Network science

Complex System :- Behind the complex system there is a "network", that defines the interactions between the components

Example :-

1. facebook
2. Structure of any organization
3. ~~Nuerons~~ Neurons of Human Brain
4. Business
5. Finance.
6. Internet
7. Human Genes

A Complex System is made up of many non-identical "elements" connected by the "interactions".

Role of Network
We will Never understand complex system unless we map out and understand the network behind them.

Network used in :- The architecture of network emerging in various domains of science, nature & technology.

Networks & Graphs

Components :- nodes, vertices
interactions :- links, edges.
System :- Network, Graph.

Example :- (i) www
(ii) Social Network
(iii) Metabolic Network.

} Network

Language (Network, ~~link~~ node, link)

~~graph~~ graph :- Mathematical representation of network :-
(i) web graph
(ii) Social graph

Language (~~Graph~~ Graph, vertex, edge)

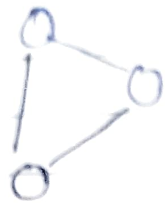
Degree Distribution :-

Average Degree :-

$$\langle K \rangle = \frac{1}{N} \sum_{i=1}^{N} K_i = \frac{2L}{N} \text{ (for undirected)}$$

where : K_i = i^{th} link.

Example



undirected

average Degree :-

$\langle K \rangle = \frac{1}{3}(2+2+2) = 3$

$= \frac{1}{3}(1+2+1) = \frac{4}{3}$

for Directed :-

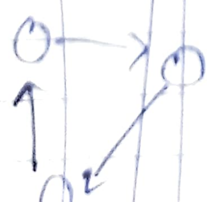
$$\langle K \rangle = \frac{1}{N}\sum_{i=1}^{N} K^{\text{in degree}} + \frac{1}{N}\sum_{i=1}^{N} K^{\text{out degree}}$$

Because $\langle K^{in} \rangle = \langle K^{out} \rangle$

the

$$\langle K \rangle = \frac{L}{N}$$

Example :-



$\langle K^{in} \rangle = \frac{1+1+1}{3} = 1$

$\langle K^{out} \rangle = \frac{1+1+1}{3} = 1$

~~total $\langle K \rangle$~~ =

average degree = 2

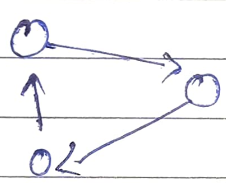
For Directed graph.

Average Digree = $\langle k^{in} \rangle$ or $\langle k^{out} \rangle$

Because: $\langle k^{in} \rangle = \langle k^{out} \rangle$

$$\boxed{\langle k \rangle = \frac{L}{N}}$$

Example



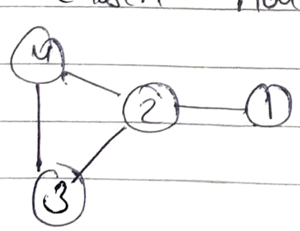
$\langle k^{in} \rangle = \frac{3}{3}$

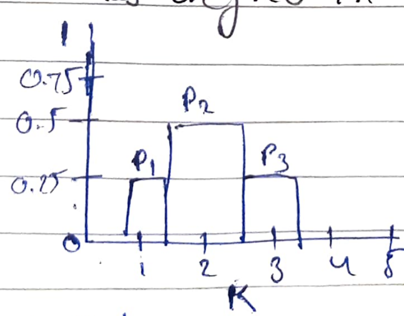
$\langle k^{out} \rangle = \frac{3}{3}$

$$\boxed{\langle k \rangle = 1}$$

Digree Distribution :-

P(K) :- Probablity that a randomly chosen node has degree K.





$P(K) = \frac{N_K}{N}$ → No. of Node with digree K

$P(1) = \frac{1}{4} = 0.25$

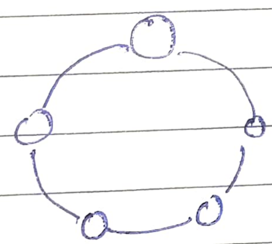
$P(2) = \frac{2}{4} = 0.5$

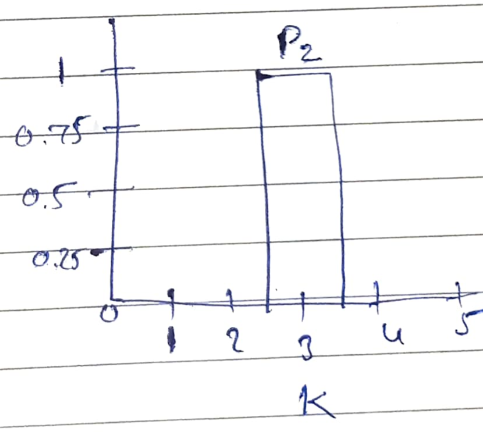
$P(3) = \frac{1}{4} = 0.25$

$P(4) = 0$

$P(5) = 0$

Example 2





$P(1) = 0$

$P(2) = \frac{5}{5} = 1$

$P(3) = 0$

$P(4) = 0$

$P(5) = 0$

Discrete Representation :-

P_K :- The Probablity that the node Have degree 'K'

$$\boxed{P_K = \frac{N_K}{N}}$$

Continous Representation :-

$$\boxed{\cancel{P(K)} = \int_{K_1}^{K_2} p(K)\, dK} \quad \text{——} \quad (1)$$

P(K) is the pdf of the digree

& the equation represents the Probablity that a node's degree is between K_1 and K_2

Normalized Condition :-

$$\boxed{\sum_{0}^{\infty} P_K = 1} \quad \text{or} \quad \boxed{\int_{K_{min}}^{\infty} P(K)\, dK = 1}$$

Where K_{min} is the minimal degree in the network.

In our case it is '0'

& K can varis to ∞

Undirected Network Example:-

1. Actor network
2. Protein interactions.

Directed Network Example:-

1. URLs on the www
2. Phone calls
3. Metabolic reactions

Representation :-

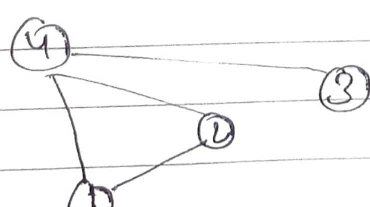
Adjacency Matrix.

$A_{ij} = 1$ if there is a link between i & j

$A_{ij} = 0$ if there is No link between i & j

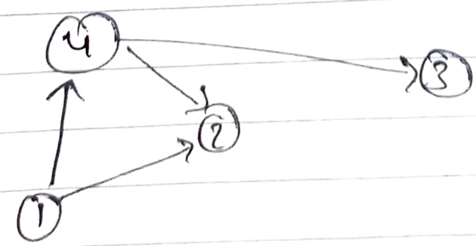
Example :-

Undirected :-



$$A_{ij} = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & 1 & 0 & 1 \\ 2 & 1 & 0 & 0 & 1 \\ 3 & 0 & 0 & 0 & 1 \\ 4 & 1 & 1 & 1 & 0 \end{array}$$

Directed :-



$$A_{ij} = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & 1 & 0 & 1 \\ 2 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 4 & 0 & 1 & 1 & 0 \end{array}$$

No. of links (L) =

$$L = \frac{1}{2} \sum_{i,j}^{N} A_{ij} \quad \text{(Undirected graph)}$$

$$L = \sum_{ij}^{N} A_{ij} \quad \text{(Directed graph)}$$

Max Links in a Network :-

$$\boxed{L_{max} = {}^{N}C_{2} = \frac{(N)(N-1)}{2}} \quad , \quad \left\{ \langle k \rangle = \frac{2L}{N} \right.$$

A graph with degree $L = L_{max}$ is a complete graph.

