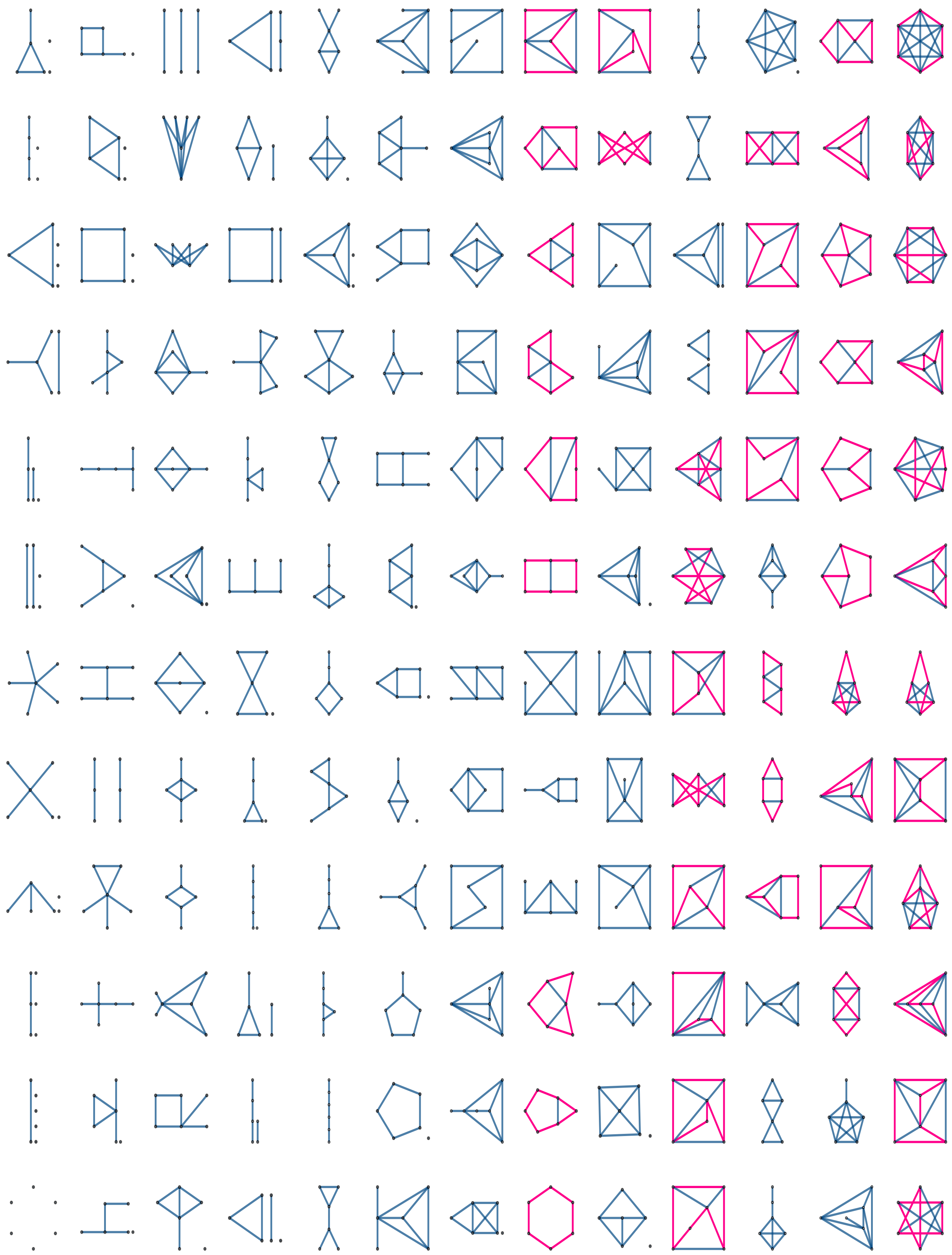


◀ *Représentation des 156 graphes à six sommets. En rouge, des exemples de chemins hamiltoniens qui relient tous les sommets.*

INFOGRAPHIES : NICOLAS CURIEN/UNIVERSITÉ PARIS-SACLAY - DR

**Nicolas Curien**
MATHÉMATICIEN, UNIVERSITÉ PARIS-SACLAY
*Professeur à l'université de Paris-Saclay, il est spécialiste des probabilités et de géométrie aléatoire. Il y donne le cours de M2 sur les graphes aléatoires.*

