

# Superpermutations

-

## Rapport de stage de L3

Guillaume DUMAS

Encadré par Ion NECHITA, LPT Toulouse

### Table des matières

<b>1 Introduction</b>	<b>1</b>
<b>2 État de l'art</b>	<b>1</b>
2.1 Définition du problème . . . . .	2
2.2 Construction standard et première conjecture . . . . .	2
2.3 Réfutation de la conjecture $\text{Opt}(n) = B_{\text{sup},1}(n)$ . . . . .	3
2.4 Une meilleure borne inférieure . . . . .	4
<b>3 Amélioration de la borne supérieure</b>	<b>6</b>
3.1 Une nouvelle borne supérieure . . . . .	6
3.2 Etat actuel . . . . .	10
<b>4 Matrices de superpermutation</b>	<b>11</b>

## 1 Introduction

Durant mon stage encadré par Ion Nechita au Laboratoire de Physique Théorique de Toulouse, j'ai étudié les superpermutations. Une  $n$ -superpermutation est un mot qui contient chaque permutation de  $n$  caractères comme facteur. Par exemple, « 121 » est une 2-superpermutation, et « 123121321 » est une 3-superpermutation. La question principale qui se pose est de trouver la taille minimale d'une  $n$ -superpermutation. Le but de mon stage a été principalement de dresser un état des lieux de la littérature sur ce sujet.

La première partie de ce rapport est un historique des avancées autour de ce problème, depuis sa formulation en 1993 par Ashlock et Tillotson [2]. Ensuite, je développerai la preuve d'une borne supérieure sur la longueur minimale par Greg Egan [4], qui est l'avancée majeure pour la résolution du problème. Enfin dans une dernière partie, je présenterai les résultats que j'ai obtenus en m'intéressant à un problème sur des matrices au lieu de mots.

## 2 État de l'art

La première partie de mon stage a été dédiée à l'étude des différentes bornes sur la longueur optimale d'une superpermutation. Je vais dans cette partie résumer les avancées et expliquer les outils de théorie des graphes.

## 2.1 Définition du problème

Un mot  $w$  sur  $A$  est une séquence finie de lettres de  $A$ . On note  $|w|$  la longueur du mot, c'est-à-dire son nombre de lettres, et  $w(i)$  la  $i$ ème lettre. On peut donc écrire  $w = w(1)w(2)\dots w(|w|)$ . On notera aussi  $[n] = \{1, \dots, n\}$ . Une permutation de  $S_n$  est alors un mot de taille  $n$  sur  $[n]$  où chaque lettre de  $[n]$  apparaît une et une seule fois (si  $\pi \in S_n$ , sa représentation sera  $\pi(1)\dots\pi(n)$  que l'on notera également  $\pi$ ).

**Définition 1.** On appelle  $n$ -superpermutation un mot sur l'alphabet  $[n]$  qui contient chaque permutation de taille  $n$  en facteur. Le but est alors de trouver pour tout  $n$  la longueur minimale d'une  $n$ -superpermutation. On la note  $\text{Opt}(n)$  (OEIS A180632).

On peut déjà donner une première borne inférieure triviale : en effet, un facteur d'un mot  $w$  est défini par la position  $i$  de sa première lettre dans le mot. Mais alors sa dernière lettre, à la position  $i + (n - 1)$  doit aussi être dans le mot. Donc le nombre maximal de facteurs que peut contenir un mot  $w$  est le nombre de  $i$  vérifiant :  $1 \leq i \leq n$  et  $1 \leq i + n - 1 \leq |w|$ , ce qui donne  $1 \leq i \leq |w| - n + 1$ . Comme on veut  $n!$  facteurs différents au moins, on doit avoir  $|w| \geq n! + n - 1 = B_{\text{inf},1}(n)$ .

Pour  $n = 1$ , on a  $w = 1$  qui atteint cette borne, pour  $n = 2$ ,  $w = 121$  atteint aussi la borne : elle est optimale pour ces deux entiers.

Pour  $n = 3$  : On a  $w = 123121321$  de taille 9, or  $B_{\text{inf},1}(3) = 8$ . La solution optimale est elle 8 ou 9? Essayons de montrer que c'est 9 (et donc que la borne n'est pas optimale).

Définissons la notion de lettre *gaspillée*.

**Définition 2.** Soit  $w$  une  $n$ -superpermutation. On dit qu'une lettre  $w(i)$  est *gaspillée* si  $w(1)\dots w(i)$  contient le même nombre de permutations distinctes en facteurs que  $w(1)\dots w(i-1)$  (c'est-à-dire si  $w(i-n+1)\dots w(i)$  n'est pas une permutation, qu'elle est apparue avant, ou que toutes ses lettres ne sont pas définies).

On a alors  $|w| = n! + |\{i \mid w(i) \text{ est gaspillée}\}|$  par définition. Dans notre cas  $n = 3$  : les 2 premières lettres sont forcément gaspillées.  $3! = 6$ , donc si 9 n'est pas optimal, il n'y a pas d'autres lettres gaspillées. Alors  $w(1)w(2)w(3)$ ,  $w(2)w(3)w(4)$ ,  $w(3)w(4)w(5)$  et  $w(4)w(5)w(6)$  doivent être des permutations distinctes. Mais cela implique  $w(4) = w(1)$ ,  $w(5) = w(2)$  et  $w(6) = w(3)$  donc  $w(1)w(2)w(3) = w(4)w(5)w(6)$  ce qui est contradictoire, d'où le résultat.

Pour  $n = 4$  : notre borne inférieure dit que  $|w| \geq 24 + 4 - 1 = 27$ . Voici une 4-superpermutation de taille 33 :  $w = 123412314231243121342132413214321$  (on peut vérifier que c'est minimal par ordinateur).

## 2.2 Construction standard et première conjecture

Ashlock et Tillotson [2] ont remarqué en 1993 que les longueurs minimales étaient  $1 = 1!$ ,  $3 = 2! + 1!$ ,  $9 = 3! + 2! + 1!$  et  $33 = 4! + 3! + 2! + 1!$  pour les premières valeurs de  $n$ .

Ils ont alors montré :

**Proposition 2.1.** *Si il existe une  $(n-1)$ -superpermutation de longueur  $m$ , alors il existe une  $n$ -superpermutation de longueur  $m + n!$*

Description de l'algorithme à partir d'une  $(n-1)$ -superpermutation de longueur  $m$  :

- Écrire les permutations dans l'ordre d'apparition dans la  $(n-1)$ -superpermutation
- Dupliquer chacune en ajoutant  $n$  entre les deux copies
- Concaténer les parties en éliminant les chevauchements

On obtient un mot  $w$  qui est une  $n$ -superpermutation de longueur  $m + n!$

Par exemple, pour passer de  $n = 3$  à  $n = 4$  :

123121321  
 123 | 231 | 312 | 213 | 132 | 321  
 123 4 123 | 231 4 231 | 312 4 312 | 213 4 213 | 132 4 132 | 321 4 321  
 123412314231243121342132413214321

On déduit alors facilement de la Proposition 2.1 (par récurrence) une borne supérieure :  $\text{Opt}(n) \leq n! + \dots + 2! + 1! = B_{\text{sup},1}(n)$ . Ashlock et Tillotson ont conjecturé que cette borne supérieure était optimale, et même que l'on avait unicité de la  $n$ -superpermutation atteignant cette borne (à un renommage des symboles près).

