

Compléter une revue de littérature en utilisant les réseaux de citation

Présentation de CitationGecko et de CoCites

Damien Belvèze

1er juin 2021

Introduction

L'essor des citations ouvertes depuis 2017 ([OpenCitations]) auquel même des grands éditeurs, par ailleurs producteurs de bases citationnelles, comme Elsevier ont contribué) ont rendu possible l'apparition de nouveaux outils permettant de compléter une [revue de littérature] fondée au préalable sur la **recherche d'articles par mots-clé** par une recherche sur les réseaux de citation constituées à partir de ces citations ouvertes. C'est le cas de deux produits qu'on va présenter dans cet atelier : [CitationGecko] et [Cocites](<https://www.cocites.com/>)

Citation Gecko : recherche d'articles en **relation bibliographique** (Si nous disposons de A et B, nous allons pouvoir découvrir C, car A et B sont tous les deux cités par C ou bien, A et B citent tous les deux C) [[Compléter une revue de littérature#Cocites trouver des articles co-cités|Cocites]] = recherche des papiers cités en même temps que le papier de départ (C fréquemment cité en même temps que A et B)

Citation Gecko : trouver les papiers cités par ou citant des articles de référence

Au SCD, c'est grâce au travail d'Aaron Tay qui a parlé à plusieurs reprises de CitationGecko (en même temps que d'autres outils de ce type) que nous avons commencé à nous y intéresser. Les articles de son blog sur le sujet (et notamment le premier[(Tay 2018)]) sont la principale source des informations qui suivent.

Comment ça marche

L'algorithme de CitationGecko utilise les données bibliométriques de [[OpenCitations]], [[CrossRef]] et [[Microsoft Academic]] pour trouver un graphe de citation à partir des articles que nous lui fournissons (seed papers)

Usages possibles :

- Utiliser pour explorer les liens (s'il y en a) entre différents domaines de recherche que vous pourriez vouloir combiner
- Repérer des articles intéressants que vous auriez pu manquer en effectuant une recherche par mots clés uniquement

comment se servir de l'outil (La recherche d'information par citation)

on part d'articles jugés importants et on va agrandir le cercles des articles intéressants en trouvant : - quels articles postérieurs ont cité ces articles - Quels article antérieurs, ceux-ci ont cité

contrairement aux outils de recommandation d'articles, Citation Gecko vous permet de visualiser sous la forme de clusters les relations (citant/cités) entre articles.

Exemple :

Chercher 5 articles sur l'obfuscation et les navigateurs. Sélectionner ces articles comme seed papers Citation Gecko va afficher les articles citant ou cités par ces articles séminaux.

vue chronologique

En jaune les 5 articles ajoutés. Les ronds gris désignent des articles cités par ou citant nos seed papers. On en a la liste en cliquant sur le bouton en haut à gauche sur le rond jaune. Si on trouve que ces articles en dépit de leur relation de citation avec les articles séminaux ne sont pas pertinents, on peut les marquer comme non pertinents (mark irrelevant) On peut choisir de filtrer les articles par leur pertinence (en affichant ou faisant disparaître les articles marqués comme non pertinents)

Les ronds gris plus volumineux désignent des articles qui ont été cités par plusieurs articles séminaux, des références communes à ces articles.

Ainsi l'article intitulé *Investigation of obfuscation-based anti-reverse engineering for printed circuit boards* a été cité par deux de nos articles séminaux. En tant que tel, il peut constituer un article intéressant à découvrir.

Je peux l'ajouter à mon ensemble de seed papers, cela me permettra de faire apparaître encore davantage d'articles et de recoupements possibles. Dans cette vision chronologique des citations, il n'y a pas d'ambiguïté possible entre les papiers qui citent et ceux qui sont cités.

