

IDENTIFICATION ET VISUALISATION DE TRAJECTOIRES ETUDIANTES

Une approche par indicateurs mixtes (numériques & catégoriels) — Application aux données de Licence à l'Université Côte d'Azur

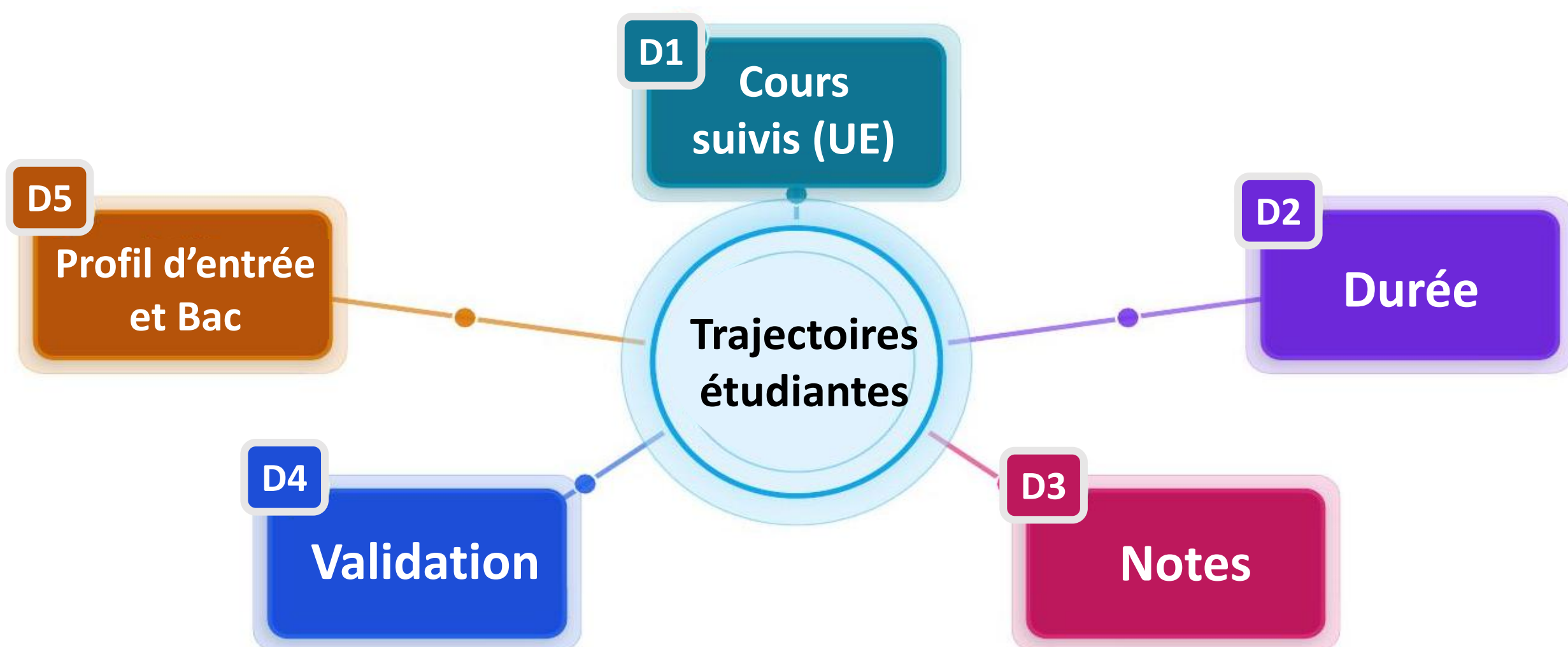
Zeyneb M'HAMED | Directrices : Isabelle MIRBEL & Christel DARTIGUES-PALLEZ | Université Côte d'Azur – Projet L@UCA – Laboratoire i3S

I. Contexte & Problématique

Projet L@UCA (Licence à l'Université, Compétences et Adaptabilité) Projet NCU — Université Côte d'Azur — 2018–2028
Objectif : Identifier et visualiser des trajectoires étudiantes pour améliorer l'offre de formation et mieux accompagner les étudiants.
Périmètre : 30 mentions, 7 portails, 13 000 étudiants

