

7. Graphes aléatoires

Dans ce chapitre, on s'intéresse à un autre modèle fondamental de graphes aléatoires : les graphes d'Erdős-Rényi. On utilisera `NetworkX` pour manipuler tous les graphes dans `Python` ; on renvoie à l'annexe pour une description des principales commandes utiles. En particulier, si `A` est une matrice d'adjacence, le graphe correspondant sera créé et affiché avec les commandes :

```
G = nx.Graph(A)
fig, ax0 = plt.subplots()
nx.draw_networkx(G, ax=ax0)
plt.show()
```

1. Modèle d'Erdős-Rényi.

Si $n \geq 1$ et $p \in (0, 1)$, le *graphe d'Erdős-Rényi* de paramètres n et p est le graphe aléatoire $G = (V, E)$ dont l'ensemble des sommets est $V = \llbracket 1, n \rrbracket$, et dont chaque arête possible $\{i, j\}$ avec $1 \leq i < j \leq n$ apparaît dans E avec probabilité p ; les variables de Bernoulli $B_{i,j}$ correspondantes sont supposées indépendantes. On note $G \sim G(n, p)$ pour indiquer qu'un graphe aléatoire est un graphe d'Erdős-Rényi de paramètres n et p .

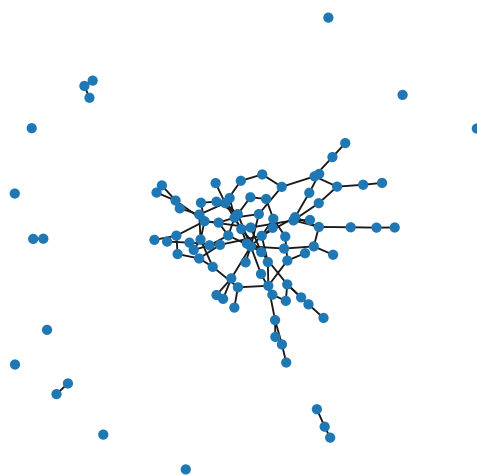


FIG. 7.1. Un graphe d'Erdős-Rényi de paramètres $n = 100$ et $p = 0.01$.

- (1) Quelle est la loi de $|E|$ si $G = (V, E) \sim G(n, p)$? Que vaut $\mathbb{E}[|E|]$? Comment choisir p en fonction de n pour que $|V|$ et $|E|$ soient du même ordre ?
- (2) Écrire des programmes `erdos_renyi(n, p)` et `dessin_erdos_renyi(n, p)` qui construisent et dessinent un graphe aléatoire d'Erdős-Rényi de paramètres n et p . On pourra utiliser l'argument `node_size` de `nx.draw_networkx` pour ajuster la taille des sommets et avoir des graphes lisibles. Expérimenter avec les paramètres $n = 100$ et $p \in \{0.005, 0.008, 0.01, 0.015, 0.02, 0.03\}$, et décrire les graphes obtenus.

(3) Pour mieux comprendre l'évolution des graphes d'Erdős–Rényi lorsque n est fixé et p augmente, on peut *coupler* les graphes $G(n, p)$ comme suit :

- on tire des variables indépendantes uniformes $U_{i,j} \in [0, 1]$;
- le graphe $G(n, p)$ est obtenu en posant $B_{i,j} = 1_{(U_{i,j} \leq p)}$.

Avec cette construction, $G(n, p_1)$ est un sous-graphe de $G(n, p_2)$ si $p_1 \leq p_2$; ceci permet de mieux comparer les graphes d'Erdős–Rényi de paramètres p différents. Écrire un programme `dessin_liste_erdos_renyi(n, P)` qui affiche plusieurs graphes d'Erdős–Rényi de paramètre p dans une liste P . Pour pouvoir suivre les sommets, on pourra :

- utiliser le même *Layout* pour les sommets de tous les graphes. Ainsi,

```
pos0 = nx.spring_layout(G)
...
nx.draw_networkx(H, pos=pos0)
```

fixe les positions de sommets dans un graphe G (on peut par exemple choisir le graphe de paramètre p maximal dans la liste P), puis impose ces positions aux sommets des autres graphes H .

- ou alors, faire évoluer le *Layout* au fur et à mesure ; on pourra explorer la documentation de `nx.spring_layout`.