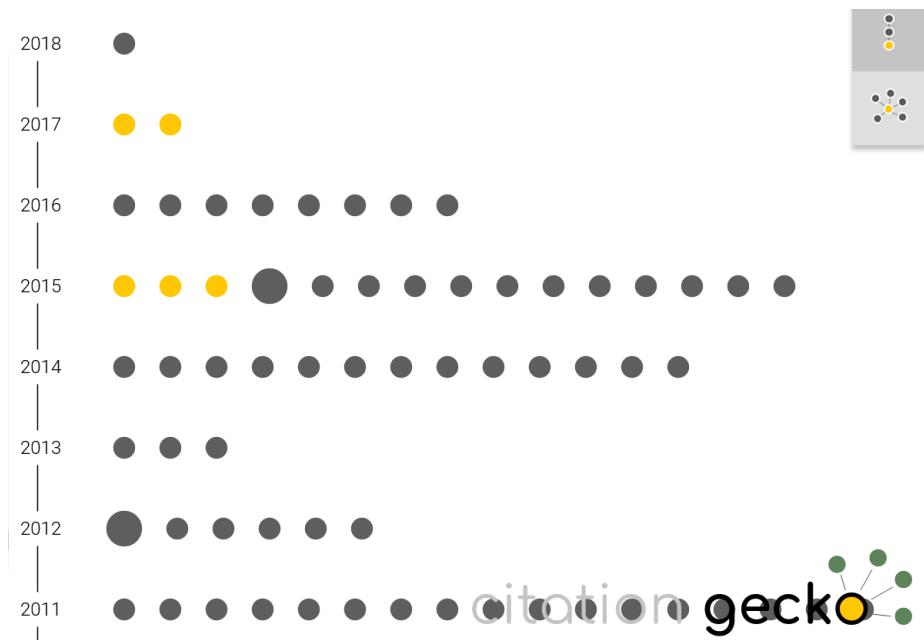


Figure 1: graphe historique 1

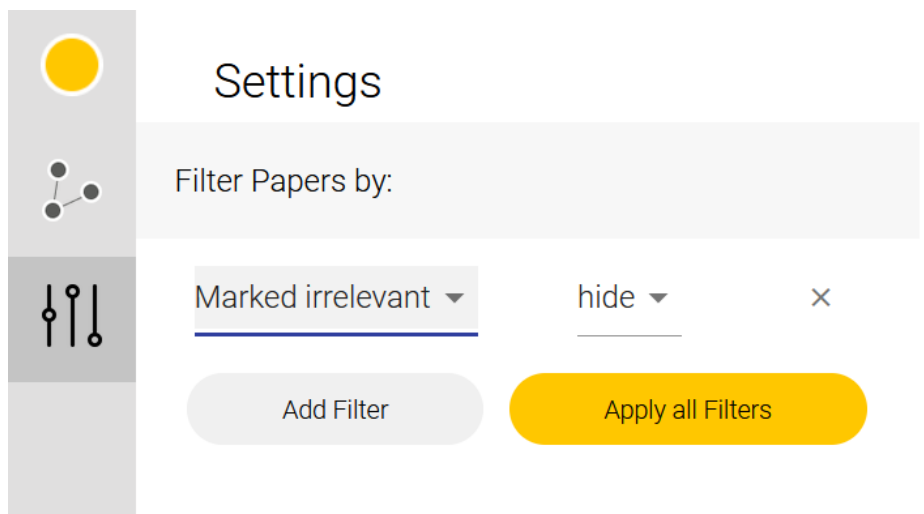


Figure 2: boutons et filtres

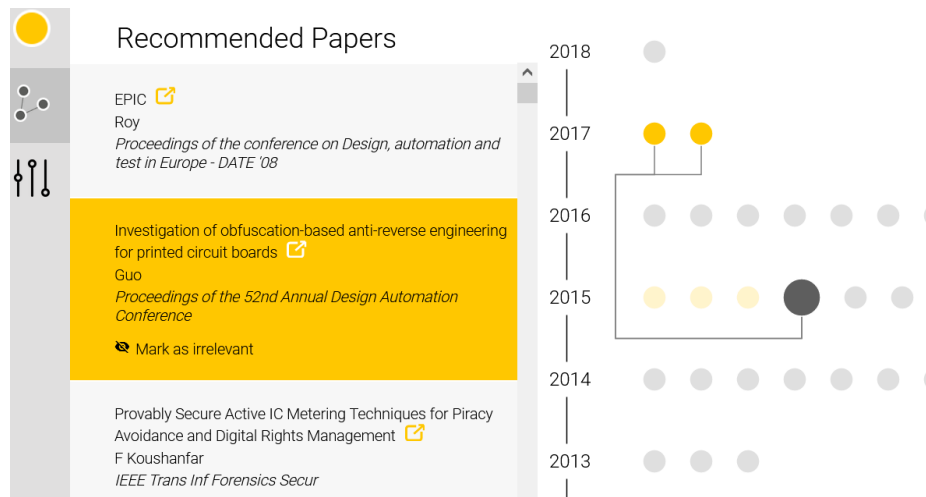


Figure 3: graphe historique 2

vue en réseau

Dans la vue en réseau, on affiche soit les articles qui sont cités par nos seed papers, soit les articles qui citent nos seed papers. Dans cette vue on affiche les articles cités par les seed papers :

Les articles qui ont le plus de poids sont ceux qui sont cités par plusieurs seed papers (au centre). On aurait pu imaginer que ce poids varie en fonction du nombre total de citations reçues (seed papers et autres), mais ce n'est pas le cas.

Et voici la vue qui affiche les articles qui citent nos seed papers :

Question : qu'est-ce qui pourrait s'avérer le plus intéressant pour nous ?

Des articles qui sont cités par un grand nombre de nos seed papers ? Des articles qui citent un grand nombre de nos seed papers ? (comme pour cet utilisateur[(Wölflé 2019)])

enregistrement

Lorsqu'on estime avoir une biblio complète, on peut sauvegarder l'ensemble des références en .bib et l'envoyer vers un éditeur de texte en Latex ou vers Zotero. (cliquer sur Save)

Différentes méthodes d'import

On peut importer ses seed papers sous la forme d'un fichier . bib téléchargé depuis Zotero ou directement depuis Zotero (Citation Gecko permet la connexion

