

# Alignement des graphes de connaissances d'activités personnelles pour la construction d'un graphe d'activités inter-personnelles

A. Choumélé Nandong Lucry Maël<sup>1</sup>, B. Pepuere Pempeme Théophile<sup>1</sup>, C. Tchotaneu Ngatcha Giresse<sup>1</sup>,  
D. Sonfack Souchio Serge<sup>2</sup>, E. Mbiaya Franck Anaël<sup>3</sup>, F. Ho Tuong Vinh<sup>1</sup>

<sup>1</sup> Université Nationale du Vietnam Hanoï -Institut Francophone International, Vietnam,

<sup>2</sup> Robert Bosch Sp. z o. o. Jutrzenki 105, POLAND,

<sup>3</sup> Univ. Orleans, INSA Centre Val de Loire, LIFO EA, 4022, France

## Résumé

*Le graphe de connaissance des activités interpersonnelles (Interpersonal Activity knowledge Graphs (IAG)) est une structure sous forme de graphe qui, au sein d'une organisation, intègre de multiples connaissances d'activités provenant de différents individus. Ce graphe repose sur l'ontologie des activités personnelles (Personal Activity Ontology (PAO)) qui offre une sémantique et un raisonnement communs pour l'acquisition et l'exploitation des activités au sein d'une organisation. Toutefois, pour parvenir à cette structure de graphe de connaissance des activités qui permettra d'exploiter pleinement la capacité du savoir-faire personnel au sein d'une organisation, il est primordial d'aligner ces graphes de connaissances sur un seul et grand graphe à partir duquel il sera possible d'obtenir davantage de connaissances.*

*Cette étude traite de l'alignement des graphes de connaissances sur les activités personnelles basées sur une approche combinant la similarité sémantique et de surface, le plongement de graphe RDF (Resource Description Framework) et une interaction humaine. L'approche proposée donne des résultats satisfaisants dans l'alignement des graphes de connaissances d'activités à partir du cas d'utilisation des activités d'enseignement dans une école. Le graphe de connaissances d'activités résultant est plus important et les requêtes permettent d'obtenir des réponses plus pertinentes.*

## Mots-clés

*Graphe d'activités interpersonnelles, ontologie, graphe de connaissances.*

## Abstract

*The interpersonal activity knowledge graph (IAG) is a single structure within an organization that embeds multiple activity knowledge from different individuals. This graph relies on the Personal Activity Ontology (PAO) that offers a common semantic and reasoning for an individual's activity acquisition and exploitation within an organization. However, to arrive at this single graph structure of the whole organization's activity knowledge that will allow to fully exploit the capacity of this personal know-how within the organization, it is paramount to align all knowledge*

*about activities to a single and large graph from which more insight can be retrieved.*

*This study elaborates on aligning personal activity knowledge graphs based on an approach combining semantic and surface similarity, resource description framework (RDF) embedding, and human interaction. The proposed approach yields satisfactory results in aligning activity knowledge from the use case of teaching activities in a school. A significant activity knowledge graph is created, and more reasoning and queries can be carried out.*

## Keywords

*Interpersonal activity graph, Ontology, Knowledge graph.*

## 1 Introduction

De nos jours, le capital intellectuel et la gestion des connaissances constituent les piliers fondamentaux de l'avantage compétitif des organisations. De ce fait, l'harmonisation, la formalisation et la valorisation ou la réutilisation des savoirs individuels émergent comme des enjeux cruciaux pour le développement des organisations.

Les graphes de connaissances d'activités personnelles, par leur capacité de représenter et de raisonner sur le savoir-faire posséder par les individus dans une organisation, s'avèrent être des outils précieux dans ce contexte. L'alignement des graphes d'activités au sein d'un seul graphe d'activités interpersonnelles (IAG :Interpersonal Activity Graph) permettra aux organisation d'accéder à une grande quantité de savoir-faire et de pouvoir y inférer de nouvelles connaissances.