# Prévoir les transitions dans les graphes aléatoires

*Les changements d'état de la matière sont régis par des règles thermodynamiques. En mathématiques, elles sont expliquées par les phénomènes de « seuil » où, dans de grands systèmes désordonnés – les graphes aléatoires –, des propriétés déterministes émergent brusquement lors d'une petite variation des paramètres. Une conjecture proposée en 2006, portant sur le calcul du point d'apparition de telles transitions, vient d'être prouvée en mars dernier par une mathématicienne coréenne et un doctorant vietnamien. Leur preuve, courte et astucieuse, ne comporte que des notions de niveau licence. Rafraîchissant, à une époque d'inflation des démonstrations.*

Les avancées en mathématiques ne sont pas toujours le labeur du travail acharné d'un génie solitaire qui, après dix ans de travail, convainc ses pairs d'un résultat marquant en échafaudant une théorie complexe, profonde… mais souvent hermétique aux non-spécialistes – on peut penser à la conjecture de Fermat démontrée par le mathématicien britannique Andrew Wiles en 1994 (*). Parfois, un théorème est démontré (ou redémontré) de manière inattendue, à l'aide d'une astuce limpide à laquelle personne n'avait pensé, et qui peut être accessible à un grand nombre de personnes. Le célèbre mathématicien hongrois Paul Erdös (1913-1996) parlait du *Livre*, *The Book*, dans lequel « Dieu » consignerait les preuves les plus élégantes et intelligibles de chaque théorème mathématique : *« Vous pouvez ne pas croire en Dieu*, lança-t-il lors d'une conférence en 1985, *mais vous devriez croire en l'existence du* Livre. » Nul doute que Jinyoung Park et Huy Tuan Pham, tous deux à l'université Stanford (États-Unis), ont découvert l'une de ces preuves récemment : dans un article mis en ligne le 31 mars 2022 sur la plateforme de prépublication Arxiv, la jeune mathématicienne coréenne et son collègue vietnamien, encore en thèse, ont résolu une conjecture proposée en 2006 par l'Américain Jeff Kahn et l'Israélien Gil Kalai. Cet énoncé portait sur l'estimation du moment où les transitions de phases apparaissent dans les graphes aléatoires. D'abord surprise, la communauté mathématique fut immédiatement convaincue par ce résultat qui figure déjà dans les livres de référence du domaine.

Revenons sur l'objet d'étude : les graphes. C'est l'un des concepts mathématiques les plus simples que l'on puisse imaginer : des points (les sommets) reliés par des lignes (les arêtes). Si le graphe possède $n$ sommets, il y a donc $n(n-1)/2$ arêtes possibles (paires de points différentes). Le nombre de « formes » de graphes avec $n$ sommets grossit très vite : avec six sommets on a déjà 156 possibilités (voir l'image d'ouverture), avec sept sommets il y a 1 044 graphes différents et, avec dix sommets, on dépasse les douze millions de graphes possibles…

Il est d'usage de faire remonter la théorie des graphes à 1736, lorsque le mathématicien allemand Leonhard Euler (1707-1783) proposa le problème des « sept ponts de Königsberg » : dans cette ville – aujourd'hui Kaliningrad, en Russie –, existe-t-il une promenade qui permette de parcourir tous les ponts une seule fois ? En représentant la ville sous la forme d'un graphe, Euler parvient à démontrer qu'un tel chemin n'existe pas, signant ainsi l'acte de naissance de la théorie des graphes.

Aujourd'hui, le domaine des théories des graphes est un vaste champ des mathématiques qui offre de multiples applications en physique ou en informatique (la structure des graphes est particulièrement bien adaptée à l'organisation des données). Il est aussi au cœur des techniques d'apprentissage statistique *(deep learning)*, qui sont en train de changer notre quotidien. De nombreux réseaux complexes tels les réseaux sociaux, les réseaux d'interactions en biologie ou le web sont modélisés par des graphes aléatoires.

