

Bibliométrie et analyse de réseaux: Quantifier les grandes tendances de la discipline économique

Séminaire MEHPERE

Aurélien Goutsmedt

08-02-2021

Introduction : Méthodes quantitatives et HPE/Philosophie économique

Objectif général de la séance

1. Introduire à la diversité de ces méthodes et aux débats en HPE/épistémologie et épistémologie économique sur leur usage.
2. Rentrer dans le détail de deux types de recherches quantitatives:
 - ▶ la bibliométrie utilisant l'analyse de réseau;
 - ▶ l'analyse de texte et la lexicométrie
3. Comprendre un peu mieux ces méthodes par la pratique.

Un intérêt récent pour les méthodes quantitatives en HPE

Numéro spécial du *Journal of Economic Methodology* de 2018

“Not Everything that can be Counted Counts: Historiographic Reflections on Quantifying Economics” Cherrier et Svorenčík (2018).

L'opposition aux méthodes quantitatives

Opposition et méfiance chez certains historiens et philosophes de l'économie. Pourquoi ça ne se justifie pas?

- ▶ manque de connaissance des méthodes;
- ▶ dimension réactionnaire d'une partie du champ : rejet de ces méthodes car on ne peut soit même les utiliser;
- ▶ arguments qui peuvent également s'appliquer aux méthodes quantitatives.

Une défense de ces méthodes

Les méthodes quantitatives sont importantes en HPE et philosophie économique pour plusieurs raisons:

- ▶ diminution du **biais de confirmation**.
- ▶ nécessité du **croisement des méthodes**.
- ▶ volonté des historiens et épistémologues/philosophes d'être des **“généralistes”** ([Trautwein 2017](#)).
- ▶ **mettre au centre les “méthodes” et “pratiques” de recherche.**

Différentes méthodes :

| Méthodes | Sources | Type de méthodes | Logiciel/outils | Exemples d'utilisation |
|---|-----------|--|---|--|
| Statistiques Descriptives | Tout type | fréquence, tri croisé. . . | Excel, langage de programmation (R, Python) | Claveau et Dion (2018) , Truc et al. (2020) , Hoover et Svorenčík (2020) |
| Econométrie et autres types de statistiques | Tout type | régressions, χ^2 , Analyse à composantes multiples, analyse séquentielle. . . | Stata, SAS, langage de programmation | Cherrier et Saïdi (2015) Goutsmedt (2021) |

Différentes méthodes (2):

| Méthodes | Sources | Type de méthodes | Logiciel/outils d'utilisation | Exemples |
|---|-------------------------|--|---|---|
| Analyse de textes assistée par ordinateur | Données textuelles | Léxicométrie, Topic-modelling | IRaMuTeQ, TXM, Voyant-tools, langage de programmation | Mohr et Bogdanov (2013), Ambrosino et al. (2018), Malaterre, Chartier, et Pulizzotto (2019) |
| Bibliométrie | Données bibliométriques | H-index, statistiques de co-écriture, de sujets/thématiques, couplage et co-citation | Excel, langage de programmation | Truc, Claveau, et Santerre (2021), Claveau et Gingras (2016), |

Les différentes sources de données :

| Sources | Types de données | Base de données & Extraction | Exemples d'utilisation |
|--|---|---|---|
| Données bibliographiques | Auteurs, date, revue, affiliations, références. . . | Web of Science ; Scopus ; extraction manuelle, ou semi-automatique via anystyle | Claveau et Gingras (2016) ; Claveau et Dion (2018) ; Goutsmedt (2021) |
| Données Prosopographiques (ou biographie collective) | lieu et date diplômes, positions occupées, associations | <i>Extraction manuelle</i> : CV et informations diverses sur internet. <i>Extraction automatisée</i> : utilisation du web-scraping <i>Population</i> : participants d'une conférence, membres d'une association ou d'une université. . . | Goutsmedt, Renault, et Sergi (2021) , Svorenčík (2018) , Hoover et Svorenčík (2020) |

