choosing $H = \mathrm{span}(\{v_1, \ldots, v_k\})$. This proves that $\lambda_k = \max_{H \in \mathcal{H}_k} \min_{x \in H} \frac{x^T A x}{x^T x}$

Application in natural language processing

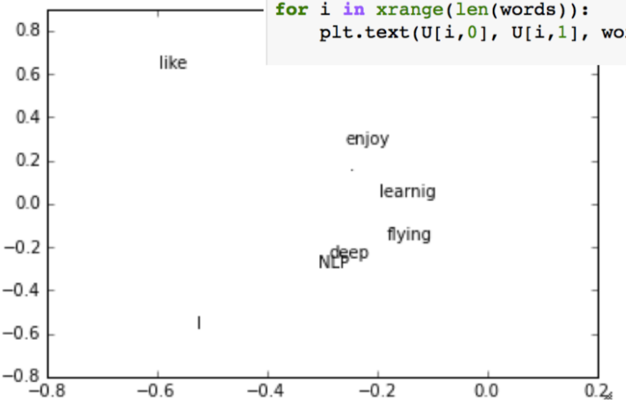
- we say two words co-occur if they are used together in a sentence
- given a dataset (e.g. Wikipedia), we can build a co-occurrence matrix of all the words in the vocabulary
- consider a corpus: { I like deep learning. I like NLP. I enjoy flying. }
- my vocabulary is {I, like, enjoy, deep, learning, NLP, flying }

$$C = \begin{bmatrix} 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Application in natural language processing

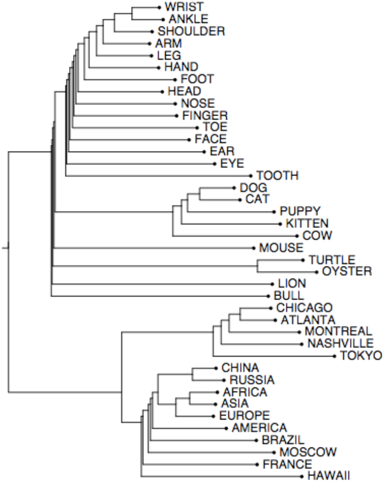
- consider a corpus: { I like deep learning. I like NLP. I enjoy flying. }

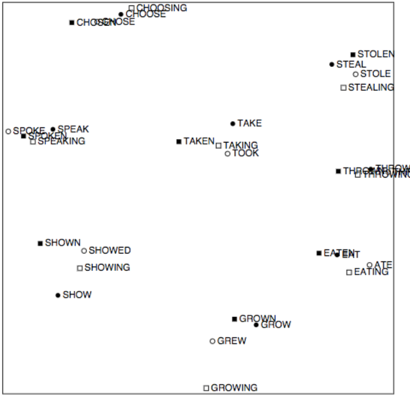
```
for i in xrange(len(words)):
    plt.text(U[i,0], U[i,1], words[i])
```



Application in natural language processing

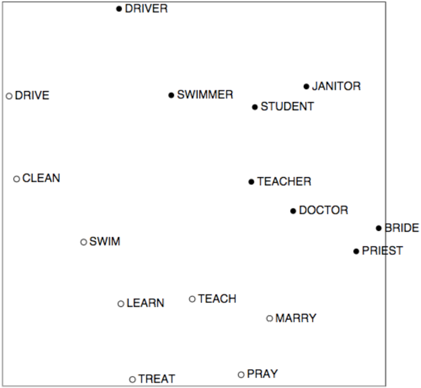
- "An improved model of semantic similarity based on lexical co-occurrence", Rhode et al. 2005





Application in natural language processing

- "An improved model of semantic similarity based on lexical co-occurrence", Rhode et al. 2005



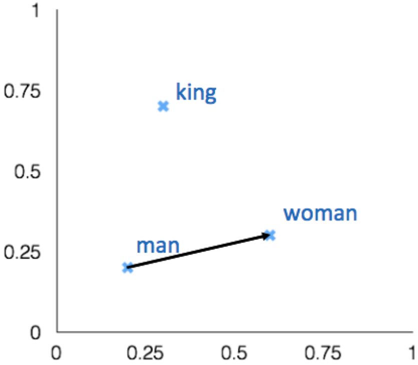
Application in natural language processing

