How to Measure Information Cocoon in Academic Environment

Jia Yuan[1], Guoxiu He[1] and Yunhan Yang[2,*]

[1]School of Economics and Management, East China Normal University, Shanghai, China
[2]Faculty of Education, The University of Hong Kong, Hong Kong, SAR, China

Presenter：**Yunhan Yang**   Email: yunhany@connect.hku.hk

Background and Motivation

Information cocoon

- Only pay attention to what they choose and the field that makes them happy[1,2].
- Affect the diversity of information acceptance[3].
- Can apply to any environment where information is generated.

in Academic Environment?

- Scholars consume homogeneous information
- Issues: Information narrowing, reduce innovation, emerge research bottlenecks, etc.
- Questions: How prevalent is the information cocoon in academia? How can it be measured? And what variations exist among different groups?
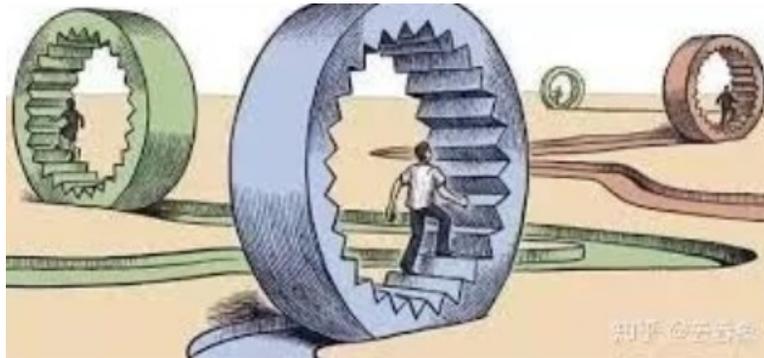
Research gaps and objectives

Research gap

- Few studies were concerned information cocoon in academia
- Few studies have measured the extent of it in academia

Objectives

- Measure the information cocoon in the academic environment
- Apply metrics to diverse groups

Context

Continuous **horizontal** (the ability to engage with a wide range of different fields and topics) and **vertical development** (in-depth exploration of specific fields) is critical for scholars.

- focuses solely on horizontal development - a superficial understanding and lack specialized knowledge
- pursuing only vertical development - deep research in a particular area but fails to integrate knowledge from different fields, thus limiting the breadth of research.

Therefore,
Scholars need to balance these two modes of development throughout their careers.
The depth and breadth of research directly impact scholars' research outcomes.

Framework

Evaluate the extent of information cocoons requires a comprehensive consideration of **both depth and breadth**.

We believe that the combination of research depth and breadth can help break through the information cocoon.

Context

Scholars with a high degree of academic cocoon may exhibit one of the following characteristics:

- prolonged focus on a single topic without achieving breakthrough innovation, leading to academic stagnation;

- or an excessive pursuit of research breadth, spanning multiple fields but struggling to generate valuable research outcomes.

Data

Source: the Semantic Scholar Open Research Corpus (S2ORC)

Time window: papers published **between 2010 and 2021**

Metadata: article **titles, first authors, reference titles, publication dates, and citation counts**.

Exclusion:
- Duplicates and incomplete entries;
- Authors with sporadic publication patterns, focusing on those who averaged more than 2 articles per year.

Final dataset size: 107,775 articles.

Metrics

Reference_depth

Self_depth

Research depth

Paper-level:
Measure the substantive variance (Sentence-BERT) between paper title and titles in reference list.

$$Ref\_depth = 1 - \frac{\sum_{i=1}^{n} R(a, b_i)}{n} \qquad (1)$$

Note: a refers to the target paper, b_i refers to the reference paper cited by paper a

Individual-level:
Scholars who achieve breakthroughs in research depth often showcase distinct differences from their prior research.

$$Self\_depth = \begin{cases} 1 - \dfrac{\sum_{j=i+1}^{i+3} R(a_i, b_j)}{3}, i+3 \leq n \\ 0, i+3 > n \end{cases} \qquad (2)$$

Note: a_i refers to the target paper, b_j refers to the previous paper published by author a

Metrics

Reference_breadth

Self_breadth

Research breadth

Paper-level:
Apply the BERTopic model to classify each reference title into topics, and record the frequency of each topic in a paper.

