

Introduction

On peut faire l'expérience de pensée suivante : on place un grand nombre d'individus qui ne se connaissent pas dans un lieu isolé et on les laisse évoluer librement. Ces individus vont interagir, et naturellement des communautés vont se former. Selon la fréquences des interactions, ces communautés vont être plus ou moins grandes. Un premier effet qu'on peut alors observer est l'apparition d'une communauté "géante" : notre population s'est alors, en quelque sorte, mondialisée. Un deuxième effet possible est l'apparition des communautés de communautés, car certains individus font le lien entre différentes communautés.

Dans cette étude, on cherchera ainsi à modéliser une population sur un graphe plus ou moins aléatoire. Le caractère déterministe permet de représenter une population au plus proche de la réalité en se fondant sur des données statistiques, mais pour avoir un modèle général simple, on fera intervenir les probabilités.

Les modèles introduits restent très généraux, mais on voit aisément que beaucoup d'applications seront envisageables : on pourra l'adapter à la propagation d'une information, d'une rumeur, au bouche-à-oreille ou encore aux réseaux sociaux. Nous ne prétendons pas ici présenter un modèle très réaliste prêt à l'emploi, cependant les concepts introduits sont féconds et peuvent être le point de départ de nombreuses études stochastiques dans des domaines pratiques.

1 L'effet de "mondialisation"

Une idée naturelle est de représenter la situation par un graphe, les individus étant les sommets et les interactions les arêtes. Tout l'enjeu de la modélisation sera la répartition de ces arêtes. Pour cela, on utilisera une approche probabiliste. Puisqu'on considère une grande population, on utilisera des graphes infinis ou du moins des graphes finis dont on étudiera les lois de probabilité dans la limite $n \rightarrow \infty$. On représentera la population par un graphe support déterministe mais sur lequel on ne conservera que certaines arêtes avec une probabilité p qu'on appellera *paramètre d'interaction*, pour modéliser la propension des individus à interagir entre eux.

1.1 Quelques résultats sur la percolation

La théorie de la percolation donne un premier résultat très fort qui s'applique à notre situation et explique l'effet de "mondialisation", qui est indépendant du choix du graphe support.

Définition Un graphe $G = (V, E)$ est la donnée d'un ensemble de *sommets* V et d'un ensemble d'*arêtes* $E \subset \mathcal{P}_2(V)$. On se limitera ici à des arêtes reliant deux sommets différents, et à E dénombrable. On note Γ l'ensemble de ces graphes.

Construction Soit $G = (V, E)$ un graphe dit *support*. Soit l'univers $\Omega := \{0, 1\}^E$, $\omega \in \Omega$ une configuration. Une arête $e \in E$ est dite ouverte si $\omega(e) = 1$ et fermée si $\omega(e) = 0$. Soit \mathcal{F} la tribu engendrée par les cylindres, c'est-à-dire les parties de Ω qui fixe l'état d'un nombre fini d'arêtes. On définit sur (Ω, \mathcal{F}) la mesure produit $P = P_p := \prod_{e \in E} \mu_e$ où l'état de chaque arête est décrit par la probabilité $\mu_e := p1_1 + (1-p)1_0$. On identifie une configuration $\omega \in \Omega$ avec le sous-graphe $G' = (V, E')$ où $E' = \{e \in E, \omega(e) = 1\}$

Définition Si $x, y \in V$, on note $x \leftrightarrow y :=$ "il existe un chemin ouvert de x à y ". Cela définit une relation d'équivalence sur V dont les classes sont appelées composantes connexes et notées $C(x)$ pour $x \in V$. On note $x \leftrightarrow \infty := \{|C(x)| = \infty\}$, et on dit qu'il y a *percolation* s'il existe $x \in V$ tel que $x \leftrightarrow \infty$. On note $\psi(p)$ sa probabilité.

