



**HAL**  
open science

## Découverte de règles causales dans les graphes de connaissances à l'aide de plongements dans les graphes

L Simonne, N Pernelle, F Saïs, Rallou Thomopoulos

### ► To cite this version:

L Simonne, N Pernelle, F Saïs, Rallou Thomopoulos. Découverte de règles causales dans les graphes de connaissances à l'aide de plongements dans les graphes. IC 2022 : 33èmes Journées francophones d'Ingénierie des Connaissances, Jun 2022, Saint-Etienne, France. hal-03820025

**HAL Id: hal-03820025**

**<https://hal.science/hal-03820025>**

Submitted on 18 Oct 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.

# Découverte de règles causales dans les graphes de connaissances à l'aide de plongements dans les graphes

L. Simonne<sup>1</sup>, N. Pernelle<sup>2</sup>, F. Saïs<sup>1</sup>, R. Thomopoulos<sup>3</sup>

<sup>1</sup> LISN, CNRS (UMR 9015), Université Paris-Saclay

<sup>2</sup> LIPN, CNRS (UMR 7030), Université Sorbonne Paris Nord

<sup>3</sup> INRAE (UMR IATE)

simonne@lisn.fr

## Résumé

La découverte de relations causales est l'objectif de nombreuses expériences. Lorsque des données observationnelles sont disponibles, l'utilisation du cadre d'étude des résultats potentiels est un des standards pour découvrir de telles relations. Dans cet article, nous nous plaçons dans ce cadre afin de découvrir des règles causales au sein de graphes de connaissances (KGs). Ces règles expriment que des différences de traitements conduisent à des différences de valeur pour une caractéristique étudiée. Cependant, ce cadre repose sur la similarité entre instances, et sa quantification dans un KG n'est pas triviale, notamment parce que leurs descriptions peuvent être incomplètes et erronées. Nous proposons une nouvelle méthode de découverte de règles causales qui exploite un appariement basé sur les plongements de graphes de connaissances. Les expérimentations menées sur deux KG de domaines différents ont montré la capacité de notre approche à découvrir des règles qui expliquent plus de différences dans la caractéristique étudiée que les approches existantes, et qui est plus robuste en cas d'information incomplète.

## Mots-clés

Règles causales, Plongements de graphes, Explicabilité, Graphes de connaissances

## Abstract

Discovering causal relationships between different observations is the goal of many experiments in science. When observational data are available, the potential outcome framework is a well-used framework for discovering such relationships. In this paper, we place ourselves in this framework to discover causal rules in Knowledge Graphs (KGs) that express that differences in treatments lead to differences in a studied characteristic. However, quantifying the similarity between individuals represented in a knowledge graph is challenging, especially because their descriptions can be incomplete and erroneous. We propose a new approach based on knowledge graph embeddings to discover causal rules in KGs. The experiments that we conducted on two KGs, including a scientific knowledge graph, showed that our approach is able to discover rules that ex-

plain much more differences in the studied characteristic than existing state of the art approaches.

## Keywords

Causal Rules, Graph Embeddings, Explainability, Knowledge Graphs

## 1 Introduction

L'adoption des technologies du web sémantique pour la représentation des données et des connaissances est en plein essor. Cela a conduit à la disponibilité de nombreux ensembles de données représentés sous forme de graphes de connaissances (KG). Ces KG peuvent contenir des données et des connaissances de différents domaines, et peuvent être exploités pour découvrir de nouvelles connaissances telles que des règles d'associations. Ces associations peuvent être utiles pour prédire de nouveaux faits, détecter des erreurs, mais elles expriment rarement des relations causales.

Une *relation causale* décrit la relation entre deux variables, où une variable nommée *traitement* a un *effet* sur une variable nommée *résultat*. Les relations causales sont intéressantes dans de nombreux domaines, comme par exemple en santé pour déterminer si un médicament traite ou non une maladie. Il existe différents cadres pour étudier le problème de la découverte de relations causales à partir de données d'observation tabulaires, tels que le *modèle causal structurel* [13] ou le *cadre des résultats potentiels* [15]. Dans ce dernier cas, l'effet d'un traitement peut être estimé en comparant des instances similaires qui diffèrent sur le traitement. Bien que de nombreuses approches existent pour découvrir les relations causales dans les données tabulaires, seules quelques approches [10, 17] se concentrent sur la découverte des effets d'un traitement dans les KG. L'approche proposée dans [17] permet de découvrir des règles causales différentielles qui expriment qu'une différence de valeurs sur le traitement explique une différence de valeurs sur le résultat. Cette approche s'appuie sur une étape d'appariement tronqué fondé sur le regroupement de certains chemins de propriétés. Les règles découvertes sont très explicites, car elles représentent explicitement le sous-ensemble de classes d'instances pour lesquelles la règle est valide. Un processus de généralisation des règles est effectué, mais

il conduit à très peu de règles générales. Ainsi, seule une faible partie des différences de résultats peut être expliquée. Dans cet article, nous présentons une nouvelle approche hybride appelée Dicare-E combinant les plongements de graphes et les techniques de fouille de règles symboliques pour découvrir des règles différentielles causales. Ce type de règle permet à la fois d'expliquer des différences de valeur sur une propriété étudiée comme la croissance de la taille d'une tumeur, ou le taux de réinsertion des chômeurs, et d'en tirer des conclusions pour effectuer des décisions. Nous adaptons le cadre des résultats potentiels pour pouvoir découvrir de telles règles. Pour appairer des instances, nous exploitons des plongements pré-entraînés et mesurons la similarité de deux vecteurs intégrés en observant la similarité des prédictions utilisant ces vecteurs. Ainsi, la méthode d'appariement est plus robuste aux données erronées et incomplètes. En outre, elle permet d'obtenir des règles plus générales qui peuvent être appliquées à un plus grand nombre d'instances. Nos contributions sont (i) la définition d'une nouvelle méthode d'appariement basée sur les plongements de graphes, (ii) un algorithme s'appuyant sur ces appariements pour découvrir des règles exprimant des effets causaux de traitements et (iii) une évaluation expérimentale sur un jeu de données réel qui montre l'efficacité de l'approche développée en comparaison avec l'état de l'art.

## 2 Travaux antérieurs

Nous présentons les travaux traitant de la découverte causale dans les données tabulaires et dans les KG, ainsi que les principales méthodes de comparaison d'instances.

**Causalité dans les données tabulaires.** De nombreuses approches existent pour la découverte de relations causales, comme le modèle causal structurel [13] ou les BN [11].