### L'EAU LIQUIDE DEVIENT BRUSQUEMENT VAPEUR AU-DELÀ DE 100 °C

Mais qu'est-ce qu'un graphe aléatoire ? En 1959, les mathématiciens hongrois Paul Erdös – celui du *Livre* – (1913-1996) et Alfred Rényi (1921-1970) ont proposé le modèle le plus simple et naturel de graphe aléatoire : à partir de $n$ sommets distincts, ajoutez des arêtes une par une, uniformément au hasard, entre ces sommets. D'abord une, puis deux, puis trois, et ainsi de suite de façon aléatoire, jusqu'à ce que toutes les arêtes soient tracées (c'est-à-dire quand tous les sommets sont reliés à tous les autres). On notera dans la suite $G(n,m)$ le graphe aléatoire obtenu lorsque $m$ arêtes ont été ajoutées.

La problématique à laquelle nous allons nous intéresser maintenant est celle des transitions de phase : quand la variation minime d'un paramètre change complètement les propriétés du système. Pour prendre un exemple physique bien connu, l'eau liquide devient subitement solide au-dessous de 0 °C. C'est encore une transition de phase qui se produit au-delà de 100 °C, lorsque l'eau liquide devient vapeur. La transition de phase se définit ainsi comme un changement brusque d'une observable suite à la modification minime d'un paramètre.

Rappelez-vous que $G(n,m)$ est un graphe aléatoire à $n$ sommets où $m$ arêtes ont été ajoutées au hasard. Il faut imaginer le nombre de sommets $n$ très grand mais fixé, alors que $m$ est un paramètre variable, analogue de la température (ou plutôt l'inverse de la température), qui peut prendre toutes les valeurs entre 0 et $n(n-1)/2$. La question qui se pose alors est de savoir quand apparaissent certaines structures dans le graphe aléatoire à mesure que $m$ croît. Par structure, on entend l'existence à l'intérieur de $G(n,m)$ de certains sous-graphes comme des triangles (trois sommets reliés les uns aux autres par trois arêtes), un chemin hamiltonien (*) ou encore une composante connexe ou amas (en anglais *cluster* – mot rendu hélas populaire par la pandémie) de grosse taille. Bien entendu, la présence d'une structure

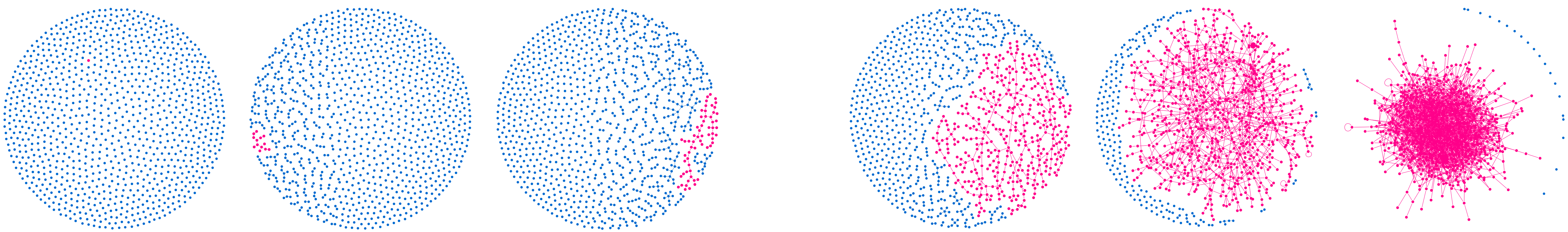
## Il suffit de quelques arêtes supplémentaires pour qu'une composante géante apparaisse avec une très grande probablité

donnée dans $G(n,m)$ est une variable aléatoire, c'est-à-dire que, pour certaines réalisations du graphe, la réponse sera oui et, pour d'autres, la réponse sera non. Mais lorsque les nombres $n$ et $m$ sont grands, la réponse devient « quasi déterministe », c'est-à-dire toujours la même avec une grande probabilité.

