

# Fouille de règles différentielles causales dans les graphes de connaissances

Lucas Simonne\* , Nathalie Pernelle\*\* , Fatiha Saïs\*

\*LRI, Université Paris Saclay  
simonne@lri.fr, saïs@lri.fr

\*\*LIPN, Université Sorbonne Paris Nord  
pernelle@lipn.univ-paris13.fr

**Résumé.** L'intérêt porté aux graphes de connaissances a grandi ces dernières années, à la fois dans le milieu académique et industriel. Nous proposons une approche permettant de déterminer des règles différentielles causales dans des graphes de connaissances. Ces règles expriment qu'un traitement différent pour deux instances d'une classe conduit à un résultat différent pour ces instances. L'approche que nous proposons est basée sur des appariements sémantiques, et des strates définies comme étant des sous-classes complexes. Une première expérimentation sur un extrait de DBPedia montre que de telles règles peuvent permettre d'expliquer certaines variabilités dans les propriétés étudiées.

## 1 Introduction

Les graphes de connaissances permettent de stocker des données et des connaissances sur différentes thématiques, ie. expériences scientifiques, politique, finance, etc. Ce type de données est aujourd'hui utilisé dans le milieu académique ou industriel. De nombreux travaux de recherche récents ont permis d'effectuer de la fouille de règles plus ou moins expressives dans les graphes de connaissances (Galárraga et al. (2013)). Ces règles permettent de corriger des faits erronés ou encore de déduire des faits manquants. Cependant, il n'existe que très peu d'approches qui ont pour objectif de découvrir des règles causales dans de tels graphes.

Un grand nombre d'expérimentations sont menées afin d'étudier l'existence de liens causaux, et ce quel que soit le domaine, qu'il s'agisse de médecine, d'environnement ou encore de marketing. De nombreux travaux de recherche ont pour objectif de découvrir des règles causales dans des données relationnelles. Ces travaux sont basés sur les modèles causaux structurels (Pearl (2009)), les réseaux bayésiens (Neapolitan (2003)) ou encore sur les modèles à résultats potentiels - i.e. *potential outcome* en anglais (Rubin D. B (1974)). Une approche standard consiste à construire des expériences où l'attribution d'un traitement est aléatoire afin de déterminer son effet au sein d'une population. Ces expériences sont toutefois compliquées à mettre en place, pour des raisons multiples et en particulier le coût de ces expériences ou encore les problèmes éthiques qu'elles

peuvent poser. Il est par exemple difficile de forcer des individus à commencer à fumer. Aussi, les données généralement utilisées dans le cadre d'études de liens causaux proviennent de données observationnelles, données pour lesquelles le chercheur n'a pas pu interférer. Dans ce cas, l'attribution du traitement n'est plus aléatoire, et d'autres méthodes sont mises en place pour déterminer l'effet d'un traitement.

L'approche que nous proposons permet de découvrir des règles différentielles causales dans des graphes de connaissances en se plaçant dans un cadre d'étude de résultats potentiels. Le but de notre approche est plus précisément de découvrir des règles permettant d'expliquer des différences de résultats par des différences de traitements. Nous pensons en effet que de telles règles peuvent être utiles en sciences ou en sciences économiques et sociales, afin d'expliquer, par exemple, les raisons possibles d'une différence de salaire, ou d'une date de floraison différente entre des plants de différents champs. Notre approche se base sur une représentation sémantique des règles causales dans laquelle l'appariement de paires d'instances et les strates sont définies par des sous-classes composant une ontologie, et des chemins de propriétés aboutissant à des valeurs littérales. Cette approche permet de sélectionner les règles les plus générales qui ne sont pas sensibles à des paradoxes statistiques où le comportement de la règle s'inverse quand l'on considère classes et sous-classes.

Après avoir présenté l'état de l'art, nous présenterons les règles différentielles causales, l'algorithme et les résultats obtenus lors d'une première évaluation.

## 2 État de l'Art

**Causalité et Graphe de Connaissances.** À notre connaissance, une seule approche (Munch et al. (2019)) traite de la découverte de liens causaux dans les graphes de connaissances. Dans ce travail, les données RDF sont transformées en un schéma relationnel en se basant sur l'ontologie et des connaissances expertes, puis un réseau bayésien est entraîné sur ces données. Cependant, les distributions conjointes résultantes n'indiquent pas de relations causales.

