Social Network

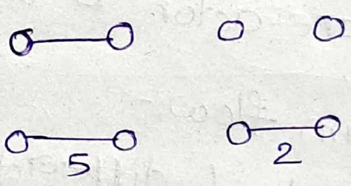
- A social network is social structure of people, related to each other through a common relation or interest.
- SNA is study of social networks to understand their structure & behaviour.

People – nodes
relations – edges

Social Network Analysis
- Sociocentric (whole)
- Egocentric

Types of social relations
- friendship
- acquaintance
- kinship
- advice
- hindrance
- sex.

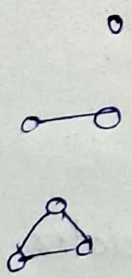
Strength of Tie



dichotomous

valued

class code :- $\boxed{\text{3c4fmde}}$



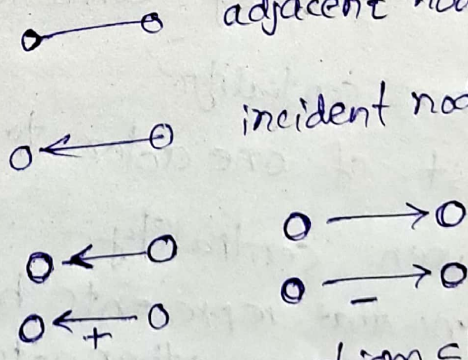
actor

dyad.

triad.

friendship Relations.

Kinship



adjacent node to

incident node to

3 isomorphism classes
- null dyad.
- mutual dyad.
- asymmetric.

converse of graph:-
Reverse direction of all arcs.

Density :- $\frac{\text{Number of ties that are present}}{\text{amount of ties that can be present}}$

out-degree :- sum of connections from an actor to others.

In-degree :- sum of connections to an actor.

---

* Distance

- walk :- sequence of actors & relations that begins & ends with actors.

- Geodesic distance :- the shortest possible walk
  Number of relationship in from one actor to another.

- Maximum flow :-
  amount of different actors in neighbourhood of a source that lead to pathways to Target.

- Degree :- sum of connections from or to an actor.

- closeness centrality :-
  - Dist. of one actor to all others.

- Between centrality :-
  no. that represents how frequently an actor is between other actor's geodesic path.

centrality Measures :-

Who is the most prominent?

- Who knows the most actors? → Degree
- ★ Who controls knowledge flow? → Between
- ★ Who has shortest distance? → closeness

Who has more power?

What are LinkedIn connections?

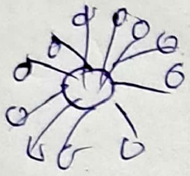
1st - directly connected

2nd - people connected to 1st degree

3rd - people connected to 2nd degree.

Why 6 degrees of separation? 6 degrees of ~~Freedom~~. separation

Each friend connected to 44 people



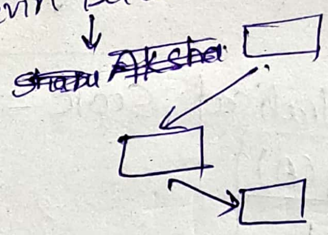
$(44)^6 \approx 7.26$ Billion people

small world theorem.

You can get job connections easily through Random acquaintances than through your close friends! - Veri~~sa~~tasium.

strength of weak Tie.

The oral of Bacon! → Kevin Bacon



Bacon! 3.

Erdös Number Project.

• Adjacency Matrix L one mode • Affiliation Matrix L Two mode.	Adjacency matrix vs Attribute Matrix

order :- no. of nodes in graph.

cardinality :- no. of edges.

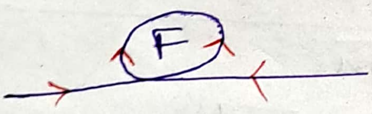
size :- max. no. of edges , ie $\frac{n(n-1)}{2}$ for undirected.

$n(n-1)$ for directed.

Degree centrality

→ for directed node → in-degree, → out-degree.

Average degree ⇒ $\frac{Deg(A) + Deg(B) + \cdots Deg(N)}{n}$.



• Loop

• Deg(F) = 4

?

Degree centralised score = $\frac{\cancel{Deg(A)}}{\cancel{max(Deg\ of\ an\ any\ node)}}$

$= \boxed{\frac{Deg(A)}{n-1}}$?

standardised score or normalised form.

