(4) → Feb4 → Python / R can use Graphic Tools

PAGE NO.:
DATE: / /

3/01/19

SNA.

→ Talk about computational part along with social science part.

→ Edge b/w 2 vertices is temporal in nature in case of n/w.
But in graph theory, we've static graph.

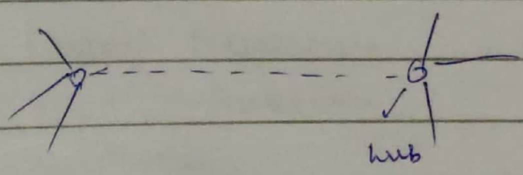
→ Degree in case of n/w may change frequently. So, not good to talk about degree over here.
Hence, we can say n/w is a completely random graph ? NO (see later)
Two vertices are connected with some probability

→ Scale free n/w : will study about it.

Active Participants in Class & Piazza	5 %.
Mid term	25%.
Project presentⁿ & Report	25%.
End term	45 %.

→ SNA : studies about relation b/w individuals (edges), not individual themselves (vertices)
↓ called
structural / topology

→ social actor : can be human / non-human



scale free n/w
(not random)

Scanned by CamScanner

③ Barabasi ⟶ N/W Science
① Zafrani, Abbasi ⟶ Social Media Mining : an Intro

PAGE NO.:
DATE: / /

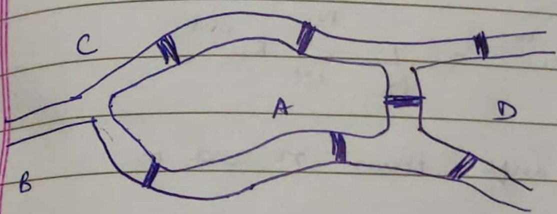
→ Weak ties are more stronger than stronger ties.
↓
a person who isn't closely connected to you will leak your info. more than " " is " " " .
↓
friend ~~p~~ of this person will also be your friend. Info will remain in closed n/w.

6/1/19

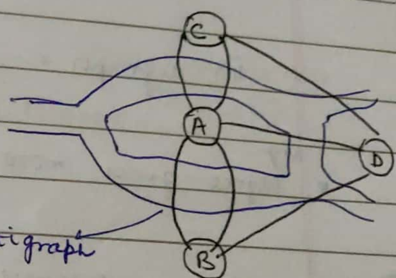
Some basic Graph Theory Concepts & Algorithms

The bridges of Königsberg

Is there a ways to go through bridges once to reach again at same point.



≡



Abstract form
(Look into only those things which matters & leave out the rest)

We've no sol^n here.

No Eulerian ~~tool~~ tour
↓
If a vertex has odd degree, it can't have Eulerian ~~tool~~ tour

Graph :

$G = (V, E)$

$V \rightarrow$ set of all vertices / nodes (n) (n, m)

$E \rightarrow$ set of all edges / links (m) N L

$e = (V_1, V_2)$

↳ Multi-graph : ≥ 1 edge for a pair of vertices

↳ Simple graph : unique edge " every "

Teacher's Signature.........

→ To analyze social n/w, we need to look at the edges / links

→ Graph is an abstracⁿ : We don't need to look what is the connecⁿ about. Graph in all case is same. ~~Con~~

Graph ↓ can be drawn in many ways (by changing node ids) but connectivity will be same

- directed graph ⇒ digraph
- no. of ~~vertices~~ edges incident on vertex ⇒ degree

Avg. degree of a node $\langle k \rangle$ in ~~or~~ : $\rightarrow$ degree of vertex i

i) undirected graph : $\frac{1}{N} \sum_{i=1}^{N} k_i$

ii.) digraph = $\frac{1}{N} \sum_{i=1}^{N} k_i^{in} = \frac{1}{N} \sum_{i=1}^{N} k_i^{out}$

* Avg. Degree gives more specific info. than n and m.

→ Dataset : Internet, WWW, Power Grid, Mobile-Phone Calls, E-mail, Science Collabⁿ, Actor N/W, Citatⁿ N/W, E coli Metabolism, Protein Interaction

| Observⁿ

No. of links ↓↓ than ~~no o~~ possible $\rightarrow \frac{N(N-1)}{2}$
degree = 6/4/... (no. of nodes ↑↑)
⇓
Sparse graph

Actor N/W : seems more closed than other graphs (degree = 83)

Graph Representation

1) Adjacency Matrix :

find ~~the~~ degree of vertex i taking into considerⁿ only i^{th} row or i^{th} col (take sum)

Q How to know indegree / outdegree in list ?

PAGE NO.:
DATE: / /

Sparse graphs → memory wasted → use list.

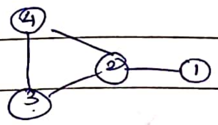
2) Adjacency list (Will use this as have sparse)

<u>Degree distrib^n</u>

- p_k : prob. that a randomly selected node in n/w has degree k.

$$p_k = \frac{N_k}{N}$$

1)

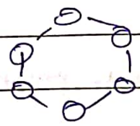


$$p_2 = \frac{2}{4} = \frac{1}{2} \; ; \; p_3 = \frac{1}{4} = 0.25$$

4 & 3

$$\sum_k p_k = 1$$

2)



<u>2-Regular Graph</u> (Every vertex has degree 2)

(general : k-Regular Graph)

↓

$$\boxed{p_k = 1}$$

- Avg. degree of n/w is : k: no. of degrees

$$\sum_{k=0}^{\infty} k \, p_k$$

(Prove Mathematically)

~~Eg:~~ :

Teacher's Signature

Power Law Distrib^n

Facebook : Users with lots of friends is very low & users with less friends is very huge. (Both globally & in U.S.)

Weighted Graph

↑ wt. ⇒ link used very frequently.