Expliquons cela en prenant l'exemple de l'existence d'une composante de grosse taille, appelée « percolation ». Dès 1959, Erdös et Rényi montrent l'existence d'une transition abrupte lorsque $m$ atteint $n/2$ par l'apparition d'une composante géante (1). Autrement dit, si le nombre d'arêtes est juste en dessous de ce seuil (par exemple 0,49 $n$), on n'aura pas de composante géante avec une très grande probabilité. Mais il suffit d'ajouter quelques arêtes et d'arriver, par exemple, à 0,51 $n$, pour qu'une composante géante apparaisse avec une très grande probabilité. La propriété « avoir une grosse composante », bien que *stricto sensu* aléatoire, devient déterministe quand $n$ est grand et, de plus, la réponse passe brusquement de « non »

(*) **Andrew Wiles** a travaillé seul pendant près de dix ans avant de produire en 1994 une preuve très longue et complexe, s'appuyant sur les avancées de ses prédécesseurs, de la conjecture proposée par Pierre de Fermat plus de quatre siècles auparavant. Pour rappel, cette conjecture, maintenant théorème, dit qu'il n'existe pas de solution à l'équation $x^n + y^n = z^n$ pour $n > 2$ et $x,y,z$ des entiers strictement positifs.

(*) **Un chemin hamiltonien** est un cycle d'arêtes qui passent par chaque sommet d'un graphe.



LA PERCOLATION EN IMAGE

*Simulation du modèle de graphe aléatoire d'Erdös-Rényi avec 1000 sommets. Le nombre d'arêtes reliant les sommets entre eux augmente jusqu'au point critique (en gros, 500 arêtes) où apparaît brusquement une composante géante, appelée percolation.*

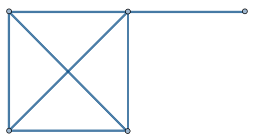
**Dans le cas du Loto, la probabilité de gagner est très faible, alors que l'espérance du gain est d'environ 1 euro pour 2 euros de mise**

quand $m \leq 0{,}49n$ à « oui » quand $m \geq 0{,}51n$. Et ceci alors même que le nombre d'arêtes n'a varié que de $\frac{0{,}51-0{,}49}{0{,}49} \approx 4\%$ !!!

C'est donc bien une transition de phase, au sens où une infime variation d'un paramètre – ici le nombre d'arêtes dans le graphe – a des conséquences globales impressionnantes.

L'exemple de la percolation présenté ci-dessus est en fait général : en 1987, le Hongrois Béla Bollobás et le Britannique Andrew Thomason montrent que, quelle que soit la propriété recherchée $\mathcal{P}$ (apparition d'un triangle, présence d'un chemin hamiltonien ou de tout autre sous-graphe), il y a toujours une transition de phase à partir d'un certain niveau de $m$ – le nombre critique d'arêtes –, mais possiblement moins abrupte que dans le cas de la percolation (2).

Nous laissons le soin au lecteur d'imaginer le cas d'une transition de phase où le paramètre variable serait la température terrestre et la propriété cible serait l'habitabilité de notre planète… Si une transition de phase abrupte a lieu à la température de 15 °C (grosso modo la température moyenne de la Terre ces dernières années), alors une infime variation de celle-ci pourrait avoir des effets catastrophiques…



LE « GRAPHE DE LA CASSEROLE »

***On montre que dans un graphe aléatoire, dès qu'apparaît le « cœur » de la casserole (le carré), on obtient beaucoup de casseroles ayant le même cœur, mais des manches différents.***

Dans notre modélisation mathématique, étant donné notre propriété $\mathcal{P}$, peut-on calculer le nombre « critique » d'arêtes $m = m_{\mathcal{P}}(n)$ autour duquel notre propriété devient très sensible ? Dans le travail dont nous venons de parler, Bollobás et Thomason montrent que l'on peut prendre $m_{\mathcal{P}}(n)$, tel que la probabilité d'apparition de la propriété $\mathcal{P}$ en mettant exactement $m = m_{\mathcal{P}}(n)$ soit égale à 1/2 (peu ou prou).