Le graphe d'activités interpersonnelles est une représentation de la connaissance sous forme de graphe qui permet de conserver du savoir-faire des personnes en ce qui concerne leurs activités dans une organisation. Conçu à partir de la théorie de l'activité utilisée dans les sciences sociales, l'IAG utilise les principes et les concepts de cette théorie ainsi que la technologie du web sémantique en général et une ontologie en particulier pour définir des axiomes et des relations qui capturent le savoir-faire des activités d'une personne [7, 8].

Le problème abordé dans cette étude est celui de l'alignement des graphes d'activités, qui ont la particularité d'avoir

des nœuds centraux qui sont les activités. Cette structure particulière des graphes personnelle d'activités les rendent singulières et difficiles à aligner avec les méthodes existantes.

Nous proposons dans cette étude une approche combinant la similarité sémantique, le plongement de graphes RDF (Resource Description Framework) et un retour humain qui permet d'améliorer les étapes précédentes.

La suite de ce travail se présente comme suite : Section 2 qui élabore sur les méthodes et techniques existantes, la section 3 décrit l'approche proposée et la section 4 et section 5 pour les résultats et la conclusion respectivement.

## 2 Etat de l'art

Dans cette section, nous explorons les avancées récentes des chercheurs dans le domaine d'alignement de graphes de connaissance personnel. Les travaux se concentrent sur l'amélioration de l'efficacité des méthodes d'alignement, en exploitant des techniques avancées, pour faciliter l'intégration et l'interprétation des vastes ensembles de données sémantiques.

Les travaux sélectionnés, couvrant la période de 2019 à 2023, présentent Les récentes avancées illustrant une tendance vers l'automatisation et la précision accrue, soulignant l'importance de ces recherches pour le développement des technologies de l'information et de la communication.

En 2019, [14] ont proposé MultiKE, un cadre pour l'alignement d'entités basé sur l'embedding multi-vue des graphes de connaissances. Leur approche, unique, intègre des vues multiples (nom, relation, attribut) pour enrichir les embeddings d'entités et faciliter leur alignement entre différents KGs. Les méthodes d'inférence cross-KG et les stratégies de combinaison des vues contribuent à une amélioration notable de la performance. Leur expérimentation sur des ensembles de données réels montre que MultiKE surpasse les méthodes existantes, démontrant l'efficacité de l'utilisation des caractéristiques multi-vues et de l'inférence cross-KG pour l'alignement d'entités.

En 2019, [12] ont proposé un modèle innovant pour l'alignement d'entités entre graphes de connaissances en utilisant des embeddings d'attributs. Publié dans le cadre de la conférence AAAI-19, leur travail se concentre sur l'amélioration de l'alignement d'entités à travers différents graphes de connaissances en exploitant les embeddings de caractères d'attributs pour surmonter les limitations des modèles basés uniquement sur les embeddings de graphes. Les expériences réalisées avec de vraies bases de connaissances ont montré une amélioration significative de plus de 50% en termes de précision d'alignement des entités par rapport aux modèles de base, démontrant l'efficacité de leur approche.

En 2020, [11] ont introduit AliNet, un réseau d'alignement de graphes de connaissances qui utilise l'agrégation de

voisinage multi-sauts avec mécanisme de porte. Publiée lors de la conférence AAAI-20, cette méthode aborde le problème de la non-isomorphie des structures de voisinage dans l'alignement des entités de graphes de connaissances. AliNet emploie un mécanisme d'attention pour sélectionner des voisins distants utiles et un mécanisme de porte pour contrôler l'agrégation des informations de voisinage, direct et distant. Des expériences approfondies sur cinq jeux de données d'alignement d'entités démontrent l'efficacité d'AliNet, illustrant une avancée significative dans le domaine.

En 2021, [5] ont introduit PRASE, un cadre novateur pour l'alignement non supervisé de graphes de connaissances basé sur le raisonnement probabiliste et l'embedding sémantique. Cette approche, combinant les avantages du système de raisonnement probabiliste PARIS avec des techniques d'embedding sémantique avancées, a prouvé sa supériorité en atteignant des performances de pointe sur plusieurs jeux de données. PRASE illustre l'efficacité d'intégrer le raisonnement et l'embedding pour l'alignement des graphes de connaissances, marquant une avancée significative dans le domaine.