& Its average degree $\boxed{\langle k \rangle = N-1}$

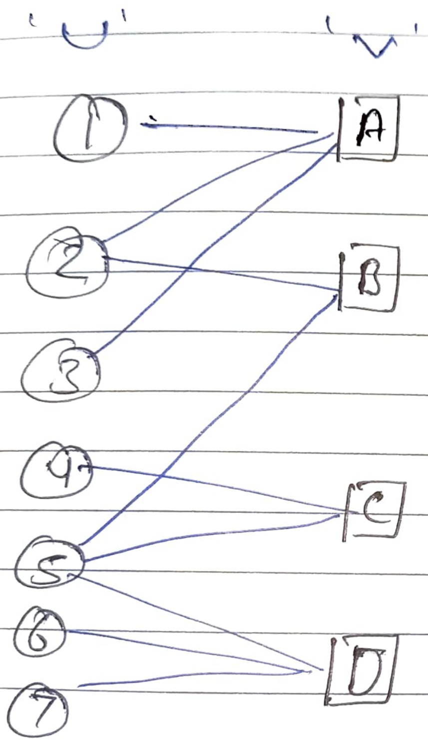
Most of the network has

$$\boxed{L \ll L_{max}}$$

&

$$\boxed{\langle k \rangle \ll (n-1).}$$

Bipartite graph:- It is a graph whose nodes can be devided into two disjoint sets 'U' & 'V' such that every link in a graph connects nodes in the set 'U' to the nodes in the set 'V'.



① Disease Network
② Collaboration Network.

A Path is a sequence of nodes in which each node is adjacent to next one.

Path

~~Path~~:- The distance between two nodes is defined as the number of edges along the sortest path connecting them.

In directed graphs each path needs to follow the direction of arrows.
Thus in a digraph the distance from node A to B is generally different from the node B to A

Number of Paths between any two node i & j

Denoted By $N_{ij}^{(L)}$

↓

Number of Path of length L from i to j.

$N_{ij}^{(1)}$ = If there is a link between i & j then $\boxed{A_{ij} = 1}$ else $\boxed{A_{ij} = 0}$

$N_{ij}^{(2)}$ = If there is a path of length two between i & j then

$\boxed{A_{ik} A_{kj} = 1}$ else $\boxed{A_{ik} A_{kj} = 0}$

Then the Number of path of ~~l~~ length 2

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik} A_{kj} = [A^2]_{ij}$$

$N^{(n)}$:- The Number of path of length 'n' from i to j

If there is a Path of length 'n' b/w ~~from~~ i & j then $A_{ik} \ldots A_{lj} = 1$

otherwise

$$\boxed{A_{ik} \cdots A_{lj} = 0}$$

then, $$\boxed{N_{ij}^{(n)} = [A^n]_{ij}}$$

Find Distance :- (Use BFS. Algorithm)

Network Diameter And Average Distance :-

Diameter (d_{max}) :- The maximum distance between any pair of nodes in the graph.

Average Path length / distance $\langle d \rangle$:- for connected graph :-

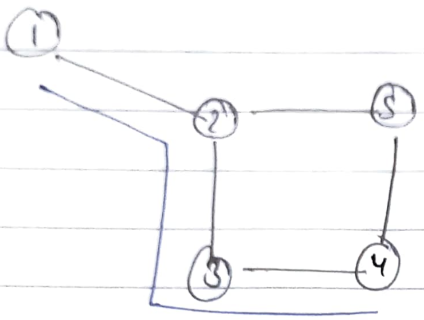
$$\boxed{\langle d \rangle = \frac{1}{2L_{max}} \sum_{i,j \neq i} d_{ij}}$$

In an undirected graph $d_{ij} = d_{ji}$, so we only need to count them once.

$$\boxed{\langle d \rangle = \frac{1}{L_{max}} \sum_{i,j>i} d_{ij}}$$

Example :-

Diameter :-



The longest Path in a graph

Average Path length :-

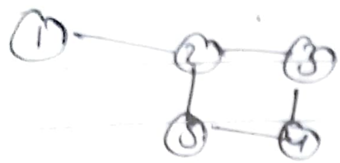
$$\frac{\text{Sum of length of } \cancel{\text{each}} \text{ path b/w each pair of node}}{\text{total no. of Path, b/w each pair of Nodes}}$$

$$\frac{[d_{1\to2} + d_{1\to3} + d_{1\to4} + d_{1\to5} + d_{2\to3} + d_{2\to4} + d_{2\to5} + d_{3\to4} + d_{3\to5} + d_{4\to5}]}{10}$$

$$\Rightarrow \frac{1+2+3+2+1+2+1+1+2+1}{10}$$

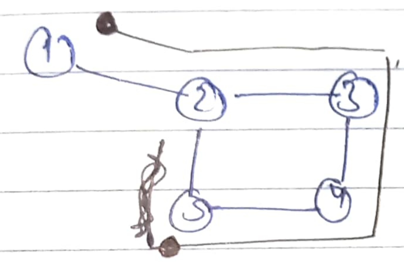
$$\Rightarrow \frac{16}{10} = 1.6$$

Cyclic Path:- A path with same start & end node

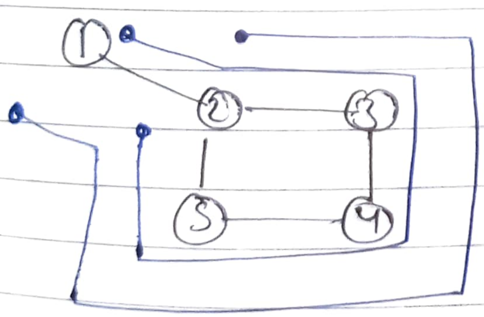


2, 3, 4, 5 cycle

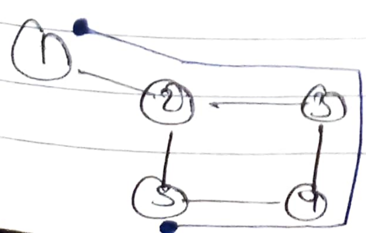
Self-avoiding path:- A path that does not intersect itself.

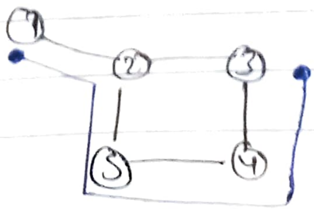


Eulerian path:- A Path that traverse each link exactly once.

