CS331
SNA
28/03/2022

* Viral Marketing
* Influence Maximisation
* Information Diffusion.

* Viral Marketing → Relies on Users
    ↳ To reach and influence maximum people
    ↳ Flows very fast
  ↳ Unpredictability
  ↳ Audience Reliability/Reliance.
  ↳ Timing.
  ↳ Risk
  ↳ Content
  ↳ Patience
  ↳ Social Influence → word-of-mouth.
    → Social Currency
    → Triggers
    → Emotion
    → Practical value
    → Public
    → Stories

* flow ~~of~~ through network of users:
  Information Cascade (Diffusion) Models:
  ↳ Independent Cascade Model
  ↳ Linear ~~...~~ Threshhold Model.
  ↳ Epidemic Model.

• Independent Cascade Model:
  ↳ when a node 'v' becomes active, it has a single chance of activating each connected inactive neighbor ω.
  ↳ The activation attempt succeedes with probability $P_{v\omega}$

• Linear Threshold Model:
  ↳ A node, v, has two random thresholds $Q_v \in [0,1]$
  ↳ A node 'v' is influenced by each neighbor x according to a weight $\omega(v,x)$, such that,

$$\sum \omega(v,x) \leq 1$$

  ↳ A node 'v' becomes active when at least (weighted) Q_v function of its neighbours are active i.e., threshold satisfies:

$$\sum \omega(v,x) \geq Q_v$$

* CS 321
SNA
31/08/2022

* Influence Maximisation Problem.

- Set of seed nodes. (S)
- Influence spread. $\sigma(S)$
- Maximise $\sigma_m(S)$

Given a social graph $G(V, E, \omega)$, directed and weighed graph, an integer $k < |V|$, and model m, find set $S \subseteq V$, and $k = |S|$, such that the expected spread of influence is $\sigma_m(S)$.

$\sigma_m(S)$ is a submodular function. Monotonic.

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$

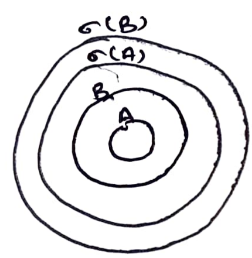
$$A \subset B \Rightarrow f(A) \geq f(B)$$

* Property of Decreasing Marginal Value: (Diminishing Theta):

(i) $A \subset B \Rightarrow \Delta_x f(A) \geq \Delta_x f(B)$, if $x \notin B$. (Submodular)

$A \subset B \Rightarrow \Delta_x f(A) \geq \Delta_x f(B)$, if $x \in B$. (Monotonic)

when, $\Delta_x f(A) = f(A \cup \{x\}) - f(A)$



* Greedy Approach:

Step I: $S \leftarrow \{0\}$

Step II: while, $|S| < k$, do,

Step III: Select any node, $u = \arg\max_{w \in V \setminus S} (\sigma_m(S \cup \{w\}) - \sigma_m(S))$

Here, $V-S$ is the set difference

$$u = \arg\max_{w \in V \setminus S} (\sigma_m(S \cup \{w\}) - \sigma_m(S))$$

Step IV: $S \leftarrow S \cup \{u\}$

CS331
SNA
07/04/2022

*Recommended System:

tailored experience
User's preference

↳ System for recommending items to users based on their preference / interests.

↳ Products, movies, videos, friends, etc.

↳ Review Mining; Opinion Mining

Ratings Sentiment Analysis

→ Collaborative filtering → Neighbourhood Approach; Matrix Factorisation; ML / DL

→ Content-based Approaches.

- Neighbourhood Approach (or Nearest neighbour):
  - → User-User
  - → Item-Item.

