* 1/11/22 CN.

Expt. No.______ Page No.______

Power law

$p_k \sim k^{-\alpha}$ $\Big\}$ Normalizing

$p_k = C \cdot k^{-\alpha}$

$\sum_k p_k = 1.$

$\sum_k k^{-\alpha} = \zeta(\alpha)$

Riemann - Zeta fx

$p(k) = Ck^{-\alpha}$ Continuum.

Normalizing

$$\int p(k)\, dk = C \int_{k_{min}}^{\infty} k^{-\alpha}\, dk = C \left[\frac{k^{1-\alpha}}{1-\alpha} \right]_{k_{min}}^{\infty}$$

$\alpha > 1$

$$= C \cdot \frac{k_{min}^{1-\alpha}}{\alpha - 1} = 1$$

$$\therefore C = (\alpha - 1)\, k_{min}^{\alpha - 1}$$ *

$$\langle k^r \rangle = C \cdot \left[\frac{k^{1+r-\alpha}}{1+r-\alpha} \right]_{k_{min}}^{k_{max}}$$

Empirically $\alpha \in [2, 3]$ across all.

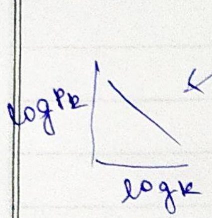
Date ________ Teacher's Signature :________

Value around the avg - $\langle k \rangle \pm \frac{\sqrt{k}}{\sqrt{n}}$ → Confidence

- Avg value - scale value.

∴ here no confidence is there as ∞ for finite value Confidence value → 0.

⇒ It's referred to as scale free dist.



$p(ak) = \underbrace{a^{-k} \cdot k^{-\alpha}}_{\text{dist remains the same.}}$

scaled value

FRACTALS → scale free construct

powerLaw dist ↔ correlated

PARETO'S 80:20

* 3/11

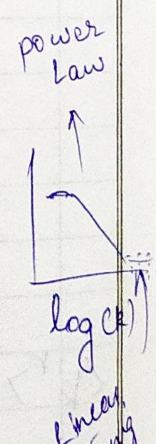
STD — Confidence

* Draw the distribution:-

$\frac{N_k}{N}$ ← $\frac{\#\text{nodes of deg } k}{\#\text{ nodes}}$

$k \in [1, k_{max}]$

$\log(P_k)$



logarithmic binning

$b_0 = 1$

$b_1 = 2, 3$

$b_2 = 4, 5, 6, 7$

b_i → 2's ith power.

↳ (>10 degrees)

Clauset / Newman (2004)

PDF - noise for hubs
→ solution is logarithmic binning

Expt. No

Page No

CDF :-

↓ See Median.

Rank-frequency Method

		r_i/n
10	1	1/7
9	2	2/7
8	3	3/7
8	4	4/7
6	5	5/7
3	6	6/7
2	7	7/7

$$\text{median} \leftarrow \frac{\int_{x_{1/2}}^{\infty} p(x)\,dx}{\int_{x_{min}}^{\infty} p(x)\,dx} = \frac{1}{2}$$

$$\frac{\int_{x_{1/2}}^{\infty} x\,p(x)\,dx}{\int_{x_{min}}^{\infty} x\,p(x)\,dx} = 2^{-\left(\frac{\alpha-2}{\alpha-1}\right)}$$

⇓

power-law is related to 80:20.

→ $k'_{max} \sim f(N)$

$$\int_{k_{max}}^{\infty} p(k)\,dk = 1/N$$ (only one node is there having deg k_{max})

$$c\int_{k_{max}}^{\infty} k^{-\alpha}\,dk = c\,\frac{k_{max}^{1-\alpha}}{\alpha-1} = k_{min}^{(\alpha-1)}\,k_{max}^{(1-\alpha)} = \frac{1}{N}$$

← $c = (\alpha-1)\,k_{min}^{(\alpha-1)}$

Date

Teacher's Signature :

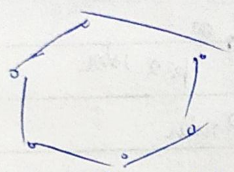
$\therefore k_{max} = k_{min} N^{1/(\alpha-1)}$

4/11 → 1. Lecture missing. - Bavit/Khush.

8/11 → I was absent — Fenil / Karan.

