



中國農業大學  
China Agricultural University

# 基于文献计量的学术评价与实践

周 群

中国农业大学

2017年10月 南京



# 主要内容

---



中國農業大學  
China Agricultural University

- 评价流程
- 实践与探索
- 操作案例

标准化

上达官听

数据来源

目的

清洗

计算方法

当我们谈评价的时候

评价指标

SCI

参照标准

我们在谈什么？

可视化

下为民用

文献集合

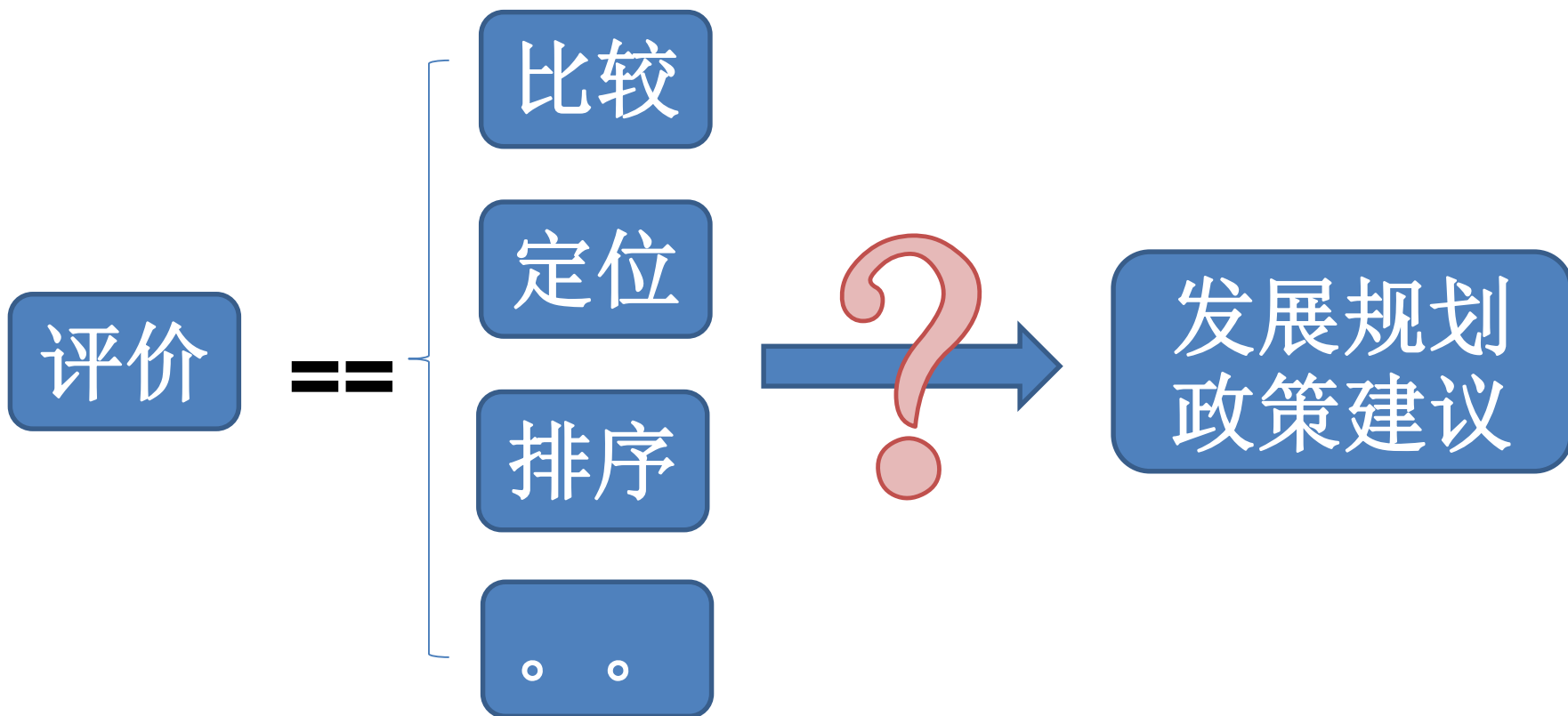
分析工具

学科分类体系

发展规划

科技创新力

学术影响力



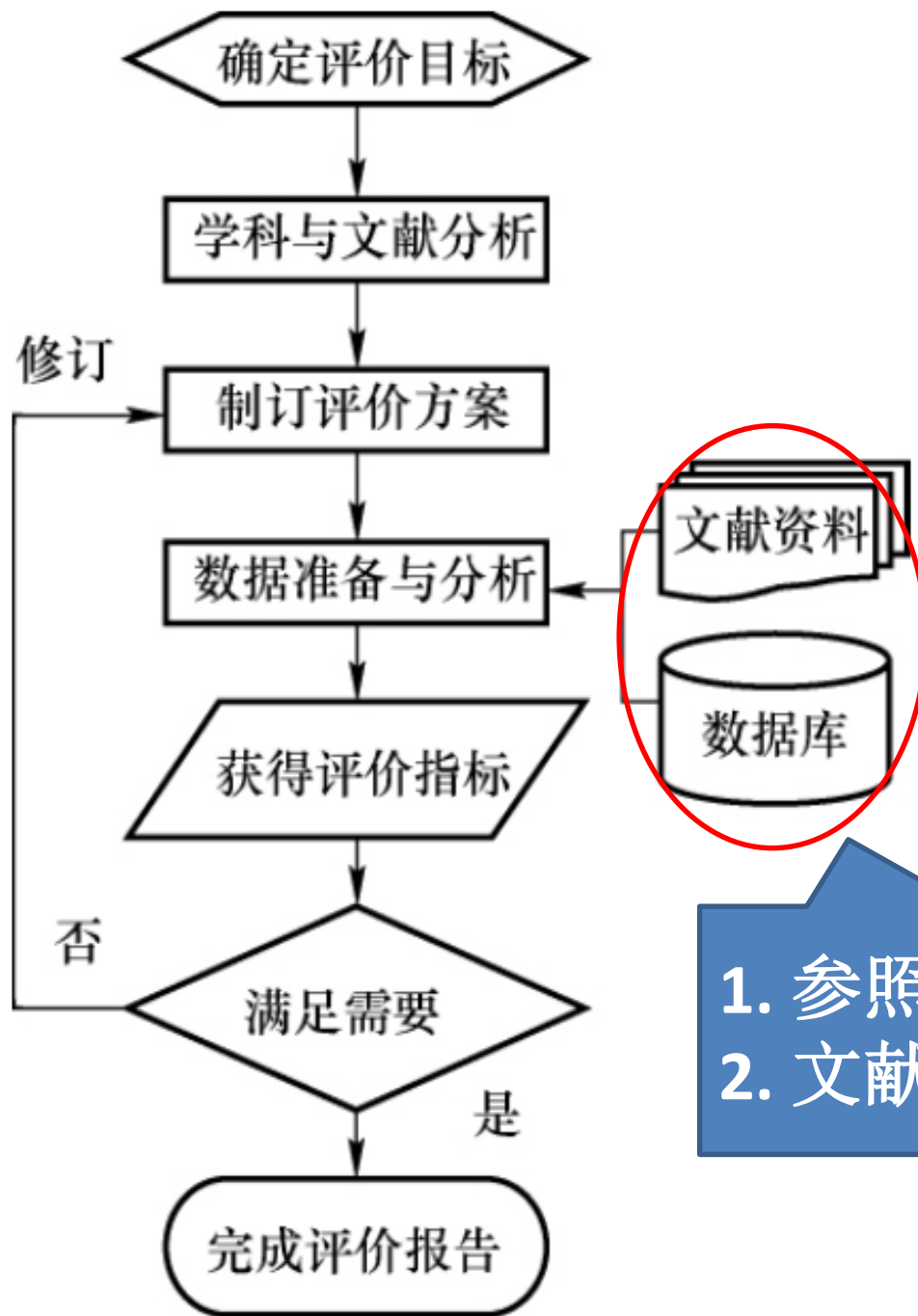
# 评价需求

---



中國農業大學  
China Agricultural University

- 对研究绩效进行评估和标杆管理
- 展示研究实力并激励高水平研究成果
- 重点学科/优势学科定位与发展
- 精确定位研究人才
- 鼓励并倡导合作研究



1. 参照标准的选择
2. 文献的归类



# 评价流程

## ● 目标

➤ 研究实体（国家、机构、学科、团队、个人...）

文献集合

## ● 评价指标

一级指标	二级指标
科研产量	论文总数
学术影响力	被引次数/高被引论文数
创新力	专利数/热点论文
发展潜力	高被引论文比例

衍生指标
H型指数
IF
CNCI
MNCS
Topx%...



# 数据准备与分析

## ● 引文分析数据库

- Web of Science----ESI/InCites
- Scopus----SciVal
- Google Scholar

## ● 数据下载和清洗

- 作者（第一或通讯）、机构和WOS等字段
- Excel（批量处理数据）、BibExcel等其他分析工具

付佳. InCites和Spotlight在学科服务中的作用比较研究[J]. 图书馆杂志, 2014, 33 (03) :37-42.

周静. 基于SciVal Spotlight的学科评估规划研究——以南开大学学科研究为例[J]. 图书馆工作与研究, 2014, (03) :105-107.

徐志玮. SciVal Spotlight评价指标原理以及同ESI的比较研究[J]. 图书情报工作, 2012, 56 (14) :86-92.





