



**HAL**  
open science

## Vers un clustering interactif assisté par l'explicabilité

Yacine Mokhtari, Robin Cugny, Nabil El Malki, Julien Aligon, Max Chevalier,  
Olivier Teste

► **To cite this version:**

Yacine Mokhtari, Robin Cugny, Nabil El Malki, Julien Aligon, Max Chevalier, et al.. Vers un clustering interactif assisté par l'explicabilité. 24ème conférence francophone sur l'Extraction et la Gestion des Connaissances (EGC 2024), Hocine Cherifi, Jan 2024, Dijon, France. pp.287-294. hal-04465483

**HAL Id: hal-04465483**

**<https://hal.science/hal-04465483v1>**

Submitted on 19 Feb 2024

**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.

# Vers un clustering interactif assisté par l'explicabilité

Yacine Mokhtari, Robin Cugny, Nabil El Malki  
Julien Aligon, Max Chevalier, Olivier Teste

Université de Toulouse, IRIT, CNRS – UMR5505, Toulouse, France  
prenom.nom@irit.fr

**Résumé.** Le clustering actif est une technique qui regroupe des données en clusters homogènes par des interactions avec l'utilisateur, qui fournit des retours sous forme de contraintes. La qualité des retours et donc du clustering dépend des connaissances de l'utilisateur. Dans cette étude, les méthodes d'intelligence artificielle explicable (XAI) sont exploitées pour enrichir l'interaction avec le clustering actif en fournissant des informations supplémentaires à l'utilisateur. Nous avons mis en œuvre cette approche avec COBRA, un algorithme de clustering actif, et l'avons évaluée sur divers jeux de données. Nos résultats expérimentaux montrent que les informations générées par les méthodes d'XAI aident un utilisateur simulé à mieux orienter le processus de clustering vers des résultats plus fidèles à la réalité terrain que dans des situations où ces informations ne sont pas disponibles.

## 1 Introduction

Les algorithmes de clustering sont des techniques largement utilisées pour extraire des connaissances dans les ensembles de données pour de nombreuses applications. Ils visent à regrouper une population (données, appelées aussi instances) en sous-groupes homogènes (partageant les mêmes "caractéristiques") (Diday et Simon, 1976). Cependant, malgré leur efficacité, un ensemble large et diversifié de résultats de clustering possibles peut être obtenu en raison de l'absence de la vérité terrain. À ce jour, il n'existe pas de fonction objective permettant d'obtenir un résultat qui correspond aux besoins et aux attentes des utilisateurs selon Bae et al. (2021). L'un des moyens de résoudre ce problème consiste à intégrer les retours de l'utilisateur dans le processus de clustering lorsqu'il est itératif. Le clustering interactif est une approche qui a été proposée pour traiter ce problème. Il a été défini par Chuang et Hsu comme "le processus piloté par l'utilisateur pour affiner et optimiser les clusters en vue d'analyses ultérieures". Bae et al. (2021) indique que les utilisateurs finaux sont généralement désireux d'interagir à la fois avec l'algorithme et les données d'une manière qui leur permette d'acquérir des connaissances à travers le clustering. Cependant, un utilisateur qui explore le jeu de données est difficilement capable de guider un algorithme. La qualité des retours de l'utilisateur dépend également de sa connaissance du jeu de données et de sa compréhension de l'algorithme de clustering. Nous supposons ici que l'utilisateur possède des connaissances d'expert métier et qu'il a une compréhension des résultats et des explications. Ceci définit notre contexte.

L'Explainable Artificial Intelligence (XAI) ou explicabilité propose des méthodes pour aider les utilisateurs à mieux comprendre le fonctionnement et les décisions des modèles d'apprentissage automatique. En particulier, les méthodes d'explicabilité post-hoc comme SHAP (Lundberg et Lee, 2017) et LIME (Ribeiro et al., 2016) sont deux techniques populaires utilisées pour rendre les modèles complexes localement interprétables. En outre, Chuang et Hsu soutient que l'explicabilité est cruciale dans le "clustering interactif", car les utilisateurs veulent comprendre comment les attributs d'entrée conduisent au résultat final. Par conséquent, les informations contenues dans les explications devraient être en mesure d'améliorer les retours utilisateurs et donc le résultat du clustering. Dans cet article, nous proposons un framework visant à intégrer les explications dans le clustering interactif et nous montrons au travers d'un utilisateur simulé qu'il peut produire des clusters plus proches de la vérité terrain. Le reste du document est organisé comme suit. Tout d'abord, nous présentons les approches de clustering interactif ainsi que les méthodes d'XAI adaptées au clustering (Section 2). Ensuite nous présentons le framework proposé pour l'intégration des explications dans les algorithmes de clustering interactif (Section 3). Les résultats expérimentaux montrent que les résultats du clustering sont plus proches de la vérité terrain lorsque les utilisateurs simulés répondent en se basant sur les explications que lorsqu'ils analysent les données brutes (Section 4). Enfin, nous concluons et donnons des perspectives (Section 5).

