



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

投入产出分析用于科学领域之间的相互影响

吴金闪

北京师范大学系统科学学院

March 13, 2017

报告的主要内容



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

- ① 这个工作所讨论的基本问题
- ② 讨论这些问题的基本思想
- ③ 来自于经济学的开放系统的投入产出分析以及在封闭系统上的拓展
- ④ 封闭系统投入产出分析方法用于科学领域
- ⑤ 其他相关的研究：汉字关系网络和学习、产业相互影响等
- ⑥ 讨论时间、致谢

问题



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

- 问题1：一个科学领域是如何与其他的科学领域关联起来的，其支撑的领域和支撑这个领域的其他领域有哪一些，如何衡量紧密关系
- 问题2：一个科学领域对整个学术界的影响力度量和排序

思想



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

- 投入产出关系与经济学的投入产出分析
- 学术世界的大投入产出表：论文发表、专利、基金、学术单位
- 某个领域的论文发表作为一个“封闭”子系统。数据呀数据！

基本思想和一个问题



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

- 在经济领域，大家关心这个问题：某产业对整个经济系统的影响，某一个产业对其他产业的影响。（顺便，这个问题还可以在产品的层次来问）
- 经济学家把它处理成一个开放系统：把人类的生产和消费把其它经济产业部门分开，时间尺度、价值衡量的方式不一样。假设在一定时期内，人类的总消费是一个可以从其他渠道预测或者估计的外生变量。
- 科学世界等其他更加一般的系统没有这样自然的区分。例如物理学的各个分支学科，哪一个做为外生变量比较合适？



简介经济学的投入产出分析

- 定义 x_j^i 是从产业部门 i 投入到产业部门 j 的产品的量（以钱为单位）
- 定义从产业部门 i 输出的总量为 X^i

$$X^i = \sum_{j=1}^{N-1} x_j^i + x_N^i = \sum_{j=1}^{N-1} b_j^i X^j + y^i, \forall i \neq N. \quad (1)$$

$$b_j^i = x_j^i / X^j, \quad (2)$$

$$X = (X^1, X^2, \dots, X^{N-1})^T \quad (3)$$

$$Y = (y^1, y^2, \dots, y^{N-1})^T \quad (4)$$

- 以上的投入产出矢量可以表达成如下关系： Y 是最终消费， X 是各个部门的总产出

$$BX + Y = X \iff X = (1 - B)^{-1} Y \implies \Delta X = (1 - B)^{-1} \Delta Y \quad (5)$$

$$L_B \triangleq (1 - B)^{-1} \quad (6)$$



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

为什么不直接用总产量？

- 其实是可以的，也是合理的。就好像网络中用顶点的度来反映顶点的重要性。
- 但是，有一些现象，需要考虑间接影响，例如部门A不直接投入到部门C，而是主要投入到部门B，然而部门B的绝大部分都投入到部门C，于是，实际上部门A对部门C具有很重要的影响。
- 一个好的度量方式需要考虑这样的间接影响。Google PageRank算法也是考虑的这个影响。
- 投入产出分析中的 B 就称为直接系数矩阵， L 就称为间接系数矩阵。 L 就考虑了这样的间接影响。
- 典型的投入产出分析就是讨论 L 的各种性质。



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

HEM方法：从去掉一个部门

- 假设去掉部门 i 同时保持 Y （除去 y^i ）不变，那么

$$\hat{X}^{-i} = L_B^{-i} Y^{-i} = (1 - B^{-i})^{-1} Y^{-i}. \quad (7)$$

这个是去掉部门 i 之后满足最终需求的总产出矢量。

- 于是 X 和 $\hat{X}^{-i}$ 之差反映了部门 i 的重要性。例如，这个差最大的分量显然受部门 i 影响较大。
- 这个方法叫做Hypothetical Extraction Method (HEM)。



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

针对封闭系统的新方法：本征分析

- 在科学领域这个问题中，再也没有合理的自然的 Y 和 ΔY 了。
- 包含各个产业部门，我们有

$$X^i = \sum_{j=1}^N x_j^i = \sum_{j=1}^N b_j^i X^j, \forall i. \quad (8)$$

$$BX = X. \quad (9)$$

- 用这个封闭系统的记号重新写一下开放系统的表达式，

$$BX + Y = X \rightarrow B^{-N}X + Y^N = X$$

- B 是一个非负元素矩阵，在一定条件下，存在最大本征值以及相应的惟一本征向量。可以证明这个最大本征值就是1。于是总产量向量 X 就是这个特殊的本征向量。这一点，非常重要。



