SOCIAL NETWORKS

FEBRUARY 24 2015

LECTURE #3

STRONG TRIADIC CLOSURE.

- In a social network, let f_1 and f_2 be two close friends of yours...
  - ⇒ They are connected by strong ties to you.
  - ⇒ Then they (f_1 and f_2) belong to a k-clan (for small k) with high prob.
  - ⇒ Then, it is likely that f_1 and f_2 are acquaintances and are connected to each other by weak ties.

[At least your social network may recommend that f_1 and f_2 explore connecting with each other.]

$$\Pr\left[(v, w) \in E \mid (u, v) \in E_S \wedge (u, w) \in E_S \right]$$

$$> \Pr\left[(v, w) \in E \right].$$

NOTE BOOK

TRIADIC CLOSURE PROPERTIES

- If f_1 and f_2 have a large subgroup of common friends (which include you)
  - ⇒ Then it is ~~possible~~ probable that they are acquaintances.
  - ⇒ The probability montonically increases with the size of the set of mutual friends.

- In a connected social network, we expect to see lots of K_3's (Cliques of size 3).

Empirical Results.

"Strength of Weak Ties" (1973)

Mark Granovetter: (American Sociologist, Currently at Stanford Univ.)

- "Weak ties enable reaching populations and audiences with much higher efficiency than what is achievable or accessible via strong ties."

WHY?

◇ Granoveller's PhD dissertation:
(Dept of Social Relation, Harvard Univ.)

"Getting a Job"

Experiment:

Location: Newton, MA.

Subjects: 282 professinal, technical & managerial workers.

Result: N = 54 = # individuals (out of 282) who found jobs through personal contacts.

- Strong Ties (16.6%)
- Weak Ties (83.4%)
  - Occasional Contacts (55.6%)
  - Rare Contacts (27.8%)

AUGMENTED GRAPH

Def^n: Consider an "augmented" graph (undir.)

$$G = (V, E, E_S)$$

in which

$$E_S \subseteq E \subseteq V \times V.$$

E: The edges/ties.
E_S: The strong ties.
} $E \setminus E_S$ = The weak ties.

$\Rightarrow (u,v) \in E \Rightarrow u$ and v are friends (acquaintances + close friends)

$(u,v) \in E_S \Rightarrow u$ and v are close friends.

The strong triadic closure property

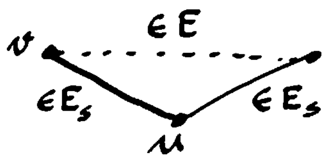
if $(u,v) \in E_S$ and $(u,w) \in E_S$
then $(v,w) \in E$ a.s.

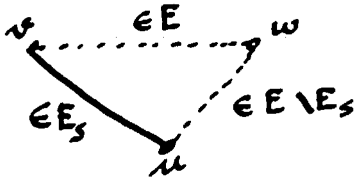
$$\Leftrightarrow \Pr[(v,w) \in E \mid (u,v) \in E_S \wedge (u,w) \in E_S] > \Pr[(v,w) \in E]$$

$\Rightarrow$ Probability Raising in Social Network.
(Influence vs Social Influence)

(I) $\Pr\left[(v,w) \in E \wedge \cancel{\Pr} \,(u,v) \in E_s \,\middle|\, (u,w) \in E_s\right]$

$> \Pr\left[(v,w) \in E \wedge (u,v) \in E_s \,\middle|\, (u,w) \in E \setminus E_s\right]$

$\Pr\Bigg[$  $\Bigg]$

$> \Pr\Bigg[$  $\Bigg]$

(II) Consider a new relation R

$\{(u,v) \in R\}$: Event u obtained a job through a "recommender" v.

[v recommended u to w;
w verified u independently;
w determined whether u is a suitable candidate.]

NOTE BOOK

$$\Pr\left[(u,v)\in R \mid (u,v)\in E_s\right]$$
$$\approx \Pr\left[(u,w)\notin E \wedge (v,w)\in E_s \mid (u,v)\in E_s\right]$$
$$< \Pr\left[(u,w)\notin E \wedge (v,w)\in E_s \mid (u,v)\in E\setminus E_s\right]$$
$$= \Pr\left[(u,v)\in R \mid (u,v)\in E\setminus E_s\right]$$

u = Applicant ; v = Recommender
w = Verifier (Employer)

Strong Ties:
$$(u,v)\in E_s \wedge (v,w)\in E_s$$
$$\Rightarrow (u,w)\in E$$
w knows about u and
can use information in addition
to what's provided in v's recommend^{n}.