Remarque La percolation est bien un événement, par réunions et intersections dénombrables sur des cylindres :

$$\bigcup_{x \in V} \bigcap_{n \in \mathbb{N}} \bigcup_{y \in V} \bigcup_{\substack{(v_1, \dots, v_n) \in (V \setminus \{x, y\})^n \\ (x, v_1) \in E, (v_1, v_2) \in E, \dots, (v_{n-1}, v_n) \in E, (v_n, y) \in E}} \{ \omega \in \Omega, \omega(x, v_1) = 1, \omega(v_1, v_2) = 1, \dots, \omega((v_{n-1}, v_n)) = 1, \omega(v_n, y) = 1 \}$$

On introduit la loi du 0-1 de Kolmogorov dans une formulation simplifiée adaptée à notre situation.

Théorème 1.1. (Loi du 0-1 de Kolmogorov) *Soit A un événement qui ne dépend pas de l'état d'un nombre fini d'arêtes. Alors, $P(A) \in \{0, 1\}$.*

Preuve Notons \mathcal{C} l'ensemble des cylindres. Soit B un événement ne dépendant que d'un nombre fini d'arêtes, i.e. $B \in \mathcal{C}$. Alors, A et B sont indépendants. Supposons que $P(A) \neq 0$. Alors, P et $P(\cdot|A)$ sont deux probabilités sur (Ω, \mathcal{F}) (où $\mathcal{F} = \sigma(\mathcal{C})$), égales sur \mathcal{C} . Soit \mathcal{F}' l'ensemble des événements E de \mathcal{F} tels que $P(E) = P(E|A)$ (on vérifie que c'est une tribu contenant \mathcal{C}). Or, \mathcal{F} est la plus petite tribu contenant \mathcal{C} , donc, $\mathcal{F}' = \mathcal{F}$. En particulier, $P(A) = P(A|A) = 1$. \square

Corollaire 1.2. $\psi(p) \in \{0, 1\}$

Proposition 1.3. $p \mapsto \psi(p)$ est croissante

Preuve Soit $p \leq p'$. On réalise un couplage croissant de P_p et $P_{p'}$. Soit $e \in E$. Il existe X_e et Y_e deux variables de Bernoulli de paramètres respectifs p et p' telles que $X_e \leq Y_e$. En effet, soit X_e et Z_e deux variables de Bernoulli de paramètres respectifs p et α indépendantes (qui existent d'après le théorème d'existence). Alors, posons $Y_e = X_e 1_{Z_e=0} + 1_{Z_e=1}$. C'est une variable de Bernoulli de paramètre $E(Y_e) = p(1 - \alpha) + \alpha$. On pose $\alpha = \frac{p'-p}{1-p}$, de sorte que $Y_e \sim \mathcal{B}(p')$, et on a bien $X_e \leq Y_e$. Il suffit alors de choisir $(X_e, Z_e)_{e \in E}$ mutuellement indépendantes (d'après le théorème d'existence, E étant dénombrable), et de poser $\omega(e) = X_e$ et $\omega'(e) = Y_e$ pour toute arête e . On définit alors la mesure de probabilité P sur $(\{0, 1\}^E)^2$ comme la loi de (ω, ω') . Alors,

$$\psi(p) = P(\exists x \in V, |C(x)| = \infty \text{ dans } \omega) \leq P(\exists x \in V, |C(x)| = \infty \text{ dans } \omega') = \psi(p') \quad \square$$

On définit la probabilité critique $p_c := \sup\{p \in [0, 1], \psi(p) = 0\}$. On déduit des résultats (1.2) et (1.3) le résultat fondamental suivant, qui caractérise la transition de phase de la percolation.

Théorème 1.4. Dans le processus de percolation de Bernoulli :

- \triangleright Si $p < p_c$, $\psi(p) = 0$: il n'y a presque sûrement pas percolation. On parle de régime sous-critique.
- \triangleright Si $p > p_c$, $\psi(p) = 1$: il y a presque sûrement percolation. On parle de régime sur-critique.

1.2 Une première modélisation : le réseau carré

Remarque On cherche maintenant à appliquer ces résultats très généraux à des graphes qui modélisent bien les interactions entre individus. On peut commencer par essayer de modéliser les interactions par proximité : plus les individus sont proches les uns des autres, plus ils vont interagir. Pour modéliser cet effet, on choisit un graphe support dont seuls les sommets les plus proches peuvent être reliés par une arête. Le plus simple est donc de partir d'un réseau carré, et d'augmenter la dimension pour augmenter le nombre d'interactions. Chaque dimension supplémentaire peut aussi représenter une nouvelle sorte d'interaction.

