

Introduction aux méthodes quantitatives en HPE
et philosophie économique
Séminaire MEHPERE - 2ème séance

Aurélien Goutsmedt et Julien Gradoz

2022-02-03

Aperçu des débats, types de données et
méthodes

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;
 - ▶ les statistiques multidimensionnelles.
3. Comprendre un peu mieux ces méthodes par la pratique.

Un intérêt récent pour les méthodes quantitatives en HPE

Numéro spéciale *Journal of Economic Methodology* de 2018

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

Une opposition qui demeure dans le champ

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

- ▶ manque de connaissance des méthodes;
- ▶ rejet de ces méthodes car on ne les utilise pas soi-même?
- ▶ 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 ;
- ▶ Un outil de “découverte” tout autant (voire plus?) qu’un outil de “preuve” ;
- ▶ nécessité du croisement des méthode ;
- ▶ volonté des historiens et épistémologues/philosophes d’être des “généralistes” (Trautwein 2017) ;
- ▶ mettre au centre les “méthodes” et “pratiques” de recherche.

Différences sources

Sources	Types de données	Base de données et Extraction	Exemples d'utilisation
Données bibliographiques	Auteurs, date de publication, revue, affiliations des auteurs, références citées. . .	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 des diplômes, positions occupées, participation à des associations	<i>Extraction manuelle</i> : CV et informations diverses sur internet. <i>Extraction automatisée</i> : utilisation du <i>web-scraping</i> <i>Population</i> : participants d'une conférence, membres d'une association ou d'une université. . .	Goutsmedt, Renault et Sergi (2021), Svorencik et Hoover (2020)
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. Web-scraping, outil essentiel pour extraire des données sur le web	The Republic of Letters Project

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), Hoover et Svorencik (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)
Analyse de textes assistée par ordinateur	Données textuelles	Léxicométrie, Topic-modelling	IRaMuTeQ, TXM, Voyanttools, langage de programmation	Malaterre, Chartier, et Pulizzotto (2019)
Bibliométrie	Données bibliométrique	H-index, statistiques de co-écriture, de sujets/thématiques, couplage et co-citation	Excel, langage de programmation	Truc, Claveau et Santerre (2018), Claveau et Gingras (2016)
Analyse de réseaux	Plutôt données bibliométriques et prosopographiques	couplage et co-citation, statistiques de réseau (degré, intermédiarité. . .), partitionnement	GEPHI, langage de programmation	Herfeld et Doehne (2019)

R vs. logiciel “boîte à outils”

Avantages R

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

Avantages logiciels

Prise en main rapide et
possibilité de jouer plus
rapidement avec les données

Manipulation plus intuitive des
données

Premier 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

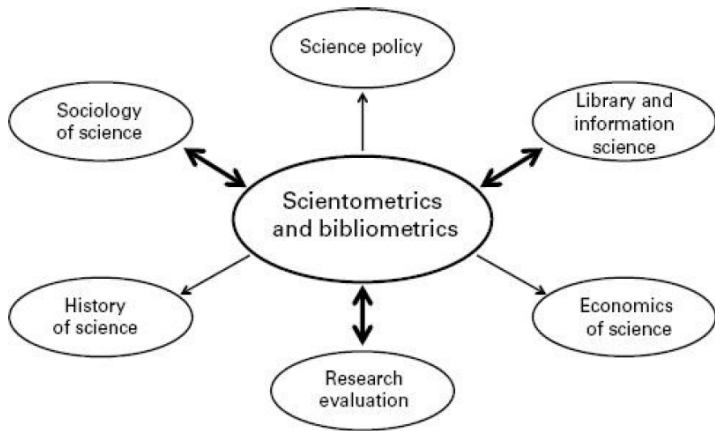


Figure 1: Usages de la scientométrie et de la bibliométrie (Gignras 2016)

La bibliométrie et l'HPE/Philo éco

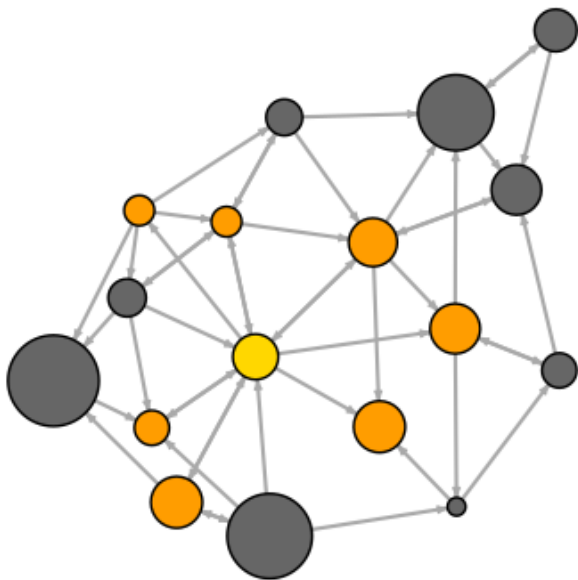
- ▶ Faire un état de lieux aujourd'hui ou à différentes époques d'une discipline, sous-discipline ou d'une thématique particulière;
- ▶ Explorer des traditions et héritages, en regardant notamment quels travaux du passé inspirent les travaux d'aujourd'hui;
- ▶ Etudier les relations et transferts disciplinaires;
- ▶ 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 nœuds (nodes ou vertices) et des liens (link ou edges) qui relient ces nœuds entre eux.

Rôle de l'analyse de réseau



Deux méthodes d'analyse de réseaux bibliométriques

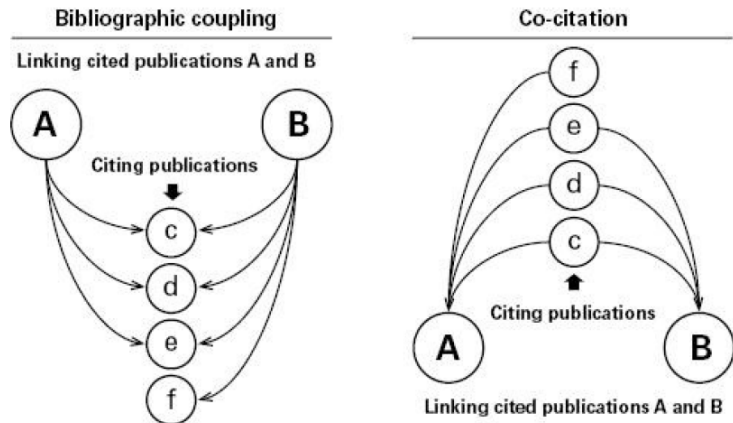


Figure 2: Couplage bibliographique et co-citation (Gingras 2016)

Où trouver les données?

Les bases de données à accès limités:

- ▶ Web of Science
- ▶ Scopus

Des bases de données ou plateformes plus récentes:

- ▶ Dimensions
- ▶ Constellate de JSTOR.

Existence d'“APIs” qui permettent d'automatiser les requêtes et facilitent les manipulations sur R.

Cas pratique

- ▶ Utilisation de Scopus:
 - ▶ Recherche des articles utilisant des modèles DSGE;
- ▶ Nettoyage des données et création d'un réseau de co-citation avec R (tutoriel plus complet [ici](#));
- ▶ Projection du réseau avec Gephi

Que faire du réseau produit ?

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/ quanti). 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

Deuxième exemple: Les statistiques descriptives multidimensionnelles en HPE