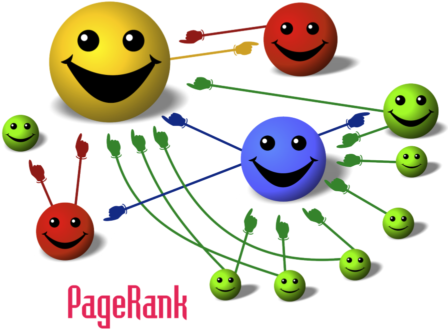
4. Spectral methods

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

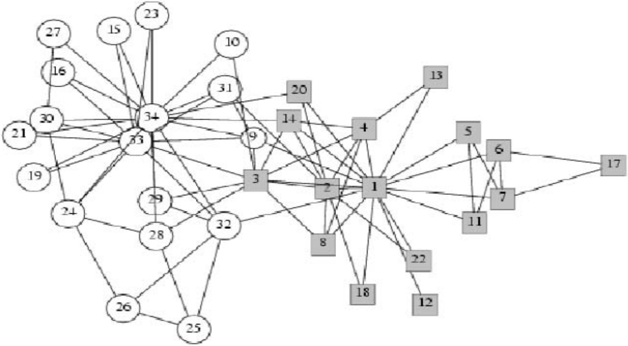
Spectral methods for network problems

- 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. $v = \text{span}(\{v_1, \ldots, v_k\})$; 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

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

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

fact 1. $\mathrm{rank}(A) = \mathrm{rank}(A^T)$
fact 2. $\mathrm{rank}(A)$ is the maximum number of independent columns (or rows) of A
fact 3. $\mathrm{rank}(A) \le \min(m, n)$
fact 4. $\mathrm{rank}(A) + \dim(\mathrm{null}(A)) = m$
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
- $\dim(\mathrm{null}(A))$ is the number of degrees of freedom crushed to zero by applying A
- $\mathrm{rank}(A)$ is the number of degrees of freedom in the output y

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

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

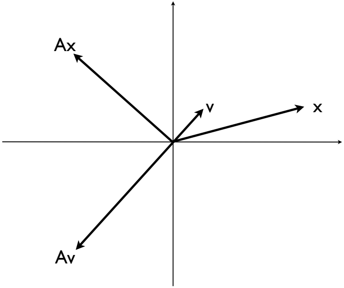
- ⋆ even if A and B are full rank, AB might not be full rank (e.g. low-rank factorization)
- ⋆ even if AB is full rank, either one of A and B might not be full rank
- ⋆ if A and B have empty null spaces, then AB has an empty null space

- ⋆ 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 $Ax = c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2$.



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

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

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

this gives

$$Ax = \Big(\frac{x^T Ax}{x^T x}\Big)x$$

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,

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

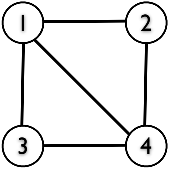
it follows that $\min_{x \in H} \frac{x^T A x}{x^T x} \leq \lambda_k$ for any k dimensional H and in particular $\min_{x \in H} \frac{x^T A x}{x^T x} \leq \lambda_k$. And we know this can be achieved with equality by choosing $H = \text{span}(\{v_1, \ldots, v_k\})$. This proves that $\lambda_k = \max_{H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

Markov chain (finite state, discrete time, homogeneous)

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

$$p(t+1) = Pp(t)$$

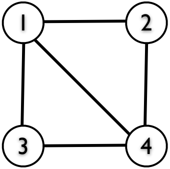
- example: random walk on a graph



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

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

- conditional probability sums to one: $\sum_j P_{ji} = 1$
- rewrite as $[1\ 1\ \dots\ 1]P = [1\ 1\ \dots\ 1]$
  $[1\ 1\ 1\ \dots\ 1]$ is a left eigenvector of P with eigenvalue 1.
- there is a corresponding eigenvector with eigenvalue 1. Let's call it $p = [p_1\ p_2\ \dots p_n]^T$.
$$p = Pp\ .$$
- **interpretation.** this eigenvector is called 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.



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

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* is 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 $A > 0$.
  - if $x \geq 0$ and $x \neq 0$, we refer to $d = (1/\mathbb{1}^T x)x$ as its distribution or normalized form.
  - $d_i = x_i/(\sum_j x_j)$ gives the fraction of the total of x, given by x_i.