- word vector analogies task

$$a : b :: c : ?$$

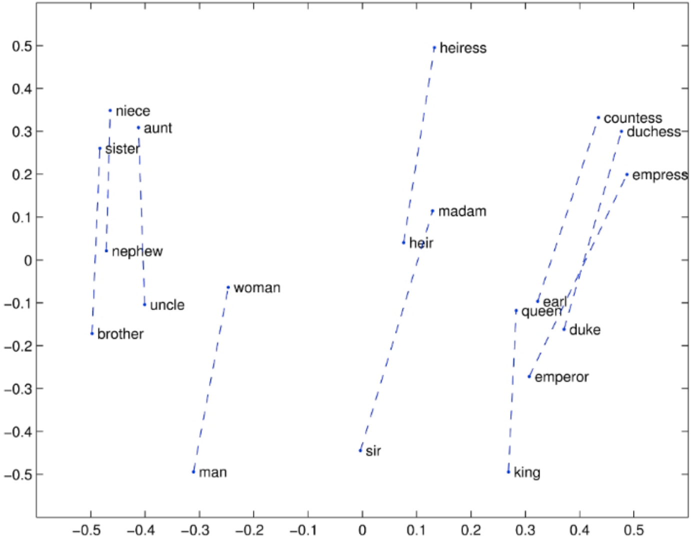
- man : woman :: king : ?
- solution

$$d = \arg\max_i (x_b - x_a + x_c)^T x_i$$



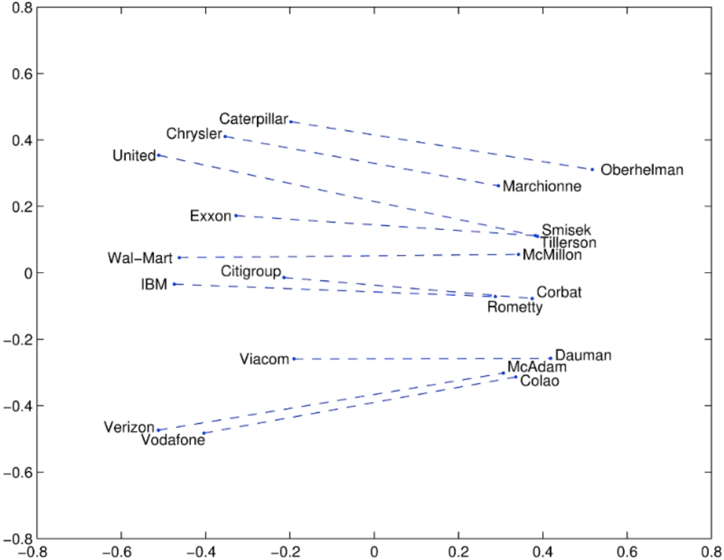
Application in natural language processing

- man : woman



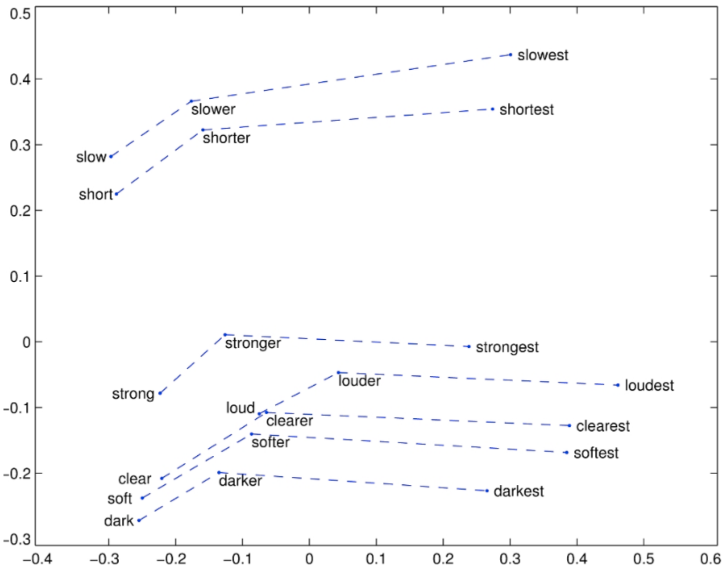
Application in natural language processing

- copmany : ceo



Application in natural language processing

- superlatives



Application in natural language processing

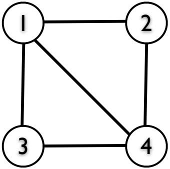
Expression	*Nearest token*
Paris - France + Italy	Rome
bigger - big + cold	colder
sushi - Japan + Germany	bratwurst
Cu - copper + gold	Au
Windows - Microsoft + Google	Android
Montreal Canadiens - Montreal + Toronto	Toronto Maple Leafs

Markov chain (discrete time, finite state, homogeneous)

- discrete time t=1,2,...
- n states
- random process X_t takes one of n states
- **Markov chain**: X_t is conditionally independent of the past give X_{t-1}
- **transition probability** $P_{ji} = \mathbb{P}(X_{t+1} = j | X_t = i)$

$$p(t+1) \;=\; P\, p(t)$$

- example: random walk on a graph

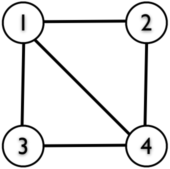


$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 1/3 \\ 1/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 1/3 \\ 1/3 & 1/2 & 1/2 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \mathbb{P}(X_t = 1) \\ \mathbb{P}(X_t = 2) \\ \mathbb{P}(X_t = 3) \\ \mathbb{P}(X_t = 4) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} 1/6+1/6+1/9 \\ 1/9 \\ 1/9 \\ 1/6+1/6 \end{bmatrix}, \ldots$$

Stationary distribution

- conditional probability sums to one: $\sum_j P_{ji} = 1$
- rewrite as $[1\ 1\ \ldots\ 1]P = [1\ 1\ \ldots\ 1]$
  $[1\ 1\ 1\ \ldots\ 1]$ is a left eigenvector of P with eigenvalue 1.
- there is a corresponding eigenvector with eigenvalue 1. Let's call it $p = [p_1\ p_2\ \ldots p_n]^T$.
$$p = Pp\ .$$
- **interpretation.** this eigenvector is called the **stationary distribution** of a Markov chain P.
  - If $p(0) = p$, then $p(t) = P^t p = P^{t-1}(Pp) = p^{t-1} p = p$ for all t.
  - in the limit as time gors to infinity, $\lim_{t\to\infty} p(t) = \lim_{t\to\infty} P^t p(0) = p$



$$P = \begin{bmatrix} 0 & 1/d_2 & 1/d_3 & 1/d_4 \\ 1/d_1 & 0 & 0 & 1/d_4 \\ 1/d_1 & 0 & 0 & 1/d_4 \\ 1/d_1 & d_2/d_2 & 1/d_3 & 0 \end{bmatrix}$$

example: random walk on an undirected graph

- $P_{ji} = 1/d_i$ if $(i,j) \in E$
- sanity check
  - ⋆ is $[1\ 1\ \cdots\ 1]P = [1\ 1\ \cdots\ 1]$?
  - ⋆ what is the right eigenvector?

- the stationary distribution unique if and only if
  - ⋆ graph is connected and
  - ⋆ graph is aperiodic

  proof uses Perron-Frobenius theorem, and we will prove it formally later in this note

Perron-Frobenius theorem

- We say a matrix or a vector is
  - **positive** if all its entries are positive
  - **nonnegative** if all its entries are nonnegative
- we use notation $x > y$ $(x \geq y)$ to mean $x - y$ is entrywise positive (nonnegative)
- Basic facts
  - if $A \geq 0$ and $z \geq 0$, then $Az \geq 0$.
  - conversely, if for all $z \geq 0$, we have $Az \geq 0$, then we can conclude $A \geq 0$.
  - in otherwords, matrix multiplication preserves nonnegativity if and only if the matrix is nonnegative
  - if $A > 0$ and $z \geq 0$, $z \neq 0$, then $Ax > 0$.
  - conversely, if whenever $z \geq 0$, $z \neq 0$, we have $Az > 0$, then we can conclude $A > 0$.

