Objectives

Random Graph Models

- Basics
- Erdos-Renyl Model

Social Network Models

Random Graph

- A random graph is a graph in which properties such as the number of graph vertices, graph edges, and connections between them are determined in some random way.

Random Graph

- Choose m pairs of vertices uniformly at random from all possible pairs and connect them with an edge.
- Strictly, in fact, the random graph model is not defined in terms of a single randomly generated network, but as an ensemble of networks, i.e., a probability distribution over possible networks.

Erdos-Renyi Model

$n \rightarrow \frac{n(n-1)}{2}$

$2m$

$V_1 \rightarrow 4$ (V_1, V_2) (V_1, V_3) (V_1, V_4) V_1, V_5

$V_2 \rightarrow 4$ (V_2, V_1) (V_2, V_3) V_4 V_5

Basics

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$
Erdos and Renyi, 1959.
Random graph models

- GNM — $G(n, m)$, a randomly selected graph from the set of C_N^m graphs, $N = \frac{n(n-1)}{2}$, with n nodes and m edges
- GNP — $G(n, p)$, each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m-random number

n vertices w/ prob p, an edge exist bet a pair of vertices

Erdos-Renyi Model ..

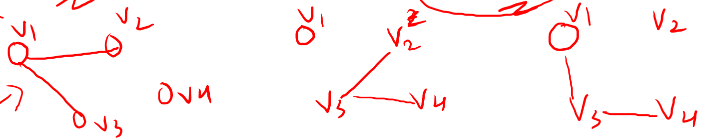
N - No. of vertices.
M - No of edges.

C^M_N

GNM Model

- Random Graph Models: GNM
  Let $n \geq 1$ and $0 \leq M \leq n(n-1)/2$. The $G(n,M)$ model for random graphs is the probability space (S,P), where S denotes the set of simple undirected graphs G on $\{1,2,\ldots,n\}$ with M edges (or equivalently, the set of all subgraphs G of K_n with n nodes and M edges) and each graph $G \in S$ is assigned equal probability $P(G) = \frac{1}{|S|}$.

vertices 4
2 edges



$\frac{1}{10}$

GNM Model ..

- Binomial Coefficient: $\binom{n}{m}$ (read as **n choose m**)

Computing $P(G)$

- The number of edges in K_n: $\binom{n}{2}$ $= \frac{n(n-1)}{2}$

  n_2 $(n-2)!\,2!$

- The number of traingle in K_n: $\binom{n}{3}$
- Note that $\binom{n}{m} = \binom{n}{n-m}$

Computing $P(G)$

- Thus the number of graphs in S in the $G(n, M)$ model is simply the number of edges in K_n choose M.
- ie., $\binom{n(n-1)/2}{M}$ as we just need to choose which M edges to include in our graph.

Choose M edges from $\frac{n(n-1)}{2}$

GNM MODEL ..

$P(G)$

- Hence $P(G) = \frac{1}{\binom{n(n-1)/2}{M}}$, $G \in S$

$G(N, M)$

$5, 3$

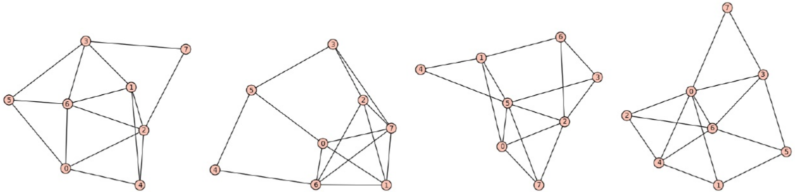
$5(4)/2 \rightarrow$ edges

$\frac{1}{\binom{10}{3}}$.

