Using WoS Keywords to Analyze Topic of Articles Collection with Different Interdisciplinary Degree

Tiantian Tong; Chengzhi Zhang; Lin Zhang

Department of Economics & Management, Nanjing University of Science and Technology, Nanjing, China, 210094
Department of Information Management, Wuhan University, Wuhan, China, 430000

ABSTRACT

Interdisciplinary scientific research (IDR) tends to be a seedbed for breakthroughs in science and technology. Topic analysis in IDR can provide a better understanding of knowledge flow among disciplines. Keywords Plus terms annotated by Web of Science (WoS) can capture articles' content and summarize their topics efficiently. In this study, we propose a two-step model to investigate the differences of IDR topics with interdisciplinary degree by WoS keywords. Our study can be applied to understand research topics of IDR and provide suggestions to interdisciplinary researchers.

INTRODUCTION

Understanding what knowledge has exchanged between disciplines and how research topics change over time can help scholars better understand knowledge structure of scientific fields and development of new disciplines. Keywords Plus terms provided by WoS, which capture articles' content with greater depth and variety, make it possible to provide systematic comparison across various disciplines. In this paper, we focus on articles covering various disciplines published in PLOS One, and investigate research hotpot based on Keywords Plus terms. We examine the distribution of keywords in research articles with different interdisciplinary degree. Our goal is to identify changes of IDR topics with time and interdisciplinary degree.
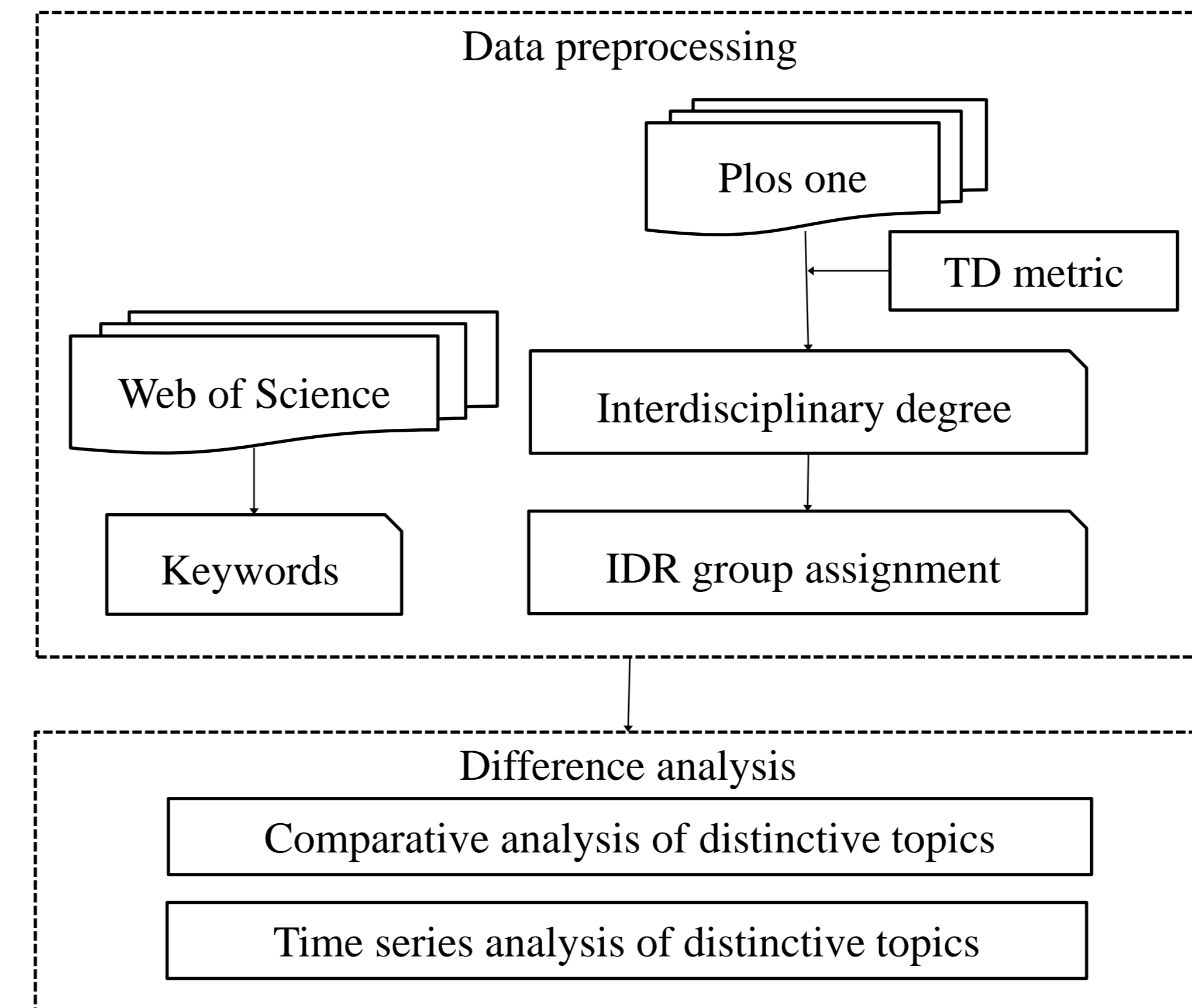
What's the status now?

