



**HAL**  
open science

## Grappe et éthique : une contribution à Parcoursup

Vincent Bebien

► **To cite this version:**

Vincent Bebien. Grappe et éthique : une contribution à Parcoursup. Recherche opérationnelle [math.OC]. 2022. dumas-03843917

**HAL Id: dumas-03843917**

**<https://dumas.ccsd.cnrs.fr/dumas-03843917v1>**

Submitted on 8 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Grappe et éthique

Une contribution à Parcoursup

---

**Vincent BEBIEN**

Master ORO - Nantes Université

Mémoire de stage

Effectué à l'IMT Atlantique

Encadré par Odile BELLENGUEZ et Gilles SIMONIN



**IMT Atlantique**  
Bretagne-Pays de la Loire  
École Mines-Télécom



**Nantes**  
**Université**

Soutenance prévue le 30 juin 2022

# Abstract

Parcoursup est depuis 2018 la plateforme nationale d'accès à l'enseignement supérieur, accueillant chaque année plusieurs centaines de milliers de candidats. Les critiques à l'encontre de ce système sont régulières depuis sa mise en place, que ce soit de la part des candidats, des établissements ou des sociologues. L'objectif de ce stage est de fournir une analyse technique de Parcoursup, et en particulier de son système d'affectation, basé sur le problème bien connu des mariages stables. Cette analyse a pour but de clarifier l'impact des algorithmes utilisés et d'identifier les problèmes qu'ils peuvent engendrer, en se basant sur différents principes de l'éthique des algorithmes. De plus, nous proposons une méthode alternative d'affectation permettant une modélisation plus fine des préférences de chacun, méthode que nous analyserons également et comparerons aux algorithmes actuellement utilisés.

**Mots-clés** : Éthique des algorithmes, Problèmes d'affectation, Mariages stables

# Table des matières

|  |           |
|--|-----------|
| <b>Introduction</b>  | <b>4</b>  |
| <b>1 État de l'art</b>   | <b>5</b>  |
| 1.1 Contexte éthique . . . . .   | 5         |
| 1.2 Problèmes de mariages stables avec ordre strict . . . . .                    | 6         |
| 1.3 Problèmes de mariages stables avec ordre partiel . . . . .                   | 9         |
| <b>2 Parcoursup</b>  | <b>13</b> |
| 2.1 Admission Post-Bac et loi ORE . . . . .                                      | 13        |
| 2.2 Fonctionnement de la plateforme . . . . .                                    | 14        |
| 2.3 Parcoursup et les mariages stables . . . . .                                 | 16        |
| <b>3 Analyse critique de Parcoursup</b>  | <b>18</b> |
| 3.1 Principes éthiques mis en avant dans Parcoursup . . . . .                    | 18        |
| 3.2 Points positifs . . . . .  | 19        |
| 3.3 Limites . . . . .  | 22        |
| 3.4 Synthèse . . . . .   | 30        |
| <b>4 Proposition d'une nouvelle modélisation : préférences avec indifférence</b> | <b>32</b> |
| 4.1 Motivation . . . . .   | 32        |
| 4.2 Expérimentation . . . . .  | 33        |
| 4.3 Avantages et limites de l'approche . . . . .                                 | 36        |
| <b>5 Conclusion</b>  | <b>38</b> |

# Introduction

Depuis 2018, Parcoursup est la plateforme nationale française dédiée à l'accès à l'enseignement supérieur. Chaque année, des centaines de milliers de personnes s'y inscrivent dans le but d'être affectés à une formation post-bac. Les inscrits doivent rentrer des vœux d'admission correspondant aux formations qu'ils souhaitent intégrer, ces vœux sont ensuite traités par les formations concernées, puis des propositions d'admissions sont envoyées aux candidats choisis par l'algorithme d'affectation. Toutes ces étapes sont concentrées sur une même plateforme, ce qui rend Parcoursup incontournable dans le processus d'admission, d'autant plus que celle-ci accueille de plus en plus de formations dans son catalogue (près de 20.000 en 2021).

Le nombre important de candidatures et de formations que concentre Parcoursup rend des traitements algorithmiques nécessaires pour assurer de bonnes affectations. Étant donné l'importance de Parcoursup dans le parcours scolaire français, le choix des algorithmes utilisés est crucial car il peut impacter le devenir de milliers d'étudiants. La plateforme est ainsi régulièrement observée et est devenu un objet important du débat public, faisant émerger diverses critiques sur les multiples aspects qu'elle touche. En particulier, les candidats se plaignent de subir un stress important lié à l'incertitude quant à leur entrée dans l'enseignement supérieur<sup>1</sup>, d'obtenir des affectations décevantes malgré leur niveau<sup>2</sup> ou encore de subir des temps d'attente très longs pour obtenir leurs affectations<sup>3</sup>. Frouillou [9] et Mizzi [29] reprennent ces arguments et critiquent l'impact sociologique que la plateforme peut avoir, notamment l'amplification des inégalités entre étudiants favorisés et défavorisés. Dès lors, on peut se demander d'où proviennent ces problèmes majeurs : viennent-ils directement de la plateforme et de l'algorithme d'affectation, des choix politiques formant le cadre de Parcoursup ou encore des usages mêmes faits par les candidats et les formations ?

Dans ce rapport, nous allons ainsi nous intéresser à l'étude de Parcoursup, et principalement à l'algorithme d'affectation qu'il embarque, en adoptant un point de vue éthique sur la procédure. Dans le chapitre 1, nous détaillerons les principaux principes éthiques liés à l'Intelligence Artificielle (IA) et à Parcoursup, ainsi que les principaux problèmes de mariages stables sur lesquels l'algorithme d'affectation se base. Dans le chapitre 2, nous expliciterons le fonctionnement global de la procédure pour en extraire les éléments principaux qui forment son mécanisme d'affectation. Dans le chapitre 3, nous reprendrons plusieurs points notables explicités précédemment afin d'analyser plus en détail les principes éthiques sous-jacents. Dans le chapitre 4, nous proposerons ensuite une alternative à ce système afin de mieux comprendre les enjeux d'un choix d'algorithme d'affectation et étudier les impacts que cela peut avoir. Enfin, dans le chapitre 5 nous ferons une synthèse des points vus précédemment avant de proposer des perspectives de travail.

---

1. Parcoursup : "C'est un moment de stress tous les matins", racontent des élèves sur liste d'attente - Franceinfo 10/06/2021
2. « Sur Parcoursup, le premier jour, c'est une déception pour presque tout le monde » - Le Monde 02/06/2022
3. Parcoursup : plus de 90 000 candidats toujours en attente de places - Le Monde 06/07/2021

# Chapitre 1

## État de l'art

### 1.1 Contexte éthique

La question de l'éthique des algorithmes est très souvent abordée avec le terme d'intelligence artificielle (IA), si bien qu'on parle souvent d'éthique de l'IA. La commission européenne<sup>1</sup> définit l'IA comme « un logiciel [...] qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit ». Cette définition inclut aussi bien des modèles de Machine Learning (ML) que des modèles d'optimisation comme ceux intégrés dans Parcoursup. Même si le ML peut poser des problèmes spécifiques dues à une approche statistique (biais d'échantillonnage, discrimination, etc...), les grands principes de l'éthique de l'IA touchent également des applications de Recherche Opérationnelle.

Le travail mené lors de ce stage reste technique mais se place dans un contexte où les enjeux éthiques liés aux algorithmes sont très étudiés, et nous nous appuyerons donc sur cette littérature pour poser un cadre éthique à notre étude. De nombreuses institutions, entreprises et organisations ont publié leur propre cadre éthique spécifique à leur domaine ou global. Au niveau national, la CNIL a écrit en 2017 un rapport nommé « Comment permettre à l'homme de garder la main ? »<sup>2</sup> présentant les grands questionnements liés à l'implantation de l'IA en France et des moyens d'action pour répondre à ces problématiques.

Floridi et Cowls [7] se basent sur six cadres éthiques différents définis par différentes institutions comme la commission européenne<sup>3</sup> (dont nous reprendrons les principes durant notre analyse). Ils identifient grâce à cela cinq grands principes généraux de l'éthique de l'IA : la bienfaisance, la non-malfaisance, l'autonomie, la justice et l'explicabilité. Les quatre premiers sont inspirés des principes traditionnels de la bioéthique définis par Beauchamp et Childress [2] et le dernier est un nouveau principe spécifique à l'IA. La bienfaisance se traduit par la volonté de placer le bien-être, le bien commun et la dignité humaine comme objectifs de la décision à prendre, tandis que le principe de non-malfaisance interdit les décisions volontairement malfaisantes, celles entraînant des effets négatifs identifiables, etc... L'autonomie humaine se traduit par la capacité d'un utilisateur à prendre des décisions autonomes face aux recommandations d'un système, et donc souhaite préserver la responsabilité de chacun sur ses actes. La justice cherche à éviter tous les biais et discriminations que peuvent engendrer l'utilisation d'une IA et à garantir un accès égal aux bénéfices qu'elle peut apporter. Enfin, le principe d'explicabilité demande à ce que les algorithmes et leurs mécanismes de décision puissent être compréhensibles pour tous (ce qui est notamment une nécessité pour préserver notre autonomie en tant qu'utilisateur), et non pas seulement pour la fraction de l'humanité qui les crée. Ce dernier principe est crucial car tous les autres principes ne peuvent être appliqués que si le fonctionnement d'un système d'IA est transparent aux yeux de tous.

---

1. [Règlement du parlement européen et du conseil établissant des règles harmonisées concernant l'intelligence artificielle, article 3](#)

2. [Comment permettre à l'homme de garder la main ? - CNIL](#)

3. [Ethics guidelines for trustworthy AI - Groupe d'Experts de Haut Niveau sur l'IA, Commission Européenne](#)

En ce qui concerne le cas particulier de Parcoursup, Tiberj [33] met en parallèle les principes algorithmiques et sociologiques de la plateforme en prenant l'exemple de l'IEP (Instituts d'Études Politiques) Bordeaux, et propose également des alternatives aux algorithmes en place. Frouillou et al. [10] explicitent l'ensemble de principes politiques et moraux auquel adhère Parcoursup, tout en relevant les conséquences sociologiques de ces principes. Mizzi [29] a décrit les différents obstacles liés à Parcoursup auxquels sont confrontés les candidats.

Depuis 2018, le Comité Éthique et Scientifique de Parcoursup (CESP) suit l'évolution de la plateforme et analyse les données de chaque campagne de Parcoursup. En se basant sur ces données, le CESP réalise une analyse globale ainsi que des études de cas pour formuler des recommandations éthiques sur la procédure et le fonctionnement de la plateforme. Ces analyses et résultats sont publiés chaque année dans un rapport au Parlement, nous nous baserons principalement par la suite sur ceux de 2021<sup>1</sup> et 2022<sup>2</sup>.

Afin d'analyser Parcoursup, nous avons également besoin de comprendre les mécanismes impliqués dans son fonctionnement, par la suite nous nous pencherons donc sur les problèmes de mariages stables qui sont à la base de la procédure d'affectation.

## 1.2 Problèmes de mariages stables avec ordre strict

Dans cette section, nous présentons les principales variantes du problème des mariages stables. Ces problèmes impliquent deux ensembles d'agents formulant chacun des préférences sur les membres de l'autre ensemble, le but étant de créer un ensemble d'affectations *stable*. Les problèmes de mariages stables ont été définis en premier par Gale et Shapley en 1962 [11], une solution est dite stable si elle n'admet pas de paire bloquante, la définition de paire bloquante change selon le problème et sera explicitée pour chaque variante.

### 1.2.1 Problème du Stable Marriage (SM)

Étant donné un ensemble d'hommes  $U = \{m_1, \dots, m_n\}$  et un ensemble de femmes  $W = \{w_1, \dots, w_n\}$ , chaque homme possède une liste de préférences strictement ordonnée sur toutes les femmes, et inversement. Étant donné un homme  $m_i \in U$  et deux femmes  $w_j, w_k \in W$ , on dit que  $m_i$  préfère  $w_j$  à  $w_k$  si  $w_j$  précède  $w_k$  dans la liste de  $m_i$ . Cette relation de préférence est définie de manière symétrique pour les femmes. Le problème du Stable Marriage consiste à vouloir affecter les hommes et les femmes ensemble de manière à créer des paires contenant chacune un homme et une femme. Une telle affectation est représentée par un ensemble  $M \subseteq U \times W$ , la personne à qui est affecté  $m_i$  (resp.  $w_i$ ) dans  $M$  est notée  $M(m_i)$  (resp.  $M(w_i)$ ). Une affectation  $M$  est dite stable si elle n'admet aucune paire bloquante.

**Définition 1** (Paire bloquante dans un SM). Dans une instance  $I$  de SM, une paire  $(m_i, w_j) \notin M$  est dite bloquante pour  $M$  si les conditions suivantes sont respectées :

- $m_i$  préfère  $w_j$  à  $M(m_i)$
- $w_j$  préfère  $m_i$  à  $M(w_j)$

La table 1.1 représente une instance du problème Stable Marriage. Le tableau de gauche représente les préférences ordonnées des hommes sur toutes les femmes, et celui de droite représente les préférences ordonnées des femmes sur les hommes. Par exemple, les préférences de  $m_1$  sont :  $w_4 > w_1 > w_2 > w_3$  (ou  $w_i > w_j$  signifie que  $w_i$  est préférée à  $w_j$ ). Une affectation  $M'$  est représentée en bleu, et il est possible de déterminer si cette affectation est stable ou non en vérifiant s'il existe une paire  $(m', w')$  respectant les conditions énoncées ci-dessus. La paire  $(m_1, w_4)$  est bloquante selon  $M'$  car  $m_1$  préfère  $w_4$  à sa partenaire et  $w_4$  préfère  $m_1$  à son partenaire ; de la même manière,  $(m_3, w_1)$  constitue également une paire bloquante.  $M'$  n'est donc pas une affectation stable.

---

1. <https://www.vie-publique.fr/sites/default/files/rapport/pdf/283849.pdf>  
2. <https://www.vie-publique.fr/sites/default/files/rapport/pdf/283851.pdf>

|       |       |       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $m_1$ | $w_4$ | $w_1$ | $w_2$ | $w_3$ | $w_1$ | $m_3$ | $m_1$ | $m_2$ | $m_4$ |
| $m_2$ | $w_1$ | $w_2$ | $w_4$ | $w_3$ | $w_2$ | $m_2$ | $m_4$ | $m_1$ | $m_3$ |
| $m_3$ | $w_2$ | $w_1$ | $w_3$ | $w_4$ | $w_3$ | $m_3$ | $m_4$ | $m_2$ | $m_1$ |
| $m_4$ | $w_3$ | $w_2$ | $w_4$ | $w_1$ | $w_4$ | $m_1$ | $m_2$ | $m_4$ | $m_3$ |

TABLE 1.1 – Listes de préférences pour un problème SM, paires bloquantes (en rouge) pour l’affectation  $M'$  (en bleu)

### 1.2.2 Problème du Stable Marriage with Incomplete lists (SMI)

Le problème du SMI est une généralisation de ce problème qui admet des listes de préférences incomplètes. Chaque homme  $m$  (resp. femme  $w$ ) possède une liste acceptable  $A(m) \subseteq W$  (resp.  $A(w) \subseteq U$ ) indiquant les femmes (resp. hommes) à qui il (resp. elle) peut être affecté. On note l’ensemble des paires acceptables  $E = \{(m, w) : w \in A(m) \wedge m \in A(w), \forall m \in U, w \in W\}$ , nécessairement on doit avoir  $M \subseteq E$  pour toute solution  $M$ . Cette généralisation implique que des personnes peuvent se retrouver sans affectation, on dit alors que  $a_i \in U \cup W$  est non-affecté si  $M(a_i) = \emptyset$ , sinon il est affecté. La définition d’une paire bloquante doit donc être modifiée pour le SMI afin de refléter ce point.

**Définition 2** (Paire bloquante dans un SMI). Dans une instance  $I$  de SMI, une paire  $(m_i, w_j) \in E \setminus M$  est dite bloquante pour  $M$  si les conditions suivantes sont respectées :

- $m_i$  est non-affecté ou préfère  $w_j$  à  $M(m_i)$
- $w_j$  est non-affectée ou préfère  $m_i$  à  $M(w_j)$

### 1.2.3 Problème du Hospitals/Residents (HR)

Dans ce problème, on cherche à affecter des internes en médecine dans des hôpitaux. On utilise pour ce problème les ensembles  $R = \{r_1, \dots, r_{n_1}\}$  et  $H = \{h_1, \dots, h_{n_2}\}$  pour ainsi représenter les internes et les hôpitaux. Cette généralisation *many/one* du problème SMI autorise les hôpitaux à avoir une capacité supérieure à 1, représentant le nombre de postes que l’hôpital a à pourvoir. Chaque hôpital  $h_j$  a donc une capacité  $c_j$  ; on dit que  $h_j$  est en sous-effectif ou rempli selon si  $|M(h_j)|$  est inférieur ou égal à  $c_j$  respectivement. Une affectation  $M$  doit faire correspondre chaque interne à au plus 1 hôpital et chaque hôpital à au plus  $c_j$  internes.

**Définition 3** (Paire bloquante dans un HR). Une affectation  $M$  dans une instance  $I$  de HR est stable si elle ne contient pas de paire bloquante  $(r_i, h_j) \in E \setminus M$ . Cette paire est bloquante si elle respecte les conditions suivantes :

- $r_i$  est non-affectée ou préfère  $h_j$  à  $M(r_i)$
- $h_j$  est en sous-effectif ou préfère  $r_i$  à l’un des membres de  $M(h_j)$  (ou les deux)

Une instance de HR peut admettre plusieurs mariages stables pouvant être obtenues au moyen de rotations entre plusieurs affectations. Les rotations consistent en un sous-ensemble d’internes intervertissant leurs affectations dans un mariage stable, menant à un nouveau mariage stable. Il est possible de parcourir tout l’ensemble des mariages stables possibles grâce à ces rotations [13]. Comme ces rotations se font uniquement entre internes affectés, une propriété importante des solutions stables est donnée par le théorème des « hôpitaux ruraux » :

**Définition 4** (Théorème des hôpitaux ruraux [32]). Pour une instance donnée de HR, les propriétés suivantes sont vérifiées :

- les internes affectés sont les mêmes dans toutes les affectations stables
- chaque hôpital accueille le même nombre d’internes dans toutes les affectations stables
- tous les hôpitaux en sous-effectif dans une affectation stable accueilleront le même ensemble d’internes dans toutes les affectations stables

Dans la suite, nous nous intéresserons particulièrement à ce problème car en plus d'être une généralisation des problèmes de mariages stables vus précédemment, il représente particulièrement bien notre cas d'étude de Parcoursup, où ici les hôpitaux correspondent aux formations d'enseignement supérieur et où les internes correspondent aux candidats. On utilisera donc par la suite les ensembles  $F = \{f_1, \dots, f_{n_1}\}$  et  $C = \{c_1, \dots, c_{n_2}\}$  pour parler respectivement des formations et des candidats.