$\binom{10}{3}$ edges

IN SAGE

```
from sage.graphs.graph_generators_pyx import RandomGNP
D = graphs.RandomGNM(8, 15)
```



OBJECTIVES

RANDOM GRAPH MODELS

- Erdos-Renyl Model
- GNP Model

Erdos Renyi Model

G NM

$p(G) \to \dfrac{1}{\binom{n(n-1)/2}{m}}$

(n, p)

GNP Model

- Second Model: GNP Model
  Let $n \geq 1$ and $0 \leq p \leq 1$. The $G(n,p)$ model for random graphs is the probability space (S, P) where S is the set of all (simple undirected) graphs on $\{1, 2, \ldots, n\}$, where for $G \in S$ with m edges, we define the probability function to be $P(G) = p^m(1-p)^{\frac{n(n-1)}{2}-m}$

no of vertices

prob. of an edge bet a pair of vertices.

$\frac{n(n-1)}{2} - m$

$p^m (1-p)^{\frac{n(n-1)}{2} - m}$

$p^x (1-p)^{n-x}$ → Bernoulli trials.

COMPUTING $P(G)$

- Let G be a random $G(n,p)$ graph.
  - For $1 \leq i < j \leq n$, let $A_{i,j}$ denote the event that the edge {i, j} is included G
  - $B_{i,j}$ the event that the edge {i, j} is not included in G.
  - Say G has m edges: $\{i_1, j_1\}, .., \{i_m, j_m\}$.
  - $\{i_{m+1}, j_{m+1}\}, .., \{i_N, j_N\}$ denote the remaining pairs of non-edges
    - $*$ where $N = \frac{n(n-1)}{2}$
- $A_{i,j}$ events and $B_{i,j}$ are pairwise independent events

edge is present | edge is not present

A_{ij} | B_{ij}

Erdos Ranyi Model: GNP ..

$p(A \cap B)$

$p(A) \cdot p(B)$

Computing $P(G)$

$\frac{n(n-1)}{2}$

- $P(A_{i_1,j_1} \cap A_{i_2,j_2} \cap \ldots \cap A_{i_m,j_m}) \cap B_{i_{m+1},j_{m+1}} \cap \ldots \cap B_{i_N,j_N}) =$
- $P(A_{i_1,j_1}) \cdot P(A_{i_2,j_2}) \cdot \ldots \cdot P(A_{i_m,j_m}) \cdot P(B_{i_{m+1},j_{m+1}}) \ldots P(B_{i_N,j_N}) = p^m(1-p)^{N-m}$

$P(G)$

- $P(G) = p^m(1-p)^{N-m}$
- $= p^m(1-p)^{\frac{n(n-1)}{2} - m}$

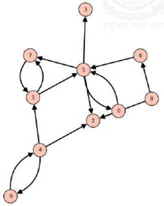
m $\quad N-m$

Erdos Ranyi Model: GNP ..

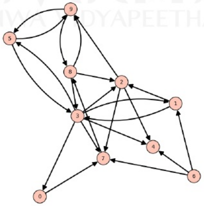
In Sage..

```
from sage.graphs.graph_generators_pyx import RandomGNP
D = RandomGNP(10, .2, directed = True)
```

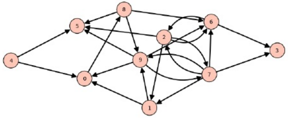
RandomDirectedGNP(10,0.200000000000000): Digraph on 10 vertices

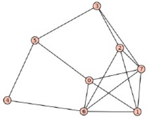


RandomDirectedGNP(10,0.200000000000000): Digraph on 10 vertices



RandomDirectedGNP(10,0.200000000000000): Digraph on 10 vertices



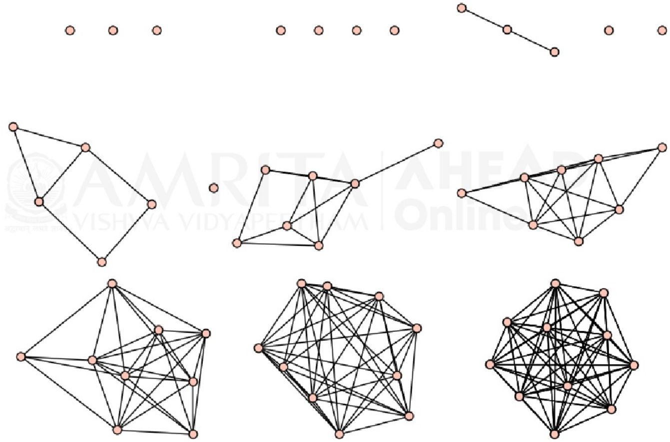


Erdos Ranyi Model: GNM & GNP .. (1/4)