METHOD

Data Preprocessing

➤Data

143, 910 articles from 2007 to 2015 - PLOS One
Keywords of articles –Web of Science.

➤Interdisciplinarity of each article

$$^2D^S = 1/(\sum_{i,j=1}^{N} (1 - d_{ij})p_i p_j)$$

➤IDR group assignment

Articles are divided into three groups according to interdisciplinarity: H-IDR group (the top 10%), M-IDR group (80%), L-IDR group (the last 10%).

Difference analysis

➤ Comparative analysis of distinctive topics

CHI feature selection method is used to select the distinctive keywords.

➤Time series analysis of distinctive topics

K-means is chosen to evaluate differences between frequency patterns of individual keywords.

Data preprocessing

Plos one
TD metric
Web of Science
Interdisciplinary degree
Keywords
IDR group assignment

Difference analysis
Comparative analysis of distinctive topics
Time series analysis of distinctive topics

RESULT

Group	Top 10 Keywords in Chi-square Score
H-IDR	Acute lung injury; Drug-delivery; *Oxidative stress*; Rna-seq; Heart-rate-variability; Metabolomics; *Gene-expression data*; *Nanoparticles;Set enrichment analysis; Regulatory networks*
M-IDR	optical coherence tomography; *Myoblast fusion*; Cystic-fibrosis; *ca2+-activated cl-channels*; Filtered back-projection; Phase-iii trial; Dendritic cells; Acute lung injury; *Messenger-Rna decay; Rna-seq;*
L-IDR	Ligand cd55;optical coherence tomography; *Ca2+-activated cl-channels*; Phase-iii trial; *receptor cd97*; Open-angle glaucoma; Filtered back-projection; Computed-tomography; Long-term potentiation; Radiotherapy;

Table 1. Top 10 keywords obtained by Chi-square across three groups.

Category	Top 10 keywords for each category
A	Cells; Gene-expression; Acute lung injury; Air-pollution exposure; Amorphous calcium-carbonate; Atomic-force microscopy; Drug-delivery; Electronic medical-records; Expression data; Female physical attractiveness;
B	Cystic-fibrosis; Dynamics; Heart-rate-variability; Inflammation; Ischemia-reperfusion injury; Mass-spectrometry; Maximum-likelihood; Respiratory-distress-syndrome; Complex networks; Human-immunodeficiency-virus;
C	Arabidopsis-thaliana; Care; Cell lung-cancer; Computed-tomography; Developing-countries; Mellitus; Open-angle glaucoma; Optical coherence tomography; Randomized controlled-trials; Risk-factors;
D	Abscisic-acid; Basal ganglia; Beta-cell function; Event-related fmri; Long-term depression; Long-term potentiation; Messenger-rna decay;Posterior parietal cortex; Transcranial magnetic stimulation; Visual-cortex

Table 2. Top 10 keywords for each category

CONCLUSION

➤Scientific articles' topics based on Keywords Plus terms across a multitude of fields can help to expand the understanding of interdisciplinary research.

➤The group of H-IDR articles tends to cover more fields and include more general keywords than the lower-IDR articles.

➤Time series analysis provides some insight into a series of groups identified by a cluster analysis.

➤Additional metrics might be considered to measure the interdisciplinarity, such as indicators based on citation.

REFERENCE

Macqueen, J. (1967).Some methods of classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, USA, 1967: 281-297.

Xu, H et al. (2014). Measurement, Visualization and Application of Interdisciplinary Research. Library & Information Service, 8 (3), 21-27

Yang, Y et al. (1997). Feature selection in statistical learning of text categorization. Planta, 230(230), 671-685

Zhang, J et al. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. Journal of the Association for Information Science & Technology, 67(4), 967-972

Zhang, L et al. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. Journal of the Association for Information Science & Technology, 67(5), 111-112

CONTACT

Tiantian Tong: tongtiantiannjau@163.com
Chengzhi Zhang : zhangcz@njust.edu.cn
Lin Zhang: zhanglin_1117@126.com