La gazette du gdr iM

15 juillet 2020
Numéro 1

www.gdr-im.fr

Éditorial

Nous sommes heureux de vous présenter le premier numéro de la gazette du GDR informatique mathématique (GDR 673 du CNRS). Elle s'adresse à toute la communauté informatique mathématique et au-delà. Nous souhaitons à travers elle mettre en avant ce qui fait le cœur de notre GDR, c'est-à-dire ses membres et leurs recherches. En exposant au fil des numéros de beaux résultats, des idées, en présentant des domaines de recherches mais aussi des parcours individuels, nous espérons montrer la richesse et la vitalité de l'informatique mathématique et de sa communauté, dans toute sa diversité.

Nous remercions chaleureusement les contributeurs de ce premier numéro (Amina Doumane, Claire Mathieu, Étienne Moutot, Alexandre Vigny, ainsi que Valérie Berthé et Nicolas Schabanel qui ont conduit les interviews), ainsi que notre cellule communication (Valérie Berthé, Julien Clément, Ines Klimann, Youssouf Oualhad et Pascal Vanier) qui nous accompagne sur ce projet depuis ses prémisses. Ines Klimann s'est tout particulièrement investie dans la maquette de ce premier numéro, qui n'aurait pas vu le jour sans elle.

Cette gazette s'adresse à vous, mais elle se fera aussi grâce à vous : n'hésitez pas à envoyer vos commentaires, vos suggestions et vos idées de contribution.

Nous vous souhaitons une excellente lecture et nous vous donnons rendez-vous au prochain numéro.

J.-M. Muller et G. Theyssier

Une place pour chaque donnée et chaque donnée à sa place, par Alexandre Vigny

Vous sortez de chez vous pour aller au cinéma. Vous arrivez à la station de métro que vous pensiez utiliser, et là le drame ! À cause d'une inondation, la station est fermée. Vous vous empressez de sortir votre téléphone à la recherche d'un trajet alternatif. Quelques clics, deux bus et un tram plus tard, vous arrivez au cinéma. Vous êtes à l'heure pour votre séance, grâce à une base de données !

Coloriages et complexité: la conjecture de Nivat, par Étienne Moutot

À quel point peut-on créer des structures complexes à l'aide de briques élémentaires ? Imaginons que l'on cherche à colorier une grille infinie avec seulement quelques couleurs à disposition. Sans restrictions, on peut facilement y dessiner des coloriages complexes et aussi grands qu'on le souhaite. Cependant, si on impose que le coloriage ne peut contenir que certains motifs, cela limite les possibilités.

Amina Doumane (CNRS, ENS Lyon), logique et liberté

Cet entretien a été conduit par Valérie Berthé le 15 janvier 2020.

Bonjour Amina, peux-tu nous dire quelques mots concernant ta situation actuelle et ta recherche ?

Je suis actuellement chargée de recherche au CNRS au LIP (Laboratoire de l'informatique du parallélisme) depuis l'automne 2019. Mes thèmes de recherche relèvent de la théorie de la démonstration et de l'étude des preuves circulaires pour la vérification, avec une évolution récente vers les algèbres de relations, la logique pour les graphes et la théorie des automates.

Peux-tu nous décrire ton parcours ?

J'ai fait mes études au Maroc, j'ai été en classes préparatoires au lycée Ibn Abdoun à Khouribga et à Ibn Youssef, à Rabat, puis j'ai intégré l'École Centrale Paris. Suite à un cours sur les machines de Turing, j'ai découvert la logique et l'informatique en me disant que wouah!!, c'était ce que j'avais vraiment envie de faire. J'ai commencé à suivre en parallèle de mes études à l'école Centrale des cours de master. J'ai d'abord suivi les cours du Master en logique mathématique LMFI (Logique mathématique et fondements de l'informatique), puis du Master en informatique MPRI (Master Parisien de Recherche en Informatique à l'Université Paris Diderot).
C'est la rencontre avec Alexis Saurin qui m'a décidée à faire une thèse. J'ai effectué celle-ci à Paris Diderot sous sa direction, ainsi que celle de David Baelde et de Pierre-Louis Curien. J'ai ensuite effectué deux postdocs, à l'ENS Lyon au LIP sous la direction de Damien Pous, puis à Varsovie, avec Mikołaj Bojanczyk.

Comment tes thèmes de recherche ont-ils évolué ?

Ma thèse portait sur les preuves circulaires. Ces preuves permettent d'utiliser le théorème que l'on a envie de prouver comme si c'était un axiome, dans le but de prouver... ce même théorème. On pourrait croire que tout cela se mord la queue, mais utilisées correctement, les preuves circulaires se révèlent très utiles. Suite à mes deux postdocs, mes thèmes de recherche ont évolué vers les algèbres de relations et les langages de graphes. Un thème qui m'attire tout particulièrement en ce moment, c'est la formalisation des graphes en Coq. Il y a un vrai besoin de formalisation, exprimé par les théoriciens des graphes.
J'en ai discuté récemment avec Nicolas Trotignon, et celui-ci voit la longueur des démonstrations dans les papiers de théorie des graphes comme un problème : difficulté de vérifier les preuves, voire risque d'erreurs. Au LIP, Damien Pous a fait des avancées récentes dans ce domaine, et j'aimerais le rejoindre dans son effort.

Penses-tu que l'utilisation des assistants de preuve va se se généraliser de telle sorte que chaque mathématicien.ne puisse s'en emparer ?

Ce serait formidable, mais cela me paraît peu réaliste. Ce qui serait plus réaliste, c'est qu'il y ait plus d'équipes qui fassent des preuves formelles et puissent aider à développer des outils à destination de tous.

Tu as fait tes études en classes préparatoires aux grandes écoles au Maroc. Est-ce que tu vois des différences significatives entre les deux systèmes ?

Les deux systèmes sont très similaires, je n'ai pas ressenti de dépaysement de ce point de vue en arrivant en France. Si je devais avoir un regret, c'est que je ne connaissais pas l'existence des ENS. Rétrospectivement, je n'aurais certainement pas pu les intégrer de toutes façons, car lorsqu'on intègre une école d'ingénieurs, on reçoit une bourse d'études du gouvernement français, mais pas pour les ENS. Je trouve que c'est très dommage.

Tu as été membre de l'équipe marocaine aux Olympiades Internationales de Mathématiques. Comment as-tu vécu cette expérience ?

J'ai toujours aimé participer aux Olympiades. Les Olympiades m'ont stimulée depuis le collège. C'était le moment de l'année scolaire que je préférais le plus. On n'a certes pas eu beaucoup d'entraînements mais on avait accès aux documents préparés par Animaths.

Tu as reçu le Prix de la revue La Recherche en 2017 pour la mention Sciences de l'information. Peux-tu nous en dire plus ?

Ce prix annuel récompense un travail dans chaque discipline scientifique. J'ai été très contente de voir que mon travail, qui est de nature fondamentale, ait plu au jury. À la suite de la remise des prix, j'ai été interviewée par les journalistes de la Recherche afin de décrire mes résultats. Faire l'effort d'expliquer mes travaux à des non experts m'a permis d'y réfléchir différemment et m'a été très utile.

L'activité de diffusion scientifique est-elle importante pour toi ?

Je n'ai pas pu faire encore d'activités de médiation. Mais j'aimerais m'impliquer plus dans des activités du genre « Pint of Science », en faisant des conférences grand public dans des lieux de vie comme des bars ou des cafés. J'ai déjà assisté à des exposés de ce type sur des disciplines qui m'intéressent et dont je ne suis pas spécialiste, comme les neurosciences et la biologie, et j'ai appris beaucoup de choses. Par ailleurs je fais de l'aide aux devoirs dans une MJC. Je me rends compte du fait que le système scolaire, sans aide extérieure, ne suffit pas, certains enfants abandonnent. Pouvoir observer l'évolution des enfants à travers l'aide aux devoirs est très utile.

Qu'aimes-tu particulièrement dans ton métier ?

La liberté. J'ai le temps de lire, d'apprendre, je peux prendre des risques dans mes choix de recherche. Je suis heureuse de ne plus avoir de problèmes d'incertitudes quant à mon avenir et ne plus avoir la pression de la sélectivité que j'ai ressentie pendant mes postdoctorats. Je découvre une autre manière de travailler.

Et qu'est-ce qui te pèse le plus ?

Certainement le système de publications en conférences privilégié dans ma communauté qui, par sa forme et sa fréquence, incite à publier vite, privilégie la forme au fond et est très largement anti-écologique. Bien sûr, nous sommes beaucoup de chercheurs/ses à être conscients du danger de cette pratique, mais nos carrières scientifiques, nos financement et notre reconnaissance restent encore très liées au publications dans certaines conférences prestigieuses. À titre personnel, je n'ai pas arrêté de participer aux conférences, mais j'essaye de publier moins et de prêter plus attention à la rédaction de mes papiers.

As-tu rencontré des difficultés concernant le fait d'être une femme dans ton parcours ?

Je n'en ai pas rencontré, pas de manière évidente en tout cas, mais je sais qu'il y a plusieurs femmes qui en souffrent. C'est pour cela qu'il est important d'éveiller les consciences. Dans ce sens, le travail fait au LIP par Natacha Portier et Nathalie Revol est particulièrement remarquable, j'aimerais être plus impliquée, comme elles. Par contre, il faut rester vigilant quant à la pression exercée sur les femmes concernant ce travail de sensibilisation et d'égalité, qui à mon sens, doit être porté aussi bien par les hommes que par les femmes.

Comment te situes-tu par rapport à l'interface mathématiques/informatique ?

Je ne fais aucune distinction entre mathématiques et informatique dans ma démarche de recherche. En tant qu'informaticienne théorique, les objets que j'étudie sont des algorithmes, je comprends leurs propriétés et je prouve des théorème les concernant, avec la même rigueur que n'importe quel autre mathématicien. C'est exactement la même démarche que fait un théoricien des nombres ou un géomètre, seul l'objet d'étude diffère.

