

Proposition thèse sur les problématiques liées à l'affectation des étudiants à des universités

Hugo Gilbert, Julien Lesca et Stefano Moretti

February 2025

Le sujet de thèse est fortement inspiré des problématiques liées à l'affectation des lycéens dans les universités, qui est effectuée en France par Parcoursup [7]. Toutefois, ces problématiques peuvent être élargies à toute procédure centralisée d'affectation de candidats à des institutions, où les candidats expriment des préférences sur les différentes institutions auxquelles ils peuvent être affectés, et où les institutions ont également des préférences sur les candidats qu'elles peuvent recruter (two-sided matching). Dans le cadre de l'affectation des lycéens aux universités, l'objectif est de trouver une affectation stable, c'est-à-dire une affectation telle qu'il n'existe pas un lycéen préférant une autre université à celle qui lui a été attribuée et que cette autre université soit prête à l'accepter. La littérature sur les algorithmes permettant de trouver des affectations stables est très riche, à la fois dans le domaine de l'économie et de l'informatique [5, 6]. La méthode standard pour calculer ce type d'affectation, utilisée par la plupart des systèmes scolaires, est l'algorithme de Gale-Shapley (https://fr.wikipedia.org/wiki/Algorithme_de_Gale_et_Shapley) [4]. Cette procédure, en plus de permettre de trouver des affectations stables, possède de nombreuses propriétés intéressantes. Elle garantit une affectation optimale pour les étudiants, c'est-à-dire qu'il n'existe pas d'autre affectation stable qui soit meilleure pour eux. De plus, elle est non-manipulable (strategy-proof) pour les étudiants [3], ce qui signifie qu'un étudiant ne peut pas améliorer son affectation en déclarant des préférences qui diffèrent de ses préférences réelles (par exemple, en mettant en avant des universités où il a plus de chances d'être accepté). Cette propriété est particulièrement appréciable lorsqu'on cherche à garantir une procédure équitable, qui ne favorise pas les lycéens capables de manipuler le système au détriment des autres.

Ce sont les bonnes propriétés de l'algorithme de Gale-Shapley qui rendent cette procédure attrayante et qui expliquent son adoption large dans l'affectation des étudiants aux universités, comme c'est le cas pour Parcoursup. Néanmoins, ces propriétés reposent sur le postulat que les préférences des lycéens sur les universités et celles des universités sur les candidats sont entièrement connues. Autrement dit, chaque lycéen doit classer l'ensemble des universités par ordre de préférence, et chaque université doit en faire de même pour tous les lycéens. Or, le nombre d'universités est trop important pour exiger des lycéens

qu'ils les classent toutes, et ce problème est encore plus marqué pour les universités, qui devraient classer l'ensemble des lycéens. C'est pourquoi la plupart des systèmes reposant sur Gale-Shapley demandent aux étudiants de classer un nombre limité d'universités et permettent aux universités de ne classer que les lycéens ayant candidaté à leurs formations. Cependant, cette version tronquée de Gale-Shapley ne conserve pas les mêmes propriétés. Elle n'est plus non-manipulable, ne garantit pas nécessairement une solution optimale pour les étudiants et peut même conduire à une affectation instable, car un étudiant peut ne pas candidater à une université qu'il ne connaît pas.

Dans l'optique de pallier ce problème structurel, nous souhaitons étudier la possibilité d'enrichir les préférences partielles des lycéens sur les universités en utilisant des mesures de similarité entre universités, qui indiquent à quel point certaines universités sont perçues comme proches dans les préférences des étudiants. Ces mesures de similarité sont bien connues dans les systèmes de recommandation [1] et peuvent être apprises à partir des données historiques contenant les préférences déjà révélées par les lycéens des années précédentes. Elles peuvent également être induites à partir de caractéristiques intrinsèques des universités, telles que leur position géographique, les programmes offerts, etc. Une autre approche, qui peut compléter l'utilisation des mesures de similarité entre universités, consiste à proposer explicitement aux étudiants de classer des universités qui n'apparaissent pas encore dans leurs préférences, afin d'améliorer l'affectation qui pourra leur être proposée. Ce processus, qui consiste à poser des questions pour inférer des préférences, est appelé élicitation [2].

References

- [1] Charu C. Aggarwal. *Recommender Systems: The Textbook*. Springer Publishing Company, Incorporated, 1st edition, 2016.
- [2] Joanna Drummond and Craig Boutilier. Elicitation and approximately stable matching with partial preferences. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 97–105, 2013.
- [3] Lester E Dubins and David A Freedman. Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly*, 88(7):485–494, 1981.
- [4] David Gale and Lloyd Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [5] D. Gusfield and R. Irving. *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, 1989.
- [6] David F. Manlove. *Algorithmics of Matching Under Preferences*. World Scientific, 2013.
- [7] Parcoursup. Parcoursup. <https://www.parcoursup.gouv.fr/>, 2025.