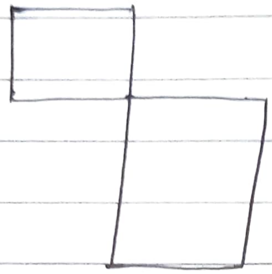


Hamiltonian Path:- A path that visits each node exactly once





If the adjcency list is like



then there is a connected component

Strongly connected Component :-

A graph has a node from each node to every other node & vice versa.

Clustering Coefficient

This clustering cofficient gives what fraction of your neighbors are connected.

Denoted by

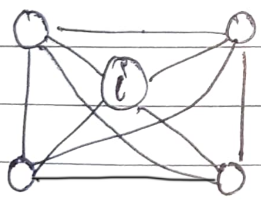
$$\boxed{C_i = \frac{2e_i}{K_i(K_i-1)}}$$

$C_i \in \{0,1\}$

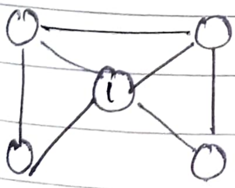
e_i = no. of edges between the neighbour of node i

K_i = degree of node i

Example :-

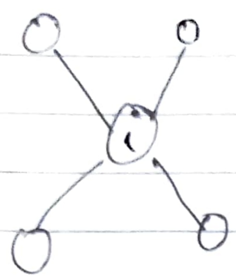


$$C_i = \frac{2 \times 6}{4 \times 3} = 1$$

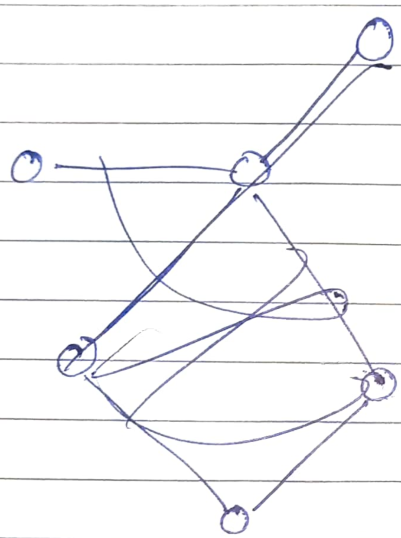


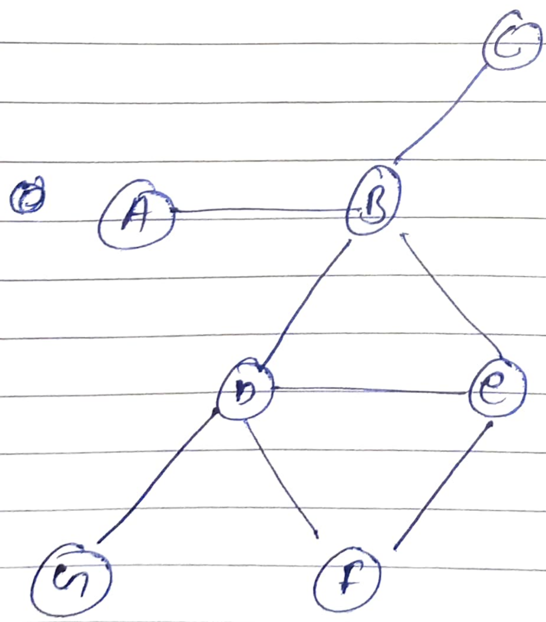
$$C_i = \frac{2 \times 3}{4 \times 3} = \frac{1}{2}$$

1



$$C_i = \frac{2\times 0}{4\times 3} = 0$$





Node	C_i
A	$\frac{2\times 0}{1\times 0} = 0$
B	$\frac{2\times 1}{4\times 3} = \frac{1}{6}$
c	$\frac{2\times 0}{1\times 0} = 0$
D	$\frac{2\times 2}{4\times 3} = \frac{1}{3}$
e	$\frac{2\times 2}{3\times 2} = \frac{2}{3}$
G	$\frac{2\times 0}{1\times 0} = 0$
f	$\frac{2\times 1}{2\times 1} = 1$

Random Network model

Definition :- A random graph is a graph of N nodes where each pair of nodes is connected by Probablity p.

$P(L)$:- The Probablity to have exactly L Links in a network of N nodes and probability P :

Binomial Distribution

$$\boxed{P(L) = \binom{L_{max}}{L} p^L (1-p)^{L_{max}-L}}$$

• $L_{max} = \frac{n(n-1)}{2}$

$$\boxed{P(L) = \binom{\frac{n(n-1)}{2}}{L} p^L (1-p)^{\frac{n(n-1)}{2}-L}}$$

No. of different ways we can choose L links among all the links.

Binomial Distribution:-

$P(x) = {}^{N}C_{x}\, p^{x}(1-p)^{N-x}$ (Probability p)

first Moment mean $\langle x \rangle = NP$

Second Moment $\langle x^2 \rangle = P(1-P)N + P^2N^2$

Variance = $\underbrace{E[x^2]}_{\text{Second moment}} - (\underbrace{E[x]}_{\text{first Moment}})^2$

$= P(1-P)N + \cancel{P^2N^2} - \cancel{N^2P^2}$

$= \cancel{PN - P^2N + P^2N^2 - NP}$

$\boxed{\sigma_x^2 = P(1-P)N} \Rightarrow NPq$

$\boxed{\sigma_x = [P(1-P)N]^{1/2}} \Rightarrow \sqrt{NPq}$

$P(L)$: Probablity that the network has L links

$$\boxed{P(L) = \binom{\frac{n(n-1)}{2}}{L} p^L (1-P)^{\frac{n(n-1)}{2} - L}}$$

Average number of Links $\langle L \rangle$ in a random network is given by

$$\boxed{\langle L \rangle = \sum_{L=0}^{\frac{n(n-1)}{2}} L \cdot P(L) = P \cdot \frac{N(N-1)}{2}}$$

Average digree $\langle K \rangle$ in random graph

$$\langle K \rangle = \frac{2L}{N} = \frac{\cancel{2}}{N} \times \frac{P \times N(N-1)}{\cancel{2}}$$

$$\boxed{\langle K \rangle = P(N-1)}$$

σ^2 (variance) $= NPq$

$$= \frac{n(n-1)}{2} \times P(1-P)$$

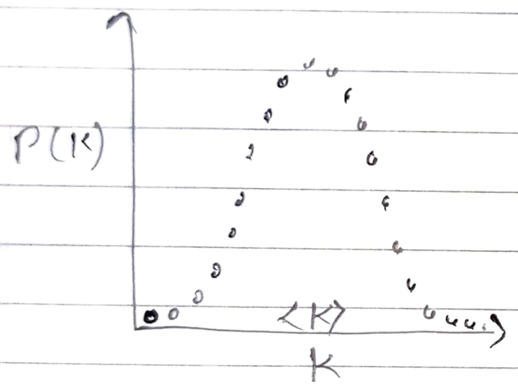
$$\boxed{\sigma^2 = P(1-P) \frac{N(N-1)}{2}}$$

In random graph the average link is

$$P(N-1)$$

But in Normal graph it is $(N-1)$

Digree Distribution



$P(K)$ = Probablity ~~the~~ ~~of~~ that the random graph has Nodes with digree K

$$\boxed{P(K) = {}^{N-1}C_K \; p^K \, (1-p)^{(N-1)-K}}$$

Selecting K nodes from N-1

Probablity of Having k edges

Probablity of missing N-1-K edges.