Paper-level:
Use Gini coefficient to measure the degree of topic distribution in one paper.

$$Ref\_breadth = \frac{ref\_topic\_counts}{10} \quad (3)$$

$$Self\_breadth = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|X_i - X_j|}{2n^2\bar{x}} \quad (4)$$

Note: ref_topic_counts is divided by 10 in (3) aims to normalize for all papers.

Note: n is the number of references, X_i, X_j represents the i-th and the j-th reference, and $\bar{x}$ is the average value of all references.

Metrics

cocoon value

- The four metrics presented in the previous section aim to break the information cocoon.

- **A decrease in the depth and breadth values** indicates that the article is limited by a single piece of information, thus leading to **an increase in the cocoon value**. Conversely, an **increase in the depth and breadth values** indicates that the literature has the potential to **break out of the information cocoon.**

$$C = Avg\left\{ \sum_{i=1}^{4}(1 - M_i) \right\} \qquad (5)$$

Note: M_i is the four indicators calculated above.
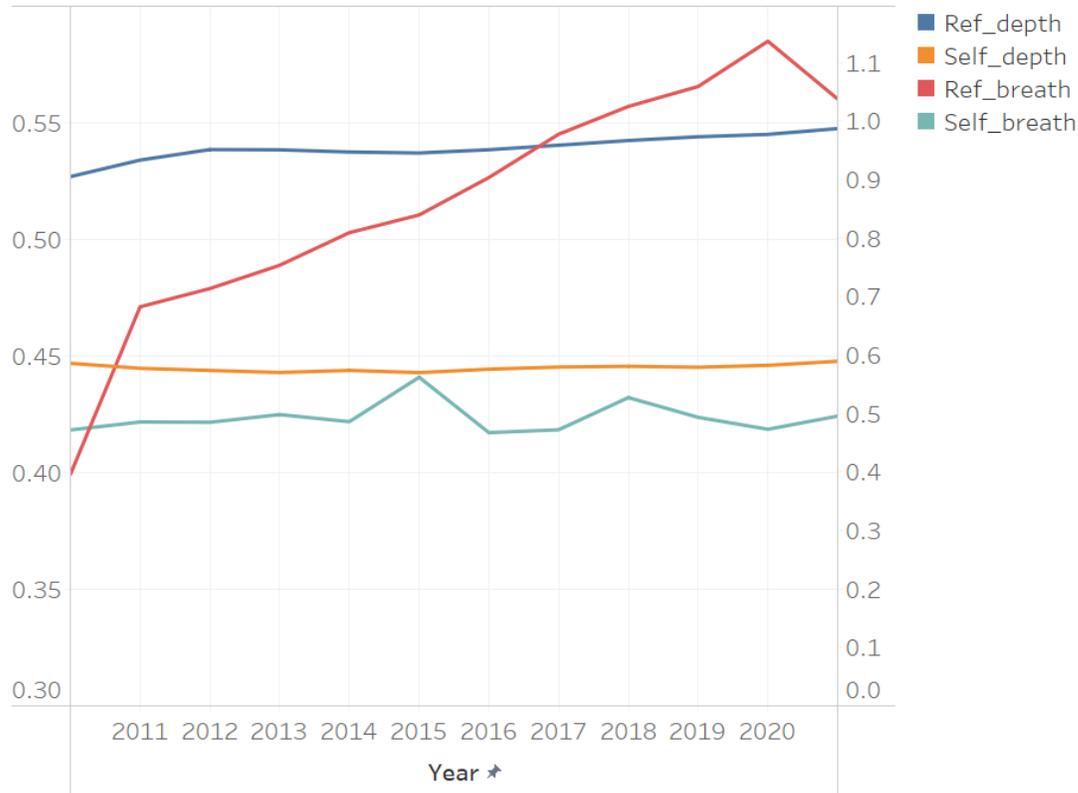
The Whole Academic Environment

Figure 1. The trend of changes in the depth and breadth dimension.