- 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}

proof of 5. We assume A is diagonalizable such that exists invertible matrix V and a diagonal matrix Λ with $A = V \Lambda V^{-1}$
then, $\frac{A^k}{\lambda_{pf}^k} = V((1/\lambda_{pf}^k)\Lambda^k)V^{-1}$
since $|\lambda_i| < \lambda_{pf}$, we have $\lim_{k\to\infty}((1/\lambda_{pf}^k)\Lambda^k) = \text{diag}([1, 0, 0, \ldots, 0])$, where one of the eigenvalue is λ_{pf} that converges to one, and the rest vanish as k grows
then we know that $\lim_{k\to\infty} \frac{A^k}{\lambda_{pf}^k} = c v w^T$, for some constant c where v is the first columns of V (and also the right eigenvector corresponding to λ_{pf}) and w is the first row of V^{-1} (and also the left eigenvector corresponding to λ_{pf})
the fact that $c = \frac{1}{v^T w}$ follows from the fact that $A^k v = \lambda_{pf}^k v$ for all k

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

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

- 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

rate of convergence to stationary distribution

rate of convergence to stationary distribution depends on the largest eigenvalue that is not λ_{pf}, i.e.

$$\mu = \max\{|\lambda_2|, \ldots, |\lambda_n|\}$$

where λ_i's are the eigenvalues of P, and $\lambda_{pf} = \lambda_1 = 1$ (to be discussed later in this note)

mixing time

the *mixing time* of the Markov chain is given by

$$T = \frac{1}{\log(1/\mu)}$$

mixing time is roughly the time it takes for the distance to stationary distribution to drop by a factor of $1/e$

Perron-Frobenius theorem for population model

- Population model
  - $(X_t)_i$ denotes the number of individuals in group i at period t
  - groups could be by age, location, health, marital status, etc.
  - population dynamics is given by $X_{t+1} = AX_t$, with $A \geq 0$.
  - A_{ij} is the fraction of members of group j that move to group i, or the number of members in group i created by members of group j (e.g. in births)
  - $A_{ij} \geq 0$ means the more we have in group j in a period, the more we have in group i in next period.

  - if $\sum_i A_{ij} = 1$ for j, population is preserved in transitions out of group j
  - we can have $\sum_i A_{ij} > 1$, if there are births from group j
  - we can have $\sum_i A_{ij} < 1$, if there are deaths from group j
- Now suppose A is regular
  - PF eigenvector v gives asymptotic population distribution
  - PF eigenvalue λ_{pf} gives asymptotic growth rate (if > 1) or decay rate (if < 1)

Perron-Frobenius theorem for path count

- Consider directed graphs on n nodes, with adjacency matrix $A \in \mathbb{R}^{n \times n}$

$$A_{ij} = \begin{cases} 1 & \text{there is an edge from node } j \text{ to node } i \\ 0 & \text{otherwise} \end{cases}$$

- $(A^k)_{ij}$ is the number of paths from j to i of length k
- Now suppose A is regular, then for large k,

$$A^k \sim \lambda_{pf}^k v w^T$$

  - total number of paths of length k: $\mathbb{1}^T A^k \mathbb{1} \sim \lambda_{pf}^k (\mathbb{1}^T w)(\mathbb{1}^T v)$
  - λ_{pf} is factor of increase in number of paths when length increases by ones
  - $w_j / (\mathbb{1}^T w)$: fraction of length k paths that start at j
  - $v_i / (\mathbb{1}^T v)$: fraction of length k paths that end at i
  - $v_i w_j / (\mathbb{1}^T v)(\mathbb{1}^T w)$: fraction of length k paths that start at j end at i

Proof of Perron-Frobenius theorem for positive matrices

- Suppose $A > 0$, and consider the optimization problem

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

  note that we can assume $\mathbb{1}^T x = 1$.
- interpretation: with $y_i = (Ax)_i$, we can interpret y_i/x_i as the 'growth factor' for component i
- problem above is to find the input distribution that maximizes the minimum growth factor
- let λ_0 be the optimal value of this problem, and let v be an optimal point, i.e. $v \geq 0$, $v \neq 0$, and $Av \geq \lambda_0 v$.