Pendant 20 ans, il n'y a pas eu d'avancées sur ce problème. Mais en 2013, Johnston [8] construit plusieurs  $n$ -superpermutations de taille  $n! + \dots + 2! + 1!$  pour  $n \geq 5$ . Ainsi, il montre qu'au moins l'une des deux conjectures d'Ashlock et Tillotson est fausse.

En 2014, Chaffin [9] vérifie le cas  $n = 5$  : il trouve les huit 5-superpermutations de longueur  $B_{\text{sup},1}(5) = 153$  et montre qu'il n'y a pas plus petit. Chaffin recherche algorithmiquement par un parcours en profondeur de la façon suivante :

- On voit facilement que l'on peut avoir au maximum 5 permutations distinctes dans un mot sans lettre gaspillée (les différentes rotations de la permutation de départ).
- On recherche en énumérant toutes les possibilités le nombre maximal de permutations distinctes dans un mot avec une lettre gaspillée (on trouve 10 avec 123451234152341)
- On recherche le nombre maximal de permutations distinctes dans un mot en autorisant 2 lettres gaspillées. Pour accélérer la recherche, on utilise les résultats précédents : après avoir trouvé un mot avec  $p$  permutations, si un mot a un gaspillage avant d'avoir  $p - 10$  permutations distinctes, on sait que l'on n'atteindra pas  $p$  donc on peut abandonner ce mot.
- On répète ceci en utilisant tous les résultats précédents

Ce système permet d'accélérer fortement les calculs, pour énumérer la totalité des résultats qui nous intéressent. On obtient ainsi qu'il faut 29 lettres gaspillées,  $\text{Opt}(5) = 153$  et il y a huit superpermutations de cette longueur. La conjecture d'unicité de la solution optimale est donc fausse, mais on ne sait toujours pas si la borne est optimale pour  $n > 5$ .

### 2.3 Réfutation de la conjecture $\text{Opt}(n) = B_{\text{sup},1}(n)$

Quelques semaines plus tard seulement, Houston [6] trouve une 6-superpermutation de longueur 872 (alors que  $B_{\text{sup},1}(6) = 873$ ) et réfute donc la conjecture de la Proposition 2.1 pour  $n \geq 6$  (le cas  $n=6$  implique les autres par la construction récursive).

Pour cela, il relie le problème à celui bien connu du *voyageur de commerce* (TSP) (où l'on recherche un chemin hamiltonien et non un cycle) : on considère un graphe complet  $X_n$  où les sommets sont les permutations de  $n$  éléments (donc  $n!$  sommets). Le poids d'une arête de  $\sigma_1$  à  $\sigma_2$  est le nombre de lettres de  $u$ , le plus petit mot tel que  $\sigma_2$  soit suffixe de  $\sigma_1 u$ . Par exemple : pour  $n=3$ , l'arête de 231 vers 312 a un poids 1 et celle de 231 vers 132 a un poids 2. Pour simplifier le graphe, on va supprimer les arêtes *impropres* :

**Définition 3.** Une arête *impropre* est une arête qui fait passer par une autre permutation, c'est-à-dire, si l'arête est  $(\sigma_1, \sigma_2)$  et  $u$  le mot minimal pour avoir  $\sigma_2$  en suffixe, alors  $\sigma_1 u$  contient une permutation autre que  $\sigma_1$  et  $\sigma_2$  en facteur. Par exemple l'arête entre 12345 et 34512 est impropre car on rajoute 2 lettres, mais en prenant seulement la première on arrive sur 23451 qui est un sommet.  $G_n$  est le graphe  $X_n$  auquel on a enlevé toutes les arêtes impropres (voir Fig. 1).

En effectuant ceci, chaque sommet a alors exactement une arête sortante de poids 1, une arête entrante de poids 1, une arête sortante de poids 2 et une arête entrante de poids 2, ainsi que d'autres arêtes de poids supérieur que l'on ne cherchera pas à compter. En effet pour le poids 1, si on veut ajouter une lettre à la fin et obtenir une permutation, on doit forcément rajouter la première car le

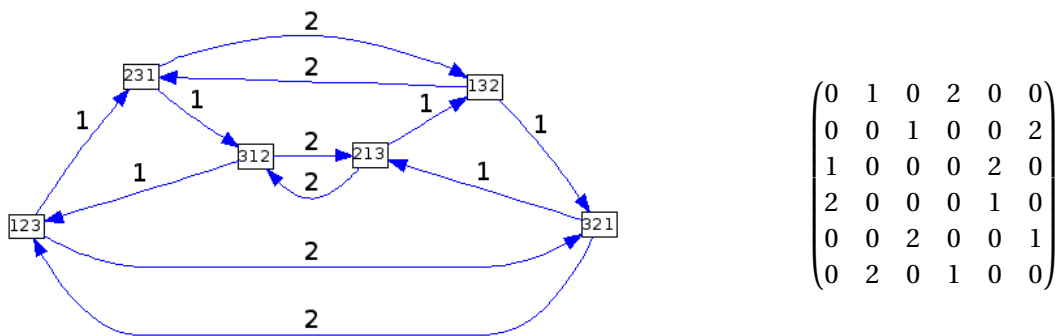


FIGURE 1 – Graphe des permutations  $G_3$  et sa matrice d'adjacence

suffixe doit être injectif. Pour le poids 2, on a deux possibilités, rajouter  $x(1)x(2)$  ou  $x(2)x(1)$  mais le premier cas donne une arête impropre.

Alors si l'on considère la solution du TSP sur  $G_n$ , c'est un chemin hamiltonien de poids minimal : or un tel chemin passe par toutes les permutations. Si l'on écrit les sommets dans l'ordre en supprimant les caractères en trop pour obtenir une superpermutation, alors la longueur de cette superpermutation est exactement le poids du chemin hamiltonien plus les  $n$  caractères du sommet initial. Cependant, étant donné une  $n$ -superpermutation  $w$ , on construit un chemin de poids  $|w| - n$  passant une fois par tous les sommets, mais pas forcément une seule fois. Donc les méthodes de résolution du TSP sont intéressantes mais a priori pas suffisantes pour trouver la longueur optimale.

Donnons un exemple du lien entre les chemins dans le graphe et les superpermutations :

- $123 \rightarrow 231 \rightarrow 312 \rightarrow 213 \rightarrow 132 \rightarrow 321$  est un chemin hamiltonien dans le graphe, de poids 6
- Si on suit la construction, on obtient la superpermutation  $123121321$  de longueur  $9=6+3$

Cependant le TSP est NP-difficile : il n'y a pas d'algorithme efficace. Mais on connaît certains algorithmes d'approximation. Houston[6] a trouvé un résultat de longueur 872, sans toutefois savoir s'il est optimal.