# 参照标准

- ≈ 学科分类体系
  - JCR学科分类体系
  - ESI 22个学科
  - 一级学科 (110/77)
  - Scopus
  - NSF、UoAs...
  - MeSH terms、Chemical Abstracts
- Related Records

# InCites 指标



中國農業大學  
China Agricultural University

论文产出和引文影响力

Web of Science文献量

被引频次

引文影响力

被引文献所占百分比

H 指数

规范化指标

百分位和平均百分位

学科规范化引文影响力

学科期望引文影响力

高被引论文

热点论文百分比

期刊规范化引文影响力

期刊期望引文影响力

相对世界平均水平影响力

高水平论文

被引次数排名前1%的论文百分比

被引次数排名前10%论文百分比

学科规范化引文影响力

高被引论文

合作指标

国际合作论文所占百分比

国际合作论文量

横向合作论文百分比



# 评价指标

- H型指数

- 被引频次标准化方法

- 被引端标准化 (cited-side normalization)

- 基于篇均被引频次

- MNCS(CWTS, 2011)、CNCI(InCites)、FSS(Abramo,2015)

- 基于百分位数----TopX%(Bornnman, 2013)

- 基于篇均参考文献数----CSNCR (Bornnman, 2016)

- 施引端标准化(citing-side normalization)

- 活跃参考文献(active references) (Zitt, 2008)

- SNCI(Waltman, 2013)

Mingers, J., A review of theory and practice in scientometrics. European Journal of Operational Research, 2015

周群. 论文被引频次标准化方法研究进展. 情报科学, 2017

周群. 基于百分位数的文献计量指标研究综述. 现代图书情报技术, 2013

# 基于篇均被引频次



## ● 相对影响指标（MNCS、CNCI）

➤ 定义：论文被引频次/期望被引频次

➤ 期望被引频次：指**参照标准**中所有论文的平均被引频次。

	论文A	全球同行论文
发表时间	2005	2005
文献类型	Article	Article
所属学科	经济学	经济学
被引频次	15	12.14（篇均被引频次）

➤ 如果相对影响指标的值大于1，则说明该论文的研究水平高于该学科领域的**平均研究水平**；反之则低于该学科领域的平均研究水平。



# 基于百分位

## ● 百分位等级分数 (PRS)

- X为从1到k的离散随机变量（在这里用于表示被引频次的百分位分区，比如美国NSF的百分位各区段的X变量为1到6，如果论文被引频次处于top-1%，那么X就为6，论文被引频次处于top1%-top5%之间，那么X为5，其他以此类推）， $p(x)$ 为X出现的概率。
- 期望PRS值即各区段的百分位离散随机量加权平均值，比如美国NSF划分的6个百分位区段的期望PRS值为 $1 \times 0.5 + 2 \times 0.25 + 3 \times 0.2 + 4 \times 0.1 + 5 \times 0.2 + 6 \times 0.1 = 1.9$ 。如果机构的值大于1.9，则表明研究实体的绩效高于平均水平。

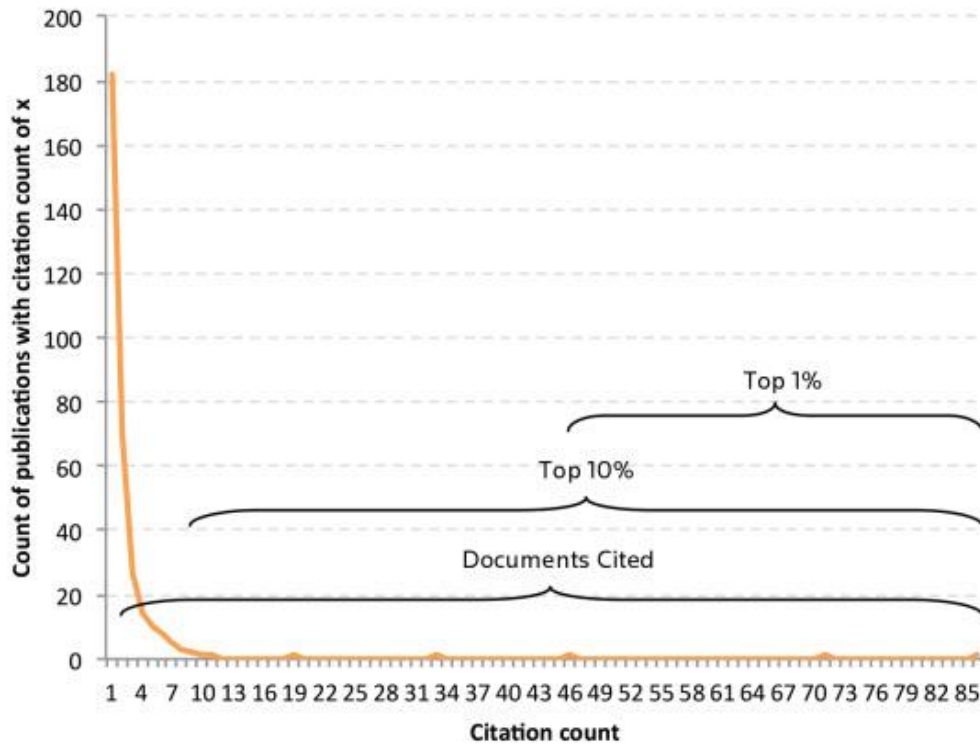
## ● 综合影响指标(I3)

$$PRS = \sum_{x=1}^k x \cdot p(x)$$

# 基于百分位



- Top10% (P top10% 和PP top10% )