---

closeness centrality :- ⇒ ~~avg (dist with all nodes)~~

⇒ inverse of sum of distance between a node & other nodes in a network.

closeness score ⇒ $\frac{1}{\text{total dist}}$

standardised score → $\frac{1}{\text{total dist}} \times (n-1)$

most closest node ⇒ ~~sta~~ highest standardised score

Betweenness centrality :-

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$$\text{normal}(g(v)) = \frac{g(v) - \min(g)}{\max(g) - \min(g)}$$

which results in max normal ⇒ 1

σ_{st} is total no. of shortest path from $s \to t$.

$\sigma_{st}(v)$ is no. of paths passing through v. ('v' is not 's' or 't')

~~v is not~~

Q] For friendship network, which is most popular person?

ans:- degree centrality (but there is never fixed answer)

Q] Find section in an information flow network that can most efficiently ~~for~~ ~~info~~ flow info?

ans:- closeness centrality.

Q] Which section controls, information flow?

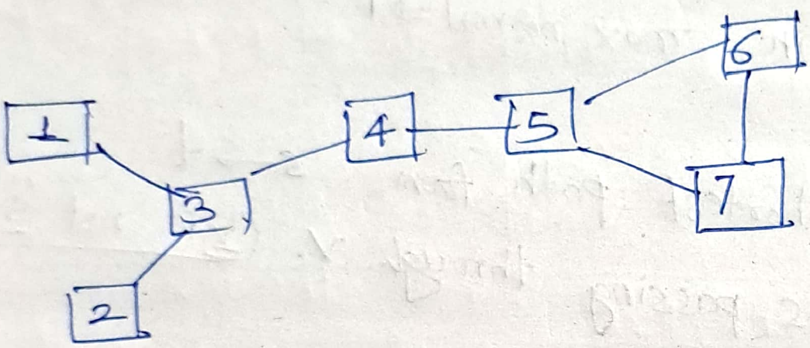
ans:- betweeness centrality

THANK YOU!

* do all maths yourself.
* GEPHI network Lab pratical.
* Install NetworkX.
* Next to Next class :- Quiz.

Data Management :-

- Adjacency Matrix.
  - Asymmetric binary
  - Asymmetric weighted.
- Affiliation Matrix.



consider only paths involving 4 in them.
As they will have contribution in Betweeness.

(1−3) ✗ (1−5) ✓

1−3−4−5

Node 3

1−2	1
1−4	1
1−5	1
1−6	1
1−7	1
2−4	1
2−5	1
2−6	1
2−7	1
	9

for directed divide by $(n-1)(n-2)$

for undirected divide by

$$\frac{(n-1)(n-2)}{2} = \frac{(7-1)(7-2)}{2} = 15$$

Hence:- $\frac{9}{15}$ (Here undirected).

Node 5

1-6	1
1-7	1
2-6	1
2-7	1
3-6	1
3-7	1
4-6	1
4-7	1
6-7	0 ⇒ not 5 in between

$\boxed{\frac{8}{15}}$

node	score
1	0
2	0
3	$\frac{9}{15}$
4	~~$\frac{8}{15}$~~ $\frac{9}{15}$
5	$\frac{8}{15}$
6	0
7	0

Another formula

$$\text{Between}(s,t) = \frac{\text{no. of paths}}{\text{no. of nodes}}$$

ex:- Between. $(1,5) \Rightarrow \frac{1}{2}$ $1 - \boxed{3-4} - 5$

$(3-5) \Rightarrow \frac{1}{1}$ $3-4-5.$

$(2-5) \Rightarrow \frac{1}{2}$ $2 - \boxed{3-4} - 5$

$(2-7) \Rightarrow \frac{1}{3}$ $2 - \boxed{3-4-5} - 7$

Graph Measures :-

★ Eccentricity ($\mathcal{E}(v)$)

$\mathcal{E}(v)$ of vertex 'v' in connected graph G is max. graph distance between v and any other vertex u of G.

$$\mathcal{E}(v) = \max_{u \in V} d(u, v)$$

For disconnected point $\mathcal{E}(v) = \infty$.