À ce stade, on pourrait se dire que c'est gagné, puisqu'ainsi il paraît théoriquement possible de trouver $m_{\mathcal{P}}(n)$ via cette équation qui fait intervenir la probabilité. Hélas, en pratique, ces probabilités sont très difficiles à calculer… En effet, l'ensemble des graphes possédant la propriété $\mathcal{P}$ s'écrit généralement comme l'union de nombreux ensembles ($A_1 \cup A_2 \cup \ldots \cup A_k$). Prenons le cas où la propriété $\mathcal{P}$ est celle d'avoir un triangle, et imaginons que les sommets de $G(n,m)$ sont numérotés $1,2,\ldots,n$. Alors $G(n,m)$ a un triangle si et seulement si les sommets 1,2,3 forment un triangle dans $G(n,m)$, ou bien les sommets 1,2,4 forment un triangle… et ainsi de suite pour tous les triplets $i,j,k$ de sommets. Même dans le cas d'une union de trois ensembles, calculer la probabilité de $A_1 \cup A_2 \cup A_3$ n'est pas si simple, car il faut calculer la probabilité de $A_1$, ajouter celle de $A_2$, celle de $A_3$, enlever celle de $A_1 \cap A_2$ (car on l'a comptée deux fois), idem pour $A_1 \cap A_3$ et $A_2 \cap A_3$, mais aussi ajouter celle de $A_1 \cap A_2 \cap A_3$ que l'on a trop soustraite. La formule générale, dite du crible, est compliquée, et son application est quelquefois humainement impossible.

En revanche, il est facile de voir que la probabilité de $A_1 \cup A_2 \ldots \cup A_k$ est toujours plus petite que la somme des probabilités $\mathbf{P}(A_1) + \ldots + \mathbf{P}(A_k)$. Cette inégalité permet de donner une borne inférieure pour $m_{\mathcal{P}}(n)$ : on cherche $M_{\mathcal{P}}(n)$ pour que la somme dans le membre de droite soit environ 1/2, et on aura alors $m_{\mathcal{P}}(n) \geq M_{\mathcal{P}}(n)$.

La borne ci-dessus est parfois appelée la « méthode du premier moment » ou « méthode de l'espérance ». Réexpliquons-la dans le cas de l'apparition d'un triangle dans $G(n,m)$ : au lieu de trouver $m = m_{\mathcal{P}}(n)$ tel que la probabilité que $G(n,m)$ possède un triangle soit égale à 1/2, on trouve $m = m_{\mathcal{P}}(n)$ tel que la moyenne (en mathématiques on parle d'espérance) du nombre de triangles dans $G(n,m)$ soit égale à 1/2.

Dans le cas de l'apparition des triangles, la méthode de l'espérance donne le bon résultat et $m_{\mathcal{P}}(n)$ est du même ordre de grandeur que $M_{\mathcal{P}}(n)$. Malheureusement, l'espérance est parfois trompeuse. Le cas du Loto est à cet égard frappant : la probabilité de gagner est extrêmement faible alors que l'espérance du gain est autour de 1 euro… pour une mise de 2 euros (la Française des jeux reste une entreprise à but lucratif…). Quand on gagne, on gagne donc beaucoup, mais cela arrive très rarement…

Dans notre cas, il se pourrait donc que $M_{\mathcal{P}}(n)$ (calculé avec l'espérance) soit un piètre indicateur de $m_{\mathcal{P}}(n)$ (calculé avec les probabilités), c'est-à-dire qu'il soit beaucoup plus faible. C'est d'ailleurs ce qui se passe avec l'apparition de certaines structures dans le graphe aléatoire, notamment la « casserole ». Ce graphe est constitué de quatre sommets entièrement reliés (le cœur de la casserole), connectés à un cinquième sommet (le manche). Dans un graphe aléatoire, on peut démontrer qu'au cours de sa construction, dès que le cœur de la casserole apparaît, on obtient en fait beaucoup de casseroles avec le même cœur mais des manches différents (c'est comme gagner au loto, quand on gagne, on gagne beaucoup !). Et, du coup, le $M_{\mathcal{P}}(n)$ pour la propriété « apparition de la casserole » est très petit par rapport au véritable $m_{\mathcal{P}}(n)$.

Pourtant, en 2006, Jeff Kahn et Gil Kalai ont proposé une conjecture audacieuse permettant de calculer le seuil de manière assez simple (3). Selon leur hypothèse, le seul écueil est le cas identifié dans la casserole : il faut appliquer la méthode de l'espérance à toutes les « sous-structures » $\mathcal{P}'$ (tous les sous-graphes) de $\mathcal{P}$ et prendre le maximum des $m_{\mathcal{P}}(n)$ obtenus. Dans le cas de la casserole, ce maximum est donné par l'apparition du cœur de la casserole. Selon leur conjecture, le seuil ainsi calculé est proche de $m_{\mathcal{P}}(n)$ (à un facteur logarithmique près).