Project Purpose ★ Metcalfe's Law : value of n/w ∝ (Nodes)^2
(Because no. of links = $O(N^2)$)

Min. edges : 0
Max. edges : $\frac{N(N-1)}{2}$] → for simple graph

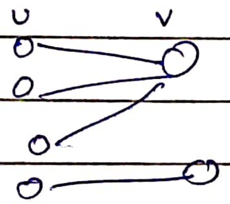
multi-graph : (Max : ∞)

↳ Cost associated with n/w ↑ linearly with no. of nodes but value will be square of it.
(That's why, Jio, etc. give offers to you. They build users once n/w is formed, it'll take care of the value).

Bipartite N/W :

Set of nodes can be partitioned into 2 in which links will only be across the 2 " , not within a part^n.

Ex.



: can have Projection U & V

→ Projection U : connects 2 U-nodes if they're linked to same V- node

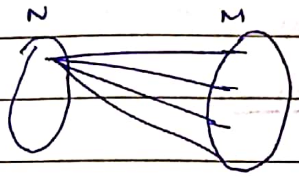
Likewise, we've Projection V.

Ex Twitter.
Join 2 accounts having common follower / following

Some basic terms:

1. Null graph : V, E empty
2. empty " : E "
3. complete " : E b/w any pair of nodes — K_n

4. Complete bipartite graphs : $K_{n,m}$



For each node in 1st, have edge with each - - 2nd
($n \times m$ edges)

4. Walk : seq. of incident edges traversed one after another
open walk (start & end not same)
closed " (" " are ")

5. Trail : walk where no edges are repeated more than 1.
Closed Trail - Tour / circuit

6. Path : " " no V or E — —
Closed Path : Cycle

7. Eulerian Tour : all edges are covered

8. Planar Graph : drawn without edges crossing over each other
K_5 : non-planar (Any graph having K_5 as sub graph is non-planar)

9. Hamiltonian ~~Graph~~ cycle : all vertices are covered [No repetition of E or V]
↓
checking this is → exponential
NP complete

Checking eulerian Tour is just O(N) [check odd degree vertices]
↓
[No repetition of E]

10/1/19

→ Open Graph : Interface (not related with graph theory)
using fb credentials for other webpages,
also uses your profile (photos, friends, etc)

∃ path

PAGE NO.:
DATE: / /

connectivity in a Graph

→ v_i is connected to v_j if ∃ path b/w them. (Both are reachable)
→ If every pair of vertices are connected wrt dir^n in directed graph : strongly connected ($v \to u$ & $u \to v$)
→ In a directed graph, if we replace all directed edges by undirected edges & graph is connected : weakly connected
* Every strongly connected is " "
→ We are dealing with large graphs. It'll not be connected. But we may have to analyze a component of graph which is connected. There, it'll be useful. (largest one to be connected)
→ To find spread of graph : Diameter

contains min. no. of edges

length of longest (shortest path) b/w any 2 pair of vertices

$u, v \in V$

$P_{u,v} \to$ shortest path b/w u & v (Dijkstra / Bellman ford)

Diameter $= \max\limits_{\forall (u,v) \in V} P_{u,v}$

Tree & Forest

Tree : Graph with no cycles
Forest : Graph not connected but has many trees

* In a tree $G = (V, E)$, E has always $(n-1)$ edges (Prove)
- spanning tree : subgraph of G which is a tree & spans all the vertices of G. ($T \subseteq G$)

appl^n → Railways want to connect each city & town.

spanning Tree
~~Both~~ It won't work when we want min. ~~cos~~ cost.
It only works when we just want the connectivity
It may not be efficient wrt time or distance.

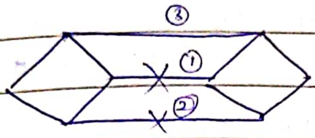
Teacher's Signature

Graph Visualization Tools:
1) NetworkX
2) igraph
3) Topolayout
4) I3.ds
5) graphviz
6) ~~To~~ Tulip

2 kind of tools:
↳ static
↳ Interactive (can interact with the visualization)

→ Min. spanning Tree : Prims, Kruskal
→ Steiner Tree : For a weighted connected graph.

• cut edge (Bridge) : ~~as~~ connects 2 components (becomes 1 component). If cut, ~~o~~ no. of comp = 2



$\{e_1, e_2, e_3\}$ is a set of all bridges
→ Have to Remove atleast 2 edges to disconnect it ()

↳ BJP & Congress : closed n/w twitters But some tweets are cross-interaction b/w 2 n/w's. This is a bridge / cut-edge.

• ~~a~~ Cut-vertex : Have to remove edges adjacent on that vertex.
Articulation pt/ vertex cut/ cut pt/
↳ if removal of that vertex disconnects the graph
↳ Here, we're removing the influential person from connecting.
set of cut-vertices : cut set

Graph Algos

1. DFS 2. BFS 3. Dijkshtra 4. Prim's
(Zafrani book) 5. Network flow algo - Ford-Fulkerson algo

Use of Traversal Algo — BFS / DFS
→ To check cycles
→ " " connectivities
~~a~~ ;

↳ In a social n/w : I want to find avg. age of users (Reason ?)
In case of Huge graph: BFS

PAGE NO.:
DATE: / /

Flow Network

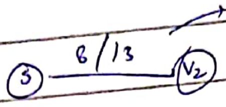
capacity of edge

→ A weighted graph $G = (V, E, c)$

→ $c(u,v) \geq 0$

→ when (u,v) in E, (v,u) not in E

→ s is source node, t is sink node



→ when there's capacity of 13, we're giving flow of 8 ($f(s, v_2) = 8$)

→ Flow func^n : $f(u,v) \geq 0$ $\forall$ (u,v) in E

→ $0 \leq f(u,v) \leq c(u,v)$ " " "

→ Flow conserv^n constraint :

$\forall\, v \in V, v \notin \{s,t\}$,

$$\sum_{k:\,(k,v)\in E} f(k,v) = \sum_{l:\,(v,l)\in E} f(v,l)$$

$\boxed{\text{Inflow} = \text{Outflow}}$

Residual graph : $c(u,v) - f(u,v)$ [can be ↑sed by this value].

but since we're getting 2, we've to send back 2 to src]

(Read) Ford - fulkerson algo

→ checks if I can ↑ flow

→ uses Residual graph, max min cut theorem

Defines a cut set for flow graph.

set of edges where path from s to t has to go from any 1 of the edges

Apply on residual graph

Can ↑se flow by 1 (min. of all → (residues) possible)

$\{(v_2, v_3), (v_2, v_4), (v_4, t)\}$

Appl^n : Want to transfer some bits of data to another system through various hops & also in ∥. Max. bits that can be transferred (through various bandwidths) at any point of time ?