- **Diameter** :- max eccentricity
- **Radius** :- min eccentricity
- **Central Point** :- $\mathcal{E}(v) =$ Radius

Eigenvector centrality (You are powerful if friends are powerful)

Influence of node in network:-

Importance of node depends on importance of its neighbours in recursive manner.

$$Av = \lambda v$$

select an eigen vector associated with largest eigen value

Eigen Vector $\boxed{A - \lambda|I| = 0}$

- A is some matrix.
- size of matrix, tells no. of Eigen values

ex:- 3x3, then 3 eigen values

Greatest Eigen value is Eigenvector Centrality

~~compute~~ A is adjacency matrix Here.

Katz centrality

- almost same as Eigen vector centrality.
- quality (who) and quantity (how many) matters
- Also considers Attenuation in influence.

Bonacich centrality

Two parametric centrality measure $c(\alpha, \beta)$.

α is normalisation factor

β can be +ve or -ve.

+ve → connected to powerful

-ve → connected to weak node.

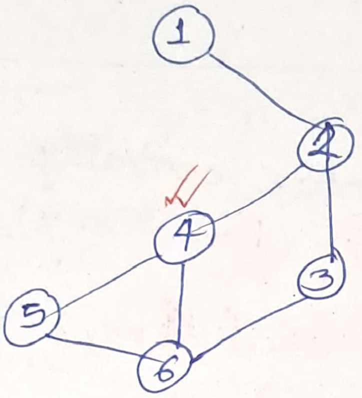
Eigen vector centrality

works well for strongly connected network.

but not for weakly connected network.

Eigenvector Centrality Example :- exam question

$$v = \begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array}\begin{bmatrix} 0.31 \\ 0.79 \\ 0.69 \\ \boxed{1.00} \\ 0.78 \\ 0.97 \end{bmatrix}\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}$$



Solve $A - \lambda I$.

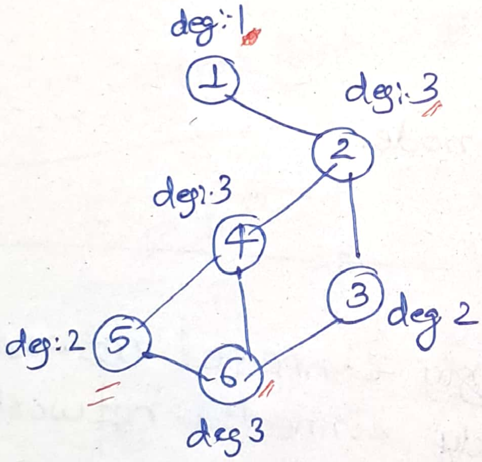
$\lambda = 2, 1, \underline{2.54}, 1.5$
max.

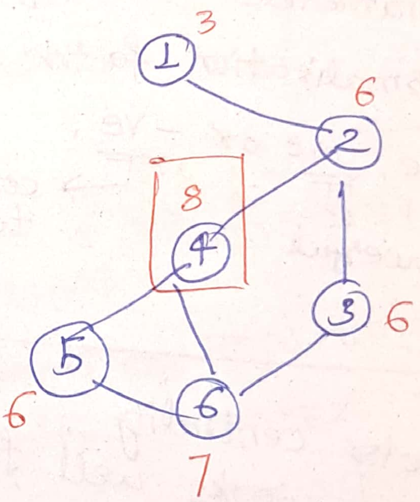
$Av = 2.54\, v$

choose eigen vector v, with highest eigen value.

(use. Matrixcalc.org
Matrix calculator)

Trick : add degree of your friends.





new power = sum of power of neighbours

Intuitive Approach (Power iteration method)

net power ← sum of power of neighbours

normalise (divide by max, Here 52

	0	1	2	3	4	
1	1	1	3	6	17	0.32
2	1	3	6	17	38	0.73
3	1	2	6	13	37	0.71
4	1	3	8	19	52	1.00
5	1	2	6	15	39	0.75
6	1	3	7	20	47	0.90

similar as eigen vector →

$$v = \begin{bmatrix} 0.31 \\ 0.79 \\ 0.69 \\ 1.00 \\ 0.78 \\ 0.97 \end{bmatrix}$$

why stop at round 4? (~~no answers~~ search at Home).

because at round 5, maybe instead of node 4, node 6 wins

	5
1	38
2	106
3	85
4	124
5	99
6	128

Stop?

$$X_t = A^T X_{t-1}$$

Stop when new−old < threshold.

or do this until highest eigen value.

