

什么是科学计量学

吴金闪

北京师范大学系统科学学院

March 14, 2017



$$\frac{e^{\pm\beta H}}{Z}$$

提纲

- 1 报告的目的
- 2 听课要求
- 3 科学计量学研究什么对象
- 4 几个研究工作
- 5 局限
- 6 参考资料和作业
- 7 为什么做这个作业
- 8 总结
- 9 致谢

整理思路启发思考促进研究

- 通过研究工作给科学计量学一个概貌
- 更一般地，如何了解一个领域如何开展工作，系联性思考和大图景

更具体的目标

- 整理和指引科学计量学的一些关键的问题、思维和分析方法
- 网络科学的思想 and 科学计量学的关系
- 学习方法：大图景大方向和小问题的结合
- 学习和思考的方法：看到联系，用联系来计算和思考

不是任何一个地方都可以听到这样的课

- 听，企图搞懂
- 随时反思
- 搞懂和反思都是在拷问和建立我讲的东西和你已经知道的东西之间的联系
- 然后问，一定要问
- 最后要做作业

举几个例子，不用担心全面和系统

- 例如期刊影响因子，为了解决什么问题

举几个例子，不用担心全面和系统

- 例如期刊影响因子，为了解决什么问题
- 例如论文的总被引次数（以及某种归一化），为了解决什么问题（信息检索，推荐）

举几个例子，不用担心全面和系统

- 例如期刊影响因子，为了解决什么问题
- 例如论文的总被引次数（以及某种归一化），为了解决什么问题（信息检索，推荐）
- 例如作者总被引次数，为了解决什么问题

举几个例子，不用担心全面和系统

- 例如期刊影响因子，为了解决什么问题
- 例如论文的总被引次数（以及某种归一化），为了解决什么问题（信息检索，推荐）
- 例如作者总被引次数，为了解决什么问题
- 例如H指数，为了解决什么问题

举几个例子，不用担心全面和系统

- 例如期刊影响因子，为了解决什么问题
- 例如论文的总被引次数（以及某种归一化），为了解决什么问题（信息检索，推荐）
- 例如作者总被引次数，为了解决什么问题
- 例如H指数，为了解决什么问题
- 例如一个国家（或者学校，在某个领域内）的总论文数量，为了解决什么问题

举几个例子，不用担心全面和系统

- 例如期刊影响因子，为了解决什么问题
- 例如论文的总被引次数（以及某种归一化），为了解决什么问题（信息检索，推荐）
- 例如作者总被引次数，为了解决什么问题
- 例如H指数，为了解决什么问题
- 例如一个国家（或者学校，在某个领域内）的总论文数量，为了解决什么问题
- 科学技术主体的评价和推荐

继续举几个例子

- 例如论文、期刊的领域标注，为了解决什么问题

继续举几个例子

- 例如论文、期刊的领域标注，为了解决什么问题
- 例如作者的领域标注，为了解决什么问题

继续举几个例子

- 例如论文、期刊的领域标注，为了解决什么问题
- 例如作者的领域标注，为了解决什么问题
- 例如领域的所谓知识图谱，为了解决什么问题

继续举几个例子

- 例如论文、期刊的领域标注，为了解决什么问题
- 例如作者的领域标注，为了解决什么问题
- 例如领域的所谓知识图谱，为了解决什么问题
- 科学家工作的主题和领域的选择，为了解决什么问题

继续举几个例子

- 例如论文、期刊的领域标注，为了解决什么问题
- 例如作者的领域标注，为了解决什么问题
- 例如领域的所谓知识图谱，为了解决什么问题
- 科学家工作的主题和领域的选择，为了解决什么问题
- 例如什么年龄最具有创新性，跟谁合作最有利等等

继续举几个例子

- 例如论文、期刊的领域标注，为了解决什么问题
- 例如作者的领域标注，为了解决什么问题
- 例如领域的所谓知识图谱，为了解决什么问题
- 科学家工作的主题和领域的选择，为了解决什么问题
- 例如什么年龄最具有创新性，跟谁合作最有利等等
- 有为科学家研究工作服务的吗？

小结

- 都在研究些什么对象

小结

- 都在研究些什么对象
- 都在回答哪些问题，为什么要问这些问题

小结

- 都在研究些什么对象
- 都在回答哪些问题，为什么要问这些问题
- 都怎么做研究，数据、分析方法

小结

- 都在研究些什么对象
- 都在回答哪些问题，为什么要问这些问题
- 都怎么做研究，数据、分析方法
- 科学家的活动、活动的成果（论文、专利、书、科学学科领域）
- 现状描述，服务科学家的研究，服务研究的组织管理资助，服务成果的出版传播
- 服务从成果到教学和学习的转化

这个思路下的其他问题举例

- 有了研究对象和主要研究目的，我们来看还有哪些要做的

这个思路下的其他问题举例

- 有了研究对象和主要研究目的，我们来看还有哪些要做的
- 科学之间的关系，学科领域之间的关系，概念之间的关系（知识框架基础）

这个思路下的其他问题举例

- 有了研究对象和主要研究目的，我们来看还有哪些要做的
- 科学之间的关系，学科领域之间的关系，概念之间的关系（知识框架基础）
- 科学和技术之间的关系，看起来没什么用的基础科学研究政府要支持吗
- 为了某个明确的技术要优先发展哪些学科和技术