# 基于篇均参考文献

- 成功论文数 (Kosmulski, 2011)

- 单篇论文的被引频次除以其参考文献数，所有比值大于1的论文被认为是成功论文。

- 成功指数 (Franceschini, 2013)

- 论文集合的平均参考文献数代替单篇论文的参考文献数，以降低人为操作的影响。

- CSNCR (Bornmann, 2016)

- 论文被引频次/篇均参考文献数量

MCSNCR (Mean CSNCR)。

$$\text{MCSNCR} = \frac{1}{n} \sum_{i=1}^n \frac{c_i}{e_i}$$



# 指标讨论 (1)

- 引文分析方法通过论文引用与被引用关系建立科学计量学指标体系，其计量因素同样受到引文关系的限制，在纳入其他因素时缺乏相应的理论基础，难以兼顾数理统计和评价实践的合理性。
- 许多科研绩效评价指标在实际科研评价过程中得不到应用，除了指标计算的复杂性等原因外，数据源的选择，数据收集的完整性和恰当性等方面也有制约。
- 随着网络与信息技术的发展，数据源和评价模式也不断拓展，基于引文内容的论文评价和替代计量学已经成为传统引文分析方法的有力补充。

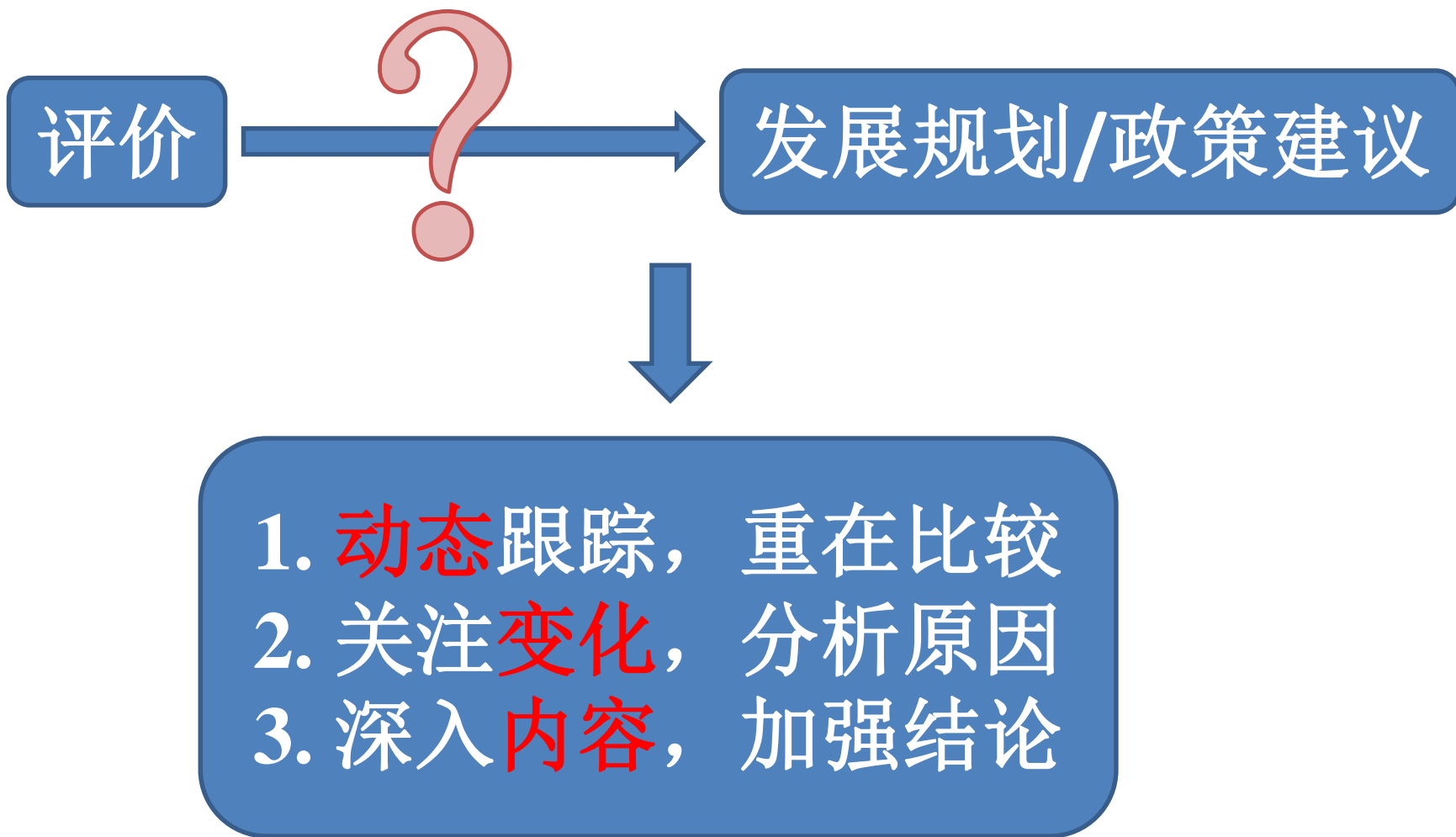




# 指标讨论 (2)

- 指标利用能力
  - 科研人员运用科学计量学指标的特征和形式来表达自己的真实科研价值的能力
- “某种评价一旦被选择用做决策时，这种评价就开始失去其价值”  
---古德哈特定律( Goodhart's law)
- 要保证新指标的提出对现有指标有绝对的附加价值;
- 更多地关注引文影响力指标的理论基础与使用策略的研究;
- 致力于开发更好的数据源以便于更全面准确地计算与应用引文指标。

(Waltman, 2016)





# 评价报告框架 (1)

## ● 机构

### ➤ 维度

- 数量指标：论文数，只含Article和Review
- 质量指标：PP top 10%，CNCI，合作情况

### ➤ 基准水平

- 全球水平
- 985高校水平
- 高水平兄弟院校水平

### ➤ 相对竞争力分析/SWOT分析（象限图）（案例）

### ➤ 结论和建议



# 评价报告框架 (2)

## ● 学科

- 年度及标杆院校
  - 分层次选择标杆3-5个
  - 基本指标比较
- 学科动态分析
  - ESI数据更新（两月一次）
  - 潜势学科判断跟踪
- 学科态势分析
  - 高被引论文及作者分布、国际合作
  - 指标：ESI Top1%、高被引论文及作者分布、国际合作论文数和比例
  - 内容分析（案例）
- 贡献度分析
  - 结合学院、团队或个人，识别主要力量
  - 国际合作机构识别
- 结论和建议



# 评价报告框架 (3)

## ● 学院

### ➤ 分析学科组成

- ESI 学科/一级学科
- 科研人员的学科组成

## ● 个人/团队

### ➤ 数据清洗

### ➤ 署名顺序、机构合作、国际合作

### ➤ 参与和跟踪研究前沿的情况（案例）

## ● 期刊

### ➤ 比较期刊分布

### ➤ 期刊发文态势和指标



# 实践探索 (1)

## ● 分析ESI优势学科中各院系的贡献情况

- 人工甄别每篇论文的第一作者、通讯作者、院系和一级学科
