SMA

For exhaustive content, read: http://faculty.ucr.edu/~hanneman/nettext/

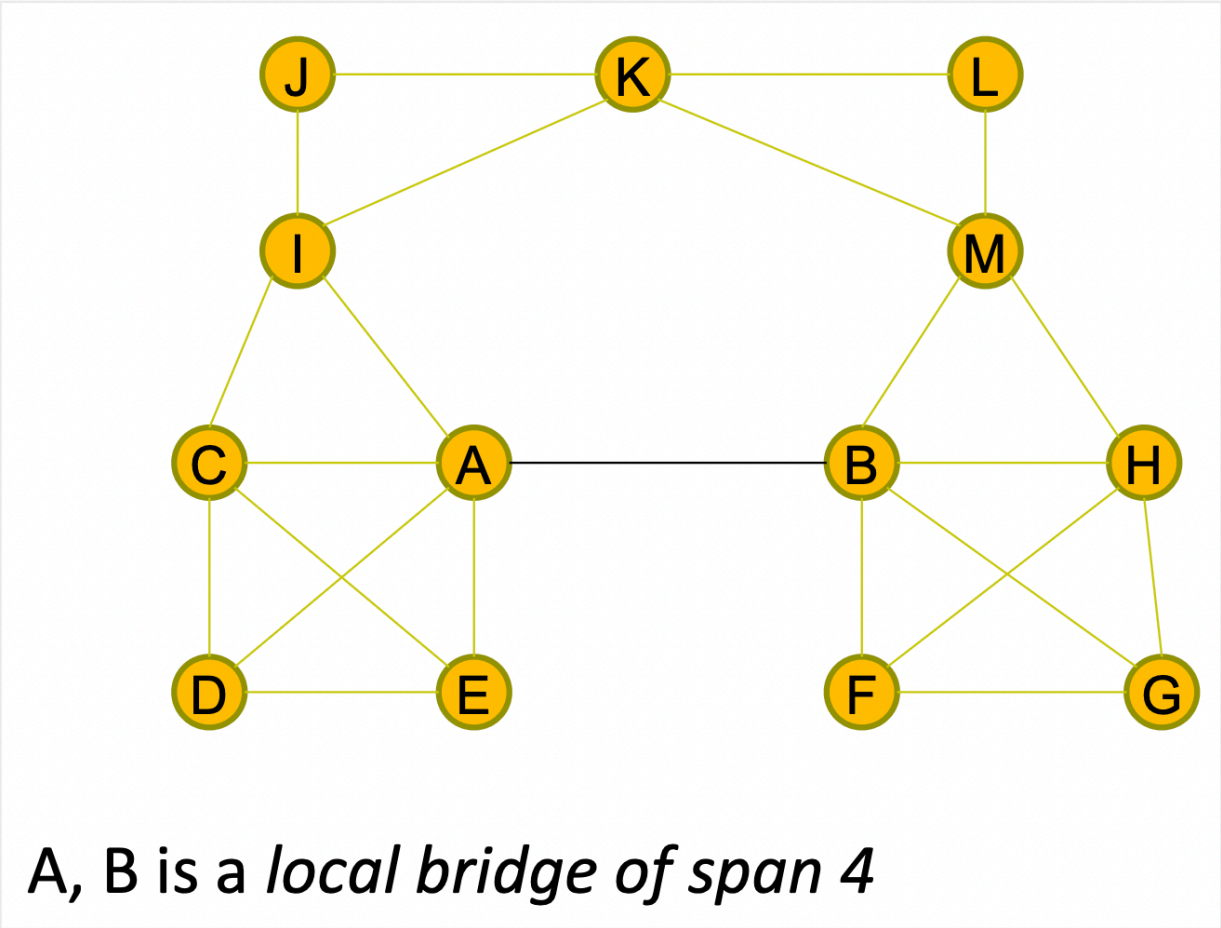
- Customer Network Lifetime Value = Customer Lifetime Value + Customer Influence Value
    - b/c Customer brings in more business (Influence Value) than just 'his' business and that should be catered

- ***Clustering coefficient*** of a user:
    - Probability that two randomly selected friends of the user are friends with each other.

- ***Degree Centrality:***
    - Number of edges connected to a node
    - Normalised Degree Centrality: DegreeCentrality/ (n-1) [where n = # nodes]
    - In Degree / Out Degree (similar concepts) for directed networks
    - A large degree centrality does not imply strong ties. Strength of ties flattens out
    - ***Degree Centrality Metric for Overall Network***
        - d* = highest degree in the network
        - di = degree of Node i
        - N = # nodes
        - xi = d* - di
        - x_max = Theoretical Max of xi (edge node of StarNetwork)
            - x_max = [(n-1) - (n-1)]+(n-1)*[(n-1) - 1]
            - x_max = 0 + (n-1)*(n-2)
            - x_max = (n-2)*(n-2)
        - Network Degree Centrality
            - (Cd = (sum[i =1 -> N](xi))/x_max)

- ***Betweenness Centrality:***
    - 'Interaction and exchange of information between two groups/ communities/clusters
    - Nodes with high betweenness are important in transmitting new information, ideas & opportunities to a wide audience
    - ***Local Bridge***
        - A-B is a local bridge of span 'X' iff the second shortest route between A-B (should the A-B link be severed) is of length 'X'

- DEFINITION. ????
- SIGNIFICANCE ????
- See video?



