

# Fouille de la littérature médicale à l'aide de graphes

Elise Bigeard<sup>1</sup> Aman Sinha<sup>1,2</sup> Marianne Clausel<sup>1</sup>  
Mathieu Constant<sup>3</sup>

<sup>1</sup>IECL, Université de Lorraine

<sup>2</sup>Indian Institute of Technology, Dhanbad

<sup>3</sup>ATILF, Université de Lorraine

`elise.bigeard@univ-lorraine.fr`, `marianne.clausel@univ-lorraine.fr`

4 Février 2021

Objectif

Visualisation

Apprentissage d'embeddings de graphes

## Problème

Veille difficile face à la quantité grandissante de publications

- ▶ Pubmed : 30 million publications
- ▶ PMC : 6 million publications
- ▶ DBLP : 5 million publications

## Demandes

### Besoins des chercheurs (Canceropole)

- ▶ J'écris un papier, quels papiers citer ?
- ▶ Je lance un projet, existe-t-il déjà un projet similaire ?
- ▶ Je lance un projet, quels chercheurs recruter ?

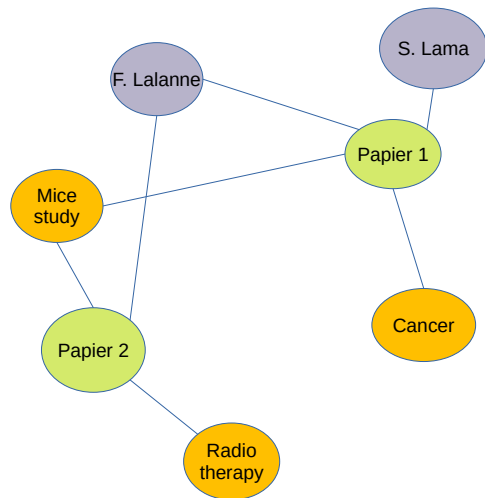
## Notre projet

Rendre la fouille de base de données de papiers plus facile

Nous créons deux outils :

- ▶ Visualisation de corpus en graphe
- ▶ Système de suggestions générique : suggestion de citations, de papiers, de collaborateurs...

# Graphe



## Travaux connexes

- ▶ Représentation de graphes [Ji et al., 2021]
- ▶ Représentation de graphes + texte  
TADW [Yang et al., 2015]  
paper2vec [Ganguly and Pudi, 2017]
- ▶ Recommandation de citations [Färber and Jatowt, 2020]
- ▶ Détection d'affirmation (claim) [Achakulvisut et al., 2019]

Objectif

Visualisation

Apprentissage d'embeddings de graphes



## Le site

<https://gremie-demonstrator.atilf.fr>

prototype !

# Le site



## REVIEW Applications of Inorganic Nanomaterials in Photothermal Therapy Based on Combinational Cancer Treatment

### Attributes

id: 957

type: article  
 fulltext link: [gremie-demonstrator.aill.fr/fulltext/nano5.html](https://www.gremie-demonstrator.aill.fr/fulltext/nano5.html)

abstract: Background: Cancer is one of the major causes of death and is difficult to cure using existing clinical therapies. Clinical cancer treatments (such as surgery, chemotherapy (CMT), radiotherapy (RT) and immunotherapy (IT)) are widely used but they have limited therapeutic effects and unavoidable side effects. Recently, the development of novel nanomaterials offers a platform for combinational therapy (meaning a combination of two or more therapeutic agents) which is a promising approach for cancer therapy. Recent studies have demonstrated several types of nanomaterials suitable for photothermal therapy (PTT) based on a near-infrared (NIR) light responsive system. PTT possesses favorable properties such as being low in cost, and having high spatiotemporal control with minimal invasiveness. However, short NIR light penetration depth limits its functions.

Methods: In this review, due to their promise, we focus on inorganic nanomaterials [such as hollow mesoporous silica nanoparticles (HMSNs), tungsten sulfide quantum dots (WS 2 QDs), and gold nanorods (AuNRs)] combining PTT with CMT, RT or IT in one treatment, aiming to provide a comprehensive understanding of PTT-based combinational cancer therapy. Results: This review found much evidence for the use of inorganic nanoparticles for PTT based combinational cancer therapy. Conclusion: Under synergistic effects, inorganic nanomaterial-based combinational treatments exhibit enhanced therapeutic effects compared to PTT, CMT, RT, IT or PDT alone and should be further investigated in the cancer field.  
 sim 1: Thomas J Webster  
 sim 2: Xia Wu

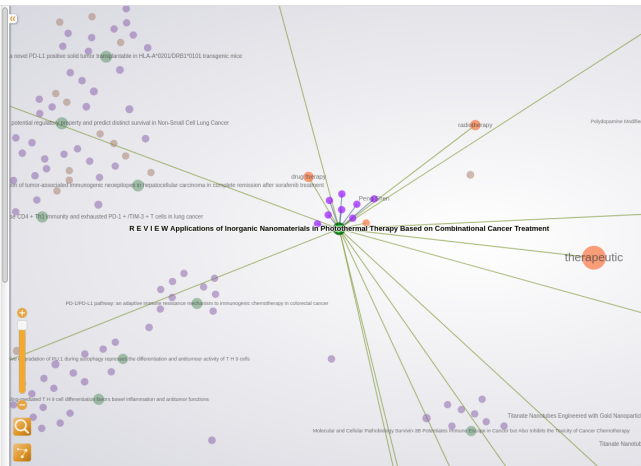
REVIEW Applications of Inorganic Nanomaterials