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

针对封闭系统的新方法：本征分析，续

- 借用HEM的思想，为了度量部门*i*的重要性，我们考虑 B^{-i} ——从总矩阵 B （不是经济学投入产出分析的 B^{-N} ）中去掉部门*i*。
- 然后，我们计算 B^{-i} 的最大本征值和本征向量： $\lambda_{\max}^{-i}$ ， $|\lambda_{\max}^{-i}\rangle$ 。是不是大概相当于去掉部门*i*以后的经济系统的最大总产出？
- 如果是，那么我们把 $\lambda_{\max}^{-i}$ 做为重要性度量，同时把 $|\lambda_{\max}^{-i}\rangle$ 当作部门*i*对其它部门的影响力的度量矢量。

$$S_{i0}^j = 1 - \lambda^{(-j)}, \quad (10)$$

$$\Delta_k^j = \frac{\langle k | X \rangle - \lambda^{(-j)} \left(\sum_{l \neq j} X^l \right) \langle k | \lambda^{(-j)} \rangle}{\langle k | X \rangle}. \quad (11)$$



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

封闭系统新方法：看做开放系统

- 还是用HEM的思想，我们考虑这样一个过程，去掉 k 部门（ k 不再接受其他部门的投入），整个系统发生什么。
- 首先，其他部门到 k 部门的直接投入都没有用了；其次，其他部门用来生产这些投入的投入也没有用了；.....
- 于是，计算过程是：首先， Bx_k 然后去掉得到的矢量的 k 分量（直接投入）；接着， B 再一次乘上上面得到的矢量，得到新的矢量以后，去掉 k 分量；.....

$$\Delta^k = \left(1 - B^{(-k)}\right)^{-1} x_k^{(-k)}, \quad (12)$$

$$S_{IO}^k = \sum_j \Delta_j^k \quad (13)$$

这里的 $x_k^{(-k)} = \left[x_k^1, \dots, x_k^{k-1}, x_k^{k+1}, \dots, x_k^N\right]^T$ 。

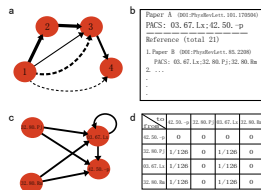
- 这个计算的结果正在整理中。



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

物理学的各个分支投入产出矩阵构建

- PACS码做为分支的标志。
- 把文章之间的引用关系转化成领域之间的依赖关系。



$$x_j^i(A \rightarrow B) = \frac{1}{P_A P_B C_A}. \quad (14)$$

Figure: P_A (P_B) 是文章A (B) 的PACS码数量, C_A 是文章A的参考文献数量。

物理学的各个分支的影响力



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

- 问：哪一个子领域在整个物理学中最有影响力，给其他的领域贡献思想、方法，或者使得其它领域的思想和方法发挥作用？
- 问：某一个领域，如果发展迟缓了，或者取得重要的发展了，对其它哪一些领域的影响最大？
- 这些问题对于科学管理部门、科学家都有重要意义。
- 这些问题与经济学投入产出分析处理的问题完全相同，仅仅系统封闭了，然后物质的量（这里是ideas）不再守恒了。
- 正好可以用我们提出来的封闭系统的投入产出分析方法。



