



# The Use of Dissimilarity Measures for the Study of Evolution in Scientific Fields

Lukun Zheng\* and Yuhang Jiang

Department of Mathematics, Western Kentucky University, Bowling Green, USA

\*Corresponding author: Lukun Zheng, Department of Mathematics, Western Kentucky University, 1916 College Heights Blvd, Bowling Green, Kentucky, USA.

Received Date: April 23, 2021

Published Date: April 29, 2021

## Abstract

One of the key issues in evolution of scientific field is to quantify the dissimilarity between two collections of scientific publications in literature. Many existing works study the evolution based on one or two dissimilarity measures, despite the fact that there are many different dissimilarity measures. Finding the appropriate dissimilarity measures among such a collection of choices is of fundamental importance to the study of scientific evolution. In this paper, we attempt to study the use of dissimilarity measures in scientific evolution.

**Keywords:** Scientific evolution; Dissimilarity measures; Principal component analysis (PCA); Dissimilarity integration

## Introduction

The scientific theory of evolution by natural selection began with Charles Darwin's *On the Origin of Species* published in 1859 [1,2]. Evolution as a scientific theory has been used in many other disciplines as well, including medicine [3,4], psychology [5,6], anthropology [6], forensics [7], agriculture [8], and other social-cultural applications [9,10]. In this paper, we aim at studying the evolution of scientific fields by investigating their developmental trends shown in scientific literature [11].

Evolution in scientific fields has been drawing attention among researchers and scientists in recent years. One of the main issues in evolution of scientific field is to quantify the dissimilarity between two collections of scientific publications in literature. The temporal evolution of scientific research can be observed in retrospective studies in many fields. What is more, the "evolution map" of scientific fields helps us understand the nature of scientific development and the relative importance of different topics or publications [11]. Innovations and scientific breakthroughs keep on emerging, leading to new or improved technology and scientific findings, which, then, shape new developmental trends in various research areas.

One of the main issues in evolution of scientific field is to quantify the dissimilarity between two collections of scientific publications in literature. There are many dissimilarity measures encountered in many different areas such as biology, computer science, mathematics, psychology, statistics, etc. Finding the appropriate measures among such a collection of choices is of fundamental importance to pattern classification, clustering, and information retrieval problems [12-14].

## Integration of Dissimilarity Measures

Over the years, there have been many approaches for measuring scientific evolution. For instance, Vargas-Quesada et al. [15] introduced a graphic representation of the intellectual structures in the form of scientograms using scientometric information such as cocitation network from a certain scientific domain. Their approach allows one to detect patterns and tendencies of scientific evolution in a scientific domain through network visualizations. Dias et al. [16] used an information-theoretic measure of linguistic similarity to investigate the organization and evolution of scientific fields based on the 20,000 most frequent words from the abstracts of the papers considered excluding a list of stop words. Jurgens et al. [17]

proposed a method on measuring scientific evolution by studying how scientific works frame their contributions through different types of citations and how this framing affects the field. Frank et al. [18] used the Microsoft Academic Graph to study the bibliometric evolution of AI research and its related fields from 1950 to 2019. The problem with these works is that they only adopted few (mostly just one single) dissimilarity measures in their study, ignoring the fact that there are many such measures [19,20].

In recent years, many attempts on integrating different dissimilarity measures have been made and they show promising results in various areas such as image classification [21], text categorization [22], and pattern recognition [23]. Most of these works combine dissimilarity measures using a weighted sum with weights determined using different algorithms such as trial-and-error method, and cross validation. In the study of scientific evolution, Zheng & Jiang [24] proposed a novel approach for the integration of twelve dissimilarity measures based on keywords distributions in scientific fields using principal component analysis (PCA). They collected a collection of bibliographic records of articles from four selected scientific fields published from 1991 to 2019 and obtain the yearly keyword distributions and calculated the values of twelve dissimilarity measures between the keyword distributions for each pair of successive years. Then PCA is utilized to combine these dissimilarity measures. Their results show a decreasing trend for the evolution between two successive years in all chosen fields during the time 1991-2019 [25,26].