Teacher's Signature........

DO exercise question : Barabasi, Zafani

PAGE NO.:
DATE: / /

② How much tweet a news media can generate to retweet all tweets its followers have tweeted.

Max^m Bipartite Matching (M males , F females → max match s.t. all M get F of their choice
↓
Convert into flow graph.)

There are Maximal Matchings.

15/1/19 2.9 Exercises

1.) Directed : A_{ij} may not be equal to A_{ji} but, in undirected

$\therefore A_{ij} = A_{ji}$

✓ (a) $\min(1, A_{ij} + A_{ji}) \longrightarrow O(n^2)$

(b) $A_{ij} \times A_{ji} \longrightarrow O(n^3)$ or 2^n

→ Because time complexity is more effecient in (a)

2.) 7 can not be a degree with 7 vertices in a simple graph.

$\sum$ degrees = even

$$\sum_{i=1}^{n} d(V_i) = 2|E|$$

Here, sum = 29 → not possible.

4.) $|E| = |V| - 1$: Induction

Consider a tree of only 1 node = 0 edges

Simple graph

$|V| = 1$, $|E| = 0$

↳ result is true : $|E| = |V| - 1$

$|V| = 2$, $|E| = 1$

Assume for $|V| = n$, $|E| = n-1$

Do this if $|V| = n+1$, what is $|E|$? should be n.

Teacher's Signature.............

Scanned by CamScanner

PAGE NO.:
DATE: / /

find answer to do this: 13) # edges = m × n

bridge = when take an edge out, so it becomes unconnected

→ Take a tree : relate it with bridges :
↳ every edge is a bridge
most minimal graph (bridge-wise)
⇓
Cut-vertex : Every parent node is a Cut vertex (Children are not)

(Zafrani Ch-3) Centrality Measures
(How important is a node in a graph)

(1) Degree Centrality
↳ more the degree - more is the connection
influential person - more interaction.

$C_d(v_i) = d_i$ — Undirected
Degree Centrality of vertex v_i is d_i

both will be equal ($C_d(v_i) = d_i^{in}$ — Pretige (more followers)
$C_d(v_i) = d_i^{out}$ — Gregarious (more following)

norm 1) maximal degree can be $n-1$

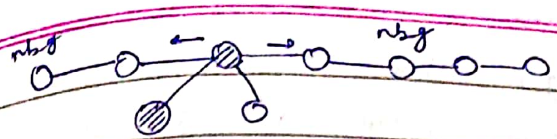
max 2) real max degree in a graph : max d

sum 3) sum of the degrees = 2m → no. of edges
↓
Pick one of the method and then use it across all n/w k (one at a time)

disadv : Only directed connected nodes are ~~consistent~~ considered and not the nbd of connected nodes.

Teacher's Signature.........

PAGE NO.:
DATE: / /



(2.) <u>Eigen vector Centrality</u>

- Having more important friends provides stronger signals

- Being connected to an important ~~person~~ friend makes you more influential
  ↳ If Modi follows a person, he will become a star figure

In directed incoming nbd → indegree

Narendra Modi → you ⇒ To become influential (if N.M. follows you)

$$\underline{\underline{C_e(V_i)}} = \frac{1}{\lambda} \sum_{j=1}^{n} A_{j,i} \, C_e(V_j)$$

↓ eigen vector centrality

for each node: Then we can denote it as a vector. $i = 1, 2, \ldots$

$$C = \left(C_e(V_1), C_e(V_2) \ldots\ldots C_e(V_n) \right)^T$$

$$\lambda C_e = A^T C_e$$

Teacher's Signature ....

PAGE NO.:
DATE: / /

Perron - Frobenius Theorem

λ more $\Rightarrow$ greatest eigen value
↳ then corresponding eigen vector v, $\forall v_i > 0$

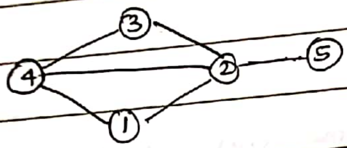
Ex. 3.2 $v_1 - v_2 - v_3$: wrt degrees centrality v_2 is imp.

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$(A - \lambda A)\, C_e = 0$$

3 roots : $\lambda = (-\sqrt{2}, 0, \sqrt{2})$ → consider max.

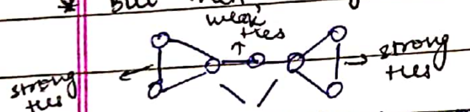
Ans. → $\left[\frac{1}{2}, \frac{\sqrt{2}}{2}, \frac{1}{2}\right]$ ← v_2 ✓

Ques  find eigen vector centrality

→ Node that has to be passed through for any communicⁿ is more influential $v_1 - v_2 - v_3 - v_4 - v_5$

Degree : 1 2 2 2 1

* But then why is v_3 more influential ?



~~degree~~ 2,5 more imp.

But if we consider them as grp then v_6 is the most imp.

Through v_6, inf. will pass to the 2 groups.

Unless you have a weak tie, you won't get inf. cascading effect
→ v_6 connects cluster with another cluster.