## 2 État de l'art

**Clustering interactif.** Les algorithmes de clustering interactifs se divisent en deux catégories principales (Bae et al., 2021). Dans la catégorie "initiative venant de l'utilisateur", les algorithmes itératifs génèrent des solutions intermédiaires à chaque itération, et l'utilisateur fournit des retours sous forme de contraintes, telles que des contraintes par paires, des contraintes d'étiquetage, ou des contraintes sur les paramètres de clustering. Ces contraintes guident le processus dans les itérations suivants du clustering. En revanche, dans la catégorie "initiative venant de la machine", l'algorithme identifie les connaissances manquantes à partir de résultats intermédiaires, demandant à l'utilisateur de fournir des réponses spécifiques à certaines requêtes. Un exemple d'approche de cette catégorie est le clustering actif contraint, où l'utilisateur fournit des contraintes pendant le processus sur des groupes d'instances sélectionnées par l'algorithme en fonction du résultat intermédiaire. Des algorithmes tels que COBRA et COBRAS reposent sur cette approche en utilisant des contraintes par paire (Cannot Link (CL)/ Must Link (ML)). COBRAS effectue des opérations de division et de fusion récursives sur des clusters en fonction des réponses de l'utilisateur, tandis que COBRA utilise un processus itératif basé sur des contraintes par paire pour décider de fusionner ou non des clusters.

**Méthodes d'XAI pour les algorithmes de clustering.** La recherche sur les méthodes d'XAI s'est longtemps concentrée sur les modèles d'apprentissage supervisé (Arrieta et al., 2020). En revanche, les travaux sur l'apprentissage non supervisé explicable (Moshkovitz et al.), y compris le clustering, sont beaucoup moins nombreux. Des travaux récents (Escriva et al., 2023) ont été proposés pour montrer l'intérêt d'utiliser des explications afin d'obtenir plus d'informations à partir de clusters obtenus par des modèles non supervisés.

Les *feature importance* sont une vaste famille de méthodes d'XAI qui vise à quantifier la contribution des attributs individuels (variables d'entrée) au résultat ou aux prédictions d'un

modèle. Deux des méthodes de cette famille les plus courantes sont LIME (Ribeiro et al., 2016) et SHAP (Lundberg et Lee, 2017). Elles produisent des explications locales, ce qui signifie qu'elles expliquent une prédiction à la fois. LIME entraîne un modèle interprétable pour approximer des modèles complexes de type boîte noire pour chaque instance. Les paramètres du modèle linéaire correspondent à l'importance des attributs du modèle pour cette instance. La méthode d'explication SHAP (Lundberg et Lee, 2017) calcule les valeurs de Shapley à partir de la théorie des jeux de coalition pour capturer les relations entre les attributs. L'explication est représentée comme une méthode d'attribution additive des attributs, elle agit comme un modèle linéaire. Il convient de noter que LIME et SHAP ont été développées initialement pour des approches supervisées. Par conséquent, Morichetta et al. a suggéré une méthodologie appelée Explain-It, une variante pour les modèles non supervisés tels que le clustering. Elle nécessite d'entraîner un modèle supervisé sur les clusters, et de générer des explications locales avec une méthode comme SHAP ou LIME sur les modèles supervisés. Ces explications répondent à la question "Pourquoi cette instance est-elle mise dans ce cluster?".

### 3 Framework de clustering actif contraint explicable

Dans le clustering actif contraint, l'utilisateur est invité à donner un retour à l'algorithme de clustering sous la forme de contraintes simples ou par paire. Cependant, pour formuler ce retour, l'utilisateur doit avoir des connaissances sur l'ensemble de données et plus particulièrement sur les instances concernées par la requête demandée. Pour aider l'utilisateur à donner une réponse éclairée, nous avons proposé d'intégrer les explications générées par les méthodes d'XAI dans le clustering actif afin qu'elles soient accessibles à l'utilisateur.