$\langle K \rangle = P(N-1)$ & $\quad \sigma_K^2 = P(1-P)(N-1)$

$$\frac{\sigma_K}{\langle K \rangle} = \sqrt{\frac{\sigma_K^2}{\langle K \rangle^2}} = \left[\frac{P(1-P)\cancel{(N-1)}}{P^2(N-1)^2}\right]^{1/2}$$

$$\frac{\sigma_K}{\langle K \rangle} = \left[\frac{(1-P)}{P} \times \frac{1}{N-1}\right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

for $n \rightarrow \infty$
$\sigma K \rightarrow 0$

then the Binomial Distribution did Not work

So we have to convert the Binomial Distribution to poision distribution.

$$P(K) = {}^{N-1}C_K \; P^K (1-P)^{N-1-K}$$

Solving each part saperatly

$${}^{N-1}C_K = \frac{(N-1)!}{K!\,(N-1-K)!}$$

$$\Rightarrow = \frac{(N-1-1)(N-1-2)(N-1-3)\cdots(N-1-K+1)\,\cancel{(N-1-K)!}}{K!\,\cancel{(N-1-K)!}}$$

$$= \frac{(N-1-1)(N-1-2)(N-1-3)\cdots\cdots(N-1-(K+}{K!}$$

$$= \frac{(N-1)^K}{K!}\left[\underbrace{\left(1-\frac{1}{N-1}\right)\left(1-\frac{2}{N-1}\right)\cdots\cdots\left(1-\frac{K+1}{N-1}\right)}_{\approx 1}\right]$$

$\because \quad N \to \infty$

$$\boxed{\binom{N-1}{K} = \frac{(N-1)^K}{K!}}$$

Now solving :-

$= (1-P)^{N-1-K}$

~~take log use~~

we can write above expression as

$$= e^{\log_e (1-p)^{N-1-K}}$$

$$= e^{N-1-K \log(1-p)}$$

$$= \cancel{e^{N-1-K}}$$

$$\left\{ \because \log(1-p) = -\left[p + \frac{p^2}{2} + \frac{p^3}{3} \cdots \infty\right] \right.$$

$$= e^{-(N-1-K)\left[p + \frac{p^2}{2} + \frac{p^3}{3} \cdots \infty\right]}$$

$$= e^{-(N-1-K)\, p \underbrace{\left[1 + \frac{p}{2} + \frac{p^2}{3} \cdots \infty\right]}_{\approx 1}}$$

$$\left\{ \because p \to 0 \right.$$

$$\boxed{(1-p)^{N-1-K} \cong e^{-(N-1-K)\,p}}$$

$$\left\{ \begin{aligned} &\because \langle K \rangle = p(N-1) \\ &p = \frac{\langle K \rangle}{(N-1)}. \end{aligned} \right.$$

$$= e^{-(N-1-K)P}$$

$$= e^{-(N-1-K)\frac{\langle K \rangle}{N-1}} \Rightarrow e^{-(\cancel{N-1})\left(1-\frac{1}{N-1}\right)\frac{\langle K \rangle}{\cancel{N-1}}}$$

$$= e^{-\left(1-\frac{K}{N-1}\right)\langle K \rangle}$$

$N \simeq \infty$

so $\left(1-\frac{K}{N-1}\right) \simeq 1$

$$= e^{-\langle K \rangle}$$

Pull all values

$$= \frac{(N-1)^K}{\underline{K}} \times p^K e^{-\langle K \rangle}$$

$$p^K = \frac{\langle K \rangle}{N-1}$$

$$= \frac{\cancel{(N-1)^K}}{\underline{K}} \times \frac{\langle K \rangle^K}{\cancel{(N-1)^K}} \quad e^{-\langle K \rangle}$$

$$= \frac{\langle K \rangle^K e^{-\langle K \rangle}}{\underline{K}}$$

$$P(K) = e^{-\langle K \rangle} \frac{\langle K \rangle^K}{\underline{K}}$$

↓

Probability k^{th} degree in any random graph from poison distribution

for Poission distribution

$$\boxed{P(K) \cong e^{-\langle K\rangle} \frac{\langle K\rangle^{K}}{\lfloor K}}$$

Distance in a random graph :-

$\langle K\rangle^{1}$ = Nodes at distance one $(d=1)$
$\langle K\rangle^{2}$ = Nodes at distance two $(d=2)$
$\langle K\rangle^{3}$ = Nodes at distance three $(d=3)$
⋮
$\langle K\rangle^{d}$ = Nodes at distance d

$$N = 1 + \langle K\rangle + \langle K\rangle^{2} + \langle K\rangle^{3} \ldots \langle K\rangle^{d_{max}}$$

$$N = \frac{\langle K\rangle^{d_{max}+1} - 1}{\langle K\rangle - 1} = \frac{r^{n} - 1}{r - 1}$$

$$N \cong \langle K\rangle^{d_{max}} \Rightarrow N = \langle K\rangle^{d_{max}}$$

log on both sides

$$\log N = d_{max} \cdot \log\langle K\rangle$$

$$\boxed{d_{max} = \frac{\log N}{\log\langle K\rangle}}$$

or

$$\boxed{\langle d \rangle = \frac{\log N}{\log \langle K \rangle}}$$

Average distance b/w 2 randomly chosen nodes.

$\frac{1}{\log \langle K \rangle}$ implies that denser the network, the smaller the the distance b/w the nodes

Clustering Cofficient for random graph:-

$$\boxed{C_i = \frac{2\langle L \rangle}{K_i(K_i-1)} = p = \frac{\langle K \rangle}{N}}$$

C decreases with the System size N.

