

# Citespace-based Recent Status of Information Systems Science and Technology Papers of China and Others

Haoyuan Lan\*, Hui Du

Beijing Jiaotong University, China

\*Corresponding author. Email: 17711011@bjtu.edu.cn

## ABSTRACT

To comprehensively and deeply understand recent status of information systems science and technology papers of China and others, following the general process of mapping knowledge domains analysis, firstly, journal papers on information systems from CSCD and SCI in the last five years were searched and cleaned respectively. Secondly, with the help of the analysis software Citespace, mapping knowledge domains were drawn, interpreted and compared. In addition to recent research status of China and other countries, especially useful reference for Chinese research was provided.

**Keywords:** Mapping knowledge domains, Information systems, Citespace, Science and technology, Recent status analysis.

## 基于Citespace的中外信息系统科技论文近况分析

蓝浩源\*, 杜晖

北京交通大学, 北京, 中国

\*通讯作者. 邮箱: 17711011@bjtu.edu.cn

## 中文摘要

为全面深入了解中外信息系统科技论文近况, 遵循科学知识图谱分析的一般流程, 检索和清洗了来自CSCD和SCI中近五年的信息系统期刊论文。借助分析软件Citespace绘制了图谱并进行了解读和中外比较。提供中外研究近况的同时, 特别为中国研究提供了有益参考。

**关键词:** 科学知识图谱; 信息系统; Citespace; 科技; 近况分析

## 1. 引言

科学知识图谱能够可视化地描述人类随时间拥有的知识资源及其载体[1]。Citespace是当前应用最为广泛的科学知识图谱分析软件。苏玥竹[2]利用从WoS (Web of Science) 搜集的2003至2012年美国的信息系统文献, 通过关键词、研究机构、作者和文献的中心度分析, 发现对未来研究具有指导作用的十篇最主要文献以及美国的研究热点。李娜娜[3]通过分析从CNKI (China National Knowledge Infrastructure) 提取的18种GIS (Geographic Information Systems) 期刊论文图谱,

发现作者在2011至2015年更趋向合作并在一些热点研究方面趋近成熟。Zongfeng Zou等[4]借助从SCI和SSCI收集的876篇商业智能文献的关键词聚类分析, 发现1997至2017年的研究热点是云计算。吴冰等[5]借助《信息系统研究》(ISR)期刊的747篇论文, 根据合作网络的中心性、聚类特点发现1998至2001年信息系统的研究热点是企业应用的信息集成, 2002至2011年的研究热点从传统组织转向虚拟组织、信任及其影响因素, 2012至2017年的研究热点从电子口碑转向网络知识社区的知识传播。但是, 上述研究所用文献均不能反映近况, 同时, 也没有进行中外比较。

本文结构如下：第2节介绍了数据的准备情况。第3节分别从作者-机构共现，关键词聚类，以及关键词时区方面进行了中外近5年数据的分析。最后，在第4节总结了全文并指出了存在的不足。

## 2. 数据准备

CSCD (Chinese Science Citation Database) 论文以Refwork格式从CNKI导出，检索条件为“篇名”或“关键词”或“摘要”包含“管理信息系统”，发文时间为2016至2021年，文献类型为“期刊论文”。SCI论文以纯文本格式从WoS中导出，包含所有记录来源，选取“核心数据库合集”，检索条件为“篇名”或“关键词”或“摘要”包含“Management Information System”或“Management Information Systems”，发文时间也为2016至2021年，文献类型为“Article”。经过清洗（去重以及去除征稿启事）得到CSCD论文182篇，SCI论文302篇。

将清洗后的数据以download\_x.txt形式重命名并创建项目文件夹后，导入Citespace进行标准化（CSCD论文选择CNKI Format，SCI论文选择Remove Duplicates (WoS)），然后选择关联强度最强的Cosine系数和TopN，

选择Pathfinder和Pruning sliced networks[6]相结合的简化方式。

## 3. 数据分析

### 3.1. 作者-机构共现分析

1. 根据CSCD论文得到的共现网络如图1所示，可以看出：存在发文量较多的机构，前三名分别是中国疾病预防控制中心妇幼保健中心、广东省疾病预防控制中心传染病预防控制所和中国疾病预防控制中心结核病预防控制中心；也存在发文量较多的作者，前五名分别是乔亚萍、王前、王潇滢、王爱玲和金曦。虽然中国疾病预防控制中心结核病预防控制中心拥有多位作者，但却没有发文量较多的作者，发文量较多的作者集中在中国疾病预防控制中心妇幼保健中心，该中心拥有发文量最多的作者乔亚萍，且与同一机构的王前等作者有合作关系，该作者还属于多个机构，并且与隶属不同机构的彭竟存在合作关系。多数作者和机构间存在合作关系；少数作者只隶属于一个机构。