## Conclusion

Most of the studies on scientific evolution have been limited to the use of single measures. Considering the successful applications of integrating dissimilarity measures, it would be a good idea to study the use of different integration techniques (e.g., ensemble methods) on scientific evolution. We believe that more efforts are needed to systematically study various properties of different dissimilarity measures in the study of scientific evolution. Future works are also needed to explore and compare the advantages and limitations of different integration approaches in this area.

## Acknowledgement

None.

## Conflict of Interest

No conflict of interest.

## References

- Darwin C, Bynum WF (2009) *The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life*. New York: AL Burt, pp. 458.
- Nesse RM (2008) The importance of evolution for medicine. *Evolutionary Medicine*, pp. 416-432.
- Grunspan DZ, Nesse RM, Barnes ME, Brownell SE (2018) Core principles of evolutionary medicine: a Delphi study. *Evolution, medicine, and public health* 1: 13-23.
- Panksepp J, Panksepp JB (2000) The seven sins of evolutionary psychology. *Evolution and cognition* 6(2): 108-131.
- Langs R (2019) *The evolution of the emotion-processing mind*. Routledge.
- Tang S (2020) *On Social Evolution: Phenomenon and Paradigm*. Routledge.
- Khan MZ, Mishra A, Khan MH (2020) Cyber Forensics Evolution and Its Goals. In *Critical Concepts, Standards, and Techniques in Cyber Forensics*. IGI Global, pp. 16-30.
- Fordham M (2019) *Britain's Trade and Agriculture: Their Recent Evolution and Future Development*. Routledge.
- Giavazzi F, Petkov I, Schiantarelli F (2019) Culture: Persistence and evolution. *Journal of Economic Growth* 24(2): 117-154.
- Donald M (1993) *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.
- Tang X, Yang C, Song M (2013) Understanding the evolution of multiple scientific research domains using a content and network approach. *Journal of the American Society for Information Science and Technology* 64(5): 1065-1075.
- Duda RO, Hart PE, Stork DG (2012) *Pattern classification*. John Wiley & Sons.
- Wei G (2018) Some similarity measures for picture fuzzy sets and their applications. *Iranian Journal of Fuzzy Systems* 15(1): 77-89.
- Jain G, Mahara T, Tripathi KN (2020) A survey of similarity measures for collaborative filtering-based recommender system. In *Soft computing: theories and applications*. Springer, Singapore, pp. 343-352.
- Vargas-Quesada B, de Moya-Anegón F, Chinchilla-Rodríguez Z, González-Molina A (2010) Showing the essential science structure of a scientific domain and its evolution. *Information Visualization* 9(4): 288-300.
- Dias L, Gerlach M, Scharloth J, Altmann EG (2018) Using text analysis to quantify the similarity and evolution of scientific disciplines. *Royal Society open science* 5(1): 171545.
- Jurgens D, Kumar S, Hoover R, McFarland D, Jurafsky D (2018) Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics* 6: 391-406.
- Frank MR, Wang D, Cebrian M, Rahwan I (2019) The evolution of citation graphs in artificial intelligence research. *Nature Machine Intelligence* 1(2): 79-85.
- Cha SH (2007) Comprehensive survey on distance/similarity measures between probability density functions. *City* 1(2): 1.
- Ibba A, Duin RP, Lee WJ (2010) A study on combining sets of differently measured dissimilarities. In *2010 20th International Conference on Pattern Recognition*. IEEE, pp. 3360-3363.
- Liu C, Wang J, Duan S, Xu Y (2019) Combining dissimilarity measures for image classification. *Pattern Recognition Letters* 128: 536-543.
- Pinheiro RH, Cavalcanti GD, Tsang R (2017) Combining dissimilarity spaces for text categorization. *Information Sciences* 406: 87-101.
- Zhang B, Srihari SN (2003) Binary vector dissimilarity measures for handwriting identification. In *Document Recognition and Retrieval X*. International Society for Optics and Photonics 5010: 28-38.
- Zheng L, Jiang Y (2021) Combining dissimilarity measure for the study of evolution in scientific fields. *DeepAI*.
- Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemometrics and intelligent laboratory systems* 2(1-3): 37-52.
- Zheng H, Zheng L (2020) An Investigation on Language Programs in US Higher Institutions-A Case Study on Chinese Language Programs. *US-China Education Review* 10(6): 257-265.