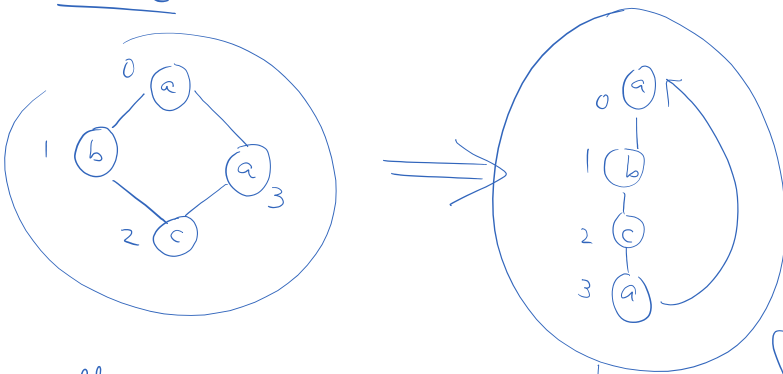
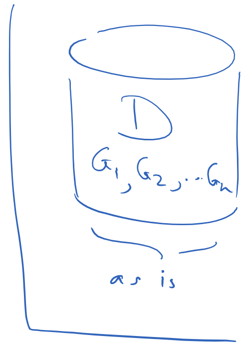
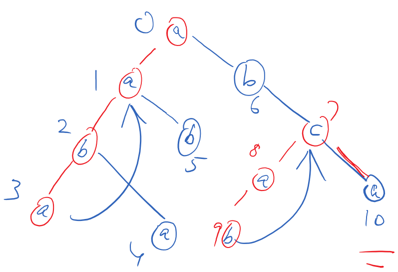
DFS code





all patterns should be kept in min DFS code format



DFS code

0	1	a	b	_
1	2	b	c	_
2	3	c	a	_
3	0	a	a	_

0	1	a	a	_
1	2	a	b	_
2	3	b	a	_
3	1	a	a	_
2	4	b	a	_
1	5	a	b	_
0	6	a	b	_
6	7	b	c	_
7	8	c	a	_
8	9	a	b	_
9	7	b	c	_
7	10	c	a	_

RMP: right most path (Pre-order tree traversal)

[0 ~~1 2 3 4 5~~ 6 7 ~~8 9~~ 10] N-L-R

0-6-7-10

see lecture24b.pdf

Graph analysis

→ graph pattern mining

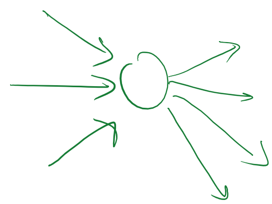
→ topological / structural characteristics

→ diffusion

struct

...pological / structural characteristics

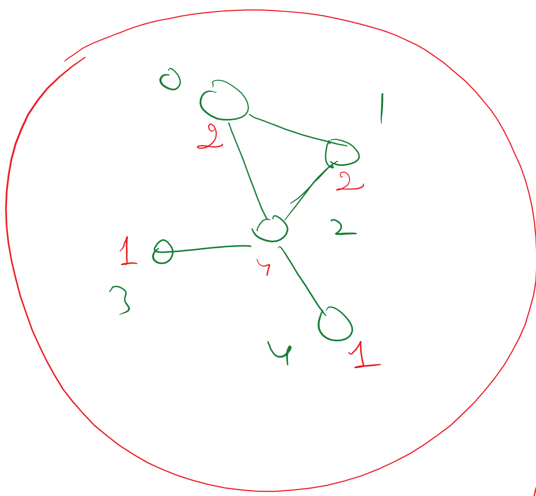
diffusion



structure + labels

Power-law degree distributions

N_i : # of nodes with degree i



N_0, N_1, N_2, N_3, N_4

0 2 2 0 1 ⟵ degree sequence frequen

degree distribution

$$\boxed{f(k) = P(\text{degree } k) = \frac{N_k}{N}}$$

$$\boxed{f(k) \propto k^{-\gamma}}$$ ⟵ degree distributions of real-world networks tend to be power-laws