Why iterations?

because wrt time, friends and connections will change

Lets check

	0	1	2	3	4
1	1	0.33	0.375	0.3	0.28
2	1	1	0.75	0.85	0.73
3	1	0.66	0.75	0.65	0.71
4	1	1	1	0.95	1
5	1	0.66	0.75	0.25	0.75
6	1	1	0.875	1	0.90
	$\sqrt{6}$	6	2.51	2.529	

reaching eigen value.

Stopping criteria

1) L_2 norm $\sqrt{\Sigma x^2}$

$(t_t - t_{t-1}) \leq \delta$

2) Fluctuation/oscillation.

3) $|X_t - X_{t-1}| < \delta$

4) Rayleigh Quotient

H.W Read Tutorial on PCA by L i smith.

$cov(x,x) = var(x)$.

$cov(x,y) = cov(y,x)$

Scanned with OKEN Scanner

In links

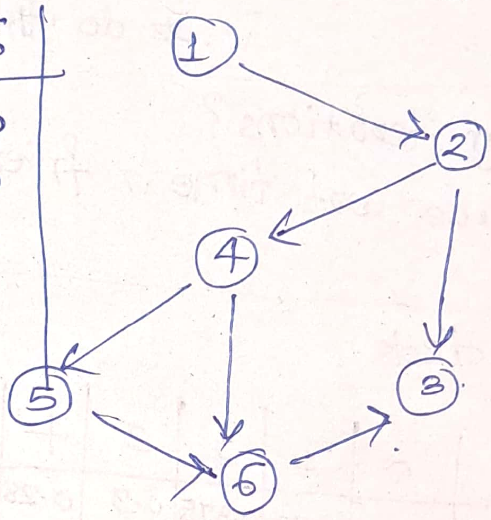
For → directed: $\boxed{AX_t = X_{t+1}}$

For → indirected: $\boxed{A^T X_t = X_{t+1}}$

Calculate Eigen vector centrality for directed graph:-

in-degree

t	0	1	2	3	4	5	6
1	1	0	0	0	0	0	0
2	1	1	0	0	0	0	0
3	1	2	3	2	2	1	0
4	1	1	1	0	0	0	0
5	1	1	1	1	0	0	0
6	1	2	2	2	1	0	0



* acyclic diagraph

• all power leaks out.

citation networks are acyclic, so eigen vector centrality is useless.

Adaptation for directed networks -

$$\boxed{X = \underbrace{\alpha A X}_{\text{power from neighbours}} + \underbrace{\beta 1}_{\text{intrisic power}}}$$

when this converges, we get:-

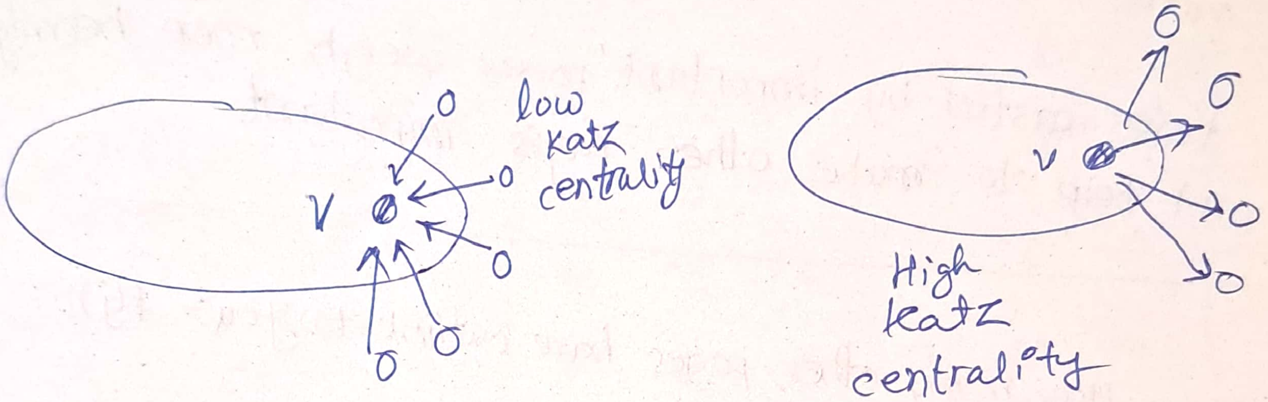
$$X = \beta(1 - \alpha A^T)^{-1} 1$$

$$\boxed{X = (1 - \alpha A^T)^{-1} 1}$$

Katz centrality
$\beta = 1$

- If $\alpha = 0$, $X = I$, centralities $= 1$
- As $\alpha \uparrow\uparrow$, centralities increase
- At $\frac{1}{\lambda_1}$, (λ_1 is largest eigen value) things diverge

Hence $\boxed{0 \leq \alpha \leq \frac{1}{\lambda_1}}$



*PageRank is most powerful measure.

