

Original Article

Mapping the Intellectual Structure of Epidemiology with Use of Co-word Analysis

Hamed Baziyad^{1,*}, Saeed Shirazi¹, Seyedmohammadreza Hosseini², Rasoul Norouzi¹¹Department of Information Technology, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.²Department of Earth and Space Science and Engineering, Lassonde School of Engineering, York University, Toronto, Canada.

ARTICLE INFO

Received 21.04.2019
 Revised 23.05.2019
 Accepted 19.06.2019
 Published 01.08.2019

Key words:

Intellectual structure of epidemiology;
 Co-word analysis;
 Text mining;
 Graph mining;
 Social network analysis

ABSTRACT

Background and Aim: The existence of an intellectual structure for every field is essential for managers and scholars. Intellectual structures provide a comprehensive map of knowledge that can guide researchers and managers to have a better view of their fields. Besides, with high-speed and massive amounts of data and information generation, reading and surveying of all resources are severely tricky. Intellectual maps solve this problem and make a situation for control and monitoring this voluminous and high-speed generated data. Epidemiology is regarded as one of the exciting fields which many researchers focused on it. A study of the structure and criteria of different epidemiological fields has not been done yet. Indeed, there is no serious effort for knowledge discovery of hidden information on epidemiological texts.

Methods: In this paper, in order to survey this field, an intellectual structure is provided using co-word analysis. Utilizing co-word analysis discloses relationships and structure among research subjects and topics in a field.

Results: Finally, four main clusters were determined, namely: genetic (with 30.53% of surveyed papers), illness (29.47%), modeling (23.16%), and prevention (16.84%).

Conclusion: According to epidemiology co-word network, epidemiology area has not been studied from enough different areas, especially from novel technologies.

Introduction

Knowledge of population-based disease aims at identifying priority health programmers for prevention and care using limited resources to the best possible effect. Epidemiological researches can provide quantitative measurements of the contribution to disease causation of different environmental variables. Epidemiology can be defined as an application of studying the distribution and determinants of health- states or events of specified populations, in the prevention

and control of health problems (1). Although epidemiology traces back to Hippocrates' observation more than 2000 years ago that environmental factors influence the occurrence of disease, and in the nineteenth century, the distribution of disease in specific human population groups was measured to any significant extent (2). In the modern form of epidemiology, it is known as a relatively new discipline and aims to study diseases in human populations to inform prevention and control effects by applying quantitative methods (3). In

* . Corresponding Author: hamed_baziyad@modares.ac.ir

some specialist fields, such as environmental and occupational epidemiology, the emphasis is on the study of populations with particular types of environmental exposure. The existing literature in a field commonly provides an excellent scientific research base for researchers to develop new frameworks and hypotheses (4). Diversity and variety of research topics in different scientific fields have made it necessary for researchers to identify the areas and domains of the current research.

Co-word analysis enables us to illustrate an overall picture of contents on a specific domain through counting and analyzing the co-occurrences of words in bibliographical units such as journal and conference papers, research reports, book chapters, and the like (5). Co-word analysis is widely used in bibliometric and scientometric studies, whose theoretical roots go back to the actor-network theory (6–8). This technique is based upon the assumption that a group of aggregated keywords could provide an adequate description of papers' content; thus, underlying themes in a research field can be characterized by a list of the most critical keywords (9). The more overlapping pair of keywords within the analyzed papers, the higher the probability of a relationship among them (7). Many researchers have used the co-word analysis to explore the intellectual structure of different subjects and disciplines.

Munoz-Leiva et al. applied a co-word analysis methodology and provided an intellectual map for Integrated Marketing Communications (IMC) research during 1991-2012 (10). Finally, four main clusters have appeared, namely integrated marketing communications, theoretical aspects, public relations and IMC, theoretical issues, and criticisms of IMC as management fashion. Xie applied co-word and co-citation analysis on anticancer drug papers of the WOS database during 2000-2014 (11). Five revealed clusters of the co-word network are chemotherapy drugs, drug delivery, bio screening, drug resistance

research, and enzyme inhibitor studies. Nguyen presented a systematic mapping of biomedical therapeutic modalities using by applying a large-scale co-word analysis between 1987 and 2017 (8). Supply Chain Management (SCM) was another area that was studied with co-word analysis methodology. Finally, four main clusters, namely: sustainable supply chain management, strategic competition, the value of information, and development of supply chain management, were found through co-word analysis (12). Scientometrics (13), the Internet of Things (14), Bioelectronics (15), and Marketing (16), were other topics that were studied based on co-word analysis methodology.