Mais en dehors de la réfutation, il n'y a pas d'avancée sur les bornes. Pour  $n = 6$ , on ne sait même pas si 872 est la solution optimale ! Actuellement, de nombreux efforts sont faits pour essayer de trouver la solution optimale pour  $n = 6$ , notamment un algorithme distribué par Egan qui utilise la méthode de Chaffin permettant à n'importe qui de prêter la puissance de calcul de son ordinateur pour aider à la recherche, mais cela est encore loin d'aboutir.

## 2.4 Une meilleure borne inférieure

Pourtant en 2011, un utilisateur anonyme de 4chan [1] a trouvé un début de réponse : en effet, un sujet demande combien d'épisodes il faut regarder pour avoir vu les 14 épisodes de *The Melancholy of Haruhi Suzumiya* dans tous les ordres possibles. Il s'agit donc effectivement de trouver une 14-superpermutation de longueur minimale.

Un utilisateur trouve rapidement une borne inférieure (pour un nombre quelconque d'épisodes) meilleure que celles connues. Cependant les mathématiciens ne s'y intéressent pas, car ils pensent que la conjecture d'Ashlock et Tillotson est vraie.

Mais les réfutations récentes apportent un nouvel intérêt à cette preuve : Houston, Pantone et Vatter rédigent un article mettant en forme les idées de cet internaute.

**Théorème 2.2.** *Toute  $n$ -superpermutation a pour longueur au moins  $n! + (n-1)! + (n-2)! + n - 3 = B_{\text{inf},2}(n)$ .*

Pour prouver le théorème, étudions d'abord différents cycles de notre graphe :

- Chaque sommet  $x$  a exactement une arête sortante de poids 1 qui va de  $x = x(1)\dots x(n)$  vers  $x(2)\dots x(n)x(1)$ . Si on suit  $n-1$  fois ces arêtes, on visite ce que l'on appelle la classe cyclique de

$x$ , c'est-à-dire l'ensemble des rotations de  $x$ . On définit alors  $\mathcal{R}$  une relation d'équivalence telle que  $x\mathcal{R}y$  si et seulement si  $x$  et  $y$  sont dans la même classe cyclique. On notera  $\text{cl}(x) = \bar{x} = \text{rot}(x)$ . On a  $(n-1)!$  classes cycliques distinctes et dans chaque chemin passant par tous les sommets, on visite toutes ces classes et tous leurs éléments.

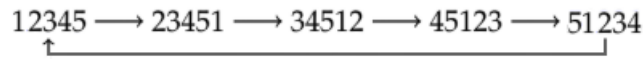


FIGURE 2 – Un exemple de 1-cycle :  $\text{rot}(12345)$

- Chaque sommet a aussi exactement une arête sortante de poids 2 qui va de  $x = x(1)\dots x(n)$  vers  $x(3)\dots x(n)x(2)x(1)$  (celle vers  $x(3)\dots x(n)x(1)x(2)$  est impropre donc on l'a supprimée). On définit le 2-cycle généré par  $x$  comme l'ensemble des sommets visités en suivant d'abord l'arête de poids 2, puis  $n-1$  arêtes de poids 1, et en répétant ceci  $n-1$  fois.

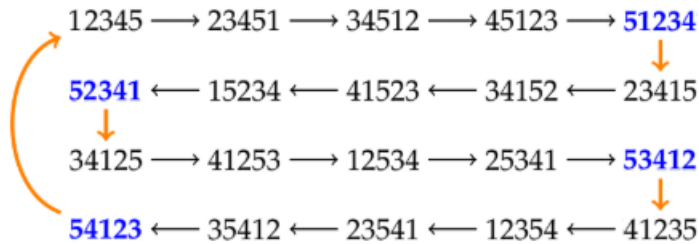


FIGURE 3 – 2-cycle engendré par 51234. Les sommets en bleu sont les générateurs, les arêtes noires ont pour poids 1 et les oranges de 2.

On constate que le 2-cycle généré par  $x$  est aussi généré par les permutations cycliques des  $n-1$  derniers éléments de  $x$ , précédées de  $x(1)$ , soit  $\{x(1)q \mid q \in \text{rot}(x(2)\dots x(n))\}$ . On a  $(n-2)!$  2-cycles distincts. On dit qu'un chemin *entre dans un 2-cycle* généré par  $x$  si on arrive sur  $x(2)\dots x(n)x(1)$  par une arête de poids au moins 2. Comme chaque 2-cycle est constitué des classes cycliques de ses générateurs, donc  $(n-1)$  éléments, un chemin visitant chaque permutation doit entrer par au moins  $(n-2)!$  2-cycles différents.

Avec ces outils, prouvons le Théorème 2.2 :

*Démonstration.* Soit  $\mathcal{C} = (x_1, \dots, x_m)$  un chemin dans le graphe. On définit :

- $w(\mathcal{C}) = \sum_{k=1}^{m-1} w(x_k, x_{k+1})$  la longueur du chemin
- $p(\mathcal{C})$  le nombre de permutations distinctes visitées
- $c(\mathcal{C})$  le nombre de classes cycliques complètes visitées dans le chemin  $x_1 \dots x_{m-1}$
- $v(\mathcal{C})$  le nombre de 2-cycles *visités* (si on est entré au moins une fois dans ce 2-cycle)

Montrons l'inégalité  $w(\mathcal{C}) \geq p(\mathcal{C}) + c(\mathcal{C}) + v(\mathcal{C}) - 2$  par récurrence sur  $m$ .

C'est vrai si  $m = 1$  car  $w(\mathcal{C}) = 0, p(\mathcal{C}) = 1, c(\mathcal{C}) = 0, v(\mathcal{C}) = 1$

Si c'est vrai pour  $m$ , soit le chemin  $\tilde{\mathcal{C}} = (x_1, \dots, x_m, x_{m+1})$  et  $\mathcal{C} = (x_1 \dots x_m)$  pour un certain  $x_{m+1}$ . Remarquons que  $p, c$  et  $v$  augmentent de 0 ou de 1 en ajoutant un nouveau sommet à la fin du chemin. On a  $w(\tilde{\mathcal{C}}) = w(\mathcal{C}) + w(x_m, x_{m+1})$ . On distingue les cas :

- Si  $w(x_m, x_{m+1}) = 1$  :  $x_m$  et  $x_{m+1}$  sont dans la même classe cyclique. Deux cas :
  - Soit on a visité  $x_{m+1}$  avant, alors  $p$  n'augmente pas,  $v$  n'augmente pas (par définition, on entre seulement avec poids au moins 2 dans un 2-cycle),  $c$  peut augmenter de 1 et  $w$  augmente de 1 : l'inégalité est conservée

- Soit on n'avait pas visité  $x_{m+1}$ , alors  $p$  augmente de 1, mais  $x_m$  ne peut avoir complété la classe (car  $x_{m+1}$  est dans la classe mais n'est pas visité) donc  $c$  n'augmente pas,  $v$  non plus : l'inégalité est conservée
- Si  $(x_m, x_{m+1})$  a pour poids 2, par unicité de l'arête,  $x_{m+1} = x_m(3)...x_m(n)x_m(2)x_m(1)$  et on entre dans un 2-cycle que l'on note  $K$ 
  - Si  $c$  augmente, cela signifie que  $x_m$  complète une classe cyclique donc n'avait pas été visité dans les  $m - 1$  premiers sommets. On a visité  $x_m(2)...x_m(n)x_m(1)$  qui est dans la même classe, mais donc pas depuis  $x_m$  qui n'avait pas été visité, donc par une arête de poids au moins 2, donc on était déjà entré dans  $K$ ,  $v$  n'augmente pas :  $p$  peut augmenter de 1 au plus, donc l'inégalité reste vraie
  - Si  $c$  n'augmente pas,  $p$  et  $v$  ne peuvent augmenter que d'un donc l'inégalité est encore vraie
- Si  $(x_m, x_{m+1})$  a pour poids 3 ou plus : la résultat est immédiat car en rajoutant une arête,  $p, c, v$  ne peuvent augmenter que d'un donc à droite la valeur augmente de 3 max, et à gauche elle augmente de 3 : l'inégalité est conservée.