L'étude de la découverte de relations causales peut être introduite par le biais du problème fondamental de l'inférence causale [15]. Étant donné  $Y_i(T)$  le résultat de l'individu  $i$  lors de l'étude du traitement  $T$ , l'effet du traitement individuel est défini comme  $TE_i = Y_i(1) - Y_i(0)$ , c.-à-d. la différence entre le résultat de  $i$  avec et sans le traitement.  $TE_i$  ne peut pas être calculé car, pour un individu donné  $i$ , si  $Y_i(1)$  est observé, le contrefactuel  $Y_i(0)$  ne peut pas être observé. Le standard pour trouver les effets du traitement est de planifier une expérience avec une attribution aléatoire du traitement. De telles expériences sont difficiles à planifier pour des raisons d'éthique et de coût. Par conséquent, la plupart des études causales sont menées en exploitant des données observationnelles, où l'attribution du traitement ne se fait pas de manière aléatoire. Dans ces approches, deux ensembles d'individus sont créés : l'ensemble *traité* des individus ( $T = 1$ ), et l'ensemble *contrôle* des individus non traités ( $T = 0$ ). Comparer naïvement les résultats des deux ensembles introduit un biais de sélection, car la distribution des covariables n'est pas la même dans les deux groupes. Par exemple, en étudiant l'effet de l'âge sur la probabilité qu'une personne ait un cancer, l'ensemble de contrôle pourrait être composé d'une majorité d'hommes, et l'ensemble traité d'une majorité de femmes, ce qui introduirait un biais.

Le cadre d'étude des *résultats potentiels* estime un effet en déterminant un contrefactuel pour les individus traités [15]. Le contrefactuel peut être estimé ou tiré parmi les individus du *contrôle* [18]. Une technique est l'appariement, qui vise à sous-échantillonner l'ensemble de données pour équilibrer la distribution des covariables et supprimer le biais de sélection. Elle consiste à comparer un individu traité à un individu contrôle similaire. Dans l'exemple, cela consiste à comparer des personnes de même sexe, taille et poids (et autres), mais qui diffèrent sur l'âge de traitement. Bien que l'appariement soit populaire, il devient de plus en plus difficile avec la complexité de la description des individus [1]. Des alternatives ont donc été proposées, comme l'appariement tronqué [6] qui consiste à relâcher la contrainte d'appariement exact. En plus de faciliter le processus d'appariement, elle minimise également le biais dans les effets découverts. L'estimation du contrefactuel doit être adaptée aux descriptions RDF des instances d'un KG, pouvant être encore plus complexes que les données tabulaires.

**Causalité dans les graphes de connaissances.** À notre connaissance, seules trois approches se sont concentrées sur l'exploration des relations causales dans les KG. Dans [10], les auteurs proposent de transformer les données du KG en utilisant des connaissances expertes et en apprenant un modèle relationnel probabiliste. Cependant, les résultats représentent la distribution conjointe entre les variables et n'indiquent pas nécessairement les relations causales. De plus, comme elle repose sur des réseaux bayésiens, l'approche ne peut pas prendre en compte de grands graphes de connaissances et elle a été évaluée sur un petit ensemble de données ( $\approx 7k$  triplets). [17] découvre des règles différentielles causales expressives, qui expriment un effet de traitement pour un sous-ensemble d'instances décrites par un motif de graphe nommé *strate*. De telles règles sont intéressantes car elles peuvent montrer des effets locaux. Cependant, une règle définie sur une strate donnée est sélectionnée si sa validité est vérifiée sur toutes les strates les plus spécifiques. Une telle contrainte stricte empêche de découvrir des règles génériques. Enfin, [5] exploite les plongements de graphes afin de découvrir de nouvelles hypothèses scientifiques à partir d'hypothèses recueillies sur un ensemble de papiers scientifiques. Bien qu'elle ne mentionne pas de causalité, cette approche compare des plongements d'une méthode qui peut s'apparenter au cadre d'étude des résultats potentiels.

**Appariement d'instances dans les graphes de connaissances.** Des approches de liaison de données visent à découvrir les liens d'identité représentés par le prédicat *owl:sameAs* dans les KG [12]. Cependant, notre objectif est de comparer des entités distinctes et non identiques. Des prédicats moins stricts comme *identiConTo* [14] expriment une relation d'identité entre des entités dans un contexte conceptuel donné (c.-à-d. une sous-partie de l'ontologie), mais ne fournissent pas de similarité quantifiée et ne sont pas adaptés aux données incomplètes. Plus récemment, [16] développe une approche reposant sur de l'appariement strict, applicable sur des graphes ayant des schémas simples mais non applicable lorsque le schéma est com-

plexe. [17] propose un appariement tronqué pour notamment traiter l'incomplétude des données en utilisant des propriétés abstraites qui sont dérivées de clusters de propriétés obtenus grâce à leur co-occurrence. Le processus de clustering doit cependant être guidé par un expert du domaine lorsque le nombre de propriétés est élevé.

La recherche par similarité a également été étudiée pour la relaxation de requêtes [3]. L'utilisation de telles approches repose sur la définition d'un espace de recherche de requêtes, ce qui n'est pas trivial. La similarité entre instances peut également être quantifiée en étudiant les clauses de Horn basée sur des prédicats et des arguments (non)-partagés par les instances [2]. Reposant sur des descriptions de logique de premier ordre complètes, cette approche n'est pas conçue pour des données incomplètes.

Enfin, les modèles de plongements présentent des caractéristiques intéressantes pour notre approche. Ces modèles capturent la sémantique des entités tels que (i) des instances RDF similaires ont des vecteurs similaires dans l'espace de plongements [19] et (ii) les vecteurs similaires dans l'espace de plongements représentent des instances RDF similaires [9, 7]. Cependant, utiliser ces approches afin de déterminer si des instances sont similaires et non identiques implique de fixer un seuil de distance entre deux vecteurs pour décider si deux instances sont suffisamment similaires.

