after MMSEM

apsara

Date:

12/3/19

Evolution of Rand. N/w :

NG : Hub : Giant component : biggest connected component.

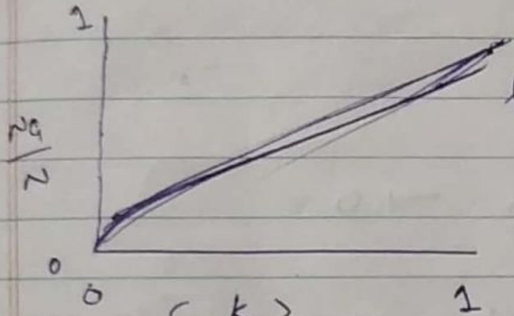
How size of NG varies, when we inc. $\langle k \rangle$ from 0 to $N-1$:

$\langle k \rangle = (n-1)p$

extreme cases

if $\langle k \rangle = 0$ then $p = 0$. – all nodes isolated
$NG = 1$, $\frac{NG}{N} \to 0$.

if $\langle k \rangle = n-1$ then $p = 1$ – complete graph
$NG = N$, $\frac{NG}{N} \to 1$.



linear (expectation of how this ratio varies with $\langle k \rangle$)

But this is not the case!

$\langle k \rangle$ when small → lacks large structure, ratio → 0
↳ isolated pairs of vertices joined together.
↳ no big clusters.

Once, $\langle k \rangle$ inc. critical value → NG/N inc.
behaves in diff. way.

$\frac{NG}{N}$ → non linear fashion, grows not gradually, but substantially.

$\langle k \rangle = 1$

Giant component $\Leftrightarrow$ each node has avg. more than 1 link

$\langle k \rangle$ → does not depend on no. of nodes in the network.

Page No.

Scanned by CamScanner

$N_G \rightarrow$ giant Component.

You need to have at least one edge to make that node connected.
↳ 1 link is sufficient for a giant comp. to emerge.

$$\underline{\underline{1}} = \langle k \rangle = (n-1)\,p$$

$$\boxed{p = \frac{1}{n-1}} \rightarrow \text{inversely proportional}$$

~~Giant~~ Huge no. of nodes → less probability → we can get a giant component.

<u>Graph</u>

$\langle k \rangle = (0, 1) \rightarrow \frac{N_G}{N} \rightarrow 0$

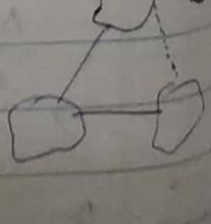
$\langle k \rangle = (1, 2) \rightarrow \frac{N_G}{N} \rightarrow$ non-linear fashion

$\langle k \rangle > 2 \rightarrow$ almost giant component.

$\langle k \rangle = 6 \rightarrow \frac{N_G}{N} \rightarrow 1$

Substantial increase

$\langle k \rangle$ beyond 4 → <u>giant component</u>



ratio

Subcritical region: < 1 $\longrightarrow$ 0 $\rightarrow$ grey part

critical pt: $\langle k \rangle = 1$

$\langle k \rangle = \log N$ } ~~yellow part. $(1, \log N)$~~ supercritical region:

Connected regime: $> \log N$ $\rightarrow$ blue part $\rightarrow$ ratio $\rightarrow 1$

$\downarrow$ every node will be in the giant component.

(avg degree $= \log N$ will be sufficient for this)

When you inc. $N \uparrow$, N_G inc. slowly ($\log N$).

$\langle k \rangle < 1$ $\rightarrow$ there might not be any ~~a~~ cycle they will be connec. like forest (4-5 nodes). many small trees.

critical pt. $\langle k \rangle = 1$ $\rightarrow$ form cycles.

- may get a giant comp.

$N_G = N^{2/3}$

~~two third of the vertices will be in giant comp.~~

still ratio will be zero $\rightarrow$ $\frac{N_G}{N} = \frac{1}{N^{1/3}} = 0$ when $N \to \infty$

Ex: $\langle k \rangle < 1$ $\longrightarrow$ $\log(7 \times 10^9) = 22.7$.

when $\langle k \rangle = 1$ $\longrightarrow$ 3×10^6 — jump of 5 orders.

substantial jump

super critical $\langle k \rangle > 1$: $(p > \frac{1}{N})$.

almost relevant to real n/w.

Vicinity of critical pt : $\frac{N_G}{N} \sim \langle k \rangle - 1$.

Away from critical pt : $N_G \sim (p - p_c) N$

$\left(p_c = \frac{1}{N}\right) \rightarrow$ critical pt.

$\langle k \rangle > \log N$ — Connected region
|
Every node is connected

$N_G \cong N$

Here, Avg. degree depends on N.