Maintenant que l'on a cette inégalité : si  $\mathcal{C}$  visite au moins une fois chaque sommet, alors  $p(\mathcal{C}) \geq n!$ ,  $c(\mathcal{C}) \geq (n - 1)! - 1$  et  $v(\mathcal{C}) \geq (n - 2)!$  donc  $w(\mathcal{C}) \geq n! + (n - 1)! + (n - 2)! - 3$ . Soit maintenant une  $n$ -superpermutation  $SP$ , on a vu que l'on pouvait en tirer un chemin  $\mathcal{C}$  passant au moins une fois par chaque sommet tel que  $w(\mathcal{C}) = |SP| - n$  donc  $|SP| = w(\mathcal{C}) + n \geq n! + (n - 1)! + (n - 2)! + n - 3$ .  $\square$

### 3 Amélioration de la borne supérieure

#### 3.1 Une nouvelle borne supérieure

En s'inspirant de cette méthode et en adaptant les travaux de Williams [11, 10] sur l'hamiltonicité du groupe de Cayley de  $S_n$  par rapport à certaines permutations, l'auteur Greg Egan [4] donne une nouvelle borne supérieure.

**Théorème 3.1.** *Pour  $n \geq 4$ , il existe une  $n$ -superpermutation de longueur au plus  $n! + (n - 1)! + (n - 2)! + (n - 3)! + n - 3 = B_{\text{sup},2}(n)$ .*

Pour trouver cette borne, Egan construit un chemin hamiltonien spécifique dont on connaît le poids exact et n'utilisant que des arêtes de poids 1 et 2. Notons que passer par une arête de poids 1, c'est faire une rotation à gauche, donc multiplier à droite par  $\sigma = 234...n1$  (l'ensemble des  $x\sigma^j$  définit la classe cyclique de  $x$ , c'est à dire la classe d'équivalence de  $x$  pour la relation  $\mathcal{R}$  définie plus haut). De même, passer par une arête de poids 2 revient à multiplier à droite par  $\delta = 345...n21$ . Williams prouve l'existence avec  $\sigma$  et  $\tau = 21345$ . Pour les superpermutations, Egan adapte la méthode de Williams pour  $\sigma$  et  $\delta$ .

On va donc travailler sur le graphe  $\tilde{G}_n = \text{Cayley}(S_n, \{\sigma, \delta\})$  (car  $\sigma$  et  $\delta$  génèrent  $S_n$ ) qui est  $G_n$  restreint aux arêtes de poids 1 et 2. La méthode consiste à trouver une couverture en deux cycles du graphe, qui peuvent ensuite être coupés et recollés en un chemin qui visite chaque sommet une fois.

**Définition 4.** Une *couverture en cycles* est un ensemble disjoint de cycles dans le graphe  $\tilde{G}_n$  tel que chaque sommet appartienne à exactement un cycle. Chaque sommet à exactement une arête qui arrive et une qui ressort :  $\deg^+(x) = 1 = \deg^-(x)$ . Par exemple, les  $(n - 1)!$  classes cycliques forment une couverture en cycles.

Pour trouver une couverture en deux cycles, on va réutiliser la notion de 2-cycles : on a vu qu'un 2-cycle engendré par une permutation quelconque  $p$  s'obtenait en prenant la permutation  $p$ , puis en

la multipliant par  $\delta$ , puis  $n - 1$  fois par  $\sigma$ , et répétant cela  $n - 1$  fois. Alors si  $p = mq$  avec  $m$  une lettre, les sommets obtenus sont

$$\text{rot}(m\text{rot}(q)) = \bigcup_{q' \in \text{rot}(q)} \text{rot}(mq')$$

et on peut noter  $m|q$  pour le 2-cycle. Par exemple, le 2-cycle de la Fig. 3 se note  $5|1234$ . Remarquons que  $\forall q' \in \text{rot}(q)$ ,  $m|q = m|q'$  et que les sommets  $m\text{rot}(q)$  sont ceux suivis par une  $\delta$ -arête.

Soient  $m_1|q_1$  et  $m_2|q_2$  deux 2-cycles distincts. Quelles sont les conditions pour avoir une intersection?

- Si  $m_1 = m_2 = m$ , l'intersection est vide car comme ils ne sont pas égaux,  $\text{rot}(q_1) \cap \text{rot}(q_2) = \emptyset$  donc  $\text{rot}(m\text{rot}(q_1)) \cap \text{rot}(m\text{rot}(q_2)) = \emptyset$
- Si  $m_1 \neq m_2$ , en prenant le mot commençant par  $m_2$  dans  $\text{rot}(q_1)$  on peut écrire  $m_1|q_1 = m_1|m_2r_1$ . De même,  $m_2|q_2 = m_2|m_1r_2$ .
  - Si  $r_1 = r_2$ , l'intersection contient deux 1-cycles :  $\text{rot}(m_1m_2r)$  et  $\text{rot}(m_2m_1r)$
  - Si  $\text{rot}(r_1) = \text{rot}(r_2)$  mais  $r_1 \neq r_2$ , l'intersection est réduite à un 1-cycle. Plus précisément si  $r_1 = ab$  et  $r_2 = ba$  avec  $a, b$  deux mots, on a  $\text{rot}(m_1bm_2a)$  dans l'intersection
  - Sinon, l'intersection est vide

On va s'intéresser à un sous-ensemble particulier de l'ensemble des 2-cycles. Donnons maintenant une autre définition caractérisant les 2-cycles de ce sous-ensemble, qui nous permettra de les manipuler plus facilement et de leur associer un générateur canonique.

**Définition 5.** On dit que  $s = s_1 \dots s_{n-1}$  est une seed si  $s$  est un mot injectif de longueur  $n - 1$  sur  $[n]$ ,  $s_1 = n$  et  $x$  la lettre manquante de  $[n]$  est  $x = (s_2 \bmod (n - 1)) + 1$ .

Si  $s$  est une seed, on définit  $\text{flower}(s) = \{xq|q \in \text{rot}(s)\}$  et  $\text{perm}(s) = \bigcup_{\pi \in \text{flower}(s)} \text{rot}(\pi)$ . Si  $S$  est un ensemble de seeds,  $\text{perm}(S) = \bigcup_{s \in S} \text{perm}(s)$ .