Définition Soit $d \in \mathbb{N}^*$. On note \mathbb{L}^d le réseau carré de dimension d (figures 1 et 2), c'est-à-dire le graphe infini $(\mathbb{Z}^d, \{(x, y) \in (\mathbb{Z}^d)^2, \|x - y\|_2 = 1\})$. Vu l'invariance par translation, on définit $\theta(p) := P(0 \leftrightarrow \infty)$.

Lemme 1.5. $\psi(p) = 0 \iff \theta(p) = 0$

Preuve Par inclusion, $\theta(p) \leq \psi(p)$. Réciproquement, si $\theta(p) = 0$, alors $\forall x \in V, P(x \leftrightarrow \infty) = 0$, donc $\psi(p) = P\left(\bigcup_{x \in V} \{x \leftrightarrow \infty\}\right) \leq \sum_{x \in V} P(x \leftrightarrow \infty) = 0$ (V dénombrable). \square

Proposition 1.6. $p_c(\mathbb{L}^1) = 1$

Théorème 1.7. Soit $d \geq 2$. Alors

$$0 < p_c(\mathbb{L}^d) < 1$$

Preuve

- \triangleright Montrons que $p_c > 0$, ce qui revient à montrer qu'il existe $p > 0$ tel que $\theta(p) = 0$. On appelle chemin auto-évitant un chemin qui ne passe pas deux fois par le même sommet. Soit $n > 0$, $\sigma(n)$ le nombre de tels chemins de longueur n depuis 0 et $N(n)$ le nombre de ces chemins ouverts. Alors $N(n) \sim \mathcal{B}(\sigma(n), p^n)$ et, par l'inégalité de Markov :

$$\theta(p) \leq P(N(n) \geq 1) \leq E(N(n)) = p^n \sigma(n)$$

Or, on a $2d$ choix pour le premier voisin et au plus $2d - 1$ choix ensuite, donc,

$$\theta(p) \leq \frac{2d}{2d-1} ((2d-1)p)^n$$

Il suffit de choisir $p < \frac{1}{2d-1}$ et de faire tendre n vers $+\infty$ pour avoir le résultat.

▷ Montrons que $p_c < 1$. Comme \mathbb{L}^d s'injecte canoniquement dans \mathbb{L}^{d+1} , $d \mapsto p_c(\mathbb{L}^d)$ est décroissante. Tout revient donc à montrer que $p_c(\mathbb{L}^2) < 1$, c'est-à-dire à trouver $p < 1$ tel que $\theta(p) > 0$ en dimension 2. Or, l'intérêt de la dimension 2 est la notion de graphe *dual* (figure 3) :

$$\mathbb{L}_d^2 := \left(\mathbb{Z}^2 + \left(\frac{1}{2}, \frac{1}{2} \right), \mathbb{B}^2 + \left(\frac{1}{2}, \frac{1}{2} \right) \right)$$

Au sous-graphe aléatoire G de \mathbb{L}^2 obtenu par percolation de paramètre p , on associe le sous-graphe G' de \mathbb{L}_d^2 dont les arêtes fermées sont celles qui coupent les arêtes ouvertes de G . G' correspond à un processus de percolation de paramètre $1-p$ sur \mathbb{L}_d^2 . La composante connexe de 0 dans G est finie si et seulement s'il existe un chemin cyclique de G' entourant 0 (notons \mathcal{C} cet ensemble). Donc :

$$1 - \theta(p) = P \left(\bigcup_{C \in \mathcal{C}} \{C \text{ est ouvert}\} \right) \leq \sum_{C \in \mathcal{C}} P(\{C \text{ est ouvert}\})$$

Or, le nombre de chemins cycliques entourant 0 de longueur n est majoré par $n3^n$: on a n choix pour l'ordonnée d'un point d'abscisse 0, puis 3 choix à chaque étape au plus. De plus, $P(\{\mathcal{C} \text{ est ouvert}\}) = (1-p)^n$, donc :

$$1 - \theta(p) \leq \sum_{n=1}^{\infty} n3^n(1-p)^n$$

Il suffit de choisir $p < 1$ assez grand de sorte que cette somme soit strictement inférieure à 1. \square

Commentaire La non-trivialité de la probabilité critique est un résultat fort : lorsqu'on modélise la population de façon très régulière avec un réseau, il peut de fait exister deux situations : une dans laquelle les communautés restent isolés, et une dans laquelle une communauté géante apparaît.