Figure 2. The trend of the overall cocoon value.

- the changes in the two **depth indicators** are relatively **stable**
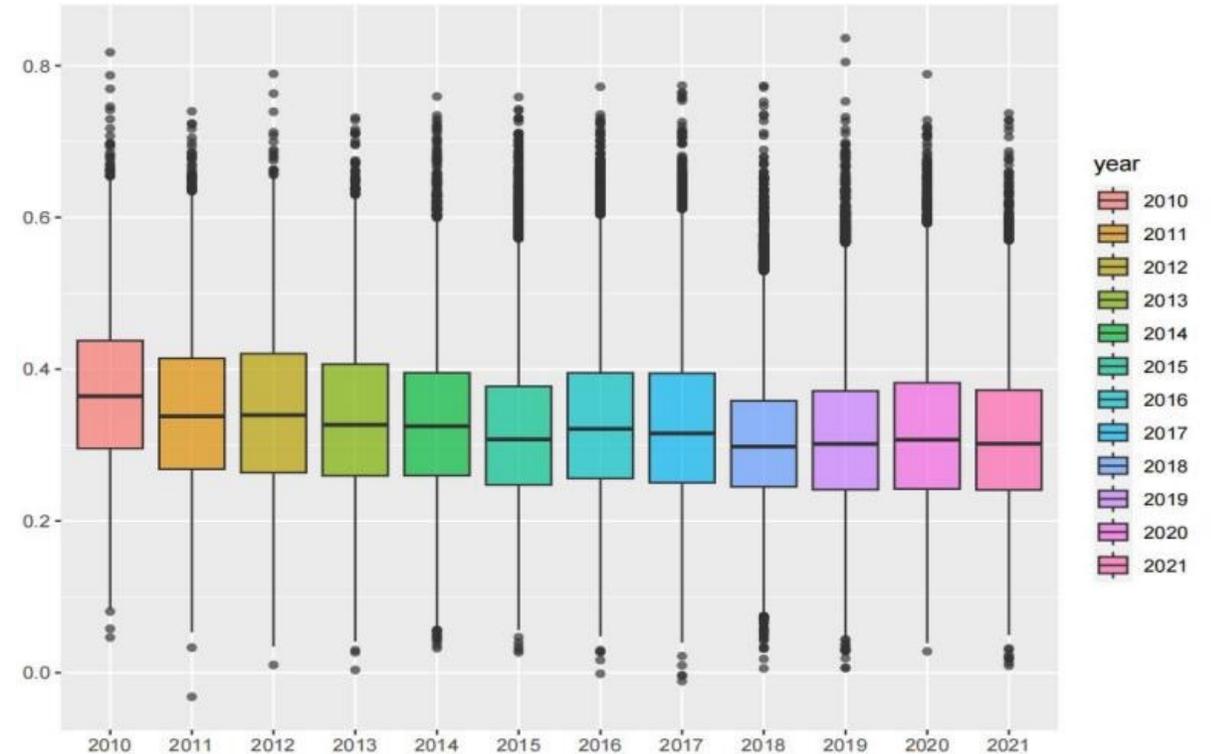- a noticeable **increase** in the "**Ref_breadth**" indicator