En fait  $\text{perm}(s) = x|s$ , donc  $|\text{perm}(s)| = n(n - 1)$  : c'est le 2-cycle engendré par  $xs$ . Dans  $\text{perm}(s)$ , il n'y a qu'une seule seed : c'est le générateur canonique de ces 2-cycles. Notons que l'on a  $(n - 1)(n - 3)!$  seeds donc on considère un ensemble d'autant de 2-cycles associés, l'ensemble  $E$  des  $m|nrq$  tels que  $m = r + 1$  et  $mnrq \in S_n$ . On peut définir une bijection de  $E$  dans les seeds, avec  $\text{seed}(m|nrq) = nrq$ . Ainsi on pourra appliquer toute fonction sur une seed ou un 2-cycle correspondant sans changer de notation.

Remarque : dans la suite, toute l'arithmétique se fera modulo  $(n - 1)$  pour simplifier les notations. Ainsi on dira simplement que la lettre manquante est  $x = s_2 + 1$ .

Notons le résultat suivant :

**Lemme 3.2.** Si  $S$  est l'ensemble de toutes les seeds,  $\text{perm}(S) = S_n$

**Définition 6.** Si  $U \subset S(G)$  l'ensemble des sommets d'un graphe  $G$ , on notera  $G[U]$  le graphe induit par la restriction de  $S$  à  $U$ , c'est-à-dire  $(U, A \cap U \times U)$

**Définition 7.** Soit  $s$  une seed de lettre manquante  $x$ . Soit  $\rho = \sigma^{-1}$  la rotation à droite. On définit pour  $1 \leq j \leq n - 1$ ,  $\pi_j(s) = x\rho^j(s)$ , que l'on note simplement  $\pi_j$  s'il n'y a pas de confusion possible.

Remarque :  $\pi_{n-1} = xs$

**Proposition 3.3.** Une règle de succession sur un graphe  $G$  est une fonction qui à un sommet  $x$  du graphe associe un sommet  $y$  tel que l'arête  $(x, y)$  existe.

On définit la règle de succession suivante sur  $G_n[\text{perm}(s)]$  :

$$f_s(\pi) =: \begin{cases} \delta(\pi) & \text{si } \pi = \pi_j \text{ pour un certain } j \\ \sigma(\pi) & \text{sinon} \end{cases}$$

Alors les applications successives de  $f_s$  construisent un cycle hamiltonien  $\text{Ham}(s)$  dans  $G_n[\text{perm}(s)]$  qui utilise exactement  $(n-1)$   $\delta$ -arêtes.

$$\begin{aligned} \text{Ham}(s) = & \sigma(\pi_{n-1}) \xrightarrow{\sigma} \sigma^2(\pi_{n-1}) \xrightarrow{\sigma} \dots \xrightarrow{\sigma} \sigma^{n-1}(\pi_{n-1}) \xrightarrow{\sigma} \pi_{n-2} \xrightarrow{\delta} \\ & \sigma(\pi_{n-2}) \xrightarrow{\sigma} \sigma^2(\pi_{n-2}) \xrightarrow{\sigma} \dots \xrightarrow{\sigma} \sigma^{n-1}(\pi_{n-2}) \xrightarrow{\sigma} \pi_{n-1} \xrightarrow{\delta} \\ & \vdots \\ & \sigma(\pi_1) \xrightarrow{\sigma} \sigma^2(\pi_1) \xrightarrow{\sigma} \dots \xrightarrow{\sigma} \sigma^1(\pi_1) \xrightarrow{\sigma} \pi_1 \xrightarrow{\delta} \end{aligned}$$

Remarquons que si  $\pi_j = xq_1\dots q_{n-1}$ , alors  $\pi_{j-1} = xq_2\dots q_{n-1}q_1$  donc  $\delta(\pi_j) = \sigma(\pi_{j-1})$ .

On va maintenant définir une structure arborescente sur les cycles :

**Définition 8.** Soit  $s$  une seed,  $\text{parent}(s)$  est la seed obtenu en enlevant  $x+1$  à  $s_1xs_2\dots s_{n-1}$

On définit aussi  $\text{ds}(s)$  la sous-séquence décroissante de  $s$  : c'est le plus long préfixe de  $(s_2-1)(s_2-2)\dots(s_2-(n-3))$  qui apparait dans  $s$  en tant que sous-séquence.

On note alors  $\text{level}(s) = (n-3) - |\text{ds}(s)|$

Enfin, le *hub* est l'ensemble des  $n-1$  seeds de niveau 0.

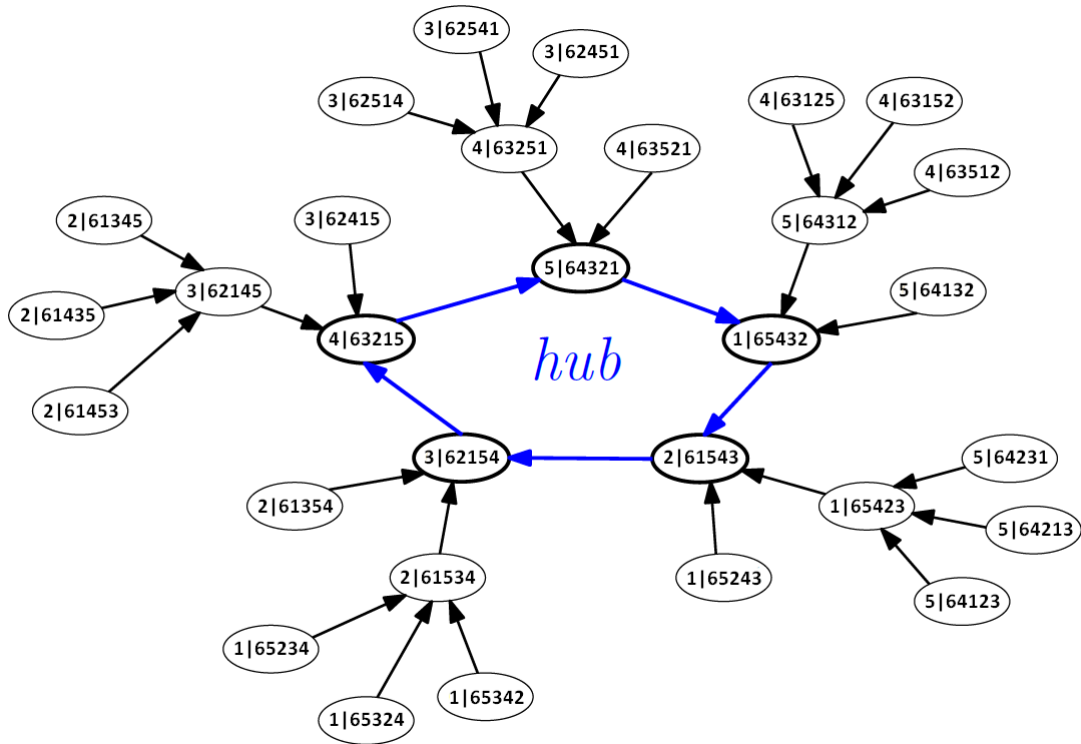


FIGURE 4 – Structure obtenue pour  $n = 6$  : on représente chacune des 30 seeds par le 2-cycle associé. Chaque flèche va d'une seed vers son parent. On a le hub qui forme un polygone au centre, et sur chacune des seeds du hub, un arbre.