$\frac{\log N}{N} \to 0$

Complete graph when $\to \langle k \rangle = n-1$.

$p = N^{z}$ $z = (-\infty, 0)$
$\downarrow$ $\downarrow$
$p=0, p=1$

$-\infty \to$ isolated
$-3/2, -4/3, -5/4 \to$ trees
$-1 \to$ cycles
:
further $\to$ more complete graphs

Imp $\to$ empirical obs. & no proof or logic

In real n/w, even when $\langle k \rangle < \log N$, we get connected n/w.
In rand. n/w when $\langle k \rangle > \log N$, ~~then~~ we get a connected N/W.

Internet - fully connected giant component,
$\langle k \rangle < \log N$
substantially less

$\langle k \rangle = 1 \rightarrow$ independent of N. apsara

but for $\langle k \rangle = \log N \rightarrow$ dependent on N Date: __.__.__

Erdos Renyi showed only when $\langle k \rangle > \log N$ then n/w is connected (giant comp.), in real n/w, we see difference.

Real N/w are Supercritical: (yellow region)

$X \rightarrow \langle k \rangle$ depends on N

when N ↑, log N depends accordingly

⇒ Why to study Rand. N/w:

to show that Rand N/w ≠ Real n/w.

In 2000, Barabasi showed that

for real n/w — Base pt. to refer to → Rand. N/w. → not same

→ it is Scale free N/w.

14/3/19

Small World Phenomenon:

Neural world, Brain research — comparison b/w an abnormal brain and an average brain.

Comparison of rand. n/w and real n/w w.r.t. different phenomenon:

1. Degree Distribution.
2. Small World Phenomenon.

Erdos → like a star, more than 500 co-authors., that's why erdos no.

↳ lived a nomadic life.

apsara

Date:

→ Graph: In around 5 intermediates, we are able to reach most of the targets.

→ WWW 2008, MS messenger.

→ If tree like str. with 5 friends / node
↳ triadic nature - small no. of hops.

→ In real n/w, triadic closure reduces hops

Diameter → max. length n/w should have:

$\langle k \rangle$ — one node is connected to k no. of nodes (on an avg.)

with 2 hops : $\langle k \rangle^2$ possibilities. → no. of nodes on level 2.

with d hops : $\langle k \rangle^d$ nodes.

avg. (capacity)

$$N(d) = 1 + \langle k \rangle + \cdots \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$$

no. of nodes it is covering. (all nodes covered : N)

$$N(d_{max}) = N$$

$$\langle k \rangle^{d_{max}} \approx N$$

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

→ it holds good in Rand. n/w, but not in Real Network.

if k fixed, vary N, $d_{max} \propto \log N$.

$$d \approx \frac{\log N}{\log \langle k \rangle}$$

→ average length in the graph will be similar to it. (Not diameter, avg length)

fb → someone has 50000 friends, some have 1,2 friends.
↳ not seen this behaviour in Rand. n/w.

→ average length : considers all avg. length
↳ central figure $\approx \log N / \log \langle k \rangle$

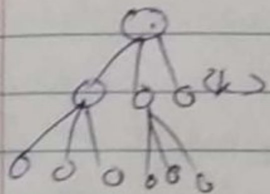
these extreme cases will give big diameters : affected by outliers

Diameter(max) is distorted due to some external factors also, so here, we talk about avg. length and not diameter.
(d)

Barabasi Table 3.2

$\langle d \rangle$ → avg. length b/w any 2 random vertices.

$\langle d \rangle \approx \ln N / \ln \langle k \rangle$ (no. of hops)



→ it is perfectly balanced, but it is not so in real graph.
↳ Rand. n/w

that's why d_{max} of Rand. n/w $\approx \langle d \rangle$ in real.

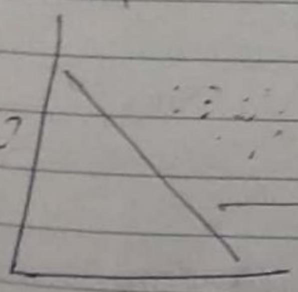
In Real life, friends of a nodes also might be friends themselves.
↳ Clustering coeff. : measure of transitivity.

Every link has prob. 'p'.
total no. of links b/w neighbours : $\frac{k_i(k_i-1)}{2}$.

$$\langle L_i \rangle = p.\frac{k_i(k_i-1)}{2} \quad , \quad p = \frac{\langle k \rangle}{N}$$

** $$C_i = \frac{2\langle L_i \rangle}{k_i(k_i-1)} = p = \frac{\langle k \rangle}{N}.$$

$\frac{1}{N} = C\langle k \rangle$



→ not following

→ expected $\left(\frac{1}{N}\right)$ behaviour.

expected clustering coeff.

no. of nodes

→ Clustering Coeff is independent of N, in Real Network

For real n/w, clustering coeff does not depend on N.

apsara

Date: ___

Watts - Strogatz Model

Small World (6) → connecting to a node outside their group.