- a **decreasing trend** year by year
- a **continuous opening up and innovation** of information in academic, which is a positive phenomenon
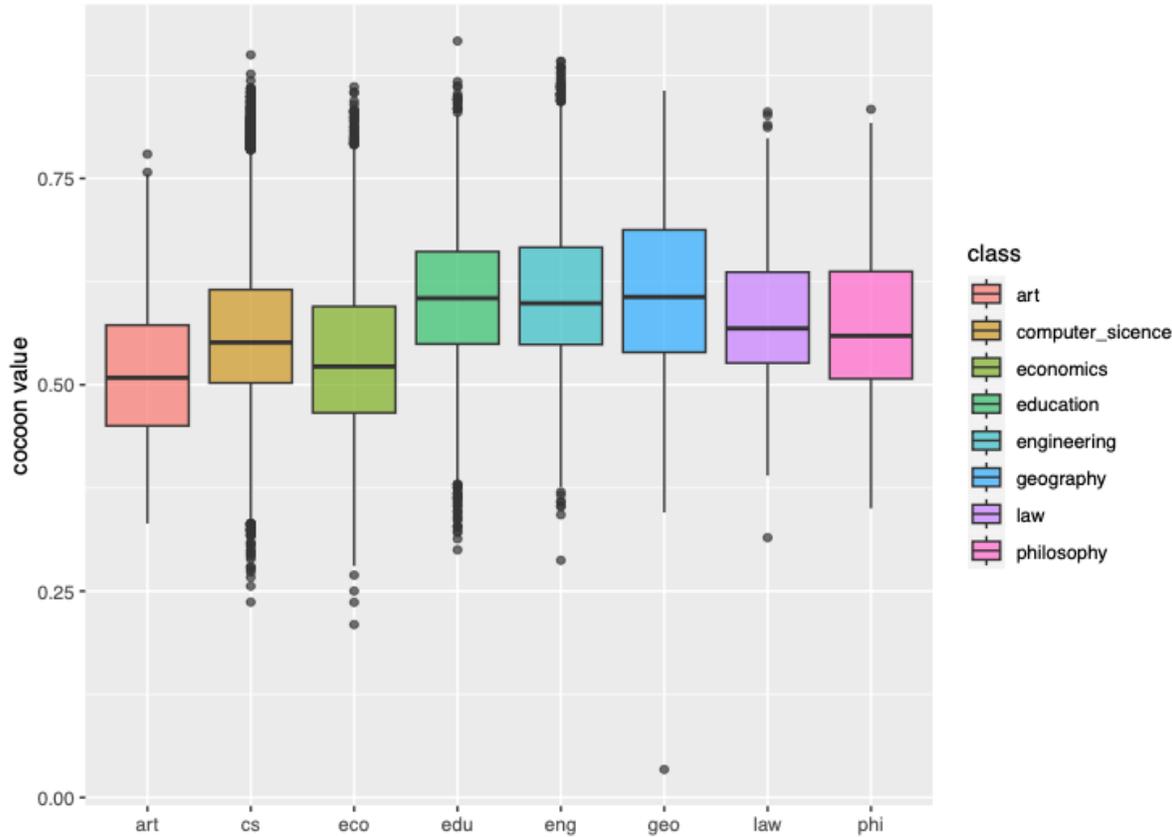
Different Disciplines

Figure 3. Information cocoon value of different disciplines.

Class	Ref_depth	Self_depth	Self_breadth	ref_breadth
art	0.6156	0.7020	0.5413	0.3356
computer_science	0.5446	0.5970	0.5172	0.4599
economics	0.5927	0.6339	0.5119	0.6375
education	0.5169	0.5469	0.4245	0.4541
engineering	0.5179	0.5082	0.4453	0.4450
geography	0.4989	0.5093	0.4219	0.6913
law	0.4977	0.6055	0.4373	0.6637
philosophy	0.5790	0.6453	0.4241	0.3327

Figure 4. Depth and breadth of different disciplines.