**Fouille de règles d'association.** Les règles d'association sont composées d'un corps  $\vec{B}$  et d'une tête  $\vec{H}$  et expriment le lien  $\vec{B} \Rightarrow \vec{H}$ . Les méthodes de fouille de règles d'association, e.g. [4], sont utilisées afin d'obtenir de nouvelles connaissances ou de supprimer des triplets erronés. Bien qu'une association soit représentée dans ces règles, elle n'indique pas nécessairement un lien de causalité. De plus, une règle d'association n'indique pas clairement l'effet d'un traitement, et les algorithmes utilisés pour les déterminer ne prennent pas en compte la similarité des instances, i.e. il n'y a pas d'étape de contrôle réalisée. Par exemple, un effet présent dans la tête d'une règle pourrait être du à un traitement non indiqué dans le corps. Ainsi, une telle règle ne peut être utilisée pour expliquer l'effet d'un traitement.

### 3 Préliminaires et Définitions

Nous cherchons des règles causales exprimées en logique du premier ordre sous la forme  $\vec{B} \Rightarrow \vec{H}$  où des différences de valeurs dans  $\vec{B}$ , représentant le *traitement* expliquent des différences de valeurs dans  $\vec{H}$  qui représente le *résultat*.

#### 3.1 Définitions

**Graphes de connaissances.** Nous considérons un graphe de connaissances  $KG$  défini par une paire  $(\mathcal{O}, \mathcal{F})$  où  $\mathcal{O}$  est une ontologie représentée en OWL composée d'un ensemble de classes et de propriétés.  $\mathcal{F}$  est un ensemble de triplets RDF décrivant des instances de classes de  $\mathcal{O}$ .

**Traitement.** Dans le cadre des KG, nous considérons des chemins de propriétés  $P_t : p_1(X, Y_1) \wedge p_2(Y_1, Y_2) \wedge \dots \wedge p_n(Y_{n-1}, Y_n)$  de longueur maximale  $l_{tmax}$ . Les propriétés se trouvant à l'extrémité de ces chemins peuvent avoir des valeurs catégorielles ou numériques, être mono-valuées

ou multi-valuées. Par abus de langage, nous considérons la (les) valeur(s) d'un chemin de propriété comme faisant référence à la (les) valeur(s) de la propriété à l'extrémité du chemin. Les règles s'appliquent à deux instances  $(X_1, X_2)$  d'une classe de  $\mathcal{O}$ , et un traitement  $T$  représente une différence de valeurs de  $P_t$  entre  $X_1$  et  $X_2$ . Nous distinguons un *traitement catégoriel*  $T_c$  impliquant un chemin de propriétés dont l'extrémité est une propriété catégorielle (e.g. littéral, date, valeurs hiérarchisées) d'un *traitement numérique*  $T_n$  impliquant un chemin de propriétés dont l'extrémité est une propriété numérique (e.g. entier, réel).

**Traitement catégoriel.** Soient  $X_1$  et  $X_2$  deux instances d'une classe, un chemin de propriétés  $P_t$  et deux ensembles de valeurs  $V_1$  et  $V_2$  du chemin  $P_t$  pour  $X_1$  et  $X_2$  respectivement. Un traitement catégoriel  $T_c$  est défini par :

$$T_c(X_1, X_2) : P_t(X_1, V_1) \wedge P_t(X_2, V_2) \wedge belongs(v_1, V_1) \wedge belongs(v_2, V_2) \wedge \neg belongs(v_1, V_2) \wedge \neg belongs(v_2, V_1)$$

où  $belongs(v, V)$  est une fonction qui vérifie que  $v$  appartient à l'ensemble des valeurs  $V$ .

**Traitement numérique.** Soient  $X_1$  et  $X_2$  deux instances d'une classe, un chemin de propriétés  $P_t$  et deux ensembles de valeurs  $V_1$  et  $V_2$  du chemin  $P_t$  pour  $X_1$  et  $X_2$  respectivement. Un traitement numérique  $T_n$  est défini par :

$$T_n(X_1, X_2) : P_t(X_1, V_1) \wedge P_t(X_2, V_2) \wedge compare_{T_n}(s(V_1), s(V_2))$$

où  $compare_{T_n}$  est une fonction de comparaison de valeurs numériques pouvant être implémentée par *lessThan* ou *greaterThan* et  $s$  est une fonction d'agrégation qui peut être par exemple *max*, *min*, *sum*, etc.

Par exemple, un traitement  $T_c$  peut être que deux athlètes ont des manualités différentes : l'un est droitier et le second est gaucher. Un traitement  $T_n$  exprime une différence sur une valeur numérique, par exemple que le budget du club d'un athlète est plus élevé que celui d'un autre athlète.

**Résultat.** Nous considérons un chemin de propriétés menant à des valeurs numériques  $P_o : p_1(X, Z_1) \wedge p_2(Z_1, Z_2) \wedge \dots \wedge p_m(Z_{m-1}, Z_m)$ . Soient  $X_1$  et  $X_2$  deux instances d'une classe,  $P_o$ , deux ensembles de valeurs numériques  $V_1$  et  $V_2$  de  $P_o$  pour  $X_1$  et  $X_2$  respectivement, et  $s$  une fonction d'agrégation. Le résultat  $O$  est défini par :

$$O(X_1, X_2) : P_o(X_1, V_1) \wedge P_o(X_2, V_2) \wedge lessThan(s(V_1), s(V_2))$$

**Règle Différentielle Causale.** Une règle causale différentielle  $RDC_T$  représente la relation de causalité entre le traitement  $T$  et son résultat. Elle exprime que le traitement, i.e., une différence de valeurs sur un chemin de propriétés  $P_t$ , explique un résultat, i.e., une différence de valeurs sur un chemin de propriétés  $P_o$  tel que  $lessThan(P_o(s(V_1)), P_o(s(V_2)))$ .

**Définition 1. (Règle Différentielle Causale).** Étant données  $X_1$  et  $X_2$  deux instances d'une classe cible de l'ontologie, le chemin de propriétés menant au résultat  $P_o$ , un traitement  $T \in \{T_n(X_1, X_2), T_c(X_1, X_2)\}$  défini par le chemin de propriété  $P_t$ , et  $s$  une fonction d'agrégation, une règle différentielle causale  $RDC_T$  est définie comme suit :

$$RDC_T : T \wedge P_o(X_1, V_1) \wedge P_o(X_2, V_2) \\ \Rightarrow lessThan(s(V_1), s(V_2))$$

Le résultat  $O$  est exprimé en partie dans le corps de la règle et dans sa conclusion. Une règle impliquant un traitement numérique (resp. catégoriel) est appelée une *règle différentielle causale numérique* (resp. *catégorielle*).