- Regular nonnegative matrices
  - suppose $A \in \mathbb{R}^{n\times n}$, with $A \geq 0$.
  - A is called **regular** if for some $k \geq 1$, $A^k > 0$
  - meaning in the example of random walk on graphs
    - there is an edge from i to j whenever $A_{ij} > 0$
    - then $(A^k)_{ij} > 0$ if and only if there is a path of length k from i to j
    - A is regular if for some k there is a path of length k from every node to every other node
- examples:
  - $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ are not regular.
  - $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ is regular.

Perron-Frobenius theorem

- Assume A is nonnegative and $A^k > 0$ for some k, then

1. there is an eigenvalue λ_{pf} of A that is real and positive, with positive left and right eigenvectors
2. for any other eigenvalue λ, we have $|\lambda| < \lambda_{pf}$
3. the eigenvalue λ_{pf} has multiplicity one
4. no other eigenvector has all positive (moreover non-negative) entries: they contain at least one negative or non real-valued entry
5. $\lim_{k\to\infty} \frac{A^k}{\lambda_{pf}^k} = \frac{1}{v^T w} v w^T$ where v and w are the left and right eigenvectors corresponding to λ_{pf}

- the eigenvalue λ_{pf} is called the Perron-Frobenius (PF) eigenvalue of A

- the associated positive (left and right) eigenvectors are called the (left and right) PF eigenvectors (and are unique, up to a scaling)

Perron-Frobenius theorem for Markov chains

- Consider a Markov chain $X_0, X_1, \ldots,$ with states in $\{1, \ldots, n\}$.
- Transition matrix P such that

$$P_{ij} = \mathbb{P}(X_{t+1} = i | X_t = j)$$

- Let p_t be the distribution of X_t, i.e. $(p_t)_i = \mathbb{P}(X_t = i)$, then

$$p_{t+1} = P p_t = P^t p_0$$

- Recall $\mathbb{1}^T P = \mathbb{1}^T$
- So $\mathbb{1}^T$ is a left eigenvector with eigenvalue 1, which in fact is the PF eigenvalue of P

an **aperiodic** and **irreducible** Markov chain has **regular** transition matrix

- ⋆ A Markov chain is aperiodic if return to state i can occur at irregular times, i.e. there exists n such that for all $n' \geq n$,

$$\mathbb{P}(x_{n'} = i | x_0 = i) > 0$$

- ⋆ A Markov chain is irreducible if there is a non-zero probability of transitioning (even if it takes more than one step) from any state to any other state.

- For a Markov chain, the right PF eigenvector is the stationary distribution

$$P\pi = \pi$$

theorem. for an *aperiodic* and *irreducible* Markov chain, there is a unique stationary distribution π that satisfy $\pi > 0$
proof. there exists an integer k such that P^k has strictly positive entries if and only if the Markov chain is aperiodic and irreducible. The stationary distribution of the Markov chain is the unique Perron-Frobenius eigenvector of P^k.

- Further, $\lambda_{pf} = 1 > |\lambda_j|$ imply $p_t \to \pi$ no matter what the initial distribution p_0

- example 1: if the Markov chain has k **disconnected components**, then there are k eigenvalues of the same value $\lambda_1 = \lambda_2 = \ldots = \lambda_k = 1$, and different stationary distribution depending on the initial position

  if each component has transition matrix P_i with stationary distribution π_i, such that

$$P = \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k \end{bmatrix}, \text{ and } \pi_i = P_i \pi_i \text{ for all } i\text{'s}$$

  precisely,
  - ⋆ we know $\mathbb{1}^T P = \mathbb{1}^T$, which implies left eigenvector $\mathbb{1}$ and eigenvalue 1
  - ⋆ therefore, there is a corresponding right eigenvector, we call it $\pi = [\pi_1, \pi_2, \ldots, \pi_k]$
  - ⋆ then it follows that any vector that can be represented as $\pi' = [a_1\pi_1, a_2\pi_2, \ldots, a_k\pi_k]$ for any real a_i's are also eigenvectors with eigenvalue 1, since

$$P\pi' = [a_1 P_1 \pi_1, \cdots, a_k P_k \pi_k] = [a_1\pi_1, \cdots, a_k\pi_k] = \pi'$$

  - ⋆ this implies that there are k linearly independent eigenvectors with eigenvalue 1, since we can choose a_i's to generate such eigenvectors

- example 2: a 2-**periodic** Markov chain has $\lambda_n = -1$, and the stationary distribution does not exist

  the transition matrix has the following structure:

$$P = \begin{bmatrix} 0 & P_1 \\ P_2 & 0 \end{bmatrix}, \text{ and } \pi_1 = P_1\pi_2 \text{ and } \pi_2 = P_2\pi_1$$

  such that $P\pi = \pi$ for $\pi = [\pi_1, \pi_2]$

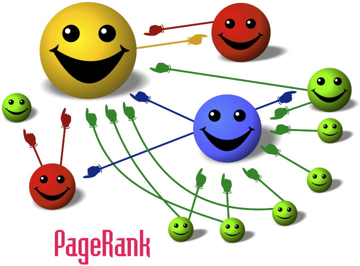
  precisely,

  - ★ we know $\mathbb{1}^T P = \mathbb{1}^T$, which implies left eigenvector $\mathbb{1}$ and eigenvalue 1
  - ★ therefore, there is a corresponding right eigenvector, we call it $\pi = [\pi_1, \pi_2]$
    then $\pi' = [\pi_1, -\pi_2]$ has the eigenvalue $\lambda_n = -1$ since

$$P\pi' = [-P_1\pi_2, P_2\pi_1] = [-\pi_1, \pi_2] = -\pi'$$

  - ★ we found two eigenvectors one with eigenvalue 1 and the other with eigenvalue -1