* if we cut v_2 then v_1 will be left alone, if we cut v_3 then 2 grps will be formed → more influential (inf. passing)

(3) **Katz Centrality** : Assume a default value (bias value) & C_e (eigen) + β (bias value) → makes value non-zero always (so that value is never 0)

$$C_{Katz}(v_i) = \alpha \sum_{j=1}^{n} A_{j,i}\, C_{Katz}(v_j) + \beta$$

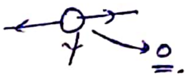
α ↓ wt. given to eigen value

$\alpha = 1 \longrightarrow$ normal

if $\alpha < 1 \longrightarrow$ less value / important

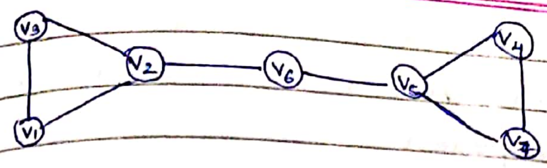
Teacher's Signature........

eigen vector disadv. : If a node only have indegrees & no outdegrees, then centrality for it comes to be 0.



PAGE NO.:
DATE: / /

7/1/19



Disadv. : Here, the degree centrality and eigen vector centrality do not give any emphasis on v_6 whereas v_6 is a very important node but these centrality will give importance to v_2 & v_5.

$$\Rightarrow \quad C_{Katz}(v_i) = \alpha \underbrace{\sum_{j=1}^{n} A_{j,i}\, C_{Katz}(v_j)}_{\text{eigen vector centrality measure}} + \beta$$

tells how much importance we have to give to this eigen vector

$$C_{Katz} = \alpha A^T C_{Katz} + \beta \cdot \mathbf{1}$$

Adjacency matrix

vector of 1

as β is constant, so to make a matrix we do this

$$\boxed{C_{Katz} = \beta (I - \alpha A^T)^{-1} \cdot \mathbf{1}}$$

- when $|(I - \alpha A^T) = 0|$ centrality value diverges
- when $\alpha = \frac{1}{\lambda}$ ~~eigen~~ the value diverges. Good value is $\alpha < \frac{1}{\lambda}$

$\lambda \rightarrow$ largest eigen value

This will avoid a node from getting a 0 centrality measure

Teacher's Signature.........

PAGE NO.:
DATE: / /

Eigen value & Katz:
Both assume that whenever there is a node which is important, its nbd also becomes important.

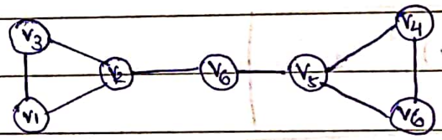
Some challenges to Katz:
→ Once a node becomes an authority, it passes all its centrality along all of its out like

d_j^{out} = no. of outgoing edges from v_j

$$C_p(v_i) = \alpha \sum_{j=1}^{n} A_{j,i} \frac{C_p(v_j)}{d_j^{out}} + \beta$$ → dividing the value wrt no. of outgoing edges

$$\boxed{C_p = \beta (I - \alpha A^T D^{-1})^{-1} \cdot \mathbf{1}}$$ Page Rank

Betweeness Centrality Measure (looks at connectivity)



↳ How many shortest path from left to right goes through V_6.

↳ Here, we talk about connectivity, not imp. to neighbour

$$C_b(V_i) = \sum_{s \neq t \neq V_i} \frac{\sigma_{st}(V_i)}{\underbrace{\sigma_{st}}_{\text{no. of paths from s to t.}}}$$ → take all the shortest paths from node s to t, that goes through V_i.

Problem: How will you find shortest path b/w various vertices?

Teacher's Signature.........

PAGE NO.:
DATE: / /

→ we will not use Dijkstras here but use Flyord's Algorithm

→ No. of shortest paths will vary according to the sparsity of graph.

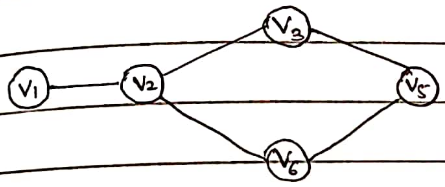
→ Normalized using maximal value of Centrality measure.

→ If every node goes through the center node than this centre node will have max. centrality.

$$C_b^{norm}(v_i) = \frac{C_b(v_i)}{\underset{\downarrow}{2}\binom{n-1}{2}}$$

for undirected graph

Q.



Calculate the $C_b(v_2)$?

$$C_b(v_2) = \underset{(v_1-v_3)}{\frac{1}{1}} + \underset{(v_1-v_4)}{\frac{1}{1}} + \underset{(v_1-v_5)}{\frac{2}{2}} + \underset{(v_3-v_4)}{\frac{1}{2}} + 0 + 0$$

Closeness Centrality

We called a node as central node in a graph, if from that node you can reach other node quickly.

$$C_c(v_i) = \frac{1}{\bar{l}_{v_i}}$$

Teacher's Signature........

Closeness & Betweeness are abt the connectivity.

Degree centrality + eigen vector centrality + Katz centrality + Page Rank ↳ = Are all similar (wrt neighbour)

PAGE NO.:
DATE: / /

22/1/19

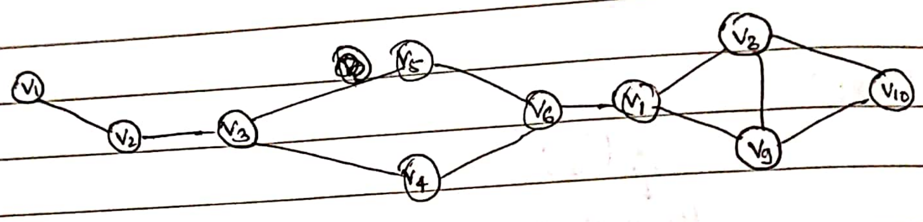
where $\bar{l}_{v_i} = \frac{1}{n-1} \sum_{v_i \neq v_j} l_{i,j}$

getting avg. distance that you need to travel frm v_i to some other node.

(Also looks at the distance)

→ Till here we have seen centrality measure for a single node.

Q.



→ Page Rank is the best Centrality Measure if we want to measure wrt nbd.