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

物理学的各个分支的影响力，续

- 领域影响力排序，考虑IOF和一个领域的被引用次数、文章数量的相关性

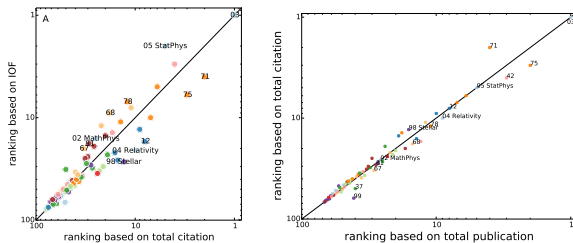


Figure: 关注05 (StatPhys)、02 (MathPhys)、04 (Relativity) 以及98 (Stellar)这些突出的领域。

物理学的各个分支的影响力，续



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

- 领域影响力排序，考虑IOF和一个领域的被引用次数、文章数量的相关性

Figure: 各领域影响力的时间演化



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

物理学分支之间的相互影响

- 给定领域，对其他领域的影响，以98（Stellar systems）和03（Quantum Mechanics）为例

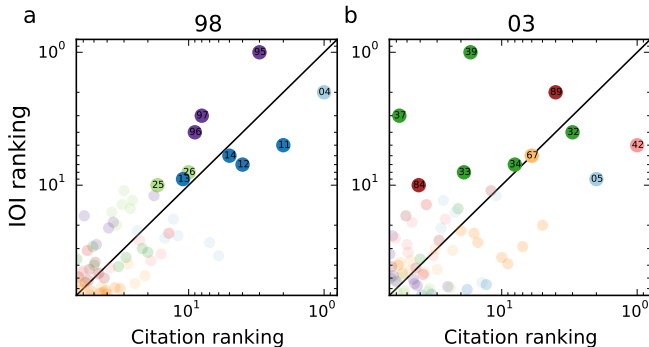


Figure: 大部分与98领域关系密切的都是天文学和相对论。直觉上这个是合理的。对于03领域，首先，影响力和引用的图相对于前者比较分散，其次引用关系不太密切的37（原子操控等）和39（原子分子仪器，后并入37）与03之间的相互影响非常密切。



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

物理学发展全局图景的把握

- 把所有物理学领域粗略地分成4个集团（按照领域排名取25%），看一看随着时间，这些领域在集团之间的运动

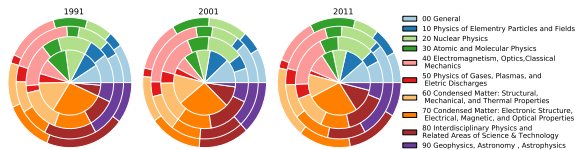


Figure: 可以关注核心25%之中的General Physics (00)和Interdisciplinary (80)。



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

影响力排名

- 把所有物理学领域粗略地分成4个集团（按照领域排名取25%），看一看随着时间，这些领域在集团之间的运动

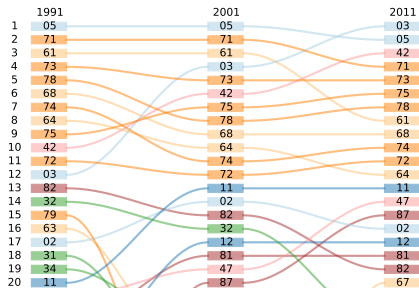


Figure: Level-2前20个领域随着时间的变化

各种“产业部门”之间的关系都可用



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

- 其它的学科（经济学、数学、计算机科学）的子领域的关系的讨论
- 还可以考虑把城市作、研究单位、基金组织、国家、期刊作为产业部门等
- 把学术单位和学科领域交叉起来分析
- 把专利技术和科学研究、基金支持都放进来，构成一个更大的系统，讨论对于某技术方向的发展，最重要的学术研究方向有哪一些
- 把新方法用回到经济学研究中去
- 实际上思想上和计算技术上，上面的分析和我们之前的汉字学习的工作、Google的PageRank的工作的是一样的，本征向量或者线性方程的解



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

这个研究在本研究之外的价值

- 以问题为导向的研究，找到和发展适合这个问题的数学结构
- 提出一个一般的方法、开辟一个发方向，有很多后续的工作可以展开：可以用于任何具有投入产出关系的系统中，考察相互影响和相对地位的问题
- 从前人那里继承了核心的思想：Leontief的投入产出、Narin的科学研究以及科学研究与技术之间的相互影响、来自于网络科学以及PageRank算法的传递
- 统计物理学就是处理多个相互作用的个体构成的系统的性质，从个体行为到群体行为的桥梁（通常表现为数学结构）
- 以前依赖于大脑的决策的各个方面，都可以数学化

请提问



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

- 团队成员：沈哲思、李梦辉、鲍建樟、韦添、裴建锁、杨立英、吴晨生、敖斌、张江、狄增如、Ronald Rousseau
- 感谢中国科学院图书馆在数据方面的协助，APS编辑部在数据方面的协助
- 感谢杨立英的邀请