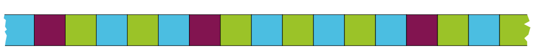
Coloriages et complexité: la conjecture de Nivat

Étienne Moutot (LIP, ENS Lyon & Univ de Turku)

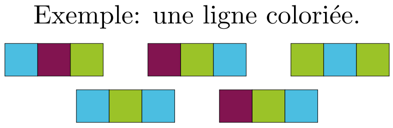
À quel point peut-on créer des structures complexes à l'aide de briques élémentaires ? Imaginons que l'on cherche à colorier une grille infinie avec seulement quelques couleurs à disposition. Sans restrictions, on peut facilement y dessiner des coloriages complexes et aussi grands qu'on le souhaite. Cependant, si on impose que le coloriage ne peut contenir que certains motifs, cela limite les possibilités. Intuitivement, si on ne dispose que d'un petit nombre de motifs, on ne pourra pas faire beaucoup plus de choses que des répétitions de ces quelques motifs. À l'inverse, si on peut utiliser de nombreux motifs, il semble possible de les assembler de manière complexe et non répétitive. Mais y a-t-il une limite précise ? Peut-on quantifier exactement le nombre de motifs suffisants pour créer des coloriages complexes ? C'est exactement à cette question que tente de répondre la conjecture de Nivat, sur laquelle nous allons nous pencher maintenant.

Colorions une ligne : le théorème de Morse-Hedlund

Avant de regarder le cas de la grille infinie, attardons-nous d'abord à un cas plus simple : une seule ligne infinie. À partir d'une ligne infinie déjà coloriée, on peut compter le nombre de motifs d'une certaine taille qui y apparaissent.



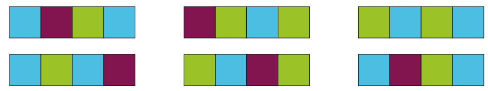
Exemple: une ligne coloriée.



Il y a 5 motifs de taille 3.

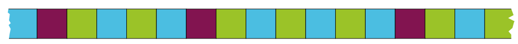
Ce nombre de motifs s'appelle la complexité de la ligne, et dépend de la taille que l'on considère. Il n'y a en général pas autant de motifs de taille *3* que de motifs de taille *4* par exemple. Ainsi, on dit que la complexité est une fonction dépendant de la taille des motifs. On note *p(n)* le nombre de motifs de taille *n* qui apparaissent dans un coloriage de la ligne.

Pour la ligne précédente, on avait donc *p(3) = 5*. Mais aussi *p(4) = 6* avec les motifs suivants.



Cette complexité permet d'exprimer de manière quantitative à quel point la ligne est constituée de motifs variés: plus la valeur prise par *p(n)* est grande, plus il existe de motifs différents qui apparaissent, et donc plus le coloriage général de la ligne est complexe.

Pour cette ligne très simple, $p(n) = 2$ pour tout n.



Pour celle ci, on a vu que $p(3) = 5 > 2$, ce qui exprime la plus grande complexité du coloriage.

Quelles sont les manières les plus élémentaires de colorier une ligne infinie ? La première idée est de colorier la ligne en utilisant une seule couleur, ce qui est effectivement très simple. Mais de manière générale, tout coloriage constitué d'un même motif répété à l'infini semble une manière simple de colorier une ligne infinie.

Exemple : une ligne périodique.

En observant la complexité d'une telle ligne périodique, un phénomène intéressant apparaît : au bout d'un moment, même si la taille des motifs *n* grandit, la complexité *p(n)* garde elle la même valeur (la fonction est bornée).

Motifs de taille 1 : $p(1) = 2$.

Motifs de taille 2 : $p(2) = 3$.

Motifs de taille 3 : $p(3) = 3$. Ensuite pour tout $n \geq 3, p(n) = 3$.

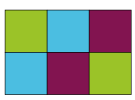
On voit donc que la fonction de complexité d'une ligne périodique est bornée. En 1938, Morse et Hedlund ont prouvé que l'implication réciproque était vraie aussi : si on a une ligne dont la fonction de complexité est bornée, alors elle ne peut être que périodique. Autrement dit, cela signifie que si on nous donne un nombre de motifs petit par rapport à la taille de ces motifs, alors il n'est possible de construire que des lignes périodiques avec ces motifs. Mais le Théorème de Morse-Hedlund [MH38] va plus loin, et donne une valeur précise au terme de "petit". En effet, il suffit qu'il existe une taille de motif n telle que $p(n) \leq n$ pour que la ligne soit nécessairement périodique. Autrement dit, dès que l'on trouve une seule taille n pour laquelle il existe moins de n motifs de taille n, alors le coloriage de la ligne est forcément périodique !

Théorème [Morse, Hedlund, 1938] Pour un coloriage d'une ligne infinie donné, s'il existe une taille de motif n telle que $p(n) \leq n$, alors le coloriage est périodique.

Le théorème de Morse-Hedlund donne ainsi un lien précis entre la périodicité d'une ligne et le nombre de motifs différents qui y apparaissent.

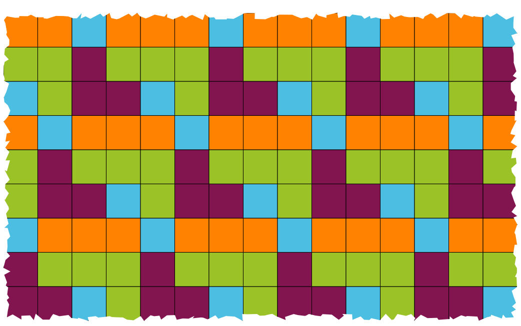
Le quadrillage : la conjecture de Nivat

Revenons maintenant à notre grille de départ en dimension *2*, et raisonnons de manière analogue au cas de la ligne (dimension *1*). Au lieu d'être un simple enchaînement de n couleurs, un motif est maintenant un rectangle, d'une certaine taille $m \times n$.



Exemple: un motif de taille 3×2.

En dimension *2*, on peut également définir la fonction de complexité $p(m,n)$, qui correspond maintenant à tous les coloriages de rectangles de taille $m \times n$ qui apparaissent dans la grille. De même, la périodicité d'un coloriage désigne le caractère répétitif de celui-ci : un coloriage est dit périodique s'il est égal à une translation de lui même. Dit autrement, en le déplaçant dans une direction, on retombe sur la même chose.



Un coloriage périodique par translation de vecteurs $(1,3)$ et $(-3,-3)$.

On a ici aussi une correspondance entre le fait que la fonction de complexité est bornée et le fait que la grille est périodique. La conjecture de Nivat [Niv97] propose une borne, et dit qu'elle devrait être similaire à la borne en dimension 1 :

Conjecture [Nivat, 1997] Pour un coloriage d'une grille infinie donné, s'il existe une taille de motif $m \times n$ telle que $p(m,n) \leq mn$, alors le coloriage est périodique.

S'il existe une taille de motif $m \times n$ telle que $p(m,n) \leq mn$, on dit que le coloriage est de faible complexité.

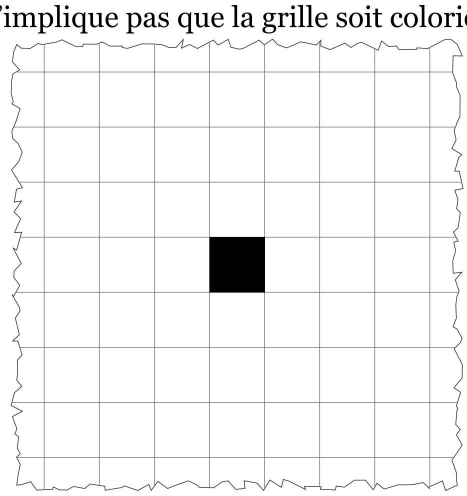
Bien que la majorité des chercheurs et chercheuses du domaine croient en la véracité de cette conjecture, et que personne n'a réussi à trouver de contre-exemple, personne n'a encore non plus réussi à trouver une preuve de cette conjecture.

La similarité avec le théorème de Morse-Hedlund ne la rend pas pour autant facile à démontrer. Le passage à la dimension *2* change plus de choses qu'il n'y paraît au premier abord. Une manière simple pour s'en convaincre est de faire un saut de dimension supplémentaire, et de regarder le cas de la dimension *3*. Cette fois, il a été prouvé que l'analogue de la conjecture de Nivat est simplement fausse : il existe une manière de colorier une grille tri-dimensionnelle qui ne soit pas périodique, mais où il existe une taille de motif n telle que $p(n,n,n) \leq n^3$. Cela montre donc que la dimension joue ici un rôle très important. Un énoncé peut passer de facilement prouvable (dimension *1* : le théorème de Morse-Hedlund), à faux (dimension *3*) en passant par extrêmement difficile mais probablement vrai (dimension *2* : la conjecture de Nivat).

Mais pourquoi s'intéresse-t-on particulièrement à cette conjecture ? Les raisons sont plus fondamentales que pratiques. Cette borne reliant la complexité d'une grille à sa périodicité quantifierait exactement la limite évoquée plus haut : combien de motifs on doit utiliser pour pouvoir former des structures non répétitives (non périodiques). Ce lien est d'autant plus intéressant qu'il relie une notion locale, la complexité, à une propriété globale de la grille, sa périodicité.