**Règles causales dans des données relationnelles.** L'étude de règles causales est un thème de recherche actif dans les données relationnelles et différents cadres d'étude existent (Pearl (2009), Neapolitan (2003)).

**Étude des résultats potentiels.** Un autre cadre d'étude important est celui de l'étude des résultats potentiels (Rubin D. B (1974), Holland (1986)). Dans ce cadre, les individus d'un système sont des unités caractérisées par un ensemble d'attributs. Son objectif est de déterminer l'effet d'un traitement  $X$  sur un résultat  $Y$ . Un traitement peut être un attribut ou une combinaison d'entre eux. Soit  $X$  un traitement binaire et  $Y$  un résultat binaire, afin d'étudier l'effet d'un traitement une solution serait de comparer le résultat et le contrefactuel sur le même individu, mais cela est impossible. L'étude des résultats potentiels a pour but de déterminer une estimation de ce contrefactuel afin de déterminer l'effet du traitement.

Dans le cas où les données utilisées sont observationnelles, plusieurs méthodes existent. Stuart (2010) propose une vue d'ensemble des méthodes utilisées, comme l'appariement exact ou tronqué (Iacus et al. (2012)), l'appariement pondéré ou stratifié et l'appariement sur score de propension (Rosenbaum et Rubin (1984)). Le but

de ces méthodes est de créer des sous-ensembles composés d'unités des deux groupes ayant des distributions similaires sur les attributs. D'autres variantes existent telle que Li et al. (2017) qui combine l'appariement sur score de propension et les arbres de décision. Cependant, King et Nielsen (2019) ont montré que l'appariement sur score de propension était n'était pas optimal, augmentant le biais. Li et al. (2013) ont développé une méthode permettant de découvrir des règles d'association causales à partir de règles d'associations générées par l'algorithme Apriori. Cette approche utilise une méthode d'appariement tronqué et une mesure de qualité inspirée du oddsratio. Le fait de se limiter aux règles d'associations permet de réduire l'espace de recherche, mais ne permet pas de générer toutes les règles causales qui pourraient être d'intérêt.

Notre approche traite de données observationnelles décrites dans des graphes de connaissances et se place dans le cadre de l'étude des résultats potentiels, en utilisant une technique d'appariement sémantique. L'approche n'a pas pour objectif de découvrir des règles permettant d'inférer un résultat particulier pour un traitement, mais d'expliquer des différences de résultats par des différences de traitement.

## 3 Fouille de règles différentielles causales

### 3.1 Définitions

Un graphe de connaissances peut être séparé en deux parties : les données et l'ontologie. Les données sont représentées en RDF<sup>1</sup> et l'ontologie en OWL<sup>2</sup>. Un tel graphe est caractérisé par un couple  $(\mathcal{F}, \mathcal{O})$  où :  $\mathcal{F}$  est une collection de triplets RDF ;  $\mathcal{O}$  est une ontologie définie par  $\mathcal{O} = (\mathcal{C}, \mathcal{DP}, \mathcal{OP}, \mathcal{A})$  tel que  $\mathcal{C}$  est l'ensemble des classes,  $\mathcal{DP}$  est l'ensemble des propriétés littérales,  $\mathcal{OP}$  l'ensemble des propriétés objets et  $\mathcal{A}$  l'ensemble des axiomes.

Dans notre approche, les valeurs associées aux traitements et au résultat sont définies par des chemins de propriétés qui décrivent la classe cible. Une règle différentielle causale ne pouvant généralement pas s'appliquer à toutes les instances d'une classe, nous définissons la notion de strate qui sera utilisée pour considérer un sous-ensemble d'instances de la classe cible pour lesquelles une telle règle s'applique.

**Definition 1. *Strate.*** Soit  $i_1$  une instance de la classe cible  $C$ . Une strate  $ST_j(i_1)$  est une conjonction de prédicats qui correspond à un motif de graphe en RDF, dont la racine est  $i_1$ , et pour lequel les feuilles représentent des valeurs littérales ou des classes.

La taille d'un motif de graphe définissant une strate peut être limité par une profondeur  $d$  correspondant à la longueur maximale d'un chemin atteint par ce motif de graphe.  $TennisMan(X)$  est une strate valide pour la classe *Sportif* (i.e.  $TennisMan \subseteq Sportif$ ). Une autre strate est la suivante :  $TennisMan(X) \wedge nationalite(X, "France")$ .