Almost a regular lattice: close to real n/w.

grid gives regularity which captures the triadic closure of n/w but doesn't capture the weak ties

Ab: 19/3/19

In random n/w a super critical region ones when $k > 1$ & $k < m.n$. but in real network the isolated nodes in super critical region is in giant component - it differentiates both.

2 Small World Phenomenon
(The graph will be very sparse)
but it is connected and it takes for any two nodes that will be connected through 5-6 hops

Small World Phenomenon
has
weak links & strong links

The real n/w are not random n/w but it has small world phenomenon. Can you proof it has small world (to model it some thing which has small world phenomenon).

Watt Strogatz Model :

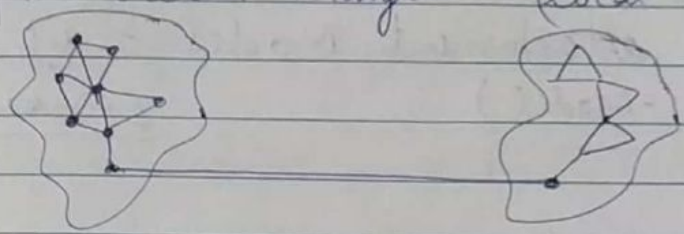
The main reason real is not random - we are not having the transition (treads) in rand. n/w

Page No.: ___

apsara

Date: ______

Homophili (close friends)
→ weak ties (global nature) as it connect to outer n/w
→ Homophili creates triangles -(local in nature).



→ A grid will solve the homophili but not capture weak tie.

** { regularity is homophili
randomness will capture the weak ties. }

⇒ till then grid also it is not equivalent to real n/w.
Milgram — decentralized search — every part where person wants to send optimizes a certain strategy.

Kleinberg n/w Random 2000

This random (wear ties) not capture by grid model
Two random nodes will be connected with a prob.

$\frac{1}{d(v,w)^q}$ $q = 0$, $\rightarrow$ 1 (same as grid)

q small you are connected to your friends (not random friends, you are connected together).

$q = 0$ to ∞ $q = 1.5$ to 2

$d^2 \cdot \frac{1}{d^2} = 1$.

Blogging - Medium.

→ Distance in graph:

(i) geographical dis.
(ii) closest friend.
(iii) 2nd level of friend.

Model which is generative in nature - closer to real world

Date:

26/3/19

WWW Network : testbed for many models.

→ Preferential attachment Model - defined by Barabasi and student Albert
(BA - Model)

→ Scale free Network :

www → imp. in development of n/w theory.

started with it as a rand. N/w, but then they observed it is not. — testbed for most n/w

main diff b/w rand. n/w and www n/w.

Graph: green: poisson
straight → real graph

for www: directed graph
k_{in}, k_{out}.

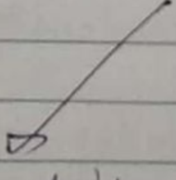
Pareto distbn : power law distbn.

$p_k = k^{-\gamma}$

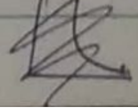
$\log p_k \sim -\gamma \log k$

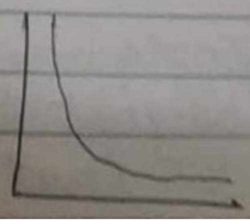
$\gamma_{in} = 2.1$
$\gamma_{out} = 2.45$

we have taken log hence the straight line with slope $-\gamma$



if we don't take log then power la distbn will be like:





Page No.

- can't avoid it when society is like that

Pareto principle - (80/20 rule) (management)

↳ 80% of wealth is with 20% ppl

→ path b/w any 2 nodes in scale free - may not have many hops.

→ More no. of nodes will have less degrees - real n/w ✓

In poisson → ⟨k⟩ will have most no. of nodes — poisson ✗

↳ not observed in real n/w.

Q) www is scale free n/w. - PL Distrbn.
Many other n/w are also scale free, how they converge to a sim. scale free n/w?

→ Erdos Renyi → G(n, p)

p fixed. (nodes: fixed)

↓

but no. of nodes ↑se

in www → nodes ↑se from 1 to trillions -
- or citations → " " .

The nodes having high degree : rich
and rich becoming rich.

- Growth of the n/w : Erdos and Renyi did not look at
- Preferential attachment.

apsara

Date: __-__-____

* BA Model

→ m_0 nodes, each nodes has atleast 1 degree

Growth : add a new node with m links

Preferential attachment : prob. of association (a node v_i with degree k_i) → $\frac{k_i}{\sum_j k_j}$

if k_i ↑ then more prob.
↳ rich becomes richer.