PageRank by GOOGLE



- Given the directed hyper link graph G and its adjacency matrix A
- Goal: score the pages according to how important it is
- Approach 1: $s(i) = \sum_j A_{ij}$ (paper with many citations is important)
- Problem: one can manipulate the score by creating two pages with lots of links in between
- Solution: $s(i) = \sum_j \frac{1}{d_j} A_{ij} s(j)$
- Interpretation 1: paper cited by important papers is important
- Interpretation 2: random walk on graphs

PageRank

- **PageRank**: the score vector of the nodes in a graph with (directed) adjacency matrix A is the top (left) eigen vector defined as

$$s = AD^{-1}s$$

  where D is the diagonal matrix with out-degree, and A is the (directed) adjacency matrix with $A_{ij} = 1$ if and only if $(i, j) \in E$
- **Irreducible PageRank**

$$s = \left(\alpha AD^{-1} \; + \; (1-\alpha)\frac{1}{n}\mathbb{1}\mathbb{1}^T\right)s$$

  interpretation: at each time, with probability $1 - \alpha$ the random surfer chooses to jump to any random node (for irreducibility)
- **Personalized PageRank** for a node i

$$s_{(i)} = \left(\alpha AD^{-1} \; + \; (1-\alpha)e_i\mathbb{1}^T\right)s_{(i)}$$

  interpretation: at each time, with probability $1 - \alpha$ the random web surfer chooses to jump to any node i so that we learn how important each node is (the score vector $s_{(i)}$) with respect to node i

Personalized PageRank

$$s_{(i)} = \left(\alpha A D^{-1} \ + \ (1-\alpha) e_i \mathbb{1}^T\right) s_{(i)}$$

- $e_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T$
- **claim 1.** for $1 > \alpha > 0$, $s_{(i)}$ is unique
- **proof.**

$$\begin{aligned} s_{(i)} - \alpha A D^{-1} s_{(i)} &= (1-\alpha) e_i \\ (\mathbf{I} - \alpha A D^{-1}) s_{(i)} &= (1-\alpha) e_i \\ s_{(i)} &= (1-\alpha)(\mathbb{I} - \alpha A D^{-1})^{-1} e_i \end{aligned}$$

- $s_{(i)}$ is uniquely determined if the matrx $(\mathbb{I} - \alpha A D^{-1})$ is invertible
- $(\mathbb{I} - \alpha A D^{-1})$ is invertible if all eigen values are non-zero
- $\lambda_i(\mathbb{I} - \alpha A D^{-1}) = 1 - \alpha \lambda_i(A D^{-1})$
- $\lambda_i(\mathbb{I} - \alpha A D^{-1}) \geq 1 - \alpha > 0$, as $|\lambda_i(A D^{-1})| \leq 1$ from Perron-Frobenius theorem

Personalized PageRank

$$s_{(i)} = (1-\alpha)(\mathbb{I} - \alpha A D^{-1})^{-1} e_i$$

- solving this is equivalent to solving a system of linear equations!
- **Spilling Paint** interpretation
  - Consider a paint that is *diffusing* over the graph, but gets stuck at each node with some probability (because it dries)
  - At any given time $t \in \{1, 2, \dots\}$ let $s^t \in \mathbb{R}^n$ denote the amount of paint that is stuck at each node that evolves according to

$$s^{t+1} = s^t + (1-\alpha) r^t$$

  - let $r^t \in \mathbb{R}^n$ denote the amount of wet paint that is remaining at each node that diffuses according to

$$r^{t+1} = \alpha A D^{-1} r^t$$

  - starting from $r^0 = e_i$, we are interested in where the paint gets stuck in the end s^∞

Personalized PageRank

$$\begin{aligned} s^{t+1} &= s^t + (1-\alpha) r^t \\ r^{t+1} &= \alpha A D^{-1} r^t \end{aligned}$$

- We can compute s^∞ as

$$s^\infty = (1-\alpha) \sum_{t \geq 0} r^t = (1-\alpha) \sum_{t \geq 0} (\alpha A D^{-1})^t e_i$$

- useful equality that holds also for matrices (proof using diagonalization): $(\mathbb{I} - M)^{-1} = \mathbb{I} + M + M^2 + M^3 + \cdots$
- so,

$$s^\infty = (1-\alpha)(\mathbb{I} - \alpha A D^{-1})^{-1} e_i$$

- which is exactly the same as personalized PageRank:

$$s_{(i)} = (1-\alpha)(\mathbb{I} - \alpha A D^{-1})^{-1} e_i$$

Personalized PageRank

- There are two ways to compute personalized PageRank $s_{(i)}$: set of linear equations, and paint diffusion process
- the paint diffusion process gives us a nice way to simulate and approximate $s_{(i)}$ as follows:
- ignore time for now, and focus on a specific point where you have $s \in \mathbb{R}^n$ dried paint on each node and $r \in \mathbb{R}^n$ wet paint
- we are interested in the following quantity, for a given s and r

  $$p_{s,r} \;=\; s + (1-\alpha)\sum_{t\geq 0}(\alpha A D^{-1})^t r$$

  for example, $s_{(i)} = p_{0,e_i}$
- we will give a sequential process that produces $s(k)$ and $r(k)$ for $k \geq 1$, that preserves $p_{s(k+1),r(k+1)} = p_{s(k),r(k)}$ but reduces $r(k+1) \leq r(k)$ such that eventually we can take $s(k)$ as our approximation for $p_{s,r}$

Personalized PageRank

- the update rule that preserves $p_{s,r}$

$$\begin{aligned} s(k+1)_u &= s(k)_u + (1-\alpha) r(k)_u \\ r(k+1)_u &= 0 \\ r(k+1)_v &= r(k)_v + \alpha \frac{1}{d_u} r(k)_u \,, \qquad \text{for } v \in N(u) \end{aligned}$$

- **Claim.** $p_{s(k),r(k)} = p_{s(k+1),r(k+1)}$
- Algorithm (**Approximate personalized PageRank**):
  - initialize $s(0) = 0, r(0) = e_i$
  - **while** exists a node with $r(v) > \varepsilon$
  - pick $u = \arg\max_v r(v)$
  - $(s(k+1), r(k+1)) \leftarrow \text{update}(s(k), r(k), u)$

