Lecture 3: Random Graphs

Radu Balan

February 14, 2017

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Definition

Today we discuss about random graphs. The *Erdöd-Rényi class* $\mathcal{G}_{n,p}$ of random graphs is defined as follows.

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Definition

Today we discuss about random graphs. The *Erdöd-Rényi class* $\mathcal{G}_{n,p}$ of random graphs is defined as follows.
Let $\mathcal{V}$ denote the set of n vertices, $\mathcal{V} = \{1, 2, \cdots, n\}$, and let $\mathcal{G}$ denote the set of all graphs with vertices $\mathcal{V}$. There are exactly $2^{\binom{n}{2}}$ such graphs.
The probability mass function on $\mathcal{G}$, $P : \mathcal{G} \to [0, 1]$, is obtained by assuming that, as random variables, edges are independent from one another, and each edge occurs with probability $p \in [0, 1]$. Thus a graph $G \in \mathcal{G}$ with m vertices will have probability $P(G)$ given by

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

(explain why)

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Probability space

Formally, $\mathcal{G}_{n,p}$ stands for the the probability space $(\mathcal{G}, P)$ composed of the set $\mathcal{G}$ of all graphs with n vertices, and the probability mass function P defined above.

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Probability space

Formally, $\mathcal{G}_{n,p}$ stands for the the probability space $(\mathcal{G}, P)$ composed of the set $\mathcal{G}$ of all graphs with n vertices, and the probability mass function P defined above.

A reformulation of P: Let $G = (\mathcal{V}, \mathcal{E})$ be a graph with n vertices and m edges and let A be its adjacency matrix. Then:

$$P(G) = \prod_{(i,j)\in\mathcal{E}} Prob((i,j)\ is\ an\ edge) \prod_{(i,j)\notin\mathcal{E}} Prob((i,j)\ is\ not\ an\ edge) =$$

$$= \prod_{1\le i<j\le n} p^{A_{i,j}}(1-p)^{1-A_{i,j}}$$

where the product is over all ordered pairs (i,j) with $1 \le i < j \le n$. Note:

$$|\{(i,j)\ ,\ 1 \le i < j \le n\}| = \binom{n}{2} \ \&\ |\{(i,j) \in \mathcal{E}\}| = |\mathcal{E}| = m = \sum_{1\le i<j\le n} A_{i,j}.$$

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Computations in $\mathcal{G}_{n,p}$

How to compute the expected number of edges of a graph in $\mathcal{G}_{n,p}$?

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Computations in $\mathcal{G}_{n,p}$

How to compute the expected number of edges of a graph in $\mathcal{G}_{n,p}$?

Let $X_2 : \mathcal{G}_{n,p} \rightarrow \{0, 1, \cdots, \left(\begin{array}{c} n \\ 2 \end{array} \right)\}$ be the random variable of *number of edges of a graph G*.

$$X_2 = \sum_{1 \leq i < j \leq n} 1_{(i,j)} \quad , \quad 1_{(i,j)}(G) = \left\{ \begin{array}{lll} 1 & if & (i,j) \text{ is edge in } G \\ 0 & if & otherwise \end{array} \right.$$

Use linearity and the fact that $\mathbb{E}[1_{(i,j)}] = Prob((i,j) \in \mathcal{E}) = p$ to obtain:

$$\mathbb{E}[Number\ of\ Edges] = \left(\begin{array}{c} n \\ 2 \end{array} \right) p = \frac{n(n-1)}{2} p$$

The Erdös-Rényi class $\mathcal{G}_{n,p}$

MLE of p

Given a realization G of a graph with n vertices and m edges, how to estimate the most likely p that explains the graph.
Concept: The Maximum Likelihood Estimator (MLE).
In statistics: The MLE of a parameter θ given an observation x of a random variable $X \sim p_X(x;\theta)$ is the value θ that maximizes the probability $P_X(x;\theta)$:

$$\theta_{MLE} = argmax_\theta P_X(x;\theta).$$

In our case: our observation G has m edges. We know

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

The Erdös-Rényi class $\mathcal{G}_{n,p}$

MLE of p

Lemma

Given a random graph with n vertices and m edges, the MLE estimator of p is

$$p_{MLE} = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}.$$

The Erdös-Rényi class $\mathcal{G}_{n,p}$

MLE of p

Lemma

Given a random graph with n vertices and m edges, the MLE estimator of p is

$$p_{MLE} = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}.$$

Why

Note $log(P(G;p)) = mlog(p) + (\binom{n}{2} - m)log(1-p)$ and solve for p:

$$\frac{dlog(P)}{dp} = \frac{m}{p} - \frac{\binom{n}{2} - m}{1-p} = 0.$$

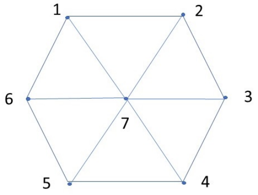
Cliques

q-cliques

Definition

Given a graph $G = (\mathcal{V}, \mathcal{E})$, a subset of q vertices $S \subset \mathcal{V}$ is called a *q-clique* if the subgraph $(S, \mathcal{E}|_S)$ is complete.