Dans cette section, nous décrivons notre framework dans lequel les méthodes d'XAI et les algorithmes de clustering actif sont combinés pour vérifier l'hypothèse suivante : soit  $q(x_A, x_B)$  une requête concernant la paire d'instances  $(x_A, x_B)$  tel que  $x_A$  et  $x_B$  sont deux représentants de deux clusters et soit  $e = (e_A, e_B)$  une explication locale de ces deux points construite sur leurs clusters d'affectation respectifs; Si l'utilisateur s'appuie sur  $e$  pour répondre à  $q$ , alors son retour conduit à un clustering plus pertinent que s'il n'avait pas utilisé  $e$ . Pour évaluer notre hypothèse, nous proposons l'algorithme XCOBRA qui intègre des explications dans les requêtes de COBRA.

#### 3.1 XCOBRA

XCOBRA est une extension de COBRA incluant un composant XAI appelé itérativement. Dans XCOBRA, lors de la phase de requêtage, l'utilisateur obtient des explications justifiant la manière dont un modèle prédictif classerait les deux représentants des clusters. Ces explications sont considérées comme des informations supplémentaires qui ne peuvent pas être vues directement à partir de l'ensemble de données. Elles permettent de formuler une réponse fiable afin d'obtenir un résultat de clustering pertinent pour l'utilisateur. Dans ce papier, nous considérons la vérité terrain comme le résultat pertinent attendu par l'utilisateur. Nous utilisons la méthode Explain-It (Morichetta et al.) qui donne à l'utilisateur final des explications sur les résultats du clustering.

Il génère des explications de type *feature importance* via LIME et/ou SHAP pour chaque représentant des clusters concernés par la requête  $q$ .

### 3.2 Simulation d'utilisateurs

Afin d'évaluer si l'utilisation par les utilisateurs des explications améliore le clustering actif contraint tout en évitant les problématiques relatives à l'expérience utilisateur et au domaine de l'interface homme-machine, nous avons conduit des expérimentations uniquement sur des utilisateurs simulés.

Notons que les explications peuvent être comparées et il est parfois utile pour ce faire de mesurer la similarité entre les explications (pour évaluer la stabilité) (Nauta et al., 2022). La comparaison entre explications ici permet à l'utilisateur de trouver des raisons pour séparer ou réunir les données.

Nous utilisons la similarité cosinus parce qu'elle est normalisée, ce qui nous permet de fixer le même seuil de similarité pour toutes les requêtes d'un ensemble de données. Deux types d'utilisateurs simulés ont été développés. Le premier (*Data User*) utilise uniquement des données brutes dans le processus de décision, tandis que le second (*XUser*) exploite les explications. Ils sont utilisés pour vérifier si le fait de n'utiliser que les explications dans la prise de décision permet d'obtenir un clustering pertinent, par opposition à l'utilisation directe de l'ensemble de données (c'est-à-dire sans explication).

Soit  $z_A$  et  $z_B$  où  $z_A$  (resp.  $z_B$ ) est soit l'instance  $x_A$  soit son explication associée  $e_A$  (resp.  $e_B$ ), une mesure de similarité  $S_C(z_A, z_B)$  et un seuil  $\theta \in [-1, 1]$ , la décision est prise par l'utilisateur simulé comme suit : `must-link` est choisi si  $S_C(z_A, z_B) \geq \theta$  sinon `cannot-link`. Nous définissons  $C(z_A, z_B) = \frac{z_A \cdot z_B}{\|z_A\| \|z_B\|} \in [-1, 1]$ , ce qui correspond à la similarité cosinus. Prenons  $\theta$  constant pour toutes les itérations, c'est-à-dire qu'il est le même pour toutes les requêtes concernant un ensemble de données. Dans nos expériences, nous avons fixé  $\theta = 0,95$  car cela correspond à un seuil strict qui est plus réaliste et qui tend à donner des résultats intéressants dans la section 4.

## 4 Expérimentations et résultats

Dans cette section, nous menons des expérimentations pour répondre aux questions de recherche suivantes : 1) Q1. Les retours de XUser permettent-ils d'obtenir des résultats plus proches de la vérité terrain que les retours de Data User ? 2) Q2. Est-ce que XUser, a un taux de retours corrects plus élevé que celui de Data User ?

jeu de données	# instances	# attributs	# classes
balance-scale	625	4	3
wisc	628	9	2
square5	1000	2	4

TAB. 1 – Jeux de données

### 4.1 Protocole expérimentale

On propose quatre variantes de *XUser* (USVM, UDT, UKNNU, UKNND) qui se distinguent par l'algorithme d'apprentissage supervisé utilisé dans Explain-It, c'est-à-dire respectivement SVM, Arbre de Décision, k plus proches voisins avec poids uniforme, et k plus

proches voisins avec poids basés sur la distance. Chaque *XUser* peut générer deux types d'explications par instance, via LIME et SHAP. Ces utilisateurs sont utilisés dans XCOBRA.