**Lemme 3.4.** Soit  $s$  une seed,  $l = \text{level}(s)$ . Si  $l > 0$ , alors  $\text{level}(\text{parent}(s)) = l - 1$

**Lemme 3.5.** Soit  $s \neq \tilde{s}$  deux seeds, alors  $\text{perm}(s) \cap \text{perm}(\tilde{s}) = \emptyset \iff \tilde{s} \neq \text{parent}(s)$  et  $s \neq \text{parent}(\tilde{s})$ . Si  $\tilde{s} = \text{parent}(s)$ , l'intersection est un 1-cycle.

*Démonstration.* On suppose que l'intersection est non vide, en utilisant les conditions d'intersection et les propriétés des seeds, on obtient le résultat, est l'intersection sera  $\text{rot}(xb\bar{x}a) = \text{rot}(x(x-1)b'(x+1)a'n)$  si  $\text{perm}(s) = x|xa'n(x-1)b'$ .  $\square$



**Définition 9.** Soit  $S$  un ensemble de permutations stable par  $\sigma$ . On définit la règle de succession sur  $G_n[S]$  par :

$$\text{next}_S(\pi) =: \begin{cases} \delta(\pi) & \text{si } (r, p_1) \in \{(1,2), \dots, (n-2, n-1), (n-1, 1)\} \text{ et } \delta(\pi) \in S \\ \sigma(\pi) & \text{sinon} \end{cases}$$

où  $\pi = p_1 \dots p_n$  et  $r$  est la lettre à droite de  $n$  dans  $\pi$ .

On peut remarquer que  $\text{next}_{\text{perm}(s)} = f_s$ . On notera  $\text{next}_{S_n} = \text{next}$ .

**Lemme 3.6.** Il y a  $(n-1)^2(n-3)!$  permutation  $\pi$  telles que  $\text{next}(\pi) = \delta(\pi)$

**Lemme 3.7.** Si  $\text{next}(\pi) = \delta(\pi)$ , il existe une unique seed  $s$  telle que  $\pi$  et  $\delta(\pi)$  sont dans  $\text{perm}(s)$ .

*Démonstration.*  $\pi$  et  $\delta(\pi)$  sont dans deux 1-cycles distincts donc le Lemme 3.5 donne l'unicité. En comptant les arêtes de poids 2 on a l'existence.  $\square$

Le théorème suivant est le résultat technique principal de cette preuve. On va le démontrer en deux lemmes.

**Théorème 3.8.** Si l'on considère l'ensemble des  $(\pi, \text{next}(\pi))$  pour  $\pi \in S_n$ , on obtient  $C(n)$  une couverture en deux cycles.

**Lemme 3.9.**  $\text{next}_{\text{perm}(\text{hub})}$  partitionne  $G_n[\text{perm}(\text{hub})]$  en deux cycles. Plus précisément, si on ordonne les  $n-1$  seeds du hub de la façon suivante :

$$\begin{aligned} h_1 &= n(n-1)\dots 32 \\ h_2 &= n1(n-1)\dots 3 \\ &\vdots \\ h_{n-1} &= n(n-2)\dots 21 \end{aligned}$$

alors  $\text{parent}(h_j) = h_{j+1}$ . Pour  $1 \leq j \leq n-1$ , on considère  $q_j = \pi_{n-1}(h_j)$  et pour  $1 \leq i \leq n-3$ ,  $q_j^{(i)} = \sigma^{n-i}(q_j)$ .

Soit  $Q = \cup_j \{q_j, q_j^{(1)}, \dots, q_j^{(n-3)}\}$ . Alors on obtient un cycle pour  $Q$  (de taille  $(n-1)(n-2)$ ) et un cycle pour  $\text{perm}(\text{hub}) \setminus Q$ .

*Démonstration.* Notons  $S = \text{perm}(\text{hub})$ .

Commençons par  $Q$  :

On constate que  $\text{next}_S(q_j) = q_{j+1}^{(n-3)}$ ,  $\text{next}_S(q_j^{(i)}) = \sigma(q_j^{(i)}) = \sigma^{n-i+1}(q_j) = q_j^{(i-1)}$  si  $i > 1$  et  $\text{next}_S(q_j^{(1)}) = \sigma(q_j^{(1)}) = q_j$  ce qui montre qu'en appliquant successivement  $\text{next}_S$  à un sommet de  $Q$ , on obtient un cycle qui parcourt tout  $Q$ .

Montrons maintenant que  $\text{next}_S$  forme un cycle dans  $S \setminus Q$  :

Soit  $\pi$  dans  $S$ , quand est-ce que  $\text{next}_S(\pi)$  vaut  $\delta(\pi)$ ? Si c'est le cas, on aura aussi  $\text{next}(\pi) = \delta(\pi)$ . Par les lemmes 3.5 et 3.7, on peut trouver une seed du hub  $h_j$  et  $i$  tels que  $\pi = \pi_i(h_j)$ . De plus on peut aussi remarquer que si l'on prend  $\pi$  dans  $\text{perm}(h_k)$  qui n'est pas un  $\pi_i(h_k)$ , il s'écrit  $\pi_i(h_j)$  ssi  $\pi = \sigma^2(\pi_{n-1}(h_j))$  ou  $\pi = \sigma^{n-2}(\pi_{n-2}(h_j))$ . En considérant aussi les  $Q \cap (j|h_j)$ , on voit que les éléments tels que  $\text{next}_S(\pi) \neq f_{h_j}(\pi)$  sont les  $\sigma^2(\pi_{n-1}(h_j)) = \pi_{n-1}(h_{j-1})$ . On vérifie alors qu'en suivant  $\text{next}$ , on suit les  $\text{Ham}(h_j)$  et on passe à  $\text{Ham}(h_{j-1})$  en arrivant à la permutation  $\sigma^2(\pi_{n-1}(h_j)) = \pi_{n-1}(h_{j-1})$ , donc on forme un cycle.  $\square$

Maintenant, il faut montrer que l'on peut ajouter les sommets des 2-cycles qui ne sont pas de le hub sans changer le nombre de cycles :

**Lemme 3.10.** Soit  $m = (n-1)(n-3)!$ . Soit  $s_1, \dots, s_m$  un ordre sur les seeds respectant le niveau, c'est à dire :  $\forall j \geq i, \text{level}(s_j) \geq \text{level}(s_i)$ . Alors si  $S_j = \text{perm}(s_1, \dots, s_j)$  pour  $n-1 \leq j \leq m$ ,  $\text{next}_{S_j}$  partitionne  $G_n[S_j]$  en deux cycles (un pour  $Q$  et l'autre pour  $S_j \setminus Q$ )

*Démonstration.* On procède par récurrence sur  $j$ .