### 1.2.4 Algorithme de Gale-Shapley

L'algorithme le plus simple pour trouver une affectation stable a été proposé par Gale et Shapley en 1962 [11]. Nous présenterons ici l'algorithme pour le problème particulier du Stable Marriage. Le principe est que chaque homme envoie des propositions aux femmes en suivant l'ordre de sa liste de préférences. Si une femme non-affectée reçoit une proposition, alors elle accepte la proposition. Si elle est déjà affectée à un autre homme, alors elle choisit son préféré selon sa propre liste et rejette l'autre. Dans une version optimisée de l'algorithme telle que présentée ci-dessous, l'homme qui vient de proposer sera toujours préféré car pour une femme donnée, tous les hommes moins appréciés que son partenaire actuel sont supprimés de sa liste. La procédure continue jusqu'à ce que chaque homme (i) soit affecté ou (ii) n'ai plus de femmes à appeler dans sa liste. La condition (i) est suffisante dans le cas spécifique du SM, mais pour des instances de SMI il est possible que tous les hommes ne soient pas affectés.

L'algorithme est représenté en détail ci-dessous.  $M$  représente l'ensemble des couples affectés; ainsi si  $(m, w)$  appartient à  $M$ , alors  $M(m) = w$  et  $M(w) = m$ , et si  $p$  n'est affecté à personne alors  $M(p) = \emptyset$ . Si une femme  $w$  préfère l'homme  $m$  à  $m'$  alors on note  $m \succ_w m'$ .

---

#### Algorithme 1 : Algorithme de Gale-Shapley

---

**Données :**  $L_m, L_w$ , listes de préférence ordonnées des hommes  $m \in U$  et des femmes  $w \in W$

$M \leftarrow \emptyset$

$\forall m \in U, t_m = 1$

**tant que**  $\exists m \in U : M(m) = \emptyset$  **et**  $t_m \leq |L_m|$  **faire**

$w \leftarrow L_m[t_m]$

$m' \leftarrow M(w)$

**si**  $m' = \emptyset$  **alors**

$M \leftarrow M \cup (m, w)$

**sinon**

$M \leftarrow M \cup (m, w) \setminus (m', w)$

$t_{m'} \leftarrow t_{m'} + 1$

**fin**

**pour**  $m^* \in L_w : m \succ_w m^*$  **faire**

$L_{m^*} \leftarrow L_{m^*} \setminus w$

$L_w \leftarrow L_w \setminus m^*$

**fin**

**fin**

**retourner**  $M$

---

La boucle "pour" est utilisée afin de supprimer des paires inutiles de l'instance : si une femme  $w$  a reçu une proposition, elle ne détériorera pas son affectation courante  $m$  en choisissant un homme qu'elle aime moins. Ainsi, tous les hommes moins bien placés que  $m$  dans  $L_w$  peuvent supprimer  $w$  de leurs listes car leurs propositions seront de toute façon refusées. Ce principe permet de réduire le nombre d'itérations de la boucle principale en ne conservant que les propositions qui peuvent être acceptées.

L'algorithme présenté ci-dessus retourne une affectation stable particulière  $M_0$  car c'est l'unique affectation qui est optimale pour les hommes. Autrement dit, il n'existe pas d'affectation stable qui serait plus satisfaisante pour les hommes que  $M_0$ . Dans cette affectation, chaque femme reçoit la pire affectation stable qu'elle pourrait avoir parmi toutes les affectations stables [13]. A l'inverse, il est possible d'exécuter l'algorithme à l'envers en permettant aux femmes de proposer aux hommes, dans ce cas l'algorithme renvoie dans

|       | $m_1$ | $m_2$ | $m_3$ | $m_4$ |
|-------|-------|-------|-------|-------|
| $M_0$ | $w_4$ | $w_1$ | $w_2$ | $w_3$ |
| $M_z$ | $w_4$ | $w_2$ | $w_1$ | $w_3$ |

TABLE 1.2 – Mariage optimal pour les hommes ( $M_0$ ) et pour les femmes ( $M_z$ )

ce cas l'affectation stable  $M_z$  qui est optimale pour les femmes. La table 1.2 représente ces deux solutions stables optimales correspondant à l'instance présentée à la table 1.1. D'autres affectations stables situées entre la solution optimale pour les hommes et celle pour les femmes peuvent exister mais l'algorithme de Gale-Shapley ne permet pas de les trouver. Ces solutions forment un treillis où  $M_0$  et  $M_z$  représentent les éléments minimum et maximum de ce treillis [13], toutes les autres solutions stables peuvent être trouvées au moyen de rotations. Il est possible que pour certaines instances, le mariage  $M_0$  soit le même que  $M_z$ , dans ce cas il est le seul mariage stable existant pour ces instances.

Afin d'adapter cette procédure pour le problème HR, il faut prendre en compte les capacités des hôpitaux : un hôpital  $h_j$  peut envoyer (ou recevoir) au maximum  $c_j$  propositions. Également, si la version de l'algorithme utilisée est l'optimale pour les internes, alors il faut qu'un hôpital soit plein avant de pouvoir supprimer des internes de sa liste.

L'algorithme de Gale-Shapley a une complexité linéaire selon le nombre de paires acceptables, et polynomiale selon le nombre d'agents (hommes, femmes, internes, places d'hôpital). Sa capacité à calculer un mariage stable en un temps très court le rend très intéressant et est donc très souvent utilisé comme base pour la résolution de problèmes de stabilité. Dans le cas de l'affectation des étudiants à des formations d'enseignement supérieur en France, c'est l'algorithme principal qui était utilisé de 2009 à 2017 avec Admission Post Bac et qui est toujours utilisé aujourd'hui de manière indirecte comme nous allons le présenter par la suite.

### 1.3 Problèmes de mariages stables avec ordre partiel

Dans cette section, nous nous intéressons au cas où les préférences des différents agents peuvent contenir des équivalences entre plusieurs choix. La généralisation du problème SM pour ce cas est le SMT (Stable Marriage with Ties) où chaque paire homme-femme est acceptable. De la même manière, HRT (Hospitals/Residents problem with Ties) généralise le problème HR et SMTI (Stable Marriage with Ties and Incomplete lists) est la restriction de HRT où les hôpitaux ont une capacité de 1.

Par la suite, nous nous intéresserons plus particulièrement au problème HRT car il constitue une généralisation des autres problèmes nommés précédemment et ses propriétés peuvent être appliquées à ceux-ci.

#### 1.3.1 Problème du Hospitals/Residents with Ties (HRT)

Cette extension du problème Hospitals/Residents autorise la présence de relations d'indifférence entre plusieurs choix. La table 1.3 montre une instance de HRT où les préférences sont ordonnées de gauche à droite avec les égalités placées entre parenthèses. La présence de cette nouvelle relation autorise des égalités de préférence qui font apparaître de nouveaux critères de stabilité car la définition classique de la stabilité pour le problème Hospitals/Residents donnée dans la définition 3 ne suffit plus. Ces nouveaux critères, définis par Irving [16] sont la stabilité faible, la stabilité forte et la super-stabilité.

**Définition 5** (Paire bloquante dans un HRT). Étant donné une affectation  $M$  d'une instance  $I$  de HRT, une paire acceptable  $(r_i, h_j) \in E \setminus M$  est bloquante si les conditions suivantes sont respectées selon le niveau de stabilité :

- Stabilité faible :
  - $r_i$  n'est pas affectée ou préfère  $h_j$  à l'hôpital auquel elle est affectée dans  $M$ , et

|     |       |       |       |       |       |       |               |       |
|-----|-------|-------|-------|-------|-------|-------|---------------|-------|
|     |       |       |       |       |       | $r_1$ | $(h_1 \ h_2)$ | $h_3$ |
| (1) | $h_1$ | $r_1$ | $r_4$ | $r_6$ |       | $r_2$ | $h_2$         | $h_4$ |
| (1) | $h_2$ | $r_1$ | $r_2$ | $r_4$ | $r_5$ | $r_3$ | $(h_3 \ h_4)$ |       |
| (2) | $h_3$ | $r_1$ | $r_3$ | $r_5$ | $r_6$ | $r_4$ | $h_1$         | $h_2$ |
| (2) | $h_4$ | $r_2$ | $r_3$ |       |       | $r_5$ | $(h_2 \ h_3)$ |       |
|     |       |       |       |       |       | $r_6$ | $h_1$         | $h_3$ |

TABLE 1.3 – Listes de préférence partiellement ordonnées, les capacités de chaque formation sont données à gauche de leurs préférences

- $h_j$  est en sous-effectif ou préfère  $r_i$  à sa pire interne affectée dans  $M$
- Stabilité forte : soit (a)
  - $r_i$  n'est pas affectée ou préfère  $h_j$  à l'hôpital auquel elle est affectée dans  $M$ , et
  - $h_j$  est en sous-effectif ou préfère  $r_i$  à sa pire interne affectée dans  $M$  ou est indifférent entre les deux
- soit (b)
  - $r_i$  n'est pas affectée ou préfère  $h_j$  à l'hôpital auquel elle est affectée dans  $M$  ou est indifférente entre les deux, et
  - $h_j$  est en sous-effectif ou préfère  $r_i$  à sa pire interne affectée dans  $M$
- Super-stabilité :
  - $r_i$  n'est pas affectée ou préfère  $h_j$  à l'hôpital auquel elle est affectée dans  $M$  ou est indifférente entre les deux, et
  - $h_j$  est en sous-effectif ou préfère  $r_i$  à sa pire interne affectée dans  $M$  ou est indifférent entre les deux

Par la suite, nous nous intéresserons principalement à la stabilité faible car elle peut être trouvée dans toutes les instances de HRT [26], contrairement aux autres [25]. Cependant, il est possible de vérifier l'existence de solutions respectant la stabilité forte et la super-stabilité en un temps polynomial [20] [17]. Ces types de solutions forment également un treillis, contrairement aux solutions faiblement stables.

### 1.3.2 Problème MAX HRT

Comme les solutions de HRT peuvent être de différentes tailles, un critère à maximiser intéressant est la taille de l'affectation : même s'il y a autant de places que d'internes, une affectation stable n'affecte pas nécessairement toutes les internes et toutes les affectations stables n'ont pas nécessairement la même taille. Dès lors, l'objectif de trouver l'affectation stable qui maximise le nombre d'internes placées est intéressante et très étudiée (MAX SMTI/MAX HRT).

L'exemple de la table 1.3 montre la différence de cardinalité qu'il peut y avoir entre plusieurs affectations faiblement stables. De manière générale, appliquer l'algorithme de Gale-Shapley en cassant arbitrairement les égalités donnera toujours une affectation faiblement stable [26]. On utilise cette méthode ici orientée pour les hôpitaux, où chaque égalité est cassée en favorisant les membres de gauche aux membres de droite (ex :  $(h_1 \ h_2)$  devient  $h_1 > h_2$ ) pour obtenir l'affectation  $M_0$ . On peut voir dans la table 1.4, le nombre de places occupées est de 4 sur les 6 au total. Cependant, une autre affectation faiblement stable représentée par  $M_1$  remplit toutes les places disponibles et maximise le nombre d'affectations. Selon la méthode employée, il est donc possible de trouver des tailles d'affectation relativement éloignées.

Le problème de trouver une affectation de taille maximale est propre aux instances avec égalité car les instances classiques du Hospitals/Residents sont de même taille comme l'indique le théorème des hôpitaux ruraux (définition 4). Les propriétés énoncées dans cette définition ne s'appliquent pas pour HRT, les deux affectations produites dans la table 1.4 en sont un bon contre-exemple. Ce théorème montre que, malgré l'existence d'algorithmes optimaux pour chacun des deux côtés, les affectations stables bénéficieront toujours

|       | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_5$ | $r_6$ |
|-------|-------|-------|-------|-------|-------|-------|
| $M_0$ | $h_1$ | $h_2$ | $h_3$ | -     | $h_3$ | -     |
| $M_1$ | $h_2$ | $h_4$ | $h_4$ | $h_1$ | $h_3$ | $h_3$ |

TABLE 1.4 – Affectations faiblement stables

aux mêmes hôpitaux et internes, et ceux qui n'ont rien dans une affectation n'auront toujours rien peu importe l'algorithme employé. Les ordres partiels des préférences peuvent ainsi donner lieu à de nouvelles affectations qui n'auraient pas été stables si les préférences étaient cantonnées à un ordre strict, et ainsi élargir le nombre de "partenaires" stables pour chaque candidat.

### 1.3.3 Méthodes de résolution de MAX SMTI/MAX HRT

Manlove et al. [26] ont montré que le problème MAX SMTI est NP-Hard. Différentes modélisations mathématiques des problèmes MAX SMTI/MAX HRT ont été présentées par Delorme et al. [5]. Plusieurs algorithmes d'approximation sur le nombre d'affectations ont également été créés avec des garanties de performance différentes. Ces algorithmes sont pour la plupart des extensions de l'algorithme de Gale-Shapley (algorithme 1).

#### Égalités des deux côtés

McDermid [28] est le premier à avoir trouvé un algorithme avec un ratio d'approximation égal à  $3/2$  s'exécutant en un temps polynomial par rapport à la taille des données. Par la suite, Paluch [31] et Király [22] ont proposé des algorithmes en temps linéaire avec le même ratio.

Sur des instances où les listes de préférence du côté "proposant" sont strictement ordonnées, les algorithmes ci-dessus sont équivalents à l'algorithme GSA1 également proposé par Király [21].

Huang et Kavitha ont également proposé un algorithme pour le cas particulier où les égalités sont de taille 2 [15] avec un ratio d'approximation amélioré à  $10/7$ . Une analyse postérieure a baissé cette borne supérieure à  $4/3$  [4], ce qui correspond également à la meilleure borne inférieure d'approximabilité connue pour le problème MAX SMTI non-restreint.

Des approches basées sur la recherche locale ont aussi été présentées par Gelain et al. [12], mais aussi par Munera et al. [30] avec une recherche adaptative.

#### Égalités d'un seul côté

Dans le cas particulier où les préférences sont strictement ordonnées d'un côté, d'autres algorithmes avec de meilleurs ratios existent. C'est le cas de GSA-LP, proposé par Iwama [19] qui utilise une relaxation linéaire de la modélisation mathématique du problème, Lam et Plaxton [23] ont amélioré cet algorithme pour obtenir un ratio d'approximation à  $1 + 1/e$ . Ces derniers ont également obtenu des résultats en bornant la taille des égalités pour obtenir un ratio d'approximation de  $1 + (1 - 1/L)^L$ , avec  $L$  la taille maximale des égalités [24]. D'autres algorithmes existent également dans le cas où les égalités sont de longueur 2 au maximum [14].

Huang et Kavitha [15] ont proposé un algorithme se basant sur l'algorithme GSA1 de Király avec un ratio qu'ils ont estimé à  $22/15$ . Une étude plus récente de l'algorithme a démontré que l'algorithme était en fait plus performant avec un ratio de  $13/9$  [1].

Tous ces algorithmes prennent en compte des égalités du côté "femme" uniquement, c'est à dire que les préférences des "proposants" (ou des hommes) doivent être strictement ordonnées.

En ce qui concerne le cas de MAX HRT, Irving et Manlove [18] ont proposé deux heuristiques (algorithmes H et R) pour le cas où les égalités sont du côté "hospitals" uniquement. L'algorithme R utilise une

modélisation à base de flots et est celui qui a présenté les meilleurs résultats empiriques comparé à l'algorithme H et celui de Király sur des instances réelles et arbitraires. Les résultats fournis par l'algorithme R sont empiriquement très bons et dépassent en moyenne les algorithmes d'approximation connus.

Comme précédemment, les proposant doivent avoir leurs listes strictement ordonnées, ce sont donc les internes qui appellent les hôpitaux au fur et à mesure.

# Chapitre 2

## Parcoursup

### 2.1 Admission Post-Bac et loi ORE

Parcoursup est une plateforme permettant de gérer l'accès en première année d'études supérieures au niveau national. Avant sa mise en place en 2018, la plateforme d'appariement des étudiants aux formations était Admission Post Bac ou APB. Cette plateforme, déployée en 2009, demandait à tous les candidats souhaitant accéder à des établissements d'enseignement supérieur de formuler des vœux et de les trier par ordre de préférence. De même, les responsables de formations sélectives pouvaient trier les candidats, mais ils pouvaient également en refuser certains dont le profil ne leur convenait pas. Une fois ces classements effectués, un algorithme de Gale-Shapley était lancé avec toutes les listes de préférences fournies. Les candidats recevaient donc au plus une proposition d'admission, qui correspondait à leur partenaire dans l'affectation stable optimale pour les formations. Il était alors possible pour les candidats d'accepter ou de refuser ces propositions. Les refus (en général, largement minoritaires) créaient des places vacantes qui étaient remises en jeu dans deux phases d'admission supplémentaires, correspondant à deux nouvelles exécutions de l'algorithme de Gale-Shapley avec des places libérées.

L'une des principales critiques d'APB ne provenait pas du système d'affectation en lui-même mais des contraintes fixées par la loi pour les formations non-sélectives. L'accès à l'enseignement supérieur pour les détenteurs d'un baccalauréat étant un droit, ces formations n'avaient pas à trier leurs candidats. Dès lors, APB ne fournissait pas aux formations les informations précises sur le profil des candidats, rendant la sélection impossible. Cependant, lors des dernières années de fonctionnement d'APB, certaines formations se sont retrouvées confrontées à un nombre de candidatures trop important par rapport à leur capacité d'accueil. Face à cela, les formations étaient obligées de classer les candidatures avec le peu d'informations qu'elles possédaient et s'en remettaient au tirage au sort lorsqu'il était impossible de départager pour les places restantes. D'autres problèmes comme la connaissance du rang des vœux des candidats par les formations ou l'opacité de la procédure ont été soulevés.