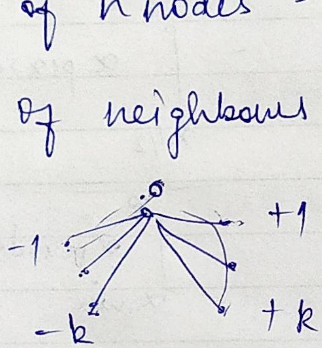
10/11. $C_{RG} = p = \frac{\langle k \rangle}{n-1}$.

Usefulness of useless research.



Ring graph of n nodes - deg $z = 2k$. (each side k-deg)

pairs of neighbours = $\binom{2k}{2}$.

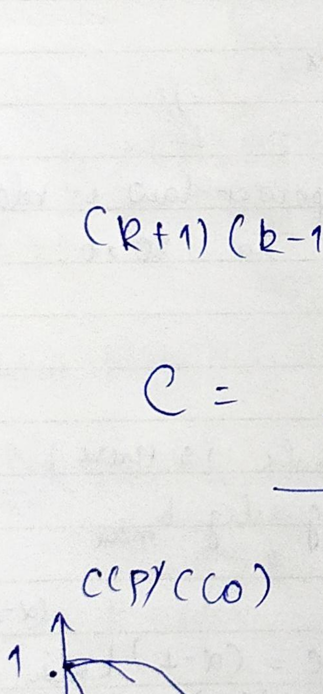


$(k+1)(k-1) + \Sigma(k-2)$ — # Triangles.

$$C = \frac{\frac{3k(k-1)}{2}}{\frac{2k(2k-1)}{2}} = \frac{3(k-1)}{2(2k-1)}$$

$$= \frac{3}{4}\left(\frac{z-2}{z-1}\right) \quad z = 2k.$$

$C(p)/C(0)$



1010

15/11 ⟵ 11/11 ⟶ Ashray's note-book [PIC in phone]

Data → Nwk → Nwk level stats → (3)

(1) Distributions, Averages, Density — why? Models — G(n,p), Watts Strogatz, Barabasi Albert

(2) Node level properties
Centrality Measures.

(3) Group of nodes.
cliques, k-plex, k-core, k-component.

Mixing: Assortative (Homophily), Disassortative

$$Q = \frac{m_c - \langle m_c \rangle}{m} \longrightarrow \text{Modularity}$$

$Q = 0$ — Random
$Q > 0$ — Assortative
$Q < 0$ — Disassortative

— m_c = number of links btw nodes that belong to the same class. C_i; $i = 1 \ldots k$

$$\therefore m_c = \frac{1}{2} \sum_{i,j} A_{ij}\, \delta_{c_i, c_j}$$

— $$\langle m_c \rangle = \frac{1}{2} \sum_{i,j} \left(\frac{k_i k_j}{2m} \right) \cdot \delta_{c_i, c_j}$$

$$\therefore Q = \frac{1}{2m} \sum_{i,j} \underbrace{\left(A_{ij} - \frac{k_i k_j}{2m} \right)}_{\text{Modularity Matrix}} \delta_{c_i, c_j} \quad \text{—— (1)}$$

* Practical scenario. * Microsoft Portal

Expt. No. Women → Check Page No.

		R_1	R_2
Men ↓	R_1	0.4	0.1
	R_2	0.2	0.3

e_r : Fraction of edges that link nodes of type r.

$$\delta_{c_i, c_j} = \sum_r \delta_{c_i, r} \cdot \delta_{c_j, r}$$

- $$\frac{1}{2m} \sum_{i,j} A_{ij} \delta_{c_i, c_j} = \frac{1}{2m} \sum A_{ij} \sum_r \delta_{c_i, r} \cdot \delta_{c_j, r}$$

$$= \sum_r \frac{1}{2m} \sum_{i,j} A_{ij} \cdot \delta_{c_i, r} \cdot \delta_{c_j, r}.$$

- fraction of edges that link to a node of type r.

$$\frac{1}{2m} \sum k_i \cdot \delta_{c_i, r}.$$

- Term (2) in Q (1).

$$\frac{1}{2m} \sum_{i,j} \frac{k_i k_j}{2m} \delta_{c_i, c_j} = \frac{1}{2m} \sum \frac{k_i k_j}{2m} \sum_r \delta_{c_i, r} \cdot \delta_{c_j, r}$$

$$= \sum_r \frac{1}{2m}$$

$$Q = \sum_r (e_r - a_r^2)$$

Date Teacher's Signature :

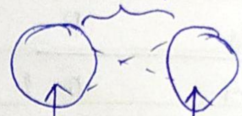
18/11

Community Detection

- 17/11 → Sir was lost on his part.