Notre approche est comparée aux concurrents COBRA et XCOBRA combiné avec *Data User*.

XCOBRA et COBRA débutent le processus par le calcul d'un clustering avec k-means ( $K = 25$ ), utilisant des centroïdes éloignés les uns des autres grâce à la méthode k-means++ (Arthur et Vassilvitskii, 2007). La valeur  $K$ , plus élevée que le nombre de classes, est choisie pour éviter l'inclusion de deux classes dans un même cluster initial. Le nombre de requêtes maximal (budget) est fixé à 100 pour COBRA et XCOBRA <sup>1</sup>.

## 4.2 Q1 : Qualité du résultat de clustering

À chaque itération, le résultat du clustering intermédiaire est évalué par rapport à la vérité terrain en utilisant l'Adjusted Rand Index (ARI) (Hubert et Arabie, 1985). L'ARI maximal obtenu au cours du processus et l'ARI final sont reportés respectivement dans les tableaux 2 et 4. En outre, le nombre de requêtes nécessaires pour atteindre l'ARI maximal et le nombre total de requêtes sont renseignés respectivement dans les tableaux 3 et 5.

Chaque combinaison d'utilisateurs simulés *XUser*, de méthodes XAI et de jeux de données a fait l'objet de 20 expérimentations. De même, *Data User* et COBRA ont été expérimentés 20 fois. La moyenne et l'écart-type sont indiqués dans les tableaux. L'introduction de COBRA vise à évaluer la limite maximale atteignable en utilisant la vérité terrain pour répondre aux requêtes. Ce dernier observe les deux instances de la requête et répond à ML si elles sont dans la même classe dans la vérité terrain.

Nous observons le meilleur *XUser* (meilleur ARI maximal pour la table 2 et meilleur ARI final pour la table 4). Nous indiquons également la moyenne de chaque *XUser* afin d'évaluer leur qualité générale.

	COBRA	Data User	Meilleur XUser	Moyenne XUsers
balance-scale	0.41 ± 0.04	0.10 ± 0.03	0.13 ± 0.04 (L-UKNNU)	0.11 ± 0.01
wisc	0.84 ± 0.02	0.29 ± 0.06	0.76 ± 0.05 (L-UKNNU)	0.68 ± 0.08
square5	0.58 ± 0.04	0.38 ± 0.05	0.45 ± 0.06 (S-USVM)	0.43 ± 0.02

TAB. 2 – Moyenne de l'ARI maximal obtenu au cours de 20 processus COBRA/XCOBRA pour chaque utilisateur. La "Moyenne *XUsers*" est la moyenne de tous les utilisateurs simulés utilisant uniquement l'explicabilité. Le nom de l'utilisateur X précédé de "L-" ou "S-" signifie qu'il est combiné avec LIME ou SHAP, respectivement.

Dans le tableau 2, nous constatons qu'en termes d'ARI maximal, il existe une marge d'amélioration par rapport à COBRA, en particulier pour le jeu de données balance-scale. Toutefois, nous constatons que les meilleurs *XUsers* ont de meilleurs scores ARI maximaux que les *Data Users* pour tous les jeux de données. Ils obtiennent également de meilleurs résultats en moyenne.

En outre, si l'on examine le tableau 3, les *XUsers* atteignent toujours leur ARI maximal avec moins de requêtes que *Data User* et COBRA (qui répond aux requêtes à l'aide de la vérité

1. Code disponible ici : [https://github.com/YacineMOK/XCOBRA\\_code](https://github.com/YacineMOK/XCOBRA_code)

## XCOBRA

	COBRA	Data User	Meilleur XUser	Moyenne XUsers
balance-scale	33.70 ± 3.70	37.30 ± 15.71	19.40 ± 4.04 (L-UKNNU)	22.54 ± 2.06
wisc	24.90 ± 0.83	31.40 ± 21.63	17.40 ± 4.90 (L-UKNNU)	23.74 ± 8.03
square5	27.95 ± 2.78	26.65 ± 7.67	20.04 ± 3.64 (S-USVM)	24.34 ± 5.78

TAB. 3 – Nombre moyen de requêtes nécessaires pour obtenir le meilleur ARI avec le meilleur utilisateur X indiqué dans le tableau 2. Cette moyenne est également indiquée pour COBRA avec la vérité terrain et *Data User*.