En réaction au tirage au sort, considéré comme inadmissible par le gouvernement, et à une mise en demeure<sup>1</sup> faite par la Commission nationale de l'informatique et des libertés (CNIL), le gouvernement a mis en place la loi relative à l'orientation et à la réussite des étudiants (ORE) permettant notamment à n'importe quelle formation de faire un classement de ses candidats lorsque sa capacité d'accueil est insuffisante. Ainsi, il n'est plus nécessaire d'avoir recours à un tirage au sort pour les formations en tension, pratique considérée comme contraire au principe de mérite [6]. Ce changement devait permettre de garantir un traitement plus juste (au sens de la méritocratie) et donc conduire à de meilleures affectations. De plus, l'établissement des classements ne doit plus être entièrement automatisé mais doit intégrer une intervention humaine, ce qui n'était pas le cas avec les tirages aléatoires. Ce point, relevé par la mise en demeure de la CNIL, met en lumière l'article 10 de la loi relative à l'informatique aux fichiers et aux libertés<sup>2</sup> qu'APB ne respectait pas.

---

1. [Décision n° MED-2017-053 du 30 août 2017 mettant en demeure X](#)

2. [Article 10 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés](#)



FIGURE 2.1 – Étapes principales de Parcoursup 2022 pour les candidats

Pour accompagner la loi ORE, Admission Post Bac a été remplacé par un tout nouveau système, Parcoursup.

## 2.2 Fonctionnement de la plateforme

### 2.2.1 Première partie : vœux des candidats

La première étape dans le processus de Parcoursup est le choix des formations souhaitées par les candidats. Dès décembre, les candidats peuvent accéder à un catalogue en ligne<sup>1</sup> répertoriant toutes les formations présentes à cet instant (certaines peuvent se rajouter au fur et à mesure). Dans ce catalogue, on peut retrouver différentes informations concernant le contenu de la formation, les attendus, les débouchés, les informations pratiques... Si la formation était déjà présente sur Parcoursup l'année passée, quelques statistiques sont disponibles sur les admissions précédentes : le nombre total de candidats, la part des candidats ayant reçu une proposition, la provenance des bacheliers (filières générale, technologique et professionnelle) ainsi que les différents quotas prévus par la loi (voir section 3.1.2).

Les responsables d'admission sont également tenus d'indiquer sur quels critères généraux les candidatures seront examinées et quelle importance est attribuée à chaque critère. Ces éléments sont donnés de manière indicative et peuvent être plus ou moins détaillés selon les établissements.

A partir de fin janvier, les candidats peuvent commencer à s'inscrire sur la plateforme et à formuler leurs vœux de formation. Ils sont limités à 10 vœux (+10 vœux en apprentissage) par personne mais certains de ceux-là peuvent être des vœux multiples. Un vœu multiple correspond à une candidature à un ensemble de formations ; si une personne candidate à deux formations du même ensemble alors cela comptera pour 1 vœu multiple comprenant 2 sous-vœux. Les vœux multiples existent sous plusieurs formes : un vœu pour une formation avec demande d'internat et un autre sans demande d'internat, des vœux pour plusieurs IEP, des vœux pour plusieurs formations d'ingénieur passant par le même concours... En tout, les candidats sont également limités à 20 sous-vœux chacun pour des formations hors apprentissage.

Les candidats ont ainsi jusqu'à début avril pour compléter leurs listes de vœux, rédiger les lettres de motivation nécessaires et confirmer finalement leurs vœux afin que leurs dossiers soient envoyés aux établissements correspondants à leurs formations souhaitées. Contrairement à la procédure d'APB, les candidats n'ont pas d'ordre à établir sur leurs vœux, ils confirment seulement un ensemble de vœux équivalents.

### 2.2.2 Deuxième partie : préférences des formations

En avril, les établissements de formation reçoivent les dossiers de tous les candidats ayant formulé un vœu pour une de leurs formations. Depuis la loi ORE, toutes les formations (même non-sélectives) ont accès aux informations pédagogiques sur leurs candidats (relevés de notes, appréciations du conseil de classe, notes du bac) ainsi qu'au profil du candidat (CV, lettre de motivation, lycée d'origine). Sauf cas particulier<sup>2</sup>, l'identité du candidat (nom, prénom, adresse) n'est pas transmise aux établissements. Les établissements peuvent ainsi classer les candidats selon les critères définis au préalable en utilisant soit un algorithme local, soit l'outil d'aide à la décision (OAD) fourni par Parcoursup<sup>3</sup>. Ce dernier permet de paramétrer la pondération des différents critères, et requiert également de rentrer une note correspondant au profil du candidat basée sur l'étude de son CV, sa lettre de motivation et ses appréciations. Les formations dites « sélectives » peuvent

1. <https://dossier.parcoursup.fr/Candidat/carte>

2. Dans le cas d'une demande d'internat ou d'alternance, les identités des candidats sont transmises. C'est également le cas pour des formations où l'admission se fait par concours ou entretien afin de convoquer les candidats.

3. [https://services.dgesip.fr/fichiers/Formation\\_OAD\\_2022.pdf](https://services.dgesip.fr/fichiers/Formation_OAD_2022.pdf)

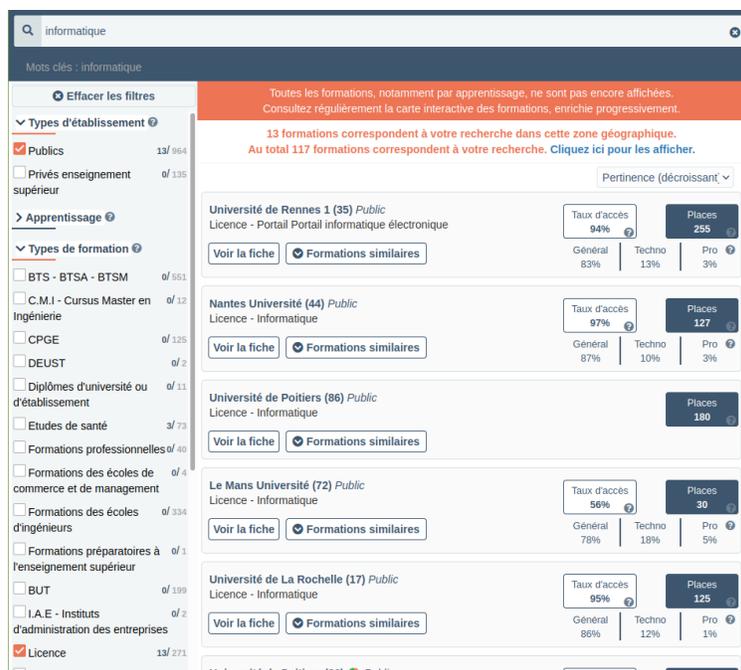


FIGURE 2.2 – Catalogue des formations de Parcoursup

également refuser des candidats ayant un niveau qu’elles considèrent insuffisant et n’ont donc pas à les classer. Quelle que soit la méthode choisie, les établissements doivent fournir un classement strict, i.e. ordre total qu’on appellera « classement pédagogique » des étudiants qu’ils jugent acceptables avant fin mai afin que les propositions puissent commencer à être envoyées début juin.

Certaines formations peuvent également émettre une condition à l’admission d’un candidat, et fournir ce que l’on appelle une réponse "Oui-si". Ces candidats font partie du classement de la formation mais celle-ci émet une condition à l’acceptation du candidat, le plus souvent cela correspond pour le candidat à un suivi de cours supplémentaires de remise à niveau.

Ces classements, ainsi que les différents coefficients utilisés pour les établir, ne sont pas accessibles publiquement. Certaines formations comme STAPS ont cependant publié leur algorithme local<sup>1</sup> afin que chaque étudiant puisse calculer son nombre total de points selon son relevé de notes.

Avant que les propositions soient envoyées aux candidats, Parcoursup va automatiquement modifier les classements pédagogiques fournis par les établissements afin que les différents quotas légaux soient respectés afin d’établir l’ordre d’appel. Ces modifications consistent en une remontée dans les classements des candidats boursiers ou résidents dans la même zone géographique et sont détaillées dans la documentation de Parcoursup<sup>2</sup>.

### 2.2.3 Troisième partie : appel des formations aux candidats

La troisième partie consiste à un envoi progressif de propositions d’admissions de la part des formations aux candidats au début de juin. L’ordre d’appel étant calculé et vérifié, les propositions d’admission commencent à être envoyées selon cet ordre. Le premier jour, chaque formation envoie un nombre fixe de propositions d’admission correspondant aux premiers candidats sur sa liste d’appel. Ce nombre correspond à la capacité d’admission de la formation, auquel est éventuellement rajouté un nombre d’admissions supplémentaires afin de faire du surbooking (voir section 3.3.1). Les candidats concernés reçoivent ces propositions

1. Algorithmes de Parcoursup 2021, Modalités de classement des nouveaux bacs - C3D STAPS

2. Document de présentation des algorithmes de Parcoursup - MESRI

et décident s'ils les conservent ou s'ils les rejettent. Ils ont droit à plusieurs jours pour délibérer (cinq jours au début puis progressivement deux) pour ne garder au maximum qu'une proposition active et ainsi faire une croix sur les autres. Ils peuvent également décider de conserver définitivement une de leurs propositions d'admission, se retirant ainsi de toutes les autres listes sur lesquelles ils étaient présents. Lorsqu'une proposition est rejetée par un candidat, elle est envoyée le lendemain au premier candidat n'en ayant pas reçu dans l'ordre d'appel de la formation. Dès qu'un candidat reçoit une nouvelle proposition, il est notifié par e-mail, sms ou via l'application mobile et est invité à choisir la proposition qu'il souhaite conserver (s'il en possède plus d'une) dans les délais impartis. Ainsi au fur et à mesure, les propositions vont être envoyées en parcourant la liste de l'ordre d'appel de haut en bas. Quoiqu'il arrive, cette procédure se termine à un instant fixe : en 2022, la phase principale de Parcoursup se termine le 15 juillet. Les candidats sont ainsi invités à s'inscrire administrativement dans la formation qu'ils ont conservé pendant la phase principale.

Cependant, après cette date les vœux en attente de chaque candidat sont archivés dans le cadre de la procédure de gestion des désistements (GDD) : au cas où un candidat se désiste finalement de la formation à laquelle il était affecté, une proposition est envoyée au suivant dans sa liste. Des propositions sont donc tout de même envoyées jusqu'en septembre, soit deux mois après la date de fin officielle communiquée dans le cadre de la phase principale. La différence entre avant et après la date de fin de la procédure principale est assez fine, les candidats n'ont simplement plus accès à leurs classements dans les listes d'attente et ne peuvent plus refuser de vœu.

Les candidats ont aussi dès le début de cette phase la possibilité d'utiliser un outil de répondeur automatique leur permettant d'ordonner leurs vœux en attente afin que les réponses aux formations soient générées directement par Parcoursup. Si une formation envoie une proposition à un candidat qu'il a classé au-dessus de sa proposition actuelle, alors une réponse positive est directement envoyée à cette formation.

## 2.2.4 Partie annexe : la phase complémentaire

En complément de l'appel des formations dans la phase dite principale, une phase complémentaire vient s'ajouter deux semaines après son commencement et se termine en septembre. Cette phase est destinée aux candidats n'ayant pas obtenu d'affectations satisfaisantes dans la procédure principale et qui souhaiteraient accéder à de nouvelles formations qu'ils n'avaient pas considéré dans un premier temps. Les formations éligibles à être proposées dans cette phase sont celles ayant appelé tous leurs candidats initiaux mais à qui il reste des places vacantes, ce sont donc ces places qui sont à pourvoir en phase complémentaire. Un candidat peut alors sélectionner une de ces formations, et elle s'ajoutera à ses vœux déjà effectués en phase principale.

Dès qu'un vœu a été formulé pour une formation, si cette dernière est sélective ou propose des aménagements (réponse "Oui-si"), elle est notifiée et possède 8 jours pour émettre une décision sur ce vœu et envoyer ou non au candidat une proposition. Pour les autres formations, une réponse est générée automatiquement au candidat le jour suivant. Le candidat possède enfin quelques jours pour se décider si une proposition lui a été faite : s'il avait une autre proposition en attente (en phase principale ou complémentaire), il doit décider laquelle il souhaite conserver. Une fois que la formation a envoyé toutes ses propositions, elle quitte la phase complémentaire mais peut revenir à tout moment si un ou plusieurs étudiants rejettent la proposition qu'ils avaient conservé.

## 2.3 Parcoursup et les mariages stables

Même si Parcoursup se différencie d'APB en partie par l'absence de classement de vœux et la non-utilisation d'un algorithme d'appariement, les mécanismes d'affectation utilisés se rapprochent fortement des algorithmes classiques de mariages stables. A l'instar d'APB, Parcoursup utilise en effet un algorithme de Gale-Shapley, effectué au travers de la procédure détaillée précédemment.

Pour synthétiser le principe de l'algorithme de Gale-Shapley (algorithme 1), il consiste au départ à ce que chaque homme envoie une proposition à sa femme préférée. Pour les besoins de l'explication, on considère que ces envois sont tous réalisés avant que les femmes puissent répondre. Les femmes ayant reçu au moins une proposition peuvent ainsi conserver leur préférée et supprimer de leurs listes les hommes qu'elles apprécient

moins. Chaque homme non retenu va ensuite envoyer une proposition à la femme suivante dans sa liste qui ne l'a pas supprimé de la sienne. Les femmes acceptent ainsi ces nouvelles propositions et rejettent les anciennes, et la boucle continue... L'algorithme s'arrête quand chaque homme est affecté ou a vidé sa liste.

En suivant cet algorithme appliqué au cas où les hommes sont des formations et les femmes sont des candidats, on retrouve bien les mêmes principes que décrits au point précédent. Comme le cas de Parcoursup se rapproche plutôt du problème Hospitals/Residents, les formations appellent plusieurs candidats d'un coup et non un seul. La différence principale entre l'algorithme de Gale-Shapley et la procédure de Parcoursup réside dans la manière d'exécuter l'algorithme : la première peut se dérouler automatiquement sur un ordinateur en ayant connaissance des listes de préférence de tout le monde au préalable, tandis que la deuxième se déroule sur plusieurs semaines, les candidats faisant avancer eux même la procédure en acceptant et refusant des formations.

Même si les candidats n'ont pas à rentrer de liste de préférences en amont, l'action de refuser une formation au profit d'une autre définit implicitement une hiérarchie entre ces deux vœux. Pendant la procédure, les candidats forment donc peu à peu leurs listes de préférence au gré des propositions envoyées par les formations, ce qui crée des classements partiels. Le fait que le classement ne soit pas ordonné totalement n'est pas important ici car ce que doit connaître l'algorithme est le vœu préféré d'un candidat parmi un sous-ensemble de ses vœux, ce sous-ensemble étant les vœux pour lesquels le candidat a reçu une proposition.

Afin d'analyser les mécanismes d'affectation de Parcoursup, nous considérerons sa procédure comme un algorithme de Gale-Shapley pour le problème Hospitals/Residents. Les listes de préférences des candidats seront définies de la même manière même si elles ne sont jamais définies explicitement sur Parcoursup, elles permettront de savoir quelle décision prendra un candidat lorsqu'il fait face à deux propositions ou plus.

On définit la vague d'appel initiale comme l'envoi simultané de propositions de la part des formations, où chaque formation envoie autant de propositions qu'elle possède de places. De la même manière, on définit les vagues de réponses comme les réponses aux propositions (acceptation/refus) faites simultanément par les candidats. Ces vagues sont seulement théoriques (dans un contexte réel, chaque candidat peut répondre à un moment différent) mais permettent de représenter simplement le déroulement global de la procédure.

## Chapitre 3

# Analyse critique de Parcoursup

Dans ce chapitre, nous cherchons à analyser les différentes caractéristiques présentées plus haut d'un point de vue technique mais également éthique. Les critères éthiques qui seront utilisés proviennent des lignes directrices pour une IA digne de confiance, rédigé par un groupe d'experts constitué par la commission européenne. Ce cadre a été choisi en particulier grâce à la globalité des sujets traités et du fait que le champ d'action de Parcoursup se situe au sein de l'Union Européenne. Les critères utilisés seront définis et mis en relation avec les différents points de discussion du chapitre.

### 3.1 Principes éthiques mis en avant dans Parcoursup

#### 3.1.1 Éthique de la stabilité

En complément du cadre éthique posé par le rapport de la commission européenne, nous pouvons nous questionner sur les valeurs morales que porte Parcoursup d'un point de vue technique. Le choix principal de la conception du système est que toute la procédure est basée sur une procédure de type Gale-Shapley, qui a pour objectif la stabilité de l'affectation obtenue. Mais quels sont les arguments justifiant la recherche de ce type de solutions ? Si l'on se retourne vers la définition standard de la stabilité, la seule chose que garantit une solution stable est l'absence de paire bloquante. Dans le cas de Parcoursup, une paire bloquante correspondrait à un candidat  $c$  qui n'a pas été affecté à une formation  $f$  alors que (1)  $c$  préfère  $f$  à ce qu'il a reçu, et (2)  $f$  préfère  $c$  à son pire candidat affecté. Dans un premier temps, on peut remarquer que ce critère est porté sur l'individuel : il est important que chaque candidat ne se retrouve pas dans une situation qu'il pourrait trouver injuste pour de bonnes raisons. Même si un candidat est très insatisfait par l'affectation produite pour son propre cas (par exemple : il n'est affecté nulle part), nous sommes sûrs que si la solution est stable, aucune formation qu'il préfère ne souhaiterait échanger un de ses étudiants pour lui. Selon ce critère, il est ainsi assuré qu'aucun candidat ne puisse obtenir mieux de manière simple. Dans un second temps, on peut interpréter une paire bloquante comme un candidat n'ayant pas obtenu une formation qu'il méritait pourtant. En effet, sachant que les formations produisent un classement basé sur des critères liés au travail et aux compétences de chacun, si un candidat souhaite obtenir une formation et que ses résultats le placent suffisamment hauts dans le classement de celle-ci, alors il devrait être en droit d'y accéder. On peut ainsi retrouver dans ce principe de stabilité appliqué à Parcoursup une envie de se conformer à un principe de méritocratie, ou du moins une volonté de distinguer les candidats selon leur mérite [10].