*Exemple.* Soit un KG décrivant des athlètes. Un résultat à étudier pourrait être le classement des athlètes, i.e., expliquer pourquoi des athlètes ont des performances différentes. Une règle différentielle causale numérique  $RDC_{age}$  peut exprimer qu'être plus jeune qu'un autre athlète peut expliquer un meilleur classement, i.e., un rang plus bas :  $age(X_1, Y_1) \wedge age(X_2, Y_2) \wedge lessThan(Y_1, Y_2) \wedge rank(X_1, Z_1) \wedge rank(X_2, Z_2) \Rightarrow lessThan(Z_1, Z_2)$ . Une règle différentielle causale catégorielle  $RDC_{manualite}$  peut indiquer qu'être gaucher plutôt que droitier pourrait être une autre explication d'une meilleure performance.

### 3.2 Effet d'un Traitement

L'effet d'un traitement d'une règle est quantifié par la mesure  $causal_T$  inspirée de [8]. Elle correspond à un odds ratio  $OR$ , qui compare les chances qu'un résultat se produise en fonction d'une exposition aux chances qu'il se produise en l'absence de cette exposition. Alors que l' $OR$  considère un ensemble d'instances,  $causal_T$  considère un ensemble de paires d'instances similaires vérifiant le traitement.

Compte tenu des définitions des règles différentielles causales, nous définissons d'abord les deux supports utilisés pour calculer  $causal_T$ . Soit  $T$  un traitement tel que  $T \in \{T_n(X_1, X_2), T_c(X_1, X_2)\}$ , et  $O$  un résultat,  $supp_{TO}$  représente le nombre de paires d'instances telles que le traitement et le résultat sont tous deux vérifiés :

$$supp_{TO} = \#(X_1, X_2) : \exists \{Y_1, \dots, Z_m\} tq T \wedge O(X_1, X_2)$$

$supp_{T\bar{O}}$  représente le nombre de paires d'instances où le traitement et l'inverse du résultat sont vérifiés, i.e. avec le chemin de propriété  $P_o$  instancié pour les deux instances mais des valeurs numériques qui vérifient le prédicat  $greaterThan$  au lieu de  $lessThan$ . Il est calculé de manière analogue à celui de  $supp_{TO}$ , sauf que le résultat  $O$  est remplacé par  $\bar{O}$ , avec  $\bar{O} \equiv P_o(X_1, V_1) \wedge P_o(X_2, V_2) \wedge greaterThan(s(V_1), s(V_2))$ .

Étant donné un ensemble de paires, la mesure de l'effet d'un traitement est défini dans l'équation 1 suivante :

$$causal_T = \frac{supp_{TO}}{supp_{T\bar{O}}} \quad (1)$$

$causal_T$  retourne un score dans  $[0, +\infty[$  et mesure la force de la relation entre le traitement et le résultat. S'il est égal

à 1, le traitement et le résultat sont considérés comme indépendants, et plus il est différent de 1, plus cette relation est forte. Une valeur supérieure à 1 montre que le traitement et le résultat sont positivement associés.

Nous utilisons cette mesure pour sélectionner les règles pertinentes et ordonner les explications fournies. Un intervalle de confiance  $CI$  est construit pour tester si  $causal_T$  est significativement supérieur à 1 :  $CI_\alpha(causal_T) = exp(\ln(causal_T) \pm u_{1-\alpha/2} \sqrt{\frac{1}{supp_{TO}} + \frac{1}{supp_{T\bar{O}}}})$  avec  $u_{1-\alpha/2}$  le  $(1 - \alpha/2)$  quantile de la loi normale  $\mathcal{N}(0, 1)$ .

## 4 Appariement d'instances fondé sur les plongements de graphes

L'effet d'une règle est calculé en considérant des paires d'instances similaires. Les méthodes classiques d'appariement symbolique pouvant échouer lorsque les descriptions des instances sont hétérogènes, erronées ou incomplètes, nous avons besoin d'une mesure de similarité capable de déterminer des instances similaires, et ce, même lorsque le KG est imparfait. Nous avons défini une nouvelle mesure de similarité exploitant les plongements de graphes.

Dans la Fig. 1, nous présentons le déroulement général de notre approche. Nous considérons un KG et entraînons un modèle de plongements de graphe qui fournit une représentation vectorielle de chaque instance et relation du KG (a). Dans (b), un ensemble de paires est sélectionné aléatoirement, puis la distance de leur vecteurs  $d$  et leur similarité  $sim_e$  (c.f. équation 2 en section 4.1) sont calculées. Dans (c), un modèle  $g(d) = sim_e$  est appris pour définir le seuil de distance  $d_{tr}$  étant donné un paramètre  $sim_{tr}$ . À (d), un ensemble d'instances appariées, i.e., avec une distance  $d$  telle que  $d < d_{tr}$ , est alors créé. Enfin (e), nous calculons l'effet de traitement sur le résultat analysé.

### 4.1 Mesures de similarité sur plongements

**Distance basée sur des vecteurs de plongements.** Étant donné un  $KG$ , nous entraînons un modèle de plongements pour obtenir des vecteurs qui représentent les entités et les relations du  $KG$  dans un espace de faible dimension.

Une distance  $d$  entre deux vecteurs  $v_{i_1}$  et  $v_{i_2}$  peut être calculée à l'aide de diverses fonctions telles que la distance euclidienne. Cette distance peut être utilisée pour décider que deux instances sont similaires en utilisant un seuil donné  $d_{tr}$  (i.e.  $i_1$  et  $i_2$  considérés similaires si  $d(v_{i_1}, v_{i_2}) < d_{tr}$ ). Ce seuil  $d_{tr}$  est à définir pour (i) sélectionner des paires d'instances suffisamment similaires pour analyser un traitement, et (ii) élaguer les paires trop différentes.

La définition d'un tel seuil  $d_{tr}$  est difficile sans connaissance de ce que les distances représentent. Sans seuil défini, l'étape d'appariement pourrait sélectionner, pour une instance  $i_1, i_2$  tel que  $\min_{\forall i_x \in I} d(v_{i_1}, v_{i_x})$ , mais  $i_1$  et  $i_2$  pourraient être très différents. Par exemple, un  $KG$  pourrait décrire des pays comme la France, les États-Unis et l'ensemble des pays d'Asie. Le pays le plus similaire à la France pourrait être les États-Unis, cependant, ces 2 pays restent très différents.