En 2023, [3] ont publié une revue expérimentale sur les méthodes d'embedding pour l'alignement d'entités dans les graphes de connaissances dans la revue "Data Mining and Knowledge Discovery". Leur recherche offre une analyse méta-niveau des différentes approches d'embedding, établissant des corrélations statistiquement significatives entre les méthodes et diverses caractéristiques des graphes de connaissances. Cette étude propose un classement des méthodes basé sur leur efficacité, ouvrant de nouvelles perspectives sur l'équilibre entre efficacité et efficience dans les techniques d'alignement d'entités.

En 2023, [15] ont présenté AutoAlign, une méthode d'alignement de graphes de connaissances entièrement automatique exploitant les modèles de langage pour l'embedding des prédicats et TransE pour les embeddings d'entités. Cette approche innovante, qui n'exige aucun alignement initial manuel, a démontré une amélioration significative dans l'alignement des entités sur des KGs réels, surpassant les méthodes existantes.

En 2023, [2] ont développé SemanticGraph2Vec, un modèle novateur pour l'embedding de graphes sémantiques visant à améliorer la représentation textuelle. Cette méthode, qui privilégie des marches sémantiques au lieu de marches aléatoires, permet de capturer les relations sémantiques importantes entre les sommets d'un graphe. Appliqué à la tâche de l'étiquetage des parties du discours, SemanticGraph2Vec a surpassé deux méthodes de pointe en termes de précision et de score F1, illustrant l'efficacité de considérer les relations sémantiques dans l'embedding des graphes pour les tâches de traitement du langage naturel.

On peut remarquer des études précédentes que les ap-

proches sont basées sur des graphes ayant une structure simple, c'est-à-dire tous les noeuds ont le même niveau d'importance. Ce qui n'est pas le cas des graphes interpersonnels d'activités.

Selon nos connaissances, ce travail représente la première initiative visant à explorer l'alignement des graphes interpersonnels d'activités. Cette exploration pionnière ouvre de nouvelles perspectives pour comprendre et exploiter la complexité inhérente aux graphes d'activité.

### 3 Approche proposée

L'approche proposée pour aligner les représentations des connaissances sur les activités isolées est décrite dans la figure 5 : (1) Sélection de sous-graphe, (2) la similarité sémantique, (3) l'intégration du graphe, (4) la similarité de surface et (5) le rétroaction humaine.

#### 3.1 Préliminaire

Un graphe d'activité est un graphe  $G$  de type RDF où les noeuds instances du concept activité sont des noeuds centraux et ont une considération particulière [9]

soit  $\langle A_1, r_1, t \rangle \in G_1$   
et soit  $\langle A_2, r_2, t \rangle \in G_2$ , où  
 $A_1, A_2$  sont des sujets,  
 $r_1, r_2$  des prédicats et  
 $t$  le prédicat.

L'alignement de  $G_1$  et  $G_2$  consiste à construire le graphe  $G$  tel que  $\langle A_1, r', t \rangle, \langle A_2, r', t \rangle$  avec  $r'$  soit égale à  $r_1$  ou  $r_2$  si  $r_1$  et  $r_2$  son identique ou soit une super relation contenant  $r_1$  et  $r_2$ .

#### 3.2 Sélection de sous-graphes

La sélection de sous-graphes est une étape cruciale dans l'alignement des graphes. Elle permet dans un premier temps de conserver les nœuds centraux qui sont les activités et par la suite d'extraire des parties spécifiques du graphe qui sont directement liées à ces activités. Cette phase implique l'identification des activités, ainsi que des entités associées telles que les ressources, les moments de début et de fin, et les lieux. De plus, elle se concentre sur l'extraction des relations entre les activités et les entités. Cette approche vise à isoler les segments du graphe pour l'analyse ultérieure, permettant une comparaison plus précise et une meilleure compréhension des similitudes et des différences entre les sous-graphes ciblés.