Les différentes sources de données (2) :

| Sources | Types de données | Base de données & Extraction | Exemples d'utilisation |
|---------------------------------|---|---|---|
| Données textuelles (texte brut) | Mots et agencement des mots dans le texte | <i>Text-mining</i> : extraction du texte brut (après usage de l' <i>OCR</i>) | Johnson, Arel-Bundock, et Portniaguine (2019), Ibrocevic et Thiemann (2018) |
| Données diverses | Projets soumis à des fonds de recherche, correspondance entre auteurs | Extraction manuelle ou automatisée. <i>Web-scraping</i> , outil essentiel pour extraire des données sur le web | The Republic of Letters Project |

1^{er} exemple de méthode: la lexicométrie

Intervention de Justine

Intermède: R vs. logiciel “boîte à outils”

| Avantages R | Avantages logiciels |
|---|--|
| Manipulation de données nécessaires pour produire votre réseau; Avantage du script (rapidité, réplication, routine, collaboration. . .); Champ des possibles plus larges; rendement d'échelle croissant: possibilité d'utiliser R pour d'autres méthodes | Prise en main rapide et possibilité de jouer plus rapidement avec les données; manipulation plus intuitive des données |

2^{ème} exemple: Bibliométrie et analyse de réseaux

De quoi parle-t-on quand on parle de bibliométrie?

Bibliométrie:

Mesure du nombre de **publications** scientifiques d'une entité donnée (individu, institution, pays. . .), et le nombre de **citations** reçu par cette entité (voir l'introduction de [Gingras 2016](#)).

Usages de la bibliométrie

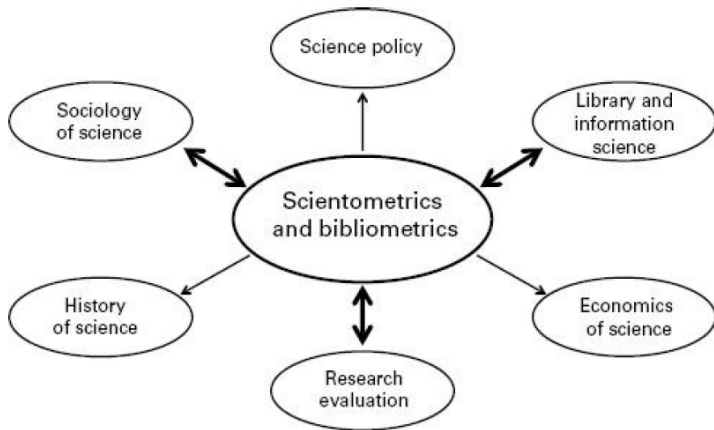


Figure 1 – Usages de la scientométrie et de la bibliométrie (Gingras 2016)

La bibliométrie et l'HPE/Philo éco

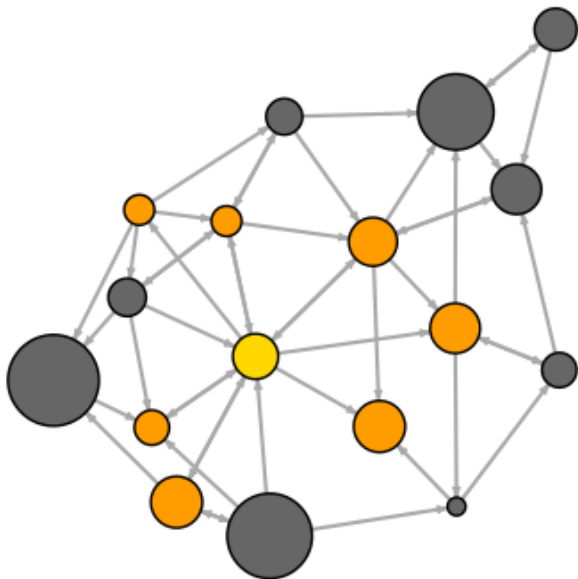
- ▶ faire un état de lieux aujourd'hui ou à différentes époques d'une discipline, sous-discipline ou d'une thématique particulière;
- ▶ explorer des traditions et héritages, en regardant notamment quels travaux du passé inspirent les travaux d'aujourd'hui;
- ▶ étudier les relations et transferts disciplinaires;
- ▶ croisement avec des questions plus sociologiques et institutionnelles, en étudiant les collaborations entre différentes institutions ou pays.