In Sage..

```
g = []
j = []
for i in range(9):
    k = graphs.RandomGNM(i+3, i^2-i)
    g.append(k)
for i in range(3):
        n = []
        for m in range(3):
            n.append(g[3*i + m].plot(vertex_size=50,
                                    vertex_labels=False))
        j.append(n)
G = sage.plot.graphics.GraphicsArray(j)
G.show()  # long time
```

ERDOS RANYI MODEL: GNM & GNP .. (2/4)

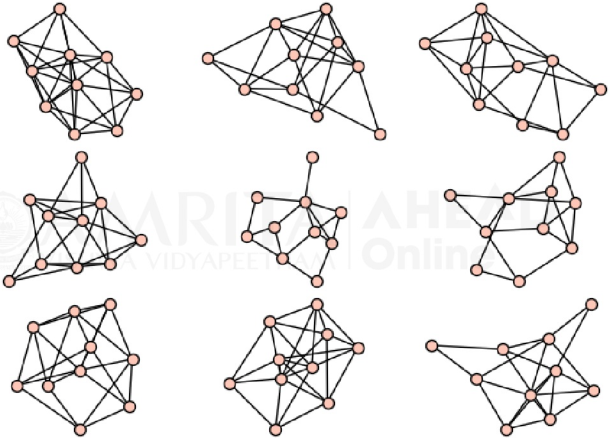


Erdos Ranyi Model: GNM & GNP .. (3/4)

In Sage..

```
g = []
j = []
for i in range(9):
    k = graphs.RandomGNP(10, 0.5)
    g.append(k)
for i in range(3):
        n = []
        for m in range(3):
            n.append(g[3*i + m].plot(vertex_size=50,
                     vertex_labels=False))
        j.append(n)
G = sage.plot.graphics.GraphicsArray(j)
G.show()  # long time
```

Erdos Ranyi Model: GNM & GNP .. (4/4)



Objectives

Random Graph Models

- Properties of GNP Model

Properties of GNP Model

Mean Number of Edges

GNP

- Consider random graph $G(n, p)$. The number of edges in the model is not fixed, but we can calculate its mean or expectation value as follows
  - The number of graphs with exactly n vertices and m edges is equal to the number of ways of picking the positions of the edges from the $\binom{n}{2}$ distinct vertex pairs.

$\binom{n}{2} \rightarrow$

2

MEAN NUMBER OF EDGES ..

- Each of these graphs appears with the same probability P(G) and hence the total probability of drawing a graph with m edges from our ensemble is

$$P(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\frac{n(n-1)}{2}-m}$$

- Binomial Distribution: mean value of m is

distinct set of vertices.

$$\sum x \cdot P(x)$$

$$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} mP(m) = \binom{n}{2} p$$

For each pair (i,j) the prob. of an edge $= p$

$\binom{n}{2} * p$

Properties of GNP Model..

A — B

Mean Degree

- Let k be the mean degree for a n-vertex graph
- then $k.n = 2m$, or $k = \frac{2m}{n}$
- hence mean degree in $G(n, p)$

$\sum x \cdot p(x)$

$$\langle k \rangle = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n}\binom{n}{2} p = (n-1)p$$

$\sum k \cdot p(k)$

- The mean degree of a random graph is often denoted **c**

m

$2m \rightarrow$ undirected

n vertices

$n.k = 2m$

$k = 2m/n$

$\langle k \rangle = \frac{2}{n} \sum m \, p(m) = \frac{2}{n}\binom{n}{2} p = \frac{2}{n} \frac{n(n-1)!}{(n-2)!\,2!} p$

PROPERTIES OF GNP MODEL..