#### 3.3 Similitude sémantique

La phase de similitude sémantique dans le processus d'alignement des graphes joue un rôle crucial en permettant une comparaison approfondie et significative entre les entités issues de différents sous-graphes. Cette étape implique l'analyse de chaque paire d'entités extraites des sous-graphes d'activités pour évaluer leur proximité conceptuelle.

La méthode ConSim de la Formule 1 (exemple de mesure en le noeud F et J Figure 1)

$$ConSim(c^*, c) = \frac{2 * N_R}{N^* + N + 2 * N_R} \quad (1)$$

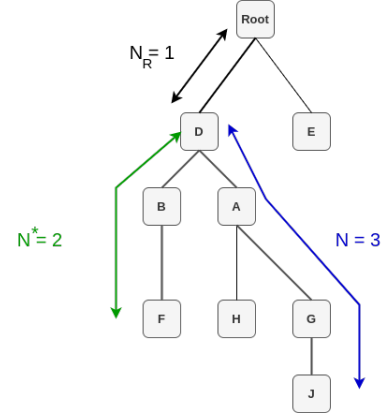


FIGURE 1 – Exemple d'application de similarité sémantique entre le noeud F et J

utilisée pour calculer la similarité entre les concepts associés à ces entités est basé sur la mesure de similarité proposée par [4, 13]. En récupérant et en comparant les concepts sous-jacents, on peut déterminer dans quelle mesure deux entités partagent des significations ou des fonctions similaires, même si elles sont nommées différemment ou si elles apparaissent dans des contextes différents.

#### 3.4 IAG embedding

L'utilisation de RDF2Vec proposé par [1], pour l'embedding des entités dans l'étape d'IAG Embedding constitue une méthode efficace pour transformer les informations structurées RDF en vecteurs denses, capturant ainsi la structure et les nuances sémantiques des entités. Cette transformation rend les données RDF exploitables pour des analyses numériques, en particulier pour la comparaison et l'analyse de similarité entre les entités des différents sous-graphes. L'ajout de la mesure de similarité cosinus (simcosine) [6] dans ce processus permet une évaluation précise de la proximité entre les vecteurs d'embedding, facilitant ainsi l'identification des correspondances potentielles entre les entités. En combinant RDF2Vec avec simcosine, l'étape d'IAG Embedding devient cruciale pour le processus d'alignement, offrant une base solide pour des comparaisons numériques précises et la détection de similarités sémantiques entre les entités des graphes.

#### 3.5 Similitude de surface

L'étape de similitude de surface agit comme un complément essentiel à l'analyse sémantique dans le processus d'alignement des graphes. En se concentrant sur les attributs textuels des entités, tels que leurs noms, et en appliquant l'indice de Jaccard [10] pour la comparaison, cette méthode permet d'évaluer la similarité littérale ou de surface entre les entités. Cela est particulièrement utile pour détecter les correspondances directes dans les noms des entités, offrant une couche supplémentaire de validation pour

les alignements proposés. En identifiant ces similitudes de surface, on peut renforcer la confiance dans les alignements sémantiques en confirmant que les entités non seulement partagent un contexte sémantique similaire mais sont également alignées au niveau de leur représentation textuelle.

### 3.6 Rétroaction humaine

Cette phase finale du processus d'alignement, marquée par l'intégration du feedback humain, est cruciale pour affiner et valider les résultats obtenus par les étapes automatisées. En invitant l'utilisateur à évaluer la qualité de l'alignement, par l'acceptation ou le rejet des liens proposés, cette approche met en lumière l'importance de la perspective humaine dans le raffinement des technologies d'alignement. Le retour humain, en étant intégré dans le processus, permet une révision ciblée des alignements automatiques, offrant ainsi une opportunité d'amélioration continue des résultats. Cette interaction entre l'homme et la machine assure que le graphe de connaissances final ne reflète pas seulement une précision algorithmique, mais aussi une pertinence et une utilité pratique pour les utilisateurs finaux. Cette étape souligne l'importance de la synergie entre l'expertise humaine et les capacités de traitement automatisé, en reconnaissant que, malgré les avancées technologiques, la contribution humaine reste indispensable pour assurer la qualité et la fiabilité des systèmes d'alignement de connaissances.