- We show that λ_0 is the PF eigenvalue of A, and v is a PF eigenvector.
- First, let's show $Av = \lambda_0 v$. i.e. v is an eigenvector associated with λ_0
- We prove this by contradiction
  - Suppose $Av \neq \lambda_0 v$
  - Then there exists k such that $(Av)_k > \lambda_0 v_k$
  - Now let's look at $\tilde{v} = v + \epsilon e_k$
  - We'll show that for small $\epsilon > 0$, we have $A\tilde{v} > \lambda_0 \tilde{v}$, which means that $A\tilde{v} \geq \delta \tilde{v}$ for some $\delta > \lambda_0$
  - This contradicts the assumption λ_0 is the maximizer
  - For $i \neq k$, we have

    $$(A\tilde{v})_i = (Av)_i + A_{ik}\epsilon > (Av)_i \geq \lambda_0 v_i = \lambda_0 \tilde{v}_i$$

    so for any $\epsilon > 0$ we have $(A\tilde{v})_i > \lambda_0 \tilde{v}_i$
  - for k-th entry,

    $$\begin{aligned} (A\tilde{v})_k - \lambda_0 \tilde{v}_k &= (Av)_k + A_{kk}\epsilon - \lambda_0 v_k - \lambda_0 \epsilon \\ &= \underbrace{(Av)_k + -\lambda_0 v_k}_{>0} - \epsilon(\lambda_0 - A_{kk}) \end{aligned}$$

    since $(Av)_k - \lambda_0 v_k > 0$, we conclude that for some small $\epsilon > 0$, $(A\tilde{v})_k - \lambda_0 \tilde{v}_k > 0$

- To show that $v > 0$, suppose that $v_k = 0$
- From $Av = \lambda_0 v$, we conclude $(Av)_k = 0$, which contradicts $Av > 0$ (which follows from $A > 0$, $v \geq 0$, $v \neq 0$)

- To show $|\lambda| \leq \lambda_0$, suppose $\lambda \neq \lambda_0$ is another eigenvalue of A, i.e., $Az = \lambda z$, where $z \neq 0$. (note that λ and z can be complex numbers)
  - let the magnitude of a vector $|z|$ denote the vector with $|z|_i = |z_i|$ (for $z_j = a_j + b_j i$ we define $|z_j| = \sqrt{a_j^2 + b_j^2}$)
  - Since $A \geq 0$ we have $A|z| \geq |Az| = |\lambda z| = |\lambda||z|$
    - ★ the first inequality follows from $(A|z|)_i^2 = (\sum_j A_{ij}\sqrt{a_j^2 + b_j^2})^2 = \sum_{j,k} A_{ij}A_{ik}\sqrt{(a_j^2 + b_j^2)(a_k^2 + b_k^2)} \geq \sum_{j,k} A_{ij}A_{ik}\sqrt{(a_j a_k + b_j b_k)^2} = \left(\sum_j A_{ij}a_j\right)^2 + \left(\sum_j A_{ij}b_j\right)^2 = |Az|_i^2$
    - ★ the last equality follows from $|xy| = |(a_x + b_x i)(a_y + b_y i)| = |(a_x a_y - b_x b_y) + (a_x b_y + b_x a_y)i| = \sqrt{(a_x^2 + b_x^2)(a_y^2 + b_y^2)} = |x||y|$
  - From the definition of λ_0 as the maximizer, we conclude that $|\lambda| \leq \lambda_0$
  - To show strict inequality is harder.

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. $P = AD^{-1}$ is diagonalizable
proof. define a symmetric matrix $M = D^{-1/2}AD^{-1/2}$

$$P = AD^{-1} = D^{1/2}(D^{-1/2}AD^{-1/2})D^{-1/2} = D^{1/2}MD^{-1/2}$$

since M is symmetries, it is diagonalizable with $M = U\Lambda U^{-1}$ where the columns of U are the left eigenvectors and the rows of U^{-1} are the right eigenvectors and Λ is a diagonal matrix with eigenvalues in the diagonals

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

it follows that P has the same eigenvalues as M, and the left eigenvectors are the columns of $D^{1/2}U$ and the right eigenvectors are the rows of $U^{-1}D^{-1/2}$
among other things, this proves that P is always diagonalizable
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