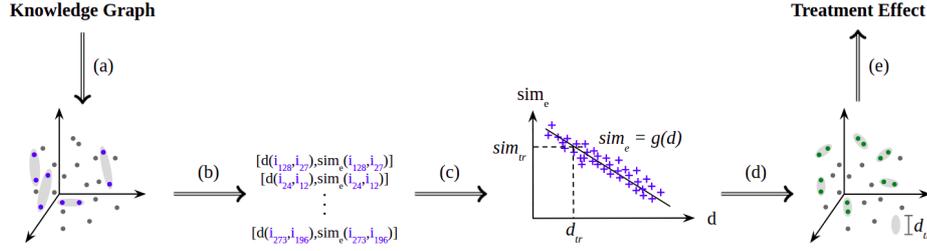


FIGURE 1 – Vision globale de l'approche

**Mesures de similarité basées sur des prédictions de plongements.** Nous proposons  $sim_e$ , une mesure de similarité reposant sur les plongements des instances.  $sim_e$  mesure la similarité de deux vecteurs appris en observant les prédictions utilisant ces vecteurs. Nous supposons que deux instances ayant deux descriptions RDF similaires conduiront à des vecteurs similaires dans l'espace de plongement [19], et que ces vecteurs produiront des prédictions similaires de la part du modèle de plongement  $f$ . Nous rappelons que  $f$  reçoit en entrée un triplet  $(h, r, t)$ , et retourne un score  $f_r(h, t)$  qui sera élevé si le modèle considère le triplet correct, et bas avec un triplet considéré faux. La comparaison des vecteurs permettra à l'approche d'être moins sensible aux données incomplètes ou erronées.

Étant donné un sujet et une propriété  $p_k$ , un modèle de plongements  $f$  prédit un score pour chaque objet possible de  $p_k$  et peut trier les objets par score décroissant. Ainsi, deux instances similaires devraient avoir des prédictions similaires sur les propriétés qu'elles instancient. Pour calculer  $sim_e$ , les prédictions sur chaque propriété  $p_k$  instanciant les instances sont étudiées. Pour une propriété  $p_k$  d'un chemin  $P$  de longueur  $l_P < l_{max}$ , un degré de fonctionnalité  $degre(p_k)$  est défini et représente le nombre moyen d'objets distincts qui instancient  $p_k$ .  $l_{max}$  est un paramètre qui définit la longueur maximale d'étude d'un chemin de propriétés. Avec  $n = degre(p_k)$ , la similarité entre deux instances sur la propriété  $p_k$  est étudiée en analysant les  $n$  premiers objets prédits par  $f$  pour chaque instance. Par exemple, pour les pays visités par une personne, nous pouvons poser le degré de fonctionnalité  $n = 3$ . Ainsi, les 3 pays les mieux classés pour chaque personne sont sélectionnés et utilisés. La similarité  $sim_{p_k}$  de deux instances  $i$  et  $j$  sur une propriété  $p_k$  de la classe cible est calculée récursivement comme suit :

$$sim_{p_k}(i, j) = \begin{cases} 1, & \text{si } p_k(i) = p_k(j) \\ 0, & \text{si } p_k(i) \neq p_k(j), i \in L, j \in L \\ \frac{\sum_{o_i \in p_k(i)} \text{Max}_{o_j \in p_k(j)} sim_e(o_i, o_j)}{degre(p_k)}, & \text{si } p_k(i) \neq p_k(j), i \in I, j \in I \end{cases}$$

où  $L$  (resp.  $I$ ) est l'ensemble des littéraux (resp. des IRIs),  $p_k(i)$  est l'ensemble des objets liés à  $i$  par  $p_k$ .

La similarité  $sim_e(i, j)$  entre  $i$  et  $j$  est obtenue en moyennant les similarités obtenues pour chaque propriété  $p_k$  de

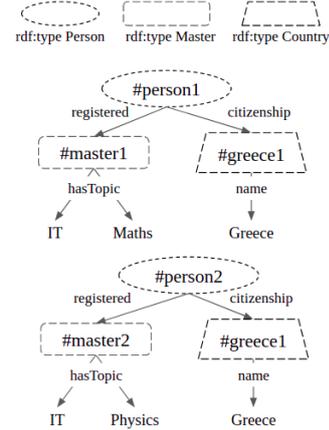


FIGURE 2 – Descriptions prédites de 2 instances

l'ensemble  $P$  des propriétés instanciées par  $i$  et  $j$  :

$$sim_e(i, j) = \frac{\sum_{p_k \in P} sim_{p_k}(i, j)}{|P|} \quad (2)$$

Un poids relatif  $w_i$  représentant l'importance d'une propriété  $p_i$  dans le calcul de similarité peut être introduit et défini par des experts du domaine (par exemple,  $w_i$  peut être fixé à 0 pour le nom d'une personne). Nous proposons un exemple pour illustrer  $sim_e$  basé sur la Fig. 2.

Premièrement,  $sim_{registered}(\#person1, \#person2) = \frac{sim_e(\#master1, \#master2)}{1} = \frac{sim_{hasTopic}(\#master1, \#master2)}{2} = \frac{(max(sim_e(IT, IT), sim_e(IT, Physics)) + max(sim_e(Maths, IT), sim_e(Maths, Physics)))}{2} = \frac{1}{2}$ .

Ensuite,  $sim_{citizenship}(\#person1, \#person2) = 1$  car  $\#greece1$  est l'objet de  $\#person1$  et  $\#person2$ . Enfin,  $sim_e(\#person1, \#person2) = \frac{1}{2} + \frac{1}{2} = \frac{3}{4}$ .

## 4.2 Définition du seuil de similarité $d_{tr}$

$d_{tr}$  est défini en analysant la distribution entre  $sim_e$  et  $d$  (parties (b) et (c) de Fig. 1). Un modèle  $g$ , tel que  $g(d) = sim_e$ , est appris. Étant donné  $sim_{tr}$  le seuil défini sur la similarité  $sim_e$  fixé par l'utilisateur, le seuil  $d_{tr}$  est défini tel que  $g(d_{tr}) = sim_{tr}$ . Des instances peuvent ensuite être appariées en fonction de leur distance et de  $d_{tr}$ .

Le calcul de  $sim_e$  étant complexe en temps, la distribution entre  $sim_e$  et  $d$  est étudiée sur un échantillon de paires.

## 5 Algorithme

L'algorithme Dicare-E est composé de 2 parties : (i) la première consiste à déterminer  $d_{tr}$ , et (ii) la seconde crée des appariements d'instances similaires en utilisant  $d_{tr}$  afin d'évaluer la règle différentielle causale pour un traitement.