Dans le lemme précédent, on a fait le cas  $j = n-1$  (car  $\text{hub} = \{s_1, \dots, s_{n-1}\}$  peu importe l'ordre choisi qui respecte le niveau)

Si  $\text{next}_{S_j}$  partitionne  $G_n[S_j]$  en deux cycles. Soit  $S = \text{perm}(s_{j+1})$ ,  $\text{next}_S = f_S$  construit  $\text{Ham}(s_{j+1})$ . Par les lemmes 3.4 et 3.5,  $\exists! i \leq j$  tel que  $s_i = \text{parent}(s_j)$  et  $\text{perm}(s_i) \cap \text{perm}(s_{j+1}) = \text{rot}(\bar{\pi})$ . Alors  $\text{perm}(s_j)$  va s'insérer dans le cycle déjà construit en coupant le 1-cycle commun.  $\square$

On a la couverture de  $G_n$  car pour  $j = m$ , on obtient  $\text{next}$ . Maintenant que l'on a obtenu une couverture en deux cycles de  $G_n$ , on peut en déduire un chemin hamiltonien.

**Théorème 3.11.** A partir de  $C(n)$ , on peut obtenir un chemin hamiltonien.

*Démonstration.* Dans les faits, on définit une nouvelle règle de succession telle que  $\delta^{-1}(\sigma(1n(n-1)\dots 32))$  n'est pas dans l'ensemble de définition et pour les autres permutations :

$$f(\pi) =: \begin{cases} \sigma(\pi) & \text{si } \pi = 1n(n-1)\dots 32 \\ \text{next}(\pi) & \text{sinon} \end{cases}$$

$q = 1n(n-1)\dots 32$  est un élément de  $Q$  tel que  $\text{next}(q) = \delta(q)$ . Donc dans  $C(n)$ , le petit cycle va de  $\delta(q)$  vers  $q$ , puis retourne à  $\delta(q)$  par  $e$  de poids 2.  $\sigma(q)$  est dans  $S \setminus Q$ , donc le grand cycle va de  $\sigma(q)$  vers  $\delta^{-1}(\sigma(q))$  puis retourne à  $\sigma(q)$  par  $f$  de poids 2. On supprime  $e$  et  $f$ , mais on rajoute  $g$  de poids 1 entre  $q$  et  $\sigma(q)$ . Donc on obtient un chemin hamiltonien : on part de  $\delta(q)$  pour arriver à  $q$ , puis  $\sigma(q)$ , puis  $\delta^{-1}(\sigma(q))$ .  $\square$

*Preuve de la borne supérieure.* Soit  $\tilde{C}$  le chemin hamiltonien obtenu. On peut en tirer une superpermutation  $SP$  telle que  $|SP| = w(\tilde{C})$ . D'après la preuve du Théorème 3.11, on supprime  $e$  et  $f$  de poids 2 et on rajoute  $g$  de poids 1 donc  $w(\tilde{C}) = w(C) - 3$  où  $C$  est la couverture en deux cycles du Théorème 3.8. Or dans une couverture, chaque sommet a un degré entrant égal à 1, donc on a exactement  $n!$  arêtes dans la couverture. Or par le Lemme 3.6, on a  $(n-1)^2(n-3)!$  arêtes de poids 2. Donc  $w(C) = n! - (n-1)^2(n-3)! + 2(n-1)^2(n-3)! = n! + (n-1)^2(n-3)! = n! + [(n-1)(n-2) + (n-2) + 1](n-3)! = n! + (n-1)! + (n-2)! + (n-3)!$ .  $\square$

## 3.2 Etat actuel

Notons que  $B_{\text{sup},2}(6) = 873$ , ce qui est plus grand que le résultat de Houston cité plus haut. Il a été conjecturé que la solution optimale pouvait être  $B_{\text{sup},2}(n) - 1$ , mais pour  $n = 7$ ,  $B_{\text{sup},2}(7) = 5908$ ,  $B_{\text{inf},2}(7) = 5884$ . Or début 2019, Coanda trouve par ordinateur un exemple de longueur 5907 raffiné par Houston en 5906.

La question de trouver la longueur minimale reste donc ouverte malgré toutes ces avancées.

On a cependant la proposition suivante :

**Proposition 3.12.** La construction de Egan ( $B_{\text{sup},2}(n)$ ) est optimale si l'on autorise seulement des arêtes de poids 1 ou 2.

*Démonstration.* On utilise l'inégalité de la preuve du Théorème 2.2. Il s'agit de remarquer que le nombre de classes cyclique visitées est majoré en fonction du nombre de 2-cycles visités, car en se restreignant aux arêtes de poids 1 et 2, on change de 2-cycle par une arête de poids 2. Alors, les 2-cycles consécutifs ont une classe cyclique en commun, donc on ne peut en visiter que  $n-1$  par 2-cycle. Comme on les visite tous, cela donne une minoration du nombre de 2-cycles visités, ce qui donne  $w(\mathcal{C}) \geq n! + (n-1)! + (n-2)! + (n-3)! - 3$  en injectant cette minoration dans l'inégalité. Alors on obtient bien que la superpermutation résultante a une longueur plus grande que  $B_{\text{sup},2}(n)$ .  $\square$

## 4 Matrices de superpermutation

Enfin, la dernière partie de mon stage a été consacré à étudier un problème connexe. En effet, je me suis demandé si l'on pouvait rechercher la taille minimale d'une matrice qui contiendrait toute matrice de permutation en tant que bloc contigu. Les différents résultats que j'ai trouvé m'ont permis d'écrire un article [3], je vais dans cette section présenter rapidement le positionnement du problème ainsi que les résultats obtenus, sans détailler les preuves.

**Définition 10.** Notons  $(e_1, \dots, e_n)$  la base canonique de  $\mathbb{R}^n$ . Soit  $\pi \in S_n$ , la matrice de permutation associée à  $\pi$  est  $M(\pi) = (\delta_{i, \sigma(j)})_{1 \leq i, j \leq n} = (e_{\pi(1)}, \dots, e_{\pi(n)})$ . Les colonnes de la matrice sont donc les vecteurs de la base réordonnés selon la permutation.  $M$  est une représentation fidèle de  $S_n$ .

En m'inspirant de la notion de tore de De Bruijn, j'ai cherché une matrice torique, c'est à dire une matrice où l'on identifie le haut et le bas, ainsi que la droite et la gauche. La notion de minimalité étant difficile à définir, j'ai choisi de me restreindre au problème suivant :

Trouver le plus petit entier  $m_1(n)$  tel que il existe  $T \in M_{n, m_1(n)}(\{0, 1\})$  avec  $T$  qui contient chaque matrice de permutation de  $S_n$  en tant que bloc.

Remarque : On aurait pu considérer des matrices où l'on fixe le nombre de colonnes à  $n$  exactement, mais les deux problèmes sont équivalent par transposition car  ${}^t M(\pi) = M(\pi^{-1})$ .