- 数据存入数据库，为以后的一级学科评价、科研绩效评估和人才评价作准备

## ● 关键问题

- 如何把ESI学科和我国一级学科联系起来
  - 虽然新版的InCites数据库提供我国学科分类体系，但期刊分类的机制并不明确，而且其分类目录中找不到植物保护学和园艺学
  - 我国一级学科评估是基于人来划分的
- 如何计算我国**一级学科**的影响力指标
  - 把我国的学科看成基于JCR或ESI分类体系的跨学科实体，以跨学科研究的计量评价方式来实现计量指标计算。

# 实践探索 (1)



## ● 操作步骤

- 从Web of Science中检索与ESI数据库**相对应时间段**的高校或科研机构发表的SCI/SSCI论文数据
- 从InCites数据库导出该机构所有SCI/SSCI论文数据
- 把两个数据库对应起来，并对遗漏的数据进行补充
- 根据论文作者所在院系和专业把论文归入我国的学科分类体系；
- 根据每篇论文的CNCI值，计算一级学科/院系/专业的CNCI值。

# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

## ● 学科计量评价存在的问题

- 学科计量评价只能从总体上反映学科的学术影响力，并不能为学科建设和规划提供更深层的建议
- 研究规划是科学研究的资助机构、管理机构和研究人員共同关注的问题。
  - 资助机构根据这方面研究去识别具有创新和有前景的研究建议和研究人員；
  - 管理者则需要了解目前的研究领域布局去决定经费的分配；
  - 研究人員则通过了解研究现状从而制定自己的研究方向。



# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

- 学科前沿性分析是基于学科研究前沿的学科发展研究规划
  - 学科前沿性分析与计量评价研究思路截然不同，其主要是以论文聚类分析和知识图谱为基础，洞悉机构所属学科的知识结构及发展动态，识别具有发展前景的研究主题。
  - 学科前沿性分析可以明确科研机构所从事的研究在全球科学研究中是否处于前沿领域（或热点领域）以及机构在这些领域的研究地位如何，揭示机构的研究的优势领域或空白领域。

# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

## ● 如何进行学科前沿性分析？

- 首先，我们需要获取学科的研究前沿
- 其次，我们可以从机构在研究前沿的论文分布情况和对前沿的引用情况来测度机构的学科领域分布

## ● ESI Research Fronts是ESI数据库的组成部分之一

- 它是在近5年的高被引论文基础上通过共被引分析和聚类分析得出各学科领域研究前沿。
- 每个研究前沿包括高被引论文数(即分析的文献)、引文数(即高被引论文被引次数，反映了研究前沿的规模)、平均引文数(表明前沿研究的集中度)和平均年份(Mean Year，越近表示研究越前沿)。

# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

- 学科前沿性可以从机构的前沿表现度和前沿关注度两个方面来分析。
  - **前沿表现度**指机构在某个学科的ESI 前沿领域的论文分布情况，具体可以通过ESI 研究前沿中机构所发表的论文数量来表示。
  - **前沿关注度**指机构对学科中前沿领域的关注程度，它不仅反映了研究人员对国际研究热点和前沿的追踪能力和关注程度，更体现了机构的研究前沿的发展潜力和布局。前沿关注度可以通过机构发表论文对前沿领域高被引论文的引用次数来测量。

# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

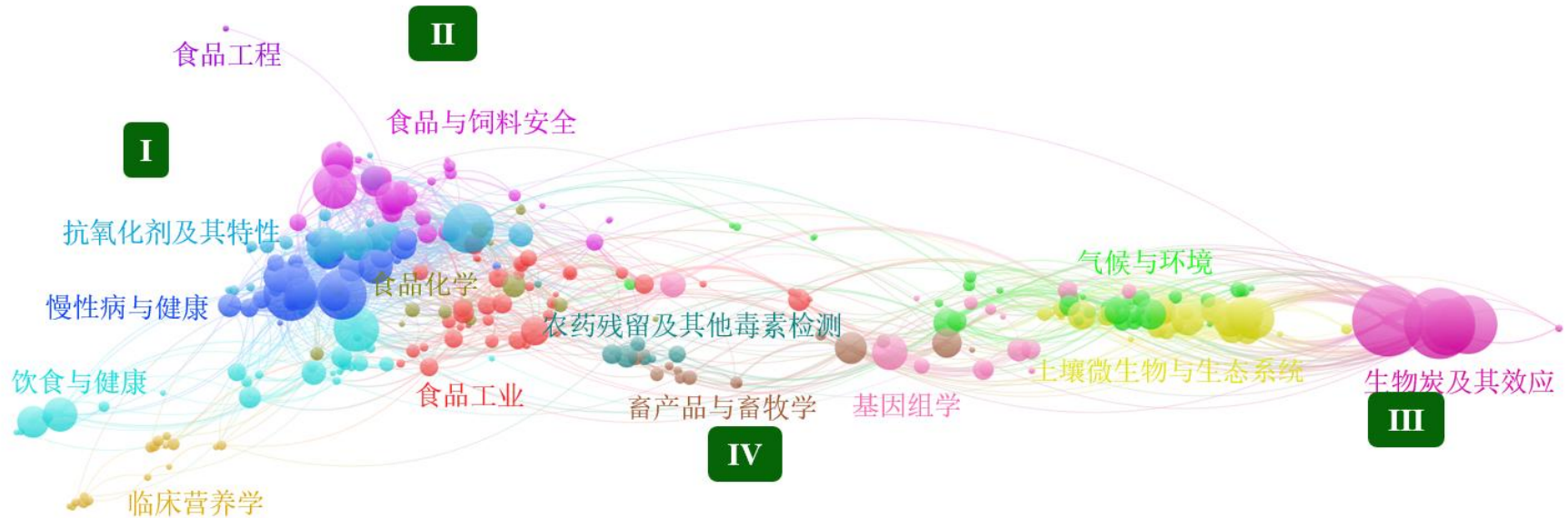
## ● 以农业科学为例

- ESI 农业科学学科共包括392个的研究前沿，高被引论文2206篇；
- 农业科学的392个研究前沿中论文全部属于农业科学的有143个（36.5%），农业科学论文比例为60%-100%的研究前沿有45个（11.5%），比例为30%-60%的研究前沿有111个（28.3%），低于30%的研究前沿93个（23.7%）。
- 本研究将农业科学论文比例在**30%**以上共299个（76.3%）研究前沿（含高被引论文1170篇）作为农业科学的研究前沿。

# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University



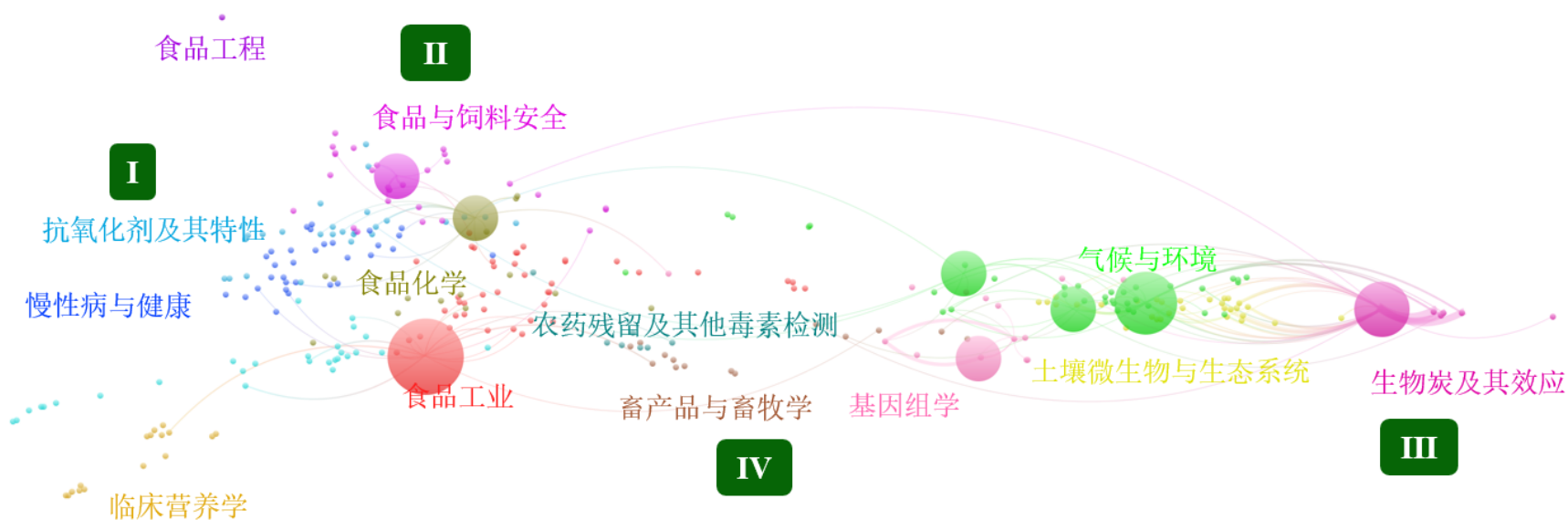
- (I) 人类健康问题和饮食相关研究，101个研究前沿，高被引论文376篇
- (II) 食品工业与安全研究领域，含88个研究前沿，高被引论文335篇
- (III) 生态环境领域，含74个研究前沿，高被引论文254篇
- (IV) 农药残留与毒素检测、基因组学和畜产品加工，分别含7、14和15个研究前沿，高被引论文分别为23、60和122篇

# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

## ● 表现度分析-以中国农大为例



中国农业大学农业科学的高被引论文中分布在II、III、IV区域，从总体上看，中国农业大学农业科学在动物营养、食品科学、气候与环境、基因组学和生物炭方面占据一定的优势地位。

# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

## ● 关注度分析-以中国农大为例



从上图可以看出，中国农业大学在农业科学领域涉及研究前沿范围较广，对研究前沿的四个区域均有引用，并且在四个区域均有引用较高的研究前沿，但主要集中在II区、III区和IV区。

# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

## ● 分析

- 优势前沿：同时具有高表现度和高关注度的研究前沿
  - 4个研究前沿，内容涉及动物氨基酸营养、中国耕地和农业系统的环境和土地酸化问题、生物炭和玉米高通量SNP分型
- 可能优势前沿：高表现度，低关注度，
  - 4个研究前沿
- 关注前沿：高关注度，无表现度
  - 7研究前沿
- 空白领域：无表现度、低关注度
  - 其他



# 实践探索（2）-学科前沿性分析



中國農業大學  
China Agricultural University

## ● 结论

### ➤ 优势：

- 中国农业大学农业科学的动物营养、气候与环境、基因组学和生物炭方面有较强的优势。
- 中国农业大学在农业科学领域涉及研究前沿范围较广，对研究前沿的四个区域均有引用，并且在四个区域均有引用较高的研究前沿

### ➤ 劣势：

- 中国农业大学在人类健康领域研究涉猎得较少，该区域为农业科学与医学交叉领域，在ESI农业科学中占据重要的地位，研究前沿相对密集，有待进一步分析。
- 农药残留与毒素检测、畜产品加工和畜牧学研究缺乏应有的表现度

- 基于WoS的学科评价

- ESI

- InCites

- 基于文献计量指标的可视化知识图谱

- 通用工具如统计分析软件SPSS和SAS;

- 专用工具如Bibexcel、CiteSpace、HiteCite、Leydesdorff系软件和TDA;

- 社交网络分析工具如Pajek和Ucinet;

- 可视化工具如VOSviewer、NetDraw等。



# 案例1 数据处理

---

## ● 主要工具：EXCEL

- 第一作者或通讯作者的提取（AU/C1/RP字段）
- 关键词提取，分词（DE字段）
- WOS号检索（UT字段）
- 词频统计（TI/DE/AB字段）

# 案例2 ESI 潜势学科预测方法



中國農業大學  
China Agricultural University

- 潜势学科指某机构未进入 ESI 引文排名，但具有潜在的发展优势且未来有可能进入 ESI 引文排名的学科领域。
- ESI Top1%学科 (Top1‰)
  - 阈值：总被引频次
  - 问题：
    - 没有排除自引，自引和他引的评价相同，导致可能有机构通过增加自引提高评价；
    - 没有区分机构是否为论文的主要负责机构，主要负责机构与参与完成的评价相同，导致可能有机构为提高影响而挂名发文；
    - 没有考虑施引文献质量差异，被质量高的文章引用与被质量差的文章引用的评价相同。

