The Fabric of Knowledge

Investigating contemporary philosophy by the methods of scientometrics

Eugenio Petrovich, PhD
Post-doctoral Researcher
Department of Economics and Statistics - University of Siena
First class: introduction to scientometrics and citation analysis
  - A brief history of scientometrics
    - The exponential growth of science and the «information crisis» (1950s)
    - Eugene Garfield and the invention of the Science Citation Index
  - Descriptive and Evaluative Scientometrics
    - The evaluative scientometrics and the advent of performance indicators
    - The descriptive scientometrics and the science of science
  - Examples of research in descriptive scientometrics
  - Scientometrics and Digital Humanities
Second class: Investigating contemporary philosophy by scientometric methods

The methodological challenge of Late Analytic Philosophy (LAP)

Scientometric methodologies to probe LAP:
1. Mapping the structure and dynamics of LAP → Co-citation analysis
2. Investigating the normalization of LAP → Citation context analysis

Some theoretical reflections
Introduction to Scientometrics

First Class, May 21
Let us start by some numbers...

- In **2009**, we passed the **50 million** mark in terms of the total number of **science papers published since 1665**.

- Today, there are **between 25 and 45 000 scientific journals active worldwide**.

- In **2016**, **2,558,812 scientific papers** were published in the journals covered by the database Scopus.

- Imagine it takes 5 minutes for reading one article and that you read 8 hours per day, all the days in the year: you will need **~27 000 days (~73 years)** for reading the scientific world production of 2016 alone.
Looking at science from a quantitative perspective: Sciento/biblio/infor-metrics

- **Scientometrics** → the quantitative study of science as an informational process
- **Bibliometrics** → the application of mathematics and statistical methods to books and other media of communication
- **Informetrics** → the mathematical study of the objects of information science
Outline

1. A brief history of scientometrics (30 min)
2. Descriptive and evaluative scientometrics (30 min)

- Break -

1. Examples of descriptive scientometrics (20 min)
2. Scientometrics and Digital Humanities: similarities and differences (10 min)
3. Questions (10 min)
Prelude: From Little Science to Big Science

Issues in post-war science:

- **Strategic role** of science and technology:
  - Manhattan Project
  - Sputnik’s launch (1958)
  - Kennedy: science had become a «national necessity»

- **1950s: scientific literature explosion → the «Information Crisis»**
  - «Because of the tremendous growth of the literature, there is the danger of science fragmenting into a mass of repetitious findings, or worse, into conflicting specialties that are not recognized as being mutually inconsistent. This is the essence of the ‘crisis’ in scientific and technical information» (PSAC, 1963)

→ How to navigate the ocean of scientific information?
A solution to the crisis: The creation of the Science Citation Index

- Eugene Garfield (1927-2017) launched the Science Citation Index as a solution to the information crisis.
- The Science Citation Index is a large-scale, multi-disciplinary database that indexes all the citation links between scientific documents.
The idea behind the SCI: the structure of a scientific paper
A network approach to the scientific literature

Example of citation network

Time
A network approach to the scientific literature

Bibliographic links (the bibliography of a paper) = outgoing links
A network approach to the scientific literature

Citation links (the citations gathered by a paper) = incoming links
The destiny of the SCI

- The SCI allowed a better and more efficient information retrieval than traditional subject headings
- However, Garfield quickly realized that it could be used for other purposes:
  A. Studying quantitatively the properties of the scientific literature
  B. Evaluating the scientific impact of papers and scientists

A → Descriptive scientometrics (quantitative study of science, «science of science»)
B → Evaluative scientometrics (science policy, research evaluation)
Evaluative scientometrics

- **Premise 1:** scientists build on the work of other scientists, as reported in the scientific communications, i.e., the scientific articles («standing on the shoulders of giants»)
- **Premise 2:** The links with previous work are marked by the citations because scientists recognize their «intellectual debts»
- Therefore, a paper that receives many citations contains some result or method that has been used by many scientists. By contrast, a paper with few citations was judged less important by the scientific community
- The number of citations gathered by an article can be used as an indicator of the impact of a paper on science

It becomes possible:
1. To compare different papers by their citation scores
2. To evaluate journals according to their average citation rates (Journal Impact Factor)
3. To assess the impact of scientists and groups of scientists (e.g., university departments)
Citations vs. peer review

Quality control in science

Peer review
- Expert-based
- Time-consuming
- Qualitative/"Subjective"
- Manageable by the non-expert
  - Quick and dirty
  - Quantitative/"Objective"

Scientometrics
The development of scientometric indicators
Descriptive scientometrics

- Descriptive scientometrics does not focus on evaluation but on the investigation of the quantitative properties of scientific information as it is embodied in the scientific literature.