	I_1	I_2	I_3	I_4	I_5
U_1	4			5	1
U_2	5	4	5		
U_3		1		2	4
		3	2		

$$J(U_1, U_2) = \frac{(U_1 \cap U_2)}{(U_1 \cup U_2)} = 1/5$$

Cosine $(U_1, U_2) \simeq 0.38$

Cosine $(U_1, U_3) \simeq 0.32$.

$$J(U_1, U_3) = \frac{2}{4} = \frac{1}{2}.$$

	I_1	I_2	I_3	I_4	I_5	
U_1	2/3			5/3	-7/3	} → $I_i - 10/3$
U_2	1/3		-2/3	1/3		} → $I_i - 14/3$
U_3		-4/3		-1/3	5/3	} → $I_i - 7/3$
U_4		1/2	-1/2			} → $I_i - 5/2$

Cosine $(U_1, U_2) \approx$

Cosine $(U_1, U_3) \approx$

* Matrix Factorisation:

$$U_{m \times k} \times I_{k \times n} = R_{m \times n}$$

→ Cold Start
→ Sparsity
→ First · Rater
→ Popularity Bias

→ Content Based Approach.
→ User Profiles
→ Item Profile

CS321
SNA
11/04/2022

Targeted Advertisement:

Trust, Reputation and Trust Propagation.

psychological definition → Trust in a person is a commitment based on a belief that action of that person will lead to good outcome.

mathematical definition → Trust is a level of subjective probability with which an agent accesses that another agent will perform a particular action, both before and independently of such an action is monitored.

psychological → Trustworthiness is a collection of qualities of an agent that leads them to be considered as deserving of trust from others

mathematical → Trustworthiness is a subjective probability that the trustee performs a particular action on which the interest of a truster depends.

→ Reputation is the community or public estimation of standing for merit, achievement and reliability
→ Rep is the opinion of a community towards a person.

Propagation of Trust in Network:

→ Eigenvector Propagation (EIG)

→ Weighted Linear Combination (WLC)

CS321
SNA
11/04/2021

* Matrix Diffusion

$T \rightarrow$ Trust Matrix. T_{ij} ; $i \xrightarrow{\text{trusts}} j$

$D \rightarrow$ Distrust Matrix. D_{ij} ; $i \xrightarrow{\text{distrusts}} j$

$B \rightarrow$ Belief ; $B = T$; $B = T - D$

$C_{B,\alpha} \rightarrow$ Combined atomic propagation matrix

$P^{(K)} \rightarrow$ K-step propagation Matrix.

$F \rightarrow$ Final belief F_{ij} ; $i \xrightarrow{\text{treats}} j$

1. Eigenvector Propagation (EIG) :
   Let K be a suitably chosen integer, then in this model, the final matrix will be

$$F = P^{(K)}$$

2. WLC :
   Let γ be a constant (i.e., smaller than the largest ~~Here~~ eigen value of $C_{B,\alpha}$), and let, K be a suitably chosen integer that, in this WLC model, then,

$$F = \sum_{k=1}^{K} \gamma^{k} P^{(k)}$$

* Atomic Propagation

C-operator ~~Cooperation~~

(i) Direct propagation — B

$i \rightarrow j \rightarrow k$

(ii) Co-citation — $B^T B$

$i_1 \rightarrow j_1$, $i_1 \rightarrow j_2$, $i_2 \rightarrow j_2$, $i_2 \dashrightarrow j_1$

(iii) Transpose Trust — B^T

(iv) Trust Coupling — $B B^T$

if, a, b trust c
Then, a trusts b.

* Combination of all four atomic propagation:

Let, $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ be a ~~factor~~ vector for ~~Let α be a factor vector~~ which to capture all the atomic propagation into a single combined atomic matrix, $C_{B\alpha}$ based on a belief matrix B and weight vector α as follows:

$$C_{B\alpha} = \alpha_1 B + \alpha_2 B^T B + \alpha_3 B^T + \alpha_4 BB^T$$

1) Trust-only, then, $B = T$, and, $p^{(k)} = C^k_{B,\alpha}$

2) One-step distrust, then, $B = T$, and, $p^{(k)} = C^k_{B,\alpha} * (T-D)$

3) Propagational distrust, then, $B = T-D$, and, $p^{(k)} = C^k_{B,\alpha}$

CS331
SNA
21/04/2022

* Rumour!