### 3.7 Cas d'étude

Pour illustrer le modèle ci-dessus du graphe de connaissances des activités interpersonnelles, le cas des activités d'enseignement au sein d'une institution est abordé.

Pour ce cas d'utilisation, nous avons des activités d'enseignement pour les cours de génie logiciel, de gestion de projet, d'interface utilisateur et de base de données.

Les personnes au sein de l'institution ont différents rôles d'enseignement, y compris l'enseignement magistral, le tutorat ou la supervision. Les bénéficiaires de ces activités sont les étudiants des niveaux 1 et 2 de l'institution.

Chaque activité a un objet défini et des ressources (livres physiques et numériques, ordinateurs, vidéos) nécessaires à sa réalisation. Les activités se déroulent soit dans les salles de classe, soit dans les salles informatiques, à un moment précis.

La Figure 2 montre le graphe issu de notre cas d'étude.

## 4 Résultat

Cette section présente les résultats obtenus à travers une analyse comparative visuelle, illustrant l'impact de l'alignement sur les entités étudiées. Les Figures 3 et 4 mettent en exergue, respectivement, un exemple d'entité non alignée et un exemple d'entité alignée, démontrant les différences significatives résultant de notre méthode d'alignement. L'analyse détaillée présentée dans le Tableau 1 vise à éclairer les aspects cruciaux de notre étude, offrant une compréhension approfondie des bénéfices et des limitations associés à l'alignement des entités dans notre contexte d'application. Il résume de manière claire et concise les résultats obtenus en fonction de différents seuils d'alignement, per-

mettant aux lecteurs de comprendre rapidement l'impact du seuil sur la qualité de l'alignement. L'interprétation aide à contextualiser les résultats et fournit des compréhensions sur la manière dont les différents seuils affectent l'alignement des nœuds.

## 5 Conclusion

Cette étude a exploré l'alignement des graphes de connaissances interpersonnels d'activités, en adoptant une approche hybride prenant en compte l'humain dans la boucle à travers l'utilisation d'une similarité sémantique, un plongement de graphes RDF (Resource Description Framework) et une similarité de surface.

Les résultats montrent que les dimensions des plongements RDF2Vec ont un impact significatif sur la précision des analyses structurales. Parallèlement, la méthode similarité sémantique a démontré son efficacité dans l'évaluation de la similarité conceptuelle, offrant une compréhension nuancée des connexions sémantiques entre les concepts.

La mesure de Jaccard a également enrichi nos analyses en fournissant une perspective supplémentaire basée sur la co-occurrence des attributs.

En perspective, ce travail ouvre de nouvelles directions de recherche. Outre l'implémentation d'un système complet d'acquisition du savoir-faire au sein d'une organisation, il sera intéressant d'extraire des connaissances plus raffinées liées au processus, à la cartographie de connaissances ou la conception des systèmes d'aide aux activités (Activity Aid/Assistant System AAS) ou Activity Knowledge As a Service.

En conclusion, bien que le projet ait mis en lumière des aspects importants de l'analyse des graphes de connaissances, il a également révélé des opportunités substantielles pour des recherches et des développements futurs dans ce domaine fascinant.