**Definition 2. *Règle Différentielle Causale.*** Soit  $C \in \mathcal{C}$  la classe cible de l'ontologie et pr le résultat étudié, une règle différentielle causale  $B \Rightarrow R(X, Y)$  est définie par :  $ST_j(i_1) \wedge ST_j(i_2) \wedge pt(i_1, v_1) \wedge pt(i_2, v_2) \wedge compare_t(v_1, v_2) \wedge pr(i_1, z_1) \wedge pr(i_2, z_2) \Rightarrow compare_r(z_1, z_2)$

1. <https://www.w3.org/RDF/>

2. <https://www.w3.org/OWL/>

où :

- $pt$  (resp.  $pr$ ) est un chemin de propriétés menant à la valeur de la propriété du traitement (resp. du résultat),
- $compare_t$  et  $compare_r$  sont des prédicats optionnels permettant de comparer  $v_1$  et  $v_2$ , et  $z_1$  et  $z_2$ , (i.e. opérateurs arithmétiques  $<$ ,  $>$  ou fonction de comparaison de valeurs symboliques hiérarchisées),
- $ST_j$  est la  $j$ -ème strate de  $C$  (i.e.  $C$  ou une expression de classe plus spécifique).

Nous présentons deux exemples de règles n'impliquant pas de strate :

(1)  $Athlete(i_1) \wedge Athlete(i_2) \wedge age(i_1, v_1) \wedge age(i_2, v_2) \wedge inferior(v_1, v_2) \wedge performance(i_1, z_1) \wedge performance(i_2, z_2) \Rightarrow superior(z_1, z_2)$

(2)  $Athlete(i_1) \wedge Athlete(i_2) \wedge Materialorigin(i_1, c_1) \wedge continentName(c_1, "Asia") \wedge Materialorigin(i_2, c_2) \wedge continentName(c_2, "Europe") \wedge performance(i_1, z_1) \wedge performance(i_2, z_2) \Rightarrow superior(z_1, z_2)$

### 3.2 Appariement d'instances et contexte d'identité

Nous utilisons la notion d'identité contextuelle définie par Raad et al. (2017) pour définir sémantiquement comment deux descriptions RDF peuvent être appariées.

**Definition 3. Contexte d'identité.** Soit  $KG$  un graphe de connaissances, le contexte d'identité  $IC$  d'une paire d'instances d'une classe  $(i_1, i_2)$  est un sous-ensemble de l'ontologie  $O$  dans laquelle les descriptions des instances sont identiques à la réécriture de l'URI près. Le prédicat  $identiConTo_{IC}(i_1, i_2)$  représente l'identité entre  $i_1$  and  $i_2$  dans le contexte d'identité  $IC$ .

Dans de nombreux contextes, deux instances de classe décrites par des URIs différentes peuvent référer au même objet. C'est pourquoi l'inégalité des URIs n'est pas prise en compte pour définir l'identité. Si l'on considère un graphe de connaissances décrivant des athlètes, un contexte d'identité peut être le sous-ensemble de l'ontologie composé des classes  $Athlete$  et  $Pays$  ainsi que des  $DP$  *nationalité*, *nomPays* et *regimeAlimentaire*. De façon analogue, le contexte de différence  $DC$  est un ensemble de chemins de propriétés pour lesquelles les valeurs littérales atteignables sont définies et différentes. Dans l'algorithme, les strates utilisent des propriétés des contextes d'identité tandis que le traitement appartient au contexte de différence.

Une paire d'instances est prise en compte lorsque la conjonction de ses contextes d'identité et de différence contient l'ensemble des propriétés pertinentes déclarées. Par exemple, le prénom d'un athlète semble non pertinent à l'étude de sa performance.

### 3.3 Mesure de Qualité

Pour évaluer la qualité des règles, nous mesurons le rapport entre le nombre d'exemples positifs et le nombre d'exemples négatifs. Cette mesure est inspirée de Sebag et Shoenauer (1988), et adaptée aux graphes de connaissances et à la prise en compte de l'appariement sémantique (i.e. seules les couples d'instances appariées sont exploitées pour évaluer la règle).

