Graph Data (pt 2)

Avg path length for connected graph

$$\mu_L = \frac{\sum_i \sum_{j>i} d(v_i, v_j)}{\binom{n}{2}} = \frac{2}{n(n-1)} \sum_i \sum_{j>i} d(v_i, v_j)$$

# of verts

Eccentricity

of v_i is the max distance between v_i and all other nodes in the graph

if graph is disconnected restrict to verts w/ path to v_i

radius of connected graph $r(G)$: min eccentricity

$$r(G) = \min_i \{\underbrace{e(v_i)}_{\text{eccentricity of } v_i}\} = \min_i \{\max_j \{d(v_i, v_j)\}\}$$

diameter of connected graph $d(G)$: max eccentricity

$$d(G) = \max_i \{e(v_i)\} = \max_{i,j} \{d(v_i, v_j)\}$$

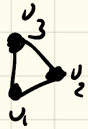
Centrality analysis

Goal! rank verts by importance (or centrality)

$C: V \to \mathbb{R}$ ← centrality or ranking function

Degree

$c(v_i) = d_i$



$c(v_i) = 2 \ \forall i$

$v_1 \leq v_2 \quad v_2 \leq v_1 \quad$ w/ $\quad v_1 \neq v_2$

example w/ total order too strong

Eccentricity centrality

$$c(v_i) = \frac{1}{e(v_i)} = \frac{1}{\max_j \{d(v_i, v_j)\}}$$

vertex w/

lowest eccentricity $r(G)$ → center node

highest eccentricity $d(G)$ → periphery node

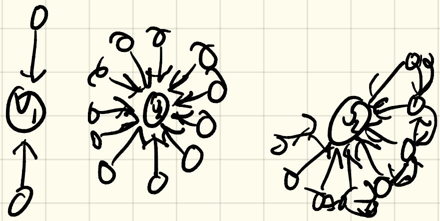
radius of graph

diameter of graph

Web Centralities

assume directed graphs

Prestige (eigenvector centrality)



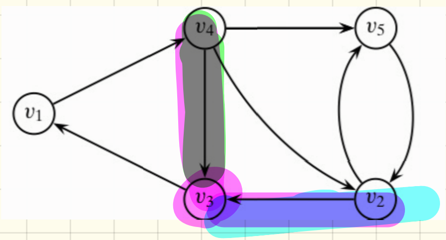
Let $G=(V,E)$ be a graph

$|V|=n$

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

let $p(u)$ be the presteige score of u

How to get presteige score



$$\mathbf{A}=\begin{pmatrix}0&0&0&1&0\\0&0&1&0&1\\1&0&0&0&0\\0&1&1&0&1\\0&1&0&0&0\end{pmatrix}$$

$$p(v_3)=A(1,3)\,p(v_1)+A(2,3)\,p(v_2)+\dots+A(5,3)\,p(v_5)$$

$$=\sum_i A(i,3)\,p(i)$$

as in general

$$p(v)=\sum_u A(u,v)\,p(u)=\sum_u A^T(v,u)\,p(u)$$

$p'=A^Tp \leftarrow$ n dim col vector w/ i^{th} entry is prestege score of v_i

$$p_k = A^T p_{k-1}$$

$$= A^T \left(A^T p_{k-2}\right) = \left(A^T\right)^2 p_{k-2}$$

$$= \left(A^T\right)^2 \left(A^T p_{k-3}\right) = \left(A^T\right)^3 p_{k-3}$$

$$= \left(A^T\right)^k p_0$$

in general

$k \to \infty \qquad p_k \to$ domanit eigenvector

$$i = \underset{j}{\operatorname{argmax}} \{p_k[j]\}$$

$$\frac{p_k[i]}{p_{k-1}[i]} \longrightarrow \lambda$$

Power iteration method for dom eigenvector and eigen value

Algorithm 4.1: Power Iteration Method: Dominant Eigenvector

POWERITERATION $(\mathbf{A}, \epsilon)$:

```
k ← 0 // iteration
p_0 ← 1 ∈ R^n // initial vector
repeat
    k ← k + 1
    p_k ← A^T p_{k-1} // eigenvector estimate
    i ← argmax_j {p_k[j]} // maximum value index
    λ ← p_k[i] / p_{k-1}[i] // eigenvalue estimate
    p_k ← (1 / p_k[i]) p_k // scale vector
until ||p_k − p_{k-1}|| ≤ ε
p ← (1 / ||p_k||) p_k // normalize eigenvector
return p, λ
```

Page rank

Idea! study prob of a very simplifide random websurfer

on a page! randomly choose a link

or w/ low prob choose a random page

Normalized prestege

assume out degree of all verts is gt 0

Recall $od(u) = \sum_v A(v,u)$

prob visiting v from u is $1/od(u)$

Init $p(v_0) = 1$

$$p(v) = \sum_u \boxed{\frac{A(u,v)}{od(u)}} p(u) = \sum_u N(u,v)p(u) = \sum_u N^T(u,v)\, p(u)$$

$$N(u,v) = \begin{cases} 1/od(u) & \text{if } (u,v) \in E \\ 0 & \text{if } (u,v) \notin E \end{cases}$$

$p' = N^T p$ ← normalized prestige vector

Next add random jump

e.g. complete graph

$$A_r = 1_{n \times n} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

$od(v) = n$

prob of jumping from u to v is $1/od(u) = 1/n$

only consider random jumps

$$p(v) = \sum_u \frac{A_r(u,v)}{od(u)} p(u) = \sum_u N_r(u,v)\, p(u) = \sum_u N_r^T(v,u)\, p(u)$$

$$N_r = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \ddots & \vdots \\ \vdots & & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

$p' = N_r^T p$ ⟸ power iteration

Full page rank

combine random jumps and normalized prestege

given α prob of random jump

$$p' = \underbrace{(1-\alpha) N^T p}_{\text{follow link}} + \underbrace{\alpha N_r^T p}_{\text{random jump}} = \underbrace{\left((1-\alpha)N^T + \alpha N_r^T\right)}_{M^T} p = M^T p$$

⟸ looks familar solve w/ power iteration

remove $od(\cdot) > 0$ assumption

define $\alpha = 1$ when $od(u) = 0$
redefine u^{th} row of M as

$$M_u = \begin{cases} M_u & \text{if } od(u) > 0 \\ \frac{1}{n} 1_n^T & \text{if } od(u) = 0 \end{cases}$$

↖ row vector of $\frac{1}{n}$