Deg. dist

DEGREE DISTRIBUTION

how many node have deg 0 = 2

- A given vertex in the graph is connected with independent probability p to each of the $n-1$ other vertices.

deg 1 = 3

- Thus the probability of being connected to a particular k other vertices and not to any of the others is $p^k(1-p)^{n-1-k}$

2 =

- There are $\binom{n-1}{k}$ ways to choose those k other vertices, and hence the total probability of being connected to exactly k others is

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Properties of GNP Model...

Directed graph

$$\frac{2 \cdot L_i}{k_i (k_i - 1)}$$

Clustering Coefficient

- Local clustering coefficient C_i measures the density of links in node $i's$ immediate neighborhood: $C_i = 0$ means that there are no links between i's neighbors; $C_i = 1$ implies that each of the i's neighbors link to each other
- To calculate C_i for a node in a random network we need to estimate the expected number of links L_i between the node's k_i neighbors.
- In a random network the probability that two of i's neighbors link to each other is say p_c

PROPERTIES OF GNP MODEL...

$\frac{n(n-1)}{2}$

CLUSTERING COEFFICIENT ..

- As there are $k_i(k_i-1)/2$ possible links between the k_i neighbors of node i, the expected value of L_i is

$$\langle L_i \rangle = p_c \frac{k_i(k_i-1)}{2}$$

- The clustering coefficient is

$$p_c = C_i = \frac{2\langle L_i \rangle}{k_i(k_i-1)}$$

Assume undirected

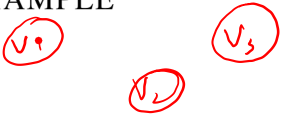
Since mean node degree is 2m/n:

- $C_i = p_c = \frac{\langle k \rangle}{n-1}$

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

- Which is nothing but simply **p**

GNP Model: Example



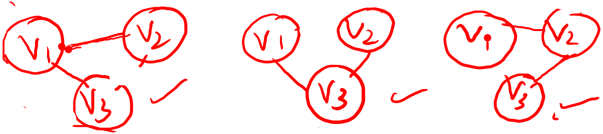
Example

Let $n = 3$ and $p = 0.5$. On $V = \{1, 2, 3\}$.

- A_m denote the subset of graphs on V with m edges.
- Hence, in the $G(n,p) = G(3, 0.5)$ model.

$\binom{3}{0}$ 1 graph with 0 edges	$P(A_0) = P(0 edges) = 1.(0.5)^3 = 0.125$
$\binom{3}{1}$ 3 graph with 1 edge	$P(A_1) = P(1 edges) = 3.(0.5)^3 = 0.375$
$\binom{3}{2}$ 3 graph with 2 edges	$P(A_2) = P(2 edges) = 3.(0.5)^3 = 0.375$
$\binom{3}{3}$ 1 graph with 3 edges	$P(A_3) = P(3 edges) = 1.(0.5)^3 = 0.125$

3 vertices and no edges



OBJECTIVES

ERDOS-RENYI GRAPHS

- Problems and Solutions

Problem 1 (1/2)

GNP Model

- Consider random graph $G(n, p)$ with mean degree c
  - Show that in the limit of large n, number of triangles in the network is $\frac{1}{6}c^3$.
  - Show that expected number of connected triplets is $\frac{1}{2}nc^2$ and hence calculate $C\ell(G)$.

Edges $\binom{n}{3}p$.

Number of triangles

- Let p be the probability of being connected

Number of triangles $= \binom{n}{3} \times p$

Given that mean degree is $c \approxeq (n-1)p$

Hence, $p = \frac{c}{n-1}$

$\therefore$ Number of triangles $= \binom{n}{3} \times \frac{c^3}{(n-1)^3} = \frac{1}{6}c^3$

Problem 1 (2/2)

Number of Connected Triplets

- Let p be the probability of being connected
- Consider a single vertex, there are $\binom{n-1}{2}$ possible pair of neighbors.