Soit la règle  $R : \vec{B} \Rightarrow compare_r(z_1, z_2)$ , la mesure *seb-kg* est :

$$seb - kg(R) = \frac{\#(i_1, i_2):identiconTo_{IC}(i_1, i_2) \wedge \exists z_1, \dots, z_n: \bar{B} \wedge compare_r(z_1, z_2)}{\#(i_1, i_2):identiconTo_{IC}(i_1, i_2) \wedge \exists z_1, \dots, z_n: \bar{B} \wedge \neg compare_r(z_1, z_2)}$$

On considère qu'une règle est valide si sa mesure de qualité est significativement supérieure à 1. Un test statistique est réalisé afin de vérifier cette hypothèse.

### 3.4 Strates et Généralisation de règles causales

L'utilisation de strates permet de découvrir des règles valides sur un sous-ensemble d'instances. Ces règles pourraient ne pas être valides sur des ensembles plus généraux (paradoxe statistique). Notre approche permet de fusionner les strates des règles spécifiées pour des strates différentes si leur effet avéré est le même sur la règle généralisée.

L'ensemble théorique des strates ne peut être exploré du fait d'un manque de données, ou car certaines d'entre elles ne peuvent être instanciées. Il sera par exemple difficile d'étudier des skieurs provenant d'un pays où il n'y a ni montagne ni neige. L'effet d'un traitement par rapport à un autre peut donc être inconnu pour certains ensembles. Dans de tels cas, le système peut se focaliser uniquement sur les strates pour lesquelles toutes les règles identiques impliquant des strates plus spécifiques sont valides (i.e. approche pessimiste), ou bien faire des hypothèses sur les strates manquantes.

### 3.5 Algorithme de fouille de règles différentielles causales

L'algorithme prend en entrée un graphe de connaissances, une classe cible  $C$ , le chemin de propriétés représentant le résultat  $pr$ , et un ensemble de propriétés non pertinentes  $NP$ . En sortie, il renvoie l'ensemble de règles différentielles causales associées aux données  $CR$ .

La première étape consiste à parcourir les paires d'instances de  $C$  afin de déterminer le contexte d'identité et de différence de chacune d'entre elles (hors  $pr$ ). L'algorithme utilisé est inspiré de Raad et al. (2017), mais renvoie un contexte de différence en plus du contexte d'identité. De plus, les valeurs des propriétés non fonctionnelles sont ramenées à une seule valeur pas des fonctions choisies par l'expert (e.g. max, avg, etc.). De plus, les propriétés numériques, excepté  $pr$ , sont discrétisées afin de définir des strates d'intérêt. Un échantillonnage est réalisé afin de réduire l'espace de recherche.

Un arbre permet de représenter les paires analysées pour chaque contexte d'identité et de différence. Seuls les noeuds dont le contexte de différence se réduit à un chemin de propriété (traitement) sont considérés. L'arbre est parcouru en largeur afin de déterminer des règles différentielles causales. Pour chaque noeud, le traitement peut être instancié par des paires de valeurs différentes. Pour chaque possibilité, l'ensemble des strates les plus spécifiques sont définies et les règles correspondantes sont évaluées en utilisant  $seb - kg$ . Un test statistique utilisant un intervalle de confiance à  $x\%$  de est réalisé et la règle est validée en fonction de ce résultat. L'algorithme est présenté ci-dessous.

1. Pour chaque paire  $(i_1, i_2) \in C \times C$ 
  - (a) calcul du contexte d'identité  $IC(i_1, i_2)$  et de différence  $DC(i_1, i_2)$
  - (b) attribution de  $(i_1, i_2)$  au noeud  $n_k$  représentant les contextes  $IC$  et  $DC$  dans l'arbre des contextes  $T$
2. Pour chaque noeud  $(n_k \in T)$  tel que  $support(n_k) > minsup$

- (a) Pour toute paire de valeurs  $(v_1, v_2)$  instanciant le traitement  $PT(DC)$ 
  - i. Pour toutes les strates  $ST_j$  de  $ST(IC)$  instanciant le contexte d'identité  $IC$  et telle que  $support(ST_j) > minsup$  :
    - Si la borne minimale  $seb - kg(R)$  de l'intervalle de confiance à  $x\% > 1$  alors ajout de  $R$  à  $CR$

La dernière étape de l'algorithme consiste à fusionner récursivement les règles afin d'obtenir des règles plus générales.