Centrality Measure

→ fake news
→ wn's Information

Easiest way : Sorting time of post

Sometimes Violates User's privacy, so not feasible every time.

Other methods → Single Source Targeting
→ Targeting Multiple sources

→ Complete Observation
→ Snapshot observation → Diffusion
→ Monitor observation.

Defining problem mathematically: In Graph $G(V,E)$, let v^* be the one to share the information first at timestamp t_0, let G_N be the number of nodes in the Graph infected in the first phase. Find v^*.

~~arg max~~ $\cancel{P(G_N \mid v)}$

~~Probability~~

$$v^* \in \arg\max_{v \in G_N} P(G_N \mid v)$$

$$v^* \in \arg\max_{v \in G_N} R(v, G_N)$$

$P(G_N \mid v)$ is Probability.

$R(v \mid G_N)$ is total number of permitted permutation.

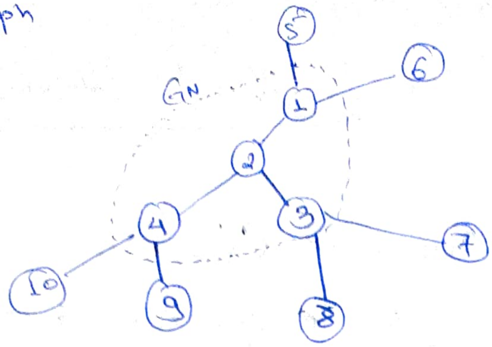
Permutation: Given a connected $G(V,E)$ and source node $v \in V$, Consider a permutation $\sigma : V \to \{1, 2, 3 \ldots |V|\}$ of nodes where $\sigma(u)$ denotes the position of node $u \in V$ in the **permutation** G

Permitted Permutation : ~~Given a P(G,v) with source~~
Given a permitted permutation $P: G(V,E)$ with source

(i) if $\sigma(v) = 1$

(ii) for any (u, u') that belongs to E $((u,u') \in E)$, if $d(v, u) < d(v, u')$, then $\sigma(u) < \sigma(u')$

Q. In the graph



G

Assuming node 1 as the source node, find the permitted permutation.

Sol^n: Permutations, taking 1 as source nod. $\{1,2,3,4\}$, $\{1,3,2,4\}$, $\{1,2,4,3\}$, $\{1,3,4,2\}$, $\{1,4,3,2\}$, $\{1,4,2,3\}$

Condition (i) Satisfied, i.e., first term is the source node.

checking for condition (ii).

Edges list out: $(1,2), (2,4), (2,3), (2,1), (4,2), (3,1)$

for each permutation, check for condition (ii) using each edges from the edge list.

Example, for permutation $\{1,2,3,4\}$ and $(u,u') = (1,2)$.

$d(1,1) < d(1,2)$ TRUE

$\sigma(1) < \sigma(2)$ TRUE

SATISFIED.

for permutation $\{1,2,3,4\}$ and $(u,u') = (2,4)$

$d(1,2) < d(1,4)$ TRUE

$\sigma(2) < \sigma(4)$ TRUE

SATISFIED

Similarly, $\{1,2,3,4\}$ and $(2,3)$ SATISFIED

$\{1,2,3,4\}$ and $(2,1)$ NOT SATISFIED

$(4,2)$ NOT SATISFIED

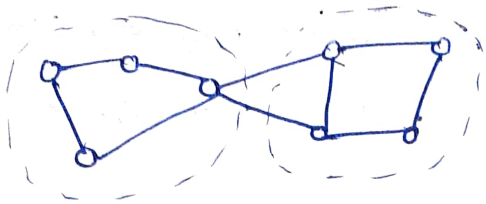
$(3,1)$ NOT SATISFIED

CS331
SNA
23/04/2022

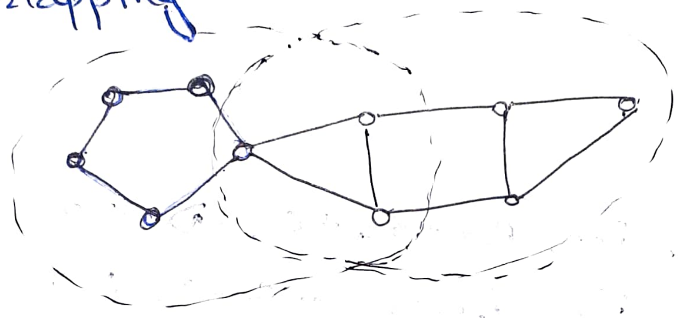
* Community Detection:

