

Scientific knowledge combination in networks: new perspectives on analyzing knowledge absorption and integration

Dr Hongshu Chen Assistant Professor

School of Management and Economics Beijing Institute of Technology, Beijing 100081, China Hongshu.Chen@bit.edu.cn

Hongshu Chen, Jingkang Liu and Zikai Liu. 2023. Scientific knowledge combination in networks: new perspectives on analyzing knowledge absorption and integration. In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries 2023 (JCDL 2023), Santa Fe, New Mexico, USA, 5 pages.

1 Introduction Background



- Recombinant innovation is considered a significant driver in generating new ideas and has been evidenced to have a higher rate of occurrence in scientific papers.
- Researchers absorb data, information and knowledge by referencing the existing literature, and subsequently generate innovative ideas and insights with knowledge combination and integration.
- Thus, the procedure of *knowledge absorption* is encoded by the **content** of the references, while the process of *knowledge integration* can be evaluated by examining the **content** of the articles themselves.



Evidence from the NSFC Major Research Plan, Q. Jin., **H. Chen.,** X. Wang., et al., working paper.

- S. Fortunato, T. Bergstrom Carl, K. Börner, et al., "Science of science," Science, vol. 359, no. 6379, pp. eaao0185, 2018.
- B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones, "Atypical Combinations and Scientific Impact," Science, vol. 342, no. 6157, pp. 468-472, 2013.
- ▶ R. J. Sternberg, "The nature of creativity," CREATIVITY RESEARCH JOURNAL, vol. 18, no. 1, pp. 87-98, 2006.
- ▶ L. Fleming, "Recombinant Uncertainty in Technological Search," Management Science, vol. 47, no. 1, pp. 117-132, 2001.

1 Introduction





• The prior literature considers IPC codes (Wang et al., 2014), keywords(Guan et al., 2017), key phrases (Jee et al., 2022), MeSH terms(Wang et al., 2023), topics or predefined tags (Li et al., 2023) as a proxy for knowledge elements.

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

• The knowledge elements in this study refer to the integral and core concepts of a scientific article.

Research Motivation

- *Citation patterns* are more frequently used than the body of knowledge to analyze evolutionary trajectories (Wang et al., 2023, Wu et al., 2019, Roach and Cohen, 2013).
- *Citation patterns* that work with rough-grained document comprehension and predefined categorizations, have been used to investigate the process of knowledge absorption (Wagner et al., 2011, Bu et al., 2021), yet both *rough-grained topics* and *explicit taxonomies* have limitations in directly indicating fundamental content and context.
- Recent studies have inspired discussions that *higher-order network structure* is also important for understanding research contents and contexts for scientific innovations (Shi and Evans, 2023). Further research on reflecting knowledge elements and their structures thus is still warranted (Seibert et al., 2017).

1 Introduction







2 Methodology

KeyBERT

- KeyBERT leverages BERT embeddings to extract the keywords most similar to a given document.
- The length of *n*-gram terms is set as one word and two words (*n*=1,2)
- TFDIF values are computed for all these terms to show their statistical significance.

Knowledge network

- Fuse knowledge elements with the same concept and similar semantic meaning in the complete process of knowledge absorption and integration
- knowledge elements are labeled as supplied knowledge elements that are provided by the references; absorbed knowledge elements that exist in both the target paper and its references; and generated knowledge elements that exist only in the title & abstract





Figure 2: Schematic diagram of the knowledge network

2 Methodology

Descriptive statistics

- Descriptive statistical metrics are used to measure the total number of nodes, edges, and components in the knowledge network to analyze the network scale.
- We compute the knowledge absorption efficiency and the knowledge integration efficiency.

 $Absorption = V_{Absorbed \ elements} / V_{elements}$ $Integration = V_{Generated \ elements} / V_{elements}$

Global network metrics

- Network density
- Average path length
- Global clustering coefficient

Knowledge proximity

- The cosine distance of embedding-based vectors derived from scientific articles and patents can be used to quantify the proximity of knowledge (Feng, 2020)
- Measure the proximity of knowledge of the target article and its references

7

3 Empirical study



Data

- In this paper, we choose the key publications of **Nobel prize in physics** .
- Only papers published before 2004 were included. Keep papers that have three or more references, and the language of writing is limited to English.
- We finalize 124 Nobel prize papers in physics as the dataset for empirical study.

Descriptive statistics of knowledge networks of Nobel Prize-winning papers

- The 124 Nobel Prize-winning papers have an average of 31.85 supplied knowledge elements, 5.54 absorbed knowledge elements and 2.64 generated knowledge elements.
- The average knowledge absorption efficiency is 0.14, and the average knowledge integration efficiency is 0.09.



3 Empirical study



Knowledge network example

• Supplied knowledge elements are marked in deep blue, absorbed knowledge elements are highlighted in light blue and generated knowledge elements are colored in green.



Figure 3: (a) One example of a connected knowledge network WOS-000201553700001; (b) One example of an unconnected knowledge network WOS-000201591300009



3 Empirical study





Figure 4: Average node degree distribution for 124 Nobel Prize-winning papers

Figure 5: Density distribution for 124 Nobel Prizewinning papers

• Figure 5 also shows a right-skewed distribution, which means that the majority of the knowledge networks are sparse ones.



0.95

3 Empirical study



Figure 6: Average shortest path length distribution

Figure 7: Average clustering coefficient distribution

• These networks demonstrate strong small-world properties, characterized by a high clustering coefficient and a short average path length.

4 Conclusion, limitations and future work



- + Explore a new perspective on modeling knowledge absorption and integration with knowledge networks.
- + Provide heuristic research that can potentially be used to model and measure the process and result of knowledge combination
- (1) At this stage, this study has not established a control group for the experimental group to further investigate whether the indicators provided by the model can effectively reflect the effects of knowledge integration and innovation;
- (2) The number of references indirectly affects the size of the current knowledge network, and this influence needs to be minimized by further adjusting the network's nodes and edges;
- (3) There are limitations in constructing network edges solely based on term co-occurrence relationships;
- (4) More metrics need to be design to measure the efficiency and structure of knowledge absorption and integration.

THANK YOU!