Average path length for random graph

$$\boxed{\langle d_{rand} \rangle = \frac{\log N}{\log \langle K \rangle}}$$

Clustering Coefficient for random graph

$$\boxed{C_i = \frac{2\langle L \rangle}{K(K-1)} = p = \frac{\langle K \rangle}{N}}$$

Degree Distribution for random graph :-

$$P(K) = e^{-\langle K \rangle} \frac{\langle K \rangle^{K}}{\lfloor K}.$$

L-4

Scale free Network

Discrete formalism:-

As node degrees are always positive integers the discrete formalism captures the probability that a node has exactly K links.

$$\boxed{P_K = CK^{-\gamma}}$$

$$\boxed{\sum_{K=1}^{\infty} P_K = 1}$$

$$\sum_{K=1}^{\infty} CK^{-\gamma} = 1$$

$$C \sum_{K=1}^{\infty} K^{-\gamma} = 1$$

$$C = \frac{1}{\sum_{K=1}^{\infty} K^{-\gamma}}$$

$$\boxed{C = \frac{1}{\zeta(\gamma)}} \qquad C = \frac{1}{\zeta(\gamma)}$$

Put in above equation

$$\boxed{P_k = \frac{k^{-\gamma}}{\zeta(\gamma)}}$$

Continous formulation:- In analytical calculations it is often convenient to assume that the degrees can take up any positive real values.

$$\boxed{P_k = C k^{-\gamma}}$$

$$\int_{K_{min}}^{\infty} P(K)\, dk = 1$$

$$\int_{K_{min}}^{\infty} C K^{-\gamma} = 1$$

$$C \int_{K_{min}}^{\infty} K^{-\gamma} = 1$$

$$C \left[\frac{K^{-\gamma+1}}{-\gamma+1} \right]_{K_{min}}^{\infty} = 1$$

$$\frac{C}{1-\gamma} \left[\frac{1}{K^{\gamma-1}} \right]_{K_{min}}^{\infty} = 1$$

$$\frac{C}{-\gamma+1}\left[\frac{1}{\infty} - \frac{1}{K_{min}^{\gamma-1}}\right] = 1$$

$$\frac{1}{\infty} = 0$$

$$\frac{\cancel{-}C}{\cancel{-}(\gamma-1)}\left[\frac{1}{K_{min}^{\gamma-1}}\right] = 1$$

$$C = (\gamma-1) \times K_{min}^{\gamma-1}$$

$$P(K) = (\gamma-1) \times K_{min}^{\gamma-1} K^{-\gamma}$$

Expected Maximum degree = K_{max}

Estimating K_{max}

$$\int_{K_{max}}^{\infty} P(K)\, dK \approx \frac{1}{N}$$

$\because$ The probablity to have a node larger then K_{max} should not exceeds by $\frac{1}{N}$

we have to prove above equation. Now we are using the $P(K)$:- Probablity of node having k links

$$= \int_{K_{max}}^{\infty} (\gamma-1)\, K_{min}^{\gamma-1}\, K^{-\gamma}\, dK$$

$$= (\gamma-1)\, K_{min}^{\gamma-1} \int_{K_{max}}^{\infty} K^{-\gamma}\, dK$$

$$= \cancel{(\gamma-1)}\, K_{min}^{\gamma-1} \left[\frac{K^{-\gamma+1}}{\cancel{-\gamma+1}} \right]_{K_{max}}^{\infty}$$

$$\Rightarrow -K_{min}^{\gamma-1} \times \left[\frac{1}{K^{\gamma-1}} \right]_{K_{max}}^{\infty}$$

$$-k_{min}^{\gamma-1} \propto \left[\frac{1}{\infty} - \frac{1}{k_{max}^{\gamma-1}}\right]$$

$$\frac{k_{min}^{\gamma-1}}{k_{max}^{\gamma-1}} \approx \frac{1}{N}$$

$$\boxed{k_{max} = k_{min} \cdot N^{\frac{1}{\gamma-1}}}$$

Scale-free Network (Definition):-

Network with Power law tail in their degree distribution are called Scale-free-networks.

Divergences in Scale free network $\langle k^m \rangle$

we know $P(k) = ck^{-\gamma}$

where $k = [k_{min}, \infty)$

$$\int_{k_{min}}^{\infty} P(k)\,dk = 1$$

$$\int_{k_{min}}^{\infty} ck^{-\gamma}\,dk = 1$$

$$\boxed{c = (\gamma - 1)\,k_{min}^{\gamma-1}}$$

then

$$\boxed{P(k) = (\gamma - 1)\,k_{min}^{\gamma-1}\,k^{-\gamma}}$$

Divergence.

$$\langle K^m \rangle = \int_{K_{min}}^{\infty} K^m P(K)\, dK$$

$$= \int_{K_{min}}^{\infty} K^m (\gamma-1) K_{min}^{\gamma-1} K^{-\gamma}\, dK$$

$$= (\gamma-1) K_{min}^{\gamma-1} \int_{K_{min}}^{\infty} K^{m-\gamma}\, dK$$

$$= \cancel{(\gamma-1) K_{min}^{\gamma-1} \int_{K_{min}}^{\infty}}$$

$$= (\gamma-1) K_{min}^{\gamma-1} \left[\frac{K^{m-\gamma+1}}{m-\gamma+1} \right]_{K_{min}}^{\infty}$$

$$= \frac{(\gamma-1) K_{min}^{\gamma-1}}{m-\gamma+1} \left[\frac{1}{K^{\gamma-m-1}} \right]_{K_{min}}^{\infty}$$

$$= \frac{(\gamma-1) K_{min}^{\gamma-1}}{m-\gamma+1} \left[\frac{1}{\infty} - \frac{1}{K_{min}^{\gamma-m-1}} \right]$$

$$= \frac{(\gamma-1)\, k_{min}^{\gamma-1} \cdot k_{min}^{m+1-\gamma}}{m-\gamma+1}$$

$$= \frac{(\gamma-1)\, k_{min}^{\cancel{\gamma}-\cancel{\gamma}+m+\cancel{1}-\cancel{\gamma}}}{m-\gamma+1}$$

$$\boxed{\langle k_0^{m} \rangle = -\frac{(\gamma-1)\, k_{min}^{m}}{m-\gamma+1}}$$

for $m-\gamma+1$ means that all moment has integral diverges

Distance in a random graph.

Random graphs tend to have a tree-like topology with almost constant node degrees.

no. of first neighbours $= \langle k \rangle$
no. of second neighbours $= \langle k \rangle^2$
no. of neighbour with distance $d = \langle k \rangle^d$

estimate maximum distance =

$$1 + \sum_{d=1}^{d_{max}} \langle k \rangle^d = N$$

$$1 + \langle k \rangle + \langle k \rangle^2 + \langle k \rangle^3 \ldots \langle k \rangle^{d_{max}} = N$$

$$\frac{\langle k \rangle^{d_{max}+1} - 1}{\langle k \rangle - 1} = N$$

$$\cong \langle k \rangle^{d_{max}} = N$$

$$d_{max} \cdot \log \langle k \rangle = \log N$$

$$\boxed{d_{max} = \frac{\log N}{\log \langle k \rangle}}$$

Small world behavior in scale free Network

$$\langle d \rangle = \quad k_{max} = k_{min} N^{\frac{1}{\gamma - 1}}$$

$$\langle d \rangle \begin{cases} \text{const} , & \gamma = 2 \\ \dfrac{\log \log N}{\log(\gamma - 1)} , & 2 < \gamma < 3 \\ \dfrac{\log N}{\log \log N} , & \gamma = 3 \\ \log N , & \gamma > 3 \end{cases}$$

generating network with predefined p(k)

$$p_{ij} = \frac{k_i k_j}{2L-1}$$

Hidden parameter model

$$p(\eta_i, \eta_j) = \frac{\eta_i \eta_j}{\langle \eta \rangle N}$$

$$P(k) = \int \frac{e^{-\eta} \eta^k}{k!} \rho(\eta)\, d\eta$$

$$P(k) = \frac{1}{N} \sum_j \frac{e^{-\eta_j} \eta_j^k}{k!} \quad \left(\eta_j = \frac{c}{i^\alpha} \right)$$

$$i = 1, 2 \ldots N$$

$$P(k) \sim k^{-\left(1+\frac{1}{\alpha}\right)}$$

Lecture - 5

BA Model

ER (Erdos & Renyi) model :- (drawback of ER model)

① This is a random network with fixed Number of Nodes (Static model)

② It Can Not be used to represent real network.