What kind of communities to identify:

Types 1) Disjoint Community



2) Overlapping



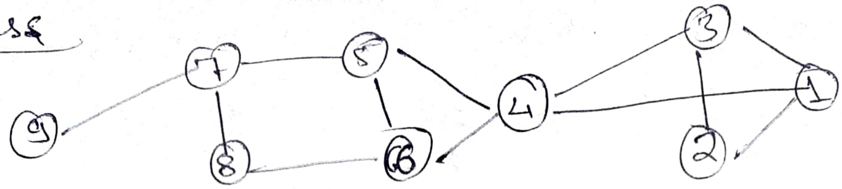
→ Crisp Overlapping.
→ Fuzzy Overlapping.

1) Node Level Properties
2) Root Level Properties
3) Network Level Properties

* Techniques:

- Heirarchical
  - (Top-down approach) Divisive
    - → Girvan Newman Algorithm.
    - → Edge-betweeness
  - (Bottom-up Approach) Agglomerative
    - → Clique Parculaslion
- Spectral
- optimisation
- ~~Statistical~~ Inference

Edge Betweenness



{ 1, 2, 3, 4, 5, 6, 7, 8, 9 }

↓

remove e(4,5) and e(4,6)

{1,2,3,4}

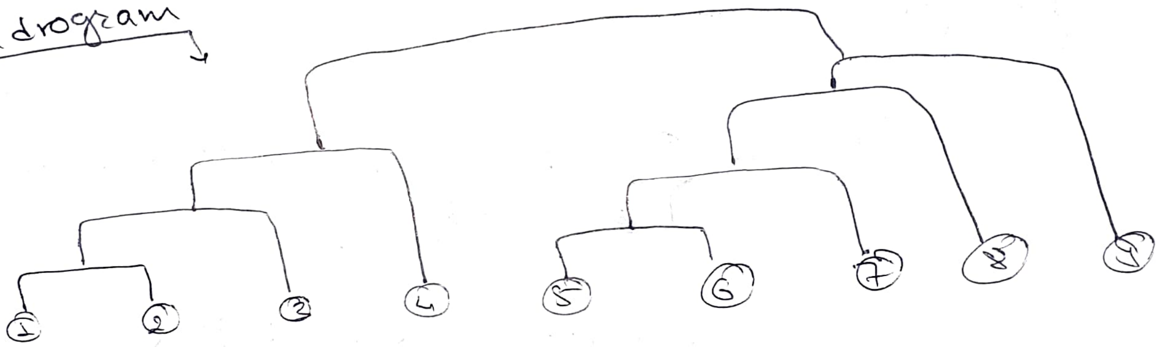
{5, 6, 7, 8, 9}

{5,6,7,8}

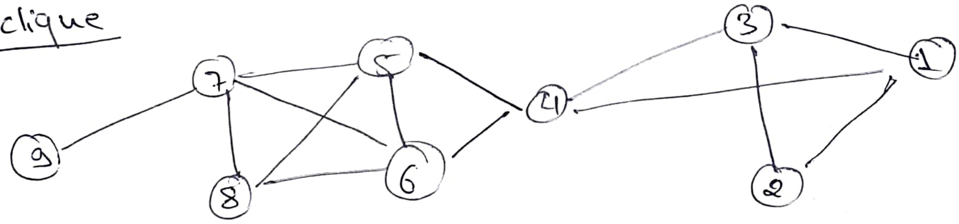
{9}

Agglomerative.

also called dendrogram



Clique → k-clique



3-clique:

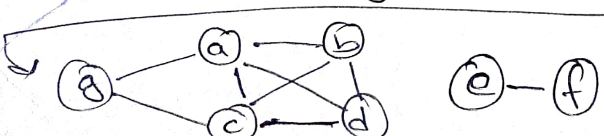
a) {5, 6, 7}
b) {5, 7, 8}
c) {5, 6, 8}
d) {6, 7, 8}
e) {1, 2, 3}
f) {1, 3, 4}
g) {4, 5, 6}

If K-1 common, they are adjacent.

i.e., in 3-clique, if 2 nodes in common, then adjacent.

So,

(a, c), (a, d), (a, b), (a, g)
(b, a), (b, c), (b, d).
(c, a), (c, b), (c, d), (c, g)
(d, a), (d, b), (d, c),
(e, f), (f, e)
(g, a), (g, c)



Let $A = \{C_1, C_2, C_3 \ldots C_k\}$ be a graph partition, such that $C_i \cap C_j = \phi$ and $\sum_{i=1}^{k} C_i = V$

Then,

$$\text{Ratio cut}(A) = \frac{1}{k} \sum_{i=1}^{k} \frac{\text{cut}(C_i, \bar{C_i})}{|C_i|}$$

~~Normalisation~~:

$$\text{Normalised Cut}(A) = \frac{1}{k} \sum_{i=1}^{k} \frac{\text{cut}(C_i, \bar{C_i})}{\text{Vol}(C_i)}$$

where, $\text{vol}(C_i) = \sum_{v \in C_i} d_v$

$$\min_{S \in (0,1)^{n*k}} Tr(S^T \tilde{L} S)$$

$$\tilde{L} = \begin{cases} D - A & \text{, if ratio part used.} \\ I - D^{-1/2} A * D^{-1/2} & \text{, if normalised part used} \end{cases}$$

S^T corresponds to the eigen vectors of $\tilde{L}$ with minimum eigen ~~values~~ vectors.

* Epidemic Models:

1) SI model.
2) SIR Model
3) SIS Model..
4) SIRS Model.

$\boxed{S} \longrightarrow \boxed{I} \longrightarrow \boxed{R} \longrightarrow \boxed{S}$

Susceptibility
Susceptible

Infected.

Recovered

* Susceptible:

An individual in the susceptible state is someone who doesn't have the disease yet, but could get it if they come into contact with someone who have the disease.

* Infected:

An individual in the infected state is someone who has the disease and can potentially pass it on, if they come into the contact with a susceptible person.

At any time t, let, $s(t)$ be the number of individuals for susceptible, and $x(t)$ be the number of individuals for infected.

CS331
SNA CSE
BTech 6th Sem

Social Network Analysis

__/__/____

* **Small World Network:**
  - Shortest path between any two nodes, meaning two nodes will be closely related with few traversals between edges. Short-cut ties that reduce the length between the clusters.
  - High local clustering (meaning all the nodes will be connected to one another in one way or the other, directly or through 1 or 2 next nodes)
  - Example: Airport Network, Websites with ~~nav~~ navigation menus, electric power grids. etc.
  - LFR Benchmark, Albert-Barabasi Model.

* **Scale-free Network (or Real World Network):**
  - Degrees follow power law distribution.
  - Common feature of real world network is the present of hubs (i.e., few nodes that are highly connected to other nodes in network).
  - Example: Social Medias.
  - Erdos-Reny Model.

NOTE: Small world Network also occur in real-world or are a real world phenomenon. Literally, small world network should have fallen under real-world network, however, by subjective definition, scale-free networks are called real world network and differs from small-world.

* **Quantifying Small world Network:**
  → Using small coefficient, σ; $\sigma = \frac{C}{C_r} * \frac{L_r}{L}$
  → Using small world measure, ω; $\omega = \frac{L_r}{L} - \frac{C}{C_t}$
  → Using Small world Index, SWI; $SWI = \frac{L-L_t}{L_r-L_t} * \frac{C-C_r}{C_t-C_r}$