ETAPE 1 : Définition et calcul des Indicateurs individuels des trajectoires étudiantes

Une trajectoire étudiante est caractérisée par **5 dimensions** :



Notations : $|X|$: le nombre d'éléments de l'ensemble X (le cardinal) et i : le $i^{\text{ème}}$ étudiant

NOM DE L'INDICATEUR	DIM couverte	TYPE	FORMULE
Évitement des UE critiques	D1	Numérique	$1 - \frac{ UE_redoublées, disti, UE_critiques }{ UE_critiques }$ UE critique si $nb_redoublements \geq \mu + \sigma$ (seuil adaptatif sur la promotion)
Complétude du parcours	D2	Numérique	$\frac{ UE_obligatoires, suivies, distinctes }{ UE_obligatoires, prévues }$
Indice de durée	D2	Numérique	$\frac{ Semestres_prévusi }{ Semestres_suivis, non disti }$
Score de performance	D3	Catégoriel	Vecteur [Score_L1, Score_L2, Score_L3] Classe 0 = Année non suivie Classe 1 = appartient au tiers inférieur Classe 2 = appartient au tiers intermédiaire Classe 3 = appartient au tiers supérieur
Validation sans compensation	D4	Numérique	$1 - \frac{ UE_compensées, distinctes }{ UE_suivies, distinctes }$
Validation académique	D4	Catégoriel	Vecteur [$S_1, S_2, L_1, S_3, S_4, L_2, S_5, S_6, L_3$] Classe 0 = Non validé Classe 1 = Non suivi Classe 2 = Équivalence Classe 3 = Après redoublement Classe 4 = Par compensation Classe 5 = Normalement
Profil d'entrée	D5	Catégoriel	Classe 1 : entrée en L1 Classe 2 : entrée en L2 Classe 3 : entrée en L3
Profil de Baccalauréat	D5	Catégoriel	Classe 0 = Bac inconnu Classe 1 = Bac Professionnel Classe 2 = Bac Technologique Classe 3 = Bac Général

II. Méthodologie



ETAPE 1 : Définir et calculer des INDICATEURS Numériques & Catégoriels pour chaque étudiant

ETAPE 2 : Choisir une mesure de dissimilarité mixte adaptée aux indicateurs (numériques + catégoriels)

ETAPE 3 : Faire ressortir les groupes homogènes de trajectoires grâce au CLUSTERING

ETAPE 2 : Choisir une mesure de dissimilarité

K-PROTOTYPE — Mesure de dissimilarité

$$d(X_i, Q_l) = \sum_{k=1}^p (x_i^k - Q_l^k)^2 + \gamma \sum_{k=p+1}^{p+q} \delta(x_i^k, Q_l^k)$$

Partie numérique

Mesure : distance euclidienne

Partie catégorielle

Mesure : Simple Matching

$$\delta(a, b) = \begin{cases} 0 & \text{si } a = b \\ 1 & \text{si } a \neq b \end{cases}$$

