STRONG TRIADIC CLOSURE

– In a social network,
  let \( f_1 \) and \( f_2 \) be two close friends of yours.
  → Connected by strong ties to you.

– Then, it is likely that \( f_1 \) and \( f_2 \) 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 set of mutual friends.

Thus one expects to see lots of \( K_3 \)'s
  → cliques of size 3.
TRIADIC CLOSURE

Define: Consider an "augmented" undirected graph

\[ G = (V, E, E') \]

in which

\[ E' \subseteq E \subseteq V \times V. \]

\[ E = \text{The edges/ties}. \]
\[ E' = \text{The strong ties}, \ E \setminus E' = \text{The weak ties}. \]

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

\[ \Pr \left[ (v, w) \in E \mid (u, v) \in E' \land (u, w) \in E' \right] \]

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

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.
\[ \Pr[(u, w) \in E \land (u, v) \in E'] \mid (u, w) \in E^e] > \Pr[(v, w) \in E \land (u, v) \in E'] \mid (u, w) \in E \times E^e] \]

⇒ "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 Relations, Harvard University.
Granovetter's Experiment.
282 professional, technical & managerial workers.
Newton, MA.

N = # individuals out of 282 who found jobs through personal contacts = 84.

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

Explanation through triadic closure property:

Consider a relation R

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

\[ \Pr \{ (u, v) \in R \} \]

Strength of weak ties: \( \sqrt{16.6\%} \)

\[ \Pr \left[ (u, v) \in R \mid (u, w) \in E \right] < \Pr \left[ (u, v) \in R \mid (u, w) \in E' \right] \approx 83.4 \% \]
\[
\Pr [(u,v) \in E \mid (u,v) \in E']
= \Pr [(u,w) \notin E \land (v,w) \in E' \mid (u,v) \in E']
< \Pr [(u,w) \notin E \land (v,w) \in E' \mid (u,v) \in E']
= \Pr [(u,v) \in E \mid (u,v) \in E \land (v,w) \in E']
\]

\(w = \text{Applicant}\)
\(u = \text{Employer, potentially}\)
\(v = \text{Recommender}\)

**Strong Ties**

\((u,v) \in E' \land (v,w) \in E'\)

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

\(w = \text{likely to be an acquaintance}\)

Can use information in addition to what \(w\) provides in the referral.

**Weak Ties**

\((u,v) \in E \land (u,w) \in E'\)

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

\(w = \text{unlikely to be an acquaintance}\)

He will go by \(v\)'s referral only.
Random Graphs.

ER - Random Graphs.

\[ E_{\text{rd}} - \text{Rényi} \]

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.

\[
\begin{align*}
0 & \quad 1 \\
2 & \quad 3 \\
1 & \quad 2 \\
1 & \quad 3 \\
1 & \quad 3
\end{align*}
\]

\( P_j = \frac{1}{3} \) for \( j = 1, 2, 3 \).

\( \exists \) Exactly three possible graphs on three vertices and two edges.

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

Equivalently, all graphs with $|V| = n$ and $|E| = M$ 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 (through sparser graphs).

At $p = \frac{1}{2}$, all graphs on $n$ vertices are chosen with equal probability.
Asymptotic Analysis.

$|V| = n \to \infty$

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

$\bar{d} = \langle d \rangle = \frac{2 \langle E \rangle}{|V|} = \frac{2 \binom{n}{2} p}{n} = \frac{2 n (n-1) p}{2 n} = (n-1) p$

$(n-1)$ possible other vertices, of which each can be adjacent with probability $p$. 
\( d(v) \sim \text{Bin}(n-1, p) \)

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

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

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

\[ P_r[d(v) = k] \approx \frac{(np)^k}{k!} e^{-np} \]

\( np = \text{const.} \approx \lambda \)

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

\( d(v) \sim \text{Poisson}(\lambda) \)

Poisson Approximation.
Phase Transition: 0-1 Laws.

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

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

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

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

$\Pr(\text{Connected})$

\[
\begin{array}{c|c}
\text{Connected} & 1 \\
\text{Disconnected} & 0 \\
\end{array}
\]

\[\frac{\ln n}{n} \rightarrow 1 \rightarrow p\]
C-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].

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

0 Assume: (a) If the recipient is already a friend, he simply ignores the request.
0 (b) But, otherwise, he receives your request for the first time, he accepts you as a friend.
0 (c) Never ignores, decline or unfriend.
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.
$n = 52$, $t = 225$

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 = \text{time to collect } i\text{th coupon after collecting }(i-1)\text{th coupon}.

t_i's \text{ are independent.}

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

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

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

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

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

\[ \text{Var}(t_i) = \frac{1-p_i}{p_i^2} = \frac{(i-1)n}{(n-i+1)^2} \]
\[ E(T) = E(\sum t_i) = \sum E(t_i) \]
\[ = \frac{\mu}{n} + \frac{\mu}{n^2} + \ldots + \frac{\mu}{1} = nH_n \]
\[ = n\ln n + \gamma + \frac{1}{2} + o(n) \]
\[ z = Euler's\; Constant \approx 0.577 \]

\[ \text{Var}(T) = \text{Var}(\sum t_i) \leq \frac{n^2}{n^2 + n(n-1)^2 + \ldots + \frac{n^2}{1}} \]
\[ = \frac{n^2}{\sigma^2} \]

\[ \sigma(T) = \frac{\sigma}{\sqrt{n}} \]

By Chebyshev Inequality
\[ \text{Pr}[|T - \mu| \geq k\sigma] \leq \frac{1}{k^2} \]
\[ \text{Pr}[|T - nH_n| \geq c \cdot n] \leq \frac{1}{c^2} \]
\[ \leq \frac{n^2}{6c^2} \]