Group Degree Centrality :

set S → called as grp $V-S$

• now we will see no. of connections from $S \rightleftarrows V-S$

$$C_d^{grp}(S) = \left| \{ v_i \in V-S \mid v_i \text{ is connected to } v_j \in S \} \right|$$

Group Between Centrality

$$C_b^{group}(S) = \sum_{\substack{s \neq t,\, s \in S \\ t \notin S}} \frac{\sigma_{st}(S)}{\sigma_{st}}$$

Group Closeness Centrality

$$C_c^{group}(S) = \frac{1}{\bar{l}_S^{group}}$$

Scanned by CamScanner

PAGE NO.:
DATE: / /

22/1/19

Centrality

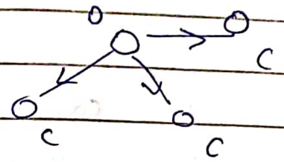
1) Degree :

disadv : don't go further than nbs, can't distinguish if degree is same

2) Eigen :

if you're an outedge to nbr with higher centrality measure, its C_e is transferred to nbr node

disadv : If there's only inedge ⇒ centrality measure : 0



3) Katz : Has biased centrality associated with it. (so, value ≠ 0)

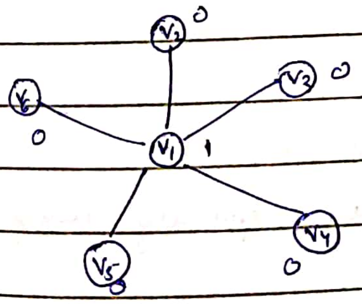
assump^n : If a node is imp, all its nbs are also imp.
↓
divided with no. of outedges.
↓
4) Pagerank.

* Above 4 : are same (considering only nbd)

* Next 3 : consider connectivity

5) Betweeness : How many paths go through a node.

$$\frac{\text{no. of shortest paths b/w A to B passing through a node}}{\text{total no. of shortest path b/w A \& B}}$$



$C(V_1) = 1$ (every path goes through V_1)

C (other) = 0 (no path goes through them)

* ~~If you want to compare to other~~ n/w, you need to normalize it.

6) Closeness : Take a node, find its ~~distance~~ shortest path to all other & take avg. Do this for each node.

Choose node with min. measure.

Teacher's Signature

All 6 → ~~ø~~ for specific node

7) Group : need to take set of vertices & generalize the def^n's.

Connectedness

→ using (any) traversal, find components
→ In social n/w : use BFS because it will reach nbs 1st (which are more imp)

→ Alternative : (if we've small graph)

Block Diagonal matrix

$$\begin{bmatrix} \boxed{\begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{matrix}} & \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \boxed{\begin{matrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}} \end{bmatrix}$$

Can conclude :
$\{v_1, v_2, v_3\}$ → 1st connected comp
$\{v_4, v_5, v_6, v_7\}$ → 2nd " "

→ can conclude from adjacency matrix (by finding no of non-zero sub-matrices)

↓ useful in case of small graphs.

* Traversal is better

Transitivity (Barabasi)

a R b
b R c $\Rightarrow$ a R c

* A friend of my friend is my friend

→ Cluster Coefficient (to find density / sparsity using transitivity)

Local
Global

① Local : measures the nbs level connectivity of a node
⇓
Find your friends & how your frnds are connected to each other

PAGE NO.:
DATE: / /

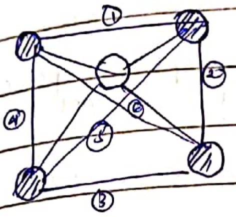
$$\boxed{C_i = \frac{2L_i}{k_i(k_i-1)}}$$

links b/w k_i

$L_i \rightarrow$ no. of nbs i has

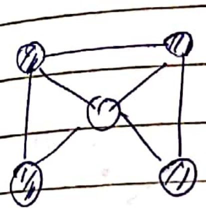
$k_i \rightarrow$ degree $\Rightarrow$ no. of nbs. $= k_i$

max. links possible among nbs. $= \frac{k_i(k_i-1)}{2}$



$\rightarrow$ All are connected

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



$\rightarrow$ How many edges among nbs in original graph (leaving centre node) = 3

max. possible = $^4C_2 = 6$

$$C_i = \frac{3}{6}.$$

* In social n/w, graphs have a certain pattern which is measured using this Local cluster Coefficient

Ex. A twitter handle : linked with Jio (Customer Service) —(1)

Another — : personal account. —(2)

We build nodes as users & make an edge only when they are communicating

LCC (1) < LCC (2)

⇓ People interacting with customer care may not communicate with each other. (like star graph)

⇓ People may interact with each other also

Teacher's Signature.........

$C_\Delta = \dfrac{\text{No. of closed paths}}{\text{No. of closed paths} + \text{No. of open paths}}$ (all of length 2)

PAGE NO.:
DATE: / /

→ Avg. $LCC_{graph} = \dfrac{\sum LCC_i}{\text{no. of nodes}}$

↓

$\langle C \rangle$

Graph (PPI) : distance travelled b/w any 2 nodes is mostly 6 hops. (with most probability) (even in millions of nodes)

⇓

small world phenomenon ⟶ Scale free graphs

w.r.t degree : As degree ↑ (nbs are ↑↑), C(k) ↓ ⇒ nbs are not connected to each other

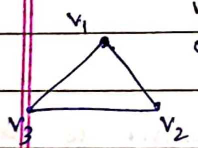
↳ shows specific pattern in graph. (not random)

• More no. of friends, interaction between them ↓↓.

② Global : One value for whole graph.

transitivity

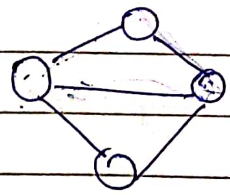
$\boxed{C_\Delta = \dfrac{3 \times \text{no. of triangles}}{\text{no. of connected triplets}}}$ → (closed)
→ Take any 3 vertices (connected) ↑ also included
→ (open or closed)



considering no. of paths with length 2 :
Possible paths : $v_1 \to v_2 \to v_3$
$v_2 \to v_3 \to v_1$
$v_3 \to v_1 \to v_2$

⟶ That's why we multiply with 3. (if we take dirⁿ into account ; 2 in numerⁿ & denominⁿ will get cancelled ans. will be 3 only)

24/1/19



$= \dfrac{3 \times 2}{8} = \dfrac{6}{8} = \dfrac{3}{4}$

↓

total paths of length 2.