Graph Partitioning

of edges should be ↓



communities

Good partitioning.

N → N1, N2 } Breaking down to true communities

$$\# \text{ partitions} = \frac{N!}{N_1!\,N_2!} \sim e^{N}$$

Bisection.

→ Bell's number.

$$B(n+1) = \sum_k \binom{n}{k} B(k)$$

- Spectral methods are not covered.

k_i^{int} : # of links within the community for vertex i.

k_i^{ext} :

Strong community : $k_i^{int}(C) > k_i^{ext}(C)$.

for each individual vertex $\in C$.

Weak Community : $\sum_{i \in C} k_i^{int} > \sum_{i \in C} k_i^{ext}$.

- Mixing ⟶ Modularity.

$$Q = \frac{1}{2m} \left\{ A_{ij} - \frac{k_i k_j}{2m} \right\} \delta(C_i, C_j)$$

Random (~~Expected links~~).

Expt. No--------- Page No---------

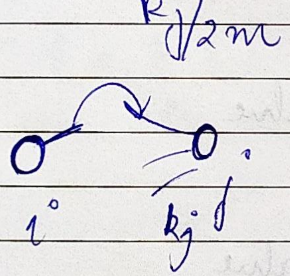
$$\frac{1}{2m} \sum_{ij} A_{ij}\, \delta(c_i, c_j)$$

$$= \frac{1}{2m} \sum_{k}^{n_k} \underbrace{\sum_{i,j \in c_k} A_{ij}}_{\text{looking at particular community } (2m_k)}$$

n_k → Total communities

k → communities.

$$= \sum_{k}^{n_k} \frac{m_k}{m}$$



⇒ for all k_i degrees → $k_i \cdot \frac{k_j}{2m}$.

$$\frac{k_i \cdot k_j}{2m} \longrightarrow \cancel{k} \; \frac{K_c^2}{2m}$$

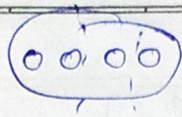
K_c → # total degree of nodes in comm. c

$$\therefore Q = \sum_{k=1}^{n_k} \left(\frac{m_c}{m} - \frac{K_c^2}{2m} \right)$$ → Bases of all community detection algorithm.

Modularity Maximization Method

Date ________________ Teacher's Signature : ________________

1) Agglomerative



- Girvan - Newman
Betweenness centrality . (Edges are ↑ important)

- single linkage (Louvain ?)

* 22/11 [Last lecture].

1) Agglomerative < single - linkage / complete - linkage / average linkage

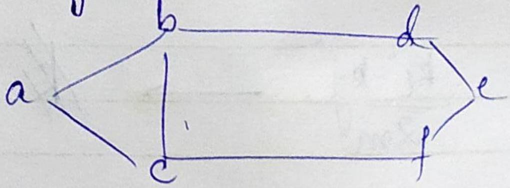
2) Louvain

3) Betweenness.

- single vs complete linkage
  ↓ minimal value — maximal value
  ↖ parameter ↗

- average linkage · average value

	a	b	c	d	e	f
a	0	1	1	2	3	2
b		0	1	1	2	2
c			0	2	2	1
d				0	1	2
e					0	1
f						0



① tie vertex - vertex pair

② complete linkage

pairing is done on closeness. (selecting to merge)

— T ①

btw groups < single / complete / avg. } Making table

– {a, b} c d e f

{a, b}	c	d	e	f
0	1	2	3	2
	0	2	2	1
		0	1	2
			0	1
				0

③ tie is btw cluster - vertex & vertex-vertex ⇒ choose cluster-vertex.

– {a, b, c} d e f

{a, b, c}	d	e	f
0	2	3	2
	0	1	2
		0	1
			0

– {a, b, c} {d, e} f

{a, b, c}	{d, e}	f
0	3	2
	0	3
		0

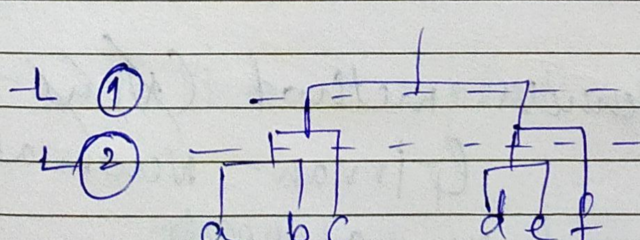
→ T(1) max value in comp.