γ : équilibre poids numérique / catégoriel

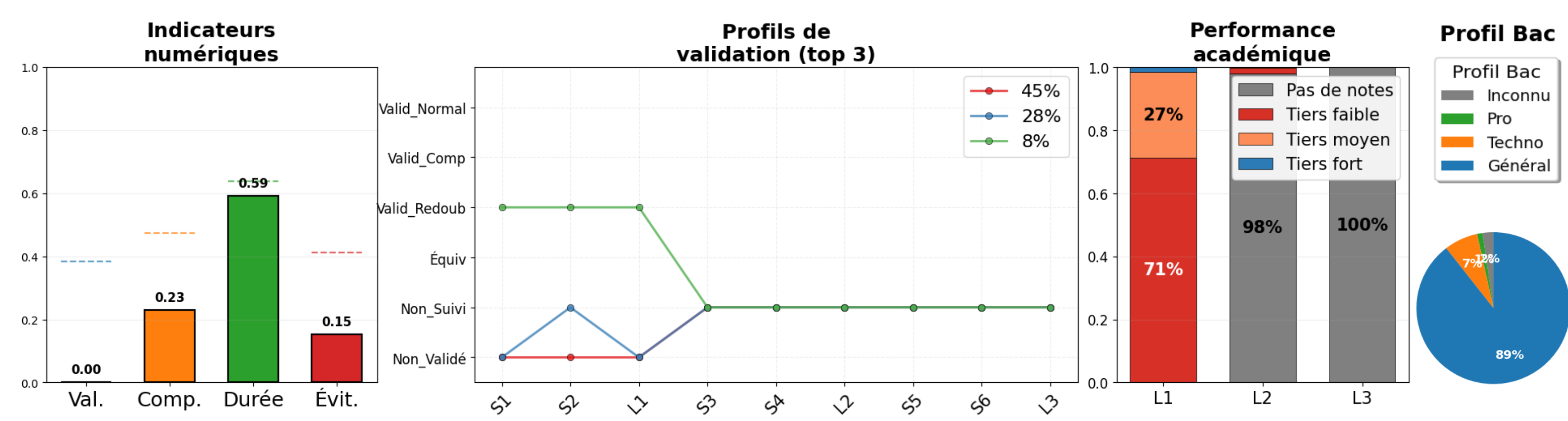
Prototype $Q_l = (\text{moy. num.}, \text{mode cat.})$

$$\text{Minimiser } E = \sum_{l=1}^k \sum_{X_i \in C_l} d(X_i, Q_l)$$

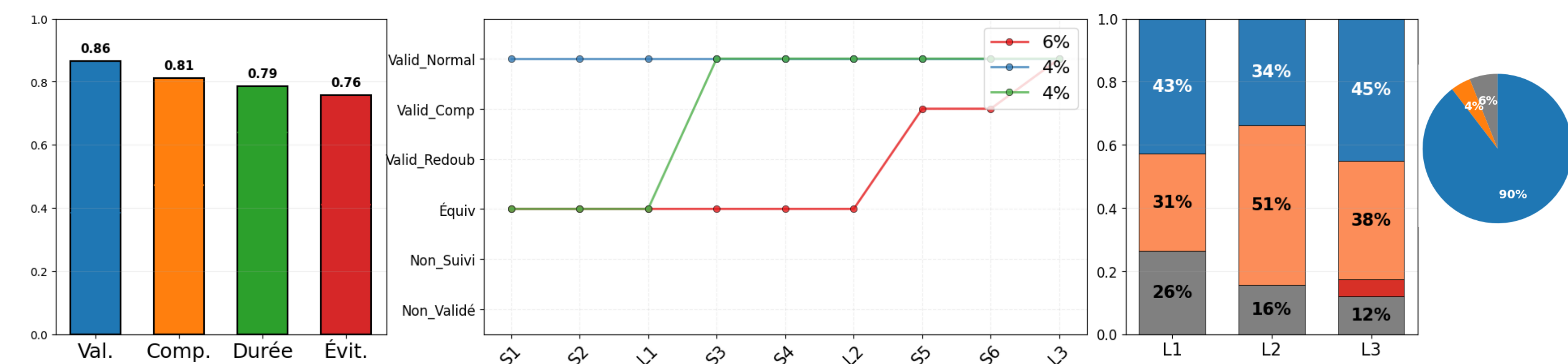
ETAPE 3 : Clustering

Clusters obtenus avec l'algorithme k-prototype appliqué sur 820 étudiants du Portail Science de la Vie