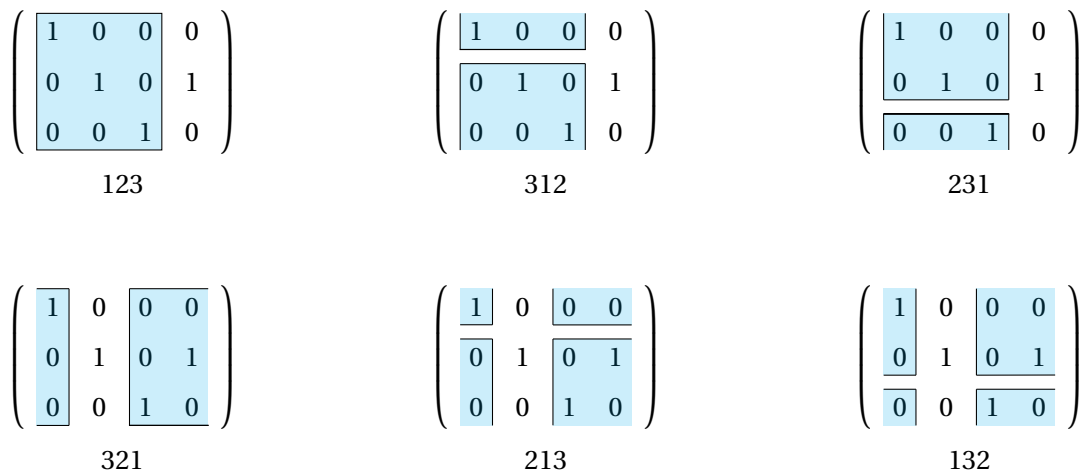


FIGURE 5 – Blocs d'une matrice  $3 \times 4$  correspondant au permutations de  $S_3$ .

Or, le caractère torique d'une matrice  $T$  implique que si  $M(\pi)$  est dans  $T$  comme bloc, alors toute matrice correspondant à un élément de  $\text{inc}(\pi) = \{\sigma^i \pi \mid 0 \leq i \leq n-1\}$  l'est aussi. Cette remarque permet de se ramener à la recherche de la longueur minimale  $m(n)$  d'un mot universel pour  $S_n / \mathcal{R}$  où  $\mathcal{R}$  est la relation d'équivalence telle que  $x \mathcal{R} y$  si et seulement si  $\exists i, y = \sigma^i x$ . On aura alors  $m(n) - (n-1) \leq m_1(n) \leq m(n)$  car on perd le caractère torique sur les bords avec le mot.

On utilise alors les mêmes idées qu'avant pour chercher une borne sur la longueur de ce mot : on définit le graphe  $H_n$  comme un graphe complet dont les sommets sont les  $(n-1)!$  classes d'équivalence pour la relation  $\mathcal{R}$ . Le poids d'une arête entre  $x$  et  $y$  est le nombre minimal de colonne à rajouter à droite d'une matrice d'un élément de  $x$  (et à enlever à gauche) pour obtenir une matrice d'un élément de  $y$ . Comme en 2.3, avec un chemin passant par tous les sommets, on peut obtenir un mot universel pour notre problème. Comme un tel chemin a donc  $(n-1)! - 1$  arêtes et qu'une arête a un poids compris entre 1 et  $n-1$ , on a les bornes triviales suivantes :

**Lemme 4.1.**

$$m(n) \leq S(n) = (n-1)(n-1)! + 1$$

$$m(n) \geq I(n) = (n-1)! + n - 1$$

Pour trouver des bornes, j'ai cherché des chemins utilisant le plus d'arêtes de poids 1. Or, une arête de poids 1 va de  $\text{inc}(\pi)$  vers  $\text{inc}(\pi\sigma)$ , donc comme  $\sigma^n = Id$ , les arêtes de poids 1 forment des cycles, nommés 1-cycles. J'ai alors montré que la taille d'un 1-cycle devait diviser  $n$ , et que si l'on note  $C(d)$  le nombre de 1-cycles de taille  $d$ , alors

$$\sum_{k|d} C(k) = \varphi\left(\frac{n}{d}\right) \left(\frac{n}{d}\right)^{d-1} (d-1)!$$

En considérant un chemin qui suit un 1-cycle entier, puis passe au suivant, on peut majorer le poids du chemin car on connaît le nombre de 1-cycles de chaque taille, et qu'entre deux 1-cycles, l'arête a un poids de  $n-1$  au plus. On obtient ainsi le résultat suivant :

**Théorème 4.2.**

$$m(n) \leq B(n) = 1 + \sum_{d|n} C(d)(d+n-2) \leq B'(n) = 1 + (n-1)! + \left( \sum_{d|n} \varphi\left(\frac{n}{d}\right) \frac{n^{d-1}(d-1)!}{d^d} \right) (n-2)$$

En utilisant une preuve similaire à celle du Théorème 2.2, on peut aussi avoir une borne inférieure :

**Théorème 4.3.**

$$m(n) \geq D(n) = (n-1)! + n - 1 + \sum_{d|n} C(d)$$

A priori, il n'est pas évident de savoir si ces bornes sont efficaces par rapport aux bornes triviales. Mais on peut voir que le rapport  $\frac{S(n)}{I(n)}$  est équivalent à  $n$  donc tend vers l'infini. Or, j'ai montré que le rapport  $\frac{B(n)}{I(n)}$  avait pour limite 2, ce qui montre que la borne supérieure est bien meilleure que la précédente.

## Références

- [1] Anonymous 4chan user, R. Houston, J. Pantone, and V. Vatter. A lower bound on the length of the shortest superpattern. <https://oeis.org/A180632/a180632.pdf>, 2018.
- [2] D. A. Ashlock and J. Tillotson. Construction of small superpermutations and minimal injective superstrings. *Congressus Numerentium*, 93 :91–98, 1993.
- [3] G. Dumas. Superpermutation matrices. *arXiv e-prints*, page arXiv:1908.04708, Aug 2019.
- [4] G. Egan. Superpermutations. <http://www.gregegan.net/SCIENCE/Superpermutations/Superpermutations.html>, 2018.
- [5] M. Engen and V. Vatter. Containing all permutations. *arXiv e-prints*, page arXiv:1810.08252, Oct 2018.
- [6] R. Houston. Tackling the Minimal Superpermutation Problem. *arXiv e-prints*, page arXiv:1408.5108, Aug 2014.
- [7] N. Johnston. The minimal superpermutation problem. <http://www.njohnston.ca/2013/04/the-minimal-superpermutation-problem/>, 2013.
- [8] N. Johnston. Non-Uniqueness of Minimal Superpermutations. *arXiv e-prints*, page arXiv:1303.4150, Mar 2013.
- [9] N. Johnston. All minimal superpermutation on five symbols have been found. <http://www.njohnston.ca/2014/08/all-minimal-superpermutations-on-five-symbols-have-been-found/>, 2014.
- [10] J. Sawada and A. Williams. A hamilton path for the sigma-tau problem. *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 568–575, 2018.
- [11] A. Williams. Hamiltonicity of the Cayley Digraph on the Symmetric Group Generated by  $\sigma = (1\ 2 \dots n)$  and  $\tau = (1\ 2)$ . *arXiv e-prints*, page arXiv:1307.2549, Jul 2013.