* It measures the transitivity in the graph.

* C_Δ & $\langle C \rangle$ are different (Barabasi)

Teacher's Signature..........

$Tr(A^2)$: summing up all

Reciprocity

- comes into picture only in ~~un~~directed graph
- considering closed loop of length 2
- If you're my friend, me too

$$R = \frac{\sum_{i,j\, i<j} A_{i,j} A_{j,i}}{m/2}$$

no of edges

max^m R possible = $\frac{m}{2}$ → no of edges

$$= \frac{2}{m} \sum_{i,j\, i<j} A_{ij} \cdot A_{j,i}$$

$$= \frac{2}{m} \times \frac{1}{2} Tr(A^2)$$

→ we only concentrate on $A_{1,1}, A_{2,2} \ldots$

↳ diagonal pe hi aayega

& ÷ by 2

A→B & B→A ↓ diagonal pe hi 1 aayega.

(A → B, B → A : One
B → A, A → B : Taken again)

That's why, we've to divide by 2.

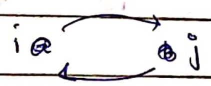
$$\boxed{R = \frac{1}{m} Tr(A^2)}$$

20/1/19

a_{ii} : starts from i & ends up at i

a_{ii}^2 : will be of length 2.

dividing by 2 →



for a_{ii} & for a_{jj} : getting 2 times.
so, have to divide

Balance and Status

→ **Balance** :

- Relationships can be positive or negative
- Not friend doesn't mean you are ~~d~~ enemies.
- +ve edge from v_1 to v_2 ⇒ v_1 considers v_2 as friend (edge is undirected, but has signs)
- -ve ⇒ " " " enemy

→ **Status** : talks about directed graph

- if $(v_1) \longrightarrow (v_2)$: v_1 considers v_2's status is higher than v_1.

(Ex. : He is richer than him, or so. → Use status in that case)
He is manager or employee under him, ...

PAGE NO.:
DATE: / /

→ can talk about how much balance is in the society ?
(equilibrium or breaking point)

→ Status : Is there an inconsistency
due to imbalance / due to status

Social Balance

↳ discusses consistency in friend / foe relationships

enemy of your enemy is friend ✓ or enemy ✗

↳ Find how many (in) consistent triads are there.

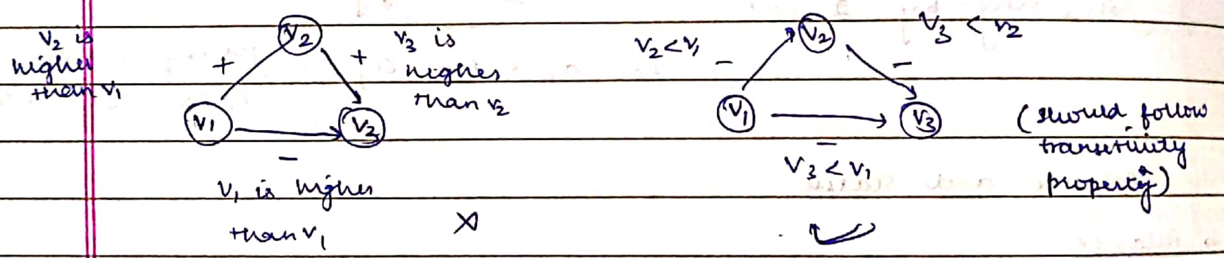
↳ Can have threshold : If ratio $\left(\frac{\text{balanced}}{\text{unbalanced + balanced}}\right) > \_\_$ → balance in society
(ex. 50%)

→ Product of all signs : +ve ⇒ balanced ; -ve ⇒ imbalanced (in case of Δ) [can't generalize about polygon]

Real life appln for social graphs

Social status ↳(more a global kind of phenomenon)

↳ If X has ↑ status than Y & Y has ↑ than Z , then X should have ↑ status than Z .



↳ In a cycle of n nodes, where n-1 consecutive ~~nod~~ edges are +ve & last edge is -ve , social status theory considers the cycle balanced.

Similarity

Check whether 2 nodes are similar in nature.

N/W similarity: Structural (wrt n/w) {I've 3 friends, he also has " } ; Regular Equivalence

Content similarity ↓ more qualitative

Teacher's Signature ....

PAGE NO.:
DATE: / /

→ finding sim. b/w 2 twitter user : N/W
" content of the tweets : Content

Structural Equivalence

↳ look at nbs of the 2 nodes & find overlapping b/w them (common nbs)

$$\sigma(v_i, v_j) = |N(v_i) \cap N(v_j)|$$

↳ varying too much for graphs (if graph is ↑↑)
↓ use (to normalize)

$$\sigma_{jaccard}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|} \quad \text{— (1)}$$

$$\sigma_{cosine}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{\sqrt{|N(v_i)|\,|N(v_j)|}} \quad \text{— (2)}$$

For very large graph, value of denom^{r} in (1) will be ↑ & thus ratio will be ↓↓. Thus, (2) is better than (1) (value: (2) > (1))
denm : almost equal to 1 of values in num^{r}

To find whether calculated measure is significant or not
↓
in random graph what is $\sigma(v_i, v_j)$
↳ Compare $\sigma(v_i, v_j)$ with expected value of $\sigma(v_i, v_j)$.
If too much diff. in 1 side } → can conclude something
~ — ~ ~ other ~

{ Talking about 1 graph (don't need norm^{n} : (1) or (2))

$|N(v_i) \cap N(v_j)|$
↓ i^{th} row, ↓ j^{th} row in Adjacency matrix
if both 1 : add

degree of i^{th} node
~ — j^{th} ~

$$\sigma_{significance}(v_i, v_j) = \sum_k A_{i,k} A_{j,k} - \underbrace{\frac{d_i d_j}{n}}_{\text{expected value}}$$

Teacher's Signature ....