Drawbacks : - not explain the initial graph to start with

- if nodes have to be added 1 by 1 or all at one time ?

↓

Refined by LCD : - start with a graph with no nodes $G(0)$

Given $G_1(t-1)$ → build $G_1(t)$ - by adding node $v(t)$

t : iteration → adding $v(t)$ node

v_i → already in the graph :

to join v_i and v_k prob :

$$\begin{cases} \dfrac{k_i}{2t-1} & \text{if } 1 \le i \le t-1 \\ \dfrac{1}{2t-1} & \text{if } i = t \end{cases}$$

Connecting node to a node already in the graph.

very less. (self loop)

C : clustering coeff

Simulated Graph : BA model

Avg. Path length is almost the same
[Small World Phenomenon] ✓
but Clustering coeff is different.

drawback

BA model is able to capture small world phenom by Growth and Preferential attachment but it is not able to produce C → triadic closure (the triangles it is not able to produce).

Fitness value of each node.

BA → rich becomes rich → earlier alta-vista search engine was very popular but then when Google came → it turned up side & down.
↳ rich becoming rich not true any more
↳ new entrant was rich now.
Fitness value of BA went down.

prob. of associating a new node with already existing n/w → need to consider both fitness values.
↳ Bianconi - Barabasi Model

⇒ Information Diffusion

→ Community Detection.

→ Through BA, we are able to produce Scale Free Graph.

BA follows approach : rich becomes richer.

↳ Eg: Windows → becoming popular and popular

↳ by this model who ever comes later on will not become popular at all.

apsara

28/8/19

Bianconi Barabasi Model

— assoc. one more value with each of the nodes : FITNESS

η_i : Fitness value

k_i : Degree

Weighted avg : $\frac{\eta_i k_i}{\sum_j \eta_j k_j}$: Preferential Attachm

Eg: NOKIA v. popular earlier → but after few years it vanished.

(external factors)

Fitness takes care of exogeneous / external factors.

Degree " " endogeneous / internal factors.

Information Diffusion in Social Networks:

Oreo → reached a lot of people without spending a penny!

Online firestorm : a neg. message of a coporate becoming viral.

info. spreading through physical / non-physical interactions.

Sociology, Epidemiology : like spread of disease, enthnography : ~~E~~ culture study.

Modelling the Communication

Page No.: ____

apsara

Date: __/__/__

When info spreads : it has a source ① senders

↳ ② receives : who receives diffused info.

In Diff. Info → set of all recv. >> set of all senders.

③ Medium : to pass the info - personal communication, or social media sites.

Types :

n/w is observable here

① Herd : $\boxed{\text{global dependence}}$ - based on other's ppl decision - they will take their decision
└ what your friend or nbg. is doing, not based on proper rational thinking.

② Info. Cascades : Just observe their nbg , in prev. case we were looking globally, at all ppl, here we just look at our immediate nbgs.

↳ Both are kind of similar.

How many ppl adopted new technology ?
└ network is not visible here, we only know rate of adoptibility of the new technology that is

③ Diffusion of Innovation.

↳ They want to diffuse the innovation at very fast rate to become popular.

④ Epidemics - similar to ③, pathogen - transmitting from one individual to another ind.

apsara

Date:__.__.__

In diffusion of innov. → you make a decision to buy a smart phone, but in epidemics you aren't making a decision, pathogen if is in your body then it will do whatever it wants.

Explicit : N/w is observable, in
Implicit : " " not "

→ Eg. of Herd Behaviours : Auction - bidding amt. of all the others (global)
(reputation and expertise also comes into picture.
Amt. of trust is also there in bidding.

→ If you want to make a decision then we go with a known devil rather than unknown devil → herd behaviour → wherever crowd goes → its evolutionary impact → not completely irrational

Eg. → Soloman Asch's Expt., - elevator video.

→ Herd behaviour → common factor - global pressure (public)

more like a complete graph.
- has to be some decision made, in seq. order.

apsara
Date:___.___.___

HERD BEHAVIOUR :- (4 conditions)

→ Decisions are not mindless → it is based on what you observe.

- no message passing is possible : Public info.

→ Urn → 3 marbles - red or blue
2 red, 1 blue or 2 blue, 1 red
decision → publically visible.

if the first two decisions are BB -- then every next dec. will be B only.
$\underbrace{BB}$ prob : > ½ = ⅔ ✓.

Bayesian Model : to study Herd behaviours.

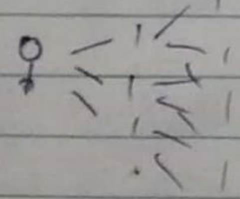
2/4/19

new slide

Herding Intervention :

↳ To stop it, you need to intervene, by passing some unknown information. - Rumours.