	COBRA	Data User	Meilleur XUser	Moyenne XUsers
balance-scale	0.41 ± 0.05	0.09 ± 0.03	0.11 ± 0.06 (L-UDT)	0.08 ± 0.02
wisc	0.83 ± 0.03	0.28 ± 0.06	0.65 ± 0.11 (S-USVM)	0.44 ± 0.15
square5	0.56 ± 0.06	0.35 ± 0.05	0.35 ± 0.08 (S-USVM)	0.30 ± 0.03

TAB. 4 – Moyenne de l'ARI final (c'est-à-dire l'ARI obtenu à la fin du processus COBRA/XCOBRA) sur 20 essais pour chaque utilisateur.

	COBRA	Data User	Meilleur XUser	Moyenne XUsers
balance-scale	36.5 ± 3.83	73.5 ± 9.21	28.3 ± 2.03 (L-UDT)	40.99 ± 13.01
wisc	26.35 ± 1.19	99.7 ± 4.21	81.80 ± 13.65 (S-USVM)	53.32 ± 25.38
square5	31.75 ± 2.64	62.65 ± 8.48	29.5 ± 2.82 (S-USVM)	34.04 ± 3.41

TAB. 5 – Nombre moyen de requêtes nécessaires pour obtenir le meilleur ARI avec le meilleur *XUser* présenté dans le tableau 4

terrain). Ce fait est intéressant dans le cadre de le clustering actif, où l'un des défis consiste à réduire le nombre de requêtes. Dans le tableau 4, les scores ARI sont légèrement inférieurs aux scores ARI maximaux (voir Tableau 2) pour COBRA, Data User et les variants XUser. Nous supposons que, dans le résultat de kmeans obtenu au début de COBRA et XCOBRA, certains clusters contiennent plusieurs classes et que leur fusion ultérieure entraîne une légère réduction de l'ARI. En effet, certaines instances sont alors intégrées à tort dans le même cluster final. Nous notons que la moyenne de tous les *XUsers* est moins bonne que celle de Data User dans plusieurs cas, mais que le meilleur *XUser* offre encore des résultats intéressants par rapport à Data User. Dans le tableau 5, le nombre total de requêtes reste inférieur à celui de Data User, ce qui signifie que les décisions prises permettent de résoudre la tâche de clustering avec moins de requêtes. Dans l'ensemble, *XUser* a réussi à obtenir de bons résultats avec peu de requêtes. Cela montre que les explications, utilisées ici comme un nouvel espace de données pour les utilisateurs simulés, ont le potentiel d'accélérer le processus de clustering et de donner des résultats plus proches de la vérité terrain. Dans un cas réaliste où l'utilisateur n'a pas de vérité terrain, nous pouvons imaginer que les explications peuvent être utilisées pour extraire les informations nécessaires pour séparer les instances qui sont plus difficiles à classer autrement.

Jeu de données	Data User	USVM		UDT		UKNNU		UKNND	
		L	S	L	S	L	S	L	S
balance-scale	51	67	45	64	52	67	58	<b>73</b>	55
wisc	45	<b>83</b>	52	67	48	76	55	79	55
square5	54	54	75	57	<b>79</b>	46	68	57	71

TAB. 6 – Pour chaque configuration impliquant un jeu de données, une méthode d’explicabilité (LIME, notée L et SHAP, notée S), et les utilisateurs simulés *XUser* et *Data User*, nous présentons le rapport cumulatif des réponses correctes à la fin du processus de clustering COBRA par rapport au nombre total de requêtes. Les jeux de données balance-scale, wisc, et square5 ont généré respectivement 33, 29, et 28 requêtes via COBRA.