④ tie is btw cluster - vertex & cluster - vertex.

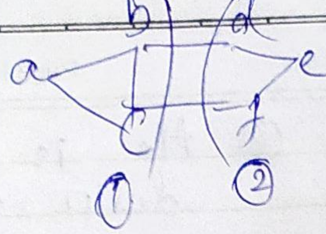
– avg values — $\frac{5}{4}$ & $\frac{3}{3}$

abcf   def

choosing minimal value.



← Dendogram.



	Community-1	Comm-2
	$m_c = 3$	$m_c = 2$
	$m = 7$	$m = 7$
	$k_c = 8$	$k_c = 6$

$$Q_1 = \frac{3}{7} - \left(\frac{8}{14}\right)^2 = \_\_\_\_$$

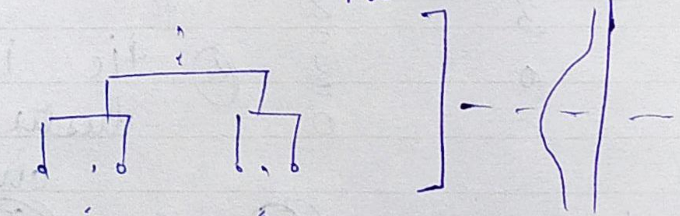
$$Q_2 = \frac{2}{7} - \left(\frac{6}{14}\right)^2 = \_\_\_\_$$

@ L_1

$Q = Q_1 + Q_2$

- same we could have done with the communities @ L_2.

$\Rightarrow \{ \; Q(L_1) \geq Q(L_2) \geq Q(L_3) \ldots \}$ here.



— max value of Q

- This is bottom-up approach.
- At that level make the partitions.

★ Edge-Betweenness method (Girvan-Newman)

- Top-down approach.

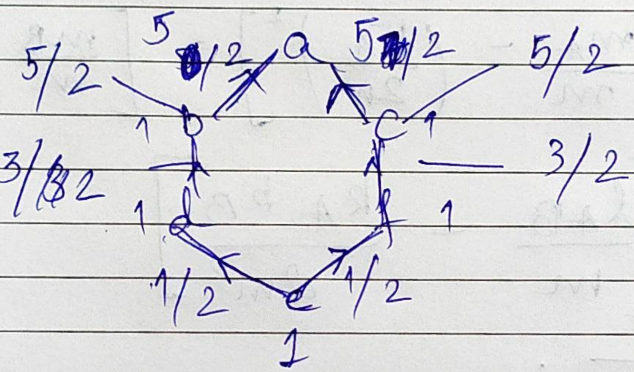
Expt. No.........

Page No.........

- Break the graph till single-node communities.

- Same example.

~~DF~~/ BFS Tree @ node-a



- Modifying the table in Newman's node betweenness

	{a, b}	{a, c}	{b, c}	{b, d}	{e, f}
→ a	2.5	2.5	0	1 / 1.5 .... all edges)	1.5

- Remove the edge whose column ~~sor~~ sum is ↑. (maximum)

- Then repeat the entri process in the remaining graph.

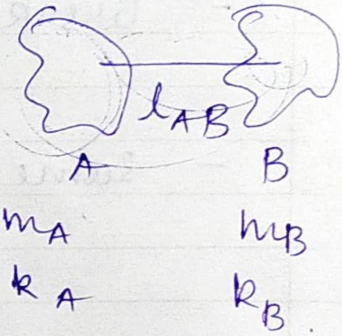
⇒ Too slow procedure, so may not be asked in the exam.

Date ....

Teacher's Signature : ....

* Louvain Algorithm (No exam questions)

<u>After pass -1</u>



$$\Delta Q = \frac{m_A + m_B + \cancel{m_e}\;^{l_{AB}}}{m} - \left(\frac{k_A + k_B}{2m}\right)^2 - \left[\frac{m_A}{m} - \left(\frac{k_A}{2m}\right)^2\right] - \left[\frac{m_B}{m} - \left(\frac{k_B}{2m}\right)^2\right].$$

$$= \left[\frac{l_{AB}}{m} - \frac{k_A k_B}{2m^2}\right]$$