There has been no research conducted to identify the hidden information such as the relationship and structure among research subjects and topics in the field of epidemiology. In this study, we aim to map the knowledge domains of epidemiology by the co-word analysis. This paper will be helpful for the researchers to quickly understanding the current circumstance and providing a basis for its future development. The remaining parts of this paper are structured as follows: Section 2, we introduce the constructing process of the co-word network and essential network theories. Section 3 presents the analytical results, elaborates on the underlying discussions and, proposes some valuable implications for further academic studies or policy formulation.

Methodology

Data Collection

In order to analyze the epidemiology field in academic researches, the co-word analysis was employed. The first 400 papers related to epidemiology were collected from Web of Science (WOS) from 5 last years. After that, the data was cleaned and preprocessed with Python 3.6.

Standardizing Keywords

In order to preprocess, keywords with the same meaning were replaced, such as standards and standards. Also, special characters such as parentheses, commas, dots, etc. were removed. After cleaning the dataset, the co-word matrix was built.

Creating Co-occurrence Matrix

Keywords that existed in the same papers are considered as co-occurrence words and were indicated as linked nodes in a graph. For example, if word A and word B were in the same paper, a linkage was drawn between nodes A and B. Finally, a graph was drawn. Each edge in the graph refers to the same co-occurrence of two linked nodes (keywords).

Columns and rows of co-occurrence matrix indicate papers' keywords, and elements of matrix show co-occurrence of the keywords in the same papers. For example, $A_{1,2}$ refers to the frequency of occurrence of both keyword1 and keyword2 in the same papers. If $A_{1,2} = 3$, it means that there are three papers in which both keyword one and keyword two are in them jointly. The main diagonal of the matrix shows the frequency of keywords.

Data Visualization

In order to find the main clusters of epidemiology, a community detection algorithm was employed, and the graph was visualized in Gephi 0.9.2. The employed algorithm of community detection was presented by Lambiotte et al. in 2009. Each community was shown separately, and finally, all of them were

shown in a standard image. Gained modularity was 0.327, which refers to the high quality of clusters. According to Xie and Szymanski, modularity between 0.3 and 0.7 indicates a strong cluster structure (17).

Results

The graph has 94 nodes (keywords) and 740 edges. After executing the community detection algorithm on the graph, three communities were determined. The red community with 30.53 nodes is the largest community that focuses on genetic issues. Keywords of this cluster were studied based on genetical points of view. The dark blue cluster was related to innless, the light green community describes the modeling of epidemiology, and finally, the orange cluster discussed the prevention of epidemiology.

Cluster 1: Genetics

Some keywords, such as genotype, phylogenetic analysis, genetic diversity, and genotyping, directly focused on genetic issues. Other keywords indirectly indicated the genetic field. For instance, diversity refers to genetic diversity, and strains refer to the talent of people who come from their genes. Emergence indicated to emergence and developments of genetics. Also, identification refers to gene identification.

Cluster 2: Illness

'Illness community' describes the exposures and threats of epidemiology and its probabilistic hazards. Some keywords like severity, exposure injury indicated problems of epidemiology.

to genetic-based epidemiology mostly have been studied in these two countries. Iran, Brazil, Pakistan, and Nigeria lay in the illness community. Epidemiology of illness is more studied in these countries. Ethiopia and Island have focused on epidemiology modeling, and Cameroon lay in reasons community. Besides, more frequent keywords of 'epidemiology' researches are strains, models, experience, resistance, infants, severity, vaccination, trends, outbreak, and surgery.