PAGE NO.:
DATE: / /

$$= \sum_k A_{i,k} A_{j,k} - n\left\{\frac{1}{n}\sum_k A_{i,k} * \frac{1}{n}\sum_k A_{j,k}\right\}$$

$$= \sum_k A_{i,k} A_{j,k} - n\bar{A}_i \bar{A}_j$$

$$= \sum_k A_{i,k} A_{j,k} - \bar{A}_i \bar{A}_j$$

$$= \sum_k A_{i,k} A_{j,k} - \bar{A}_i \bar{A}_j - \bar{A}_i \bar{A}_j + \bar{A}_i \bar{A}_j$$

$$= \sum_k (A_{i,k} A_{j,k} - A_{i,k}\bar{A}_j - \bar{A}_i A_{j,k} + \bar{A}_i \bar{A}_j)$$

$$= \sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j) \longleftarrow \text{correl}^n \text{ b/w } A_i \,\&\, A_j$$

$$\sigma_{pearson}(v_i, v_j) = \frac{\sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j)}{\sqrt{\sum_k (A_{i,k} - \bar{A}_i)^2}\sqrt{\sum_k (A_{j,k} - \bar{A}_j)^2}}$$

$= 0$ → No rel^n

$= 1$ → positively correlated (significantly diff' in +ve dir)

$= -1$ → -vely " (" " " -ve ")

~ 1 : almost similar

~ -1 : " dissimilar

~ 0 : random

Regular Equivalence (Zafrani)

→ we don't look at nbd shared b/w 2 nodes, but how nbd themselves are similar

→ If 2 nodes have to be similar → their nbd has to be similar
↓
their nbd has to be similar (Recursion)

2 parameters are varying here

$$\sigma_{regular}(v_i, v_j) = \alpha \sum_{k,l} A_{i,k} A_{j,l} \, \sigma_{regular}(v_k, v_l)$$

α ↓ parameter; $A_{i,k}$ ↓ nbs of i^{th} node; $A_{j,l}$ ↓ nbs of j^{th} node; $\sigma_{regular}(v_k, v_l)$ ↓ σ of nbs.

Teacher's Signature.........

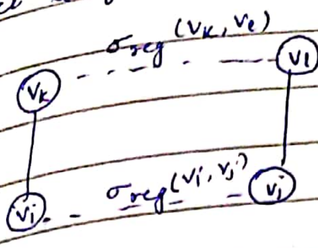
Read ex
Next s
(Strong w
reduce
Ex 3.15

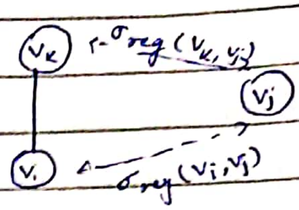
Read examples in Zafrani (3.3)

Next section : N/w, Crowds & Markets (Book)
(Strong Ties & weak Ties)

PAGE NO :
DATE : / /

reduce to formula so that only 1 variable is there :





Original formlutation

Relaxed formul^n

$$\sigma_{regular}(v_i, v_j) = \alpha \sum_k A_{i,k} \, \sigma_{regular}(v_k, v_j)$$

In vector format :

$$\sigma_{regular} = \alpha A \, \sigma_{regular}$$

$$\sigma_{regular} = (1 - \alpha A)^{-1}$$

added 1
because every
node is similar to itself

$\alpha < \frac{1}{\lambda}$

(like Katz)

Ex. 3.15

leave main diagonal

conclusion : 0.73 ~ 1 : v_1 is sim to v_2

0.80 : v_2 ~ v_3

0.73 : v_1 ~ v_3

v_1, v_2, v_3 are ~~similar to each other~~

from graph : v_2 & v_3 : similar
v_1 is almost similar] structural

Teacher's Signature..........

29-1-19

PAGE NO.:
DATE: / /

Ties and Triadic closures
↓
links/edges

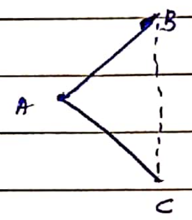
→ Triadic Closures:

If 2 people in a social n/w have a friend in common, then there is an ↑ed likelihood that they'll become friends themselves at some point in the future.

→ Connections:

1) structural (more structured)

2.) Interpersonal (closely connected)



Reason for Δdic closure:

1) Oppurtunity a common frnd A gives for frnds B & C to meet
2) Trust factor is good if we've common frnd
3) Incentive that A brings - latent stress if B & C are not frnds.

Bridges and Local Bridges
↓
~~Cut-edge~~
⇓

is deleted, g n/w is disconnected

↳ Finding a bridge in social n/w (very ↑) is very rare.

Figure : if A & B ka edge x, it won't be a bridge, but a Local bridge (if edge is cut, distance to be travelled to reach B from A will ↑} more than by 2)

Local Bridge : if A & B have no common friends Or, if deleting edge will ↑ distance b/w A & B by strictly more than 2

Teacher's Signature.........

Scanned by CamScanner

Two types of ties
↳ Strong ties
(close friendship)
it is ↓ for you to define acc. to the n/w (quantifiable manner)

↳ Weak ties
Weight associated with the links ↓↓.

Strong Triadic Closure Property : (STCP)

If node A has edge to B & C then B - C edge is especially like to form if A's edges to B & C are both strong ties.

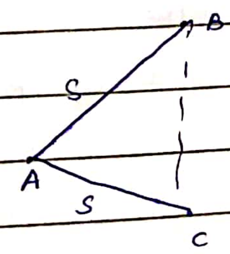
↳ A node violates STCP if it has 2 strong ties with B & C & ∄ no edge at all b/w B & C
→ If you try it, graph will almost be complete

↳ STCP is too extreme. → Thus, it'll become too extrem
We will try to weaken ~~loosen~~ it a bit

Local Bridges and Weak ties (we're trying to connect local property to global property)
Weak Tie (talking abt nodes)
↓
global property (talking about graph)

Claim If node A satisfies STCP & is involved in atleast 2 strong ties, then any local bridge it is involved in must be a weak tie.