2 L'effet "Milgram"

2.1 Présentation du modèle

On peut cependant choisir de modéliser différemment la situation en considérant que les individus ne sont pas influencés par la proximité de leurs voisins pour interagir. Ainsi, dans le monde d'aujourd'hui, des milliers de personnes se croisent dans la rue sans interagir, et parcourent des distances toujours plus grandes pour se rendre sur leur lieu de travail ou de vie pour interagir. Il faut donc un modèle permettant de modéliser des interactions indépendantes de la distance. Plutôt qu'un réseau, on modélise donc la population par un graphe complet (toutes les interactions sont envisageables) sur lequel on va aléatoirement ouvrir et fermer des arêtes.

Définition Soit $n \in \mathbb{N}^*$, $p \in [0, 1]$. Soit $\mathcal{G}(n)$ le graphe complet à n sommets et $\binom{n}{2} = \frac{n(n-1)}{2}$ arêtes. On appelle graphe d'Erdős-Rényi $\mathcal{G}(n, p)$ le sous-graphe aléatoire obtenu comme précédemment (figure 4).

Remarque Soit N_{p_n} le nombre de voisins d'un sommet choisi aléatoirement sur $\mathcal{G}(n, p_n)$. N_{p_n} suit une loi binomiale $\mathcal{B}(n-1, p_n)$, et le degré moyen du graphe est $\Delta := E(N_{p_n}) = (n-1)p_n$. Dans la suite, on considère que $\forall n \in \mathbb{N}^*$, $p_n = \frac{\Delta}{n}$ pour Δ fixé (le degré moyen n'est *a priori* pas une fonction de la taille de la population).

2.2 Probabilité critique sur $\mathcal{G}(n, p_n)$

Dans cette section, on cherche à montrer le résultat intuitif selon lequel le seuil critique sur le graphe d'Erdős-Rényi correspond à un degré moyen égal à 1.

Méthode On se place en régime sous-critique, et on se concentre sur les arêtes plutôt que sur les sommets, pour simplifier la généralisation du raisonnement qu'on fera en 3. On procède à un algorithme de parcours : on choisit une arête aléatoirement, puis on visite les arêtes qui lui sont adjacentes (qui ont un sommet en commun avec elle) et on itère ce processus jusqu'à avoir parcouru toute sa composante connexe. On appelle sous-graphe envahi depuis un couple (sommet, arête) le sous-graphe parcouru par cet algorithme qui visite à un certain moment l'arête considérée en arrivant d'une arête adjacente par rapport à l'autre sommet.

On introduit cinq variables aléatoires :

- R : nombre d'arêtes adjacentes à une arête choisie aléatoirement
- Q : nombre d'arêtes adjacentes à un couple (arête, sommet)
- I : taille de la composante connexe d'une arête choisie aléatoirement
- H : taille du sous-graphe envahi par un couple (sommet, arête) choisi aléatoirement
- H_m : taille sommée des sous-graphes envahis par m couples (sommet, arête) choisis aléatoirement

On admet le résultat suivant, d'une part car il n'est absolument pas évident, quoiqu'intuitif, d'autre part car la taille des composantes connexes n'est pas l'objet de cette étude. Il donne une condition suffisante pour connaître la probabilité critique.

Lemme 2.1. $\forall p < p_c, E(I_p) < \infty$. S'il existe $p_0 \leq p_c$ tel que $E(I_p) \xrightarrow{p \rightarrow p_0^-} +\infty$, alors $p_c = p_0$.