Because Networks expand through the addition of new nodes.

③ In this model links are added randomly to the network.

While in the real network new node Prefer to connect to the more connected nodes.

Growth & Preferential attachment :-

The random network model differs from real networks in two importent characterstics :-

① Growth :- The random networ assume that the number of nodes is fixed. While in the real network the nodes are continously increasing.

② Preferential Attachment :- In a random netwo[rk] ~~nod~~ a new node connects to or prefer to connect other random node in the network.

But in the real network The new node ~~or~~ Prefer to connect ~~to the whith~~ with more connected nodes.

Because all above requirements are Not fullfill by the rendom network so we Prefer to Barabasi - Albert model (BA-model)

Digree Distribution of this model :-

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

$$\# \quad \because \overset{\text{node}}{\frac{\partial k_i}{\partial t}} \propto \overset{\text{degree}}{\pi(k_i)}$$

As a no. of links (degree of node) ki change with the time the probability of node ~~haven~~ having degree ki $\pi(k_i)$ will be proprtionally changing.

$$\frac{\partial k_i}{\partial t} = A \frac{k_i}{\sum_j k_j}$$

use

$$\sum_j k_j = 2mt$$

$$\frac{\partial k_i}{\partial t} = A \frac{k_i}{2mt}$$

$$\boxed{\because A = m}$$

$$\frac{\partial k_i}{\partial t} = m \frac{k_i}{2\cancel{m}t}$$

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}$$

$$\frac{\partial k_i}{k_i} = \frac{\partial t}{2t}$$

$$\int_m^K \frac{\partial k_i}{k_i} = \frac{1}{2}\int_{t_i}^{t} \frac{1}{t}\, dt$$

$$\left[\log k_i\right]_m^k = \frac{1}{2}\left[\log t\right]_{t_i}^t$$

$$\log \frac{k}{m} = \frac{1}{2}\log\left(\frac{t}{t_i}\right)$$

$$\log \frac{k}{m} = \log\left(\frac{t}{t_i}\right)^{1/2}$$

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \boxed{\beta = \frac{1}{2}}$$

Degree Distribution :-

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}$$

$$\boxed{\beta = \frac{1}{2}}$$

$$P(k) = 1 - \frac{m^{1/\beta}\, t}{k^{1/\beta}\,(t+m_0)}$$

P(K) Probability that node has K links

derivate w.r.t K

$$\frac{d}{dK} P(K) = \frac{d}{dK}\left(1 - \frac{m^2 t}{K^2(t+m_0)}\right)$$

$$P(K) = \frac{d}{dK} - \frac{m^2 t}{K^2(t+m_0)} \frac{d}{dK}$$

$$P(K) = 0 - \left(\frac{m^2 t}{(t+m_0)}\right)\frac{d}{dK} K^{-2} \quad \left[\frac{m^2 t}{(t+m_0)} = \text{constant}\right]$$

$$P(K) = -\left(\frac{m^2 t}{(t+m_0)}\right) - 2K^{-3}$$

$$P(K) = \frac{2m^2 t}{K^3(t+m_0)} \qquad t = 1$$

$$P(K) = \frac{2m^2}{K^3(1+m_0)} \qquad \frac{2m^2}{1+m_0} = \text{Constant}$$

$$P(K) = K^{-3}$$

$$P(K) \simeq K^{-3}$$

$$P(K) \cong K^{-\gamma} \qquad \gamma = -3$$

Digree Distribution Rate Equation :-

$$\langle N(k,t) \rangle = t\, P(k,t)$$

$\downarrow$

Number of node with digree K at time t.

At each timeStamp t we add one new node.

So

$$\boxed{N = t}$$

(total number of node = number of timestamps)

$$\boxed{\Pi(k) = \frac{k}{\sum_j k_j} = \frac{k}{2mt}}$$

2m = each node add m link
But the link is Bidirection
So 1 ~~libne~~ link has ~~digree~~
contribute in 2 digree.

t = timestamp t

Number of link added to degree K nodes after the arrival of a new node :-

Prefrental Attachmend $\pi(k)$ X Number of Nodes with digree K

X New node with m new link

$$= \frac{k}{2\cancel{m}\cancel{t}} \times \cancel{t}\, p(K,t) \times \cancel{m}$$

$$\boxed{= \frac{k}{2} p(K,t).}$$

Because the mean field theory offers the correct scaling, But it provides the wrong cofficient of the degree distribution

So To fix it, we need to calculate P(K) exactly, which we will ~~did~~ ~~next~~ using a rate equation based approch.

So

No. of digree k nodes that acquire a new link , become digree K+1

$$= \frac{K}{2} p(K,t)$$

Similarly :-

No. of digree K-1 nodes that acquire a new link become degee K

$$= \frac{K-1}{2} p(K-1, t)$$

So.

No. of K-nodes at time t+1 =

No. of digree K nodes at time t

+

No. of digree K nodes at which become digree K from K-1

− No of digree K nodes which become digree K+1 from K.

$$(N+1)P(k, t+1) = NP(k,t) + \frac{k-1}{2} P(k-1, t) - \frac{k}{2} P(k,t)$$

Now we assuming that all node have digree 'm'.

And a new node came with the digree 'm'

then from Previous equation.

$$(N+1)\, P(m, t+1) = NP(m,t) + 1 - \frac{m}{2} P(m,t)$$

$(N+1)\,P(m,t+1)$: No. of m digree node at time t+1

$NP(m,t)$: No. of m digree node at time t

1: ~~No of~~ Add one m-digree node

$\frac{m}{2}P(m,t)$: Loss of an m-node which become m+1 from m

Converting Rate equation to "stationary state":-

Rate equation:-

$$(N+1)\, P(K, t+1) = N P(K,t) + \frac{K-1}{2} P(K-1,t) - \frac{K}{2} P(K,t)$$

Stationary state:- The digree distribution $P(K,t)$ describes the probablity that a node in the network has K connection (digree K) at time t.