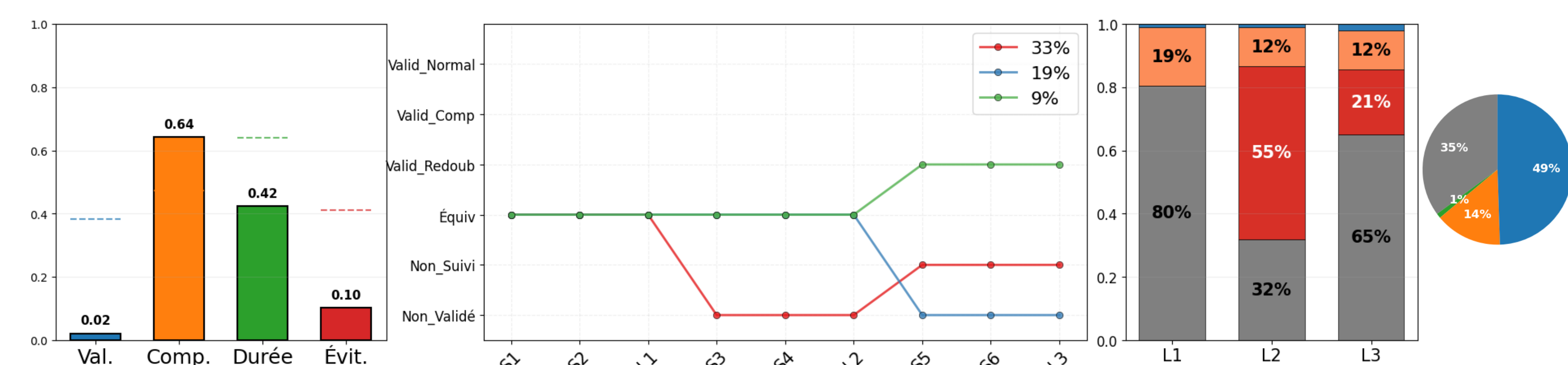
Cluster 1 (352 étudiants) : Profil des « décrocheurs L1 »



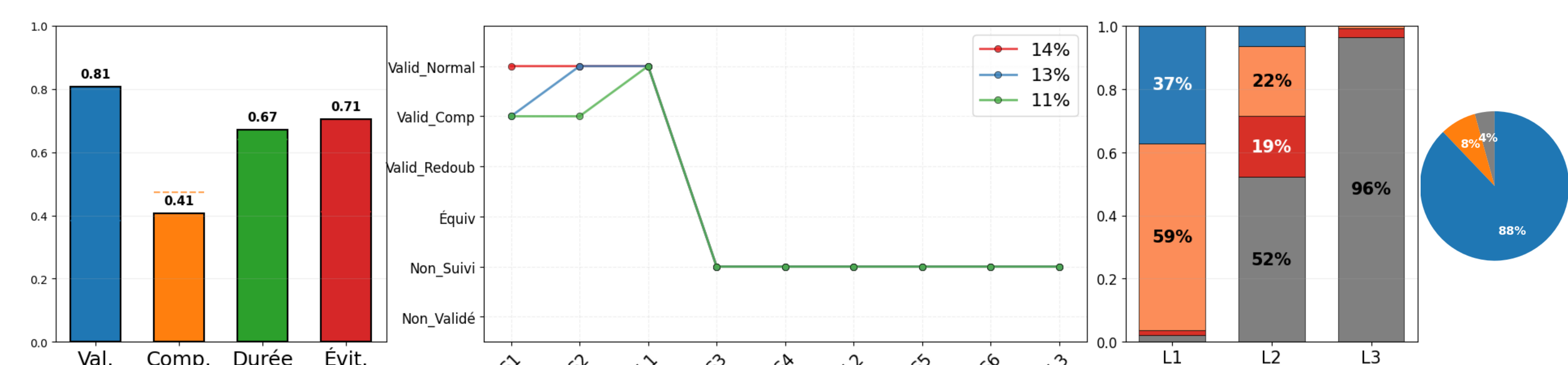
Cluster 2 (231 étudiants) : Profil des « étudiants performants »



Cluster 3 (97 étudiants) : Profil des « entrants tardifs en difficulté »



Cluster 4 (140 étudiants) : Profil des « bons en L1 qui disparaissent »



III. Perspectives

- Déterminer le nombre optimal de clusters par une série d'expérimentations faisant varier k , évaluées via des métriques de validation
- Comparer le clustering selon trois configurations : indicateurs numériques seuls, catégoriels seuls, puis mixtes
- Définir une mesure de dissimilarité selon la nature des indicateurs
- Visualisation des groupes homogènes obtenus
- Comparaison des groupes obtenus