## Références

- [1] Mona Alshahrani, Mohammad Asif Khan, Omar Maddouri, Akira R Kinjo, Núria Queralt-Rosinach, and Robert Hoehndorf. Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics*, 33(17):2723–2730, 2017.
- [2] Wael Etaiwi and Arafat Awajan. Semanticgraph2vec : Semantic graph embedding for text representation. *Array*, 17:100276, 2023.
- [3] Nikolaos Fanourakis, Vasilis Efthymiou, Dimitris Kotzinos, and Vassilis Christophides. Knowledge graph embedding methods for entity alignment : experimental review. *Data Mining and Knowledge Discovery*, 37(5):2070–2137, 2023.
- [4] Fabien Gandon, Olivier Corby, Rose Dieng-Kuntz, and Alain Giboin. Proximité conceptuelle et distances de graphes. In *Atelier Raisonner le Web Sémantique avec des Graphes, Plateforme AFIA 2005*, 2005.

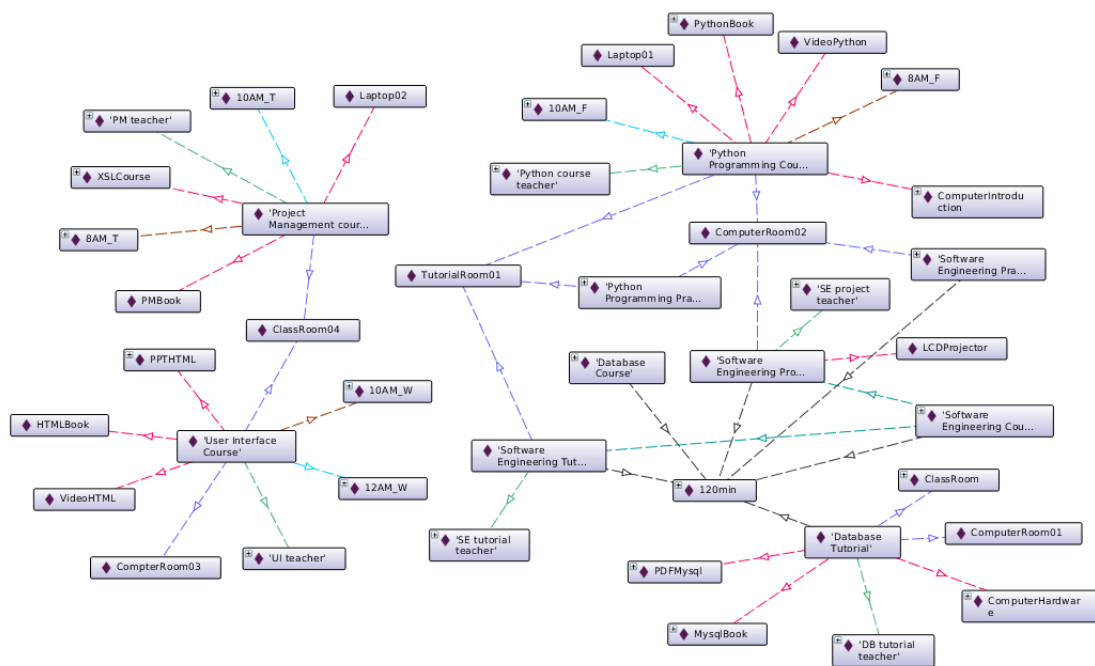


FIGURE 2 – Graphe de connaissances des activités dans le domaine de l’enseignement

Entrée	Seuil	Résultat	Interprétation
Entrez le nom du nœud : PythonCourse	0,1	19 paires de Nœud	Alignant même des nœuds qui n’ont pas de correspondances significatives
Entrez le nom du nœud : PythonPractical	0,3	6 paires de Nœud	Alignant même des nœuds qui n’ont pas de correspondances significative
Entrez la relation usesResource	0,5	4 paires de Nœuds	Nœuds Sensiblement Bien Aligné
	0,9	0 paire de Nœuds	Conduit à une sélection très stricte,excluant même des paires qui pourraient être des correspondances potentielles.