Rate of convergence

- ★ Let $p_0 = a_1 v_1 + \cdots + a_n v_n$, where v_i's are eigenvectors of P such that

$$Pv_i = \lambda_i v_i$$

$$\begin{aligned} p_t &= P^t p_0 = P^t(a_1 v_1 + \cdots + a_n v_n) \\ &= P^{t-1} \sum_k a_k \lambda_k v_k \;=\; \sum_k a_k \lambda_k^t v_k \\ &= a_1 \lambda_1^t v_1 + \sum_{k=2}^{n} a_k \lambda_k^t v_k \\ &= v_1 + \sum_{k=2}^{n} a_k \lambda_k^t v_k \end{aligned}$$

  where $\lambda_1 = 1$. We can also show that $a_1 = 1$, but we omit the proof.
- ★ The error term decays as $|\lambda_2|^t$ (where $|\lambda_2| \geq \cdots \geq |\lambda_n|$)
- ★ Mixing time is when the error decays by $1/e$
- ★ We want $|\lambda_2|^t < 1/e$, then we need $t > 1/\log(1/|\lambda_2|)$.

$$\text{Mixing Time} = \frac{1}{\log(1/|\lambda_2|)}$$

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 complete binary tree

- $\lambda_2 \simeq 1 - 1/n$
- Then, mixing time is $1/\log(1/(1-1/n)) \simeq n$
- Intuition: A random walk is twice more likely to move towards a leaf than towards the root. So it takes about $O(n)$ time to reach the root.

a dumbbell

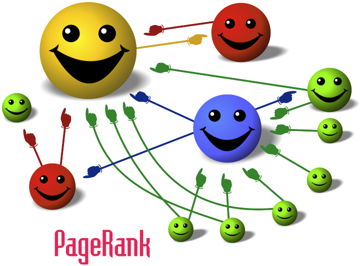
- The dumbbell graph consists of two complete graphs on n vertices, joined by one edge.
- A complete graph with n vertices is a graph with n nodes that are connected to all the other nodes in the graph.
- $\lambda_2 \simeq 1 - 1/n^2$
- Then, mixing time is $1/\log(1/(1-1/n^2)) \simeq n^2$
- Intuition: 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 $1/n^2$.

exercise: let π be the stationary distribution of a natural random walk on a directed graph with n nodes. Prove that for every node i,

$$\pi_i \geq n^{-n}$$

proof:
we know $\pi_i's$ satisfy $\pi_i = \sum_{j \in P(i)} \frac{1}{d_j} \pi_j$, where $P(i)$ is the set of predecessors of node i
we also know that $d_j \leq n$ and let π's be sorted such that $\pi_1 \leq \pi_2 \leq \ldots \leq \pi_n$
then, $\pi_1 = \sum_{j \in P(1)} \frac{1}{d_j} \pi_j \geq \frac{1}{n} \pi_2$ (when there is a self loop, it only makes π_i larger)
similarly, $\pi_1 \geq \frac{1}{n} \pi_2 \geq \left(\frac{1}{n}\right)^2 \pi_3 \geq \cdots \geq \left(\frac{1}{n}\right)^{n-1} \pi_n$
since π_i's sum to one, at least one of them has to be larger than $\frac{1}{n}$, hence $\pi_n \geq \frac{1}{n}$
this finishes the proof

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

Axiomatic approach to PageRank algorithm

we study an **axiomatic approach** to PageRank

- ⋆ L. Page, S. Brin, R. Motwani, T. Winograd, "The PageRank citation ranking: Bringing order to the web", 1999
- ⋆ Altman, Tennenholtz, "Ranking systems: the pagerank axioms", 2005

PageRank algorithm

- ⋆ input: a directed graph $G = (V, E)$, where a directed edge (i, j) indicates that a hyperlink (in the context of web pages) or support/approval/trust (in social contexts)
- ⋆ output: scores for each node $\pi = (\pi_1, \ldots, \pi_n)$
- ⋆ algorithm:

1. compute the transition matrix of a natural random walk on the directed graph as

$$P = A^T D^{-1} \text{, where}$$

$$P_{ij} \triangleq \mathbb{P}(X_{t+1} = i \,|\, X_t = j) = \begin{cases} \frac{1}{\text{out-degree of node } j} & \text{if } (j,i) \in E \\ & \text{otherwise} \end{cases}$$

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

and D is a diagonal matrix with D_{ii} equal to the out-degree of node i.