The system evolves over time as nodes and connections are added and we aim to derive steady-state distribution:-

~~P(K)~~ $\boxed{P(K,t) = P(K)}$ or $\boxed{P(K,\infty) = P(K)}$

where $t \to \infty$

Now from above equation

$$(N+1)\, P(K, t+1) = N P(K,t) + \frac{K-1}{2} P(K-1,t) - \frac{K}{2} P(K,t)$$

for steady-state distribution $t \to \infty$

$$(N+1)\, P(K,\infty) = \cancel{N+1}\; N P(K,\infty) + \frac{K-1}{2} P(K-1,\infty) - \frac{K}{2} P(K,\infty)$$

$$\left\{ \because \; P(K,\infty) = P(K) \right\}$$

so $$(N+1)\,P(K) = N\,P(K) + \frac{K-1}{2}\,P(K-1) - \frac{K}{2}\,P(K)$$

$$\cancel{N\,P(K)} + P(K) = \cancel{N\,P(K)} + \frac{K-1}{2}\,P(K-1) - \frac{K}{2}\,P(K)$$

$$\cancel{P(K) =} \quad P$$

$$P(K) = \frac{K-1}{2}\,P(K-1) - \frac{K}{2}\,P(K)$$

$$P(K) + \frac{K}{2}\,P(K) = \frac{K-1}{2}\,P(K-1)$$

$$\frac{(K+2)\,P(K)}{\cancel{2}} = \frac{(K-1)\,P(K-1)}{\cancel{2}}$$

$$\boxed{P(K) = \frac{K-1}{K+2}\,P(K-1)}$$

Similarly for the equation for 'm' (Priviously seen)

~~N+1~~(

$$(N+1)\,P(m,t+1) = N\,P(m,t) + 1 - \frac{m}{2}\,P(m,t)$$

we can calculate steady state (for $t \to \infty$)

$$(N+1)\,P(m) = N P(m) + 1 - \frac{m}{2} P(m)$$

which leads $\cancel{P(m)N} + P(m) = \cancel{NP(m)} + 1 - \frac{m}{2} P(m)$

$$P(m) = 1 - \frac{m}{2}\,P(m)$$

$$P(m) + \frac{m}{2} P(m) = 1$$

$$P(m)\left(1 + \frac{m}{2}\right) = 1$$

$$P(m)\left(\frac{2+m}{2}\right) = 1$$

or $\boxed{P(m) = \frac{2}{2+m}}$

Now solving a recurrence relation

$$P(K) = \frac{K-1}{K+2}\,P(K-1)$$

with Base case minimum digree m

$\boxed{P(m) = \frac{2}{2+m}}$ put $K = m+1$

$$P(m+1) = \frac{(m+1)-1}{(m+1)+2}\,P((m+1)-1)$$

$$P(m+1) = \frac{m}{m+3} \times P(m)$$

$$P(m+1) = \frac{2m}{(m+2)(m+3)}$$

similarly

$$P(m+2) = \frac{m+1}{m+4} P(m+1) = \frac{2m(m+1)}{(m+2)(m+3)(m+4)}$$

$$P(m+3) = \frac{m+2}{m+5} P(m+2) = \frac{2m(m+1)}{(m+3)(m+4)(m+5)}$$

$$\vdots$$

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$$

$$P(k) = \frac{1}{k^3}\left[\frac{2m(m+1)}{\left(1+\frac{1}{k}\right)\left(1+\frac{2}{k}\right)}\right]$$

$$P(k) \cong \frac{1}{k^3}$$

or

$$\boxed{P(k) \cong k^{-3}} \qquad \boxed{\gamma = 3}$$

Another way to solve this problem

We know

$$P(K) = \frac{K-1}{2} P(K-1) - \frac{K}{2} P(K)$$

$$\left\{ \because P(K-1) \approx P(K) - \frac{\partial P(K)}{\partial K} \right\}$$

continous approximation for large K

$$2P(K) = (K-1)\left[P(K) - \frac{\partial P(K)}{\partial K}\right] - K\,P(K)$$

$$2P(K) = (K-1)P(K) - (K-1)\frac{\partial P(K)}{\partial K} - K\,P(K)$$

$$\left(2-(K-1)+K\right)P(K) = -(K-1)\frac{\partial P(K)}{\partial K}$$

$$3P(K) = -(K-1)\frac{\partial P(K)}{\partial K}$$

$$\boxed{P(K) = \frac{-(K-1)}{3}\frac{\partial P(K)}{\partial K}}$$

further solving

$$3\int \frac{-\partial K}{K-1} = \int \frac{\partial P(K)}{P(K)}$$

$$-3\log(K-1) = \log P(K) + \log C$$

$$\log(K-1)^{-3} + \log C = \log P(K)$$

$$\log\left[C\cdot(K-1)^{-3}\right] = \log P(K)$$

$$P(K) = C\cdot(K-1)^{-3}$$

for large K, $K-1 \approx K$

So $\boxed{P(K) \sim K^{-3}}$

$$\boxed{\gamma = 3}$$

Chapter - 6

Bianconi - Barabasi model :- It is advanced version of the Barabasi - Albert model. used to explain how networks grow over time.

Think of a network as a collection of nodes.

Each node has a "fitness" value, which measures how attractive or competitive it is.

Ex :- ~~The~~ A. person which is famous in a Social network might have high fitness making them more likely to get connections

Over a time, the network forms a structure where both "connections" and "fitness" determines the most influential nodes.

$$\boxed{\text{fitness}(n) = \Pi(k_i) \cong \frac{\eta_i k_i}{\sum_j \eta_j k_j}}$$

In each timestemp a new node j ~~with~~ with m links and fitness η_j is added to a network.

The Probablity that a link of a new node connects to node i is proportional to the product of node i's degree k_i & its fitnes η_i is

$$\pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

Chapter 7

Hierachical clustering :-

Ravasz Algorithm :-

$$x_{ij} = \frac{J_N(i,j)}{\min\{k_i, k_j\}}$$

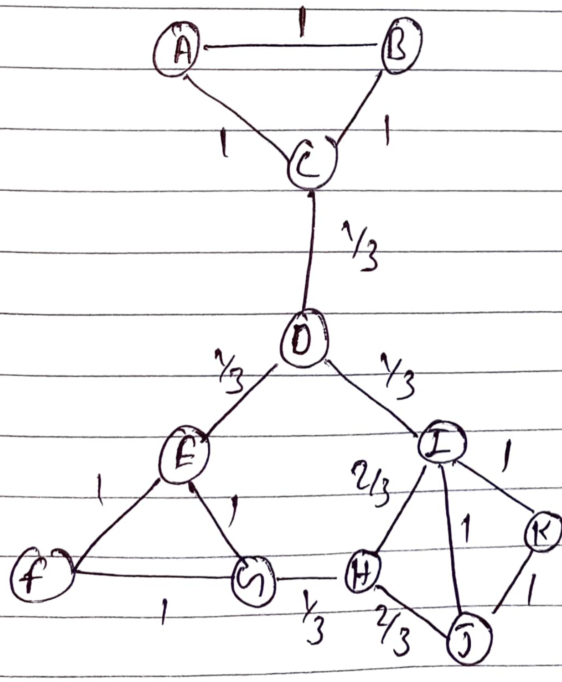
where $J_N(i,j)$ is the number of common neighbour of node i & j

+1 if there is direct link b/w i & j

or

$$x_{ij} = \frac{\text{Common Node B/w } i \& j + \cancel{+1} \text{ if common link } i \& j}{\min(\cancel{k_i}\ \text{dig}(i),\ \text{dig}(j))}$$

Example :-



$$r_{A,B} = \frac{1+1}{\min(2,2)} = \frac{2}{2} = 1$$

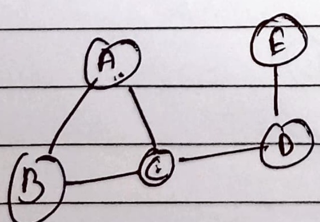
$$r_{A,C} = \frac{1+1}{\min(2,3)} = \frac{2}{2} = 1$$

$$r_{H,I} = \frac{1+1}{\min(3,4)} = \frac{2}{3}$$

$$r_{G,H} = \frac{0+1}{\min(3,3)} = \frac{1}{3}$$

etc.