Par la suite, nous considérerons donc la méritocratie comme un critère éthique car c'est celui-ci qui est recherché au travers de l'utilisation d'une procédure de type Gale-Shapley. Même si ce principe peut être remis en question de plusieurs manières, nous cherchons simplement ici à produire une analyse critique cherchant à vérifier l'adhésion de Parcoursup à des principes éthiques choisis en amont. Le principe de méritocratie étant recherché de par le fonctionnement de la procédure, il est ainsi intéressant de l'y intégrer afin d'analyser la plateforme selon les intentions initiales de sa mise en place.

### 3.1.2 Quotas légaux

Afin de tendre vers des promotions avec des profils divers, le gouvernement a décidé d'imposer certains quotas et taux d'appels qui concernent différents types de candidats. Un taux d'appel minimum signifie que pour une formation donnée, on souhaite qu'une proportion fixe d'un certain type de candidats reçoive une proposition, et ce, à tout moment de la procédure. Il faut bien différencier ce taux avec le taux de candidats qui accepteront au final la proposition : Parcoursup assure simplement que la population appelée par chaque formation respecte certains taux.

Tout d'abord, les lycéens bénéficiant d'une bourse nationale de lycée (lycéens boursiers) sont concernés : chaque formation doit appeler un taux minimum de boursiers, ce taux est fonction de la proportion de candidats boursiers ayant postulé à cette formation. Pour assurer l'application de ce taux, Parcoursup a recours à un algorithme faisant remonter automatiquement les boursiers dans les classements s'il y en a trop peu en tête. La raison de l'existence de ce taux part du constat que ces candidats partent à la base avec ce qui est considéré un handicap social et économique pour accéder à l'enseignement supérieur. La mise en place de ce taux a donc pour but de contrebalancer ce désavantage en favorisant l'accès de ces bacheliers aux formations supérieures. Il peut de la même manière aider des boursiers à être acceptés dans des formations plus exigeantes. Ce taux est le seul à être appliqué pour toutes les formations, sélectives comme non-sélectives.

Les formations non-sélectives ont également un taux supplémentaire à prendre en compte, celui des résidents : chaque formation doit appeler au minimum un certain pourcentage de candidats résidant dans la même académie qu'elle. La méthode pour assurer un taux minimum de résidents est similaire à celle pour les boursiers. Ce taux a pour but de garantir l'accès à des formations proches du domicile des candidats afin de diminuer les cas où des candidats sont forcés à partir loin pour pouvoir étudier. Le rapport du CESP de 2021 a pris comme exemple la région parisienne qui voit l'arrivée d'un grand nombre d'étudiants venant d'autres régions à cause de l'attractivité des formations de la région. Les bacheliers franciliens se retrouvent donc en compétition avec de très bons élèves de province et sont donc pour certains d'entre eux obligés de quitter la région pour faire leurs études ailleurs.

Enfin dans le cas particulier des formations de Sections de Technicien Supérieur (STS) et des Instituts Universitaires de Technologie (IUT), des quotas sont appliqués en fonction du type de bac des candidats. Ainsi, les formations des IUT doivent accueillir au moins 50% de bacheliers technologiques<sup>1</sup>, et les STS ont également un certain pourcentage de bacheliers professionnels à accueillir, fixé par le recteur de l'académie. Le but de ces quotas est, comme pour le taux de boursiers, de faciliter l'accès à l'enseignement supérieur de ces profils qui ne sont souvent pas valorisés lors de l'admission dans ces établissements, malgré le fait que ces formations sont sensées être dans la continuité de ces types de bac.

Pour chacun de ces cas, le choix politique est de faire de la discrimination positive pour mettre en valeur des profils qui ont moins de chance d'être sélectionnés et garantir une meilleure répartition des places. Ces actions ont pour but d'offrir une plus grande égalité des chances : les candidats partant avec des handicaps sont revalorisés lors de la sélection par les établissements et sont sensés revenir à égalité avec les autres. Cette discrimination positive est cependant un point très critiqué par certains étudiants : des étudiants aux résultats moyens qui ne bénéficient pas de ces taux ont l'impression de se faire passer devant par des boursiers, des bacheliers technologiques... ce qui peut faire monter un sentiment d'injustice chez ces personnes.

## 3.2 Points positifs

### 3.2.1 Liberté d'action

#### D'un algorithme à une plateforme

---

1. [Article 17 de l'arrêté du 6 décembre 2019 portant réforme de la licence professionnelle](#)

**Critère 1** (Action humaine et contrôle humain). Les systèmes d'IA devraient soutenir l'autonomie et la prise de décisions humaines, conformément au principe du respect de l'autonomie humaine, en vertu duquel les systèmes d'IA devraient être à la fois les vecteurs d'une société démocratique, prospère et équitable en se mettant au service de l'utilisateur et favoriser les droits fondamentaux, ainsi que permettre un contrôle humain.

"Lignes directrices en matière d'éthique pour une IA de confiance"  
- Groupe d'Experts de Haut Niveau sur l'IA - Commission Européenne

Comme souligné dans la mise en demeure de la CNIL, le traitement totalement automatisé des candidatures par les formations est très critiquable sur le plan du contrôle humain. La prise de décision via un tirage au sort telle que parfois faite par APB ne respectait pas les principes énoncés par la loi relative à l'informatique, aux fichiers et aux libertés. En effet, le tirage au sort qui pouvait être fait dans les formations non sélectives en tension pour l'inscription de candidats dans l'enseignement supérieur relevait d'une décision purement algorithmique où aucun examen humain n'était effectué. Désormais, Parcoursup n'utilise plus le tirage au sort en permettant à toutes les formations de pouvoir classer leurs étudiants quand il y a trop de candidatures. De même, ces candidatures doivent depuis la loi ORE faire l'objet d'un traitement humain pour pouvoir être classées, ce qui peut néanmoins être contraignant pour l'examen des vœux (voir section 3.4).

Le passage d'APB à Parcoursup a également grandement changé la place des étudiants dans le processus d'affectation. En effet, la mise en place de ce système permet aux étudiants de faire leurs choix eux-même en répondant directement aux propositions qui leur sont faites. Ils peuvent également décider au bout d'un certain temps de laisser un algorithme, nommé le répondeur automatique, répondre aux propositions qui lui sont faites en ayant fourni au préalable un classement des vœux en attente. De plus, ils ont accès à un grand nombre d'informations comparé à APB. Que ce soit sur les formations et leurs attendus ou encore sur les places en temps réel dans la liste d'attente, les candidats possèdent un grand nombre de renseignements leur permettant de mieux faire leurs choix et de mieux comprendre leurs situations respectives.

La plateforme devient un outil très dépendant des candidats, ces derniers peuvent prendre des décisions de manière autonome selon leurs préférences actuelles. Parcoursup permet ainsi d'être simplement un support pour que les candidats puissent simplement faire leurs démarches plutôt qu'un algorithme statique, permettant également un accès plus égalitaire aux informations (voir section 3.2.2).

### Exemple : prise de décisions collectives

Si on admet qu'il est possible que des groupes de candidats puissent partager entre eux leurs situations respectives et qu'ils aient des intérêts à se retrouver affectés au même endroit, alors il leur est possible d'utiliser le système d'affectation de Parcoursup à leur avantage. La table 3.1 permet de visualiser un exemple de préférences pouvant être altérées par des éléments extérieurs susceptibles d'influencer les arbitrages individuels. On considère ici en bleu les propositions envoyées par les formations, et en rouge les propositions qui ont été refusées. Seul  $c_1$  a refusé deux de ses propositions à cet instant. Nous allons nous intéresser au cas de  $c_2$  et  $c_5$  qui vont constituer un groupe de candidats souhaitant se retrouver dans la même formation. Actuellement,  $c_2$  a reçu des propositions pour tous ses vœux mais  $c_5$  n'a reçu qu'une proposition de  $f_1$ .

Malgré le fait que  $f_1$  soit la formation la moins désirée par  $c_2$  et  $c_5$  et qu'ils l'auraient placé en dernière position s'ils avaient été amenés à classer leurs vœux, le fait que ce soit la seule formation qui ait proposé une place à chacun pourrait les amener à reconsidérer leurs préférences afin de se retrouver ensemble. De manière générale, un groupe de candidats pourrait aussi vouloir intégrer un même sous-ensemble de formations (qui seraient dans une même ville ou une même région par exemple), dans ce cas là il est intéressant pour eux de ne pas avoir trié leurs vœux à priori. On pourrait également penser à d'autres comportements :  $c_2$  pourrait par exemple refuser  $f_2$  pour laisser sa place à  $c_5$ , le fait de savoir que  $c_2$  est premier dans la liste d'attente assure qu'il recevra forcément une proposition de  $f_2$  si  $c_2$  a refusé la sienne. Même si ce cas a peu de chances de se produire avec des listes d'attente composées de milliers de candidats, on peut penser qu'il existe d'autres configurations particulières pouvant pousser les candidats à modifier leurs préférences initiales.

|     |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|     |       |       |       |       |       |       |       |       |       | $c_1$ | $f_4$ | $f_2$ | $f_1$ |       |
|     |       |       |       |       |       |       |       |       |       | $c_2$ | $f_2$ | $f_3$ | $f_1$ |       |
|     |       |       |       |       |       |       |       |       |       | $c_3$ | $f_4$ | $f_2$ | $f_3$ |       |
|     |       |       |       |       |       |       |       |       |       | $c_4$ | $f_2$ | $f_3$ |       |       |
| (4) | $f_1$ | $c_1$ | $c_2$ | $c_6$ | $c_5$ | $c_7$ | $c_8$ | $c_9$ |       | $c_5$ | $f_2$ | $f_3$ | $f_1$ |       |
| (2) | $f_2$ | $c_1$ | $c_2$ | $c_4$ | $c_5$ | $c_3$ | $c_6$ | $c_8$ | $c_9$ |       | $f_4$ | $f_2$ | $f_1$ |       |
| (2) | $f_3$ | $c_3$ | $c_2$ | $c_5$ | $c_4$ | $c_7$ | $c_9$ | $c_8$ |       |       | $f_4$ | $f_2$ | $f_1$ |       |
| (1) | $f_4$ | $c_1$ | $c_3$ | $c_6$ | $c_8$ |       |       |       |       |       | $f_3$ | $f_1$ |       |       |
|     |       |       |       |       |       |       |       |       |       | $c_6$ | $f_4$ | $f_2$ | $f_1$ |       |
|     |       |       |       |       |       |       |       |       |       | $c_7$ | $f_3$ | $f_1$ |       |       |
|     |       |       |       |       |       |       |       |       |       | $c_8$ | $f_4$ | $f_3$ | $f_2$ | $f_1$ |
|     |       |       |       |       |       |       |       |       |       | $c_9$ | $f_2$ | $f_3$ | $f_1$ |       |

TABLE 3.1 – Exemple de décisions collectives :  $c_2$  et  $c_5$  cherchent à se retrouver au même endroit

Toutes ces stratégies seraient difficilement envisageables si les préférences étaient figées et qu'un algorithme de Gale-Shapley était exécuté ensuite. Parcoursup autorise donc les candidats qui le souhaitent à prendre des décisions collectivement pour s'assurer qu'un groupe puisse obtenir des affectations qui satisfont tous les membres. Cette liberté peut même mieux affecter les autres candidats par répercussion : comme  $c_2$  et  $c_5$  vont choisir  $f_1$  qui est une formation globalement peu appréciée par les candidats, ils libéreront respectivement une place dans  $f_2$  et une place dans  $f_3$  qui seront proposées à des candidats qui ne les auraient pas reçu autrement. Plus généralement, de très nombreux facteurs sont susceptibles de faire évoluer les représentations que les candidats se font des formations (forum de discussion, promotion des établissements, critères géographiques et financiers, etc...), et donc les classements retenus. C'est pour leur laisser le temps d'affiner les informations reçues que le classement est retardé par rapport à APB, les ajustements, non modélisés dans la démarche algorithmique, étant alors à la charge des candidats en temps réel face aux réponses reçues.

Ce système peut donc avoir des effets vertueux pour bon nombre de candidats, mais ces bienfaits sont moins évidents pour les formations. On pourrait légitimement trouver injuste qu'une formation accueille des candidats qu'elle apprécie moins à cause de candidats mieux classés ayant préféré se tourner vers une formation qui pouvait accueillir tous leurs groupes d'amis. Dans notre exemple,  $f_2$  et  $f_3$  se retrouveront avec des candidats moins bien classés que  $c_2$  et  $c_5$  alors que ces derniers les préféreraient initialement. En ça la stabilité peut aussi être questionnée, car même si chacun de ces candidats choisit sa meilleure option à un instant donné, la solution ne sera pas stable selon les préférences individuelles de chacun.

### 3.2.2 Simplicité de la procédure

**Critère 2** (Explicabilité). L'explicabilité est essentielle pour renforcer et conserver la confiance des utilisateurs envers les systèmes d'IA. Cela signifie que les processus doivent être transparents, que les capacités et la finalité des systèmes d'IA doivent être communiquées ouvertement, et que les décisions – dans la mesure du possible – doivent pouvoir être expliquées aux personnes directement et indirectement concernées. [...]

L'explicabilité technique suppose que les décisions prises par un système d'IA puissent être comprises et retracées par des êtres humains. Par ailleurs, des arbitrages peuvent s'avérer nécessaires entre le renforcement de l'explicabilité d'un système (qui pourrait réduire sa précision) et l'amélioration de sa précision (au détriment de l'explicabilité). Dès qu'un système d'IA a une incidence importante sur la vie des personnes, il devrait être possible d'exiger une explication appropriée du processus de décision du système d'IA. Ces explications devraient être présentées en temps opportun et adaptées à l'expertise de la partie prenante concernée (par exemple, non-spécialiste, autorité de réglementation ou chercheur).

"Lignes directrices en matière d'éthique pour une IA de confiance"  
- Groupe d'Experts de Haut Niveau sur l'IA - Commission Européenne

Même si Parcoursup est une plateforme récente et unique en son genre, ses mécanismes ont beaucoup de

points communs avec ce qui pouvait être fait avant même l'utilisation d'une plateforme centrale d'affectation. Les procédures d'envoi de candidatures à chaque formation et l'attente des réponses que pouvaient connaître les étudiants avant la mise en place d'APB restent en effet d'actualité avec Parcoursup. Même si le passage par une procédure centralisée change beaucoup d'aspects, notamment les délais et le fait de pouvoir "conserver" une réponse positive, le principe très simple de demande/réponse est inchangé. La procédure se rapproche fortement d'un algorithme de Gale-Shapley mais une personne n'a nul besoin de le connaître pour comprendre les mécanismes de Parcoursup.

Cette simplicité est un point essentiel car Parcoursup s'adresse à un très grand nombre de personnes non-expertes et fournit un résultat affectant grandement la vie de ses utilisateurs. Les candidats peuvent simplement comprendre les enjeux liés à leurs choix de vœux et leurs décisions ; les encadrants dans les lycées et les écoles peuvent être également plus à même d'expliquer le fonctionnement global de la plateforme à des élèves. De même, l'explicabilité du système permet de tendre vers une compréhension globale identique pour tous ses utilisateurs quel que soit leur degré d'expertise informatique, ce qui renforce le principe d'égalité d'accès à l'information. Ainsi, chacun peut par exemple faire un recours quant à l'affectation qui lui a été faite par Parcoursup en toute connaissance du fonctionnement du système.

### 3.2.3 Maturation du choix

L'argument principal porté par le CESP en faveur de Parcoursup est la possibilité pour les candidats de mûrir leurs choix d'orientation. Là où un système plus traditionnel demanderait le préférences de tout les candidats jusqu'à un certain temps après lequel aucune modification n'est permise, Parcoursup permet de délayer ce choix sur une plus grande durée. Si l'on prend comme exemple APB, les candidats pouvaient trier leurs vœux jusqu'à fin mai, désormais un candidat peut recevoir des propositions pour ses vœux jusqu'à mi-juillet. Lors de la dernière campagne d'APB, les candidats pouvaient réordonner leurs choix jusqu'au 31 mai au plus tard, tandis qu'en 2022 les candidats devront répondre aux premières propositions avant le 7 juin. Cette fin de période plus éloignée offre à l'étudiant un temps de réflexion supplémentaire qui peut être utile pour revoir ses choix et mûrir son projet de formation.

Seulement, il existe une certaine inégalité sur ce point. Déjà, les propositions peuvent arriver à n'importe quel moment de la troisième partie, une fin de phase principale en juillet ne garantit donc pas qu'un candidat n'aura pas déjà eu à faire tous ses choix au début de juin. De plus, certains candidats obtiendront des propositions beaucoup plus rapidement que d'autres : les étudiants bien classés dans les listes d'appel recevront plusieurs propositions d'admission en début de procédure, et ils auront moins d'une semaine pour délibérer. A l'inverse, des étudiants moins bien classés recevront leurs propositions au fur et à mesure et auront donc en tout plus de temps pour se décider. L'exemple présenté à la table 3.2 met en lumière ces disparités. Ici, on peut remarquer que les candidats avec un faible indice sont en général préférés par les formations. Les propositions envoyées lors de la première vague d'appel sont représentées en bleu : on peut voir que  $c_1$  et  $c_2$ , candidats parmi les mieux classés par les formations, doivent directement choisir leurs formations définitives dès la première vague de réponses, tandis qu'aucun autre n'a à prendre de décision. Le temps de maturation disponible pour chaque candidat peut donc être très variable selon les cas et est en général moins long pour les candidats les mieux classés.