TABLE 1 – Tableau des résultats

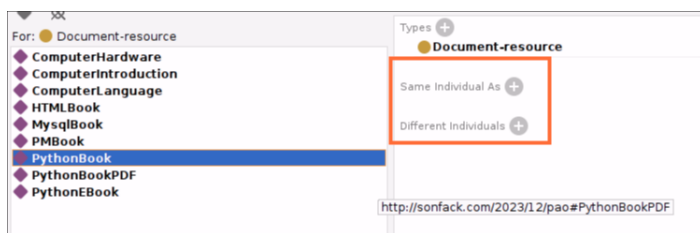


FIGURE 3 – Exemple d’entité non alignée

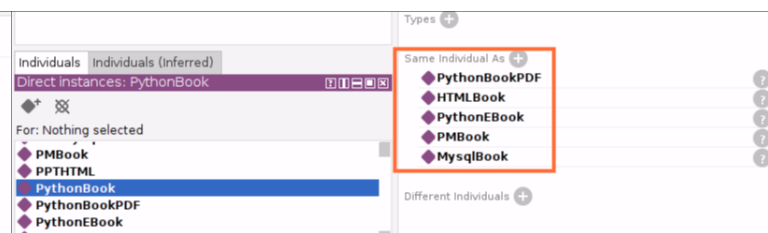


FIGURE 4 – Exemple d’entité alignée

- [5] Zhiyuan Qi, Ziheng Zhang, Jiaoyan Chen, Xi Chen, Yuejia Xiang, Ningyu Zhang, and Yefeng Zheng. Un-supervised knowledge graph alignment by probabilistic reasoning and semantic embedding. *arXiv preprint arXiv :2105.05596*, 2021.
- [6] Faisal Rahutomo, Teruaki Kitasuka, and Masayoshi Arisugi. Semantic cosine similarity. In *The 7th international student conference on advanced science and technology ICAST*, volume 4, page 1, 2012.
- [7] Serge Sonfack Souchio, Thierry Coudert, Bernard

Kamsu Foguem, Laurent Geneste, Cédric Beler, and Sina Namakiaraghi. Organizations’ interpersonal activity knowledge representation. In *HHAI 2023 : Augmenting Human Intellect*, pages 254–262. IOS Press, 2023.

- [8] Serge Sonfack Souchio, Laurent Geneste, Bernard Kamsu-Foguem, Cédric Béler, Sina Namaki Araghi, and Muhammad Raza Naqvi. An enterprise architecture for interpersonal activity knowledge management. In *Iberoamerican Knowledge Graphs and Se-*