## TROP NAÏF, TROP BEAU POUR ÊTRE VRAI

Voici une autre manière de présenter la conjecture à l'aide de « recouvrements », dans l'esprit de la preuve qui vient d'être publiée par Jinyoung Park et Huy Tuan Pham. Il faut d'abord se rappeler qu'en général, l'ensemble des graphes qui satisfont la propriété $\mathcal{P}$ s'écrit naturellement comme une union d'ensembles $A_1 \cup A_2 \cup \ldots \cup A_k$ et que la méthode de l'espérance est fondée sur l'inégalité entre la probabilité de l'union et la somme des probabilités, qui, en général, est assez grossière. La conjecture de Kahn et Kalai dit qu'on peut en fait faire un autre recouvrement par des ensembles croissants de graphes $B_1 \cup \ldots \cup B_l$, pour lesquels cette inégalité n'est pas si mauvaise. Ce nouveau recouvrement est dit « parcimonieux », au sens où les ensembles qui le composent ne s'intersectent pas trop.

Lorsque cette conjecture (dite du « seuil en espérance », *expectation threshold* en anglais) a été rendue publique, peu de gens y croyaient : cela paraissait trop naïf, trop beau pour être vrai. Ne pourrait-il pas exister des cas beaucoup plus compliqués où le simple calcul de l'espérance ne donnerait pas du tout le bon seuil ? Les auteurs de la conjecture eux-mêmes la qualifiaient d'« osée » et pensaient qu'elle pourrait être fausse. Mais la recherche d'un contre-exemple, en vain, a conforté leur idée.

La conjecture a résisté… Mais les premières fissures apparurent en 2019, lorsqu'une version affaiblie de celle-ci fut prouvée. Puis, le 31 mars dernier fut publié sur le site Arxiv un article de Jinyoung Park et Huy Tuan Pham prouvant en toute généralité la conjecture de Kahn et Kalai (4). Le papier est très court (six pages), la preuve à proprement parler n'en occupe que deux et n'utilise aucune sorte de mathématiques difficiles – uniquement de la combinatoire élémentaire. C'est remarquable. Cela a permis à beaucoup de personnes de vérifier sa justesse – un étudiant en licence de mathématiques serait tout à fait capable de suivre la démonstration ligne à ligne. Reste qu'elle est très astucieuse et fait elle-même appel à l'aléatoire pour produire des recouvrements parcimonieux $B_1 \cup \ldots \cup B_l$ comme décrits ci-dessus.

Le parcours des deux auteurs de la preuve mérite d'être évoqué. Huy Tuan Pham, un brillant étudiant vietnamien venu aux États-Unis pour suivre des études de mathématiques, est actuellement en thèse à l'université Stanford sous la direction de Jacob Fox. Quant à Jinyoung Park, elle raconte dans une vidéo de l'Institute for Advanced Study, à Princeton (États-Unis), qu'elle n'était pas spécialement douée en maths au collège et qu'elle a décidé de changer sa façon d'étudier en lisant des livres (5). Après une licence de professorat de maths, elle enseigne cette matière durant sept ans dans un collège à Séoul, avant de suivre son époux qui venait d'obtenir un travail aux États-Unis. C'est là qu'elle décide de reprendre des études de maths et soutient une thèse sous la direction de Jeff Kahn (l'un des auteurs de la conjecture de 2006) à l'université Rutgers en 2020. Elle est actuellement en poste à l'université Stanford. Elle raconte qu'en thèse, elle se trouvait trop lente par rapport à d'autres, ce à quoi Jeff Kahn aurait répondu : *« C'est bien d'être rapide, mais c'est plus important d'être profond »*. Et élégant, ajouterai-je ! ■

(1) P. Erdös et A. Rényi, *Zastos. Mat., 10*, 47, 1969.
(2) B. Bollobas et A. G. Thomason, *Combinatoria, 77*, 35, 1987.
(3) J. Kahn et G. Kalai, doi : 10.48550/arXiv.math/0603218, 2006.
(4) J. Park et H. T. Pham, doi : 10.48550/arXiv.math/.2203.17207, 2022.
(5) www.youtube.com/watch?v=7H2YvxefFro