这个思路下的其他问题举例

- 有了研究对象和主要研究目的，我们来看还有哪些要做的
- 科学之间的关系，学科领域之间的关系，概念之间的关系（知识框架基础）
- 科学和技术之间的关系，看起来没什么用的基础科学研究政府要支持吗
- 为了某个明确的技术要优先发展哪些学科和技术
- 自动摘要，自动综述？

主要思维方式

- 引文是基本关系，论文之间，论文-专利之间，书以及各种文档之间

主要思维方式

- 引文是基本关系，论文之间，论文-专利之间，书以及各种文档之间
- 共被引、共施引，能用于分类吗？

主要思维方式

- 引文是基本关系，论文之间，论文-专利之间，书以及各种文档之间
- 共被引、共施引，能用于分类吗？
- 作者-主题共现用于审稿人的选择，作者-引文呢？
- 作者-作者合作关系实际上可以看做从作者到论文再从论文到作者的关系的耦合

主要思维方式

- 引文是基本关系，论文之间，论文-专利之间，书以及各种文档之间
- 共被引、共施引，能用于分类吗？
- 作者-主题共现用于审稿人的选择，作者-引文呢？
- 作者-作者合作关系实际上可以看做从作者到论文再从论文到作者的关系的耦合
- 有基本的数据描述模型吗？

主要思维方式

- 引文是基本关系，论文之间，论文-专利之间，书以及各种文档之间
- 共被引、共施引，能用于分类吗？
- 作者-主题共现用于审稿人的选择，作者-引文呢？
- 作者-作者合作关系实际上可以看做从作者到论文再从论文到作者的关系的耦合
- 有基本的数据描述模型吗？
- 有基本的分析计算方法吗？
- 分析方法有大概一致的思路思维方式吗？

主要思维方式，续

- 一切从数据开始

主要思维方式，续

- 一切从数据开始
- 一切都是关系

主要思维方式，续

- 一切从数据开始
- 一切都是关系
- 表现数据和关系的一般方式是网络
- 网络的核心思想是间接联系和直接联系都重要，而且间接联系可以计算

PageRank用于论文期刊排序

- 首先排序的目的基本上是为了信息获取
- 其次，可以直接统计被引次数来排序
- 但是，被重要网站引用和被一般网站引用应该不一样。怎么办？
- 传播起来，跃迁几率矩阵和不变向量
- Eigenfactor、Scimago

引文网络直接用于领域标注

- 以前，共被引被经常用于聚类，也就是领域标注
- 了解共被引实际上是引用网络的投影（两步耦合）之后，自然地，可以考虑直接用引文网络
- 怎么用？网络聚类算法初步，介数算法和间接关系

广义投入产出用于领域关联

- 另一个“转移矩阵”以及他的背景
- 注意其中的思想：间接联系
- 还可以用于什么其他地方
- 现在打开另一个pdf文件“广义投入产出用于科学领域关联分析”
- 课程的第二部分，一会还回来

其他几个工作

- 个性化PageRank用于作者、论文等排序
- 科学学数据的多层网络描述
- 进一步的研究：讨论什么问题、算法是什么、如何检验
- 试着举几个例子？

科学计量学的困难

- 没有好的度量单位
- 引用是准确的吗？是可以当做一个单位的吗？
- 概念网络基本没有
- 计算机还不能理解论文
- 结果基本只能看做一个大概，但是却有可能被当做标准来使用
- 服务科学家做的还太慢
- 数据公开和深入

- 见“什么是科学计量学——吴金闪的工作和思考”的总结和里面的文献
- 以其中的文献[1]（或者其中任何一个）为起点，整理一个文献报告，需要包含主要（什么意思，你需要提供说明和选择的方法）的往前和往后的论文，需要对论文做一个大概的分类和综述。
- 在前面整理的从[1]（或者其中任何一个，但是请保持两个作业的起点文献相同）往后的论文中，请选出来主要（什么意思，你需要提供说明和选择的方法）的来自于科学计量学（如何确定，你需要提供说明和选择的的方法）的论文，对论文做一个大概的分类和相比上一题更加详细的综述。
- 思考题：了解一下科学计量学的历史之后，请回答，之前主要做什么发展情况，现在主要做什么发展情况如何，这个转变是因为什么。如果有数据，请用数据来支持你的论证。

- 玩玩科学计量学的工具（例如“主要”的衡量）和数据（从数据开始是科学计量学的核心理念，引文数据库的熟悉和使用是最基本的技能）
- 熟悉一下间接联系的计算和意义这个主题
- 学会做文献综述（整理出来线索和分类远远比阅读文献本身难和有意义）

读文章作总结：WHWM问题

- 文章最主要的信息是什么(What)
- 这个信息如何通过具体的例子、概念和思考来构建 (How)
- 作者为什么想传达它，为什么这样传达(Why)
- 对于我这个读者来说意味着什么 (Meaningful)

今天学了什么

- 抓住一个学科的大图景：基本问题、典型思维、典型分析方法，和它们的例子，从大到小从小到大
- 科学学的典型思维：联系和数据，以及一般描述、一般分析方法
- 更一般的学习怎么学：系联性思考，进去-跳出来
- 文章怎么看WHWM
- 从大图景和基本问题中提出新的问题、提炼新的分析方法，甚至思想的创新，一般描述框架的创新
- 促进学科的发展
- 定位自己的研究工作

感谢大家，问题时间

- 感谢这些相关工作的合作者们
- 感谢同学们
- 感谢杨立英老师的邀请