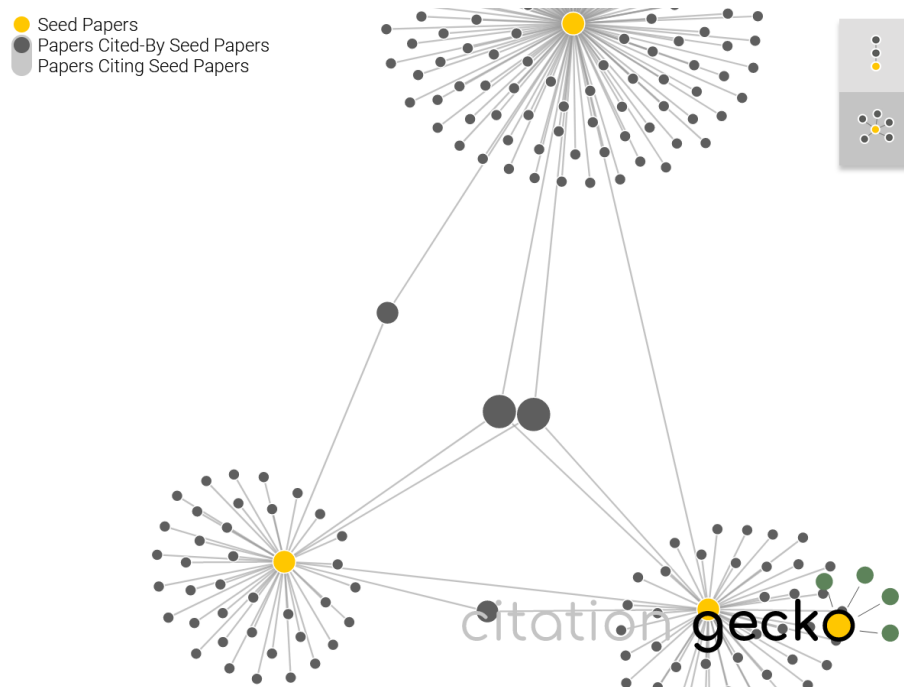


Figure 4: graphe cited by seed papers

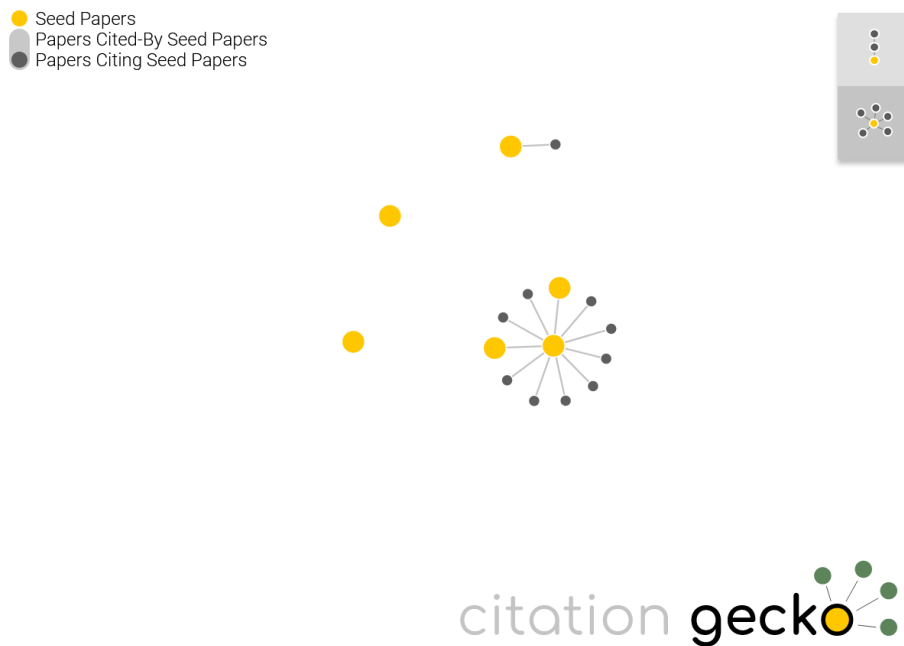


Figure 5: graphe citing seed papers

ion avec Zotero)

Pour importer depuis Zotero, il faut s'identifier à son compte Zotero et permettre la création d'une clé assurant la liaison entre Zotero et CitationGecko

Si un message d'erreur s'affiche, revenir en arrière à l'accueil du site CitationGecko. Le processus a normalement quand même eu lieu.

Choisir la collection de références à importer depuis Zotero. Seuls les items que CitationGecko pourra retrouver dans les bases de Crossref, Microsoft Academic ou OpenCitations s'afficheront

Cocites : trouver des articles co-cités

Dans ce cas, on part du principe que pour compléter une recherche d'articles basée sur des mots-clé, on ne va pas se tourner vers les articles qui citent ou sont cités par des articles considérés comme références mais on va chercher quels articles sont le plus souvent cités en même temps que ces articles de référence. C'est le travail que permet de faire l'extension Cocites

une fois chargée dans le navigateur, cette extension permet de visualiser dans Pubmed pour chaque article :