Random walk on a graph

- Undirected Graph $G = (V, E)$
- Markov chain with $n = |V|$ states

$$P_{ij} = \mathbb{P}(X_{t+1} = i | X_t = j) = \begin{cases} 1/d_j & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

where d_j is the degree of node j

- Distribution at time t

$$p_t(i) = \mathbb{P}(X_t = i) = \sum_j \underbrace{\mathbb{P}(X_t = i | X_{t-1} = j)}_{P_{ij}} \underbrace{\mathbb{P}(X_{t-1} = j)}_{p_{t-1}(j)}$$

- Matrix form of $p_t(i) = \sum_j P_{ij} p_{t-1}(j)$

$$p_t = P p_{t-1}$$

- Stationary distribution

$$\pi = P\pi$$

- Unique if the random walk is aperiodic and the graph is connected (Perron-Frobenius Theorem)

- π is the right Perron-Frobenius eigenvector corresponding to $\lambda_1 = 1$ and the left eigenvector is $w = \mathbb{1}$

$$\pi = P\pi \ , \qquad \mathbb{1}^T = \mathbb{1}^T P$$

- **claim.** $\pi_i = d_i / \sum_k d_k$.
- **proof.** we need to show that $d_i / \sum_k d_k = \sum_j P_{ij}(d_j / \sum_k d_k)$ for all i

$$\begin{aligned} \sum_j P_{ij}\pi_j &= \sum_{j:(i,j)\in E} \frac{1}{d_j}\frac{d_j}{\sum_k d_k} \\ &= d_i / \sum_k d_k \ = \ \pi_i \end{aligned}$$

this proves that $\pi = P\pi$ for the choice of $\pi_i = d_i / \sum_k d_k$ and therefore this is a stationary distribution
by Perron-Frobenius theorem, it is unique if the Markov chain is aperiodic and the graph is connected

claim. A symmetric matrix M has eigenvectors that are orthogonal to each other, and can be factorized as

$$M = U \Lambda U^T$$

where Λ is a diagonal matrix with the eigenvalues in the diagonal, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, and $U = [U_1, \ldots, U_n]$ is an orthonormal matrix, where $U_i^T U_j = 0$ and $\| U_i \| = 1$ for all $i \neq j$ such that $UU^T = U^T U = \mathbf{I}$. Further, U_i's are the eigenvectors of M. (we omit the proof here)

claim. $P = AD^{-1}$ is diagonalizable

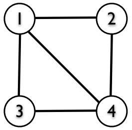
proof. define a symmetric matrix $M \triangleq D^{-1/2}AD^{-1/2}$

$$P = AD^{-1} = D^{1/2}(\underbrace{D^{-1/2}AD^{-1/2}}_{=M})D^{-1/2} = D^{1/2}MD^{-1/2}$$

since M is symmetries, it is diagonalizable with $M = U\Lambda U^T$ where the columns of U are the eigenvectors and Λ is a diagonal matrix with eigenvalues in the diagonals. It follows that

$$P = D^{1/2}U\Lambda U^T D^{-1/2}$$

among other things, this proves that P is always diagonalizable, i.e. can be decomposed into $P = Q\Lambda Q^{-1}$, for a diagonal matrix Λ with eigen values in the diagonals



$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 1/3 \\ 1/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 1/3 \\ 1/3 & 1/2 & 1/2 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 3 & & & \\ & 2 & & \\ & & 2 & \\ & & & 3 \end{bmatrix}$$

claim. $P = AD^{-1}$ is diagonalizable

$$P = D^{1/2} U \Lambda U^T D^{-1/2}$$

corollary. $P = AD^{-1}$ has the same eigenvalues as $M = D^{-1/2} A D^{-1/2}$, and the left eigenvectors are the columns of $D^{1/2} U$ and the right eigenvectors are the rows of $U^{-1} D^{-1/2}$

corollary. Since the first left eigenvector of P is $\mathbb{1}$, we know that the first eigenvector of $D^{-1/2} A D^{-1/2}$ corresponding to eigenvalue one is

$$u_1 = \frac{1}{\|d^{1/2}\|} d^{1/2} = \frac{1}{\sqrt{\sum_{i=1}^{n} d_i}} d^{1/2}$$

for $d^{1/2} \triangleq [\sqrt{d_1}, \ldots, \sqrt{d_n}]$, and $\pi = D^{1/2} u_1$.

in general, reversible Markov chains are diagonalizable

- a Markov chain P with stationary distribution π is **reversible** if and only if it satisfies the following *detailed balance equation*

$$P_{kj}\pi_j = P_{jk}\pi_k$$

- for a reversible Markov chain P with stationary distribution π,

$$N = \Pi^{-1/2} P \Pi^{1/2}$$

is always a symmetric matrix, where $\Pi = \mathrm{diag}(\pi)$

claim. a reversible Markov chain P is diagonalizable
proof. we have for some symmetric matrix N

$$P = \Pi^{1/2} N \Pi^{1/2}$$

since N is symmetric, it can be diagonalized such that $N = U \Lambda U^{-1}$

$$P = \Pi^{1/2} U \Lambda U^{-1} \Pi^{1/2}$$

P has the same eigenvalues as N, and the above factorization gives a eigen value decomposition of P
this implies P is diagonalizable

claim. Let $p^{(t)}$ be the distribution of a (aperiodic and irreducible) random walk $P = AD^{-1}$ with stationary distribution π after t time steps. Then,

$$p^{(t)} \;\; = \;\; \pi + D^{1/2} \sum_{i \geq 2} \lambda_i^t u_i (u_i^T D^{-1/2} p^{(0)})$$

where u_i's are the normalized eigenvectors of $D^{-1/2}AD^{-1/2}$.