Switch view :

[Full text](#) | [abstract](#) | [similarity](#)

Visible nodes :

[all nodes](#) | [articles-authors](#) | [articles-keywords](#) | [keywords](#)



# Objectif

Faciliter l'exploration de publications

## Features

- ▶ Thèmes centraux
- ▶ Recherche par mot-clé, auteur
- ▶ Nœuds les plus similaires
- ▶ Exploration MeSH
- ▶ Lecture facilitée des articles : Mots-clé et parties importantes du texte

## Sous le capot

- ▶ Point de départ : articles PDF
- ▶ Grobid : infos biblio depuis PDF [GRO, 2021]
- ▶ Détection des mots-clé : Mesh, TF-IDF, Rake [Rose et al., 2010], méthodes plus avancées à venir (embeddings etc)
- ▶ Gephi : visualisation

## Particularités de notre méthode

- ▶ Utilisation du texte complet
- ▶ Graphe hétérogène, metapath, embeddings de graphes avec plusieurs types de nœuds + embeddings de texte
- ▶ Full stack, complètement automatique, depuis les PDF.  
Utilisable facilement par une unité de recherche

## À venir

- ▶ Résumé automatique
- ▶ Meilleure détection des mots-clé
- ▶ Aider les citoyens à appréhender la science (projet Olki)

Objectif

Visualisation

Apprentissage d'embeddings de graphes



## Objectif

Calculer finement des similarités pour :

- ▶ Recommandation de citation, co-auteur, etc
- ▶ Catégoriser les articles , organiser une collection par thèmes, attribuer des tags

Tâche pour le développement : prédire la conférence d'un article

# Corpora

	documents	abstract	texte complet
Cancéropole	305	✓	✓
ACM	11 050	✓	
ACL Anthology	10 566	✓	✓
DBLP	529 498	✓	
PMC	à trier	✓	✓

## Représentations de graphes

- ▶ Deepwalk [Perozzi et al., 2014]
- ▶ GraphSage [Hamilton et al., 2017]
- ▶ GCN [Kipf and Welling, 2017]

## Représentations de texte

Preprocessing : lemmatisation, suppression des mots vides

- ▶ TF-IDF
- ▶ Glove embeddings (moyenne des mots du texte)
- ▶ Doc2vec
- ▶ Bert, Scibert

## Représentation des auteurs

- ▶ Concaténation des titres de leurs articles
- ▶ Concaténation de leurs meilleurs mots-clé dans l'ordre d'importance

Le résultat est traité comme un texte par Doc2vec

## Résultats

Tâche : prédiction de conférence

Corpus : ACM

Classes : 4 groupes de conférences

Texte disponible : Titre + abstract

## Résultats

Résultats préliminaires, en développement

Feature	algo	rappel	précision	F-mesure
scibert	neurones	.83	.83	.83
tfidf	SVC	.77	.77	.77
glove	SVC	.74	.74	.74
doc2vec	SVC	.75	.76	.75

## Embeddings de graphe

Résultats moyens pour le moment avec solution out of the box.  
L'améliorer plus tard ?



## Conclusion

- ▶ Prototype fonctionnel de visualisation
- ▶ Résultats prometteurs pour l'apprentissage sur graphe



(2008–2021).

Grobid.

<https://github.com/kermitt2/grobid>.



Achakulvisut, T., Bhagavatula, C., Acuña, D. E., and Körding, K. P. (2019).

Claim extraction in biomedical publications using deep discourse model and transfer learning.  
*CoRR*, abs/1907.00962.



Färber, M. and Jatowt, A. (2020).

Citation recommendation: approaches and datasets.  
*International Journal on Digital Libraries*, 21(4):375–405.



Ganguly, S. and Pudi, V. (2017).

Paper2vec: Combining graph and text information for scientific paper representation.  
In *ECIR*.



Hamilton, W. L., Ying, R., and Leskovec, J. (2017).

Inductive representation learning on large graphs.  
*CoRR*, abs/1706.02216.



Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S. (2021).

A survey on knowledge graphs: Representation, acquisition and applications.



Kipf, T. N. and Welling, M. (2017).

Semi-supervised classification with graph convolutional networks.



Perozzi, B., Al-Rfou, R., and Skiena, S. (2014).

Deepwalk.

*Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*.



Rose, S., Engel, D., Cramer, N., and Cowley, W. (2010).  
*Automatic Keyword Extraction from Individual Documents*, pages 1 – 20.



Yang, C., Liu, Z., Zhao, D., Sun, M., and Chang, E. Y. (2015).  
Network representation learning with rich text information.  
*In Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, page 2111–2117.  
AAAI Press.