图1 CSCD论文作者-机构共现网络

2. 根据SCI论文得到的共现网络如图2所示，可以看出：存在发文量较多的机构，前两名分别是Minist Hlth和London Sch Hyg & Trop Med；不存在发文量较多的作者。以上两家机构的发文作者非常多，但却没有发文量较多的作者。LISA R HIRSCHHORN等作者隶属于多个机构，也与其他多个作者存在合作关系。多数作者和机构间存在合作关系；多数作者隶属于一个机构。

3. 将图1和图2进行比较发现：中外都存在发文量较多的机构，存在机构间及同一机构内和不同机构间的作者合作，一位作者可以隶属于多个机构。但是，国内存在发文量较多的作者，并且这些作者隶属于同一机构。

### 3.2. 关键词聚类分析

1. 根据CSCD论文得到的关键词聚类网络如图3所示，可以看出：按关键词数量从多到少的顺序排列共有9个聚类，代表信息系统研究的9个主要方向。

2. 根据SCI论文得到的关键词聚类网络如图4所示，可以看出：按关键词数量从多到少的顺序排列共有13个聚类，代表信息系统研究的13个主要方向。

3. 将图3和图4进行比较发现：中外不存在相同聚类。中国的聚类数量少于外国，且主要偏向医疗，而外国的聚类则涉及更多主题。



图2 SCI论文作者-机构共现网络

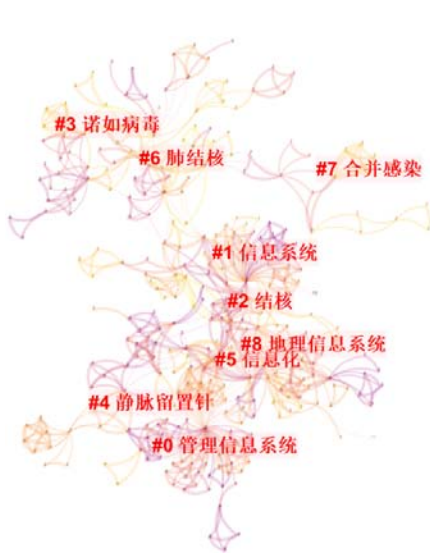


图3 CSCD论文关键词聚类网络

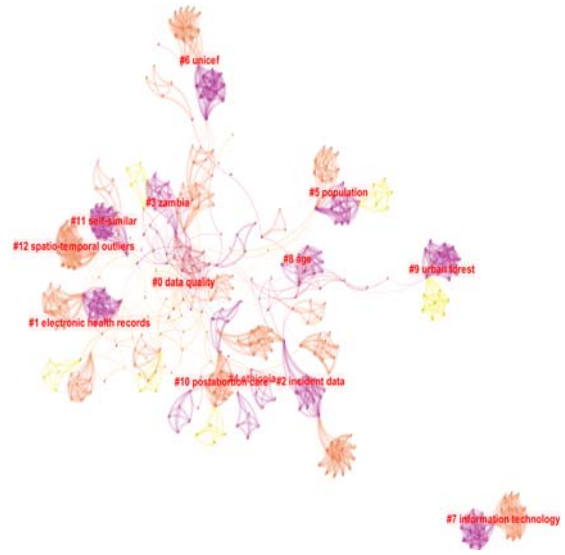


图4 SCI论文关键词聚类网络



图5 CSCD论文关键词时区网络

### 3.3. 关键词时区分析

1. 根据CSCD论文得到的关键词时区网络如图5所示, 可以看出: 除2021年由于没有采集到全年数据外, 其余每年的关键词数量没有明显变化。同时, 虽然有新主题的出现, 但多为既有医疗主题的延续。