- ***Betweenness Centrality of a Node:***

$$b_i = \sum_{s,t\ (s \neq i \neq t)} \frac{g_{st}(i)}{g_{st}}$$

- ***Betweenness Centrality of a Network:***
  - yi = b* - bi
  - y_max = Theoretical max vale of numerator (for a star network)

- b* = 1 *(n-1 C 2) -> for central node
- b* = (n-1)*(n-2)/2
- bi = 0 -> for peripheral node
- y_max = (b* - b*) + (b* - 0)*(n-1)
- y_max = (n-1)^2*(n-2)/2

$$Cb = \frac{\sum_{i=1}^{N} y_i}{y_{max}}$$

- ***Closeness Centrality:***
  - How close a node is to all others
  - = 1/(Sum of shortest distance from a node to all others
  - In a Star Network:
    - Closeness of central node (c*) : 1/(n-1)
    - Closeness of peripheral node (ci) : 1/[1 + 2*(n-2)] = 1/(2n - 3)
    - c_max = (c* - ci) = [1/(n-1) - 1/(n-1)] + (n-1)*[1/(n-1) - 1/(2n - 3)]
    - c_max = 0 + (n-1)*[2n - 3 - (n-1)]/(n-1)(2n-3)
    - c_max = (n-2)/(2n - 3)
  - Thus, Denominator of any Network level Closeness:
    - c_max = (n-2)/(2n - 3)

$$\text{Closeness centrality of node } i\text{: } c_i = \frac{1}{\sum_{j\,(j\neq i)} l_{ij}} = \frac{1}{(N-1)\bar{d}_i}$$

Where l_{ij} is the distance between nodes i and j.

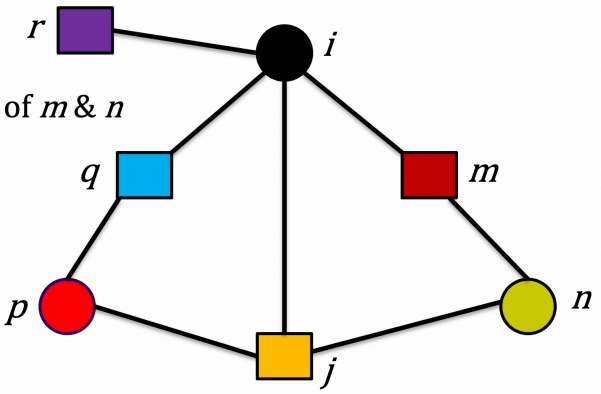
$\bar{d}_i$ is the average geodesic length from node i to all other nodes.

- ***Eigenvector Centrality:***
  - One's importance is partly determined by "the company one keeps"
  - If one has many important friends, s/he should be importantJ
  - Eigenvector centrality considers not only your degree, but your friends' degree

- Eg: PageRank
    - In order to ensure we don't reach 0-vectors when transitions matrices are disconnected, we transition Matrix
    - ***A* = (1-p)A + p*B*** [B = np.ones(shape of A)]
    - Indicates there's a small probability that people start randomly no matter what

- ***Community Detection***
    - Cluster similar users together
    - Girvan-Newman Algo (Divisive Algo) [Uni-variate Network]
        - Cut link with Highest betweenness centrality
        - Recalculate Betweenness for all remaining links
        - Repeat till network completely disjoint
    - Clustering in Bi-partite network
        - Cut links with **lowest** clustering coefficient.
        - Aliter, Separate into 2 uni-modal networks and apply Girvan-Newman
            - Easier but information Loss

Detecting Communities in Bi-partite Networks



k_m & k_n degrees of m & n

For a pair of nodes, i and j, let m and n be neighbors of i and j respectively

q_{ijmn} = 1 if m and n are connected, 0 otherwise. θ_{ijmn} has the opposite definition.

Edge clustering coefficient $C(i,j) =$

squares that currently include i-j / possible # squares that include i-j

$$\frac{\sum_{m=1}^{k_i}\sum_{n=1}^{k_j} q_{ijmn}}{\sum_{m=1}^{k_i}\sum_{n=1}^{k_j} \theta_{ijmn} + \sum_{m=1}^{k_i}(k_m - 1) + \sum_{n=1}^{k_j}(k_n - 1) - \sum_{m=1}^{k_i}\sum_{n=1}^{k_j} q_{ijmn}}$$

- ***Cliques / Cores***
    - ***N - Clique***
        - A network on n fully connected nodes

        - Usually spammers/fake accounts are all densely/fully connected
    - **K - Cores**
        - A node must be connected to all but k other actors in the group
        - Feeling of belongingness depends on number of connections
        - If an actor has ties to a sufficient number of members of a group, they may feel tied to that group -- even if they don't know many, or even most members.
        - It may be that identity depends on connection, rather than on immersion in a sub-group.
    - Hypothesis: Spammers have larger k-cores than non spammers
    - Hypothesis: Spammers have larger n-cliques than non spammers
    - Hypothesis: Spammers have higher network density

- ***Homophily (Similarity)***
    - Connect similar people together
    - Social Influence v/s Homophily
        - If sales driven by Influence
            - Invest big money in small set of Influencers
        - If sales driven by homophily
            - Invest small money in big set of grass-roots
    - Tests for Homophility:
        - *p*(Becoming friends in *t*+1 where attributes were same in *t*) > *p*(Becoming friends in *t*+1 where attributes were different in *t)*
        - *p*(Break friendships in *t*+1 where attributes were same in t) < *p*(Break friendships in *t*+1 where attributes were different in *t*)
    - Checking which is dominant:

Homophily: $C(X_t, G_{t+1}) > C(X_t, G_t)$

Social influence: $C(X_{t+1}, G_t) > C(X_t, G_t)$

$$Relational\ autocorrelation\ C(X,G) = \chi^2 = \frac{N.(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

	Improved Outcome	Didn't improve	Total
Treatment	36	14	50
No Treatment	30	25	55
Total	66	39	105

$$\chi^2 = ?$$

$$\chi^2 = \frac{105*(36*25-30*14)^2}{50*55*66*39} = 3.42$$