### 5.1 Définition de $d_{tr}$

Dans la première partie,  $d_{tr}$  est déterminé. L'algorithme prend en entrée le  $KG$ , la classe cible  $t_c$ , les chemins de propriétés liés au résultat et au traitement étudiés  $P_o$  et  $P_t$ , un modèle de plongement  $f$  et le paramètre  $sim_{tr}$  pour définir  $d_{tr}$ . Le graphe de connaissances  $KG_{tr}$ , utilisé pour entraîner  $f$ , est obtenu en retirant les derniers prédicats de  $P_o$  et  $P_t$  de  $KG$ . Étant donné  $\mathcal{I}_{t_c}$  l'ensemble des instances de  $t_c$ , un échantillon de paires de  $t_c$ ,  $\{(i_i, i_j) \in \mathcal{I}_{t_c} \times \mathcal{I}_{t_c}\} \subset \mathcal{I}_{t_c} \times \mathcal{I}_{t_c}$ , est tiré, et les distances entre les vecteurs correspondants  $d(v_{i_i}, v_{i_j})$  et les  $sim_e$  sont calculées pour chaque paire grâce au modèle appris  $f$ .  $g$  est appris pour obtenir les paramètres de la distribution entre  $d$  et  $sim_e$  pour obtenir  $d_{tr}$  tel que  $g(d_{tr}) = sim_{tr}$ .

### 5.2 Création des paires et découverte de règles $RDC_T$

La deuxième partie consiste à déterminer des règles causales différentielles. En entrée, l'algorithme considère la classe cible  $t_c$ , les chemins  $P_o$  et  $P_t$ ,  $d_{tr}$  et  $\alpha$ .

L'extraction des règles se fait en deux étapes. La première étape consiste à construire les paires d'instances similaires vérifiant  $T$  grâce à  $d_{tr}$ . Premièrement, pour chaque paire  $(i_i, i_j) \in \mathcal{I}_{t_c} \times \mathcal{I}_{t_c}$ , la distance euclidienne entre leurs vecteurs  $d(v_{i_i}, v_{i_j})$  est calculée. Les distances sont stockées dans une matrice de distance  $\mathcal{D}$ . Ensuite, l'ensemble d'instances appariées  $\mathcal{M}$  est créé en sélectionnant à chaque itération la paire présentant la plus petite distance et vérifiant le traitement,  $\min_{(i_i, i_j) \in \mathcal{I}_{t_c} \times \mathcal{I}_{t_c}} \mathcal{D}$  et en l'enlevant de  $\mathcal{D}$  ensuite. Ce processus est appliqué jusqu'à ce qu'il ne reste aucune paire  $(i_i, i_j)$  dans  $\mathcal{D}$  telle que  $d(v_{i_i}, v_{i_j}) < d_{tr}$ .

La deuxième étape consiste à calculer l'effet du traitement. Pour cela, l'algorithme prend en entrée  $\mathcal{M}$ , les chemins  $P_o$  et  $P_t$ , et  $\alpha$ , et est initialisé en fixant à 0 les deux supports décrits dans l'équation 1. Ensuite, pour chaque paire, le traitement et le résultat sont obtenus en utilisant  $P_o$  et  $P_t$ , et les supports sont modifiés en conséquence. Une fois que toutes les paires ont été traitées, l'effet du traitement peut être calculé et son intervalle de confiance construit.

Cet algorithme permet de déterminer l'effet d'un traitement. Afin de déterminer l'effet d'un autre traitement,  $KG_{tr}$  doit être mis à jour en enlevant le nouveau traitement et en ajoutant l'ancien, et  $f$  réentraîné en conséquence.

## 6 Expériences

### 6.1 Données Utilisées

Il n'existe pas de référence pour la découverte de causalité dans les KGs. Nous avons exploité deux KGs déjà utilisés décrits dans Tab. 1 : *Vitamin* [17], et un extrait relativement simple de *DBPedia*, que l'on nomme *DBPediaW* [10, 17].

TABLE 1 – Description des Données

	<i>DBPediaW</i>	<i>Vitamin</i>
# Triplets	6908	86006
# Classes	4	19
# Instances $t_c$	185	1714
# Propriétés	8	22

TABLE 2 – Résultats

	<i>DBPediaW</i>	<i>Vitamin</i>
# Règles ([16])	12	0
# Règles ([17])	12	77
# Règles Dicare-E	3	48
% Paires expliquées ([16])	21,2	0
% Paires expliquées ([17])	21,2	50,1
% Paires expliquées Dicare-E	<b>78,1</b>	<b>92,8</b>
% Règles pertinentes ([17])	83,3	66,6
% Règles pertinentes Dicare-E	<b>100</b>	<b>76,6</b>

*Vitamin* décrit des personnes et leurs caractéristiques socio-économiques telles que leur âge, travail, sexe, régime alimentaire actuel et idéal, opinions sur des faits liés au bien-être animal et au changement climatique. Ce graphe a une profondeur de 2 et 1714 instances de la classe cible *Person*. Nous souhaitons expliquer la différence entre le régime actuel et idéal d'une personne, i.e., sa volonté de réduire sa consommation de viande, indiquée par le prédicat *reduceMeat* ayant des valeurs  $\in \mathbb{N}$ .

*DBPediaW* [10] a une profondeur de 2 et décrit des auteurs, leurs parcours académique et livres. La classe *Writer* est composée de 185 instances. Nous cherchons pourquoi des auteurs publient leur premier livre plus jeunes que d'autres.

### 6.2 Évaluation et Résultats

L'objectif est de montrer l'efficacité et la robustesse de l'utilisation de modèles de plongements de graphes pour la découverte de règles causales différentielles dans les KGs. Nous avons entraîné et évalué plusieurs modèles de plongements. Ensuite, nous avons étudié la distribution entre la distance euclidienne  $d$  et la métrique de similarité  $sim_e$  afin de montrer que cette distribution peut être estimée par un modèle. Enfin, nous avons appliqué notre algorithme pour découvrir des règles causales différentielles et comparons les résultats obtenus à ceux de [16] - appariement strict - et [17] - appariement par communautés.

**Entraînement des modèles de plongements.** *ConvE* est le plus performant pour *Vitamin*, i.e. avec les *MRR* et *Hits@n* les plus élevés, et est donc utilisé par la suite. Pour *DBPediaW*, *DistMult* obtient les meilleures performances. Les performances sont présentées à ce lien <sup>1</sup>.