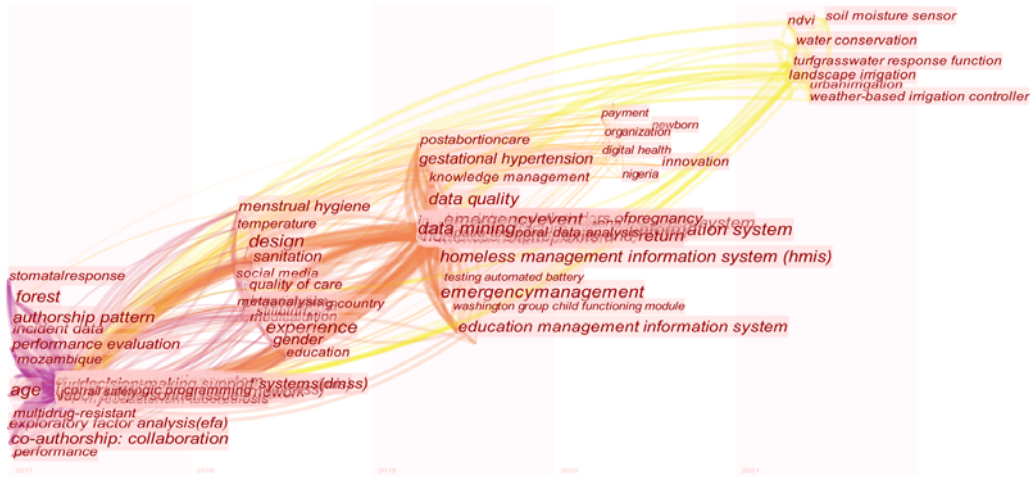


图6 SCI论文关键词时区网络

2. 根据SCI论文得到的关键词时区网络如图6所示, 可以看出: 每年的关键词数量呈多少交替出现特征, 且总体上有减少趋势。2017和2019年的关键词数量较多, 2018和2020年的关键词数量较少, 特别是2020年, 只出现6个关键词。虽然2021年尚未结束, 但其关键词数量已多于2020年。在新主题出现的同时, 既有医疗主题也在延续。

3. 将图5和图6进行对比发现: 中国的关键词数量总体多于国外, 且每年的数量比较稳定。两者在延续既有医疗主题的同时, 都在向其他主题扩展, 但国外的扩展主题更加广泛。

### 4. 结论

遵循科学知识图谱分析的一般流程, 检索和清洗了来自CSCD和SCI中近五年的信息系统期刊论文。借助分析软件Citespace绘制了图谱并进行了解读和中外比较。

由于信息系统涉及主题众多, 文献形式多样, 因此, 本文用于分析的数据并不全面。另外, 由于专业知识和能力的限制, 对图谱的解读和对比也较肤浅。

### REFERENCES

- [1] Yue Chen, Yuanze Liu, Jin Chen and Jianhua Hou (2008), History and theory of mapping knowledge domains, *Studies in Science of Science* (03), 449-460. DOI: 10.16192/j.cnki.1003-2053.2008.03.025.
- [2] Yuezhu Su (2014), Research Status – quo Analysis of IS in the last ten Years of U.S. based on Mapping Knowledge, *Journal of Modern Information* (08), 154-159. DOI: CNKI: SUN: XDQB.0.2014-08-031.
- [3] Nana Li (2018), Analysis on the Research Status of Geographic Information System in China in Recent Five Years: Bibliometric Analysis Based on Citespace, *Geomatics & Spatial Information Technology* (07), 168-171+175+179. DOI: CNKI: SUN: DBCH.0.2018- 07-047.
- [4] Zongfeng Zou, Jie Cheng, Kugan Huang and Yonghan Gui, Research on the Developments of Business Intelligence and its Enlightenment based on Bibliometric Statistics and Knowledge Map Analysis, *Journal of Physics: Conference Series*, Volume 1176, Issue 4, 2019.
- [5] Bing Wu and Peng Li (2018), Bibliometrics Analysis of Information Systems Research during 1998~2017—Based on Citespace Viewable Spectrum, *Advances in Social Sciences*, Volume 7, Issue 9. 2018. PP 1521-1530. DOI: 10.12677/ASS.2018.79225.
- [6] Chen, C. and Morris, S (2003), Visualizing evolving networks: Minimum spanning trees versus Pathfinder networks, *Proceedings of IEEE Symposium on Information Visualization*, (Seattle, Washington, 2003), IEEE Computer Society Press, 67-74.