Total connected triplets
$= \binom{n}{2} \times p^2 \times n$
For large n, $\approxeq \binom{n-1}{2} \times p^2 \times n$
$= \frac{n(n-1)}{2} \times \frac{c^2}{(n-1)^2} \times n$
$= \frac{1}{2}nc^2$

Clustering Coefficient

- $C\ell(G) = \frac{Number\ \ of\ \ triangles \times 3}{Number\ \ of\ \ connected\ \ triplets}$
  $= \frac{\frac{1}{6}c^3 \times 3}{\frac{1}{2}nc^2}$
  $= \frac{c}{n}$

Problem 2 (1/4)

$F_1 \to F_2 \to F_3 \to F_4$

Transaction Network

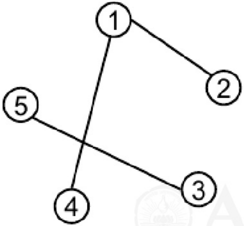
N firms produce goods by buying and selling intermediate products from and to each other. The resulting inter-firm transaction network is modeled as an (undirected) Erdos-Renyi random graph. Based on this model, the expected average nearest-neighbor degree is estimated as $\langle k^{nn} \rangle \approx 2$ and the expected average clustering coefficient as $\langle C \rangle \approx 0.01$

- Determine the total number N of firms and the probability p that any two firms are connected.
  - $N = \frac{\langle k^{nn} \rangle}{\langle C \rangle} = \frac{2}{0.01} = 200$
  - $p = \langle C \rangle = 0.01$

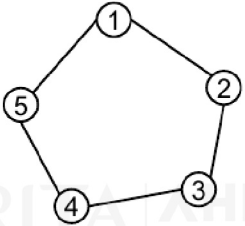
PROBLEM 2 (2/4)

- Determine the average number $\langle k \rangle$ of firms each firm is connected to, the total expected number $\langle L \rangle$ of links, and the average firm-firm distance $\langle D \rangle$ in the network.
  - $\langle k \rangle = (N-1)p \approxeq 2$
  - $\langle L \rangle = \frac{N(N-1)}{2} p = \frac{200 \times 199}{2} \times 0.01 = 199$
  - $\langle D \rangle = \frac{logN}{logk} = 7.64$
- Consider the following three possible configurations for the subgraph involving the first 5 firms only. Calculate the probability of occurrence for each of the three possibilities in the inter-firm network considered above.

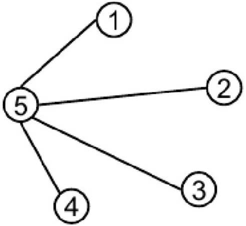
PROBLEM 2 (3/4)



Subgraph 1



Subgraph 2



Subgraph 3

- Total possible edges $= \binom{5}{2} = 10$
- Subgraph 1 - $p^3(1-p)^7 = (0.01)^3(0.99)^7$
- Subgraph 2 - $p^5(1-p)^5 = (0.01)^5(0.99)^5$
- Subgraph 3 - $p^4(1-p)^6 = (0.01)^4(0.99)^6$

Problem 2 (4/4)

Discussion

- Critical condition probability $p_c \approxeq {}^1/_N = {}^1/_{200}$
  - if $p < p_c$ - Network is divided into many small connected components
  - if $p > p_c$ - Network has a gaint component and have some smaller components
- In this example $p = {}^1/_{100} > p_c$, hence, there must be a gaint component.

References

Hagberg, A., Swart, P., and S Chult, D.
Exploring network structure, dynamics, and function using networkx.
Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

Newman, M.
Networks.
Oxford university press, 2018.

Zinoviev, D.
Complex network analysis in Python: Recognize-construct-visualize-analyze-interpret.
Pragmatic Bookshelf, 2018.