Expression de $E(I)$ On va utiliser les fonctions génératrices. En régime sous-critique, la probabilité que deux arêtes choisies au hasard soient dans la même composante connexe tend vers 0, donc $G_{H_m}(x) = [G_H(x)]^m$. Il en est de même de la probabilité qu'on ait une boucle dans une composante connexe, donc :

$$\forall k \in \mathbb{N}^*, P(H = k) = \sum_{l=0}^{\infty} P(Q = l)P(H_l = k - 1) \text{ et } P(I = k) = \sum_{l=0}^{\infty} P(R = l)P(H_l = k - 1)$$

Et :

$$\begin{aligned} \forall x \in [-1, 1], G_H(x) &:= E(x^H) = \sum_{k=0}^{\infty} \left[\sum_{l=0}^{\infty} P(Q = l)P(H_l = k - 1) \right] x^k \\ &= \sum_{k=0}^{\infty} \left[\sum_{l=0}^{\infty} P(Q = l) \frac{1}{(k-1)!} \left(\frac{d^{k-1}}{dx^{k-1}} [G_H(x)]^l \right)_{x=0} \right] x^k \\ &= \sum_{l=0}^{\infty} x P(Q = l) \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{d^k}{dx^k} [G_H(x)]^l \right)_{x=0} x^k \\ &= \sum_{l=0}^{\infty} x P(Q = l) [G_H(x)]^l \\ &= x G_Q(G_H(x)) \end{aligned}$$

Notons que, puisque cette double somme converge avec des valeurs absolues, on a pu intervertir les deux symboles somme. En faisant le même raisonnement avec I , on aboutit aux deux équations fonctionnelles :

$$G_H(x) = x G_Q(G_H(x)) \text{ et } G_I(x) = x G_R(G_H(x))$$

Alors, comme les variables étudiées sont positives, considérons :

$$E(I) = G'_I(1) = G'_R(G_H(1)) + G'_H(1)G'_R(G_H(1)) = G'_R(1) + G'_H(1)G'_R(1) \quad (1)$$

$$E(H) = G'_H(1) = G'_R(G_H(1)) + G'_H(1)G'_R(G_H(1)) = G'_R(1) + G'_H(1)G'_R(1)$$

Donc :

$$E(H) = \frac{1}{1 - G'_Q(1)}$$

Et en réinjectant dans (1),

$$E(I) = 1 + \frac{G'_R(1)}{1 - G'_Q(1)} = 1 + \frac{E(R)}{1 - E(Q)} \quad (2)$$

Or, $Q \sim \mathcal{B}(n-2, p)$ et dans la limite $n \rightarrow \infty$, $E(Q) = np$. Donc (2) diverge pour $E(Q) = 1$.

Théorème 2.2. La probabilité critique sur le graphe $\mathcal{G}(n, p)$ dans la limite $n \rightarrow \infty$ est :

$$p_c = \frac{1}{n}$$

Commentaire Ainsi, la probabilité critique sur le graphe d'Erdős-Rényi tend vers 0. Même si le paramètre d'interaction p est très faible, la structure du graphe entraîne la mondialisation.

3 L'effet "communautés de communautés"

L'inconvénient de la modélisation précédente est sa structure support de graphe complet. Lorsqu'on considère une très grande population, des communautés se forment et interagissent ensuite préférentiellement entre elles. Un moyen de modéliser ce phénomène sans changer la structure du graphe est de considérer non pas les composantes connexes des arêtes, mais celles des k -cliques, qui vont représenter des communautés. Ce modèle permet aussi de contrer l'inconvénient d'une modélisation statique d'un phénomène dynamique, en repérant des communautés qui ont de fortes chances de correspondre à des liens fréquents.

Définition Soit $k \leq 2$. Une k -clique du graphe $\mathcal{G}(n, p)$ est un sous-graphe de k sommets complet. Deux k -cliques différentes sont dites adjacentes si elles ont en commun $k - 1$ sommets.