→ A child shouting : emperor doesn't have any clothe was an intervention.



→ a person starts rumours, then it spreads. Graph - may not be complete.

≠ info. cascades ↔ b/w friends. (immediate)

Diff. b/w herding and cascading?

Hotmail - free mail service - "Get your free email at Hotmail" → Advertising through email itself!!
↳ It became VIRAL.

Page No.:

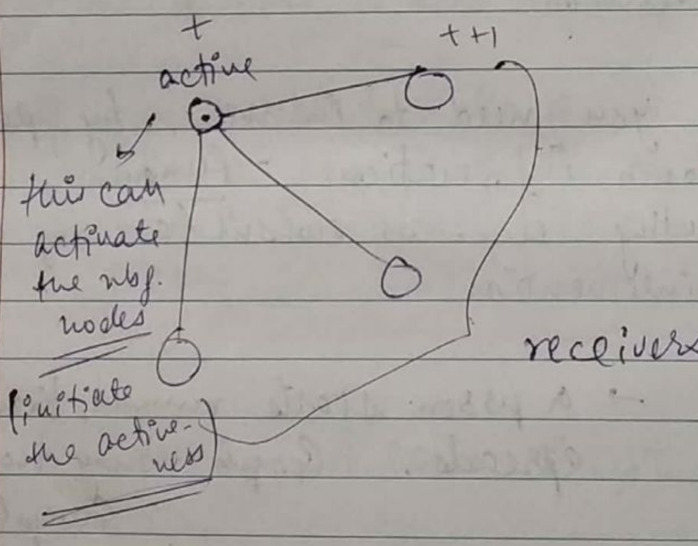
N/W → directed graph → if undirected then assume there are 2 edges - (⇄)

Decisions – binary (node)

Active node : adopted the behaviour
Inactive : has not got the info.

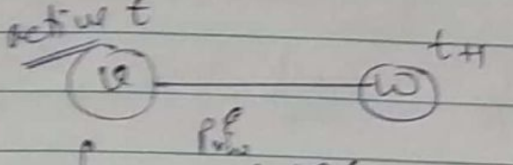
Can activate the nbg. nodes.
Activation → inactive → active, not vice versa
↓
How can you make it other way also true ? - Ppl are studying this

→ ICM : sender centric → focus on senders rather than receivers.
└ every nodes get 1 chance to activate its nbg.



Temporal → t : 1 node active, at t+1
↓
another node(s) activates

Q) max. no. of nodes that you can reach which has a particular info.?



→ p(node v activating w)

assumption: has only 1 chance to activate its nbg.

If a person is not retweeting at that time, it won't retweet that after 2 days. So, this assumption works!

Under cluster model: ICM: on an avg. what are max. no. of nodes it can initiate in a set of nodes.

Eg:- news agency ← pass info., select nodes that will help you to cascade that info.

- ppl who have good reach among public
- your reach wants to go to the max. ppl
- marketing strategy

What if we started with v_3 → ?

Q). Set of initiators to reach max. no. of ppl with less no. of steps?

Maximising the spread of cascades.

S : initial set

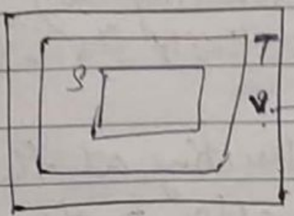
f(S) : spread
↳ maximisation problem.

S: $\underline{v_1}$ $f(S) = \{v_1, 2, 4, 5, 6\}$ (5)

Date:

$f(S) \rightarrow$ at any time it is going to ↑se.
(non-(-ve))

Submodular:
func.



$S \subset T \subset N$

$\rightarrow \forall v \in N \setminus T$

$V = N$

not in T

$f(S) \rightarrow$ set of all nodes that can reach when start with S.

$$f(S+v) - f(S) \geq f(T+v) - f(T)$$

$$f(S) < f(T)$$

↳ impact that is created on adding v to S is more than adding v to T.

Greedy set $\rightarrow$ S : Add v to S to max: $f(S \cup \{v\}) - f(S)$.

↳ 63% of optimal soln. we will be able to achieve

Intervention: – By limiting no. of outlinks from
– " " " inlinks from senders. receivers pt of view.
– By dec. activating probability. prob

objective: to spread to max. nodes

{ ① Herd
② Info. Cascading Effect } - n/w were observable - in ① & ②
- public observable in herd - ①
but not in cascading - ②

③ Diffusion of Innovations : (not only info. passing)
— an idea - that can be adopted.

different

n/ws are not observable , pub. observable
(unlike ① & ②)

objective: why and how info. spreads.

