# Graph

→ Adjacency: graph is rather two categories or type, or rather have a adjacency matrix of a subgraph of this
→ multiplex graph: is rather having the types where they related to each other (e.g. transportation system, interactions, mode of transportation)
→ Node identification: graph link assumption or on the
→ Graph embedding: embedding nodes
→ link prediction & link/entity disambiguation
Statistical interactions generalization

→ Homophily: is when nodes share attributes with their neighbors
→ Structural equivalence: if nodes with similar neighborhood structure will have similar labels
→ Community detection: is unsupervised and we use it find latent community structure in the input graph
Graph clustering: is an attempt to find distinctly dense between graphs

Node function
→ Node degree $d_v = \sum_u A[u,v]$ → how many neighbor
→ Node centrality
- Eigenvector centrality
  Centrality → importance
  $e_v = \frac{1}{\lambda_1 \sum_u} \sum_u A[u,v] e_u \ \forall v \in V$ → constant
  $\underbrace{Ae = \lambda_1 e}_{}$ → Eigenvector equation
  Intuitive: node is valuable if it is close to valuable nodes or is close to valuable nodes

$A^{k+1} = A \cdot A^k$

Power iteration → stability for eigenvector

$A^{k+1}$ my matrix $\neq$ $A^k \neq 0$ $\left(\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}\right)$

$A \cdot x^k = \lambda^k$
$A^k = \begin{bmatrix} \lambda_1^k & \\ & \lambda_2^k \end{bmatrix}$

Eigenvalue of the $n \times n$ matrix are obtained by solving the characteristic equation $\det(A - \lambda I) = 0$
$\det(A - \lambda I) = \lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1 \lambda + c_0 = 0$
for $n \geq 5$, solving polynomial for eigenvalues is numerical iteration for approximating roots of polynomial are at high degree
$n \geq 5$ → unstable or numerically unstable

An eigenvector corresponding to one and it has the dominant/leading eigenvalue of $A$
find $\det(A - \lambda I) = 0$ → $\lambda_1$ → leading eigenvalue

$f(\lambda) = \det(A - \lambda I)$ → characteristic polynomial
$A = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}$
$f(\lambda) = \det\left(\begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right)$
$= \begin{vmatrix} 2-\lambda & -3 \\ 1 & -2-\lambda \end{vmatrix} = (2-\lambda)(-2-\lambda) + 3$
$= -4 - 2\lambda + 2\lambda + \lambda^2 + 3$
$= \lambda^2 - 1 = 0$

$\lambda_1 = 1$, $\lambda_2 = -1$
dominant $\lambda = \lambda_1 = 1$
Power method → iteration → approximate the dominant eigenvalue of a matrix

$A \cdot x = \lambda x$
$x_1 = A \cdot x_0 = A \cdot (A x_0) = A^2 x_0$
$x_2 = A x_1 = A(A^2 x_0) = A^3 x_0$
$x_k = A x_{k-1} = A(A^{k-1} x_0) = A^k x_0$

for $A = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}$
initial, you are region
$x_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$x_1 = A x_0 = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \rightarrow \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$x_2 = A x_1 = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \rightarrow \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$x_k = A x_{k-1} = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \rightarrow \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$\lim = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for $\lambda_1 = 1$

$\lambda = \frac{A x \cdot x}{x \cdot x}$ → Rayleigh quotient
$\lambda = \frac{x^T A x}{x^T x}$ → Rayleigh quotient
$x, \lambda$ eigen value and eigenvector

$A x = \lambda x$
$\frac{A x \cdot x}{x \cdot x} = \frac{\lambda x \cdot x}{x \cdot x} = \frac{\lambda (x \cdot x)}{x \cdot x} = \lambda$

$x = \begin{bmatrix} 2.97 \\ 1 \end{bmatrix}$
$A x = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 2.97 \\ 1 \end{bmatrix} = \begin{bmatrix} 5.94 - 3 \\ 2.97 - 2 \end{bmatrix} = \begin{bmatrix} -0.06 \\ 0.97 \end{bmatrix}$
$A x \cdot x = \begin{bmatrix} -0.06 \\ 0.97 \end{bmatrix} \cdot \begin{bmatrix} 2.97 \\ 1 \end{bmatrix} = (-0.06)(2.97) + (0.97)(1) = -0.20 + 0.97$
$x \cdot x = 2.97^2 + 1^2 = 9.94$
Rayleigh coefficient $= \frac{-0.20}{9.94} \approx -2.01 \approx -2$

## Power method with scaling

→ Goal is to scale down representation. Use the component of $Ax$ with largest absolute value and multiply $Ax$ by the reciprocal of that component

$A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
$x_0 = (1,1,1)$
$A x_0 = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$ → largest absolute entry has reciprocal $= \frac{1}{3}$
by scaling $x_1 = \frac{1}{3}\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
$A x_1 = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$ $x_2 = \frac{1}{3}\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

## Convergence of the power method

$A$ = $n \times n$ diagonalizable (i.e. has $n$ linearly independent eigenvalues)
with unique eigenvalues
$\lambda_1$ is dominant $\lambda$
Then there exists $x_0$ (an seed)
such that the sequence of vectors
$Ax_0, A^2x_0, A^3x_0, \dots A^kx_0 \rightarrow$ a multiple of the dominant eigenvector of $A$

