Random graphs

Leo Davy

Toulouse, May 27th 2019

Definition

A graph is a pair $G = (V, E)$:

- V is the vertex set
- E is the edge set, $E \subseteq \{\{x, y\} : x, y \in V\}$

Interest :

- Simple graphical representation
- Allows to describe phenomenas
- and much more

An example of real world graph

A collaboration graph:

- V the set of scientists
- $E = \{\{x, y\} : x \text{ and } y \text{ co-authored a paper}\}$

Experimental results:

- Positive density of triangles
- Degree distribution follows a power law

Remark

The Erdős collaboration graph is the subgraph of the above collaboration graph s.t. $d(\text{Erdős}, x) < \infty$ for all $v \in V$.

Another example of real world graph

The Web-graph:

- V the set of web pages
- $E = \{(x, y) : x, y \in V, x \text{ contains a link to } y\}$
- The graph is directed
- The degrees follow a power law

Definition

A random graph is a graph where nodes, edges are both are selected by a random procedure.

Restrict ourselves to *static* random graphs on n vertices.

The model:

- V is fixed ($V = [n] = \{1, 2, \ldots, n\}$).
- No loops.
- No parallel edges.
- Total number of possible edges : $N = \binom{n}{2} = \frac{n(n-1)}{2}$.

The Erdős-Rényi model

Definition (ER 1960)

$\mathcal{G}_{n,M}$: A graph is picked uniformly at random among all graphs with M edges on n vertices.

Let G a graph with M edges and n vertices, then

$$\mathbb{P}_M(G) = 1/\binom{N}{M} \tag{1}$$

The Erdős-Rényi model

Definition (Gilbert 1959)

$\mathcal{G}_{n,p}$: Each of the $N = \binom{n}{2}$ possible edges is chosen independently with probability $p \in [0,1]$.

Let G a graph with n vertices, then

$$\mathbb{P}_p(G) = p^{e_G}(1-p)^{N-e_G} \tag{2}$$

The stability number

Definition

$\alpha(G) :=$ largest set of vertices s.t. no vertices are adjacent.

Theorem

$\alpha(G \in \mathcal{G}_{n,p}) \leq \lceil 2p^{-1} \log n \rceil$ *w.h.p.*

Let $S \subseteq V$, $|S| = k+1$,

$$\mathbb{P}(\text{"S is a stable set"}) = (1-p)^{\binom{k+1}{2}} \tag{3}$$

Let $X_{k+1} = \sum_S \mathbb{1}(\text{"S is a stable set"})$.

Remark

We have $X_k = 0$ if $k > \alpha$.

$$\mathbb{E}X_{k+1} = \binom{n}{k+1}(1-p)^{\binom{k+1}{2}} \leq \frac{(ne^{-p\frac{k}{2}})^{k+1}}{(k+1)!} \tag{4}$$

The RHS goes to 0 if $k \geq 2p^{-1} \log n$.

Is $\mathcal{G}_{n,p}$ connected ?

Definition

A graph is connected if there is a path linking any two vertices.

Theorem

Let $p = \frac{\log n + c}{n}$ *with* $c \in \mathbb{R}$, *then*
$\mathbb{P}(G \in \mathcal{G}_{n,p} \textit{ is connected }) \longrightarrow_n e^{-e^{-c}}$.

Remark

Let $\epsilon > 0$

- If $p = (1+\epsilon)\frac{\log n}{n}$, then $\mathcal{G}_{n,p}$ is connected w.h.p.
- If $p = (1-\epsilon)\frac{\log n}{n}$, then $\mathcal{G}_{n,p}$ is disconnected w.h.p.

Is $\mathcal{G}_{n,p}$ connected ?

Definition

A graph is connected if there is a path linking any two vertices.

Theorem

Let $p = \frac{\log n + c}{n}$ *with* $c \in \mathbb{R}$, *then*
$\mathbb{P}(G \in \mathcal{G}_{n,p}$ *is connected* $) \longrightarrow_n e^{-e^{-c}}$.

In order to prove this result, we need to know the probabilities that $\mathcal{G}_{n,p}$ contains :

- Isolated vertex
- Isolated edges
- No isolated component of size between 3 and $\lceil \frac{n}{2} \rceil$

X_k : number of connected component of size k.
Does $\mathcal{G}_{n,p}$ contain an isolated edge ?

$$\mathbb{P}(X_2 \geq 1) \leq \mathbb{E}X_2 = \binom{n}{2} p((1-p)^{n-1})^2 \quad (5)$$
$$= \mathcal{O}(n^2 p e^{-2pn}) = \mathcal{O}(p) = o(1). \quad (6)$$

N

o isolated edge with high probability.

Does $\mathcal{G}_{n,p}$ contain an isolated component of size $3 \leq k \leq n/2$? First moment method requires the number of connected components of size k.

Remark

Any connected component on k vertices contains a tree on k vertices.

Theorem

Cayley's formula There are k^{k-2} possible spanning trees on k vertices.

Let q_k the probability that a set of k vertices doesn't connect with any other vertex outside the set.

$$\mathbb{P}(X_k \geq 1) \leq \mathbb{E}X_k \leq \binom{n}{k} k^{k-2} q_k \tag{7}$$

$$= p^{-1}(\mathcal{O}(\frac{\log n}{\sqrt{n}}))^k \tag{8}$$

$$\sum_{k=2}^{\lceil \frac{n}{2} \rceil} \mathbb{P}(X_k \geq 1) \leq o(1) + p^{-1} \sum_{k=3}^{\lceil \frac{n}{2} \rceil} A^k \longrightarrow 0 \tag{9}$$

With high probability $\mathcal{G}_{n,p}$ doesn't contain isolated component larger than 1.

What is the distribution of isolated vertices ?

Theorem

In $\mathcal{G}_{n,p}$ the number of isolated vertices follows a Poisson distribution of mean e^{-c}.

Then, $\mathbb{P}(X_1 = 0) = e^{-e^{-c}}$.
The probability that $\mathcal{G}_{n,p}$ is connected tends to $e^{-e^{-c}}$.

The phase transition What does $\mathcal{G}_{n,\frac{\lambda}{n}}$ look like ?

The subcritical phase: $\lambda < 1$

- Tree like components (well approximated by a GW process).
- Largest component of size $\mathcal{O}(\log n)$.

The supercritical phase: $\lambda > 1$

- A single complex Giant Component of size $\theta(n)$.
- All other components are simple and of size $\mathcal{O}(\log n)$.