**Association entre distance  $d$  et similarité  $sim_e$ .** Pour *Vitamin*, la distribution entre  $d$  et  $sim_e$  est présentée dans la Fig. 3. Comme prévu par [19], plus la distance  $d$  entre les vecteurs de deux instances est élevée, moins les instances sont similaires. La distribution entre  $d$  et  $sim_e$  de la Fig. 3

1. <https://github.com/IC2022RuleEmbeddings/Soumission>

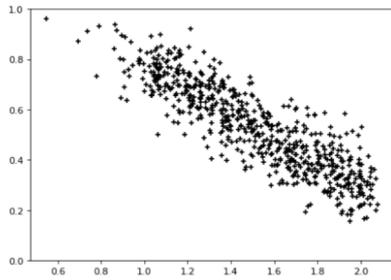


FIGURE 3 – Distribution des valeurs de distance  $d$  (axe  $x$ ) et de similarité  $sim_e$  (axe  $y$ ) pour un ensemble de paires

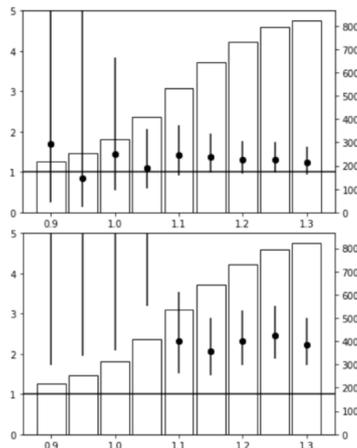


FIGURE 4 –  $causal_{OR_T}$  (axe  $y$  gauche, points) et nombre de paires créées (axe  $y$  droite, histogramme) selon  $d_{tr}$  (axe  $x$ ). Traitements : (haut) genre ( $X_1$  : femme,  $X_2$  : homme) ; (bas) lieu d'habitation ( $X_1$  : campagne,  $X_2$  : aire urbaine)

est modélisée par un modèle linéaire avec ( $r^2 = 0,78$ ). En utilisant ce modèle, le seuil de distance  $d_{tr}$  est fixé à 1,1 pour obtenir une similarité  $sim_{tr}$  de 0,75.

**Règles découvertes et effet de  $d_{tr}$ .** Deux traitements différents sont analysés pour illustrer l'importance de la détermination de  $d_{tr}$  dans la Fig. 4. : le sexe et le lieu de vie. Une règle est valable si la barre d'erreur ne croise pas la barre horizontale placée à 1. Cette figure montre que plus  $d_{tr}$  est faible, plus les barres d'erreurs sont larges. Les paires d'instances des ensembles obtenus sont très similaires, mais la taille de ces ensembles diminuent avec  $d_{tr}$ , résultant en une variance élevée dans l'effet estimé. Inversement, plus  $d_{tr}$  est élevé, plus les ensembles sont grands mais avec des paires moins similaires, menant à une plus faible variance de l'effet et à un biais plus élevé [18]. Le fait que  $d_{tr}$  soit fixé avant le calcul des effets de traitement évite l'introduction d'un biais de la part de l'utilisateur. La Fig. 4 nous indique que, pour  $d_{tr}$  fixé à 1,1, la volonté de réduire sa consommation de viande est indépendante du genre car la barre d'erreur correspondante croise la barre horizontale. Il semble en revanche y avoir un effet du lieu d'habitation, habiter en campagne plutôt qu'en aire urbaine pourrait expliquer une volonté de réduire sa consommation de viande.

$$RDC_{hasDiet} : hasDiet(X_1, omnivorous) \wedge hasDiet(X_2, vegetarian) \wedge reducing(X_1, Z_1) \wedge reducing(X_2, Z_2) \Rightarrow lessThan(Z_2, Z_1) \quad (3)$$

$$RDC_{bornIn} : bornIn(X_1, Y_1) \wedge bornIn(X_2, Y_2) \wedge publishedIn(X_1, Z_1) \wedge publishedIn(X_2, Z_2) \wedge greaterThan(Y_2, Y_1) \Rightarrow lessThan(Z_2, Z_1) \quad (4)$$

### Évaluation qualitative et quantitative et comparaison.

Trois critères ont été évalués. Qualitativement, les règles des deux approches ont été évaluées. Quantitativement, nous avons étudié le pourcentage de paires pour lesquelles les approches peuvent fournir une explication concernant une différence de régimes, et testé la robustesse des deux approches aux données incomplètes.

Notre approche découvre 48 règles différentielles causales pour *Vitamin*, et 3 pour *DBPediaW*. 2 règles sont présentées dans les équations 3 et 4<sup>2</sup> et invitons le lecteur à visiter le GitHub pour obtenir la liste exhaustive des règles. La règle de l'équation 3 exprime qu'être omnivore plutôt que végétarien explique une volonté plus forte de réduire sa consommation de viande. Cette règle fait sens car, pour  $A$  et  $B$  deux personnes similaires, si  $A$  consomme déjà moins de produits animaux que  $B$ , alors  $B$  est plus susceptible de réduire sa consommation de produits animaux. La règle de l'équation 4 exprime qu'être né plus tard qu'un autre auteur explique une publication à un âge plus jeune de son premier livre. Une explication peut être le nombre grandissant d'éditeurs ou les réductions de coûts de publication.

*Vitamin*. Les règles découvertes par [17] et Dicare-E ont été évaluées par un expert du domaine. Aucune instances n'ont pu être appariées avec [16] qui nécessite que leurs représentations soit isomorphes au traitement et au résultat près. En conséquence, aucune règle n'a pu être extraite (voir table 6.1). Pour faciliter la comparaison des approches, les 30 meilleures règles de [17] et Dicare-E ont été évaluées. Les règles générées dans [17] peuvent utiliser un motif de graphe plus ou moins spécifique, nommé strate, définissant l'ensemble des instances auxquelles la règle est appliquée. L'expert a souligné que plus la strate est spécifique, plus elles sont difficiles à interpréter, alors que la plupart des règles de Dicare-E sont faciles à interpréter.