proof. since $D^{-1/2}AD^{-1/2}$ is a symmetric matrix, we know from the previous claim that $D^{-1/2}AD^{-1/2} = U\Lambda U^T$, where $U = [u_1, \ldots, u_n]$ is the orthonormal matrix with eigenvector as each column. Since $p^{(t)} = P^t p^{(0)}$, we have

$$\begin{aligned}
p^{(t)} &= P^t p^{(0)} \\
&= (D^{1/2} D^{-1/2} A D^{-1/2} D^{-1/2})^t p^{(0)} \\
&= D^{1/2} U \Lambda^t U^T D^{-1/2} p^{(0)} \\
&= D^{1/2} \Big(\sum_{i=1}^n \lambda_i u_i u_i^T \Big) D^{-1/2} p^{(0)} \\
&= D^{1/2} \Big(\lambda_1 u_1 u_1^T \Big) D^{-1/2} p^{(0)} + D^{1/2} \Big(\sum_{i=2}^n \lambda_i u_i u_i^T \Big) D^{-1/2} p^{(0)}
\end{aligned}$$

we know that $\lambda_1 = 1$ since the largest eigenvalue of M is the same as the largest eigenvalue of $D^{-1/2}AD^{-1/2}$, and we also know that $u_1 = \frac{1}{\sqrt{\sum_{i=1}^n d_i}} d^{1/2}$

this shows that the first term can be simplified as

$$\begin{aligned} D^{1/2}\left(\lambda_1 u_1 u_1^T\right) D^{-1/2} p^{(0)} &= \left(\frac{1}{\sum_{i=1}^n d_i}\right) d \; \mathbb{1}^T p^{(0)} \\ &= \pi \end{aligned}$$

this proves the desired claim

Rate of convergence

theorem. consider a random walk on an undirected graph $G = (V, E)$ that starts at a node i. Then, after k steps of the random walk, the distribution of the random walk is $p^{(k)} = P^k e_i$, where e_i is the standard basis vector that has one in the i-th entry and zeros everywhere else. Then, after k steps, the probability that the random walk is at a particular node j is $p_j^{(k)}$. The distance of this probability from the stationary distribution is then bounded by

$$|p_j^{(k)} - \pi_j| \quad \leq \quad \sqrt{\frac{d_j}{d_i}} |\lambda_2(P)|^k$$

proof. with $p^{(0)} = e_i$, we know that

$$p_j^{(k)} \quad = \quad e_j^T \left(\pi + D^{1/2} \sum_{\ell=2}^{n} \lambda_\ell^k u_\ell (u_\ell^T D^{-1/2} e_i) \right) \text{, then}$$

$$\begin{aligned} |p_j^{(k)} - \pi_j| \quad &= \quad \left| e_j^T D^{1/2} \sum_{\ell=2}^{n} \lambda_\ell^k u_\ell (u_\ell^T D^{-1/2} e_i) \right| \\ &\leq \quad \sqrt{\frac{d_j}{d_i}} \left| \sum_{\ell=2}^{n} \lambda_\ell^k (e_j^T u_\ell)(u_\ell^T e_i) \right| \leq \sqrt{\frac{d_j}{d_i}} |\lambda_2|^k \sum_{\ell=2}^{n} |e_j^T u_\ell| |u_\ell^T e_i| \end{aligned}$$

we finish the proof by showing that $\sum_{\ell=2}^{n} |e_j^T u_\ell||u_\ell^T e_i| \leq 1$
applying Cauchy-Schwarz, we know that

$$\begin{aligned} \sum_{\ell=2}^{n} |e_j^T u_\ell||u_\ell^T e_i| &\leq \sqrt{\sum_{\ell=2}^{n} |e_j^T u_\ell|^2} \sqrt{\sum_{\ell=2}^{n} |e_i^T u_\ell|^2} \\ &\leq 1 \cdot 1 \end{aligned}$$

since U is an orthonormal matrix and $\sum_{\ell=1}^{n} u_{\ell i}^2 = 1$.

- in the worst case, the error decays as $\max_{i,j} \sqrt{\frac{d_j}{d_i}} |\lambda_2|^t$
- Mixing time of a random walk is the minimum time that the error is less than $1/e$ in the worst case starting node and ending node
- in order to guarantee $\max_{i,j} \sqrt{\frac{d_j}{d_i}} |\lambda_2|^t < 1/e$, we need

$$t > \frac{1 + \frac{1}{2} \log\left(\frac{d_{\max}}{d_{\min}}\right)}{\log(1/|\lambda_2|)}$$

examples on how fast a random walk converges to the stationary distribution

a complete graph

- $P = (1/n)\mathbb{1}\mathbb{1}^T \Rightarrow \lambda_2 = 0$
- Stationary distribution is $\pi = (1/n)\mathbb{1}$
- Then, mixing time is 1
- Intuition: When a graph is well connected, it can reach any node fast.

a cycle graph

- $\lambda_2 \simeq 1 - 1/n^2$
- Stationary distribution is $\pi = (1/n)\mathbb{1}$
- Then, mixing time is $1/\log(1/(1 - 1/n^2)) \simeq n^2$
- Intuition: Random walk on a line after time t converges in the limit of $n \to \infty$ to a Gaussian distribution $N(0, t)$

a dumbbell

- ⋆ The dumbbell graph consists of two complete graphs on n vertices, joined by one edge.
- ⋆ A complete graph with n vertices is a graph with n nodes that are connected to all the other nodes in the graph.
- ⋆ $\lambda_2 \simeq 1 - 1/n^2$
- ⋆ Then, mixing time is $1/\log(1/(1-1/n^2)) \simeq n^2$
- ⋆ Intuition: Consider starting the random walk at some node that is not attached to the bridge. After one step, the random walk mixes well on one side of the graph. There is a $1/n$ chance that the random walk reaches the node attached to the bridge. And only $1/n$ chance that it crosses the bridge. So overall the probability of crossing is about $1/n^2$.