Rôle de l'analyse de réseau

Qu'est-ce qu'un réseau?

Un réseau est une structure constituée de deux types d'éléments: des noeuds (*nodes* ou *vertices*) et des liens (*edges*) qui relient ces noeuds entre eux (voir l'exemple de la figure 2).

Rôle de l'analyse de réseau



Deux méthodes d'analyse de réseaux bibliométriques

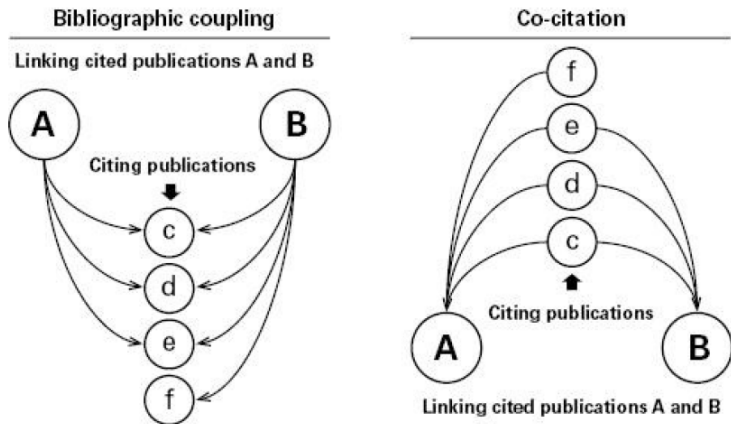


Figure 3 – Couplage bibliographique et co-citation ([Gingras 2016](#))

Le projet Macro-histoire de la macroéconomie

Brève introduction du projet

- ▶ projet financé par l'*History of Economics Society* "New Initiatives Fund";
- ▶ produire une histoire bibliométrique des grandes évolutions de la macroéconomie depuis les années 1970 + une plateforme interactive et dynamique pour diffuser les résultats obtenus;
- ▶ objectifs:
 1. être capable d'évaluer les différentes spécialités d'un champ et leur poids respectifs;
 2. Comment ces spécialités et leur poids respectifs évoluent au cours du temps?
 3. Quelles spécialités sont les plus proches et les plus éloignées?
 4. Quelle est l'importance de certains articles ou certains auteurs à différents moments de l'histoire.

Constitution du corpus et données collectées

- ▶ Utilisation des JEL codes;
- ▶ information sur les articles et sur les auteurs;
- ▶ segmentation chronologique du corpus:
 1. Le corpus n'est pas distribué de manière similaire au cours du temps: on a de plus en plus d'articles plus on avance dans le temps;
 2. Les articles les plus récents ont tendance à avoir des bibliographies plus conséquentes;
 3. On tend à citer des articles qui ont été publiés récemment, et pas de vieux articles.