## 4 Evaluation expérimentale

Le jeu de données utilisé provient de l'unique approche combinant graphe de connaissances et causalité : Munch et al. (2019). Le schéma de cet extrait de DB-Pedia est présenté en figure 1. Il comporte 6908 triplets. La classe cible est Auteur (Writer) et le résultat étudié est l'âge auquel un auteur publie son premier livre, défini par l'expression :  $dbp :auteur / datesortie - datenaissance$ . Les propriétés  $dbp :numberOfPages$  et  $dbp :genre$  sont considérées comme non pertinentes pour cette analyse, et la discrétisation a été réalisée en suivant les expériences de Munch et al. (2019). Les tests statistiques ont été réalisés en construisant des intervalles de confiance à 90%.

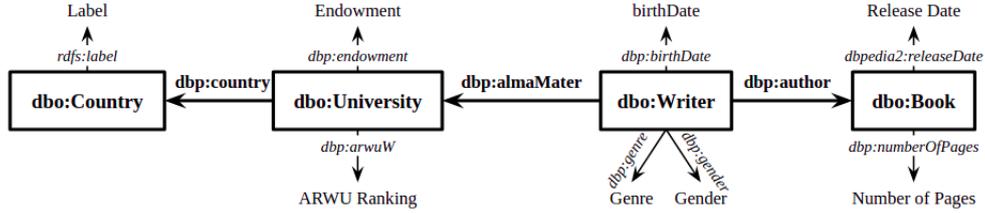


FIG. 1: Schéma du graphe de connaissances utilisé.

L'approche a permis de déterminer 14 règles avant fusion. Certaines de ces règles sont présentées dans le tableau TAB. 1<sup>3</sup>. Certaines règles sont plus faciles à appréhender que d'autres. Ainsi, il semble raisonnable qu'étudier dans les meilleures universités favorise une publication plus tôt du premier livre. Par contre, il est plus compliqué d'expliquer en quoi être né après 1960 permette de publier plus tôt que les personnes nées avant 1934, sauf si l'on considère que l'accès et le processus d'édition est d'autant plus facile que les années sont récentes.

TAB. 1: Exemples de règles différentielles causales obtenues sur un extrait de DBPedia.

Strate	Traitement	Tête
$Uni.(endowment < 2.10^9) \wedge Author(birthDate) \in [1934,1959] \wedge (Country=U.S.)$	$arwu(x, <100) \wedge arwu(y, >100)$	$agePublish(x) < agePublish(y)$
$Uni.(endowment < 2.10^9) \wedge Author(birthDate) \in [1960,2000] \wedge (Country=U.S.)$	$arwu(x, <100) \wedge arwu(y, >100)$	
$Uni.(endowment < 2.10^9, arwu >100) \wedge (Country=U.S.)$	$birthDate(x, >1960) \wedge birthDate(y, <1934)$	
$Uni.(endowment < 2.10^9, arwu >100) \wedge (Country=U.S.)$	$birthDate(x, >1960) \wedge birthDate(y, [1934,1960])$	
$Uni.(endowment > 2.10^9, arwu < 100) \wedge Author(birthDate < 1934)$	$country(x, England) \wedge country(x, U.S.)$	

3. La représentation des propriétés a été simplifiée afin de faciliter la lecture des règles pour les lecteurs.

L'étape de fusion a permis d'obtenir un ensemble de 11 règles qui sont identiques ont qui généralisent les 14 règles les plus spécifiques (6 règles ont pu être fusionnées). Dans cette étape, nous avons considéré les strates non instanciées comme invalides.

L'approche a été testée en faisant varier la taille de l'échantillon et ce pour une série d'exécutions. Les règles obtenues sur l'ensemble du jeu de données sont considérées comme références, et les règles obtenues après échantillonnage sont comparées à ces références. Plus l'échantillon est grand, plus le nombre de règles obtenues est important. Cette évolution n'est pas linéaire, le nombre de règles augmente rapidement puis atteint un plateau. Néanmoins, l'échantillonnage semble permettre la découverte de règles causales différentielles qui restent valides sur tout le jeu de données. Par exemple, en moyenne on découvre 10,67 règles en parcourant la moitié des paires et 80% d'entre elles sont valides. Les règles qui ont été obtenues permettent d'expliquer les différences de 4092 des 13016 ayant des résultats différents, soit environ une sur trois.