- disciplines with **smaller information cocoons tend to show a higher level of interdisciplinary and openness**, such as computer science and art;
- **traditional disciplines tend to have large information cocoons**.
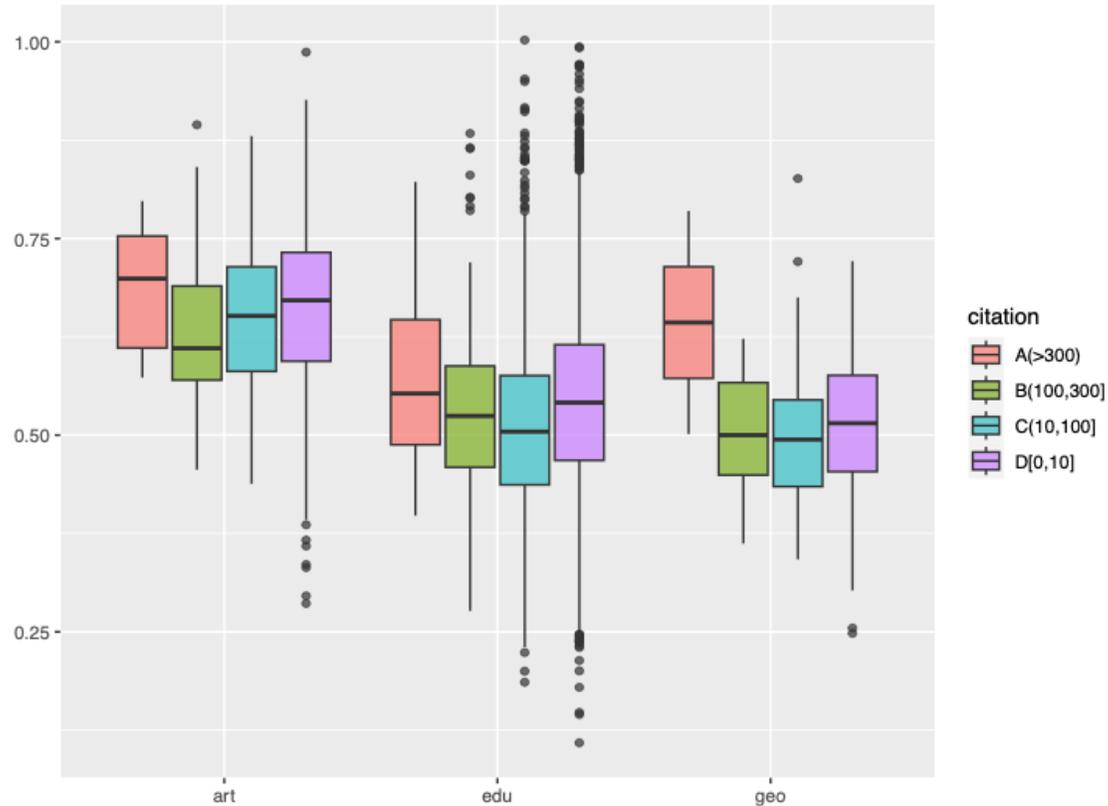
Citations Classification