The Laplacian Matrix

The adjacency matrix A of a graph is natural but not the most useful. Eigenvalues and eigenvectors of a matrix is most useful when associated with the natural operator or the natural quadratic form

A natural **operator** associated with an undirected graph is the transition matrix of a **natural random walk** on the graph

$$P = D^{-1}A$$

where D is a diagonal matrix with the degree of each node in the diagonal

$$D_{ij} = \begin{cases} d_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

where d_i is the degree of node i, and A is the adjacency matrix

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

A natural **quadratic form** associated with an undirected graph is the **Laplacian matrix** L_G, defined as

$$L_G = D - A$$

quadratic form of L_G is useful in capturing the structure of the graph:

$$\begin{aligned} x^T L_G x &= \sum_i d_i x_i^2 - \sum_{(i,j)\in E} 2x_i x_j \\ &= \sum_i \sum_{j:(i,j)\in E} x_i^2 - \sum_{(i,j)\in E} 2x_i x_j \\ &= \sum_{(i,j)\in E} 2x_i^2 - x_i x_j \\ &= \sum_{(i,j)\in E} x_i^2 + x_j^2 - x_i x_j \\ &= \sum_{(i,j)\in E} (x_i - x_j)^2 \end{aligned}$$

it measures how smooth the function x is: $x^T L_G x$ small for smooth x

a few properties

- L_G is positive semidefinite, i.e. $x^T L_G x \geq 0$ for all x
- $\mathbb{1}$ is in L_G's null space, i.e. $L_G \mathbb{1} = 0$, since $\mathbb{1}$ is the most smooth
- for a set $S \subseteq V$, let $x \in \{0,1\}^n$ be the indicator of the set such that $x_i = 1$ if $i \in S$. Then, $x^T L_G x$ is the cut value $|c(S, S^c)|$. Precisely,

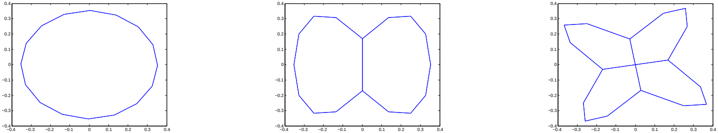
$$x^T L_G x = \frac{1}{2}\Big\{ \sum_{i\in S, j\in S^c} 1^2 + \sum_{i\in S^c, j\in S} 1^2 \Big\} = |c(S, S^c)|$$

Graph Laplacian for graph visualization

- drawing graphs is assigning coordinates to nodes (x_i, y_i)
- we might want to assign coordinates such that connected nodes are close to each other
- idea: use eigenvectors corresponding to the smallest eigen values (other than $\mathbb{1}$, which will give a trivial coordinates of placing all nodes inthe same place)
- the second smallest eigenvalue and the corresponding eigenvector minimizes the following

$$\min_{\|x\|=1, x \perp \mathbb{1}} x^T L_G x \;=\; \min_{\|x\|=1, x \perp \mathbb{1}} (x_i - x_j)^2$$

  the third smallest eigenvector minimizes the same function subject to being orthogonal to $v_1 = \mathbb{1}$ and v_2
- use v_2 and v_3 corresponding to λ_2 and λ_3, which are the smallest eigenvalues other than zero



for weighted graphs with weights w_{ij}'s, we define Laplacian matrix as

$$L_G \quad = \quad D - A$$

where $D_{ii} = \sum_k w_{ik}$ and $A_{ij} = w_{ij}$ such that

$$x^T \, L_G \, x \quad = \quad \sum_{(i,j) \in E} w_{ij}(x_i - x_j)^2$$

Graph partitioning

how well we can separate a subset S from a graph can be represented by the **isoperimetric ratio** of S

$$\theta(S) \triangleq \frac{|c(S, S^c)|}{|S|}$$

and the **isoperimetric number** of a graph is defined as

$$\theta_G \triangleq \min_{|S| \leq n/2} \theta(S)$$

theorem the second smallest eigenvalue of the graph Laplaican matrix lower bounds the isoperimetric number as

$$\frac{1}{2}\lambda_2(L_G) \leq \theta_G$$

proof of the lower bound.

consider a vector I_S indicating the set S such that

$$I_S = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}$$

for a vector x orthogonal to $\mathbb{1}$, we know that $x^T L_G x \geq \lambda_2 x^T x$. Consider $x = I_S - \frac{|S|}{|V|}\mathbb{1}$ which is orthogonal to $\mathbb{1}$. We know that

$$x^T L_G x = I_S^T L_G I_S = \sum_{(i,j) \in E} ((I_S)_i - (I_S)_j)^2 = c(S, S^c)$$

also, we know that

$$x^T x \;=\; |S| - |S|^2/|V| \;=\; |S|\left(1 - \frac{|S|}{|V|}\right)$$

this finishes the proof of the lower bound

exercise.

Problem 1.

(a) Suppose that $A \in \mathbb{R}^{7\times 5}$ has rank 4, and $B \in \mathbb{R}^{5\times 7}$ has rank 3. What are the possible values of $\mathtt{rank}(AB)$? For each value r that is possible, give an example, i.e., a specific A and B with the dimensions and ranks as given above, for which $\mathtt{rank}(AB) = r$. Try to give simple examples, and explain for each example for each value of r why AB has a rank of r.

(b) If V is a subspace in $\mathbb{R}^n$, we define $V^\perp$ as the set of vectors orthogonal to every element in V, i.e.

$$V^\perp \triangleq \{x \in \mathbb{R}^n \,|\, x^T y = 0 \text{ for all } y \in V\}\,.$$

For example if $V = \text{span}\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right)$ then $V^\perp = \text{span}\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right)$,

where $\text{span}(v_1,\ldots,v_k) = \{x \in \mathbb{R}^n \,|\, \sum_{i=1}^k a_i v_i \text{ for } a_1,\ldots,a_k \in \mathbb{R}\}$ is the subspace spanned by the set of vectors. Verify that $V^\perp$ is also a subspace.

exercise.

