Lecture #5

February 26 2013

STRONG TRIADIC CLOSURE

◇ In a social network,
let f_1 and f_2 be two close friends of yours.

→ Connected by ~~two~~ strong ties to you.

◇ Then, it is likely that f_1 and f_2 are are acquaintances.
→ Connected by weak ties to each other.

(At least, your social network may recommend that f_1 and f_2 explore contacting each other.)

→ If f_1 and f_2 have a large subgroup of common friends (including you), it is probable that they are acquaintances – the probability increasing with the size of the ~~nu~~ set of mutual friends.

Thus one expects to see lots of K_3's – cliques of size 3.

TRIADIC CLOSURE

Defn: Consider an "augmented" undirected graph

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

in which

$$E' \subseteq E \subseteq V \times V.$$

E: The edges/ties.

E': The strong ties. $E \setminus E'$: The weak ties.

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

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

The strong triadic closure property states that:

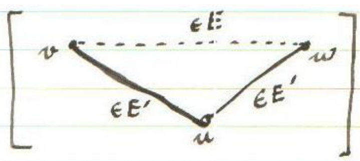
if $(u,v) \in E'$ and $(u,w) \in E'$, then $(v,w) \in E$, a.s.

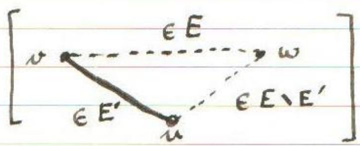
$$\Pr[(v,w) \in E \mid (u,v) \in E' \wedge (u,w) \in E'] > \Pr[(v,w) \in E].$$

◇ The knowledge that v and w have a common close friend, namely u, raises the (conditional) probability that v and w are at least acquaintances.

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

$\square$

Pr 

> Pr 

⇒ "STRENGTH OF WEAK TIES" (1973)

American Sociologist (currently at Stanford Univ.) : Mark Granovetter.

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

"Getting a Job" Granovetter's PhD Dissertation
Dept of Social Relation,
Harvard University.

pp16

<u>Granovetter's</u> <u>Experiment</u>.

282 professional, technical &
managerial workers.

Newton, MA.

N = # individuals out of 282 who found jobs through personal contacts = 54.

- Strong ties (16.6%)
- Weak ties
  - (55.6%) Occasional contact
  - (27.8%) Rare contact.

Explanation through triadic closure property:

Consider a relation R

$\{(u,v) \in R\}$ = Event u obtained a job through a referral by v.

$$Pr\{(u,v) \in R\}$$

Strength of Weak Ties:

16.6%

$$Pr[(u,v) \in R \mid (u,v) \in E'] < Pr[(u,v) \in R \mid (u,v) \in E \setminus E']$$

83.4%.

$Pr[(u,v) \in R \mid (u,v) \in E']$

$= Pr[(u,w) \notin E \wedge (v,w) \in E' \mid (u,v) \in E']$

$< Pr[(u,w) \notin E \wedge (v,w) \in E' \mid (u,v) \in E \setminus E']$

$= Pr[(u,v) \in R \mid (u,v) \in E \setminus E']$

u = Applicant
w = Employer, (potentially)
v = Recommender.

Strong Ties

$(u,v) \in E' \wedge (v,w) \in E'$

$\Rightarrow (u,w) \in E$

w = Likely to be an acquaintance
Can use information in addition to what v provides in the referral.

Weak Ties

$(u,v) \in E \setminus E' \wedge (v,w) \in E'$

$\Rightarrow (u,w) \notin E$

w = Unlikely to be an acquaintance
He will go by v's referral only

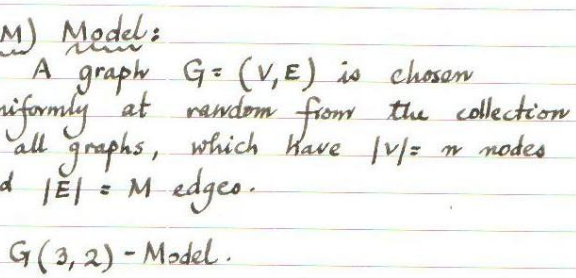
Random Graphs.

ER - Random Graphs.
↳ Erdös - Rényi

TWO WAYS OF DESCRIBING RANDOM GRAPHS.
Closely Related Variants of ER - Random Graphs

$G(n, M)$ Model:
A graph $G = (V, E)$ is chosen uniformly at random from the collection of all graphs, which have $|V| = n$ nodes and $|E| = M$ edges.

$G(3, 2)$ - Model.



Pr = $1/3$ $\quad$ $1/3$ $\quad$ $1/3$

- Exactly three possible graphs on three vertices and two edges.
- Each is assigned a probability $= \frac{1}{3}$.

G(n, p) - Model.

A graph $G = (V, E)$ is constructed by connecting every pair of nodes uniformly randomly.

For every pair of vertices ~~(u, v)~~ $u, v \in V$, an edge $(u, v) \in E$ is included in the graph with probability p independent from every other edge.

Equivalently,

All graphs with $|V| = n$ & $|E| = M$ have equal edge probability of

$$p = \frac{M}{\binom{n}{2}}$$

Parameter p = Density of the graph.

As p increases from 0 to 1, the model produces denser graphs with higher likelihood (than sparser graphs).

At $p = 1/2$, all graphs on n vertices are chosen with equal probability.

