Random Graphs and Networks – 1° LECTURE

martedì 25 gennaio 2022 09:46

A. Introduction: real-world networks

Examples

- social networks (fb, twitter, ...)
- collaboration networks (scientists, actors, ...)
- internet / www
- telecommunication networks / Electrical power grids
- neural networks

↳ Deterministic structures made by individuals and connections

↳ $G = (V, E) \qquad E \subset V \times V$

Comment: though deterministic, they are huge and their detailed structure is unpredictable. However, one can achieve

- local structure (typical)
- global properties → connectivity, average distance between points

"Universal" behavior: many real-world networks display similar behavior

- small-world property → average distance btw points is very small
- scale-free property → # connections of an individual has no typical scale

↳ hence abstract models → random graphs.

B. Graph setting

Definitions: $G = (V, E)$ where

Definitions: $G = (V, E)$ where

$V = V(G)$, $E = E(G)$, $E = \{(v,w) : v,w \in V , i \neq j\}$

- The edge $e = (v,w)$ is undirected ().

  It is possible to consider directed graph with edge-set $\vec{E} \ni \vec{e} = [v,w]$ ().

- v and w are neighbors if $(v,w) \in E(G)$: $v \sim w$

- $e_1, e_2 \in E$ are neighbors if incident to a common vertex (), then we write $e_1 \sim e_2$.

- A path from v to w, $v,w \in V$, is a sequence $\gamma = (v_0, \dots, v_k)$ s.t. $v_0 = v$, $v_k = w$ and

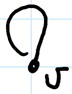
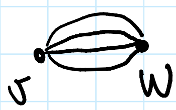
  $v_j \sim v_{j+1}$ $\forall$ $j = 0, \dots, k-1$ ()

  It can be viewed also as $\gamma = (e_1, \dots, e_k)$ where

  $e_j = (v_{j-1}, v_j)$ $\forall$ $j = 1, \dots, k$. Denoted as $\gamma : v \longrightarrow w$

- Length of a path $\gamma : v \longrightarrow w$ is $|\gamma| = \#$ edges
- If $\exists$ $\gamma : v \longrightarrow w$, hence v and w are connected and write

  $$v \longleftrightarrow w$$

- It is possible to consider also multigraphs G that may contain
  - loops :  (v,v)
  - multiple-edges 

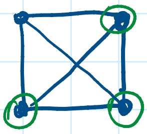
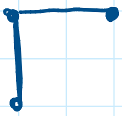
hence $E \subseteq V \times V$ with not necessarily distinct edges.

A multigraph without loop or multiple-edges is called a

A multigraph without loop or multiple-edges is called a <u>simple graph</u>.

- H is a <u>subgraph</u> of G $(H \subseteq G)$ if
  $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G|_{V(H)})$

  When $E(H) = E(G|_{V(H)})$ then H is an <u>induced sub-graph</u>

  Example: G:  H:  $\Rightarrow$ "$H \subseteq G$" "$H \not\subseteq G$"

- $\forall v \in V$, the degree of v is # neighbors of v:

$$d(v) \equiv d_G(v) = |\{w \in V : w \sim v\}|$$

<u>Note</u>: $$\sum_{v \in V} d_G(v) = \sum_{v \in V} \sum_{w \in V} \mathbb{1}\{v \sim w\} = 2|E|$$

- $\forall v, w \in V$, their <u>distance in G</u> is:

$$\mathrm{dist}_G(v,w) = \min_{\gamma : v \to w} |\gamma|$$ with convention $\min \emptyset = +\infty$

- $\mathrm{diam}(G) = \max\limits_{v,w \in V} \mathrm{dist}_G(v,w)$, <u>diameter of G</u>

- $\forall v \in V$, <u>the cluster of v</u> in G:

$$\mathcal{C}(v) = \{w \in V : v \leftrightarrow w\}$$

- The <u>largest component of G</u> (though not unique) is denoted $\mathcal{C}_{max}$ and is s.t

$$|\mathcal{C}_{max}| = \max_{v \in V} |\mathcal{C}(v)|$$

$v \in V$

ASSUMPTION: $V = [n] := \{1, 2, \dots, n\}$

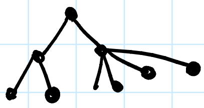
hence $E \subseteq \mathcal{E}_n = \{(i,j) : i,j \in [n],\ i \neq j\}$

Example:

- The complete graph on $[n]$ is $K_n = ([n], \mathcal{E}_n)$

  Note $|\mathcal{E}_n| = \binom{n}{2}$

- A tree (graph) over $[n]$ is s.t. $\forall\, i,j \in [n],\ i \neq j$ $\exists$ a unique path $\gamma : i \to j$. Hence it is connected and has no cycles:



Remember:

- $|E(\text{Tree on } [n])| = n-1$ (proof by induction)
- Cayley formula: $|\{\text{trees on } [n]\}| = n^{n-2}$

TOOL: To describe "typical properties" of a graph, we often consider the point of view of a vertex chosen u.a.r. in $[n]$. Let $U \sim \text{Uniform}[n]$.