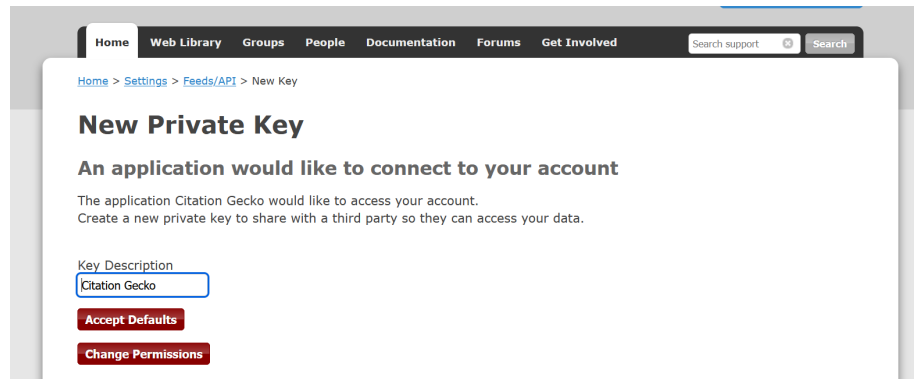


Figure 6: création clé

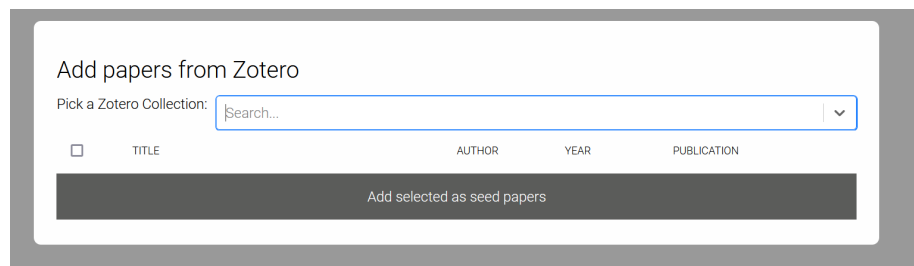


Figure 7: choix de la collection



Figure 8: fonctionnement de cocites 1

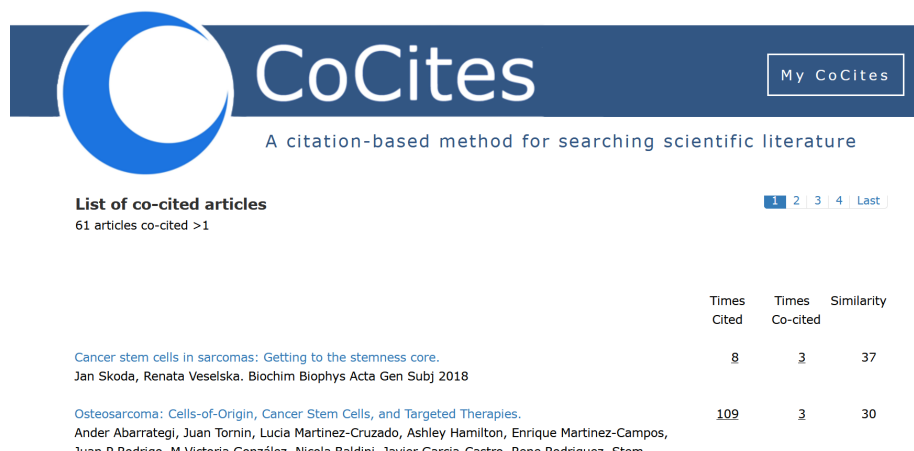


Figure 9: fonctionnement de cocites

ci-dessus le score de l'article Sarcoma Stell Cell heterogeneity

Cet article a été cité 10 fois Il a été cité en même temps que 61 autres articles Il a été trois fois cité en même temps que l'article Cancer stem cells in sarcomas: Getting to the stemness core. la similarité entre les deux articles est calculée comme suit

nombre de co-citations / le nombre de citations de celui des deux articles co-cités qui a eu le moins de citations.

Dans le cas présent : $3/8 = 37\%$. Ainsi si un article est peu cité mais souvent en même temps qu'un autre article, cela renforce la probabilité que cet article soit ressemblant avec l'article avec lequel il est régulièrement co-cité.

Possibilité de créer un compte pour exporter en .bib les références obtenues avec Cocites

conclusion

La recherche par co-citations ou bien la recherche par liens de citation (cité par / citant) sont des moyens prometteurs de compléter une collection de références d'abord élaborée par une recherche par mots-clé. Notons toutefois que ces relations purement citationnelles ou bibliographiques ne qualifient pas la citation. Une étude peut être critiquée ou infirmée par plusieurs autres en raison de ses défauts, cela comptera pour autant comme des citations dans le sens positif (telle étude s'appuie sur les conclusions de telle autre). Pour aller plus loin, il faudrait ajouter une couche de traitement automatique de la langue, ce que fait le service Scite

Références

- Tay, Aaron. 2018. “Knowtro and Citation Gecko - 2 Tools for Literature Review with a Twist for Phd Students & Researchers.” Blog. *Musings About Librarianship*.
- Wölfe, Tim. 2019. “Local Citation Network and Citation Gecko: Making Literature Discovery Fun.” *Leiden Matrics*.