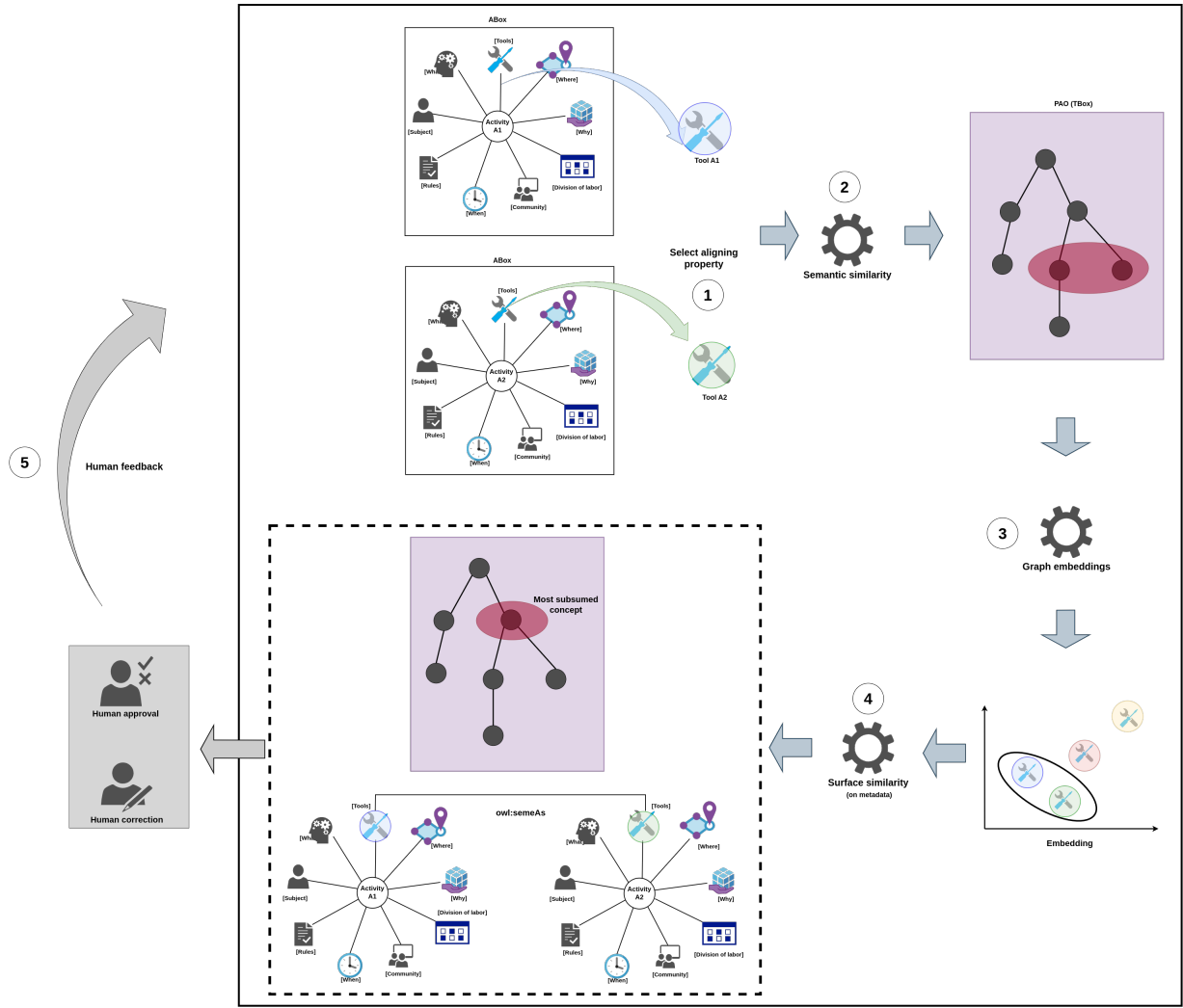


FIGURE 5 – Résumé des différentes étapes de l'alignement des graphes d'activités

- mantic Web Conference*, pages 66–81. Springer, 2023.
- [9] Serge SONFACK SOUNCHIO, Trawina HAL-GUIETA, Baudelaire Ismael TANKEU NGUEKEU, Laurent GENESTE, and Bernard KAMSU-FOGUEM. Interpersonal activity knowledge graph for organizations. 2023.
- [10] Shuang-Bo Sun, Zhi-Heng Zhang, Xin-Ling Dong, Heng-Ru Zhang, Tong-Jun Li, Lin Zhang, and Fan Min. Integrating triangle and jaccard similarities for recommendation. *PloS one*, 12(8) :e0183570, 2017.
- [11] Zequn Sun, Chengming Wang, Wei Hu, Muhao Chen, Jian Dai, Wei Zhang, and Yuzhong Qu. Knowledge graph alignment network with gated multi-hop neighborhood aggregation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 222–229, 2020.
- [12] Bayu Distiawan Trisedya, Jianzhong Qi, and Rui Zhang. Entity alignment between knowledge graphs using attribute embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 297–304, 2019.
- [13] Zhibiao Wu and Martha Palmer. Verb semantics and lexical selection. *arXiv preprint cmp-lg/9406033*, 1994.
- [14] Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. Multi-view knowledge graph embedding for entity alignment. *arXiv preprint arXiv :1906.02390*, 2019.
- [15] Rui Zhang, Yixin Su, Bayu Distiawan Trisedya, Xiaoyan Zhao, Min Yang, Hong Cheng, and Jianzhong Qi. Autoalign : Fully automatic and effective knowledge graph alignment enabled by large language models. *IEEE Transactions on Knowledge and Data Engineering*, 2023.