For example: we can consider $d_G(U)$ and study its distribution: For $k \in \mathbb{N}$

$$P(d_G(U) = k) = \underbrace{\frac{1}{n} \overbrace{\sum_{j \in [n]} \mathbb{1}_{\{d_G(j) = k\}}}^{= n_k}}_{\text{empirical distribution of the vertex degree}} = \frac{n_k}{n}$$

= proportion of vertices with degree $= k$

C. Properties of real-world networks

1. **large size** $\longrightarrow$ Consider sequence of graph $(G_n)_{n\in\mathbb{N}}$ of size n, and study the limit $n \to \infty$

2. **Sparsity**: it means that $|E(G_n)| = \mathcal{O}(n)$.
In other words, that the vertex-degree is "tipically" finite indep. of n. $\longrightarrow$ Formally, we will often look at the following convergence:

$$P(d_{G_n}(U)=k) = \frac{1}{n}\sum_{j\in[n]} \mathbb{1}_{\{d_{G_n}(j)=k\}} \xrightarrow[n\to\infty]{} p_k, \quad k\in\mathbb{N} \quad (1)$$

where $(p_k)_{k\in\mathbb{N}}$ is a probability density $\left(\sum_{k\in\mathbb{N}} p_k = 1\right)$ with finite average μ. This is equivalent to ask

$$d_{G_n}(U) \xrightarrow[n\to\infty]{d} D, \quad D \overset{d}{\sim} (p_k)_{k\in\mathbb{N}}.$$

Indeed: $|E(G_n)| = \frac{1}{2}\sum_{j\in[n]} d_{G_n}(j) = \frac{n}{2}\sum_{k\in\mathbb{N}} k\cdot \frac{n_k}{n}$ $\quad = P(d_{G_n}(U)=k)$

$$\overset{\text{by }(1)}{\approx} \frac{n}{2}\sum_{k\in\mathbb{N}} k\cdot p_k = \frac{n}{2}\mu$$

$\longrightarrow$ sparse behavior.

3. **highly connected**: there is a component of size $O(n)$.
$\longrightarrow$ It amounts to verify: $\frac{|\mathcal{C}_{max}|}{n} \xrightarrow[n\to\infty]{} c > 0$

$\left(\text{or at least } \liminf_{n\to\infty} \frac{|\mathcal{C}_{max}|}{n} > 0\right)$

(or at least $\liminf_{n\to\infty} \frac{|\mathcal{E}_{max}|}{n} \to 0$)

4. <u>Scale-free property</u>: vertex-degree have no typical scale

$\longrightarrow$ it requires the analysis of $d_{G_n}(U)$, and it is modeled by <u>power-law distribution</u>.

<u>Def</u>: A real r.v. $X \geq 0$ has <u>power-law distr.</u> with exponent $\tau > 1$ if

$$\underline{\mathbb{P}(X \geq x)} \simeq c \cdot x^{-(\tau-1)}, \quad \begin{matrix} c>0 \\ \text{as } x \to \infty \end{matrix}$$

Formally:
$$\boxed{\lim_{x\to\infty} \frac{\log \mathbb{P}(X\geq x)}{\log\left(\frac{1}{x}\right)} = \tau - 1}$$

- If X has a density (discrete or cont.) f_X, we can equivalently say that $f_X(x) \simeq c \cdot x^{-\tau}$, as $x \to \infty$

Formally:
$$\lim_{x\to\infty} \frac{\log f_X(x)}{\log\left(\frac{1}{x}\right)} = \tau$$

<u>Remark</u>: Notice that $\forall\, a \geq \tau - 1$

$$\mathbb{E}\left(X^a\right) = +\infty \qquad \left(= \int_0^{+\infty} x^a \cdot f_X(x)\,dx \geq c \int_{x_0}^{+\infty} x^{a-\tau}\,dx = \infty \right)$$

Hence in general we will have to verify that

$d_{G_n}(U) \xrightarrow{d} D$ s.t.

$$\mathbb{P}(D=k) = p_k \quad (\text{as in } (1)) \quad \text{with } (p_k)_{k\in\mathbb{N}} \text{ s.t.}$$

$$\mathbb{P}(D \geq k) = \sum_{j\geq k} p_j \simeq c \cdot k^{-(\tau-1)}.$$

- Small-world: "typical distance" between vertices is very small w.r.t. n $\longrightarrow$ Take $U_1, U_2 \overset{d}{\sim} \text{Uniform}\,[n]$, independent, and study $d_{G_n}(U_1, U_2)$

Def: $(G_n)_{n\in\mathbb{N}}$ is small-world if $\exists\ K \in (0,\infty)$ s.t.

$$\lim_{n\to\infty} \mathbb{P}\left(d_{G_n}(U_1,U_2) \leq K \log n \right) = 1$$

Notice: $\log 10^6 = 6 \log 10$; $\log \log 10^6 = 0{,}95$

$\hookrightarrow$ In some cases, one is interested in studying the ultra-small-world property:

$$\mathbb{P}\left(d_{G_n}(U_1,U_2) \leq K \log\log n \right) \xrightarrow[n\to\infty]{} 1$$