Weaker property, talking just about a node, not the whole graph



suppose A - B : local bridge → strong tie
⇒ A & B must've no frnds in common
⇒ B - C must not exist
but STCP states B - C must exist } → Contradiction
⇒ ~~A-B~~ local bridge must be a weaker tie.

N/W Models → Random Graph

PAGE NO.:
DATE: / /

Generalizing notions of Weak Ties & Local Bridges

Generalizing notions of weak ties to be ratio

- Define nbd overlap of A – B edge
  - no. of $|N(A) \cap N(B)| = n$
  - no. of nbs of atleast A or B $|N(A) \cup N(B)| = m$
  - nbd ratio $= \frac{n}{m}$

↳ Nbd ratio for bridge : 0 $(|N(A) \cap N(B)| = 0)$
local bridge : may be very close to 0.

↳ for ratio very small : ~~local bridge~~ "almost" local bridge

Graph : when Tie Strength ↑, Overlapping b/w nbd also ↑es. (linearly)

Empirical Results

↳ Weak ties serve to link together diff. tightly ~~[illegible]~~ -knit communi[ties] that each contain a large no. of stronger ties.

(If you know about a job, your friend will also be knowing ab[out] it. But an acquaintance can provide you a job links which you don't know about)
↳ weak ties

(PhD students go out to get some new information)

* Weak ties are more imp in global sense & Strong ties in local sen[se]
Both are imp. for our social n/w.

* If there is a weak tie, there is a high prob. that it may be a local bridge.

Teacher's Signature

DATE: / /

Q. segment sizes are: 1250 bytes (n_f) & 25 byte overhead & 25 bytes ack. R = 1 Mbps. $2\times(t_{prop} + t_{proc})$ = i) 1 ms, 2) 10 ms, 3) 100 ms. Find effeciency of stop & wait protocol.

• Try for R = 1 Gbps also.

i) $$R_{eff} = \frac{n_f - n_o}{2\times(t_{prop}+t_{proc}) + \frac{n_f}{R} + \frac{n_a}{R}}$$

$$= \frac{1250 - 25}{1\,ms + \frac{(1250+25)\times 8}{1\times 10^6}} = \frac{1225\times 8}{1\,ms + 10.2\,ms} = 875\times 10^3$$

$$\eta_{sw} = \frac{1250-25}{1250+25+1\,ms\,(1\times10^6)} = \frac{1225\times 8}{1275\times 8 + 1000} = \frac{9800}{11200} = 0.875$$

ii) 10 ms

R_{eff}

→ Considering with errors:

• The effect of Transmission Errors on the effeciency of S&W ARQ

1) If a frame/segment incured errors during transm^n, the time out Mech will cause retransm^n of frame/segment. In that case, t_o is the transmis^n/retransm^n time (sec)

P_f → probability of segment transmitted has errors & it needs retransm^n

Then,

$$\text{Avg. succesfull Retransm}^n = \frac{1}{1-P_f} \quad \text{(1 in 10 segments)}$$

$$t_{sw} = \frac{t_o}{1-P_f} = t_o\left(\frac{1}{1-P_f}\right) \quad \text{——— (I)}$$

time for Stop & wait

$$\boxed{\eta_{sw} = \frac{(n_f - n_o/t_{sw})}{R}}$$

Teacher's Signature........

PAGE NO.:
DATE: / /

Que. same values as prev. quesⁿ, for 1ms only.
P_f = 10^{-6}, 10^{-5}, 10^{-4} → Bit error rate

i) $P_f = 10^{-6}$

$$t_{sw} = \frac{t_0}{1-10^{-6}} = 2\times(t_{prop}+t_{proc}) + n_f/R + n_o/R \left[\frac{1}{1-10^{-6}}\right]$$

• To calculate the Transmⁿ effeciency for S & W ARQ :

1) We need to calculate an avg. total time req. to deliver correct frame.

let n_t → i transmⁿ's are req. to deliver correct segment/pkt
→ (i−1) retransm'n's are req. ⇒ (i−1) TO are req.

$$P[n_t = i] = (1-P_f)\cdot P_f^{\,i-1} \quad \text{for } i = 1,2,3\ldots$$

↳ i-1 times, need timeout

$$\underbrace{E[t_{sw}]}_{\text{expected } t_{sw} \text{ time}} = t_0 + \sum_{i=1}^{\infty} (i-1)\, P[n_t = i]\, t_{out}$$

$$= t_0 + \sum_{i=1}^{\infty} (i-1)\, t_{out}\, (1-P_f)\, P_f^{\,i-1}$$

$$= t_0 + \frac{t_{out}\, P_f}{1-P_f}$$

if $t_{out} = t_0$:

$$E[t_{sw}] = \frac{t_0}{1-P_f}$$

→ that's how, we get t_{sw} in ①

Teacher's Signature.........

Concurrent & Parallel Computing → Pipelining

PAGE NO.:
DATE: / /

A system app. model : CN by Peterson & David

Quesⁿ: Bandwidth = 50 Kbps
1 way transmⁿ time = 250 ms. Find out Bandwidth delay product (BDP).
Window size = W = 2 (Bandwidth delay) + 1
Find link utilization

delay = 250 ms +

↳ Drawbacks:
- For single pkt, we're waiting
- To very huge

↓ solⁿ

Use pipelining concept
~~need not to wait till a task is completed.~~ Can be introduced while a task is going on.
Need to choose a buffer which will be transmitted.

- Sliding window Protocol : used in GoBack N & Selective Repeat

GoBack N → Problem: whenever error, whatever successful pkt have been sent needs to be retransmitted

Selective Repeat → send ack for pkt which is lost.