De manière très informelle, imaginons que la grille

représente l'univers et qu'on cherche à comprendre quelles possibilités offre cet univers. Un univers possible correspondrait à un coloriage de la grille, et les motifs autorisés les parties finies que l'on peut observer. La conjecture de Nivat exprime alors le fait que s'il n'y a qu'un petit nombre de possibilités locales, cet univers ne peut être que périodique, et donc très simple à décrire. Cette analogie est d'autant plus intéressante que l'on a en général accès qu'à des portions finies d'un univers, d'où l'intérêt de pouvoir déduire des choses sur l'univers en entier en n'en observant que de petites parties.
Depuis qu'elle a été énoncée en 1997, il y a eu de nombreuses tentatives de prouver cette conjecture, menant à des résultats de plus en plus proches sans encore l'atteindre. Une des manières de s'en approcher est de supposer que le coloriage a une complexité encore plus faible que mn. Par exemple, en 2003 Epifanio, Koskas et Mignosi prouvent que s'il existe m et n tels que $p(m,n) \leq mn/144$, alors le coloriage doit être périodique [EKM03]. Cette condition a ensuite été améliorée en $p(m,n) \leq mn/16$ par Quas et Zamboni en 2004 [QZ04], et enfin en $p(m,n) \leq mn/2$ par Cyr et Kra en 2015 [CK15] ; ce qui est aujourd'hui un des résultats s'approchant le plus de la conjecture de Nivat.
On sait également que la conjecture de Nivat est "optimale", dans le sens où c'est la plus grande borne plausible. En effet, on sait que $p(m,n) \leq mn+1$ n'implique pas que la grille soit coloriée



Exemple de coloriage de complexité $nm + 1$: il y a en effet nm motifs de taille $m \times n$ contenant la cellule noire (la cellule peut se trouver en n'importe quelle position du motif), plus un entièrement blanc (ne contenant pas la cellule noire). De plus, ce coloriage n'est pas périodique (la cellule noire n'est répétée nulle part).

périodiquement, puisqu'il existe un coloriage non périodique pour lequel il existe m et n tels que $p(m,n) = mn+1$.

Les coloriages sont des polynômes

En 2015, Kari et Szabados ont initié une nouvelle approche pour s'attaquer à la conjecture de Nivat. L'originalité de leur idée repose sur la reformulation du problème dans un contexte différent : celui de l'algèbre commutative. Cela permet d'utiliser des théorèmes d'algèbre, à priori sans lien direct avec la conjecture de Nivat, pour obtenir des résultats très proches de la conjecture (sans pour autant réussir à la prouver).
La première étape est de désigner chaque couleur par un nombre, puisque c'est plus facile de faire des raisonnements avec des nombres que des couleurs. En utilisant ces couleurs, on représente les motifs comme des polynôme à deux variables. Un polynôme à deux variables est exactement comme un polynôme habituel, sauf que deux variables, x et y, sont utilisées, par exemple $x^2y^4 + 2.5x + 4y^2 + 5.21$. On appelle monôme un polynôme de la forme Ax^ay^b, par exemple $25.4x^4y^2$. Pour représenter une case coloriée, on multiplie le nombre (ou la couleur) présent sur la case (a,b) par le monôme x^ay^b. Ainsi, l'exposant de x représente l'abscisse d'une cellule et l'exposant de y son ordonnée. Par exemple si la couleur 2 est présente sur la case $(3,1)$, elle sera représentée par le monôme $2x^3y^1$. En additionnant plusieurs monômes, on peut ensuite représenter des motifs entiers.

y	xy	x^2y
1	x	x^2

Les positions des différents monômes sur un motif 3×2.

1	3	2
3	2	1

Le motif représenté par $3 + 2x + x^2 + y + 3xy + 2x^2y$.

Comme un coloriage complet n'est finalement qu'un motif infini, on peut le représenter par un polynôme avec un nombre infini de termes, avec potentiellement des exposants négatifs (cela s'appelle une série de Laurent). En toute généralité, un coloriage pourra donc être représenté par un "polynôme infini" *c(x,y)*:

$$c(x,y) = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} c_{i,j} x^i y^j,$$

avec $c_{i,j}$ la couleur en position *(i,j)*. Un polynôme infini peut sembler un peu étrange au premier abord, mais peut cependant être défini rigoureusement. Nous ne nous y attarderons pas puisque l'intuition suffira pour comprendre ce qui suit.

On peut ensuite s'intéresser aux opérations sur ces "polynômes infinis", et leurs significations pour les coloriages représentés.

L'addition de deux polynômes correspond à superposer deux coloriages et additionner les nombres qui se recouvrent, puisque

$$\begin{aligned}&\sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} c_{i,j} x^i y^j + \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} d_{i,j} x^i y^j \\ &= \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} (c_{i,j} + d_{i,j}) x^i y^j.\end{aligned}$$

Notons qu'à cause de cette somme, le résultat de cette opération peut produire des nouvelles couleurs, sous la forme de nouveaux nombres, qui n'étaient pas présentes dans les deux configurations initiales.

De la même manière, on peut soustraire deux polynômes, ou deux coloriages.

La multiplication de deux polynômes infinis n'a pas forcément de résultat en général. Cependant, la multiplication **par un monôme** est bien définie.

$$\begin{aligned}&x^a y^b \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} c_{i,j} x^i y^j \\ &= \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} c_{i,j} x^{i+a} y^{j+b} \\ &= \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} c_{i-a,j-b} x^i y^j\end{aligned}$$

Pour comprendre le résultat de cette opération, regardons la couleur en position *i,j* de ce produit. Par définition, c'est le coefficient devant $x^i y^j$, qui est $c_{i-a,\,j-b}$. Le coloriage possède donc les mêmes couleurs que *c*, mais ses indices sont décalés. La multiplication par $x^a y^b$ a donc fait une translation de notre coloriage par le vecteur *(a, b)*.

En utilisant ces deux opérations, on peut reformuler algébriquement une notion clé de la conjecture de Nivat : la périodicité. Un coloriage est périodique si et seulement s'il existe $a,b \in \mathbb{Z}$ non nuls tels que le coloriage est égal à lui-même translaté par *(a,b)* (ou *(−a,−b)*). Reformulé en terme de polynômes, cela donne :

$$\begin{aligned} c \text{ périodique} &\Leftrightarrow \text{il existe } a,b, \quad x^a y^b c = c \\ &\Leftrightarrow \text{il existe } a,b, \quad x^a y^b c - c = 0 \\ &\Leftrightarrow \text{il existe } a,b, \quad (x^a y^b - 1)c = 0. \end{aligned}$$

Ainsi, pour prouver qu'un coloriage est périodique, il suffit de trouver un polynôme de la forme $x^a y^b - 1$ tel que son produit avec le "polynôme infini" du coloriage donne le coloriage nul (on dit que $(x^a y^b - 1)$ annule *c*). Appelons Ann*(c)* l'ensemble des polynômes qui annulent *c*.

On peut donc maintenant se concentrer sur l'étude de cet ensemble puisque que pour prouver la conjecture de Nivat, il "suffirait" de prouver que si *c* est de faible complexité, alors Ann*(c)* contient un polynôme de la forme $x^a y^b - 1$. Au premier abord, cela peut semble plus complexe d'étudier cet ensemble, Ann*(c)*, que de prouver que *c* est périodique. Cependant, Ann*(c)* n'est pas un ensemble quelconque, c'est ce qu'on appelle un idéal polynomial. Cela signifie qu'il possède une structure algébrique forte, que l'on peut exploiter pour obtenir des résultats très intéressants.

C'est en étudiant cette structure que Kari et Szabados ont obtenu un certain nombre de théorèmes très proches de la conjecture de Nivat. La première étape de leur raisonnement est de montrer que si l'on dispose d'un coloriage de faible complexité, alors Ann*(c)* contient nécessairement un polynôme annulateur. De plus, ils prouvent que l'on peut trouver un polynôme annulateur ayant une certaine forme, assez proche de ce qu'on souhaite obtenir pour prouver que le coloriage est périodique. Soit c un coloriage de faible complexité. Alors il existe un polynôme de la forme suivante qui annule *c* :

$$\left(x^{a_1} y^{b_1} - 1\right)\left(x^{a_2} y^{b_2} - 1\right) \cdots \left(x^{a_r} y^{b_r} - 1\right).$$

Il est intéressant de remarquer que si r=1, alors le coloriage est périodique, puisque le polynôme est exactement de la forme $x^a y^{b-1}$. Kari et Szabados ont aussi montré que si $r=2$ et que c est de faible complexité, alors c est encore périodique, mais de manière moins évidente [Sza18b]. Dans leur article de 2015, ils prouvent une version "asymptotique" de la conjecture de Nivat : au lieu d'avoir l'hypothèse de l'existence d'une valeur de m et n tel que $p(m,n) \leq mn$, ils prouvent un énoncé similaire lorsque l'on dispose d'une infinité de telles valeurs de m et n [KS15].

Théorème [Kari & Szabados, 2015] Pour un coloriage 2D donné, s'il existe une infinité de $m,n \in \mathbb{Z}$ *tels que* $p(m,n) \leq mn$*, alors le coloriage est périodique.*

En utilisant ces techniques algébriques, Szabados propose dans un article publié en 2018 [Sza18b] une preuve simplifiée du théorème de Cyr et Kra [CK15].

Théorème [Cyr & Kra, 2015; Szabados, 2018] Pour un coloriage 2D donné, s'il existe une taille de motif $m \times n$ *telle que* $p(m,n) \leq mn/2$*, alors le coloriage est périodique.*

Toujours en utilisant ces techniques algébriques et en y incorporant des techniques de dynamique symbolique utilisées notamment par Cyr et Kra, nous avons, avec Kari, obtenu un théorème similaire à la conjecture de Nivat, mais portant sur une classe de coloriages particuliers : les coloriages uniformément récurrents [KM19]. On dit qu'un coloriage est uniformément récurrent si tout motif qui y apparaît y apparaît "partout", dans le sens où on peut le trouver à distance bornée de tout point du plan. Le coloriage donné en exemple à la page 5 est un exemple typique de coloriage qui n'est pas uniformément récurrent : la cellule noire se trouve uniquement au centre du coloriage, et ne peut être retrouvée nulle part ailleurs. Un coloriage uniformément récurrent permettrait de retrouver cette cellule noire "à tout endroit" du plan.