In other words, S is a q-clique if for every $i \neq j \in S$, $(i,j) \in \mathcal{E}$ (or $(j,i) \in \mathcal{E}$), that is, (i,j) is an edge in G.



- Each edge is a 2-clique.

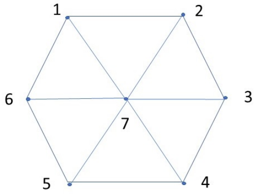
Cliques

q-cliques

Definition

Given a graph $G = (\mathcal{V}, \mathcal{E})$, a subset of q vertices $S \subset \mathcal{V}$ is called a *q-clique* if the subgraph $(S, \mathcal{E}|_S)$ is complete.

In other words, S is a q-clique if for every $i \neq j \in S$, $(i,j) \in \mathcal{E}$ (or $(j,i) \in \mathcal{E}$), that is, (i,j) is an edge in G.



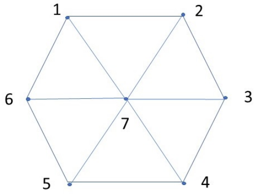
- Each edge is a 2-clique.
- $\{1,2,7\}$ is a 3-clique. And so are $\{2,3,7\}, \{3,4,7\}, \{4,5,7\}, \{5,6,7\}, \{1,6,7\}$

Cliques
q-cliques

Definition

Given a graph $G = (\mathcal{V}, \mathcal{E})$, a subset of q vertices $S \subset \mathcal{V}$ is called a *q-clique* if the subgraph $(S, \mathcal{E}|_S)$ is complete.

In other words, S is a q-clique if for every $i \neq j \in S$, $(i,j) \in \mathcal{E}$ (or $(j,i) \in \mathcal{E}$), that is, (i,j) is an edge in G.



- Each edge is a 2-clique.
- $\{1,2,7\}$ is a 3-clique. And so are $\{2,3,7\}, \{3,4,7\}, \{4,5,7\}, \{5,6,7\}, \{1,6,7\}$
- There is no k-clique, with $k \geq 4$.

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Computations in $\mathcal{G}_{n,p}$: q-cliques

How to compute the expected number of q-cliques?

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Computations in $\mathcal{G}_{n,p}$: q-cliques

How to compute the expected number of q-cliques?
For $k = 2$ we computed earlier the number of edges, which is also the number of 2-cliques.
We shall compute now the number of 3-cliques: triangles, or 3-cycles.
Let $X_3 : \mathcal{G}_{n,p} \to \mathbb{N}$ be the random variable of number of 3-cliques. Note the maximum number of 3-cliques is $\binom{n}{3}$.
Let S_3 denote the set of all distinct 3-cliques of the complete graph with n vertices, $S_3 = \{(i,j,k)\ ,\ 1 \leq i < j < k \leq n\}$.
Let

$$1_{(i,j,k)}(G) = \begin{cases} 1 & if \quad (i,j,k)\ is\ a\ 3-clique\ in\ G \\ 0 & if \quad otherwise \end{cases}$$

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Expectation of the number of 3-cliques

Note: $X_3 = \sum_{(i,j,k)\in S_3} 1_{(i,j,k)}$. Thus

$$\mathbb{E}[X_3] = \sum_{(i,j,k)\in S_3} \mathbb{E}[1_{(i,j,k)}] = \sum_{(i,j,k)\in S_3} Prob((i,j,k) \text{ } is \text{ } a \text{ } clique).$$

Since $Prob((i,j,k) \text{ } is \text{ } a \text{ } clique) = p^3$ we obtain:

$$\mathbb{E}[Number \text{ } of \text{ } 3 - cliques] = \left(\begin{array}{c} n \\ 3 \end{array} \right) p^3 = \frac{n(n-1)(n-2)}{6} p^3.$$

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Number of 3 cliques

Assume we observe a graph G with n vertices and m edges. What would be the expected number N_3 of 3-cliques?

$$\mathbb{E}[X_3 | X_2 = m] = \frac{1}{L} \sum_{k=1}^{L} X_3(G_k)$$

where L denotes the numbe of graphs with m edges and n vertices, and $G_1, \cdots, G_L$ is an enumeration of these graphs.

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Number of 3 cliques

Assume we observe a graph G with n vertices and m edges. What would be the expected number N_3 of 3-cliques?

$$\mathbb{E}[X_3|X_2 = m] = \frac{1}{L}\sum_{k=1}^{L} X_3(G_k)$$

where L denotes the numbe of graphs with m edges and n vertices, and $G_1, \cdots, G_L$ is an enumeration of these graphs.
We approximate:

$$\mathbb{E}[X_3|X_2 = m] \approx \mathbb{E}[X_3; p = p_{MLE}(m)]$$

and obtain:

$$E[X_3|X_2 = m] \approx \frac{4(n-2)}{3n^2(n-1)^2}m^3.$$

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Expectation of the number of q-cliques

Let $X_q : \mathcal{G}_{n,p} \rightarrow \mathbb{N}$ be the random variable of number of q-cliques. Note the maximum number of q-cliques is $\begin{pmatrix} n \\ q \end{pmatrix}$.