## 3.3 Limites

### 3.3.1 Sous-optimalité

#### Non-stabilité de la solution

La procédure principale de Parcoursup ressemble fortement à l'algorithme de Gale-Shapley appliqué au problème Hospitals-Residents : les formations proposent des places à leurs candidats les mieux placés, les candidats ne conservent à la fois qu'une proposition et libèrent des places dans les formations pour lesquelles ils déclinent une proposition. Parcoursup propose donc aux étudiants d'être les acteurs de cet algorithme en leur faisant choisir entre plusieurs propositions au fil du temps. Comme on souhaite que

|     |       |       |       |       |       |       |       |       |       |       |       |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| (4) | $f_1$ | $c_1$ | $c_2$ | $c_6$ | $c_5$ | $c_7$ | $c_8$ | $c_9$ |       | $f_4$ | $f_2$ | $f_1$ |       |
| (2) | $f_2$ | $c_1$ | $c_2$ | $c_4$ | $c_5$ | $c_3$ | $c_6$ | $c_8$ | $c_9$ | $f_2$ | $f_3$ | $f_1$ |       |
| (2) | $f_3$ | $c_3$ | $c_2$ | $c_5$ | $c_4$ | $c_7$ | $c_9$ | $c_8$ |       | $f_4$ | $f_2$ | $f_3$ |       |
| (1) | $f_4$ | $c_1$ | $c_3$ | $c_6$ | $c_8$ |       |       |       |       | $f_2$ | $f_3$ | $f_1$ |       |
|     |       | $c_1$ | $c_3$ | $c_6$ | $c_8$ |       |       |       |       | $f_3$ | $f_1$ |       |       |
|     |       | $c_1$ | $c_3$ | $c_6$ | $c_8$ |       |       |       |       | $f_4$ | $f_3$ | $f_2$ | $f_1$ |
|     |       | $c_1$ | $c_3$ | $c_6$ | $c_8$ |       |       |       |       | $f_2$ | $f_3$ | $f_1$ |       |

TABLE 3.2 – Vague d’appel initiale de Parcoursup

|     |       |       |       |       |       |       |       |       |  |       |       |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|--|-------|-------|-------|-------|
| (1) | $f_1$ | $c_1$ | $c_2$ | $c_4$ | $c_5$ |       |       |       |  | $f_1$ | $f_3$ | $f_2$ |       |
| (1) | $f_2$ | $c_1$ | $c_2$ | $c_6$ | $c_3$ | $c_7$ | $c_9$ |       |  | $f_1$ | $f_4$ | $f_2$ | $f_3$ |
| (1) | $f_3$ | $c_3$ | $c_1$ | $c_2$ | $c_4$ | $c_5$ | $c_8$ | $c_7$ |  | $f_2$ | $f_5$ | $f_3$ | $f_6$ |
| (2) | $f_4$ | $c_2$ | $c_5$ | $c_9$ |       |       |       |       |  | $f_1$ | $f_4$ | $f_3$ | $f_6$ |
| (2) | $f_5$ | $c_4$ | $c_3$ | $c_7$ | $c_6$ | $c_8$ |       |       |  | $f_5$ | $f_6$ | $f_2$ |       |
| (2) | $f_6$ | $c_4$ | $c_3$ | $c_5$ | $c_8$ | $c_6$ | $c_9$ |       |  | $f_2$ | $f_3$ | $f_5$ |       |
|     |       | $c_4$ | $c_3$ | $c_5$ | $c_8$ | $c_6$ | $c_9$ |       |  | $f_3$ | $f_5$ | $f_6$ |       |
|     |       | $c_4$ | $c_3$ | $c_5$ | $c_8$ | $c_6$ | $c_9$ |       |  | $f_2$ | $f_4$ | $f_6$ |       |

(a) Affectation produite par Gale-Shapley (optimale pour les formations)

|     |       |       |       |       |       |       |       |       |  |       |       |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|--|-------|-------|-------|-------|
| (1) | $f_1$ | $c_1$ | $c_2$ | $c_4$ | $c_5$ |       |       |       |  | $f_1$ | $f_3$ | $f_2$ |       |
| (1) | $f_2$ | $c_1$ | $c_2$ | $c_6$ | $c_3$ | $c_7$ | $c_9$ |       |  | $f_1$ | $f_4$ | $f_2$ | $f_3$ |
| (1) | $f_3$ | $c_3$ | $c_1$ | $c_2$ | $c_4$ | $c_5$ | $c_8$ | $c_7$ |  | $f_2$ | $f_5$ | $f_3$ | $f_6$ |
| (2) | $f_4$ | $c_2$ | $c_5$ | $c_9$ |       |       |       |       |  | $f_1$ | $f_4$ | $f_3$ | $f_6$ |
| (2) | $f_5$ | $c_4$ | $c_3$ | $c_7$ | $c_6$ | $c_8$ |       |       |  | $f_5$ | $f_6$ | $f_2$ |       |
| (2) | $f_6$ | $c_4$ | $c_3$ | $c_5$ | $c_8$ | $c_6$ | $c_9$ |       |  | $f_2$ | $f_3$ | $f_5$ |       |
|     |       | $c_4$ | $c_3$ | $c_5$ | $c_8$ | $c_6$ | $c_9$ |       |  | $f_3$ | $f_5$ | $f_6$ |       |
|     |       | $c_4$ | $c_3$ | $c_5$ | $c_8$ | $c_6$ | $c_9$ |       |  | $f_2$ | $f_4$ | $f_6$ |       |

(b) Affectation produite par Gale-Shapley au fil de l’eau après 4 vagues de réponse

TABLE 3.3 – Différences d’affectation entre deux méthodes

chaque candidat dispose d’un temps minimum pour prendre sa décision, cela impose une temporalité très différente de l’algorithme de Gale-Shapley. Là où ce dernier dure moins d’une heure pour plusieurs centaines de milliers de candidats, la phase principale de Parcoursup s’arrête au bout d’un mois et demi sans garantie que les conditions d’arrêt classiques de l’algorithme soient respectées. Nous expliquerons ici en quoi les affectations faites à partir de la phase principale de Parcoursup peuvent également ne pas être stables puis observerons ce phénomène sur des données réelles.

Nous allons ici modéliser un cas de non-stabilité pouvant être obtenu avec un algorithme de Gale-Shapley incomplet en prenant un cas à 6 formations et 9 candidats. Les préférences de chacun sont exprimées en table 3.3. On suppose ici que la procédure s’arrête après la quatrième vague de réponses, l’affectation résultante est celle présente en table 3.3b et les résultats que l’on aurait obtenu après un algorithme de Gale-Shapley orienté du côté des formations si toutes les vagues avaient été déroulées sont représentés à la table 3.3a.

On peut remarquer assez facilement que la solution en 3.3b est sous-optimale comparé à celle fournie par Gale-Shapley. En effet, cette affectation n’est pas stable car  $\{f_3, c_7\}$  et  $\{f_3, c_8\}$  sont des paires bloquantes. En réalité si on avait laissé aux candidats 8 vagues de réponses possibles au lieu de 4, la solution aurait été

la même que Gale-Shapley car on serait arrivé au bout de l'algorithme (chaque formation est soit remplie ou soit à court de candidats à appeler).

Une observation intéressante que l'on peut tirer de cet exemple est le contraste entre la baisse de qualité globale des affectations du côté des étudiants et la hausse de la qualité des affectations pour une grande partie des formations. Toutes les formations ayant rempli autant que ce qu'elles auraient eu avec l'algorithme de Gale-Shapley peuvent obtenir de meilleurs candidats qu'avec ce dernier ; à l'opposé, les candidats ne peuvent obtenir qu'au maximum ce qu'ils auraient eu avec Gale-Shapley. Dans cet exemple,  $f_2$ ,  $f_5$  et  $f_6$  sont plus satisfaites par l'affectation en 3.3b,  $f_1$  et  $f_4$  sont autant satisfaites et seule  $f_3$  est moins satisfaite car elle perd une affectation. De l'autre côté,  $c_1$ ,  $c_2$ ,  $c_4$ ,  $c_5$  et  $c_7$  sont autant satisfaits mais  $c_3$ ,  $c_6$ ,  $c_8$  et  $c_9$  sont moins satisfaits, en particulier  $c_9$  qui n'est affecté nulle part.

Cette disparité peut s'expliquer par le déroulement de l'algorithme de Gale-Shapley : dans Parcoursup, les premiers candidats appelés sont ceux préférés par les formations, et au fur et à mesure du temps ils appellent des candidats de plus en plus en bas de liste. Les formations dégradent donc leurs potentielles affectations au fil du temps. Du côté des candidats, chaque proposition reçue leur permet potentiellement d'améliorer leur satisfaction si cette proposition est meilleure pour eux que ce qu'ils avaient conservé. Les candidats gagnent donc en satisfaction au fil du temps.

L'affectation renvoyée par Parcoursup après la phase principale est donc particulière :

- Sa stabilité n'est pas garantie car les formations n'ont pas nécessairement le temps d'envoyer des candidatures à tout le monde
- Elle défavorise globalement les candidats en les affectant comme dans l'algorithme de Gale-Shapley dans le meilleur cas jusqu'à ne plus du tout les affecter dans le pire cas. Le fait d'avoir des affectations en moins défavorise également les formations concernées mais la qualité globale du résultat est meilleure pour les formations.

De plus, ce système crée des places vacantes menaçant la pérennité des formations concernées, et donc plus globalement entraîne une baisse du nombre total d'étudiants admis. On constate également que l'intérêt collectif est que chacun réponde rapidement aux propositions faites et supprime le plus de voeux possibles de leurs listes pour accélérer la procédure, ce qui n'est individuellement pas préférable si l'on préfère se laisser le plus de temps et d'options disponibles. Il existe donc également des tensions entre intérêts individuels et intérêts collectifs pour les candidats.

### Observation sur les données de Parcoursup

Depuis la création de Parcoursup en 2018, les calendriers et les délais de chaque campagne ont été adaptés selon la situation. La période sur laquelle sont envoyées les proposition d'admission a été diminuée de moitié (3 mois pour la campagne 2019, 1 mois et demi pour 2022) suite aux différentes critiques liées à la longueur de la procédure et au stress engendré. Le délai dont dispose un candidat pour répondre à une proposition a donc également baissé en conséquence (7 jours en juin 2018, 2 jours en juin 2022). S'il était difficile de reconnaître des cas tels que vus précédemment, l'ouverture au public des données concernant les formations et leurs admissions a permis de repérer des cas montrant une sous-optimalité quasi certaine des affectations. Des exemples sont illustrés dans le rapport de 2021 du CESP concernant la campagne 2020, nous nous intéresserons ici aux données de la campagne 2021 qui a duré plus de 7 semaines et qui permettait aux candidats une réponse aux propositions à J+2 après réception (soit entre deux et trois jours de délai).

Nous nous intéressons ici à deux licences présentes sur Parcoursup en 2021 dont les données sont disponible à la table 3.4 et 3.5.

| Capacité | Candidats classés | Rang du dernier appelé | Admis |
|----------|-------------------|------------------------|-------|
| 518      | 3747              | 1875                   | 501   |

TABLE 3.4 – Données de la phase principale de Parcoursup 2021 sur la licence A

Sur cet exemple, la licence A n'a pu appeler des candidats que jusqu'au rang 1875, laissant 1872 candidats non appelés à la fin de la procédure principale. Cependant, il reste 17 places non attribuées qui ne pourront

pas être ouvertes à nouveau en phase complémentaire (voir section 3.3.2) et qui seront donc vacantes dès la rentrée.

| Capacité | Candidats classés | Rang du dernier appelé | Admis |
|----------|-------------------|------------------------|-------|
| 725      | 6202              | 2652                   | 695   |

TABLE 3.5 – Données de la phase principale de Parcoursup 2021 sur la licence B

Sur cet exemple, la licence B n’a pu appeler des candidats que jusqu’au rang 2652, laissant 3550 candidats non appelés à la fin de la procédure principale. Cependant, il reste 30 places non attribuées qui ne pourront pas être ouvertes à nouveau en phase complémentaire et qui seront donc vacantes dès la rentrée.

Ces deux exemples permettent d’illustrer que l’algorithme Gale-Shapley déroulé par les candidats n’était pas terminé lors de l’arrêt de la procédure principale le 16 juillet 2021 : il existe des formations en sous-effectif n’ayant pas appelé tous leurs candidats, la condition d’arrêt prévue n’est donc pas respectée.

## Expérimentation

Afin d’avoir un aperçu de l’impact que peut avoir un arrêt prématuré de l’algorithme de Gale-Shapley, nous utiliserons des instances générées aléatoirement. Les instances utilisées se basent sur le principe de préférences de popularité [27] : à chaque formation est assignée une certaine popularité  $p \in \mathbb{N}$ , et il en va de même pour les candidats. On souhaite qu’une formation avec une grande popularité soit choisie plus fréquemment et ait plus de chances d’être placée en tête de liste d’un candidat qu’une formation moins populaire. Ainsi, chaque candidat va constituer sa liste de vœux de taille  $n$  en tirant successivement  $n$  formations sans remise. La probabilité de tirer une formation est proportionnelle à sa popularité. Pour chaque candidat, la première formation tirée correspondra à son premier choix, la deuxième à son deuxième choix, etc...

Une fois que tous les candidats ont établi leurs listes de vœux, les formations doivent classer les candidatures ; de la même façon que les candidats, les formations vont tirer successivement chacune des candidatures, les candidats les plus populaires ayant plus de chances d’être placés haut. Dans ces instances, on ne considère pas les cas où les formations rejettent directement des étudiants comme les formations sélectives peuvent le faire avec Parcoursup.

Les instances utilisées ici comportent 100.000 candidats pour 100.000 places réparties entre 1.000 formations et chaque candidat formule en moyenne 12 vœux, le nombre de vœux de chaque candidat étant tiré aléatoirement selon une loi normale. Pour chaque instance, on mesure le nombre de candidats affectés selon l’avancement de l’algorithme. L’avancement correspond au nombre de vagues de réponses effectuées par rapport au nombre total de vagues nécessaires pour obtenir une affectation stable. Dans le pire cas, il est possible qu’il n’y ait que 48 vagues de réponses de juin à septembre (au minimum 15 vagues pendant la phase principale et 34 pendant la gestion des désistements) pour la campagne Parcoursup 2022, nos instances ont en moyenne un nombre de 372.8 vagues au maximum. Ainsi, si la procédure Parcoursup se déroulait avec l’une de nos instances, l’algorithme aurait progressé jusqu’à 12.9% en moyenne. La figure 3.1 compile nos résultats liés à l’avancement de la procédure.

La figure 3.1a montre que dès 5% de progression, plus de 90% des candidats qui seront affectés en fin d’algorithme ont reçu une proposition, autour de 13% d’avancement, environ 97% de ces candidats auront reçu une proposition. Cependant, est-ce que ce sont les meilleures propositions qu’ils peuvent obtenir ? La figure 3.1b montre qu’environ un quart des candidats n’ont pas reçu leur affectation définitive à 13% de progression. Afin d’atteindre le palier de 90% d’affectations définitives, cela nécessiterait environ deux fois plus de vagues de réponses.

On peut remarquer que lors de la deuxième moitié du déroulement de l’algorithme, moins de 0.1% des candidats qui seront affectés n’ont pas reçu d’affectations ; également, moins de 1% d’entre eux n’ont pas reçu leurs affectations définitives. Les vagues de réponses dans cette moitié sont moins grandes car les places

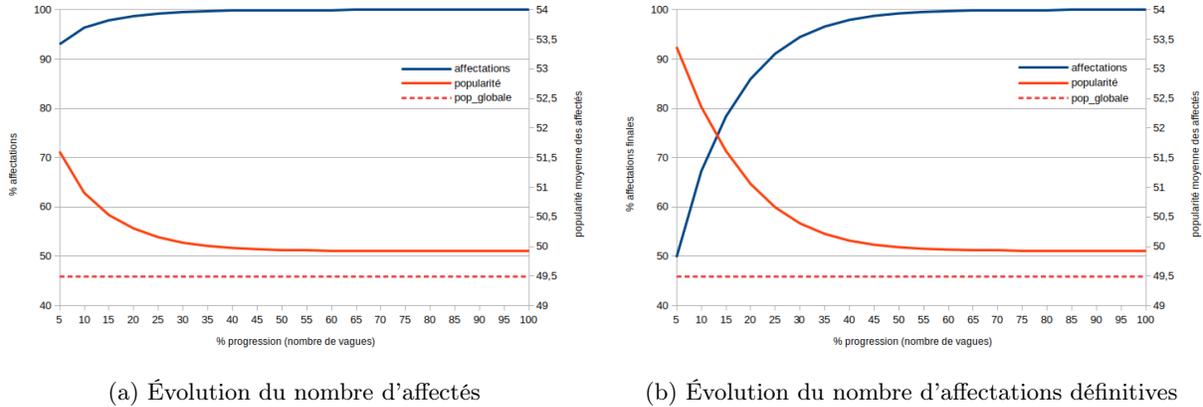


FIGURE 3.1 – Nombre de candidats affectés et leur popularité selon la progression de l'algorithme

disponibles se réduisent au fur et à mesure, ce qui conduit à moins d'appel de la part des formations. Selon le comportement des candidats (délais de réponses, suppression des voeux inutiles), la procédure peut avancer plus ou moins rapidement et créer ou non de meilleures affectations pour les candidats. Avec le fonctionnement actuel de Parcoursup, il est cependant impossible de garantir que la procédure atteigne plus de 48 vagues.

En ce qui concerne la popularité, on peut noter que les candidats les plus populaires sont en moyenne mieux affectés et de manière plus rapide. Cela va dans le même sens que l'exemple de la table 3.2 : dès les premières vagues, la popularité moyenne des affectés est de 53 contre un peu moins de 50 en fin de procédure. Les candidats affectés sont également plus populaires en moyenne que la totalité des candidats, ces derniers ayant une popularité de 49.5 en moyenne (représentée par des tirets).

Il est à noter que ces instances restent relativement éloignées de la réalité : la taille du problème est moins grande, les candidats font en moyenne plus de voeux, le nombre de places correspond exactement au nombre de candidats... Cependant, le déroulement d'un algorithme Gale-Shapley suit généralement cette évolution en affectant une grande partie des candidats dès les premières vagues pour n'améliorer sa solution que pour quelques candidats dans les dernières.

### Utilisation du surbooking

Pour diminuer l'impact de ce phénomène, il faut trouver un moyen d'appeler plus de candidats. Il serait possible par exemple d'augmenter le temps d'appel disponible, mais cela irait directement à l'encontre de ce qui est fait depuis plusieurs années pour réduire le temps total de la procédure (voir ultérieurement). Un autre moyen serait d'appeler plus de candidats durant cette période, et c'est précisément ce qui a été mis en place dans Parcoursup. La documentation de Parcoursup ainsi que la note de cadrage sur les données d'appel<sup>1</sup> détaillent les deux possibilités disponibles pour appeler plus de candidats : via l'appel d'un bloc de candidats, ou un taux d'appel supplémentaire.