- Some topics in descriptive scientometrics:
  - Distribution of scientometric properties (bibliometric laws, citation patterns)
  - Science mapping (visualizing the structure of science)
  - Differences in citation behavior among the scientific disciplines
  - Evolution of the quantitative properties of science over time
  - Citation context and content analysis
  - Co-word analysis
  - Citation theory
  - ...
BREAK TIME
Bibliometric laws

“...inequality is the pattern of information processes under which a few authors are responsible for most of the scientific literature in a given research field [Lotka’s law]; a few scientific journals publish the majority of the papers relevant to any given subject [Bradford’s law]; and a relatively small number of scientists producing a given individual linguistic behavior in scientific communication....” (De Bellis, 2014, p. 37)
Dynamic mapping

https://www.leydesdorff.net/journals/nanotech/
Citation theory

- What do citations measure?
  - Impact/quality/use/...?

- Which motivations do scientists have for citing?

- What are the determinants of the scientists’ citing behavior?

Two competing paradigms:

1. The normative theory of citations (Merton) → scientists cite for paying their intellectual debts, citations go to the best papers in the field

2. The socio-constructivist theory of citation (STS) → scientists cite for strategic reasons, to persuade their peers, citations go to authoritative scientists
They are two distinct fields, with different targets and different origins. Still, they hold similar perspectives on the human intellectual production (be it the sciences or the humanities):

1. Quantitative standpoint
2. Data-driven epistemology
3. Statistical methods for handling the data
4. Computer-aided analysis
Investigating contemporary philosophy by scientometric methods

Second Class, May 22
Outline

1. Numbers and methods: the challenge of contemporary philosophy (10 min)
2. Late Analytic Philosophy: an object difficult to handle (15 min)
   - Scales in the history of philosophy
3. Mapping LAP: Co-citation analysis (20 min)
   - Break -
4. The normalization of LAP: Citation context analysis (20 min)
5. Questions (20 min)
Let us start (again) by some numbers...

"The number of journal articles and book chapters in the Philosopher's Index (PI) doubled every five years between 1945 and 1965, doubled again at seven-year intervals from 1966 to 1980, and doubled again in the next ten years (an exponential growth rate of 7% a year) - to about 156,000 works. By 1995 the number exceeded 173,000. Publication output of monographs having Philosophy as a Library of Congress subject descriptor shows a rather steady growth rate: it doubled every 25 years between 1900 and 1975 and doubled again in the past 18 years, approaching 174,000 books in 1995."

Schwartz, 1995, p. 147

APA membership over time. Source: elaboration from (Soames, 2008) and (Schwartz, 1995)

History of Philosophy - University of Turin, 21-22 May
The methodological challenge

- The traditional method of the historiography of philosophy is the close reading of texts
- This method is inadequate for handling the thousands of texts produced by contemporary philosophers
  → We need a massive approach
The object of investigation: Late Analytic Philosophy

- Tripodi 2015 periodization:
  - Early, Middle, and Late Analytic Philosophy
  - LAP = analytic philosophy developed approximately over the last forty years

- LAP is a **difficult object** to handle with the traditional methods of the historiography of philosophy:
  1. Quantitative growth
  2. Field-level phenomena (as perceived by historians and philosophers)
     1. Specialization and fragmentation
     2. Professionalization, technicalization, a scientific style of intellectual production
Scales in the history of philosophy: micro → macro

- Local → Global
- Space
- Short → Longue durée
- Time
- Individual → Collective
- Social
- Growth
- Specialization
- Fragmentation
- Technicalization
First study | Co-citation analysis

1. The basic idea behind co-citation analysis is that two publications that are frequently cited together are more «similar» than publication that are only seldom cited together.

2. The Citation Index allows to generate the co-citation matrix that can be visualized as a network (a «science map»).

3. The graph shows the citation network among the publications P1, P2, P3, P4, P5, and P6.
Methodology

1. Operationalization of «analytic philosophy» as the set of papers published in five prestigious journals: «Top Five» journals

2. Definition of the timespan of the analysis: 1985-2014

3. Download of the records (metadata + cited references) from Web of Science: 4,966 articles, 58,281 references

4. Co-citation analysis and visualization: VOSviewer tool
On the operational definition of LAP

Operationalization = “The process of transforming an abstract concept or theory into an empirical, testable subject of research. Proper operationalization is therefore crucial to obtaining relevant results and is especially at stake in the formulation of research methods” (Oxford Dictionary of the Social Sciences)

→ Operational definition (OD)

OD of LAP:

All the articles published in five “top” journals in analytic philosophy (Philosophical Review, Noûs, Journal of Philosophy, Mind, Philosophy and Phenomenological Research) between 1985 and 2014 → 4 966 articles, 58 281 cited references

Losses in the operationalization:

- Published side vs. philosophy in the making
- Explicit citations vs. implicit references
- Focus on the metadata vs. focus on the full text
The structure of LAP