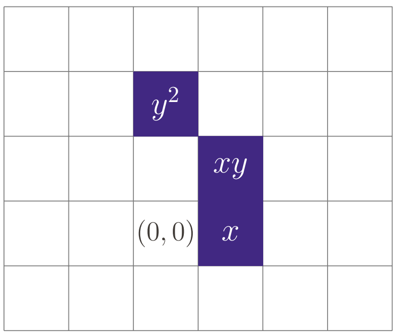
Théorème [Kari & Moutot, 2019] Pour un coloriage 2D uniformément récurrent, s'il existe une taille de motif $m \times n$ *telle que* $p(m,n) \leq mn$*, alors le coloriage est périodique.*

Ce théorème est particulièrement intéressant, puisque intuitivement les coloriages uniformément récurrents sont ceux où l'on pouvait le plus espérer trouver un contre-exemple à la conjecture de Nivat. En effet, on peut avoir l'intuition que les coloriages qui ne sont pas uniformément récurrents vont plus généralement être de grande complexité à cause des endroits où certains motifs peuvent être "pointés". Ainsi, il paraît peu probable de trouver un contre-exemple à la conjecture de Nivat qui ne soit pas uniformément récurrent. Et pourtant c'est ce dernier cas, celui des coloriages non-uniformément récurrents, qui reste encore ouvert, bien que semblant intuitivement plus simple.

Pour conclure, quelques mots sur cette nouvelle méthode, dont l'efficacité peut être surprenante au premier abord. En effet, avec l'approche algébrique on s'intéresse à un problème plus général que la "simple" périodicité d'un coloriage. Il semblerait donc logique que ce problème soit plus difficile à résoudre, alors pourquoi s'y intéresser ? C'est en fait une approche courante en mathématiques : prendre un problème et le voir comme un cas particulier d'un problème plus général. Mais une généralisation quelconque n'aurait en effet pas d'intérêt. Ici, comme dans beaucoup de cas, cette généralisation est utile car elle propose une structure supplémentaire, dont le problème initial ne disposait pas. Ici cette structure vient de l'opération d'addition que l'on autorise, et qui confère une structure algébrique au problème. Cette structure supplémentaire permet d'étudier la généralisation avec des outils de mathématiques puissants et généraux. Tout le problème est ensuite de revenir dans le contexte initial, puisque l'on obtient des conclusions qui ne parlent que de la généralisation. Dans le cas de la conjecture de Nivat, cela peut se faire dans la plupart des cas, permettant ainsi à cette approche d'être fructueuse, même si cela ne suffit pas (encore ?) à résoudre complètement la conjecture.

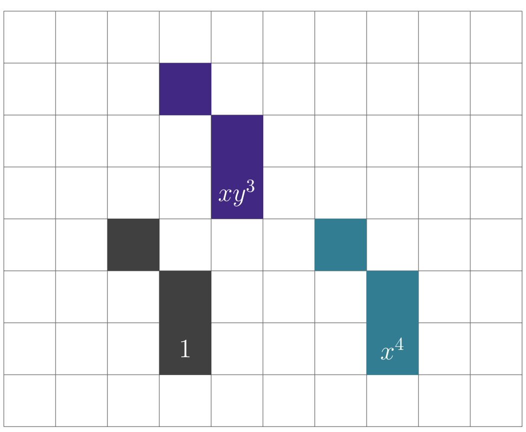
Inversement : Les polynômes sont des coloriages

Dans cette partie, nous revenons à des concepts plus simples et visuels que la précédente. On a vu plus haut que les coloriages du plan pouvaient être vus comme des polynômes, mais l'inverse est vrai aussi: tout polynôme peut être dessiné comme un coloriage ! Pour faire simple, prenons des polynômes dont les coefficients ne sont que *0* ou *1*. Alors, un polynôme développé peut être facilement colorié : on colorie en noir les cellules correspondant aux monômes non-nuls, et en blanc le reste.



Coloriage représentant $x + xy + y^2$.

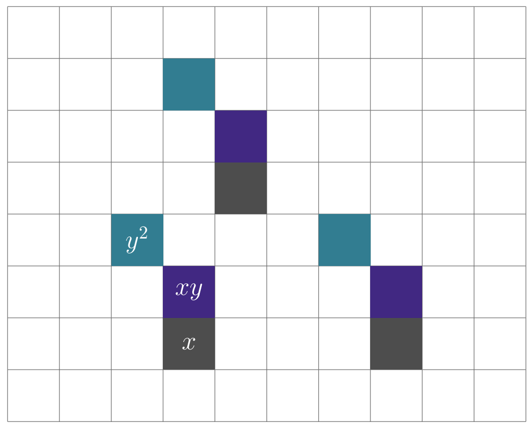
Pour interpréter le produit de deux polynômes, on va développer le produit et effectuer les translations du coloriage comme vu précédemment. Reprenons le polynôme de l'exemple précédent, $x+xy+y^2$ et multiplions-le par $(1+x^4+xy^3)$. Alors, puisque $(x+xy+y^2)(1+x^4+xy^3) = (x+xy+y^2) + x^4(x+xy+y^2) + xy^3(x+xy+y^2)$, le coloriage représentant ce produit sera égal à la somme de trois coloriages : celui représentant $(x+xy+y^2)$, $x4(x+xy+y^2)$ et $xy^3(x+xy+y^2)$. Or, ces trois coloriages ne sont autres que des translatés du polynôme initial par les vecteurs *(0,0)* (donc pas translaté), *(−4,0)* et *(−1,−3)*. Ainsi, le coloriage représentant le produit de deux polynômes est simplement un polynôme recopié, translaté, et sommé, les translations dépendant du second polynôme.



Coloriage représentant $(x + xy + y^2)(1 + x^4 + xy^3)$.

Toutes les couleurs représentent des "1" sur les cellules, elles sont là pour faire apparaître les trois translations de $(x+xy+y^2)$ selon les monômes de $(1+x^4+xy^3)$.

De manière moins évidente, on peut aussi voir $(x+xy+y^2)(1+x^4+xy^3)$ comme trois translations de $(1+x^4+xy^3)$ selon les monômes de $(x+xy+y^2)$.

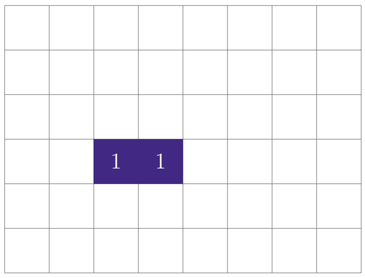


Le même coloriage, mais avec cette fois les couleurs choisies pour mettre en évidence les trois copies de $(1 + x^4 + xy^3)$ translatées selon les monômes de $(x + xy + y^2)$.

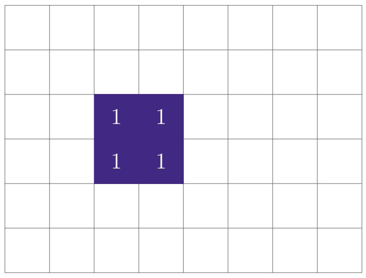
Évidemment cet exemple reste simple puisque les différentes translations ne se superposent pas. Dans le cas général le coloriage résultat peut être plus complexe, puisque si les translations se chevauchent il faut additionner (ou soustraire) les coefficients qui se superposent. Mais dans tous les cas cela nous donne un bel outil pour visualiser n'importe quel polynôme !

Pour aller plus loin

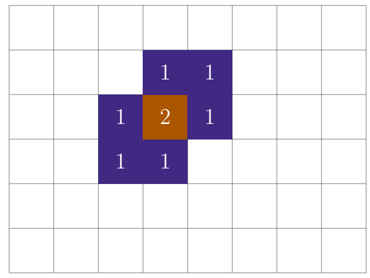
Si le sujet vous a intéressé et que vous n'avez pas peur de lire un peu d'anglais, voici quelques références qui permettent de rentrer dans les détails de ce qui a été abordé ici.
Une bonne introduction à la notion de complexité est donnée dans le début du chapitre "Factor complexity" du livre "Combinatorics, Automata and Number Theory" [CN10]. Le chapitre dans son ensemble est extrêmement complet sur cette notion.



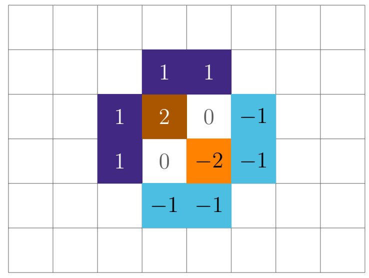
Coloriage représentant $(1+x)$.



Coloriage représentant $(1+x)(1+y)$.



Coloriage représentant $(1+x)(1+y)(1+xy)$.



Et enfin un coloriage représentant $(1+x)(1+y)(1+xy)(1-xy^{-1})$.

L'approche algébrique à la conjecture de Nivat est expliquée de manière très pédagogique dans la thèse de Szabados, "An Algebraic Geometric Approach to Nivat's Conjectur" [Sza18a]. Pour une liste plus exhaustive des avancées apportées par cette approche, Kari a publié en 2019 un court article les résumant toutes [Kar19].

Références

[CK15] Van Cyr and Bryna Kra. Nonexpansive $\mathbb{Z}^2$-subdynamics and nivat's conjecture. *Transactions of the American Mathematical Society*, 367(9) :6487–6537, 2015.

[CN10] Julien Cassaigne and François Nicolas. *Factor complexity*, page 163–247. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2010.

[EKM03] Chiara Epifanio, Michel Koskas, and Filippo Mignosi. On a conjecture on bidimensional words. *Theor. Comput. Sci.*, 299(1) :123–150, 2003.

[Kar19] Jarkko Kari. Low-complexity tilings of the plane. *arXiv :1905.04183*, 2019.

[KM19] Jarkko Kari and Etienne Moutot. Decidability and periodicity of low complexity tilings. *arXiv :1904.01267*, 2019.

[KS15] Jarkko Kari and Michal Szabados. An algebraic geometric approach to nivat's conjecture. In Magnús M Halldórsson, Kazuo Iwama, Naoki Kobayashi, and BettinaEditors Speckmann, editors, *Automata, Languages, and Programming*, volume 9135, page 273–285. Springer Berlin Heidelberg, 2015.