Expérimenter avec $n = 100$ et un échantillonnage P des intervalles $[0.003, 0.007]$, $[0.007, 0.02]$ et $[0.02, 0.06]$.

2. Composantes isolées.

Dans cette section, $p = \frac{c}{n}$ avec $c \in [0, 1)$. On va voir que dans ce cas, le graphe d'Erdős–Rényi est constitué de petites composantes connexes qui sont presque toutes des arbres.

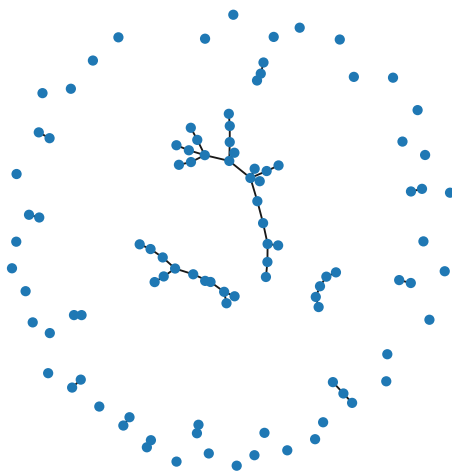


FIG. 7.2. Aspect d'un graphe d'Erdős–Rényi avant l'apparition d'une composante géante : $p = \frac{c}{n}$ avec $c < 1$. Toutes les composantes connexes sont de petits arbres.

(1) Si G est un graphe `NetworkX` et L est une partie de l'ensemble des sommets de G , `G.subgraph(L)` est le graphe induit par G sur cette partie. En utilisant la commande `nx.connected_components(G)`, écrire un programme `proportion_arbre` qui calcule la proportion de sommets d'un graphe $G = (V, E)$ qui sont dans des composantes

connexes qui sont des arbres :

$$P_{\text{arbre}} = \sum_{c \text{ composante connexe de } G} 1_{(c \text{ est un arbre})} \frac{|c|}{|V|}.$$

On pourra tester si un graphe est un arbre avec `nx.is_tree`.

(2) Si les graphes $G(n, p)$ sont couplés comme dans la section précédente, montrer que la proportion d'arbres est une fonction (aléatoire) décroissante de p . Avec $n = 10000$ et $p = \frac{c}{n}$, dessiner la fonction $c \mapsto P_{\text{arbre}}(G(n, \frac{c}{n}))$ pour $c \in [0, 2]$. Qu'observe-t-on ?

(3) Si $I = \{i_1, i_2, \dots, i_k\}$ est une partie de taille $k \geq 1$ de $\llbracket 1, n \rrbracket$ et T est un arbre sur cet ensemble de sommets, montrer que la probabilité pour que :

- I soit une composante connexe de $G(n, p)$ (donc, déconnectée de $\llbracket 1, n \rrbracket \setminus I$) ;
- le graphe induit par $G(n, p)$ sur I est T

est donné par la formule :

$$p_k = p^{k-1} (1-p)^{k(n-k) + \frac{(k-2)(k-1)}{2}}.$$

On admet la formule combinatoire suivante (*formule de Cayley*) : le nombre d'arbres sur un ensemble de sommets de taille k est k^{k-2} . En déduire une formule pour l'espérance de T_k , le nombre de composantes connexes de $G(n, p)$ qui sont des arbres de taille k .

(4) Montrer que si $c > 0$, alors pour tout $k \geq 1$,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[T_k(G(n, \frac{c}{n}))]}{n} = \frac{1}{c} \frac{k^{k-2}}{k!} (ce^{-c})^k.$$

Pour $|z| < e^{-1}$, on pose $T(z) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k$; la série est convergente par la formule de Stirling. Comme k^{k-1} est le nombre d'arbres étiquetés enracinés sur k sommets (c'est-à-dire avec une racine distinguée), $T(z)$ est la série génératrice exponentielle des arbres enracinés. Or, un arbre enraciné peut être vu comme une racine reliée à une union éventuellement vide d'autres arbres enracinés ; ceci implique la formule

$$T(z) = ze^{T(z)}$$

pour tout z complexe de module inférieur à e^{-1} . Ceci implique en particulier que

$$\lim_{\substack{s \rightarrow e^{-1} \\ s < e^{-1}}} T(s) = 1$$

et que T établit une bijection croissante entre $[0, e^{-1})$ et $[0, 1)$. Si $c = T(s)$ avec $c \in [0, 1)$, on a alors :

$$\frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k = \frac{1}{T(s)} T(T(s) e^{-T(s)}) = \frac{T(s)}{T(s)} = 1.$$