γ is constant

$$f(k) = c \cdot k^{-\gamma}$$

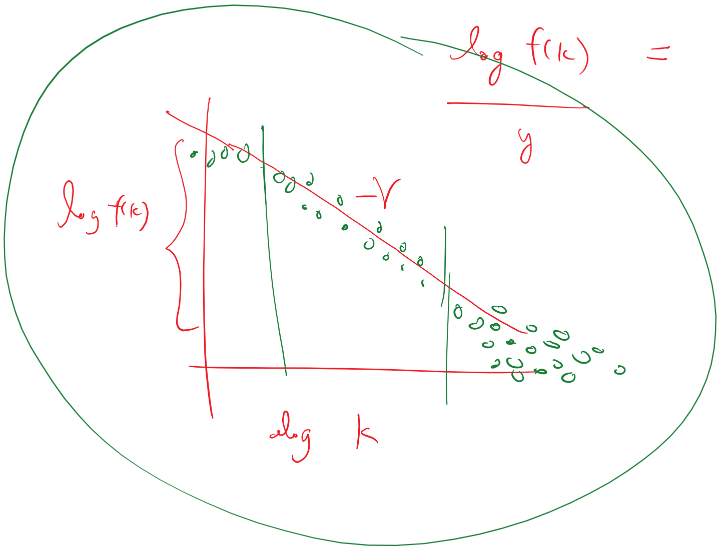
↑ proportionality constant

proportionality constant

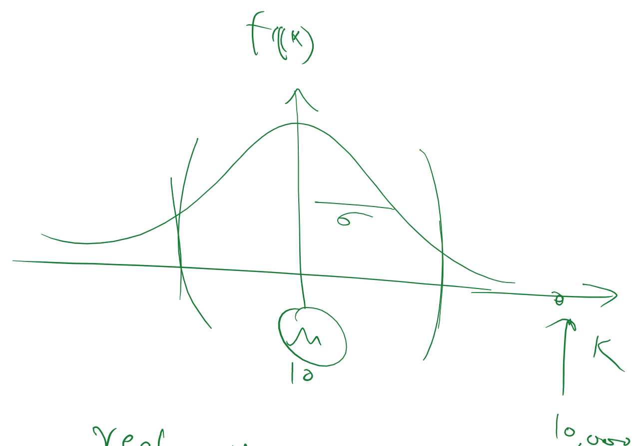
$$\log f(k) = \underbrace{\log(c \cdot k^{-\gamma})}$$

$$= \log(c) + \log(k^{-\gamma})$$

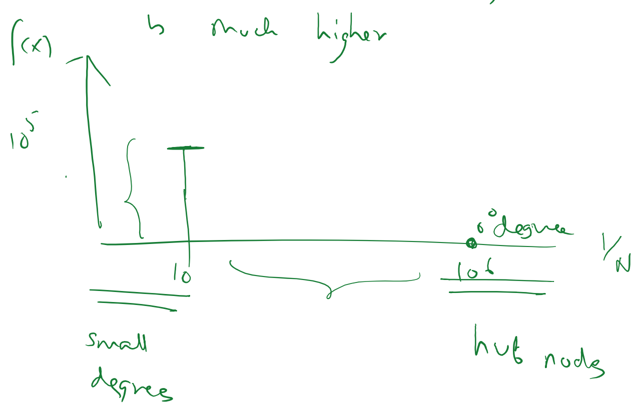
$$\underbrace{\log f(k)}_{y} = \underbrace{\log(c)}_{b} - \underbrace{\gamma}_{\text{slope}} \underbrace{\log k}_{x}$$



Power-law / scale-free network



Prob of seeing high-degree nodes (hub nodes) is much higher



real-world graphs are far from normal

$$\underbrace{\mu = \frac{10 \cdot 10^5 + 3 \cdot 10^6}{10^5} =}$$

mean degree does not convey much information

Path-lengths

Given "Undirected" graph G

$$\underbrace{d(x,y)}_{\text{distance}} = \underbrace{\text{length of the shortest path from } x \text{ to } y}_{\#\text{ of hops}}$$