[MH38] Marston Morse and Gustav A. Hedlund. Symbolic dynamics. *American Journal of Mathematics*, 60(4) :815–866, 1938.

[Niv97] Maurice Nivat. Keynote address at the 25th anniversary of eatcs, during icalp 1997. Bologna, 1997.

[QZ04] Anthony Quas and Luca Zamboni. Periodicity and local complexity. *Theor. Comput. Sci.*, 319(1–3) :229–240, 2004.

[Sza18a] Michal Szabados. *An algebraic approach to Nivat's conjecture*. PhD thesis, University of Turku, 2018.

[Sza18b] Michal Szabados. Nivat's conjecture holds for sums of two periodic configurations. In A Min Tjoa, Ladjel Bellatreche, Stefan Biffl, Jan van Leeuwen, and Jiří Wiedermann, editors, *SOFSEM 2018 : Theory and Practice of Computer Science*, page 539–551. Springer International Publishing, 2018.

Entretien avec Claire Mathieu, DR (CNRS, IRIF)

Cet entretien a été conduit par Nicolas Schabanel le 5 novembre 2019, à l'occasion de la remise de la médaille d'argent du CNRS à Claire Mathieu.

Bonjour Claire, j'ai le plaisir de m'entretenir avec toi à l'occasion de la remise de ta médaille d'argent du CNRS. Nous aimerions revenir ensemble sur ta carrière et ta vision actuelle sur l'informatique et l'algorithmique en particulier. Tout d'abord, pourrais-tu retracer les étapes clés de ta carrière, les moments qui ont le plus compté dans ta vie scientifique ?

Le première a été le concours d'entrée de l'École Normale Supérieure de Jeunes Filles. Car pour le préparer, j'ai passé deux années en classes préparatoires à travailler dur, à apprendre, à découvrir les maths et cela m'a été utile tout au long de ma carrière. En particulier, au début, quand j'étais avec d'autres doctorants aux États-Unis, je calculais plus vite qu'eux, je savais faire les petites preuves plus vite aussi, et cela m'a donné un avantage pour trouver des professeur.e.s avec qui travailler.

Ensuite, grâce à un programme d'échange entre l'École Normale Supérieure et Princeton, j'ai eu l'occasion de travailler avec Andy Yao, sur un problème de calcul de fonction booléenne où il faut interroger les valeurs des variables pour pouvoir la calculer et où les réponses que l'on reçoit peuvent être erronées. C'était mon premier vrai travail de recherche, où j'ai découvert le plaisir de travailler sur une question ouverte puis de réfléchir, de m'y plonger et après d'avoir une idée, et de trouver une solution.

L'étape suivante a été à l'occasion de deux séjours à Berkeley, en 1992 et 1995. J'étais au CNRS, qui m'a mise à disposition de Berkeley pendant 1 an en 1994-1995, donc en plus je gardais mon salaire. Là, j'ai découvert la recherche en informatique avec les n°1 mondiaux de l'époque. Les étudiants travaillaient les uns avec les autres dans une atmosphère effervescente. Là, J'ai eu l'occasion de travailler avec Dick Karp sur deux ou trois projets et j'ai découvert comment il s'y prenait pour attaquer un problème de recherche qui résistait. Il cherchait systématiquement des exemples, des simplifications, des contre-exemples, de façon très organisée et en essayant à chaque fois de poser quelque chose de rigoureux de façon à ce que l'on ne se quitte jamais au bout d'une réunion de travail sans avoir un résultat écrit, même si la preuve prenait deux lignes, une brique sur laquelle s'appuyer pour continuer. Avec lui, j'ai appris une méthode qui a transformé ma vie de chercheur.

Ensuite, ma carrière s'est développée, j'ai eu des thésards, puis on m'a proposé de rejoindre le comité de programme de la conférence SODA, et j'ai découvert comment les chercheurs évaluent le travail des autres chercheurs, les critères qui font un bon article ou pas. J'ai travaillé sur les problèmes de bin packing avec l'équipe de David Johnson à AT&T et j'ai eu la chance de travailler sur les algorithmes d'approximation au moment où ce domaine démarrait, au début des années 1990, lors de mon séjour à Berkeley. Le domaine était alors en pleine explosion avec de nombreux “low hanging fruits”. Ma spécialité dans le domaine était plutôt les schémas d'approximation, c'est-à-dire les algorithmes qui produisent en temps polynomial une solution arbitrairement proche de l'optimum. J'ai participé pleinement à ces études de l'approximation des problèmes d'optimisation combinatoire dont on ne savait pas quoi faire rigoureusement avant, à part dire qu'ils étaient NP-complets.

Tu as exploré différents aspects de l'informatique théorique et de l'algorithmique : les pavages, l'ordonnancement, des problèmes de graphes, de clustering, etc. Quelle a été la logique de ces choix ?

Je pense rétrospectivement que le centre de mes activités est la conception et l'analyse d'algorithmes. La conception d'algorithmes est quelque chose qui nous est propre : quand je travaille avec des mathématiciens, ils s'intéressent principalement à démontrer des théorèmes, prouver des choses, mais la construction algorithmique, c'est secondaire pour eux ; alors que pour moi, c'est central.

Pour en revenir à ta question, voici quelques enchaînements de sujets. Au tout début, j'ai travaillé avec Andy Yao sur le problème de calcul de la valeur d'une fonction booléenne en présence d'un certain nombre d'erreurs, fixé, sur les valeurs des variables. Et je me suis dit que ce modèle d'erreurs n'était pas bon et j'ai travaillé au début des années 1990 avec Valerie King sur des modèles probabilistes d'erreur. Ensuite, je me suis dit que si on fait un tri, on peut faire des erreurs en comparant deux nombres, mais ces erreurs sont beaucoup moins probables si les nombres sont éloignés que s'ils sont proches : par exemple, le résultat d'un match de tennis entre les 2 premiers mondiaux est bien plus incertain qu'entre le premier et le 100ème mondial où le résultat est presque certain. Ainsi lorsque je suis allée à Berkeley, j'ai posé cette question à Dick Karp et nous avons conçu ensemble un modèle où la probabilité d'erreur dépend de la distance entre les deux éléments comparés. Un résultat dans cette veine va d'ailleurs être publié en janvier à SODA 2020, sur le calcul en présence d'erreur où l'on s'intéresse à l'évaluation d'algorithme en tenant compte de la taille du résultat : si le résultat est court à décrire, il est possible qu'il soit plus facile à calculer. Ceci est d'ailleurs en lien avec la complexité paramétrée, en pleine explosion ces dernières années, qui permet désormais de faire des choses bien plus difficiles. Voici un premier enchaînement de sujets.

Ensuite, les pavages. Quand j'étais en post-doc, je travaillais avec un mathématicien qui était mon mari à l'époque et qui avait fait sa thèse sur les pavages du plan par des tuiles. Thurston avait proposé un algorithme très original basé sur une représentation de groupe pour le cas des dominos. Nous en avons discuté et il s'intéressait à paver le plan avec des formes plus compliquées. J'ai proposé d'essayer d'abord avec des formes juste un peu plus compliquées, des rectangles 3x1, 4x1, etc. C'était un problème mathématique fascinant, un algorithme complètement inattendu et d'impact pratique… nul. La résolution de ce problème ne m'a pas satisfaite car il y a de nombreux cas où la figure est presque pavable et cet algorithme ne produit alors rien. Je me suis donc intéressée au pavage approximatif, où l'on s'autorise à casser quelques carreaux à la fin. J'ai décidé de commencer par le problème en une dimension: il s'agit d'un cas particulier du bin packing. On m'a déconseillé de me lancer dans ce domaine déjà très balisé, mais j'y suis allée quand même et il restait des choses à faire. À mon retour de Berkeley, je me suis dit qu'il y a au moins un cas que l'on devrait réussir à approcher en dimension 2 : lorsque la figure à paver est elle même très simple, un rectangle long et étroit, c'est-à-dire une bande. Cela nous a conduit avec mon collègue, Éric Rémila, à résoudre une version du bin packing bidimensionnel. C'est cet article qui a fait que d'un seul coup, le gens du domaine des algorithmes d'approximation ont repéré mon nom. Avant, j'étais invisible. J'avais tenté de parler de ce résultat avant sa publication à David Johnson lors d'une conférence, mais il m'avait ignorée. J'avais été un peu vexée, mais c'est normal, en fait, il ne me croyait pas, c'était un problème réputé difficile et l'article n'était pas encore écrit. Quand il a paru, 6 mois plus tard, j'ai reçu de nombreuses invitations à venir travailler avec untel, des personnes voulaient me rendre visite, des conférences m'ont invitée dans leur comité de programme etc.. Les techniques que l'on avait développées pour ce problème s'appliquaient tout naturellement à des problèmes d'ordonnancement et donc de là, ça a été un tout petit pas de côté pour travailler sur l'ordonnancement (ton sujet de thèse entre autres !). Comme tout était basé sur l'approximation des nombres de façon plus ou moins difficile, ça a fait une autre série d'articles qui ont eu du succès pendant de longues années.

Vers 1997, j'ai vu que je comprenais bien comment marchaient les problèmes de bin packing. Je me suis dit qu'en fait les données de tous ces problèmes n'étaient que des nombres, mais qu'il y avait tout un pan de l'algorithmique où les données sont autre chose : les graphes, et ça, je ne savais pas faire. J'ai eu alors la chance de rencontrer Laló Fernandez de La Vega. C'était l'une des toutes premières personnes à avoir trouvé un schéma d'approximation pour le bin packing. Quand on s'est rencontré pour parler du bin packing 2D, il m'a dit qu'il s'intéressait maintenant à autre chose : il avait trouvé un schéma d'approximation pour le problème de la coupe maximum dans les graphes denses. Nous avons travaillé ensemble sur le cas métrique. L'article a été publié à FOCS. Les gens ont été surpris par ce résultat inattendu. De plus, entre le moment où il a été accepté et la rédaction finale, nous avons trouvé comment simplifier la preuve qui était très technique, au point qu'elle tenait sur une seule page. Cela m'a permis de l'exposer entièrement lors de la conférence, dans les 20 minutes allouées. Cet article a été beaucoup enseigné par la suite. Voici encore un exemple d'enchainement de problèmes.