Pour chaque règle, quatre possibilités sont étudiées par l'expert : la règle (1) *semble être pertinente*, (2) *pourrait être pertinente*, (3) *l'expert ne sait pas si la règle est pertinente* et (4) *semble être fausse*. Pour Dicare-E, 11 règles ont été classées dans (1), 12 dans (2), 6 dans (3) et 1 dans (4). Pour [17], 12 règles ont été classées en (1), 8 en (2), 6 en (3) et 4 en (4). 66,6% des règles de [17] semblent donc pertinentes ou pourraient l'être, contre 76,6% pour Dicare-E. De plus, Dicare-E n'a qu'une seule règle qui semble fausse, contre 4 pour [17]. Par ailleurs, notre approche fournit un ensemble

2. La représentation des chemins a été simplifiée, les *belong* sont omis, et la fonction  $s$  n'est pas définie car les propriétés numériques sont monovaluées dans *Vitamin* et *DBPediaW*.

plus large de traitements, des chemins de propriété supplémentaires par rapport à [17] ont été déterminés.

Le pourcentage de paires pour lesquelles au moins une explication du résultat peut être fournie est de 92,8% pour notre approche et de 50,1% pour [17]. En supprimant 15% des triplets de *Vitamin* pour tester la robustesse des deux approches, le nombre de règles extraites et le pourcentage de paires expliquées ont été réduits dans les deux approches : le pourcentage de paires expliquées est passé de 92,8% à 89,5% pour notre approche, et de 50,1% à 24,7% pour [17]. Ceci s'explique du fait qu'une instance avec une description incomplète ne sera pas utilisée pour la détermination de règles dans [17] alors qu'elle le sera dans notre approche.

*DBPediaW*. Les 3 règles obtenues par Dicare-E sont compréhensibles et semblent pertinentes. De plus, elles comportent des traitements également obtenus dans [17]. Sur les 12 règles obtenues par [17], 9 semblent pertinentes. Les résultats de [16] sont les mêmes que [17] car, le schéma étant simple, aucune détection de communauté n'a été réalisée. Il est intéressant de voir que, par opposition à *Vitamin*, pour un *KG* ayant un schéma relativement simple comme *DBPediaW*, les règles issues de [17] sont facilement compréhensibles car contenant peu d'éléments dans leurs strates. Elles sont plus expressives que Dicare-E, qui expriment un effet moyen d'un traitement sur les instances de *DBPediaW*. Dicare-E permet d'expliquer 78,1% des paires contre 21,2% pour [17]. En enlevant 15% des triplets, les règles obtenues avec notre approche sont les mêmes. Avec [17], ce pourcentage passe à 6,0%. Comme sur *Vitamin*, notre approche permet d'expliquer plus de paires et est plus robuste aux données manquantes.

## 7 Conclusion

Nous avons proposé une approche combinant des modèles de plongements de graphes et des techniques de fouille de règles symboliques pour la découverte de règles différentielles causales dans les graphes de connaissances. Cette approche hybride est capable de traiter des données incomplètes tout en fournissant des règles interprétables pouvant expliquer les différences dans une caractéristique numérique étudiée. Elle peut être utilisée pour apparier des instances similaires grâce à leur représentation dans un espace de plongements, permettant ainsi l'application du cadre des résultats potentiels. La métrique de similarité proposée, basée sur les prédictions du modèle de plongements, garantit la création de paires similaires uniquement. Notre expérience montre que notre approche peut être utilisée sur des domaines variés ainsi que sur des *KGs* complexes.

## Références

- [1] Althausen, R.P., Rubin, D. : The Computerized Construction of a Matched Sample. *American Journal of Sociology* 76(2), 325–346 (1970)
- [2] Ferilli, S., Basile, T.M., Biba, M., Di Mauro, N., Esposito, F. : A general similarity framework for horn clause logic. *Fundamenta Informaticae* 90(1-2), 43–66 (2009)
- [3] Ferré, S. : Answers Partitioning and Lazy Joins for Efficient Query Relaxation and Application to Similarity Search. *Lecture Notes in Computer Science* 10843 LNCS, 209–224 (2018)
- [4] Galárraga, Luis Teflioudi, Christina Hose, Katja Suchanek, Fabian. (2013). *AMIE : Association rule mining under incomplete evidence in ontological knowledge bases*. WWW 2013. 413-422.
- [5] Haan, Rosaline Tiddi, Ilaria Beek, Wouter. *Discovering Research Hypotheses in Social Science Using Knowledge Graph Embeddings*. *The Semantic Web* 477-494 (2021)
- [6] Iacus, S.M., King, G., Porro, G. : Causal inference without balance checking : Coarsened exact matching. *Political Analysis* 20(1), 1–24 (2012)
- [7] Jain, N., Kalo, J.C., Balke, W.T., Krestel, R. : Do embeddings actually capture knowledge graph semantics ? *The Semantic Web*. pp. 143–159.
- [8] Li, Jiuyong le, Thuc Liu, Lin Liu, Jixue Jin, Zhou Sun, Bingyu. (2013). *Mining Causal Association Rules*. ICDMW 2013. 114-123.
- [9] Moon, C., Jones, P., Samatova, N.F. : Learning entity type embeddings for knowledge graph completion. *CIKM '17* p. 2215–2218. ACM, NY, USA (2017)
- [10] Munch, M., Dibie, J., Wuillemin, P., Manfredotti, C.E. : Towards interactive causal relation discovery driven by an ontology. In : *FLAIRS* (2019)
- [11] Neapolitan, R.E. : *Learning Bayesian Networks*. Pearson Prentice Hall. (2003)
- [12] Nentwig, M., Hartung, M., Ngomo, A.N., Rahm, E. : A survey of current link discovery frameworks. *Semantic Web* 8(3), 419–436 (2017)
- [13] Pearl, J. : *Causality*. Cambridge University Press (2009)
- [14] Raad J., Pernelle N., Saïs F. *Detection of Contextual Identity Links in a Knowledge Base*. In *Proceedings of the Knowledge Capture Conference (K-CAP 2017)*. Association for Computing Machinery
- [15] Rubin D. B : Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology* 66(5), 688–701 (1974)
- [16] Simonne, L., Pernelle, N., Sais, F. : Fouille de règles différentielles causales dans les graphes de connaissances, *EGC 2021*, pp.293-300
- [17] Simonne L., Pernelle N, Saïs F, Thomopoulos R. *Differential Causal Rules Mining in Knowledge Graphs*. In *Proceedings of the Knowledge Capture Conference (K-CAP '21)*. Association for Computing Machinery.
- [18] Stuart, E.A. : Matching methods for causal inference : A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics* 25(1), 1–21 (2010)
- [19] Wang, C., Pan, S., Hu, R., Long, G., Jiang, J., Zhang, C. : Attributed graph clustering : A deep attentional embedding approach. *IJCAI 2019* (2019), 3670-3676.