Problem 1. (continued)

- (c) Orthonormal basis of a subspace V of rank r in $\mathbb{R}^n$ is defined as a set of r vectors $\{u_1, \ldots, u_r\}$ such that each vector is normalized, i.e. $u_i^T u_i = 1$ and each pair is orthogonal, i.e. $u_i^T u_j = 0$ for any $i \neq j$, and they span the subspace, i.e. $\mathrm{span}(u_1, \ldots, u_r) = V$.
  Projection of a vector x onto a subspace V given an orthonormal basis matrix $U = [u_1 \;\cdots\; u_r]$ is defined by a projection matrix
  $$P \triangleq UU^T ,$$
  and the projection of a vector x is $Px = UU^T x$. Prove that all projection matrices satisfy $P^2 = P$ and $P^T = P$.
- (d) Show every $x \in \mathbb{R}^n$ can be represented as $x = v + v^\perp$ where $v \in V$ and $v^\perp \in V^\perp$.
- (e) Show that $\dim(V) + \dim(V^\perp) = n$.

exercise.

Problem 2.

Consider a tall measurement matrix $A \in \mathbb{R}^{m \times n}$ with $m > n$. Given a signal $x \in \mathbb{R}^n$, the output of the measurement is $y = Ax$. However, instead of y itself, we observe a corrupted version of y, which we denote by z. z and y differ only in one entry. For example, if the 4th entry is corrupted, then $y_i = z_i$ for $i \neq 4$ and $y_4 \neq z_4$.

Given A and z, we want to figure out which entry in z is the corrupted one. Use MATLAB to figure out which entry is corrupted, given the following measurement matrix `A` and corrupted measurement `z` in the file `corrupt.m`.
To check if a vector v is in a subspace spanned by the columns of V, you can use the MATLAB script: $\texttt{rank}([\,V\ v\,]) == \texttt{rank}(V)$, which returns 1 if and only if v is in the subspace.

exercise.

Problem 3.

- Consider a network of n smartphones that can transmit and receive radio signals. A smartphone i can choose the transmit power $P_i > 0$. When this signal reaches a smartphone j that is different from i, the received signal power is $G_{ji}P_i$.
- The signal power of i at receiver i is $S_i = G_{ii}P_i$.
- Assume all entries of G are positive
- The interference power received at smartphone i caused by interference from all other signals transmitted from other smartphones is $I_i = \sum_{k \neq i} G_{ik}P_k$.
- Signal to interference ratio (SIR) is

$$\frac{S_i}{I_i} = \frac{G_{ii}P_i}{\sum_{k \neq i} G_{ik}P_k}$$

- We want to set transmit powers P_i's such that the minimum SIR is maximized

exercise.

Problem 3. (continued.)

- We are going to minimize the maximum *interference to signal ratio*, i.e.

$$\begin{array}{ll} \text{minimize} & \max_i \frac{(\tilde{G}P)_i}{P_i} \\ \text{subject to} & P > 0 \end{array}$$

where

$$\tilde{G}_{ij} = \begin{cases} G_{ij}/G_{ii} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

- We saw in the proof of Perron-Frobenius theorem that the optimal solution of the following problem is the PF eigenvalue λ_{pf} and the corresponding eigen vector

$$\begin{array}{ll} \text{maximize} & \delta \\ \text{subject to} & Ax \geq \delta x \text{ for some } x > 0 \end{array}$$

- The solution to the above problem is also the solution to the following problem:

$$\begin{array}{ll} \text{minimize} & \max_i \frac{(Ax)_i}{x_i} \\ \text{subject to} & x > 0 \end{array}$$

exercise.

Problem 3. (continued.)

- Then, the solution of minimizing the maximum interference problem can be solved by computing the PF eigenvector of $\widetilde{G}$ and using it to assign power P_i's.
- It follows that the maximum possible SIR is $1/\lambda_{pf}$, and with optimal power allocation, all SIR's are the same.

(a) For two matrices $G1$ and $G2$ given in the file `power.m`, use MATLAB to compute $\widetilde{G}1$ and $\widetilde{G}2$. Using the function `eig(·)`, compute the spectral gap of two matrices $\widetilde{G}1$ and $\widetilde{G}2$:

$$\frac{\lambda_1(\widetilde{G}1) - \lambda_2(\widetilde{G}1)}{\lambda_1(\widetilde{G}1)}$$

Feel free to use the skeleton given in `power.m`.

exercise.

Problem 3. (continued.)

(b) Start with two random vectors of dimension 20: `x=rand(20,1)` and `y=rand(20,1)`. For each matrices $\widetilde{G}1$ and $\widetilde{G}2$, use the following algorithm to compute the Perron-Frobenius eigen vector and plot the residual error as a function of the number of iterations.
At each iteration compute $x = \widetilde{G}1x$ and $y = \widetilde{G}1y$. Compute the residual error at iteration i: `e(i) = norm(`x`/norm(`x`) - `y`/norm(`y`))`. Plot `e(i)` as a function of i for $i \in \{1, 2, \ldots, 100\}$ for both $\widetilde{G}1$ and $\widetilde{G}2$.

(c) Using the result on the spectral gap, explain why one converges faster to the Perron-Frobenius eigenvector than the other.

exercise.

Problem 4. For an undirected graph $G = (V, E)$, let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of the adjacency matrix A, where

$$A_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

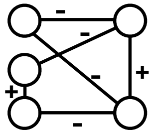
Let $d_{\text{ave}} = \frac{1}{n} \sum_i d_i$ be the average degree of the graph and $d_{\max}$ be the maximum degree.
Prove that

$$d_{\text{ave}} \;\leq\; \lambda_1 \;\leq\; d_{\max} \; .$$

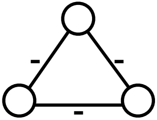
exercise.

Problem 5.

Social balance theory studies relationships between pairs of people in a group. There are two types of relationships between a pair, positive and negative. Such relationships are represented using signed undirected graph $G = (V, E, S)$ where V is the set of nodes representing each person in the group, E is the set of edges representing interactions between pairs of people, and $S : V \times V \to \{+1, -1\}$ where $S_{ij} \in \{+1, -1\}$ is the type of the relationship between a pair $(i, j) \in E$.



a balanced signed graph



an unbalanced signed graph

A signed graph is said to be **balanced** if any cycle in the graph has even number of negative edges. Prove that a signed graph is balanced if and only if there exists a partition of the edges into two sets A and B such that every edge within A are positive edges, every edge within B are also positive edges, and every edge across A and B are negative edges.