### 4.3 Q2 : Taux de retours corrects

Dans cette section, nous comparons les utilisateurs *XUser* avec *Data User* en termes de taux de retours corrects. Pour ce faire, les retours des utilisateurs simulés sont comparés à la vérité terrain (ensemble de tous les contraintes ML/CL correctes) à chaque itération. Un retour est considéré comme correct si la contrainte ML ou CL proposée pour la paire d’instances concernée dans l’itération en cours est valide dans la vérité terrain. Il est à noter que seule la vérité terrain est utilisée pour décider de fusionner ou non les clusters. Donc, le même processus de COBRA est exécuté pour les utilisateurs simulés *XUser* et *Data User*. Les résultats sont reportés dans le tableau 6. Nous constatons que, la plupart du temps, les *XUsers* ont un taux plus élevé de réponses correctes que celui de *Data User*. Ces explications représentent ainsi un nouvel espace de données qui apparaît être plus informatif que l’utilisation unique des données originales pour regrouper ou séparer les instances. Aucune méthode d’apprentissage (comme SVM, KNN, etc.) ne se distingue de manière constante par rapport aux autres sur l’ensemble des jeux de données. Cette observation s’applique également aux méthodes d’explicabilité (LIME et SHAP). Le choix des méthodes d’apprentissage et d’explicabilité conduit à des explications différentes qui peuvent être plus ou moins simples.

## 5 Conclusion et perspectives

Cette étude a mis en lumière un domaine relativement peu étudié : les explications locales dans le contexte des modèles non supervisés, avec une application au clustering actif. Nous avons proposé et évalué une approche pour l’intégration des explications locales dans les algorithmes de clustering interactif. Grâce à des expériences, guidées par des utilisateurs simulés pour la validation, nos résultats soulignent le potentiel de l’exploitation des explications pour augmenter les connaissances extraites des données brutes. En effet, les explications fournies à l’utilisateur lui permettent d’identifier des informations cachées telles que les relations entre les attributs caractérisant les instances. Ainsi, avec un utilisateur simulé, elles aident l’algorithme de clustering via l’interaction ainsi augmentée par les explications à se rapprocher de la vérité terrain.

À la lumière de ce travail, nous identifions plusieurs directions prometteuses pour des travaux futurs. Le premier consiste à effectuer des tests auprès des utilisateurs en évaluant la qualité des clusters d’utilisateurs avec et sans explications. Ensuite, il conviendrait de tester

## XCOBRA

le système sur des ensembles de données plus importants et plus diversifiés afin d'évaluer sa scalabilité et ses capacités de généralisation.

## Références

- Arrieta, A. B., N. Díaz-Rodríguez, et al. (2020). Explainable artificial intelligence (xai) : Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*.
- Arthur, D. et S. Vassilvitskii (2007). K-means++ : The advantages of careful seeding. Volume 8, pp. 1027–1035.
- Bae, J., T. Helldin, M. Riveiro, S. Nowaczyk, M.-R. Bouguelia, et G. Falkman (2021). Interactive Clustering : A Comprehensive Review. *ACM Computing Surveys* 53(1), 1–39.
- Chuang, J. et D. J. Hsu. Human-centered interactive clustering for data analysis.
- Diday, E. et J. Simon (1976). Clustering analysis. In *Digital pattern recognition*, pp. 47–94. Springer.
- Escriva, E., J. Aligon, J.-B. Excoffier, P. Monsarrat, et C. Soulé-Dupuy (2023). How to make the most of local explanations : effective clustering based on influences. In *ADBIS*.
- Hubert, L. et P. Arabie (1985). Comparing partitions. *Journal of classification* 2, 193–218.
- Lundberg, S. M. et S.-I. Lee (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30.
- Morichetta, A., P. Casas, et M. Mellia. Explain-it : Towards explainable ai for unsupervised network traffic analysis. In *Big-DAMA '19*.
- Moshkovitz, M., S. Dasgupta, C. Rashtchian, et N. Frost. Explainable k-means and k-medians clustering. In *ICML'20*.
- Nauta, M., J. Trienes, S. Pathak, E. Nguyen, M. Peters, et al. (2022). From Anecdotal Evidence to Quantitative Evaluation Methods : A Systematic Review on Evaluating Explainable AI. *arXiv :2201.08164 [cs]*. arXiv : 2201.08164.
- Ribeiro, M. T., S. Singh, et C. Guestrin (2016). " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD*, pp. 1135–1144.

## Summary

Active clustering is a technique that groups data into homogeneous clusters through interactions with the user, who provides feedback in the form of constraints. The quality of the feedback, and therefore of the clustering, depends on the user's knowledge. In this study, explainable artificial intelligence (XAI) methods are employed to enhance the interaction with active clustering by providing additional information to the user. We implemented this approach with COBRA, an active clustering algorithm, and evaluated it on various datasets. Our experimental results show that the information generated by XAI methods assists a simulated user in better guiding the clustering process towards results that are more faithful to the ground truth compared to situations where such information is not available.