# 案例2 ESI 潜势学科预测方法



中國農業大學  
China Agricultural University

## ● 注意问题

- 由于ESI数据库与InCites数据库更新频率不一致，。因此，两个数据库的数据源略有差距，但对ESI学科发展趋势评测影响不大。
- ESI统计的文献类型仅为Article和Review，因此InCites数据库获取数据的文献类型也限制为Article和Review。



# 方法 1：学科欠缺度

- 计算方法

$$Pd = \frac{Ce}{Ci} - 1$$

Pd即学科欠缺度指标，Ce指当前进入该学科ESI引文排名的最低引文次数，Ci指机构在该学科实际获得的引文次数。

- Pd值越小表明机构在该学科与ESI引文排名差距越小，从而越接近ESI引文排名。



# 方法2：反向百分位数指标

## ● 计算方法

$$PRS = \sum_{x=1}^k x \cdot p(x)$$

## ● 步骤：

- 从InCites 数据库获取单篇论文的百分数
- 计算PRS值
- 与PRS期望值比较



# 其他预测方法

---

- 定标比超法
- 灰色系统预测方法
- 曲线拟合预测方法

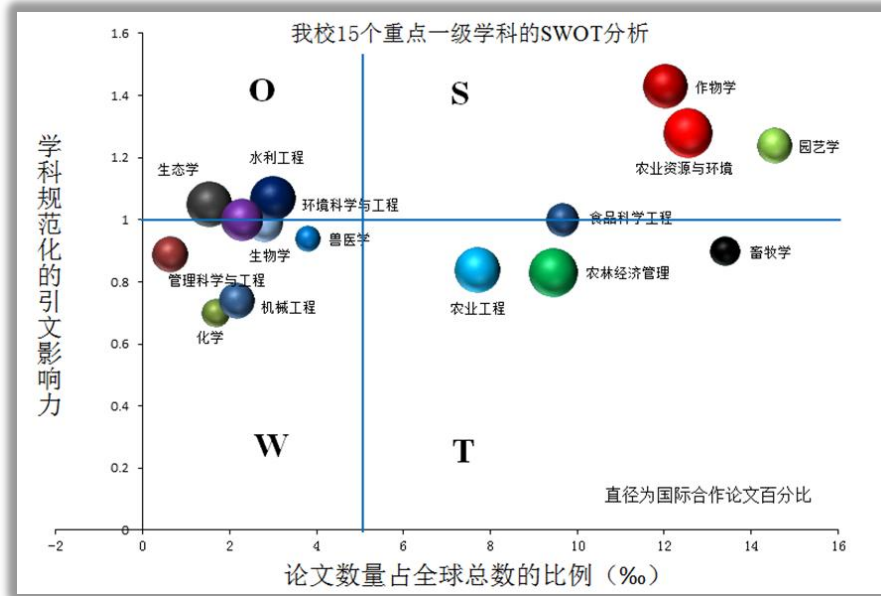
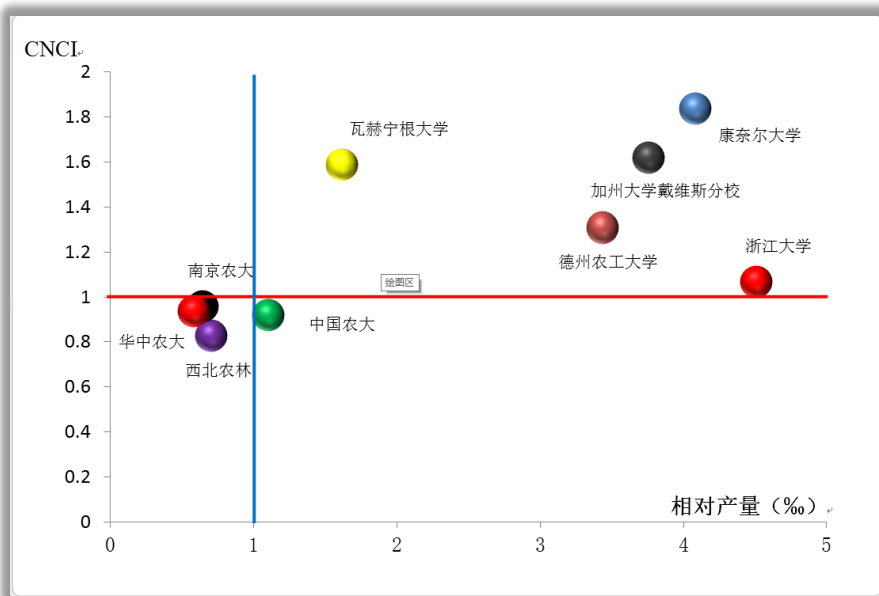


# 案例 3 机构/学科SWOT分析



中國農業大學  
China Agricultural University

- 产量（相对产量）
- 影响力（标准化影响力）



谢谢  
请批评指正！