Weak Ties:
$$(u,v)\in E\setminus E_s \wedge (v,w)\in E_s$$
$$\Rightarrow (u,w)\notin E$$
w will go by v's recommendation only.

$G = (V, E)$ = Undirected / Directed } graph.

Connected Component:

A connected component of a graph is defined as a maximal subgraph in which path exists from every node to every other.

◇ A path in a graph is closed if its start and end vertices coincide:

$$v_0, v_1, \dots, v_n \equiv v_0$$

$$\forall_i \ (v_i, v_{i+1}) \in E \qquad 0 \le i < n$$

$$\forall_{i,j} \ v_i \ne v_j \qquad 0 \le i < j \le n$$

◇ A cycle is defined as a closed path in which $n \ge 3$.

Strongly Connected Component

A strongly connected component of a graph is defined as a maximal subgraph in which cycle exists connecting every node to every other.

TREE: A tree is a connected graph that contains no cycle.

ADJACENCY MATRIX.

Every graph $G = (V, E)$ (Undir) with $|V| = n$ has associated with it a symmetric adjacency matrix $A \in \{0,1\}^{n\times n}$

Binary $n \times n$ matrix A in which

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \text{ \& } v_j \text{ are adjacent} \\ & (v_i, v_j) \in E \\ 0 & \text{otherwise.} \end{cases}$$

Since in an undirected graph

$$(v_i, v_j) \equiv (v_j, v_i) \qquad a_{ij} = a_{ji}$$

$$A^T = A. \quad \leftarrow \text{A real-valued symmetric matrix.}$$

$$d_{ii} = d(v_i) = |\{ v_j \mid (v_i, v_j) \in E \}| \text{ : Degree.}$$

$$D = \begin{bmatrix} d_{11} & & & 0 \\ & d_{22} & & \\ & & \ddots & \\ 0 & & & d_{nn} \end{bmatrix} = \text{Diagonal Matrix.}$$

$$\text{Trace } D = \sum_{v_i} d(v_i) = 2m.$$

Boundary Matrix $\mathcal{B}$. $\in \{-1,0,+1\}^{m \times n}$

Columns are indexed by the vertices of G.
Rows are indexed by the edges of G

$$\mathcal{B}(e,v) = \begin{cases} +1 & \text{if } v \text{ is the head of } e \\ -1 & \text{if } v \text{ is the tail of } e \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{B}^T B = L = D - A.$$

Choose edge directions arbitrarily if G = undir.

$$L = D - A.$$

Random Surfing on G.

$$P(u,v) = \begin{cases} \frac{1}{d_v} & \text{if } (u,v) \in E \\ 0 & \text{otherwise} \end{cases}$$

$$P = D^{-1} A.$$

$$L = D(I - P) = D\Delta ; \quad \Delta \equiv I - P$$

$\Delta = I - P \equiv$ (Discrete) Laplace operator.

Your position within the social network assigns a social-value.

$\Rightarrow$ Rank

(e.g. page rank).

$f(x)$

$f : V \to \mathbb{R}$ a scalar rank function.

Dirichlet Sum of G

$$\sum_{(u,v)\in E} (f(u)-f(v))^2$$

You'd like to choose ranks so that the sum is minimized:

Avoid the trivial rank: $f(x)=1 \ \forall x$.

$\Rightarrow$ Focusing on the relative values.

$$\Delta f(x) = \frac{1}{d_x} \sum_{(y,x)\in E} (f(x)-f(y))$$

$$\Downarrow$$

$$(I-P) f.$$

RANDOM SURFER MODEL.

Imagine a web surfer bouncing along randomly following the graph ("hyperlink graph of the web.)

When the surfer arrives at a node he chooses at random hyper-links (directed edge) to a new node.

Asymptotically, the proportion of time the random surfer spends on a given node/page is a measure of

$\boxed{\text{Relevance (Relative Importance)}}$

Dangling Nodes - Sinks
Periodicity in the graph

Problems.

Stochastic Teleportation.