(5) Quel est le lien entre la proportion P_{arbre} et les variables T_k , $k \geq 1$? Déduire du calcul ci-dessus le fait suivant : si $c \in [0, 1)$, alors

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[P_{\text{arbre}}\left(G\left(n, \frac{c}{n}\right)\right)\right] = 1.$$

(6) Montrer aussi que si $c < 1$, alors il n'y a pas de *composante géante* dans $G(n, p)$: pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[G(n, p) \text{ contient une composante connexe de taille plus grande que } \varepsilon n] = 0.$$

3. Composante géante.

Dans cette section, $p = \frac{c}{n}$ avec $c > 1$ fixé. Dans ce cas, le graphe d'Erdős–Rényi est constitué d'une composante géante de taille $O(n)$, et d'autres composantes connexes beaucoup plus petites.

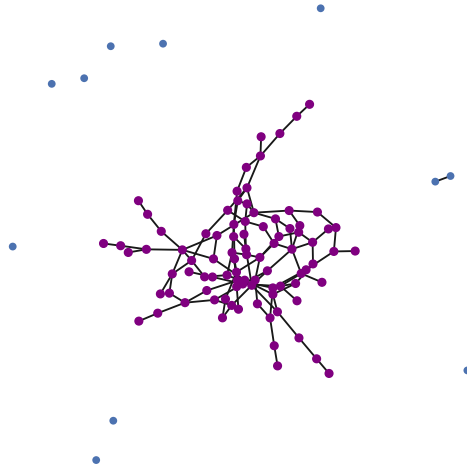


FIG. 7.3. Aspect d'un graphe d'Erdős–Rényi après l'apparition d'une composante géante : $p = \frac{c}{n}$ avec $c > 1$. Il y a une composante géante d'ordre $O(n)$, et les autres composantes sont de taille $O(\log n)$.

- (1) Modifier le programme `dessin_erdos_renyi` pour que la plus grande composante connexe (ou l'une des plus grandes composantes connexes s'il y en a plusieurs de même taille) soit affichée avec une autre couleur. Expérimenter avec $n = 100$ et $p = \frac{c}{n}$ avec $c \in \{0.005, 0.006, \dots, 0.015\}$.

- (2) Si les graphes $G(n, p)$ sont couplés comme précédemment, montrer que la fonction aléatoire

$$p \mapsto \frac{\text{taille de la plus grande composante connexe de } G(n, p)}{n}$$

est croissante. Dessiner cette fonction pour $n = 10000$ et $p = \frac{c}{n}$ avec c variant entre $[0.5, 3]$. Qu'observe-t-on ?

- (3) Remarquons que la fonction $x \mapsto xe^{-x}$ est croissante sur $[0, 1]$ et décroissante sur $[1, +\infty)$; par conséquent, si $c > 1$, il existe un unique c^* tel que $c^*e^{-c^*} = ce^{-c}$. Comparer la fonction précédente sur $[1, 3]$ à la fonction $c \mapsto (1 - \frac{c^*}{c})$. Pour trouver c^* , on pourra utiliser `fsolve`, qu'on importera avec `from scipy.optimize import fsolve`.
- (4) Vérifier par la simulation que les autres composantes connexes de $G(n, \frac{c}{n})$ avec $c > 1$ sont d'ordre $O(\log n)$. Ainsi, il y a une unique composante connexe géante pour $c > 1$.

On peut montrer que si $c > \log n$, alors la composante géante absorbe tous les sommets : le graphe est connexe avec grande probabilité pour $p > \frac{\log n}{n}$. Plus précisément, si $c = \log n + d$ avec $d \in \mathbb{R}$ fixé, alors

$$\lim_{n \rightarrow \infty} \mathbb{P}\left[G\left(n, \frac{c}{n}\right) \text{ est connexe}\right] = e^{-e^{-d}}.$$