Let S_q denote the set of all distinct q-cliques of the complete graph with n vertices, $S_q = \{(i_1, i_2, \cdots, i_q)\ ,\ 1 \leq i_1 < i_2 < \cdots < i_q \leq n\}$. Note $|S_q| = \begin{pmatrix} n \\ q \end{pmatrix}$.

Let

$$1_{(i_1,i_2,\cdots,i_q)}(G) = \left\{ \begin{array}{lll} 1 & if & (i_1, i_2, \cdots, i_q)\ is\ a\ q - clique\ in\ G \\ 0 & if & otherwise \end{array} \right.$$

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Expectation of the number of q-cliques

Since $X_q = \sum_{(i_1,\cdots,i_q)\in S_q} 1_{i_1,\cdots,i_q}$ and

$Prob((i_1,\cdots,i_q)\ is\ a\ clique) = p^{\binom{q}{2}}$ we obtain:

$$\mathbb{E}[Number\ of\ q-cliques] = \binom{n}{q} p^{q(q-1)/2}.$$

The Erdös-Rényi class $\mathcal{G}_{n,p}$

Expectation of the number of q-cliques

Since $X_q = \sum_{(i_1,\cdots,i_q)\in S_q} 1_{i_1,\cdots,i_q}$ and

$Prob((i_1,\cdots,i_q) \text{ is a clique}) = p^{\binom{q}{2}}$ we obtain:

$$\mathbb{E}[Number\ of\ q - cliques] = \binom{n}{q} p^{q(q-1)/2}.$$

Using a similar argument as before, if G has m edges, then

$$\mathbb{E}[X_q|X_2 = m] \approx \binom{n}{q} \left(\frac{2m}{n(n-1)}\right)^{q(q-1)/2}.$$

Computation of Number of Cliques
An Iterative Algorithm

We discuss two algorithms to compute X_q: iterative, and adjacency matrix based algorithm.

Framework: we are given a sequence $(G_t)_{t\geq 0}$ of graphs on n vertices, where G_{t+1} is obtained from G_t by adding one additional edge: $G_t = (\mathcal{V}, \mathcal{E}_t)$, $\emptyset = \mathcal{E}_0 \subset \mathcal{E}_1 \subset \cdots$ and $|\mathcal{E}_t| = t$.

Iterative Algorithm: Assume we know $X_q(G_t)$, the number of q-cliques of graph G_t. Then $X_q(G_{t+1}) = X_q(G_t) + D_q(e; G_t)$ where $D_q(e; G_t)$ denotes the number of q-cliques in G_{t+1} formed by the additional edge $e \in \mathcal{E}_{t+1} \setminus \mathcal{E}_t$.

Computation of Number of Cliques

An Analytic Formula

Laplace Matrix $\Delta = D - A$ contains all connectivity information.
Idea: Note the (i,j) element of A^2 is

$$(A^2)_{i,j} = \sum_{k=1}^{n} A_{i,k}A_{k,j} = |\{k : i \sim k \sim j\}|.$$

This means $(A^2)_{i,j}$ is the number of paths of length 2 that connect i to j.
Remark: The diagonal elements of $A(A^2 - D)$ represent twice the number of 3-cycles ($=$ 3-cliques) that contain that particular vertex.
Conclusion:

$$X_3 = \frac{1}{6} trace\{A(A^2 - D)\} = \frac{1}{6} trace(A^3).$$

Computation of Number of Cliques
An Analytic Formula

Laplace Matrix $\Delta = D - A$ contains all connectivity information.
Idea: Note the (i, j) element of A^2 is

$$(A^2)_{i,j} = \sum_{k=1}^{n} A_{i,k} A_{k,j} = |\{k : i \sim k \sim j\}|.$$

This means $(A^2)_{i,j}$ is the number of paths of length 2 that connect i to j.
Remark: The diagonal elements of $A(A^2 - D)$ represent twice the number of 3-cycles (= 3-cliques) that contain that particular vertex.
Conclusion:

$$X_3 = \frac{1}{6} trace\{A(A^2 - D)\} = \frac{1}{6} trace(A^3).$$

Exercise: Generalize this formula for X_4.

References

- B. Bollobás, **Graph Theory. An Introductory Course**, Springer-Verlag 1979. **99**(25), 15879–15882 (2002).
- F. Chung, **Spectral Graph Theory**, AMS 1997.
- F. Chung, L. Lu, The average distances in random graphs with given expected degrees, Proc. Nat.Acad.Sci. 2002.
- R. Diestel, **Graph Theory**, 3rd Edition, Springer-Verlag 2005.
- P. Erdös, A. Rényi, On The Evolution of Random Graphs
- G. Grimmett, **Probability on Graphs. Random Processes on Graphs and Lattices**, Cambridge Press 2010.
- J. Leskovec, J. Kleinberg, C. Faloutsos, Graph Evolution: Densification and Shrinking Diameters, ACM Trans. on Knowl.Disc.Data, **1**(1) 2007.