Innov. - Observable : Apple builds a new phone.
Advantage : Adv. from your competitor.
Compatible : cannot change society to completely new.
(connec. b/w old and new)
Trialability : limited basis (good)
Not Complex : should be easy to use.

Segregated the adopters!

Adoption: External as well as internal influence
(mass) — weak tie (global input) — more information
(personal) — strong tie (close friend) — more of influence not information

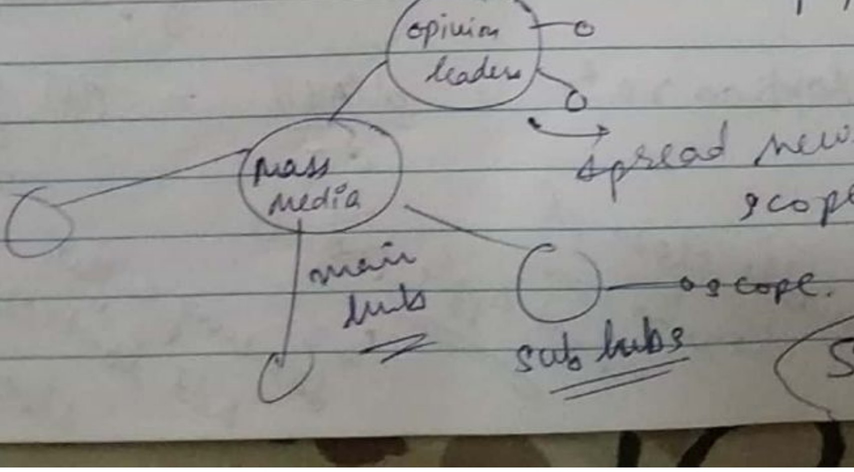
exogeneous factors - endogeneous factors

①

S-shaped {
Innovators : ppl. who innovate (min. no. of ppl)
Early Adopters : Fans of the product : immediately purchase it.
Early Majority : based on reviews - if positive
Late " : a little late
Laggards : wait for 2-3 years, then ~~wait~~ observe if product is going good.
}

② Katz

Two step



spread news to their scopes.

→ it may not have triadic closure

Scale Free Graph

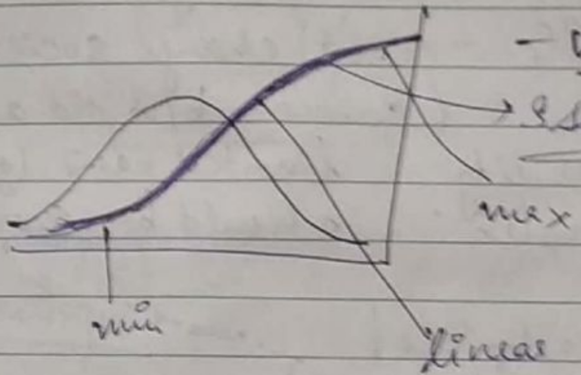
apsara

Date: ___

(3) Rogers : Awareness : a new product is there
Interest : show some interest
Evaluation : reviews, analysis.
Trial : if friend got it, then decide
Adoption : after checking

4/4/19

Marketing people → benefit from Early Adopters.



→ modeling of Diffusion of Innov :

differential eq → to find rate of growth.

$A(t)$ → ppl : adopted the product, t : time
$i(t)$ → coeff. of diffusion : innov. rate - measure
↳ we can define it. (Company - good brand value → high $i(t)$ (Apple, Google))

define this to no. of ppl who might adopt it — P → till time t → no. of potential adopters. (potential adopters)

Adoption rate $= \frac{d(A(t))}{dt} = i(t)\,[P - A(t)]$

↳ upper middle class ppl.
↳ who can afford Apple products.
& Samsung might add lowerclass ppl also.

Page No.: ___

$i(t)$ → how innov the product is.
rate affects the remaining ppl who might adopt the product.

$i(t)$:

Innov. factor

- External Influence Model → $i(t) = \alpha$
- Internal " → $i(t) = \beta A(t)$ — imitation factor
- Mixed " → $i(t) = \alpha + \beta(A(t))$.

based on product reviews - (websites) — α
(not under any peer pressure or friends influence)

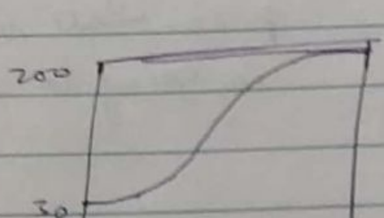
ppl who will not see any external factors, and rely on their friends - influence (Herd Behaviour)
↳ not much of rational. — $\beta(A(t))$

both : $\alpha + \beta(A(t))$

→ Internal Influence Model — Pure imitation

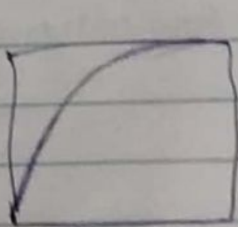
getting losses and losses

starts with 30 → then reaches max. - flat



→ Mixed →

most of the cases



apsara

Date: __/__/__

→ <u>Intervention</u> : if a faulty product - address it

Reduce interest - don't advertise

<u>Epidemics</u> : - not only spreading of disease
→ viral attack, worm attack on computers

Some cascading effects without n/w observable

→ Segment the phenomenon - classify what types among these (?) types can be combination also

→ Pathogen → virus transmitted among PCs
→ retweeting of tweets.