Next class:- 1, 2, 3 chapter.
Eigen, Katz, bonachich.
Hard copy Quiz.

Note:- We stop when Eigen vector values converge.

Scanned with OKEN Scanner

PAGE RANK

For web link analysis models.

PageRank interprets a hyperlink from page x to page y as a vote, by page x, for page y.

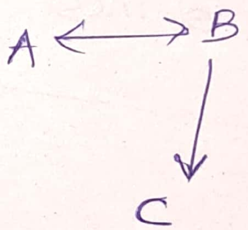
However, Page Rank looks at more than sheer number of votes; it also analyses the page that casts the vote.

Votes casted by "important" pages weigh more heavily and help to make other pages important.

How many other pages have outlink to you :- $P(j)$

How many votes they give :- O_j

B. votes = 2.



$P(A) = \frac{P(B)}{2} = 1.5.$

A. votes = 1

$P(B) = \frac{P(A)}{1} = 1.5$

Lets assume $P(B) = 3$.
Then calculate $P(A)$ & $P(C)$.

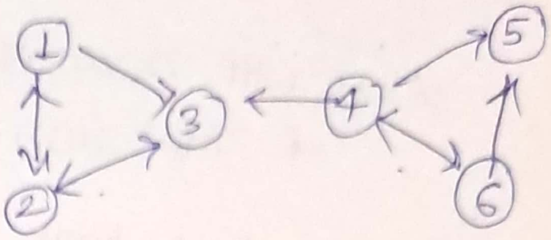
"If your network has low Rank pages,
Your page Rank will also decrease"

Solving Page Rank

$$\boxed{P_{i+1} = A^T P_i}$$

stopping criteria $\Rightarrow$ $P_{i+1} - P_i < 0.1$.

$$\bar{A} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$



Let $P_0 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix}$

$$A^T = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/3 & 0 & 1/2 \\ 0 & 0 & 0 & 1/3 & 0 & 0 \end{bmatrix}$$

$P_1 = A^T P_0$

$P_2 = A^T P_1$

$\vdots$

$P_9 = A^T P_8$

$$P_8 = \begin{bmatrix} 4.518 \\ 8.898 \\ 6.784 \\ 0.210 \\ 0.315 \\ 0.182 \end{bmatrix} \qquad P_9 = \begin{bmatrix} 4.501 \\ 9.096 \\ 6.830 \\ 0.173 \\ 0.213 \\ 0.122 \end{bmatrix}$$

$P_9 - P_8 < 0.1$, so stop.

$\therefore$ Rank :- 2, 3, 1, 5, 4, 6

Irrespective of P_0, same Rank will be achieved.

How compare importance of pages?.

start with normalisation.
& Total sum = 1

∴ 6 pages, Each Rank is 1/6.
Total sum = 1.

∴ Updated page Rank :-

$$P(i) = (1-d) + d \sum_{(j,i \in E)} \frac{P(j)}{O_j}$$

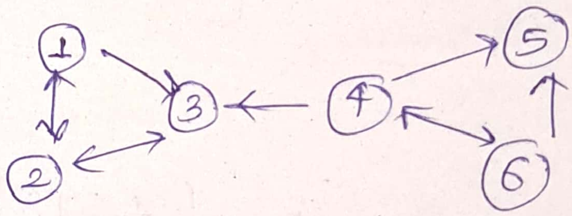
Also $d = 0.8$ is recommended.