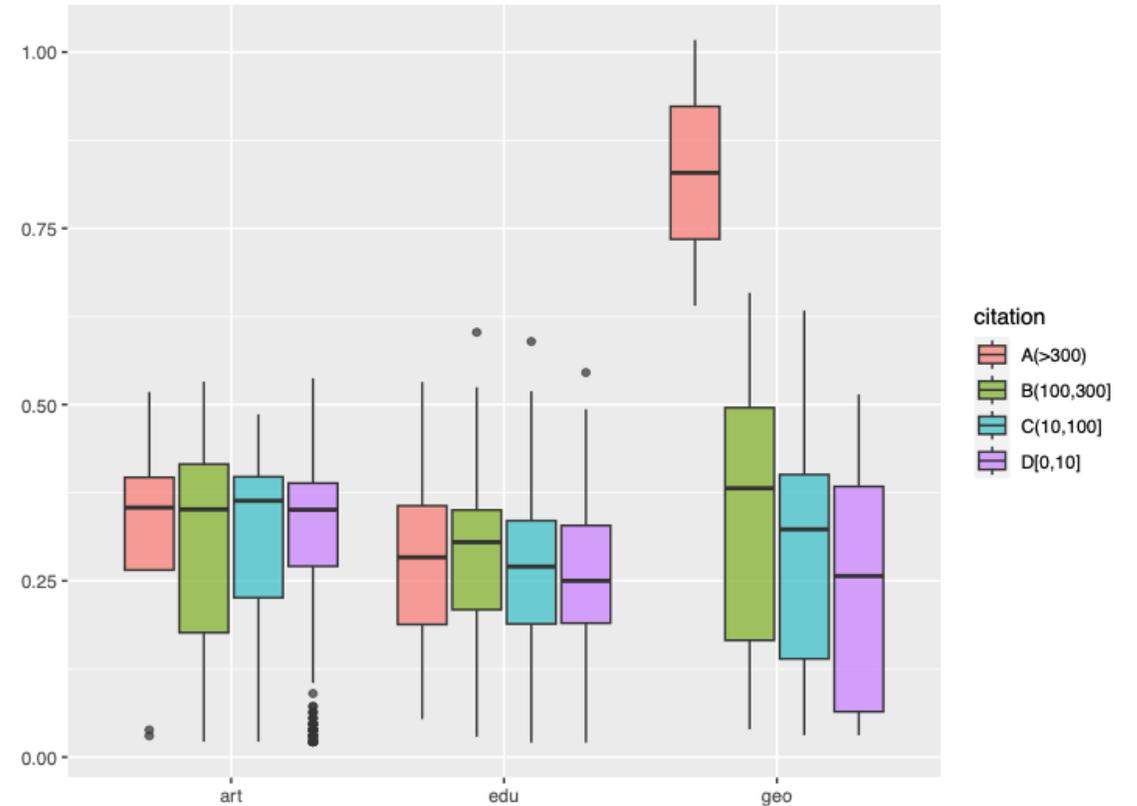Figure 5. The depth value of different citations.

Figure 6. The breadth value of different citations.

- **The most cited articles** usually has **higher research depth and breadth** (in red color);
- **Highly cited articles** in group B (in green color) **focuses on academic hotspots** and attracts scholars with great breadth, but the depth may be insufficient;
- **The less-cited articles** (in blue and purple) is **limited in research breadth but high in depth,** it may be difficult to be accepted because it explores a niche issue or has a high degree of depth.

Findings

- Firstly, we observe **a gradual breakdown of information cocoon** within the overall academic environment, presenting a trend towards greater comprehensiveness and innovation.

- Secondly, significant **disparities exist** in terms of depth, breadth, and cocooning across **traditional, creative, and technological disciplines**.

- There are **variations** in breadth and depth among **different citation groups**, with literature possessing **greater breadth and lesser depth** often garnering **wider acceptance**.

Researchers should adeptly utilize extensive and intricate academic information, continually assessing whether their research processes are constrained by information cocoon.

Shortcomings and Prospects

FUTURE FORECASTS

- **Abstracts** have not been included in the calculation
- **Subject coverage** could be expanded further
- **Regression analyses** can be used to further explore the underlying influence mechanisms

Thank you for your attention

How to Measure Information Cocoon in Academic Environment

Presenter：**Yunhan Yang**   Email: yunhany@connect.hku.hk

Reference

[1] Yuan, X., and Wang, C. (2022). Research on the Formation Mechanism of Information Cocoon and Individual Differences among Researchers Based on Information Ecology Theory, Frontiers in Psychology (13).

[2] Sanstan, (2008). Information Utopia: How People Produce Knowledge", Translated by Bi Jingyue Beijing: Law Publishing House(8).

[3] Falck, A., and Boyer, K. 2022. Online Filters and Social Trust: Why We Should Still Be Concerned about Filter Bubbles.