$x_0 = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$
$A x_0 = A(c_1 x_1 + c_2 x_2 + \dots + c_n x_n)$
$= c_1 A x_1 + c_2 A x_2 + \dots + c_n A x_n$ $[A x_i = \lambda_i x_i]$
$= c_1(\lambda_1 x_1) + c_2(\lambda_2 x_2) + \dots + c_n(\lambda_n x_n)$
following the upper alone in power lower
we get
$A^k x_0 = c_1(\lambda_1^k x_1) + c_2(\lambda_2^k x_2) + \dots + c_n(\lambda_n^k x_n)$
$A^k x_0 = \lambda_1^k \left[c_1 x_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k x_2 + \dots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k x_n\right]$

$\lambda_1 \gg \lambda_i \ \forall i = 2, 3, \dots, n$
$\left|\frac{\lambda_2}{\lambda_1}\right|, \left|\frac{\lambda_3}{\lambda_1}\right|, \dots \left|\frac{\lambda_n}{\lambda_1}\right| < 1$
$\therefore \left(\frac{\lambda_2}{\lambda_1}\right)^k, \left(\frac{\lambda_3}{\lambda_1}\right)^k, \dots \left(\frac{\lambda_n}{\lambda_1}\right)^k \rightarrow 0$ as $k \rightarrow \infty$
$A^k x_0 \approx \lambda_1^k c_1 x_1 \quad c_1 \neq 0$
$\therefore A^k x_0 \rightarrow$ multiple of $\lambda_1 \rightarrow$ dominant eigenvector of $A$

$A^k x_0 \rightarrow \lambda_1^k c_1 x_1 \quad c_1 \neq 0$

If $|\lambda_2| / |\lambda_1|$ is small power method will converge quickly since remainder is close to 1

$e^{t+1} = A e^t$
$e^0 = (1, 1, 1, \dots, 1)^T$
$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ → graph with cut line, undirected
$\rightarrow \begin{bmatrix} 1+0+0 \\ 1+1+0 \\ 0+0+0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$
$e^1 = A e^0$
$= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$
$e^2 = A e^1 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0.5 \\ 1 \end{bmatrix} = \begin{bmatrix} 0+0+0 \\ 0+0+0 \\ 0+0+0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$e^t$ = number of length-$t$ paths arriving at each node

- The dwelling coefficient
  $c_u = \frac{|(v_1, v_2) \in E : v_1, v_2 \in N(u)|}{\binom{d_u}{2}}$ → neighborhood of $u$

## Graph kernels

- Bag of nodes → based on node-level attributes and can miss global graph prop
- The Weisfeiler-Lehman kernel
  iteratively relabel nodes in being neighborhood
  ↳ compute summary statistics over labeled nodes
  ↳ repeat for graph
  ↳ if we want to look for graph isomorphisms

## Neighborhood overlap statistics

- $S[u,v] = |N(u) \cap N(v)|$
  ↳ relationship between node $u$ & $v$

$S \in \mathbb{R}^{|V| \times |V|}$ → structurally matrix summarizing all the pairwise node scores

$\underbrace{P(A[u,v] = 1)}_{\text{likelihood of an edge b/w } u \text{ \& } v} \propto \underbrace{S[u,v]}_{\text{neighborhood overlap measure}}$

- Local overlap measures → link prediction, identification of community detection

Sorensen index
$S_{Sorensen}[u,v] = \frac{2|N(u) \cap N(v)|}{d_u + d_v}$
$S_{Sorensen} \in \mathbb{R}^{|V| \times |V|}$
$S_{Salton}[u,v] = \frac{2|N(u) \cap N(v)|}{\sqrt{d_u d_v}}$
$S_{Jaccard}[u,v] = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$

Normalization is to prevent edges with nodes that have $d \gg 1$ from dominating the overlap measure

Resource Allocation index → counts inverse degrees
$S_{RA}[v_1, v_2] = \sum_{u \in N(v_1) \cap N(v_2)} \frac{1}{d_u}$ → low-degree neighbors is more informative than a high degree neighbor

Adamic-Adar index
$S_{AA}[v_1, v_2] = \sum_{u \in N(v_1) \cap N(v_2)} \frac{1}{\log(d_u)}$

- Global overlap → good for community detection
  - Katz index $S_{Katz}[u,v] = \sum_{l=1}^{\infty} \beta^l A^l[u,v]$ → based by node degree
    $\beta \in \mathbb{R}^+$ $\beta < 1$ → minimize contribution of long-paths
    for $d_u, d_v \gg 1$
    $S_{Katz}[u,v]$ will be larger, compared to other pairs
  - Leicht et al
    $E[A]$ = number of expected paths b/w 2 nodes
    Configuration model → assumes that we know a random graph with the same set of degrees as our given graph.
    $E[A[u,v]] = \frac{d_u d_v}{2m}$ → $d_u$ edges leaving $u$ have a $\frac{d_v}{2m}$ chance of ending at $v$
    m: $|E|$ = total number of edges in the graph

$E[A^2[v_1, v_2]] = \frac{d_{v_1} d_{v_2}}{(2m)^2} \sum_{u \in G} (d_u - 1) d_u$

$E[A^3[v_1, v_2]] = \frac{d_{v_1} d_{v_2}}{(2m)^3}$