Le premier consiste à déterminer au préalable quels candidats seront appelés dans les premiers jours. Par exemple, une formation A qui a une capacité de 100 places peut décider d'appeler un bloc de 600 candidats : les 600 premiers candidats de sa liste recevront donc dès le premier jour une proposition qui pourra être acceptée ou refusée. Cette procédure n'est accessible qu'aux formations qui ont au moins une année d'expérience sur la plateforme car, selon la note de cadrage, elle nécessite une connaissance précise du recrutement.

Le taux d'appel supplémentaire est accessible à toutes les formations et leur permet d'envoyer plus de propositions que leur nombre de places disponibles. Chaque formation peut décider de son propre taux, mais il est limité à 20% (ou 50% sur demande) de la capacité de la formation. Par exemple, la formation A qui

1. [https://services.dgesip.fr/fichiers/Guide\\_accueil\\_\\_\\_DA\\_ET\\_ADMISSION\\_MAI\\_2021.pdf](https://services.dgesip.fr/fichiers/Guide_accueil___DA_ET_ADMISSION_MAI_2021.pdf)

possède 100 places disponibles peut décider d'appeler au maximum 120 personnes à la fois. Ce taux est appliqué tout au long de la procédure et peut être modifié en cours si besoin.

Ces deux méthodes, pouvant être utilisées conjointement, peuvent permettre à une formation de remplir plus rapidement en accélérant le processus d'appel. Ces outils doivent cependant être manipulés avec prudence car tous les candidats appelés peuvent accepter les propositions qui leur sont faites, il est donc possible que des formations se retrouvent avec plus d'étudiants qu'elles ne puissent en accueillir. Les formations sont donc mises à contribution pour résoudre le problème de lenteur de la plateforme, mais l'incertitude qu'elles ont sur le comportement des candidats les force à prendre le risque de se retrouver avec des classes surchargées. Le CESP va néanmoins dans ce sens en proposant dans son rapport de 2021 de renforcer l'utilisation du surbooking et de mieux accompagner les formations dans ce processus dans le but d'appeler plus de candidats et de pourvoir plus de places. Cependant, même si les responsables d'admission sont sensibilisés à ce problème, déterminer le comportement d'un groupe de candidats a priori implique une incertitude relativement élevée.

De plus, comme illustré par la table 3.3, l'intérêt pour les formations de parcourir rapidement la liste est relativement restreint. Seules les formations ayant des difficultés à remplir seraient vraiment gagnantes à parcourir plus rapidement leurs listes de candidats car elles s'assureraient d'avoir plus de gens affectés. Pour ce qui est des autres, les changements d'affectations qu'implique un parcours rapide ne peut que détériorer leur satisfaction en leur faisant perdre des candidats bien classés. L'intérêt du surbooking est avant tout pour les candidats, pour qui son utilisation serait au contraire bénéfique. Cependant, ce système est piloté uniquement par les formations qui décident combien de candidats elles veulent appeler, sans aucune obligation d'en appeler plus que nécessaire. Un effort de sensibilisation et de responsabilisation doit donc être nécessaire pour que toutes les formations puissent coopérer à ce que la procédure progresse plus rapidement si l'on veut obtenir une solution stable et plus d'affectations.

A travers le constat de la sous-optimalité et des tentatives de contournement de ce problème, on peut en déduire que l'on touche à la limite de ce système d'affectation. Ce fonctionnement d'appel par les formations et de réponses par les candidats ne peut pas assurer une affectation globale stable et les tentatives d'amélioration de ce point amènent à considérer que l'on puisse accepter un surplus de candidats dans certaines formations afin d'accélérer le processus. Le risque de ne pas pouvoir appeler tous les candidats est inhérent à ce fonctionnement et les méthodes utilisées pour amoindrir ce risque peut engendrer de nouveaux problèmes au-delà de l'aspect technique.

Ce problème met en lumière la contradiction entre les différents critères éthiques : le système actuel donne un grand champ libre à l'action humaine et une bonne lisibilité mais en contrepartie fournit un résultat largement améliorable (notamment sur la stabilité et le nombre d'affectations) pour le plus grand nombre. Comme formulé dans la définition de l'explicabilité : « des arbitrages peuvent s'avérer nécessaires entre le renforcement de l'explicabilité d'un système (qui pourrait réduire sa précision <sup>1</sup>) et l'amélioration de sa précision (au détriment de l'explicabilité). »

### 3.3.2 Inéquité de la phase complémentaire

La phase complémentaire de Parcoursup a pour but d'affecter des candidats qui n'ont rien obtenu de satisfaisant dans des formations où des places sont disponibles. Les candidats peuvent formuler 10 nouveaux vœux (différents de leurs vœux formulés en phase principale) dans un catalogue de formations restreint. Ce catalogue est composé des formations qui ont appelé tous leurs candidats mais qui n'ont pas réussi à tout remplir. Les formations avec des places vacantes n'ayant pas appelé tous leurs candidats ne sont pas disponibles afin qu'aucun candidat ne puisse passer devant d'autres plus méritants. Pour illustrer ce fait, nous pouvons nous baser sur l'exemple présenté sur la table 3.3 en imaginant que la phase complémentaire se lance au bout de la quatrième période de réponse et que la phase principale s'interrompt. Grâce à la table 3.3b, nous pouvons voir que  $f_3$  est la seule formation à avoir une place vacante. Si cette formation était disponible en phase complémentaire, alors  $c_6$  et  $c_9$  pourraient formuler un nouveau vœu pour remplir cette formation, et si l'un des deux est accepté alors cela voudra dire qu'il sera passé devant tous les candidats

---

1. Ici, le sens de précision est celui utilisé dans le domaine du machine learning, dans le cas d'un algorithme d'appariement on peut remplacer ce terme par optimalité

$$(1) \quad f_1 \mid c_1 \quad c_2 \quad c_3 \quad \begin{array}{l|l} c_1 & f_1 \\ c_2 & f_1 \\ c_3 & f_1 \end{array}$$

TABLE 3.6 – Préférences (théoriques) des formations et des candidats en phase complémentaire

non-appelés, et notamment  $c_7$  et  $c_8$  qui préfèrent  $f_3$  à leurs affectations respectives. Cela peut créer une injustice car  $c_6$  et  $c_9$  seraient acceptables sans même avoir été comparés aux autres candidats qui auraient peut-être plus mérité leur place. Ainsi, c'est pourquoi ces formations là ne sont pas disponibles en phase complémentaire. On peut voir donc ici que le critère du mérite semble être plus important que celui du nombre de candidats placés en ne voulant pas placer plus de candidats pour qu'aucune injustice ne soit subie à ce niveau là.

Cette procédure commence avant que la phase principale ne se termine (en 2022, elle commence le 16 juin), un candidat peut donc cumuler des voeux en phase principale et en phase complémentaire en même temps. Cependant, si un candidat choisit de conserver une proposition d'une formation en phase complémentaire, il doit abandonner celles qu'il a déjà reçu en phase principale de manière à ne garder qu'un seul voeu en attente au total. Les délais de réponse pour les candidats sont similaires au début mais deviennent de plus en plus serrés au fil du temps (en 2021, les candidats avaient 3 jours de délai avant le 26 août et seulement 1 jour après cette date). Des formations peuvent se rajouter au fur et à mesure si elles ont appelé tous leurs candidats de la phase principale après le début la phase complémentaire.

Cette phase se différencie de la première par plusieurs points. Dans un premier temps, les candidats peuvent formuler des voeux à n'importe quel moment de la procédure. De même, des formations peuvent apparaître ou disparaître au fil de la procédure, rendant les choix possibles différents pour deux candidats se connectant à des moments différents.

Une fois qu'une formation a reçu une candidature, elle a également un délai de réponse défini selon son type (en 2021, ce délai était de 8 jours pour les formations sélectives ainsi que les formations non-sélectives proposant des "Oui-si"<sup>1</sup> et 1 jour pour les autres). Cependant, comme les formations non-sélectives ne proposant pas de réponse "Oui-si" ne refusent aucune candidature, celles-ci génèrent automatiquement une réponse le lendemain de la réception d'une proposition.

Si la phase principale a pour but d'obtenir une affectation stable, la phase complémentaire a plutôt pour objectif d'affecter le plus d'étudiants possible, et ce, sans que les préférences des formations aient réellement d'importance. Un problème éthique se pose en terme d'équité : les candidats formulant les premiers des voeux pour une formation seront avantagés par rapport aux autres. Ce problème est surtout flagrant pour les formations non-sélectives : une réponse est générée automatiquement à J+1 par rapport à la formulation d'un voeu. Ainsi, les candidats les plus rapides à formuler leurs voeux auront facilement une place, et ce, indépendamment de leurs dossiers. L'exemple de la table 3.6 illustre le fait qu'un candidat peu apprécié pourra avoir une place dans une formation, tandis que des candidats théoriquement mieux placés n'auront plus de place, ce qui constitue plusieurs paires bloquantes.

Dans cet exemple, trois candidats souhaiteraient une place dans la formation  $f_1$  mais il ne lui reste qu'une place en phase complémentaire. Les préférences (théoriques) de la formation sont représentées dans la table 3.6 et il est clair que  $\{f_1, c_1\}$  est la seule affectation stable. Cependant, si  $c_3$  est seul à postuler en premier, alors il sera pris par  $f_1$  et l'affectation sera donc  $\{f_1, c_3\}$ .  $f_1$  n'aura donc pas connaissance de  $c_1$  et  $c_2$  même s'ils auraient été préférés s'ils avaient pu postuler. Cette affectation contient deux paires bloquantes  $\{f_1, c_1\}$  et  $\{f_1, c_2\}$  et n'est donc clairement pas stable, et va donc à l'encontre du principe de mérite qui est pourtant recherché.

Ce type de cas peut se présenter par exemple car des candidats vérifient moins régulièrement que d'autres les formations disponibles en phase complémentaire, car ils arrivent dans la procédure après tout le monde

1. Ces formations nécessitent un examen humain des candidatures et disposent donc d'un mois de "gel" pendant l'été pendant lequel elles ne sont pas tenues de répondre aux candidatures qui leur sont adressées, cf. [arrêté relatif au calendrier de Parcoursup](#).

(c'est le cas notamment pour des étudiants en réorientation), ou car des candidats ne sont pas encore sûrs de leurs choix. On peut donc questionner le fonctionnement de cette procédure : certes, elle permet d'affecter plus rapidement des étudiants en complément de la phase principale mais cette rapidité implique une qualité d'affectation aléatoire (« premier arrivé, premier servi »). L'urgence de choisir de nouvelles formations que crée cette procédure pour un candidat est contradictoire avec la volonté que chacun puisse mûrir sa décision. Un candidat ne peut pas savoir à l'avance quelles formations seront disponibles en phase complémentaire, les décisions qu'il aura à prendre devront donc être rapides pour maximiser ses chances d'être sélectionné. De plus, le catalogue de formations disponibles lors de cette phase évolue chaque jour : des formations disparaissent mais d'autres apparaissent également. C'est le cas des formations ayant fini d'appeler tous leurs candidats en phase principale après le début de la phase complémentaire, mais aussi des formations ayant été précédemment remplies en phase complémentaire pour qui des places se sont à nouveau libérées. Pour maximiser ses chances de trouver une formation qui lui correspond en phase complémentaire, un candidat devra se connecter quotidiennement sur la plateforme afin de ne rien rater. Ce type de comportement que provoque l'aspect arbitraire de la procédure peut engendrer un stress sur une durée assez conséquente, ce qui est peu souhaitable (voir critère 3 ci-dessous).

### 3.3.3 Longueur de la procédure

**Critère 3** (Prévention de toute atteinte). Les systèmes d'IA ne devraient ni porter atteinte, ni aggraver toute atteinte portée, ni nuire aux êtres humains d'une quelconque autre manière. Cela englobe la protection de la dignité humaine ainsi que de l'intégrité mentale et physique.

"Lignes directrices en matière d'éthique pour une IA de confiance"  
- Groupe d'Experts de Haut Niveau sur l'IA - Commission Européenne

Une des principales critiques adressées de la part des candidats et de leurs parents concerne le stress qui peut être développé pendant toute la longueur de la procédure. Même si l'entrée dans les études supérieures peut être source d'anxiété pour de multiples facteurs indépendamment du système utilisé, il est clair que la durée totale que peuvent attendre les candidats pour connaître leurs affectations définitives peut avoir un impact important sur leur mental. De plus, l'accès aux positions dans les files d'attente qui sont constamment mises à jour incite certains candidats à se connecter régulièrement sur la plateforme pour regarder leur avancement et juger leurs chances d'être acceptés.

Un autre problème est l'inégalité des candidats face aux délais d'arrivée d'une réponse : s'ils doivent déménager, les candidats les plus défavorisés peuvent être très contraints sur les logements qui leur sont disponibles. Ces candidats là ont donc un vrai besoin d'avoir une réponse rapide afin de pouvoir s'organiser et des propositions de formations peuvent malheureusement arriver trop tard. Par exemple, si un candidat décide de valider définitivement sa proposition le 30 juin pour se laisser le temps de s'organiser et qu'une proposition qu'il aurait préféré arrive en juillet, il ne pourra pas la choisir malgré ses préférences. A l'inverse, il peut décider d'attendre pour obtenir mieux mais il sera plus compliqué de trouver un logement sur place. Ce type de problèmes peut arriver fréquemment sachant qu'en 2020, 11% des lycéens boursiers ont attendu plus de deux mois avant d'obtenir la proposition qu'ils accepteront finalement. La procédure est également très longue comparée à d'autres pays ayant des systèmes d'admission similaires : là où les procédures à l'étranger durent généralement moins d'un mois<sup>1</sup>, celle de Parcoursup peut durer jusqu'à 4 mois, de juin à septembre.

Comme évoqué en section 3.2.3, les candidats les plus touchés par les longueurs de procédure sont ceux qui ne se trouvent pas en haut des listes des formations. Là où un candidat très bien classé pour tous ses vœux connaîtra rapidement son affectation définitive, un candidat qui n'est bien classé que sur une partie de ses vœux ou sur aucun vœu devra patienter en file d'attente pour connaître son affectation. Le critère de "maturation du choix" rentre d'ailleurs en conflit avec le bien-être des candidats, l'un profite à ceux qui ont du temps pour choisir et l'autre profite à ceux qui doivent tout choisir dès le début.

1. voir [rapport du CESP de 2021](#), annexe 3

Une idée mise en avant par le CESP pour déplacer ce curseur entre ces deux principes est la mise en place d'un répondeur automatique obligatoire pour tous les candidats avec des vœux en attente à partir d'une certaine date. Le principe du répondeur automatique consiste à ce que chaque candidat fournisse un classement strict de ses vœux en attente afin qu'un algorithme puisse répondre à sa place aux propositions qui lui sont envoyées. Dans cette configuration où les préférences des formations et des candidats sont connues par le système, il s'agit juste d'une exécution d'un algorithme de Gale-Shapley sur un sous-ensemble de candidats et de vœux, et en quelques minutes les candidats pourront connaître avec certitude leur affectation donnée en phase principale. Une mise en place de ce système à la fin officielle de la phase principale en mi-juillet pourrait largement accélérer le processus et limiter le nombre de propositions faites après cette date dans le cadre de la GDD.

### 3.3.4 Manque de transparence

**Critère 4 (Traçabilité).** Les ensembles de données et les processus permettant au système d'IA de rendre une décision, y compris les processus de collecte et d'étiquetage de données, ainsi que les algorithmes utilisés, devraient être documentés selon les normes les plus strictes afin de permettre la traçabilité ainsi qu'une amélioration de la transparence.

"Lignes directrices en matière d'éthique pour une IA de confiance"  
- Groupe d'Experts de Haut Niveau sur l'IA - Commission Européenne

Malgré la bonne explicabilité et la simplicité de la procédure mentionnés en section 3.2.2, Parcoursup manque d'une réelle transparence sur la façon dont sont classés les candidats. Actuellement le gouvernement rend obligatoire pour les formations de présenter plus ou moins succinctement les critères d'évaluation des dossiers (voir figure 3.2) mais aucune obligation n'existe quant aux barèmes utilisés pour la notation. Les algorithmes « locaux » utilisés par les formations pour trier leurs candidats ne sont pas publiés et cela vient entraver l'explicabilité globale de l'algorithme, ces algorithmes étant clés dans la procédure globale. Il en va de même pour les quotas de boursiers dans chaque formation qui sont fixés arbitrairement par le recteur de chaque académie et sur lesquels aucune justification n'est fournie. Cette opacité, mêlée aux différents quotas qui viennent perturber les classements, peut provoquer une incompréhension et une méfiance vis à vis de l'entièreté du système. Comment un candidat peut-il connaître les raisons pour lesquelles il n'a pas été accepté dans une formation? Pour quelle raison un autre candidat est-il placé plus haut alors que ses résultats sont moins bons? Si les critères publiés par les formations peuvent apporter une réponse partielle, force est de constater que certaines formations restent très floues quant aux méthodes de classement comme montré dans la figure 3.2. Un premier pas vers la transparence serait de contrôler plus strictement les critères publiés afin que les candidats soient mieux informés quels que soient les formations qu'ils souhaitent, ou encore publier les barèmes utilisés sur les critères quantitatifs (notes de lycée et notes du baccalauréat).

## 3.4 Synthèse

Suite aux critiques publiques visant le tirage au sort et le manque de traitement humain des candidatures, le gouvernement a réagi en proposant cette toute nouvelle plateforme. Tandis que le tirage au sort a été rendu impossible suite à la loi ORE, l'examen humain des candidatures reste complexe à mettre en œuvre malgré les modifications de la loi. Si l'on prend l'exemple des universités qui doivent désormais trier des candidats affluant par milliers en quelques mois, le traitement individuel qu'elles peuvent effectuer reste malheureusement très restreint et il est impossible de passer plus de quelques minutes sur chaque candidature. De plus, établir un classement strict sur un tel nombre de candidats peut s'avérer très arbitraire, car ces traitements vont au final faire partie d'un classement automatique basé sur des pondérations restant à la discrétion du comité d'examen des vœux. Si des formations recevant peu de propositions, comme des formations privées, peuvent passer plus de temps à l'examen de chaque candidature, la tâche est plus complexe pour les formations universitaires par exemple.