→ Population - can be electronic devices also

→ Spreading

Implicit n/w - sim. to innov. diffusion model

person is not going to decide whether you want to carry pathogen or not.

person who is adopting is going to decide

→ <u>Contact N/w</u> : connection - medium : air sneeze

→ <u>Fully Mixed</u>

Basic Epidemic Models :

- → SI, SIR, SIS, SIRS.
- ppl types:
  - susceptible
  - infected
  - recoverd

} based on this we have defined 4 models.

1. In SI → no recovery, like $\boxed{\text{AIDS}}$.

no. of susceptible individuals → normalized = $\frac{S(t)}{N}$.

.. .. infected .. → $I(t)/N$

β: Contact Probability.

if $\beta = 1$ → everyone contacts everyone.

if $\beta = 0$ → no connections.

$$N = S(t) + I(t).$$ – only two sets of ppl.

Model: infected ~~ppl~~ person meet βN people on an avg. and infect βS of them.

rate of change of no. of susceptible ppl → $\frac{dS}{dt} = -\beta I S$, $\frac{dI}{dt} = \beta I S$

decreasing → they are infected now. by same value, # infected ppl will increase.

rate of infection: $I(t)$ ✓.

$$S \xrightarrow{\beta} I$$

② **SIR model** : we can be cured, and then further ~~to~~ we will not be infected.

$$S \xrightarrow{\beta} I \xrightarrow{\gamma} R$$

once you get recovered, you will hold the immune.

but immune can be time dependent.

SIRS

for some time (1 year)

$$I + S + R = N$$

$$\frac{dS}{dt} = -\beta I S$$

$$\frac{dI}{dt} = \beta I S - \gamma I$$

γ → recovery rate.

$$\frac{dR}{dt} = \gamma I$$

③ **SIS model** : no recovery - you recover but become susceptible again.

④ **SIRS model**.

\# Intervention : - Immunisation

In Gujarat rat plague 90s

Mad cow disease → anyone can sell anything to anyone - animals affected

weak ties

remove weak ties → animals were killed — intervene, soil where animals were living

SMM - An Intro Slides

apsara

Date: __.__.___

11/9/19

Evolution

Community Detection / Analysis / Evaluation :

group of ind. - having common interests (economical, social, political, personal) or similar characteristics.

→ late 198 → Dutch, police sys. - bots - 30 million - sending spam mails (billions)
↳ this was found out by community detection algo
Spam filters → very accurate these days.
→ Bot detection - not very easy - find bots irrespective of geographical loc or anything, irrespective of OS, type of sys.
find an algo - which detects a bot which has some kind of characteristics.
bot detection is diff from people detection (community)

→ Positive and Negative aspects of finding a community.
group of ind. or sys.
this bot one, and sys. infected with virus.

→ Liberal / conservative
left / right / middle

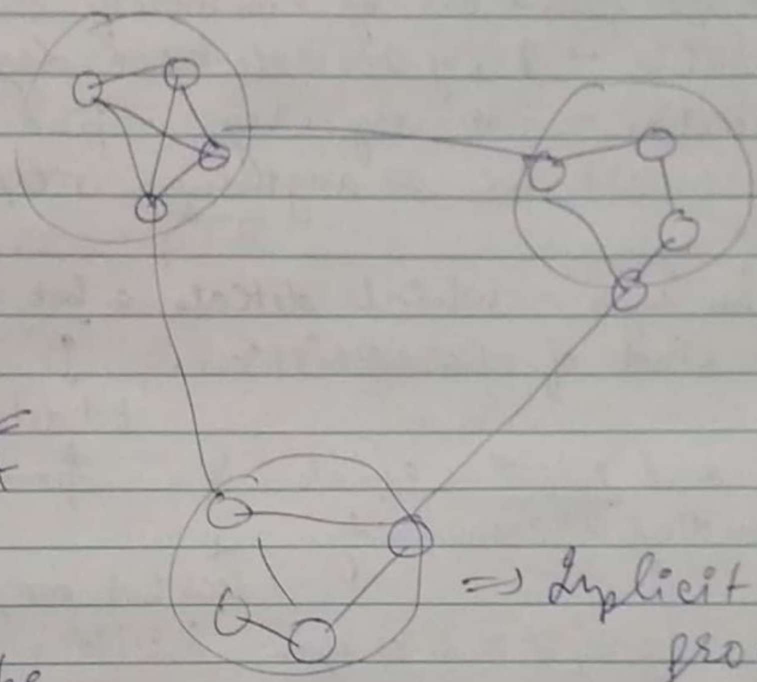
→ group atmosphere study : noise ↓
individual " " : noise ↑

→ Recomm. sys.