Exemple pratique: manipuler les données sur GEPHI

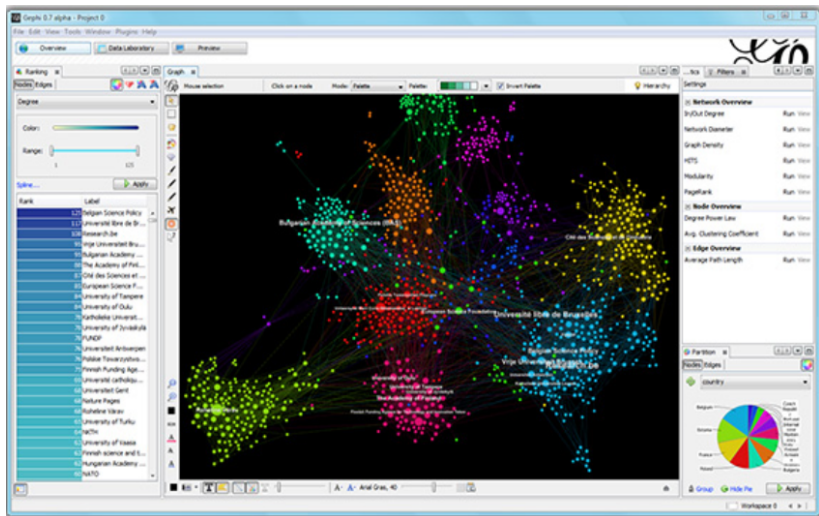


Figure 4 – Logiciel GEPHI

Exemple pratique: manipuler les données sur GEPHI

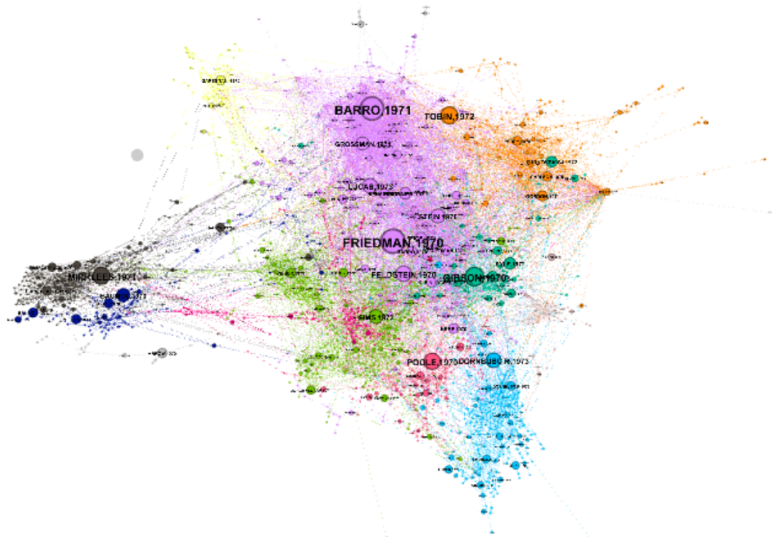


Figure 5 – Réseau produit avec GEPHI

Que faire de cela?

1. la disposition centre-périphérie, qui établit ce qui va être au coeur de la discipline, et les travaux plus périphériques, qui partagent moins de liens avec les autres.
2. la taille des communautés, via le nombre de noeuds qui les composent.
3. la proximité des communautés entre elles.

Quelles étapes ensuite?

- ▶ étudier plus précisément la composition de ces communautés (travail quali/quantitatif). Construire une catégorisation générale des communautés;
- ▶ analyser de manière plus précise les relations entre les communautés, en construisant des indicateurs de proximité, pour étudier les communautés les plus “autonomes” et les plus reliées;
- ▶ pousser l'analyse en “dynamique”, en étudiant l'évolution dans le temps de nos différentes catégories.

References

- Ambrosino, Angela, Mario Cedrini, John B. Davis, Stefano Fiori, Marco Guerzoni, et Massimiliano Nuccio. 2018. « What Topic Modeling Could Reveal about the Evolution of Economics ». *Journal of Economic Methodology* 25 (4): 329-48.
<https://doi.org/10.1080/1350178X.2018.1529215>.
- Cherrier, Beatrice, et Aurélien Saïdi. 2015. « The Indeterminate Fate of Sunspots in Economics ». SSRN Scholarly Paper ID 2684756. Rochester, NY: Social Science Research Network.
- Cherrier, Beatrice, et Andrej Svorenčík. 2018. « The Quantitative Turn in the History of Economics: Promises, Perils and Challenges ». *Journal of Economic Methodology* 0 (0): 1-11.
<https://doi.org/10.1080/1350178X.2018.1529217>.
- Claveau, François, et Jérémie Dion. 2018. « Quantifying Central Banks' Scientization: Why and How to Do a Quantified Organizational History of Economics ». *Journal of Economic Methodology* 0 (0): 1-18.
<https://doi.org/10.1080/1350178X.2018.1529216>