Metaphysics
- Lewis, 1986, plural worlds
- Lewis, 1974, counterfactuals

Epistemology
- Harman, 1986, change view
- Nozick, 1981, philos. explanations

Ethics and Political Philosophy
- Williams, 1981, moral luck

Philosophy of Mind
- Kripke, 1980, naming necessity

Philosophy of Language
- Kripke, 1980, word objects

Overall map (1985-2014)

History of Philosophy - University of Turin, 21-22 May
The dynamics of LAP

The **pattern** visible in subsequent science maps (**from a sparse to a clusterized network**) can be interpreted as a **dynamic of specialization** occurring in the field (**structuration in sub-disciplinary specialized literature**).
BREAK TIME
Citations have an **epistemological function** in the citing texts. For instance, they can be used:

- To **overview** the state-of-the-art (state of the art citation)
- To **support** a claim (positive citations)
- As a target of **criticism** (negative citations).

The citation network of a **Kuhnian normal science** is characterized by a shared state-of-the-art and by a high rate of positive, instead of negative, citations.

→ The **normalization** of a field (= its becoming closer to a normal science) can be investigated by studying the **transformation of citation functions in its literature**.

Citation context analysis is a technique that allows to classify citations according to their epistemological function.
Methodology

1. Definition of the sample: 60 most cited articles published in the «Top Five» journals between 1950 and 2010, divided in 6 10-year timespans of 10 articles each.

2. Download of full texts: 60 articles, 1,293 references.

3. Design of the Classificatory Scheme: 7 functions of references.


5. Statistical analysis of the results: aggregate and longitudinal analysis.
# The classificatory scheme

<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>State of the art</td>
<td>The reference is used to provide an overview of the state of the art of the field. It is «neutral»: the citing author does not use it either to support her argument or to criticize the cited document.</td>
</tr>
<tr>
<td>2</td>
<td>Supporting</td>
<td>The reference supports the argument of the citing author, either because it brings additional arguments to the stated claims or because it strengthens the author’s arguments since the authors agree on specific topics.</td>
</tr>
<tr>
<td>3</td>
<td>Supplementary/Perfunctor</td>
<td>The reference to the cited document is accidental. The cited work it is not essential to follow or support the citing author’s argument.</td>
</tr>
<tr>
<td>4</td>
<td>Acknowledgment</td>
<td>The citing author pays a form of tribute or acknowledgement towards the cited document.</td>
</tr>
<tr>
<td>5</td>
<td>Critical</td>
<td>The cited document is criticized, the citing author disagrees with the cited document.</td>
</tr>
<tr>
<td>6</td>
<td>Documental</td>
<td>The cited document is used to support the historical reconstruction of the argument of the cited author provided by the citing author.</td>
</tr>
<tr>
<td>7</td>
<td>Unknown</td>
<td>The citation does not fit into any of the previous categories.</td>
</tr>
</tbody>
</table>
# Epistemic functions

<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>State of the art</td>
<td>The reference is used to provide an overview of the state of the art of the field. It is «neutral»: the citing author does not use it either to support her argument or to criticize the cited document.</td>
</tr>
<tr>
<td>2</td>
<td>Supporting</td>
<td>The reference supports the argument of the citing author, either because it brings additional arguments to the stated claims or because it strengthens the author’s arguments since the authors agree on specific topics.</td>
</tr>
<tr>
<td>3</td>
<td>Supplementary/Perfunctory</td>
<td>The reference to the cited document is accidental. The cited work is not essential to follow or support the citing author’s argument.</td>
</tr>
<tr>
<td>4</td>
<td>Acknowledgment</td>
<td>The citing author pays a form of tribute or acknowledgement towards the cited document.</td>
</tr>
<tr>
<td>5</td>
<td>Critical</td>
<td>The cited document is criticized, the citing author disagrees with the cited document.</td>
</tr>
<tr>
<td>6</td>
<td>Documental</td>
<td>The cited document is used to support the historical reconstruction of the argument of the cited author provided by the citing author.</td>
</tr>
<tr>
<td>7</td>
<td>Unknown</td>
<td>The citation does not fit into any of the previous categories.</td>
</tr>
</tbody>
</table>
Results: State of the art citations

Emergence of a shared state-of-the-art

Building of a «knowledge base»
Results: Critical citations

- The disagreement in the community decreases over time
Results: Supporting citations

However, Positive citations did not raise but show an unstable trend.

The paradoxical situation can be explained by two hypotheses:

- AP underwent a process of fragmentation.
- A soft, instead of strong, paradigm has taken over.
Some theoretical considerations

- **The Agent/Structure problem:**
  - what is the role of the individual philosopher faced with the philosophical field?
  - How does the individual contribute to philosophical change?
  - How do individual agents perceive the structure of the field on themselves?

- **History of philosophy as a social science?**
  - Social Studies of philosophy
  - Quantitative methods (scientometrics, text analysis, ...)
  - Qualitative methods (interviews, ethnographies, ...)

History of Philosophy - Università di Torino, 21-22 May