→ Explicit Community / Group v/s Implicit
↓ ppl who know they are forming this group, know about the members.
↳ Google group, Yahoo group.
↳ finding bots, need to find characteristics.

emic , etic

In Social media, we will find many Implicit groups. - confining with those who have similar idea only 'left right', and blocking others with diff. idea - kind of bubble.



disjoint

|

can be overlapping → set of ppl in more than 1 community

⇒ Implicit set of groups.

Amazon - recomm sys. - 1 group: action - other group: thriller, suspense ⟷ oriented these movies.

can be overlap as well

→ Protein interaction only happens when they are similar in nature.

→ www comm.: ppl having interest in same topic

→ How can you evaluate comm? - good/bad/neutral/ values it is bringing
→ Clustering method in Graphs

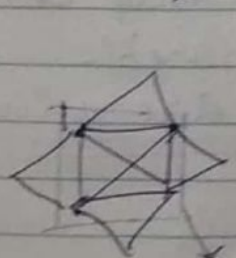
Strong connec : Internal in a comm.
Weak " : external ties.

→ Partition the graph into various components : comm.
Ideal decomposition - completely disjoint comm.

Zachary's Karate Club - from 2000, citation of this paper grew exponentially.
Know: In past 2 years, if they communicated more than 10 times then they have a link b/w them.
Tussle between trainer and admin.
PPl who had more interaction with node 1 moved with him and others " node 34.
↳ It is implicit at beginning but it ended as explicit.

→ Clustering v/s Community.

→ member based comm. | group based.
1.
nodes with similar char.
[Eg: all the nodes having k degrees. - clique
+ sub graph which is completely connected



8 nodes
clique over 4 nodes.

apsara

Date: ___

→ finding a clique – mostly used in SNA.
↳ NP Hard problem

maybe we can relax it a bit,
when n → very huge, then
if one node is connected to n-1 or n-2 --- n-k nodes then
(k-cliques) subgraph which are almost a clique → also fine.
→ T.C. : quicker
because clique is NP-Hard

reachability
k-clubs
k-clans

→ similarity measures

Q) → (maximum clique, maximal clique)
↳ one way of finding communities.

→ maximum, maximal ($\mathbb{N}$, a|b) → a div. b (a≤b)

minimum = 1 ← 1 divides everything
maximum = ∞

($\mathbb{N}$ - {1}, a|b)

~~Minimum~~

minimal → prime nos. (all) (2, 3, 5, 7 ---)
they are not minimum because they can't be compared.
All these are minimal

not comparing these ele with other eles. and can't find lesser ele. that is minimum

→ minimum ele is a minimal ele.
→ but minimal is not minimum
→ only 1 minimum but minimal ≥ 1

v_x → starting vertex. — Brute force technique
— it will give maximal clique but may not maximum because starting vertex is arbitrary.

2^n — Complexity. — if $n >>$ huge then high complexity.

Pruning → clique → nodes should have degree of $k-1$ or more.
prune all nodes with degree $< k-1$
↳ to reduce no. of nodes we need to go through.

Even after pruning, complexity is undesirable

→ k-plex : set of vertices which have $\boxed{d_v \geq |V|-k}$
previously → $d > k-1$ but now →

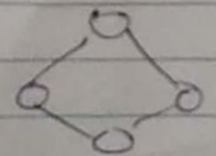
when k↑
$(v-k)$↓

~~100 nodes~~

$6-1 = 5$, $d_v \geq 5$
$|V|$ ↓ k
↓ 1-plex

2-plex → $d_v \geq 6-2 = 4$

⇒ k-core — degree at least k.



⇒ k-shell — nodes part of k-core but not of (k+1) core. | relaxed clique

↓
it is 2-plex as well as 2-core.

maximal connected

→ 0-core : same graph, atleast 0 degree

0-shell : graph itself

↳ isolated pts.

→ 1-core : nodes with atleast 1 edge → then maximal connected.

→ k-cores of complete graph : — same complete graph.

n nodes → $k < n$

16/4/19

1-plex → complete graph.

k plex → also NP Hard

(III) Using clique as a seed of a Comm :

1st CPM : clique as a seed (smaller clique)