## Critères généraux d'examen des voeux

### RÉSULTATS ACADÉMIQUES

#### Ensemble des notes

- Ce critère est **Important** pour la commission d'examen des voeux  
Retenues de notes des années en cours et antérieures

#### Résultats obtenus aux épreuves anticipées du baccalauréat de français.

- Ce critère est **Important** pour la commission d'examen des voeux  
Retenue de notes du baccalauréat de français

### COMPÉTENCES ACADÉMIQUES, ACQUIS MÉTHODOLOGIQUES, SAVOIR-FAIRE

#### Cohérence du projet de formation

- Ce critère est **Important** pour la commission d'examen des voeux  
La fiche avenir et le projet motivé

## Critères généraux d'examen des voeux

### RÉSULTATS ACADÉMIQUES

#### Bulletins de Première, Terminale, Résultats des épreuves anticipées au baccalauréat français et, dans les cas de réorientation universitaire, résultats au baccalauréat et résultats dans l'enseignement supérieur

- Ce critère est **Essentiel** pour la commission d'examen des voeux  
Notes de : Français, histoire/géographie, philosophie, science économique, science sociale, mathématiques.

#### Bulletins de Première, Terminale, Résultats des épreuves anticipées au baccalauréat français et, dans les cas de réorientation universitaire, résultats au baccalauréat et résultats dans l'enseignement supérieur

- Ce critère est **Complémentaire** pour la commission d'examen des voeux  
Options : Droit et grands enjeux du monde contemporain, sciences sociales et politiques, sciences économiques et sociales, latin.

### COMPÉTENCES ACADÉMIQUES, ACQUIS MÉTHODOLOGIQUES, SAVOIR-FAIRE

#### Maîtrise de la langue française

- Ce critère est **Essentiel** pour la commission d'examen des voeux  
Note de français et de philosophie

### SAVOIR-ÊTRE

#### Aptitude au travail universitaire

- Ce critère est **Important** pour la commission d'examen des voeux  
Appréciation des enseignants

### MOTIVATION, CONNAISSANCE DE LA FORMATION, COHÉRENCE DU PROJET

#### Motivation

- Ce critère est **Très important** pour la commission d'examen des voeux  
Projet de formation motivé

### ENGAGEMENTS, ACTIVITÉS ET CENTRES D'INTÉRÊT, RÉALISATIONS PÉRI OU EXTRA-SCOLAIRES

#### Engagement citoyen

- Ce critère est **Complémentaire** pour la commission d'examen des voeux  
Champ "Engagement citoyen" de la fiche Avenir

FIGURE 3.2 – Différence de détails sur les critères d'examen entre deux licences portant la même mention

Au delà de ces problèmes, Parcoursup a apporté avec lui un tout nouveau système d'affectation, toujours basé sur l'algorithme de Gale et Shapley mais faisant participer directement les candidats à l'intérieur de la procédure. Cela permet de produire des effets vertueux liés à l'action humaine et à l'explicabilité, principes quasiment absents de l'ancienne procédure APB. Au delà de leurs mises en œuvre, ces principes sont clés dans l'éthique des IA et constituent de très bons arguments en faveur de ce système, car ils sont directement liés à la démocratie et à un accès équitable aux informations.

Cependant, la mise en œuvre de l'amélioration de ces principes se fait au prix d'un algorithme plus faible, fournissant des affectations globalement moins satisfaisantes pour les candidats et produisant probablement moins d'affectations au total. L'absence de garantie de stabilité est une conséquence difficilement évitable, et elle peut légitimement créer un sentiment d'injustice chez des milliers de candidats. Ces points faibles seront principalement ressentis par les candidats les moins bien classés par les formations, ceux-là sont en effet largement désavantagés par une procédure de type Gale-Shapley. De plus, les longueurs de procédure que peuvent subir ceux qui seront tout de même affectés creusent un écart plus grand entre les meilleurs candidats, très rapidement affectés, et les autres. La phase complémentaire, visant les étudiants n'ayant rien ou ne pouvant se permettre d'attendre, va les reconduire vers des formations qu'ils n'ont pas sélectionnés en amont et les incite à choisir une formation peu attractive par dépit. De plus, l'accès aux formations dans cette phase est assez arbitraire en comparaison avec la phase principale et ne permet pas nécessairement aux étudiants les plus méritants d'avoir une seconde chance. D'autres éléments se situant hors du domaine technique comme les lettres de motivation, la fiche Avenir ou l'auto-censure peuvent avoir comme effet de désavantager les candidats défavorisés [9]. Les algorithmes n'ont dans ce cas pas d'effet sur ces points, qui sont traités et notés exclusivement par des humains. Pour aider à contrebalancer ces biais sociaux, le gouvernement a mis en place différents quotas ayant permis un meilleur accès à la plateforme pour les candidats visés. Cependant, l'opacité des critères fixant les taux de boursiers ainsi que la pertinence d'utiliser le critère « boursier » pour discriminer l'origine sociale des candidats sont critiqués [10].

Ainsi, les nouveaux éléments positifs apportés par Parcoursup comportent des problèmes aux conséquences importantes sur une partie non négligeable des candidats qui semblent difficiles à effacer. Dans le chapitre suivant, nous étudions une approche différente qui pourrait être mise en place en essayant de contourner les principaux problèmes de Parcoursup.

## Chapitre 4

# Proposition d'une nouvelle modélisation : préférences avec indifférence

### 4.1 Motivation

Le ressenti global des candidats sur la procédure Parcoursup est mitigé : selon un sondage IPSOS sur la campagne Parcoursup de 2021, 82% des sondés considèrent l'adjectif "stressant" comme s'appliquant bien à la procédure Parcoursup. Même si par leur nature, les procédures relatives à l'orientation peuvent être considérées comme stressantes en tant que tel, la durée relativement longue de la procédure Parcoursup est à questionner. En effet, si l'on souhaite réduire le stress et l'anxiété générés par cette procédure, il semble naturel de la faire subir le moins longtemps possible aux candidats. D'ailleurs, c'est probablement pour cette raison que la durée de la phase principale de Parcoursup se voit abaisser d'année en année. Dans ce même sondage, 47% des candidats considèrent que la procédure n'est pas rapide, cependant, 68% des candidats ont apprécié le fait de ne pas avoir eu à classer leurs vœux. L'exemple de la table 3.3, ainsi que les données présentées en table 3.4 et 3.5 montrent la difficulté à obtenir une affectation globalement stable en ne rendant pas obligatoire un classement préalable. Ne pas avoir à faire ceci mènera forcément à des temps de procédure plus longs. En outre, dans notre contribution nous chercherons à revenir à une procédure à la fois rapide et qui permette une garantie sur l'optimalité de l'affectation des candidats. En ce sens, faire tourner l'algorithme de Gale-Shapley (comme ce qui était fait précédemment avec le système Admission Post Bac) en demandant les préférences des étudiants au préalable semble être une bonne solution.

Cependant, si le retour à un système plus classique retire les points positifs identifiés relatifs à l'action humaine sur la procédure, nous voulons tout de même garantir une certaine liberté d'action pour les candidats. Dans ce type de procédure, l'action du candidat est directement liée à la façon dont il va trier ses vœux. Les méthodes de classement traditionnelles obligent les candidats et les formations à classer de manière stricte tous leurs vœux et candidatures. Si on admet que certains d'entre eux peuvent être indifférents face à plusieurs propositions, alors il peut être intéressant que ce type de préférences soient pris en compte par l'algorithme d'affectation. Si on souhaite conserver le principe de stabilité, alors une modélisation par un problème Hospitals/Residents with Ties (HRT) correspond parfaitement au système que l'on veut implémenter. Comme vu en section 1.3.2, les solutions à ce problème peuvent avoir des tailles variables et par souci de vouloir affecter un maximum de candidats possible tout en respectant la stabilité, nous pouvons utiliser la résolution d'un problème MAX HRT pour trouver la meilleure solution possible.

A travers ce changement de problème, nous souhaitons voir à quel point les solutions trouvées sont différentes par rapport à une approche de mariages stables traditionnelle et si une modélisation plus fine des préférences peut donner de meilleurs résultats. Nous nous pencherons également sur les conséquences de l'utilisation d'un tel système par rapport aux pointés éthiques situés plus hauts afin de comprendre les avantages et les inconvénients d'une telle méthode par rapport au système existant.

Nous modéliserons ici le problème où seuls les candidats peuvent insérer des égalités entre leurs vœux,

les formations ayant toujours à fournir un classement strict. Ce choix s’est fait pour plusieurs raisons : la présence d’un classement partiel d’un seul côté permet d’utiliser des algorithmes avec de meilleures bornes d’approximation, et la mise en place d’un tel système à l’échelle nationale semble plus accessible. En effet, il peut être plus complexe pour des formations de formuler des égalités dans leurs classements pouvant contenir des milliers d’étudiants en comparaison avec des candidats qui n’ont qu’une dizaine de vœux à classer. Cependant, la possibilité pour les établissements d’effectuer un classement partiel pourrait avoir un effet bénéfique et renforcer la qualité de traitement de chaque candidature par les établissements : s’il peut être compliqué de comparer deux étudiants avec des profils similaires, la possibilité de les représenter comme équivalents pour une formation peut éviter d’avoir recours à des pondérations arbitraires qui en placeraient un au dessus d’un autre. Ce type de stratégies est déjà appliqué par certaines formations [10] afin de se concentrer principalement sur les candidats de niveau intermédiaire qui sont « sur le fil » de l’admission.

## 4.2 Expérimentation

Dans l’optique de vérifier l’efficacité d’un changement de problème de Hospitals/Residents à MAX HRT, nous comparons le nombre d’affectations produits par une procédure Gale-Shapley ne prenant pas en compte les égalités de préférences, avec un algorithme d’approximation conçu pour MAX HRT. Nous nous intéresserons ici au cas où les candidats peuvent formuler des égalités dans leurs préférences, les préférences des formations restant strictes. Dans un premier temps, nous détaillerons la construction des instances utilisées pour comparer les deux approches, puis nous verrons comment l’algorithme de Gale-Shapley peut être utilisé dans le contexte de classements avec égalités, nous détaillerons ensuite l’algorithme de Huang et Kavitha pour les égalités d’un seul côté [15], et nous comparerons enfin les résultats obtenus selon les différentes approches.

Les instances utilisées sont similaires à celles décrites en section 3.3.1. Aux listes de préférences déjà établies, nous devons ajouter des égalités : selon un paramètre  $p\_ties$ , pour chaque formation  $f$  dans la liste d’un candidat  $c$ ,  $f$  a une probabilité  $p\_ties$  de se retrouver au même rang que la formation la précédant dans la liste de  $c$ . Ainsi, on obtient une instance de HRT avec des égalités présentes uniquement du côté des candidats.

### 4.2.1 Algorithme de Gale-Shapley et égalités

Afin de pouvoir utiliser ces instances avec l’algorithme de Gale-Shapley pour simuler le contexte de Parcoursup, il est nécessaire de savoir par quel moyen les égalités peuvent être tranchées. En effet, un candidat pourrait dans ce contexte se retrouver face à un choix obligatoire entre deux formations qu’il apprécie autant. Quoiqu’il arrive, la manière de départager n’influera pas sur la stabilité de la solution : peu importe la manière de trancher, la solution obtenue restera faiblement stable [26]. Comme les moyens de départager deux formations peuvent être différents selon le candidat (proximité de la formation, camarades fréquentant l’établissement, popularité globale de la formation...), on décide de départager aléatoirement pour chaque candidat.

Afin de simuler d’autres comportements possibles, trois autres méthodes basées sur la popularité et le nombre de places sont testées. La première consiste à choisir tout le temps la formation la plus populaire, la deuxième consiste au contraire à préférer la moins populaire et la troisième va choisir la formation qui dispose de plus de places. La première peut se rapprocher d’un cas plus réel (dans le doute, les étudiants peuvent être influencés par les popularités globales) mais semble peu efficace à première vue. La deuxième semble moins réaliste mais serait potentiellement plus efficace sur le nombre total d’affectés, les formations moins populaires étant celles qui peinent le plus à remplir. Au contraire des deux premières, la troisième est dynamique car les égalités ne peuvent être cassées qu’au moment d’un choix à faire entre deux formations. Dans tous les cas, ces trois méthodes peuvent servir d’heuristiques simples pour départager les égalités d’un problème HRT.

## 4.2.2 Algorithme de Huang et Kavitha

Afin de démontrer l'efficacité d'un algorithme conçu pour maximiser le nombre d'affectations, nous utiliserons celui décrit par Huang et Kavitha [15] pour le cas où les égalités ne sont que d'un côté. Quelques modifications de l'algorithme ont dû être faites pour le cas de MAX HRT, celui de base étant conçu pour le problème MAX SMTI. Pour notre cas, les formations correspondent aux hommes et appelleront donc tour à tour les candidats de leur liste à la manière de l'algorithme de Gale-Shapley.

La procédure se base sur un système de promotion initialement proposé par Király [21] : dès qu'une formation a appelé tous ses candidats et qu'il lui reste des places à pourvoir, elle repart au début de sa liste et devient promue (1-promue). Une promotion lui permet d'être favorisée si un candidat a le choix entre deux formations à égalité et que l'autre formation est non-promue. Si elle arrive à nouveau à la fin de sa liste, elle est promue une nouvelle fois (2-promue) et est prioritaire sur les non-promues et les 1-promues. Après ça, la formation ne peut plus être promue et une arrivée en fin de liste signifiera qu'elle ne pourra plus rien proposer, on dit qu'elle abandonne.

De plus, au lieu de ne faire qu'une proposition par place disponible, les formations ont le droit de faire deux propositions par place (sachant qu'une formation peut envoyer deux propositions à un seul candidat). On note une proposition provenant de la formation  $f_j$  comme  $p_{j,k}^i$  où  $i \in \{1, 2\}$  et  $k \in \{1, \dots, c_j\}$ . De leur côté, les candidats ne peuvent conserver qu'au plus deux propositions à la fois. Si un candidat en reçoit une troisième il doit éliminer l'une des moins désirables parmi les trois, c'est à dire une proposition qui n'est supérieure à aucune des deux autres.

**Définition 6** (Supériorité d'une proposition). Pour un candidat  $c$ , la proposition  $p_{j,k}^i$  est supérieure à  $p_{j',k'}^{i'}$  si dans la liste de  $c$  :

- $f_j$  est mieux classée que  $f_{j'}$
- $f_j$  et  $f_{j'}$  sont à égalité ;  $f_j$  est 2-promue tandis que  $f_{j'}$  est 1-promue ou non-promue
- $f_j$  et  $f_{j'}$  sont à égalité ;  $f_j$  est 1-promue tandis que  $f_{j'}$  est non-promue
- $f_j$  et  $f_{j'}$  sont à égalité et les deux sont non-promues ; le candidat  $c$  a déjà rejeté une proposition de  $f_j$  mais n'a rejeté aucune proposition de  $f_{j'}$

L'algorithme 2 décrit la procédure complète.  $t_j^i$  représente pour la formation  $f_j$  le prochain candidat à recevoir la  $i$ -ème proposition. La principale modification par rapport à l'algorithme de Huang et Kavitha réside dans le fait que chaque formation  $f_j$  peut envoyer  $2c_j$  propositions, même quand  $c_j$  est supérieur à 1.

---

**Algorithme 2** : Algorithme de Huang et Kavitha

---

**Entrée** : Listes de préférences et capacités de chaque formation

**Sortie** : Un ensemble de propositions faites aux candidats (au maximum 2 par candidat)

$\forall i \in \{1, 2\}, \forall f_j \in F, t_j^i \leftarrow 1$

**tant que** une formation  $f_j \in F$  n'a pas abandonné et a une proposition  $p_{j,k}^i$  qui n'a pas été acceptée

**faire**

$f_j$  envoie une proposition  $p_{j,k}^i$  au  $t_j^i$ -ième candidat  $c$  de sa liste

**si**  $c$  a actuellement deux propositions ou moins ( $p_{j,k}^i$  inclus) **alors**

$c$  accepte  $p_{j,k}^i$

**si**  $k < c_j$  **alors**

$t_j^i \leftarrow t_j^i + 1$

**fin**

$i' \leftarrow i$

$j' \leftarrow j$

**sinon**

$c$  rejette l'une de ses propositions les moins désirables  $p_{j',k'}^{i'}$

$t_{j'}^{i'} \leftarrow t_{j'}^{i'} + 1$

**fin**

**si**  $t_{j'}^{i'} >$  nombre de candidats dans la liste de  $f_{j'}$  **alors**

$t_{j'}^{i'} \leftarrow 1$

**si**  $f_{j'}$  est non-promue **alors**

$f_{j'}$  devient 1-promue

**sinon si**  $f_{j'}$  est 1-promue **alors**

$f_{j'}$  devient 2-promue

**sinon si**  $f_{j'}$  est 2-promue **alors**

$f_{j'}$  abandonne

**fin**

**fin**

**fin**

---

Une fois l'algorithme arrivé à la fin de son exécution, chaque candidat possède 0, 1 ou 2 propositions et chaque formation  $f_j$  a au plus  $2c_j$  propositions qui ont été acceptées. On construit ainsi le graphe biparti  $G$  reliant les candidats aux formations pour lesquels ils ont accepté une proposition. On peut ainsi chercher une affectation  $M$ , un sous-graphe de  $G$  où les sommets correspondant aux candidats ont un degré inférieur ou égal à 1 et ceux correspondant aux formations ont un degré inférieur ou égal à la capacité de la formation. On cherche évidemment ici une affectation de cardinalité maximale, celle-ci peut être calculée en temps polynomial et sera nécessairement stable [15]. Dans notre implémentation, nous transformerons le graphe pour pouvoir résoudre un problème de flot maximal qui sera résolu avec l'algorithme de Ford-Fulkerson [8], ce qui nous permet d'obtenir toutes les affectations finales.

### 4.2.3 Résultats expérimentaux

Afin de tester l'efficacité de l'algorithme de Huang-Kavitha comparé à des approches plus simples basées sur Gale-Shapley, nous avons mesuré le nombre d'affectations effectuées par chacun des algorithmes sur des instances comportant de 5.000 à 100.000 candidats, ces résultats sont compilés dans la figure 4.1. Nous pouvons voir que l'implémentation de Huang-Kavitha fournit toujours des solutions contenant en moyenne plus d'affectations qu'avec n'importe quelle heuristique simple. De plus, le nombre de candidats non-affectés augmente bien plus en fonction du nombre de candidats total avec une procédure Gale-Shapley qu'avec l'algorithme Huang-Kavitha. Même si l'augmentation du nombre d'affectés reste faible ( $< 1\%$  du nombre total de candidats), l'application d'un tel algorithme au niveau national pourrait affecter plus d'un millier de candidats supplémentaires.