2. output the stationary distribution π satisfying $\pi = P\pi$

an axiomatic approach to ranking algorithms focus on characterizing a set of axioms that a ranking algorithm of interest must satisfy, and tries to identify the unique ranking algorithm that satisfies the set of proposed axioms

Altman and Tennenholtz studied five axioms satisfied by PageRank and proved that PageRank is the only algorithm satisfying all five axioms

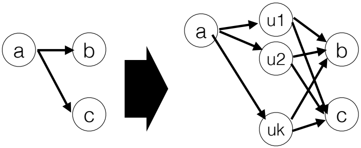
we will first prove that PageRank satisfies all five axioms

axiom 1: isomorphism

- a ranking system satisfies *isomorphism* if its ranking does not depend on the name of the nodes
- in particular, if two nodes have the same predecessor and successors, then their score and ranking must be the same

axiom 2: vote by committee

- ⋆ a ranking system satisfies *vote by committee* if a node can indirectly vote through a committee which casts the same set of votes as the original voter and the resulting ranking does not change
- ⋆ precisely, consider a voter a and its successors $S(a)$ in G
- ⋆ create a new graph G' by adding committee $\{u_1, \ldots, u_k\}$ of size k and removing all the outgoing edges of a and adding edges from a to all committee and from all committee to all nodes in $S(a)$



★ **lemma.** PageRank satisfies vote by committee.

consider the random walk P on $G = (V, E)$ with stationary distribution π which satisfies the following (P is for the example in the figure above)

$$
\begin{aligned}
\pi_a &= \sum_i P_{ai}\pi_i \\
\pi_b &= \frac{1}{d_a}\pi_a + \sum_{i \neq a} P_{bi}\pi_i \\
\pi_c &= \frac{1}{d_a}\pi_a + \sum_{i \neq a} P_{ci}\pi_i
\end{aligned}
\qquad
P = \begin{array}{cc}
\begin{array}{cccc} & a & b & c \end{array} & \\
\begin{bmatrix}
\ddots & \vdots & \vdots & \vdots \\
\cdots & 0 & 0 & 0 \\
\cdots & 1/2 & 0 & 0 \\
\cdots & 1/2 & 0 & 0
\end{bmatrix} & \begin{array}{c} \\ a \\ b \\ c \end{array}
\end{array}
$$

we next show that the new random walk P' on the new graph $G' = (V \cup \{u_i\}_{i \in [k]}, E')$ with added committee has a Perron-Frobenius eigenvector whose entries corresponding to the nodes in the original graph does not change.

consider a new vector $v = [\pi\,,\ \frac{1}{k}\pi_a\,,\ \cdots\,,\ \frac{1}{k}\pi_a]$, and we show that this new vector is the PF eigenvector of P'

$$
\begin{aligned}
\pi_a &= \sum_i P_{ai}\pi_i \\
\pi_{u_i} &= \frac{1}{k}\pi_a \\
\pi_b &= \sum_{j=1}^{k} \frac{1}{2}\pi_{u_j} + \sum_{i\in V\setminus\{a\}} P_{bi}\pi_i \\
\pi_c &= \sum_{j=1}^{k} \frac{1}{2}\pi_{u_j} + \sum_{i\in V\{a\}} P_{ci}\pi_i
\end{aligned}
$$

$$
P' = \begin{array}{c} \begin{array}{cccccccc} & a & b & c & u_1 & \cdots & u_k \end{array} \\ \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \cdots & 0 & 0 & 0 & 1/2 & \cdots & 1/2 \\ \cdots & 0 & 0 & 0 & 1/2 & \cdots & 1/2 \\ \cdots & 1/k & 0 & 0 & 0 & \cdots & 0 \\ \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 1/k & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \end{array}
$$

hence, the stationary distribution of P' is proportional to v up to a normalization, and the relative ordering does not change after adding the committee axiom 3: self-edge

- a ranking system satisfies *self-edge* if adding a self-edge strengthens the node but does not change the ranking of the rest of the nodes
- the new stationary distribution is $\pi_i' = \frac{d_i+1}{d_i}\pi_i$ and the rest unchanged

$$\pi_i = \frac{s_i}{d_i}\pi_i + \sum_{j\in P(i)\setminus\{i\}} P_{ij}\pi_j$$

and for a $j \in S(i)$,

$$\pi_j = \frac{1}{d_i}\pi_i + \sum_{k\in P(j)\setminus\{i\}} P_{jk}\pi_k$$

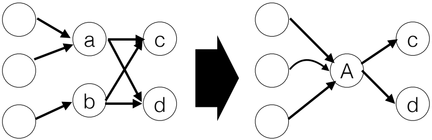
$$\pi_i' = \frac{1+s_i}{1+d_i}\pi_i' + \sum_{j\in P(i)\setminus\{i\}} P_{ij}\pi_j$$

and for a $j \in S(i)$,

$$\pi_j = \frac{1}{1+d_i}\pi_i' + \sum_{k\in P(j)\setminus\{i\}} P_{jk}\pi_k$$

axiom 4: collapsing

- a ranking system satisfies *collapsing* if nodes with the same set of successors can be merged into a single giant node with the same set of successors as the original nodes and all the incoming edges of the original nodes, and the ranking of the nodes that was not merged remains unchanged
- precisely, in the figure below, ranking of nodes other than a and b should not be changed after collapsing a and b into a giant node A