Méthode L'étude est la même que précédemment en transposant l'étude des arêtes (qui sont des 2-cliques) aux k -cliques. La condition (2) est toujours vérifiée, et, pour $n \rightarrow \infty$, $Q \sim \mathcal{B}(n - k, (k - 1)p^{k-1}(1 - p) + p^k)$. Donc,

$$E(Q) = (n - k)[(k - 1)p^{k-1}(1 - p) + p^k] \approx n(k - 1)p^{k-1}$$

Théorème 3.1. La probabilité critique de k -clique percolation sur le graphe $\mathcal{G}(n, p)$ est :

$$p_c = \frac{1}{[n(k - 1)]^{\frac{1}{k-1}}}$$

Commentaire On a toujours $p_c \rightarrow \infty$ pour un paramètre k faible, mais l'autre cas extrême, à savoir un paramètre k qui varie dans les mêmes proportions que n , entraîne une probabilité critique proche de 1 (figure 5).

4 Un modèle plus réaliste

Jusqu'ici, on a modélisé la population par un graphe complet, mais on peut envisager des distributions plus réalistes, qui pourraient se fonder sur des données connues. Le raisonnement effectué en 2.2 se généralise en fait aisément à tout graphe dont la distribution de degrés est donnée. Pour distinguer des variables aléatoires, on notera ses moments $(\langle k^m \rangle)_{m \in \mathbb{N}^*}$. L'équation (2) est toujours vérifiée. Notons S le nombre d'arêtes adjacentes à un couple (arête, sommet) choisi aléatoirement dans le graphe *support*.

$$\begin{aligned} G_Q(x) &= \sum_{k=0}^{\infty} P(Q = k)x^k \\ &= \sum_{k=0}^{\infty} \sum_{l=k}^{\infty} P(Q = k | S = l)P(S = l)x^k \\ &= \sum_{k=0}^{\infty} \sum_{l=k}^{\infty} \binom{l}{k} p^k (1 - p)^{l-k} P(S = l)x^k \\ &= \sum_{l=0}^{\infty} P(S = l)(1 + p(x - 1))^l \\ &= G_S(1 + p(x - 1)) \end{aligned}$$

D'où le résultat tout à fait intuitif :

$$E(Q) = p G'_S(1) = p E(S)$$

Or, contrairement à précédemment, on a plus de chances de tomber sur un sommet à haut degré en choisissant une arête aléatoirement. Donc : $\forall i \in \{0, \dots, n - 1\}$,

$$\begin{aligned} P(S = i) &= \frac{|\{\text{couples (arête, sommet) dont le sommet est de degré } i + 1\}|}{|\{\text{couples (arête, sommet)}\}|} = \frac{(i + 1)p_{i+1}}{\langle k \rangle} \\ E(S) &= \sum_{i=0}^{n-1} i \frac{(i + 1)p_{i+1}}{\langle k \rangle} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \end{aligned}$$

D'où :

Théorème 4.1. La probabilité critique sur un graphe de distribution de degrés donnée est :

$$p_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

Commentaire Ce résultat est bien cohérent avec le résultat 2.2. Ici, en changeant la structure du graphe support, on n'a plus de dépendance directe en n . Avec ce modèle plus réaliste, on se ramène à une gestion purement statistique des données. Il suffira donc de jouer sur $\langle k \rangle$ et $\langle k^2 \rangle$ pour augmenter le seuil critique. Notons que si $2\langle k \rangle > \langle k^2 \rangle$, alors il n'y aura presque sûrement jamais percolation.

Conclusion

La fécondité du modèle de la percolation duquel nous sommes partis est très importante. Ici, nous avons seulement vu comment les résultats mathématiques de la transition de phase de percolation pouvaient s'appliquer à la modélisation de populations sur différents graphes. Cette application peut en fait s'étendre à toute étude globale d'un phénomène à propagation locale, et des modèles plus poussés donnent aujourd'hui des résultats utiles à une meilleure connaissance de ces phénomènes.

Bibliographie

- [1] G.Grimmett. *Percolation*. Springer-Verlag, 1989.
- [2] M.E.J. Newman, *Physical Review* (E 66, 016128), 2002.
- [3] G.Palla, I. Derényi, T.Vicsek. "The critical point of k-clique percolation in the Erdős-Rényi graph". *Journal of Statistical Physics* (128, 219-227), 2007.