Tu as passé une part importante de ta vie de chercheuse entre la France et les États-Unis, pour revenir finalement en France. Qu'est-ce que ces deux pays t'ont apporté ?

Au début de ma carrière, l'analyse d'algorithmes était bien représentée en France, mais la conception d'algorithmes assez peu. C'était aux États-Unis que ça se faisait. Le programme d'échange de l'École Normale Supérieure et le fait d'être au CNRS ensuite, m'ont donné la liberté de faire des visites aux États-Unis, m'ont permis de mûrir, et d'apprendre à mieux faire de la recherche. Ces contacts fréquents au début de ma carrière et cette liberté apportée par la structure de la recherche française m'ont permis de me lancer. Sans eux, je ne serai pas ici maintenant.

Encore aujourd'hui, c'est clair, le pays le plus actif dans mon domaine sont les États-Unis. Il est donc important de maintenir des liens forts avec eux et j'y suis allée fréquemment tout au long de ma carrière. De 2004 à 2012, j'ai été professeur à l'université de Brown, près de Boston et c'est là que j'ai appris le vrai métier de professeur. C'est une petite université où les professeurs font une quantité de cours modérée, un cours par semestre (2 séances par semaine) et où la priorité est donnée à la qualité des cours ainsi qu'à leur renouvellement. Nous passions donc beaucoup de temps à réfléchir sur la pédagogie. Nous étions aussi tenus de passer au moins la moitié, plutôt les deux tiers, de notre temps à la recherche. Et le directeur de laboratoire, les collègues étaient là pour créer les conditions optimales pour bien avancer en recherche. J'appréciais aussi beaucoup le fait que dans cette université, c'était le corps des professeurs (qui se réunissait 2 fois par an) qui en guidait les choix stratégiques. Nous prenions les décisions de façon collégiale pour que l'on fasse le meilleur boulot possible pour l'enseignement et la recherche. C'était le travail parfait pour moi. Je m'y suis beaucoup plu.

C'est d'ailleurs ce qui a été le plus difficile pour moi à mon retour des États-Unis. Là-bas, à Brown, j'étais partie prenante des décisions. Le doyen prenait l'avis des directeurs de laboratoires, et consultait certains professeurs à l'occasion. On discutait ensemble en essayant de discerner les changements à faire pour l'évolution nécessaire avec le temps. À mon retour en France, j'ai eu la sensation d'être un pion et non un acteur de la recherche en France, un simple exécutant censé se plier à des directives venant "d'en haut", alors que c'est nous qui faisons la recherche.

Mais alors, pourquoi rentrer ?

J'aime la France, ses paysages, ses bâtiments, son épaisseur historique. Les villes y sont plus humaines, la société moins dure. Il y a aussi les liens familiaux bien sûr… Concernant le travail, j'ai aussi eu des opportunités ici que je n'aurais pas eues aux États-Unis. J'ai été professeur au Collège de France, un institut qui n'a pas d'équivalent aux États-Unis. Ça a été pour moi l'occasion d'enseigner au grand public, et de rentrer en contact avec des gens hors informatique, d'autres disciplines scientifiques, et avec les médias et les ministères.

Quels conseils pourrais-tu donner aux chercheurs, les jeunes en particulier, à qui on demande de plus en plus d'expérience à l'étranger ?

J'ai beaucoup voyagé, certes, mais j'ai eu la chance de pouvoir le faire dans de bonnes conditions. J'avais la garantie de l'emploi et j'ai fait mes voyages par choix, non par contrainte.

Une chose importante que je souhaite dire est qu'en tant que femme, j'ai pu avoir mes enfants relativement jeune car j'étais au CNRS. Quand on a un bébé à la maison, un enfant, c'est un fait, la productivité décroît. Titulaire au CNRS, on n'a pas à s'inquiéter, c'est temporaire, après ça va remonter. C'est une sérénité bienvenue. De plus, comme on n'est pas tenue d'enseigner, si on est fatiguée pendant 6 mois car on ne dort pas bien, qu'on ne s'en sort pas, ce n'est pas grave. Ça donne un peu de mou et ça a été vraiment une chance exceptionnelle pour réussir à combiner ma vie de famille et mon travail.

La plupart des chercheurs et enseignants chercheurs que l'on recrute ont maintenant 4 à 5 ans de post doc. As-tu des conseils à leurs donner ?

Nous assistons à une dérive qu'il faut stopper. Je crois que les comités de sélection ne devraient pas recruter des personnes qui ont 4 ou 5 ans de post-doc. C'est un engrenage. À un moment, il faut que le.a post-doctorant.e se résolve à passer à autre chose. Ça éviterait que la précarité se prolonge indéfiniment. Quand je prend mes étudiants en thèse, je leur demande toujours "Avez-vous un plan B qui vous satisferait ?", car peu importe d'où ils viennent, qu'ils aient été brillants ou non, on ne sait jamais à l'avance comment ça va se passer.

Toi qui a étudié la mise en place et la pérennisation de la parité dans les milieux compétitifs, quelle est ton analyse de la situation en sciences et quelles seraient tes recommandations ?

On voit que dans les sciences dures, il y a très peu de femmes. Je suis convaincue qu'il n'y a pas de raison intrinsèque et qu'en se restreignant à ne choisir qu'essentiellement des hommes, on se prive de la moitié du vivier. À partir du moment où le nombre de femmes est inférieur à, disons, 20%, cela veut dire que lors des réunions avec une demi-douzaine d'enseignants, on est la seule femme, et que dans une classe de 30 élèves, il n'y a que quelques filles dans le groupe. Et ça c'est très mauvais pour l'ambiance et… pour tout le monde à mon avis. Il faut donc essayer d'être pro-actif pour tenter d'augmenter le nombre de femmes. Maintenant, comment faire ? Voici ce qui se faisait à Brown. Tout d'abord il y avait dans chaque comité de recrutement un.e responsable "parité et minorité". On regardait la proportion de femmes parmi les candidats et cette proportion devait être au moins la même parmi les auditionnés. Sinon il fallait écrire un argumentaire. Il n'y avait pas de pourcentage imposé, mais une vigilance, une demande de justifier systématiquement par un paragraphe écrit lorsque les pourcentages n'étaient pas respectés ou que le poste n'était pas proposé d'abord à une femme. C'était assez efficace car ça obligeait les gens à garder cela à l'esprit. Cela dit, j'ai l'impression que la situation s'améliore en France. J'ai entendu des mauvaises blagues, des remarques qui dénigrent les femmes, mais… c'était plutôt des vieux qui sont partis ou vont bientôt partir à la retraite. Je crois que la génération montante a intégré la parité. Bien sûr, il reste encore quelques remarques déplacées. Je crois qu'il faut qu'on apprenne déjà aux gens à ne plus les dire quand elles leurs viennent à l'esprit, et ensuite, à ne plus les penser. Les mentalités doivent encore évoluer mais ça, ça prendra du temps.

Les petites phrases peuvent avoir un impact important, positif aussi. Par exemple, en Math Sup, j'ai eu une mauvaise note en maths et mon professeur m'a demandé "Ça ne va pas très fort Mlle Mathieu en ce moment ?" et il m'a encouragée et m'a demandé ce que je voulais faire. Je lui ai répondu que je souhaitais intégrer l'École Normale Supérieure de Jeunes filles mais que je ne savais pas si je réussirais. Et il m'a répondu "Mais vous réussirez". Pendant toute l'année suivante, chaque fois que je me sentais découragée, je me rappelais que mon prof m'avait dit que j'allais réussir. Donc un mot, une phrase, ça peut avoir un impact important.

Quand on voit les carrières tout à fait comparables aux hommes des anciennes étudiantes de l'École Normale Supérieures de Jeunes Filles, on est en droit de se demander si ce n'était pas une erreur d'avoir supprimé le concours séparé…

Dans les concours en France, l'écrit est anonyme et on corrige les copies indifféremment que ce soit une fille ou un garçon. On peut espérer qu'à l'oral, la proportion d'admises soit au moins celle des admissibles. Malgré tout, il semble que pour des raisons que je ne m'explique pas, les garçons s'y retrouvent mieux dans la conception très axée sur la compétition et l'individualisme des classes préparatoires.

Tu t'es beaucoup investie dans Parcours Sup'.

Ecole Jeunes Chercheurs Jeunes Chercheuses en Informatique Mathématique 2020

L'Ecole Jeunes Chercheurs Jeunes Chercheuses en Informatique Mathématique 2020 s'est déroulée en ligne (une première !) du 8 au 18 juin 2020. Félicitations aux organisateurs pour un travail d'une très grande qualité. Les livres regroupant les cours des EJCIM passées peuvent être téléchargés à https://www.gdr-im.fr/im-photographie/ .

Quel a été ton retour d'expérience sur cette confrontation avec le terrain ?

J'ai trouvé cela passionnant, c'était la première fois que je faisais quelque chose qui a eu un impact direct sur les gens. Avec Hugo Gimbert, nous étions les consultants côté algorithmique. Les décisions étaient prises en groupe, chacun avait une voix. Le fait de travailler avec des gens d'autres domaines, d'autres milieux, m'a permis de me rendre compte de la complexité du problème qui faisait que les décisions prises n'étaient pas forcément optimales du point de vue algorithmique, mais les bonnes quand on prenait en compte l'ensemble des facteurs. Je me suis fait insulter occasionnellement sur internet. Ça surprend les premières fois, mais j'assume, ça ne me dérange pas. Ce qui était intéressant c'était que parmi les opposants à Parcours Sup', certains sont clairement biaisés par leurs opinions et quand on leurs explique trois choses, ils ne retiennent que celle qui sert leurs objectifs, qui n'ont rien à voir avec la réalité. En revanche, il y en a d'autres que je respecte tout à fait, qui cherchent à comprendre, ont des objections légitimes et qui sont opposés à Parcours Sup' pour des raisons sérieuses. C'était une expérience très riche.