Asymptotic Analysis.

$$|V| = n \longrightarrow \infty$$

Random Graphs are often studied in the asymptotic case, as $|V| = n$ (the number of vertices) tends to infinity.

Expected Number of Edges:

$$\langle |E| \rangle = \binom{n}{2} p$$

Expected Degree

$$\bar{d} = \langle d \rangle = \frac{2 \langle |E| \rangle}{\langle |V| \rangle} = \frac{2 \binom{n}{2} p}{n}$$

$$= \frac{2\, n(n-1)\, p}{2n}$$

$$= (n-1)\, p.$$

$(n-1)$ possible other vertices, of which each can be adjacent with probability p.

(pp21)

$d(v) \sim \mathrm{Bin}(n-1, p)$

The degree of a vertex in a graph
$G \in \mathcal{G}(n,p)$
is distributed as a Binomial.

$$\Pr[d(v) = k] = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

If expected degree $\bar{d}$ is held constant
(independent of n)
n = large np = const.

$$\Pr[d(v) = k] \approx \frac{(n-1)^{(k)}}{k!} p^k (1-p)^{\frac{1}{p}(n-1-k)p}$$

$$\approx \frac{(np)^k}{k!} e^{-np}$$

np = const = λ.

$$\Pr[d(v) = k] = e^{-\lambda} \frac{\lambda^k}{k!}$$

$d(v)$ = Poisson(λ)

Poisson Approximation.

pp22

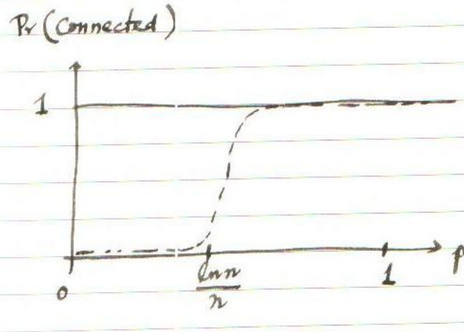
Phrase Transition: 0-1 Laws.

Small p: If $p < \frac{(1-\varepsilon)\ln n}{n}$

Then a graph in $G(n,p)$ will a.s. contain isolated vertices $\Rightarrow$ DISCONNECTED

Large p: If $p > \frac{(1+\varepsilon)\ln n}{n}$

Then a graph in $G(n,p)$ will a.s. be CONNECTED



0-1 Laws.

Describe a phenomenon where an event either occurs or does not occur
- Almost Surely.

TIPPING POINTS
PHASE TRANSITION } With a small increase in a critical parameter the event of interest very quickly goes from probability 0 [Almost Never] to probability 1 [Almost Sure].

- Imagine sending a friend request randomly to $(n-1)$ other individuals in a network (with n individuals)

- Assume: (a) If the recipient is already a friend, he simply ignores the request
  (b) But, otherwise, he receives your request for the first time, & he accepts you as a friend.
  (c) Never ignores, declines or unfriends.

(pp24)

- After $\Theta(n \ln n)$ requests, one will have a.s. befriended all the other $(n-1)$ individuals.
- If every one in the network behaves this way, then with $\Theta(n^2 \ln n)$ messages, the social network will be completely connected, K_n.

- However, with only $\Theta(n \ln n)$ messages, the graph will be in $G(n, \frac{\ln n}{n})$ and a.s. connected.

→ Coupon Collector's Problem.

Collect-All-Coupons-And-Win-Contest.

Problem Statement.

There are n distinct coupons

Coupons can be collected with replacement.

→ What is the probability that more than t sample trials are needed to collect all coupons?

Given n coupons, how many coupons are expected to be drawn with replacement, before each coupon has been drawn at least once.

(pp25)

$n = 52$, $t = 225$

If you draw a card randomly (with replacement) from a deck, then after 225 draws you would have seen every card at least once almost surely.

$$t = \Theta(n \ln n).$$

t_i = time to collect i^{th} coupon after collecting $(i-1)^{th}$ coupon.

t_i's are independent.

$$T = \sum_{i=1}^{n} t_i = \text{Time to collect all coupons.}$$

$$p_i = \Pr[\text{Collect a new coupon after } (i-1)]$$

$$= \frac{n-i+1}{n}$$

$$t_i = \text{Geometric}\left(\frac{1}{p_i}\right)$$

$$\Pr[t_i = k] = (1-p_i)^k p_i$$

$$E(t_i) = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$\text{Var}(t_i) = \frac{1-p_i}{p_i^2} = \frac{(i-1)n}{(n-i+1)^2}$$

(pp26)

$E(T) = E\left(\sum t_i\right) = \sum E(t_i)$

$= \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} = nH_n$

$= n \ln n + \gamma n + \frac{1}{2} + o(n)$

γ = Euler's Const = 0.577

$Var(T) = Var\left(\sum t_i\right) \leq \frac{n^2}{n^2} + \frac{n^2}{(n-1)^2} + \cdots + \frac{n^2}{1}$

$= \frac{\pi^2}{6} n^2$

$\sigma(T) = \frac{\pi n}{\sqrt{6}}$

By Chebyshev Inequality

$\left\{ Pr\left[|T - \mu_T| \geq k\sigma_T\right] \leq \frac{1}{k^2} \right.$

$Pr\left[\,|T - nH_n| \geq c.n\,\right]$

$\leq \frac{\pi^2}{6c^2}$

$\square$