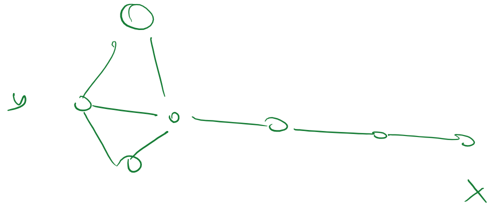
$$\mu_{dist} = \frac{\sum_{x} \sum_{y>x} d(x,y)}{\underbrace{\binom{n}{2}}_{\text{assuming } G \text{ is connected}}}$$

$${}^{n}C_{2} = \frac{n.(n-1)}{2}$$

$d(x,y) = \infty$ if there is no path between x & y

Eccentricity of vertex x

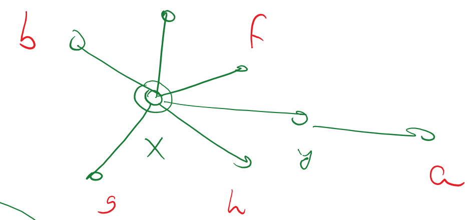
$$e(x) = \max_{y} \{ d(x,y) \}$$



Radius (G) :

$$\min_{x} \left\{ \max_{y} \{ d(x,y) \} \right\}$$

e



$e(x) = 2$

$e(y) = 2$

Radius(G) = 2

x, y are called "center" nodes

diameter(G) : $\max_{x} \{ \max_{y} \{ d(x,y) \} \}$

$= \max_{x,y} \{ d(x,y) \}$

dia(G) = 3

Such vertices are "peripheral"

"effective" radius / diameter

↓

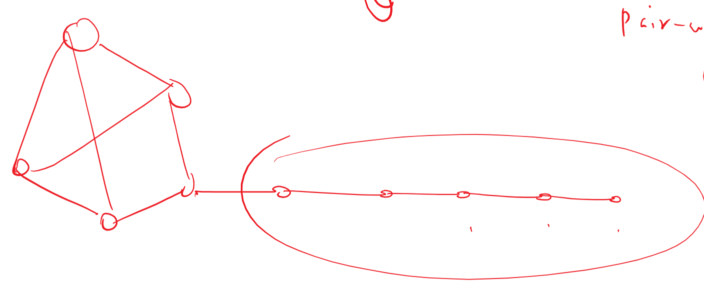
trimmed

90% radius

90% diameter

θ

⇒ remove $1-\theta$ of the largest pair-wise distances & then compute the radius / diameter



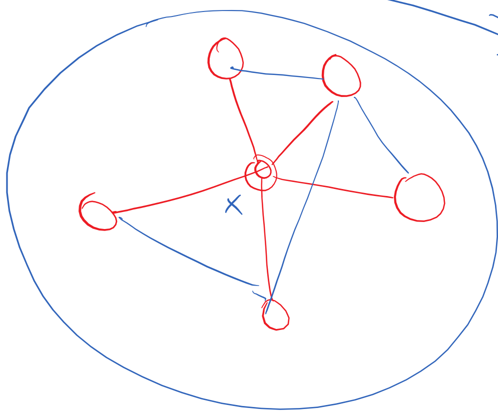
Real-world network are "small-world"

avg path length / diameter $\propto O(\log n)$

→ also in time

→ densification in time

Clustering coefficient / transitivity



ego-network

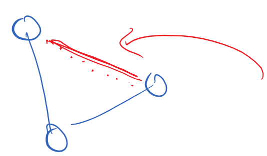
$$CC(x) = \frac{\text{\# of edges among } x\text{'s neighbors}}{\text{\# of possible edges between them}}$$

$$= \frac{4}{{}^5C_2}$$

$$= \frac{4}{10} = 40\%$$

$$\boxed{CC(G) = \frac{1}{n} \sum_x CC(x)}$$

avg. clustering coef



V — closure or triangle closure

$$\boxed{\text{transitivity} = \frac{\text{\# closed triangles} \times 3}{\text{\# of V's}}}$$

real-world networks have higher clustering coefficient than purely random networks