L'informatique a révolutionné notre perception du monde ces dernières années, encore plus récemment avec l'intelligence artificielle… Est-ce la fin de l'algorithmique ?

Il est clair que le monde dans lequel grandissent nos enfants n'a pas grand chose à voir avec celui dans lequel nous nos parents ont grandi. La capacité des systèmes à récueillir des données sur nous, les possibilités de fichage (reconnaissance faciale, ADN), la dépersonnalisation de ces systèmes peuvent engendrer beaucoup d'inquiétudes. Cependant, il existe de nombreuses avancées très encourageantes. Notre manque d'imagination nous limite. Moi même, j'utilise depuis quelques temps une app qui compte mes pas chaque jour et m'indique si je marche assez ou non. Je ne ressentais pas le besoin d'une telle app, mais maintenant que je l'ai essayée je suis très contente de l'avoir. Qu'y a-t'il d'autre qui serait techniquement faisable et qui améliorerait ma qualité de vie, mais que je n'arrive même pas à imaginer ?

Maintenant, concernant l'intelligence artificielle, c'est bien sûr tout sauf la fin de l'algorithmique. Dans de nombreux contextes, il est inacceptable de ne pas pouvoir expliquer comment le résultat est obtenu, ni de pouvoir en assurer la conformité à certaines propriétés (la conformité aux textes de loi par exemple). Depuis 5 ans environ, de nombreuses recherches en algorithmique progressent dans cette direction, en cherchant à réduire cet aspect "boîte noire" de l'intelligence artificielle. Lorsque s'est posée la question des recommandations aux candidats dans Parcours Sup', nous avons fait le choix de pouvoir expliquer aux candidats son fonctionnement et nous avons volontairement limité les possibilités.

La confiance du grand public dans la science a été entachée par de nombreux scandales (laboratoire pharmaceutiques, engrais, surveillance, …). L'image de la science s'est trouvée mise à mal et on constate une désaffection pour les carrières scientifiques. As-tu un avis sur cette question ?

Il me semble que cela est surtout vrai aux États-Unis, moins en France. Pour la première fois depuis longtemps, les jeunes adultes aux USA ne sont pas plus riches que leurs parents, et leur niveau de vie ne progresse plus. Si les progrès scientifiques, les machines ne servent que les personnes les plus fortunées et pas la majorité de la société, on a un gros problème. Il faut s'assurer que les progrès scientifiques et techniques profitent à la majorité des gens. Par exemple, si les investissements dans les trains consistent à privilégier des lignes de TGV toujours plus rapides et plus chères, en délaissant les lignes rurales, il est normal que les gens des campagne protestent contre ces "progrès". Il faut que les changements, les méthodes d'éducation, les programmes, l'organisation générale soient d'abord bénéfiques à la majorité des gens. Sinon, ça ne fonctionne pas. Ensuite, on peut s'occuper des 10% les plus démunis, et après seulement, des 10% les plus riches.

Pour en revenir à la science, quelles sont pour toi les grandes directions de la recherche en algorithmique ? Quel impact sur la société ?

C'est peut-être lié à mon biais personnel envers la recherche interdisciplinaire, mais je crois beaucoup dans le travail algorithmique en liaison avec les économistes qui permet de revisiter des problèmes en y incorporant d'autres critères, comme par exemple la rationalité, ou que les agents aient des stratégies ou non, etc. Cela s'est concrétisé par la création de conférences EC il y a une dizaine d'années. Je crois que nous avons besoin de travailler avec des gens d'autres disciplines pour apprendre d'eux ce qui est important de leur point de vue. En effet, la plupart des problèmes que l'on résout n'ont pas d'applications directes, car nos solutions ne sont pas… pertinentes ! Mais il n'empêche que notre travail a quand même une valeur, car des éléments dedans peuvent être utilisés par d'autres personnes, à condition que l'on ait conscience de leurs priorités, de ce qui est important pour elles. J'ai d'ailleurs pris récemment conscience de l'importance de la simplicité des solutions proposées. Avant, pour moi, la simplicité n'était pas un but en soi, cela permettait par exemple d'avoir des preuves plus simples. Désormais, je crois que c'est essentiel pour que d'autres personnes puissent s'en inspirer. Je suis d'ailleurs beaucoup plus intéressée qu'il y a 15 ans par l'analyse d'algorithmes très simples, en démontrant qu'ils sont efficaces dans certains cadres, même restreints. Il me semble également que les algorithmicien.ne.s ne doivent pas trop s'éloigner des systèmes réels. Je me réjouis de la création de la conférence APOCS qui réunit algorithmique, systèmes et bases de données pour étudier les algorithmes, peut-être pas les plus élégants, mais qui fonctionnent dans des conditions réelles.

Pour finir, sur quoi travailles-tu en ce moment ?

Avec des collègues du LIP6, nous travaillons autour du redécoupage des circonscriptions électorales. Grâce à etalab.gouv.fr, data.gouv.fr et au site de l'INSEE, nous avons désormais accès à de nombreuses données et nous pouvons travailler à améliorer les méthodes utilisées. C'est un problème de clustering très différent de ceux que l'on a étudiés jusqu'à présent. Il s'agit en effet, conformément à la loi, de découper le territoire en parties connexes qui soient toutes équilibrées démographiquement à ±20% suivant un certain nombre de critères. Bon, il y a peu de chance que l'on arrive à toucher directement les politiques. En revanche, nous pensons que proposer au grand public un moyen d'évaluer la qualité du découpage des circonscriptions via une app par exemple pourrait avoir un impact.

Merci beaucoup Claire et bonne chance avec ce projet !

Contributions Futures

Vous voulez participer à un futur numéro en proposant une contribution ?

Contactez les responsables du gdr-im : Jean-Michel Muller <jean-michel.muller@ens-lyon.fr> et Guillaume Theyssier <guillaume.theyssier@cnrs.fr>.

Une place pour chaque donnée et chaque donnée à sa place

Alexandre Vigny (Universität Bremen, Allemagne)

Vous sortez de chez vous pour aller au cinéma. Vous arrivez à la station de métro que vous pensiez utiliser, et là le drame ! À cause d'une inondation, la station est fermée. Vous vous empressez de sortir votre téléphone à la recherche d'un trajet alternatif. Quelques clics, deux bus et un tram plus tard, vous arrivez au cinéma. Vous êtes à l'heure pour votre séance, grace à une base de données !

Qu'est ce qu'une base de données ?

Une base de donnée est une structure qui enregistre et organise des informations, de manière à répondre à toutes nos questions. L'arrivée de nouvelles technologies a généré un nombre croissant de données, qu'il faut à présent stocker et organiser. De nos jours, en seulement quelques clics beaucoup trouvent sur Wikipédia, sur des forums, ou sur leurs sites préférés, toutes les informations dont ils ou elles ont besoin. Ces informations sont stockées sur des bases de données qui sont accessibles au public. Il ne reste plus qu'à poser les bonnes questions !

Un peu d'histoire.

Malgré un essor plutôt récent, le concept de bases de données ne date pas d'hier. Un annuaire qui associe des noms, des numéros de téléphones, et des adresses n'est rien d'autre qu'une base de données. Un registre qui contient les horaires de départ de train avec leur destination et le nombre de places disponibles à bord est également une base de données.

Concrètement, une base de donnée est un objet qui permet de stocker des informations, de sorte que lorsque qu'une requête (une question) est posée, il est possible de calculer l'ensemble de toutes les solutions (c'est à dire toutes les réponses à la question). On peut par exemple calculer la liste de tous les trains qui partent de la gare de Lille entre 9h et 10h.

En 1970, dans *A relational model of data for large shared data banks* [https://dl.acm.org/citation.cfm?doid=362384.362685], E. F. Codd pose les bases de ce qui deviendra les bases de données modernes. Plus précisément, il définit le *modèle relationnel.*

Dans ce modèle, les bases de données sont vues comme une collection de tableaux. Chaque tableau est appelé une *relation*. Par exemple un tableau dont chaque ligne correspond à un trajet de train. Les colonnes pourraient être: le numéro du train, la ville de départ, la ville d'arrivée, l'heure de départ, la durée du trajet, la capacité du train, ...

Dans la même base de données, on peut définir une autre relation qui, elle, contiendrait des informations à propos des passagers. Chaque billet correspondrait à une ligne et les colonnes seraient: nom, abonnement, numéro du train, voiture, siège, ...

Le modèle relationnel ne se cantonne pas à définir la structure des bases de données. Il définit aussi la structure des requêtes qui peuvent être posées : l'algèbre relationnelle. Peu après avoir été définie par Codd, cette notion est implémentée et commercialisée pour la première fois en 1979, sous

Trains	Départ	Destination	Horaire	Durée	Capacité
ICE669	Paris-Est	Frankfort	13h10	4h05	550
TGV 2234	Lille	Marseilles	9h03	5h27	610
TGV 8900	Nantes	Paris	6h06	2h06	245

Nom	Abonnement	Train	Voiture	Siège
Tom	Carte Jeune	TGV 8900	3	32
Belinda	Carte Senior	ICE669	2	10
Phil	Aucun	TGV 8900	3	34
Anne	Carte Week-end	TGV 2234	5	2

Tables des trains et des usagers

le nom de SQL (pour Structured Query Language, ou en français Langage de Requête Structurée) [https://fr.wikipedia.org/wiki/Structured_Query_Language].
Depuis les années 80 et jusqu'à aujourd'hui, de nombreux systèmes de gestion de bases de données tels que MySQL [https://www.mysql.com/fr/] et PostgreSQL [https://www.postgresql.org] ont été développés, commercialisés, ou mis gratuitement à disposition du public. Bien que très nombreux, ces systèmes ne contiennent que de légères différences de fonctionnalités et quelques différences de syntaxe. Tous se conforment globalement aux standardisations internationales de l'ANSI (American National Standards Institute) et de l'ISO (International Organization for Standardization). Au cœur de cette standardisation on retrouve l'algèbre relationnelle de Codd.