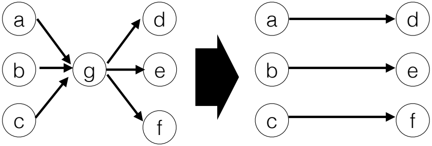
on the original graph, the Markov chain P and stationary distribution π, with two nodes a and b that have the same set of successors of size k, satisfy

$$\begin{aligned}
\pi_a &= \sum_{i \in S(a)} P_{ai}\pi_i \\
\pi_b &= \sum_{i \in S(b)} P_{bi}\pi_i \\
\pi_c &= \frac{1}{k}(\pi_a + \pi_b) + \sum_{i \neq a,b} P_{ci}\pi_i \\
\pi_d &= \frac{1}{k}(\pi_a + \pi_b) + \sum_{i \neq a,b} P_{di}\pi_i
\end{aligned}$$

$$\begin{aligned}
\pi_A &= \sum_{i \in S(a)} P_{ai}\pi_i + \sum_{i \in S(b)} P_{bi}\pi_i \\
\pi_c &= \frac{1}{k}\pi_A + \sum_{i \neq a,b} P_{ci}\pi_i \\
\pi_d &= \frac{1}{k}\pi_A + \sum_{i \neq a,b} P_{di}\pi_i
\end{aligned}$$

looking at the equation on the right for the stationary distribution on the new collapsed graph G', we see that the Pf eigenvector for the new P' is $\pi' = [\pi_1 \cdots (\pi_a + \pi_b) \cdots \pi_n]$ where a and b are merged and the rest of the vector is unchanged axiom 5: proxy

a ranking system satisfies *proxy* if k nodes of equal rank and out-degree one that voted for k nodes via a proxy can achieve the same result by voting for one node each



$$
\begin{aligned}
\pi_g &= \pi_a + \pi_b + \pi_c = 3\pi_a \\
\pi_d &= \frac{1}{3}\pi_g + \sum_{i \notin \{a,b,c,g\}} P_{di}\pi_i
\end{aligned}
$$

$$
\pi_d = \pi_a + \sum_{i \notin \{a,b,c,g\}} P_{di}\pi_i
$$

theorem. (Altman and Tennenholtz, 2005) ranking system satisfies axioms 1–5 if and only if it is a PageRank algorithm.

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

the most 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

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

the most 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,

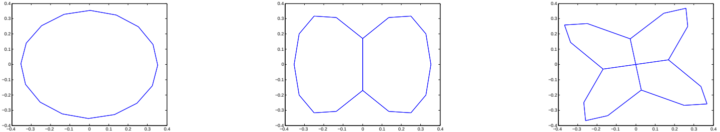
$$x^T L_G x = \frac{1}{2}\Big\{ \sum_{i\in S, j\in S^c} 1^2 + \sum_{i\in S^c, j\in S} 1^2 \Big\} = |c(S, S^c)|$$

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

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

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

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

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

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

proof of the lower bound.

consider a vector I_S indicating the set S such that

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

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

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

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

Homework 4.

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.

Homework 4.

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. $\text{span}(u_1, \ldots, u_r) = V$.
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 $\dim(V) + \dim(V^\perp) = n$.

Homework 4.

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: `rank`$([\,V\ v\,]) == $`rank`$(V)$, which returns 1 if and only if v is in the subspace.

Homework 4.

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

Homework 4.

Problem 3. (continued.)

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

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

where

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

- 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:

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

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`.

Homework 4.

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 $i \in \{1, 2, \dots, 100\}$ 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.

Homework 4.

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

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

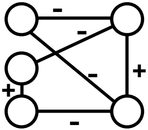
Let $d_{\text{ave}} = \frac{1}{n}\sum_i d_i$ be the average degree of the graph and $d_{\max}$ be the maximum degree.
Prove that

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

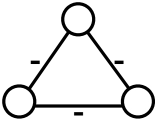
Homework 4.

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.