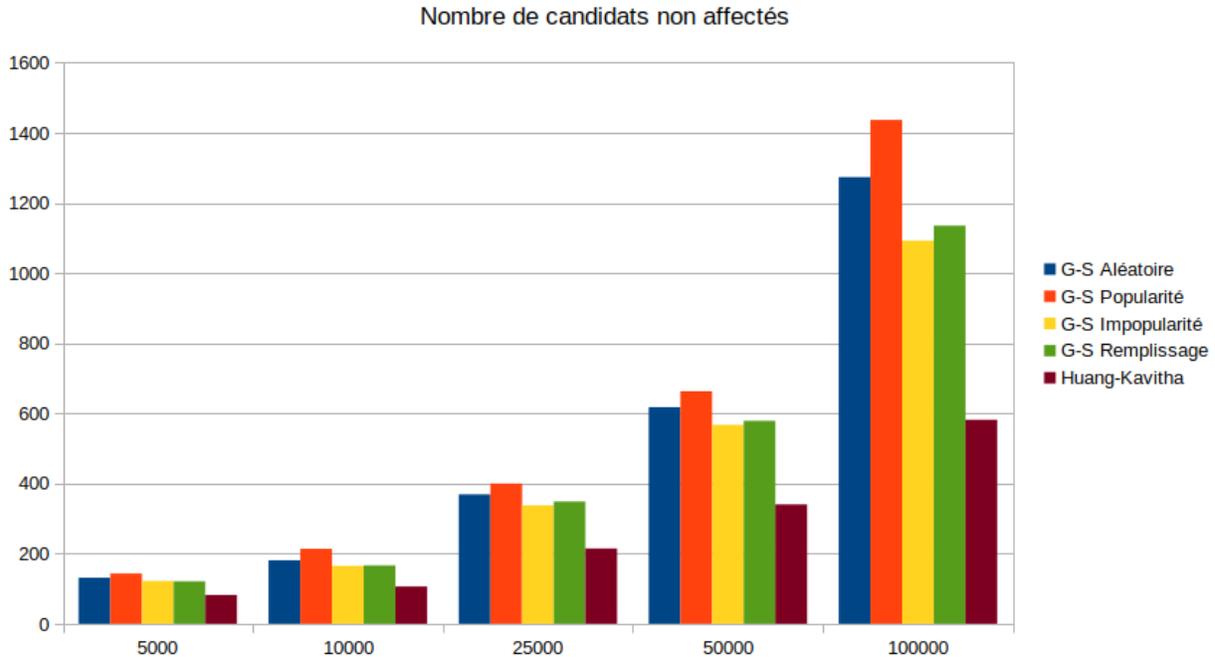


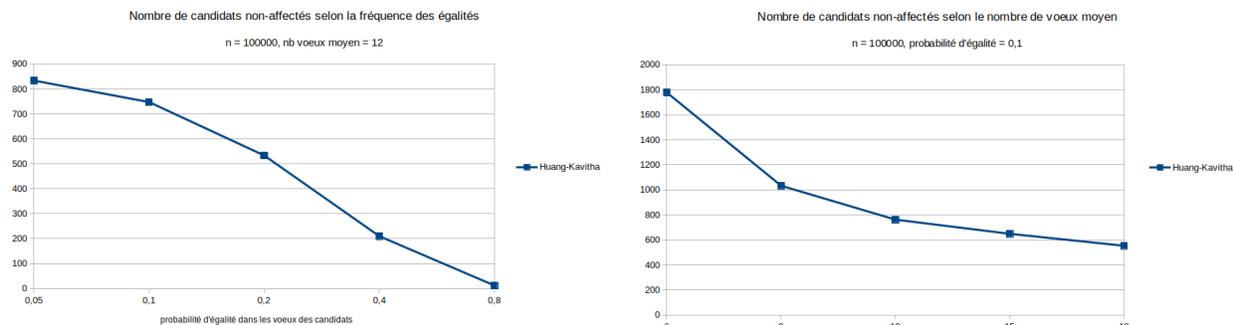
FIGURE 4.1 – Comparaisons des différents algorithmes de résolution de MAX HRT selon le nombre de candidats

En ce qui concerne les heuristiques, celle choisissant les formations les plus populaires est bien sous-effective par rapport aux autres. La meilleure possible semble être celle qui au contraire choisit la moins populaire pour casser les égalités, qui reste légèrement meilleure que l’heuristique privilégiant les formations les moins remplies.

Bien sûr, la forme ainsi que la taille des listes de voeux produites par les candidats influenceront beaucoup sur le nombre d’affectations finales. Les figures 4.2a et 4.2b montrent l’évolution du nombre d’affectations avec l’algorithme de Huang-Kavitha selon la probabilité  $p\_ties$  et selon le nombre de voeux moyen que les étudiants font. Si la montée du nombre de voeux moyen permet d’affecter un peu plus d’étudiants, la montée de la présence des égalités permet à l’algorithme de Huang et Kavitha de fonctionner pleinement en ne laissant qu’en moyenne une dizaine de candidats sans affectations lorsque  $p\_ties = 0.8$ . Même si cette probabilité est très haute et peu réaliste (tous les candidats avaient des listes de voeux composées d’au maximum 4 groupes de formation équivalentes pour plus d’une dizaine de voeux), il peut être intéressant d’inciter les candidats à mettre le plus de formations à égalité possible pour augmenter les chances de tous les candidats.

### 4.3 Avantages et limites de l’approche

Comme vu avec l’expérimentation, le changement de problème permet d’affecter plus de candidats tout en conservant la propriété essentielle de stabilité, permettant de garantir un certain niveau de satisfaction individuelle. Même si la part des individus concernés par ces changements est relativement faible, cela reste néanmoins une bonne amélioration sur ce critère. Comme dit précédemment, une telle modélisation implique de retourner vers un système où les candidats devront classer en amont leurs préférences avant de les envoyer aux formations. Ainsi tous les avantages, mais aussi les inconvénients, liés au fonctionnement particulier du système d’affectation de Parcoursup disparaissent. Les problèmes de sous-optimalité (section 3.3.1) et de



(a) Effet de la variation de la probabilité d'égalités sur le nombre d'affectation (b) Effet de la variation du nombre de voeux moyen sur le nombre d'affectation

FIGURE 4.2 – Efficacité selon le type d'instance

durée de la procédure (section 3.3.3) n'existent donc plus, mais la liberté d'action offerte par ce système (section 3.2.1) n'est plus comparable.

La possibilité nouvelle de modéliser des égalités semble être un élément permettant aux candidats de modéliser de façon plus précise leurs véritables préférences. Cependant, est-ce que cela les avantage-t-il pour autant? Laisser plusieurs formations à égalité dans sa liste de voeux revient à laisser décider un algorithme laquelle serait optimale pour maximiser le nombre d'affectations totales. Or, cela semble contradictoire avec la volonté que chaque candidat puisse décider lui-même de sa formation. Finalement, rentrer des égalités revient à créer un classement avec des informations incomplètes qu'un algorithme va compléter selon son objectif, ce qui peut être vu comme une perte d'autonomie de l'homme face à l'algorithme (voir critère 1).

Cet algorithme est également plus complexe et beaucoup moins intuitif que la procédure actuelle : même sans avoir à comprendre la preuve de la stabilité de la solution sortante, des connaissances en algorithmique et en théorie des graphes doivent être mobilisées pour comprendre le fonctionnement de l'algorithme. Même l'algorithme de Gale-Shapley dans sa version standard est relativement simple à comprendre car son fonctionnement reste intuitif. Ainsi, sur le critère de l'explicabilité (section 3.2.2), le système proposé est bien moins bon que ce que Parcoursup propose actuellement. Cela met encore une fois en avant des incompatibilités entre optimalité et explicabilité du système, comme mentionné dans la définition du critère 2.

Cette proposition permet de cerner certains problèmes qui se posent lors du choix d'un algorithme. Ici, la volonté que les candidats puissent mieux exprimer leurs préférences motive la création d'un système se basant sur la modélisation du problème MAX HRT. Les avantages sont clairs et vérifiés par l'expérience, ce qui peut nous permettre de conclure positivement sur l'efficacité de notre système, notamment en terme du nombre de personnes affectées. Cependant, les avancées obtenues par cette proposition provoquent également un recul du point de vue éthique sur des points cruciaux comme la transparence du système ou l'autonomie humaine.

Ainsi, les défis éthiques liés à l'implémentation d'une plateforme comme Parcoursup sont complexes car les avancées dans un sens impliquent souvent un recul sur plusieurs points. La particularité de Parcoursup étant son mode de fonctionnement permettant une meilleure lisibilité de la procédure, elle peut aussi impliquer des inconvénients moins visibles comme des inégalités et des affectations moins satisfaisantes. Une fois ce constat fait, on peut voir que la question de savoir quels algorithmes utiliser dépend en très grande partie des préférences éthiques et des buts à atteindre des décideurs.

## Chapitre 5

# Conclusion

A travers notre analyse de Parcoursup, nous avons pu extraire différents points d'intérêts sur le plan éthique, qu'ils soient positifs ou négatifs, au sens des critères issus du rapport Trustworthy AI sur lequel nous nous sommes basés. Dans une volonté d'apporter des éléments bénéfiques liés à l'ouverture du système au plus grand monde et de rendre les procédures moins mécaniques et plus humaines, les nouveautés mises en place dans Parcoursup nuisent à l'efficacité de l'algorithme d'affectation et conduisent à des longueurs de procédure qui sont devenus d'importants facteurs de stress se rajoutant à l'anxiété globale que génère l'entrée dans l'enseignement supérieur. D'autres points négatifs liés à la transparence et à la phase complémentaire peuvent faire monter un sentiment d'injustice chez les candidats, et sont donc des éléments prioritaires à regarder pour améliorer la confiance portée à la plateforme.

Nous avons établi une proposition de changement de système qui consiste à modéliser les préférences des étudiants et/ou des formations par des ordres partiels plutôt que des ordres totaux. Par une volonté initiale de laisser un plus grand champ libre aux candidats pour exprimer leurs réelles préférences, nous avons vu que ce choix pouvait amener des effets vertueux en affectant potentiellement un plus grand nombre de candidats que ce qui est fait dans Parcoursup. Nous avons également remarqué que, tout comme Parcoursup, les changements bénéfiques effectués impliquent un recul non désiré sur certains acquis. Dans notre cas, la transparence du système subit un grand recul par rapport à Parcoursup du fait d'algorithmes peu accessibles pour des non spécialistes. A travers ce cas d'étude, nous avons pu mettre en lumière les dilemmes qui peuvent se poser lors des choix de modélisation d'un système, et les différentes contreparties à équilibrer pour obtenir un algorithme le plus fidèle aux volontés du décideur pouvant elles-mêmes être discutées et modélisées de différentes façons...

Si Parcoursup reste un système perfectible et modifiable sur plusieurs points, il n'est que la partie visible des politiques menées en terme d'admissions dans l'enseignement supérieur. Les algorithmes utilisés par la plateforme marchent correctement selon les spécifications [3], ce sont donc bien ces spécifications qui sont à débattre plutôt que le fonctionnement technique de la plateforme. Bien d'autres curseurs modifiables sont disponibles pour régler les problèmes évoqués plus haut, et ce sous différentes formes. En 2021, plus de 900.000 personnes étaient candidates à des formations post-bac, malgré le fait que celles-ci ne comportent en tout que 770.000 places (cf. données publiques de Parcoursup 2021). Même si une partie des candidats n'utilise Parcoursup que comme une solution de secours, cette asymétrie entre nombre de candidats et nombre de places créera nécessairement des tensions et une concurrence entre étudiants, menant à un stress plus important. Un nombre important de facteurs, rappelés régulièrement par le CESP, peuvent jouer sur la manière dont tous ces candidats seront affectés, bien au-delà du choix d'un algorithme qui n'est finalement qu'un outil permettant d'encadrer la concurrence entre les établissements et entre les étudiants [10].

# Bibliographie

- [1] Felix BAUCKHOLT, Kanstantsin PASHKOVICH et Laura SANITÀ. « On the approximability of the stable marriage problem with one-sided ties ». In : *CoRR* abs/1805.05391 (2018).
- [2] Tom L BEAUCHAMP et James F CHILDRESS. *Principles of Biomedical Ethics*. 8<sup>e</sup> éd. Oxford University Press, 2019.
- [3] Pierre CASTÉLAN et al. *Vérification de l'algorithme de calcul des ordres d'appel dans Parcoursup*. Research Report. Oct. 2021.
- [4] Robert CHIANG et Kanstantsin PASHKOVICH. « On the approximability of the stable matching problem with ties of size two ». In : *Algorithmica* 82.9 (2020), p. 2668-2686.
- [5] Maxence DELORME et al. « Mathematical models for stable matching problems with ties and incomplete lists ». In : *European Journal of Operational Research* 277.2 (2019), p. 426-441.
- [6] François DUBET. « Injustice et reconnaissance ». In : *Esprit* 7 (2008), p. 144-159.
- [7] Luciano FLORIDI et Josh COWLS. « A unified framework of five principles for AI in society ». In : *Ethics, Governance, and Policies in Artificial Intelligence*. Springer, 2021, p. 5-17.
- [8] Lester Randolph FORD et Delbert R FULKERSON. « Maximal flow through a network ». In : *Canadian journal of Mathematics* 8 (1956), p. 399-404.
- [9] Leïla FROUILLOU. « Parcoursup : quelles sélections à l'entrée dans le supérieur ? » In : *26 es Journées du Longitudinal*. 26. 2021, p. 43-50.
- [10] Leïla FROUILLOU, Clément PIN et Agnès van ZANTEN. « Les plateformes APB et Parcoursup au service de l'égalité des chances ? » In : *L'Année sociologique* 70.2 (2020), p. 337-363.
- [11] David GALE et Lloyd S SHAPLEY. « College admissions and the stability of marriage ». In : *The American Mathematical Monthly* 69.1 (1962), p. 9-15.
- [12] Mirco GELAIN et al. « Local search approaches in stable matching problems ». In : *Algorithms* 6.4 (2013), p. 591-617.
- [13] Dan GUSFIELD et Robert W IRVING. *The stable marriage problem : structure and algorithms*. MIT press, 1989.
- [14] Magnús M HALLDÓRSSON et al. « Randomized approximation of the stable marriage problem ». In : *Theoretical Computer Science* 325.3 (2004), p. 439-465.
- [15] Chien-Chung HUANG et Telikepalli KAVITHA. « Improved approximation algorithms for two variants of the stable marriage problem with ties ». In : *Mathematical Programming* 154.1 (2015), p. 353-380.
- [16] Robert W IRVING. « Stable marriage and indifference ». In : *Discrete Applied Mathematics* 48.3 (1994), p. 261-272.
- [17] Robert W IRVING, David F MANLOVE et Sandy SCOTT. « The hospitals/residents problem with ties ». In : *Scandinavian Workshop on Algorithm Theory*. Springer. 2000, p. 259-271.
- [18] Robert W. IRVING et David F. MANLOVE. « Finding Large Stable Matchings ». In : *ACM J. Exp. Algorithmics* 14 (2010). ISSN : 1084-6654. DOI : [10.1145/1498698.1537595](https://doi.org/10.1145/1498698.1537595).

- [19] Kazuo IWAMA, Shuichi MIYAZAKI et Hiroki YANAGISAWA. « A  $25/17$ -approximation algorithm for the stable marriage problem with one-sided ties ». In : *Algorithmica* 68.3 (2014), p. 758-775.
- [20] Telikepalli KAVITHA et al. « Strongly stable matchings in time  $O(nm)$  and extension to the hospitals-residents problem ». In : *ACM Transactions on Algorithms (TALG)* 3.2 (2007), 15-es.
- [21] Zoltán KIRÁLY. « Better and simpler approximation algorithms for the stable marriage problem ». In : *Algorithmica* 60.1 (2011), p. 3-20.
- [22] Zoltán KIRÁLY. « Linear Time Local Approximation Algorithm for Maximum Stable Marriage ». In : *Algorithms* 6 (sept. 2013), p. 471-484. DOI : [10.3390/a6030471](https://doi.org/10.3390/a6030471).
- [23] Chi-Kit LAM et C Gregory PLAXTON. « A  $(1+1/e)$ -approximation algorithm for maximum stable matching with one-sided ties and incomplete lists ». In : *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2019, p. 2823-2840.
- [24] Chi-Kit LAM et C Gregory PLAXTON. « Maximum stable matching with one-sided ties of bounded length ». In : *International Symposium on Algorithmic Game Theory*. Springer. 2019, p. 343-356.
- [25] David MANLOVE. *Algorithmics of matching under preferences*. T. 2. World Scientific, 2013.
- [26] David MANLOVE et al. « Hard variants of stable marriage ». In : *Theoretical Computer Science* 276 (avr. 2002), p. 261-279. DOI : [10.1016/S0304-3975\(01\)00206-7](https://doi.org/10.1016/S0304-3975(01)00206-7).
- [27] Simon MAURAS. « Analysis of Random Models for Stable Matchings ». Thèse de doct. HEC, Paris, France, 2021.
- [28] Eric MCDERMID. « A  $3/2$ -Approximation Algorithm for General Stable Marriage ». In : *Automata, Languages and Programming*. Sous la dir. de Susanne ALBERS et al. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009, p. 689-700.
- [29] Alban MIZZI. « Ressources et incertitudes des lycéens aux heures de Parcoursup ». In : *26èmes journées du Longitudinal : Sélections, du système éducatif au marché du travail*. 2020.
- [30] Danny MUNERA et al. « Solving hard stable matching problems via local search and cooperative parallelization ». In : *Proceedings of the AAAI Conference on Artificial Intelligence*. T. 29. 1. 2015.
- [31] Katarzyna PALUCH. « Faster and simpler approximation of stable matchings ». In : *Algorithms* 7.2 (2014), p. 189-202.
- [32] Alvin E ROTH. « The evolution of the labor market for medical interns and residents : a case study in game theory ». In : *Journal of political Economy* 92.6 (1984), p. 991-1016.
- [33] Vincent TIBERJ. « Parcoursup ou la sélection par les algorithmes ». In : *La vie des idées* (2021), p. 1-15.