Example :- Do Hierachical clustering of the following graph.



Solve: finding r_{ij} B/w each pair of Node to make a similarity matrix.

$$r_{A,B} = \frac{1+1}{\min(2,2)} = 1$$

$$r_{A,C} = \frac{1+1}{\min(2,3)} = \frac{2}{2} = 1$$

$$n_{A,D} = \frac{1+0}{\min(2,2)} = \frac{1}{2}$$

$$n_{A,E} = \frac{0,0}{\min(2,1)} = \frac{0}{1} = 0$$

$$n_{B,C} = \frac{1+1}{\min(2,3)} = \frac{2}{2} = 1$$

$$n_{B,D} = \frac{1+0}{\min(2,2)} = \frac{1}{2} = \frac{1}{2}$$

$$n_{B,E} = \frac{0+0}{\min(2,1)} = \frac{0}{1} = 0$$

$$n_{C,D} = \frac{0+1}{\min(3,2)} = \frac{1}{2} = \frac{1}{2}$$

$$n_{C,E} = \frac{1+0}{\min(3,1)} = \frac{1}{1} = 1$$

$$n_{D,E} = \frac{0+1}{\min(2,1)} = \frac{1}{1} = 1$$

we got similarity matrix :-

	A	B	C	D	E
A	-				
B	1	-			
C	1	1	-		
D	1/2	1/2	1/2	-	
E	0	0	1	1	-

Important formula :-

for Single Linkage

$$Sim(AB, C) = Max(sim(A,C), sim(B,C))$$

for Complete linkage

$$Sim(AB, C) = Min(Sim(A,C), Sim(B,C))$$

& Take Highest value for Both Single & Complete linkage when merging.

Now Create a dendograme using Single linkage method.

Choose max value which is 1
take any of the link that has a similarity 1

we took A, B

	AB	C	D	E
AB	–			
C	1	–		
D	1/2	1/2	–	
E	0	1	1	–

$Sim(C, AB) = Max(sim(C,A), Sim(C,B))$

$= max(1, 1) = 1$

$sim(D, A,B) = Max(Sim(D,A), Sim(D B)) = Max(\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}$

$Sim(E, AB) = Max(Sim(E,A), Sim(E,B)) = max(0,0) = 0$

Now again choosing maximum value 1 (C, AB) combining

	ABC	D	E
ABC	–		
D	1/2	–	
E	1	1	–

$Sim(D, ABC) = max(sim(AB, D), Sim(C,D))$

$= max(\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}$

$sim(E, ABC) = max(0, 1) = 1$

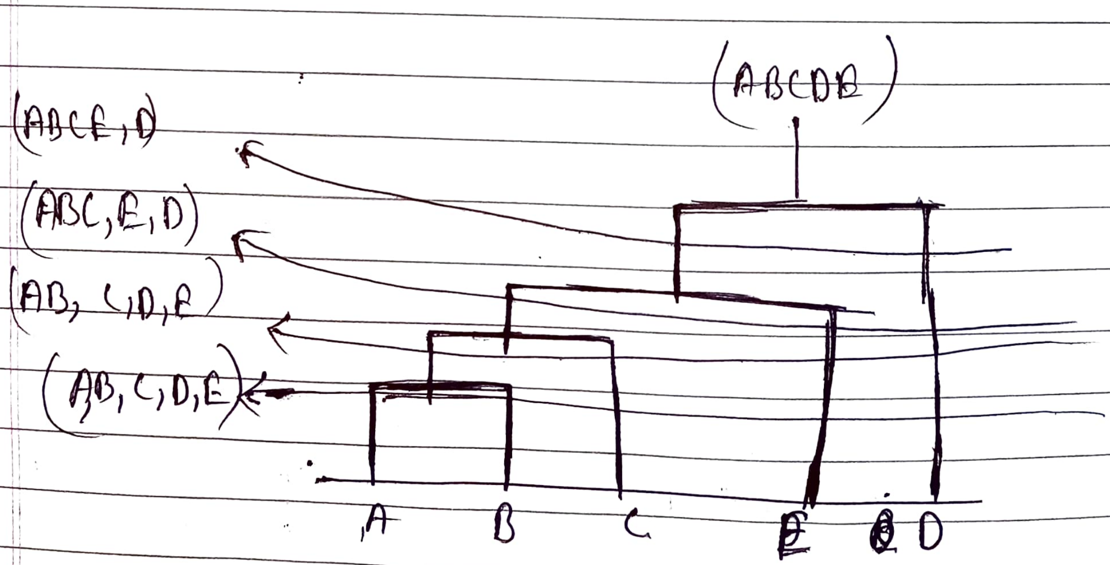
Now again taking maximum value /
~~thi~~ combining (ABC & E) we can also (E,B)

	ABCE	D
ABCE	—	
D	1	—

$$sim(ABCE, D) = Max(sim(ABC, D); sim(\cancel{ABC}, E, D))$$
$$= Max(1, 1)$$
$$= 1$$

finally ABCE & D is combined

Dendogram for this.



Construct Dendogram using complete linkage

Take a maximum value which is 1
Take Any of which has similarity 1
& combine them

	AB	C	D	E
AB	–			
C	1	–		
D	1/2	1/2	–	
E	0	1	1	–

Again taking max value link which is ~~ABC~~ AB & C

~~ABC~~	ABC	D	E
ABC	–		
D	1/2	–	
E	0	1	–

$$\text{sim}(ABC, D) = \min(\text{sim}(AB, D), \text{sim}(C, D))$$
$$\min\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}$$

$$\text{sim}(ABC, E) = \min(\text{sim}(AB, E), \text{sim}(C, E))$$
$$= \min(0, 1) = 0$$

Again taking max value which is B/W E & D so combining E & D

	ABC	ED
ABC	—	
ED	0	—

$$Sim(ABC, ED) = Min(Sim(ABC, E), Sim(ABC, D))$$
$$= min\left(\frac{1}{2}, 0\right) = 0$$

Dendogram for this