$$= (1-d) + d(A^T P)$$

---

ex:- let $P_0 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$

$d = 0.8$



~~P(1) = (1-0.8)~~

$P(1) = (1-0.8) + 0.8\,(2/2) = 1$

$P(3) = (1-0.8) + 0.8\,(1/2 + 3/2 + 4/3) = 2.86$

---- Likewise

Ranking :- (Iteration 1) $[5, 2, 3, 4, 1, 6]$

Iteration 2) :- $[2, 3, 1, 5, 4, 6]$.
which is answer

Scanned with OKEN Scanner

Next improvement (improved normalisation)

$$P = (1-d)\frac{1}{n} + d\left(\frac{PR(i)}{O(i)} + \frac{PR(j)}{O(j)} + \cdots \frac{PR(n)}{O(n)}\right)$$

---

cutpoint - delete node to increase components.

bridge - delete edge, to increase components

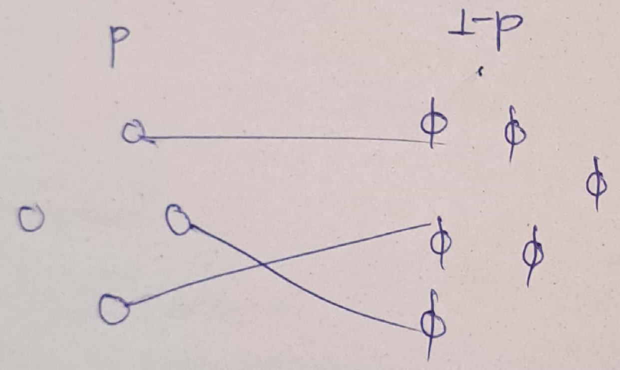
Homophily :- Birds of feather flock together.

Netlogo Web → Social science → segmentation

consider random network $G(V, E^r)$
where each node is assigned ~~most~~ male :- p
female :- 1-p.

Let $G = (V, E)$ be random sample of R with p fraction of male and 1-p fraction of female.

- consider any edge $(i, j) \in E^r$ of random network R.



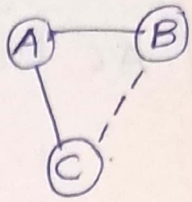
$$E(x^2) - (E(x))^2 = p(1-p) + (1-p)p = 2p(1-p)$$

Computer networks :- Hoax calls, mails (phishing links)

uptill Homophily mid-sem

Social vs selection Influence

* Triadic Closure.
* Focal closure
* membership closure



Community Homophily :-

Asian origin vertices:- 5, 7, 8, 10, 11, 12.

Caucasian origin vertices :- 1, 2, 3, 4, 6, 13, 14.

Hispanic origin vertices:- 9, 15, 16, 17, 18, 19, 20.

$\# \text{Links} = \frac{\text{sum of degrees}}{2} = 47$

A : $\frac{6}{20} = 0.3$

C : $\frac{7}{20} = 0.35$

H : $\frac{7}{20} = 0.35.$

A & C	H
p	(1-p)
$\frac{13}{20}$	$\frac{7}{20}$

$2p(1-p) = 2 \times \frac{13}{20} \times \frac{7}{20} = \frac{182}{400}$

A & H	C
$\frac{13}{20}$	$\frac{7}{20}$

$2p(1-p) = \frac{182}{400}$

C & H	A
$\frac{14}{20}$	$\frac{6}{20}$

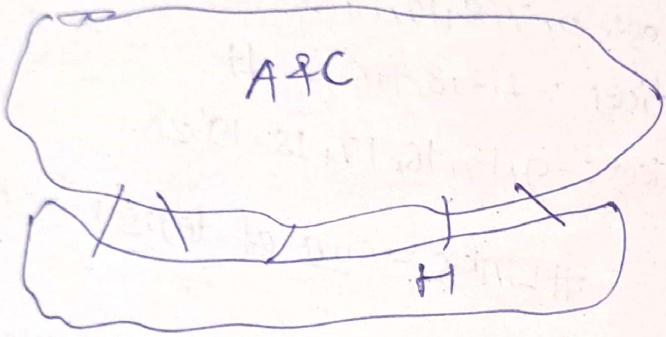
$2p(1-p) = 2 \times \frac{14}{20} \times \frac{6}{20} = \frac{168}{400}.$

consider any 2 groups.

Fraction of A-C Links $= \frac{12}{47} = 0.26$

$> 2*A*C = 2*0.3*0.5 = \boxed{0.21}$

~~Fraction~~ Final community detected



2 communities → $\boxed{A\&C}$ $\boxed{H}$

Mid-sem exam syllabus over

~~Not taken~~ Homophily.

Assignments :- 35 marks