Discussion

This paper provided an intellectual structure of epidemiology researchers based on 400 Web of Science's papers from five last years. Co-word analysis was utilized, and community detection was applied to studied keywords. Finally, four main clusters were determined, namely, genetics, illness, modeling, and prevention. Genetic with 30.53% of surveyed papers had the most volume of researches. Illnesses with 29.47%, modeling with 23.16%, prevention with 16.84% were lied in other positions, respectively.

The 'prevention' field seems to have received less attention, and it can be regarded as one of the fields for more attempts and innovation. This section can be combined with the 'modeling' category for more achievements in future works. For example, mathematical modeling can be used for simulation of epidemiology and facilitating the prevention of illnesses process.

Results show that there is almost no trace of new technologies in epidemiology papers. Study on some novel technologies such as the Internet of Things (IoT), big data, cloud computing, and Artificial Intelligence (AI) can increasingly revolutionize the epidemiology area.

Conclusion

Utilizing co-word analysis based on publications of the Web of Science (WoS) database service, disclosed research topics of epidemiology area. Besides, relationships among found topics were shown within an intellectual structure. The results indicated that epidemiology includes an interdisciplinary diversity of knowledge fields. The intellectual map of this paper shows four main clusters, namely a) genetics, b) illness, c) modeling, and d) prevention. This paper enables researchers to understand the current status of epidemiology and its evolution. Besides, this study can be utilized as a roadmap for other applications in the field of epidemiology.

References

1. Last JM, Spasoff RA, Harris SS. A Dictionary of Epidemiology, Fourth Edition. 2001;154(1).
2. R. Bonita, R. Beaglehole TK. Basic Epidemiology. World Health Organization; 2006. 213 p.
3. Beaglehole R BR. Public Health at the Crossroads: Achievements and Prospects. 2nd Editio. Cambridge University Press; 2004. 318 p.
4. Hoz-correa A De, Mu F. Past themes and future trends in medical tourism research: A co-word analysis. 2018;65:200–11.
5. Wang GLJHH. A co-word analysis of digital library field in China. 2012;(August 2011):203–17.
6. Yang Y, Wu M, Cui L. on co-word analysis. 2012;659–73.
7. Piedad M, Igami Z, Mugnaini R. A new model to identify the productivity of theses in terms of articles using co-word analysis. 2014;(October).
8. Nguyen D. Social Science & Medicine Mapping knowledge domains of non-biomedical modalities: A large-scale co-word analysis of literature 1987 – 2017. Soc Sci Med [Internet]. 2019;233(May):1–12. Available from: <https://doi.org/10.1016/j.socscimed.2019.05.044>

9. Limoges C, Courtial JP, Laville F, Nationale E. HISTORICAL SCIENTOMETRICS? MAPPING OVER 70 YEARS OF BIOLOGICAL SAFETY RESEARCH WITH CO-WORD ANALYSIS 1. 1993;27(2):119–43.
10. Muñoz-leiva F, Porcu L, Barrio-garcía S, Porcu L, Barrio-garcía S. Discovering prominent themes in integrated marketing communication research from 1991 to 2012: a co-word analytic approach. 2016;0487(January).
11. Xie P. word and document co-citation visualization analysis. *Scientometrics*. 2015;
12. Yogesh WS, Tsai KDC. Supply chain management: exploring the intellectual structure. *Scientometrics*. 2015;
13. Ravikumar S, Agrahari A, Singh SN. Mapping the intellectual structure of scientometrics : 2014;
14. Yan B, Lee T, Lee T. of Things (IoT) field (2000 – 2014): a co-word analysis. *Scientometrics*. 2015;
15. Hinze S. BIBLIOGRAPHICAL CARTOGRAPHY OF AN EMERGING. 1994;29(3):353–76.
16. Wang Z, Zhao H, Wang Y. Social networks in marketing research 2001–2014: a co-word analysis. *Scientometrics*. 2015;
17. Xie J, Szymanski BK. Community Detection Using A Neighborhood Strength Driven Label Propagation Algorithm. 2011;(978):188–95.