<b>Param. échant. (%)</b>	12.5	25	37.5	50
<b># Règles</b>	3.33 ± 0.48	8.33 ± 0.48	10 ± 2.16	10.67 ± 1.25
<b>Précision (%)</b>	80.56 ± 14.16	75.93 ± 10.29	83.30 ± 4.08	81.23 ± 6.94
<b>Param. échant. (%)</b>	62.5	75	87.5	100
<b># Règles</b>	12 ± 1.41	12.67 ± 0.94	12.67 ± 1.25	14
<b>Précision (%)</b>	81.39 ± 7.96	70.64 ± 5.61	89.47 ± 3.52	100

TAB. 2: Nombre et précision des règles en fonction du paramètre

Notre approche permet donc de découvrir des règles causales, et ce même en étant très sélective (intervalle de confiance à 90%, heuristique pessimiste pour la fusion). Contrairement à Munch et al. (2019) dont le but est de prédire le résultat (discrétisé auparavant), notre approche vise à expliquer des différences. Les résultats montrent que nos règles sont plus expressives que les règles probabilistes induites par le réseau. Leur approche découvre que la date de publication peut dépendre de la date de naissance et de l'université de l'auteur. Notre approche confirme ces résultats, mais montre que d'autres éléments de description tels que le pays et la richesse de l'université peuvent expliquer des différences. Notre approche n'impose pas de discrétiser les propriétés ayant des valeurs numériques (traitement et résultat), ce qui permet à la fois de découvrir des règles où la contrainte est une simple comparaison numérique ou de s'intéresser à l'impact de la propriété selon des tranches de valeurs.

## 5 Conclusion

Nous avons proposé une approche originale pour déterminer des règles différentielles causales dans des graphes de connaissances. Les expériences réalisées sur un extrait de DBpedia ont permis de montrer l'efficacité de l'approche pour déterminer de telles règles, permettant d'expliquer certaines variations dans les ages de première publication. Dans nos futurs travaux, nous souhaitons intégrer des contraintes sémantiques afin d'optimiser la fouille de règles et leur évaluation. Nous souhaitons également tester cette approche sur des jeux de données plus volumineux.

## Références

- Galárraga, L., C. Teflioudi, K. Hose, et F. Suchanek (2013). Amie : Association rule mining under incomplete evidence in ontological knowledge bases. pp. 413–422.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* 81(396), 945–960.
- Iacus, S. M., G. King, et G. Porro (2012). Causal inference without balance checking : Coarsened exact matching. *Political Analysis* 20(1), 1–24.
- King, G. et R. Nielsen (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis* 27(4), 435–454.
- Li, J., T. D. Le, L. Liu, J. Liu, Z. Jin, et B. Sun (2013). Mining causal association rules. *Proceedings - IEEE 13th International Conference on Data Mining Workshops, ICDMW 2013*, 114–123.
- Li, J., S. Ma, T. Le, L. Liu, et J. Liu (2017). Causal Decision Trees. *IEEE Transactions on Knowledge and Data Engineering* 29, 257–271.
- Munch, M., J. Dibie, P. H. Willemin, et C. Manfredotti (2019). Interactive causal discovery in knowledge graphs. *CEUR Workshop Proceedings 2465*, 78–93.
- Neapolitan, R. E. (2003). Learning Bayesian networks with integration of indirect prior knowledge. *Prentice Hall*.
- Pearl, J. (2009). *Causality*. Cambridge University Press.
- Raad, J., N. Pernelle, et F. Saïs (2017). Detection of contextual identity links in a knowledge base. In *Proceedings of the Knowledge Capture Conference, K-CAP 2017*.
- Rosenbaum, P. R. et D. B. Rubin (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79(387), 516–524.
- Rubin D. B (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology* 66(5), 688–701.
- Sebag, M. et M. Shoemaker (1988). Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. *Proc. of the European Knowledge Acquisition Workshop Conference*, 28–1–28–20.
- Stuart, E. A. (2010). 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.

## Summary

In this paper, we present an approach that discovers differential causal rules in Knowledge Graphs. Such rules express that for two class instances, a different treatment leads to different outcomes. The proposed approach is based on a semantic matching and strata that can be defined as complex sub-classes. A first experimental evaluation on a DBpedia extract showed that such discovered rules can help to explain a significant number of variability in outcomes.