Plus récemment.
On pourrait croire que 50 ans après son introduction, tout ce qu'il y a à savoir sur le modèle relationnel est aujourd'hui connu. C'est loin d'être le cas. Son étude reste un domaine de recherche industriel et académique des plus actifs.
Parmi les problèmes les plus étudiés, se trouve *l'évaluation des requêtes* : étant donné une basse de données et une requête, comment calculer l'ensemble des solutions le plus efficacement possible ?
Par exemple: ``Comment aller de Gare de Lyon à Gare de l'Est en moins de 2 changements ?'', ``Est ce qu'il est possible d'aller à Montparnasse en moins de 20 minutes ?'' , ou encore ``Quels sont les sièges non réservés du Paris-Lille de 9h17 ?''.

Des solutions à ne plus savoir quoi en faire.
Avant d'ouvrir une nouvelle ligne, la SNCF peut se demander ``Quelles sont les paires de villes qui corespondent à un trajet de plus de 1h30, parcouru quotidiennement par plus de 1000 usagers ?'' Il se peut que la plupart des paires de villes soient solutions. Or, si le nombre de villes est élevé, le nombre de paires de villes l'est plus encore ! On peut facilement imaginer des requêtes (peut être plus abstraites) pour lesquelles les solutions seraient des triplets de villes, ou bien des quadruplets, ...
De manière plus rigoureuse, si notre réseau contient n villes et que chaque solution correspond à un ensemble de k villes, le nombre de solutions peut être n^k. Donc avec une base de données de 50 villes (qu'on peut donc raisonnablement décrire sur une feuille de papier) et des solutions impliquant une dizaine de villes chacune, le nombre de solutions peut atteindre 50^{10}, ce qui est plus que le nombre de lettres dans tous les livres jamais écrits mis bout à bout !
En effet, d'après le Guinness World Records [http://www.guinnessworldrecords.com/world-records/longest-novel], le plus grand livre écrit est *À la recherche du temps perdu* par Marcel Proust avec environ 10 millions (10^7) de lettres. Le nombre de livres jamais écrits a été estimé par une étude Google [http://booksearch.blogspot.fr/2010/08/books-of-world-stand-up-and-be-counted.html] à environ 130 millions ($1,3x10^8$), soit moins de $1,3x10^{15}$ lettres dans tout les livres combinés, alors que $50^{10} \simeq 10^{17}$.

Et si on ne veut pas toutes les solutions ?
Fort heureusement, il est rarement nécessaire de calculer toutes les solutions. La plupart du temps on peut se contenter de certaines informations sur cet ensemble de solutions. On peut par exemple demander ``Existe-t-il une solution ?'' dans ce cas la réponse ne peut être que oui ou non. On parle alors de *problème de décision*. Il est aussi possible de demander ``Combien y a t'il de solutions ?''. On parle alors de *problème de comptage*.
Il n'est souvent pas utile d'avoir toutes les solutions en même temps. Dans ce cas on peut se contenter de les calculer une par une. On parle alors du *problème d'énumération*. Ce qui signifie que la personne qui pose la requête recevra assez rapidement une première solution, puis, à la demande, une seconde solution, puis une troisième, etc. Le processus s’arrête si il n'y a plus de solutions ou si la personne est satisfaite par les premières solutions proposées.

Répondre aux questions naïvement, c'est toujours possible.
Mais quand peut-on faire mieux ?
Dans le modèle relationnel de Codd, pour n'importe quelle base de données et n'importe

quelle requête, il est toujours possible de calculer les solutions. Il suffit de tester toutes les combinaisons ! C'est ce qu'on appel un algorithme naïf. L'objectif devient donc : est il possible de calculer les solutions plus rapidement que naïvement?

Au delà même du domaine des base de données, un grand champ d'étude en informatique théorique consiste à trouver des algorithmes qui résolvent des problèmes complexes de la manière la plus efficace possible. Cependant, après la conception d'un algorithme, il est très difficile d'être sûr de ne pas pouvoir l'améliorer. Dans certains cas il existe des preuves qui montrent qu'il n'existe pas d'algorithme plus efficace. Dans d'autres cas, ce n'est qu'une conjecture.

Pour le problème d'évaluation de requêtes, ainsi que ses dérivés (problèmes de décision, de comptage et d'énumération), nous nous trouvons dans le second cas.

Nous ne connaissons pas d'algorithme qui soit plus efficace qu'un algorithme naïf et qui fonctionnerait avec n'importe quelle requête et n'importe quelle base de donnée.

Il se pourrait que des tels algorithmes existent, mais il est globalement accepté par la communauté scientifique que ce n'est pas le cas.

Que faire si tout est trop compliqué ?

Il ne faut cependant pas désespérer. En attendant que des théoriciens prouvent ou invalident ces conjectures, il reste du pain sur la planche. Notamment, on peut se contenter d'algorithmes qui ne fonctionnent que pour certaines requêtes et certaines bases de données.

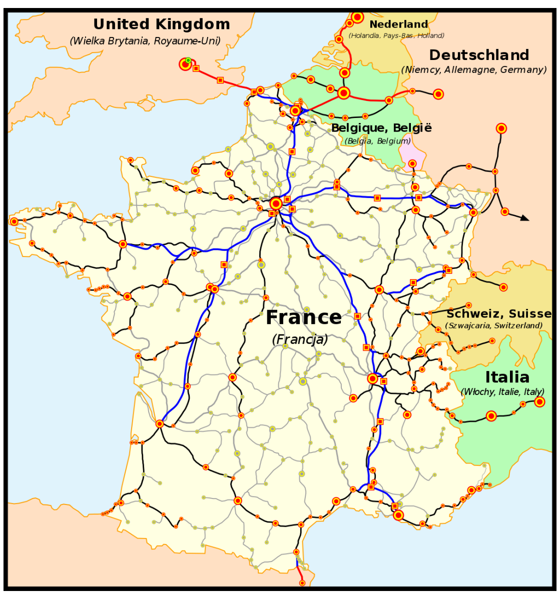
Retournons à notre exemple de réseaux de chemins de fer. Un réseau peut avoir des particularités exploitables astucieusement par un algorithme. Par exemple, un réseau ``en étoile'', assez proche du système de train français où quasiment tout passe par Paris. Dans ce cas, pour aller d'un point A à un point B, le plus court chemin est probablement un train direct ou un changement à Paris. Il devient alors assez facile de générer des premières solutions.

Un autre exemple peut être un réseau ``en grille'' comme à New-York, où les rues et la plupart des lignes de métro ne sont que Nord-Sud ou Est-Ouest. Un algorithme saura tirer partie de ces spécificités et sera plus efficace.

Et pendant ce temps, dans les laboratoires.

Caractériser précisément les propriétés des bases de données qui permettent d'obtenir des algorithmes efficaces est un domaine de recherche à part entière.

Il y a un peu plus d'une dizaine d'années, J. Nešetřil et P. Ossona de Mendez ont défini la notion de graphes *nulle-part denses*, une caractérisation mathématique des réseaux ``pas trop touffus'' [https://www.sciencedirect.com/science/article/pii/S0195669811000151]. Cette notion, généralise



Réseaux "en étoile" à la française ou "en grille" comme à New York

Nouveau site du gdr-im

Le site a été remanié : https://www.gdr-im.fr/

beaucoup de notions comme celles des réseaux ``en étoile'' ou ``en grille''.
Ces quelques dernières années, l'étude de ces structures nulle-part denses a attiré l'œil de la communauté scientifique des bases de données. Il a été prouvé que le problème de décision peut être résolu efficacement sur les base de données qui ont une structure nulle-part dense [https://dl.acm.org/citation.cfm?id=3051095&dl=ACM&coll=DL]. Pour les problèmes de comptage [https://dl.acm.org/citation.cfm?doid=3196959.3196970] et d'énumération [https://dl.acm.org/citation.cfm?doid=3196959.3196971], l'existence d'algorithmes performants a été prouvée à son tour en 2018.

En attendant que les algorithmes issus des laboratoires théoriques soient mis en application, ne vous impatientez pas trop s'il faut du temps pour trouver un itinéraire de substitution !

Programme des journées nationales du gdr-im
(du 23 au 26 mars 2021)

Conférenciers invités

Jean-Daniel Boissonnat, Inria Sophia Antipolis Méditerranée, DataShape
Triangulations des variétés : de l'existence à la construction
Peter Burgisser, Technical University of Berlin, Institute of Mathematics
Towards a theory of non-commutative optimization: geodesic first and second order methods for moment maps and polytopes
Assia Mahboubi, Inria, Laboratoire des Sciences du Numérique de Nantes
Méthodes numériques rigoureuses et formellement vérifiées
Claire Mathieu, CNRS, IRIF
Problèmes de partitionnement

Demi-journée thématique Informatique Quantique

La demi-journée thématique, qui aura lieu le jeudi 25 mars matin, sera consacrée à l'Informatique Quantique. Elle est organisée et animée par Simon Perdrix (CNRS, Nancy, LORIA) et Benoît Valiron (LRI, Université Paris-Saclay).

Orateurs des GTs, posters

et plus d'informations sur https://jnim2020.sciencesconf.org/

Comité éditorial : Jean-Michel Muller, Guillaume Theyssier et la cellule communication du gdr-im (Valérie Berthé, Julien Clément, Ines Klimann, Youssouf Oualhadj et Pascal Vanier).
Maquette et production : Ines Klimann.

gdr iM