计算理论之美

南京大学
计算机科学与技术系
尹一通

“计算机科学？是研究计算机的吗？”

“Computer science is no more about computers than astronomy is about telescopes.”

--- Edsger Dijkstra

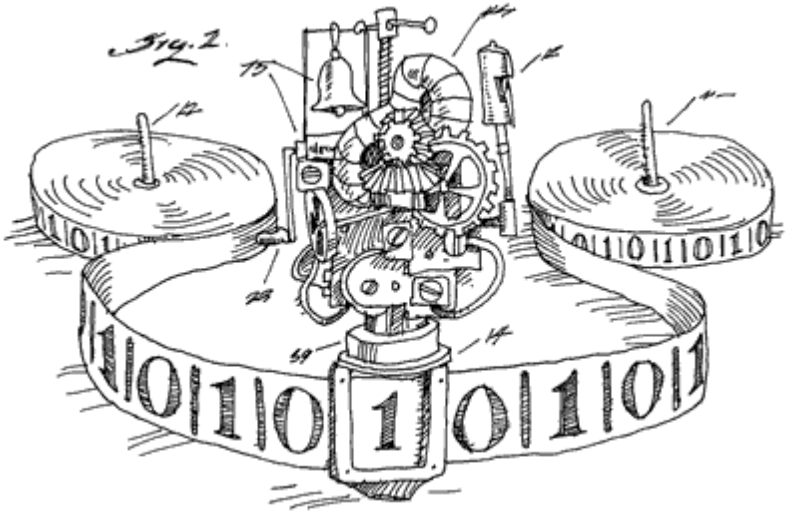
Alan Turing

什么是计算？

“What is computation?”

阿兰·图灵
Alan Turing (1912-1954)

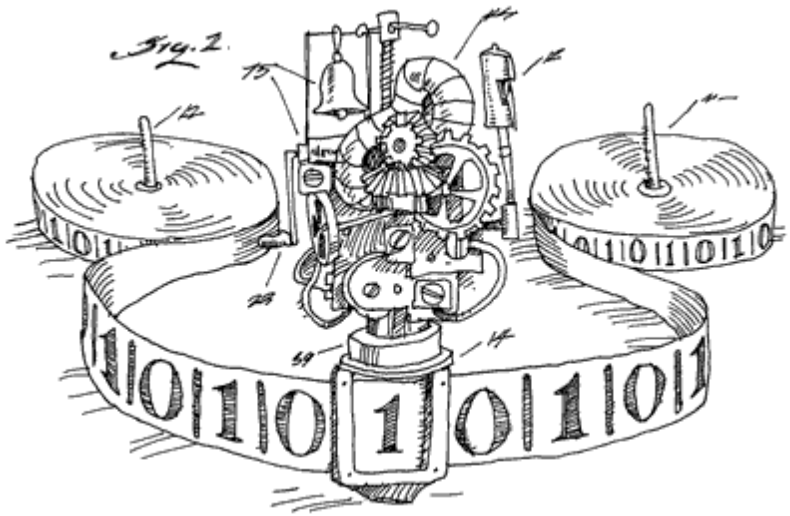
Turing Machine



图灵机
Turing machine

阿兰 · 图灵
Alan Turing (1912-1954)

Turing Machine

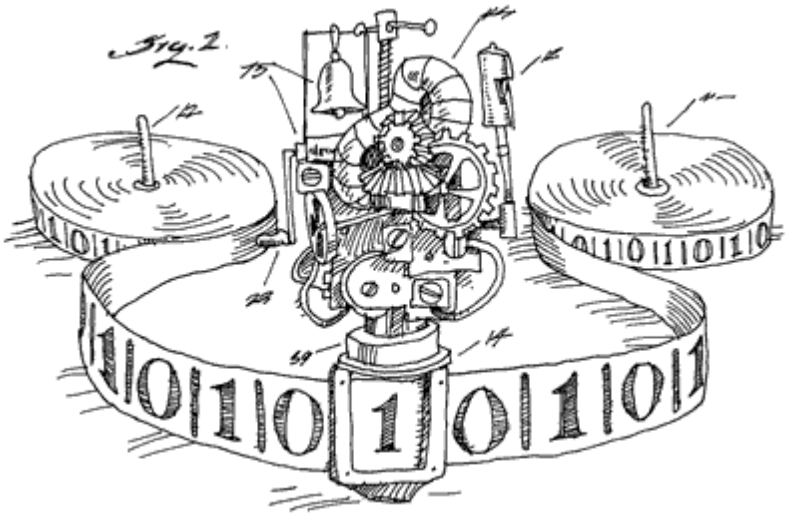


图灵机
Turing machine

图灵机 $M = \langle Q, \Gamma, \square, \Sigma, \delta, q_0, F \rangle$

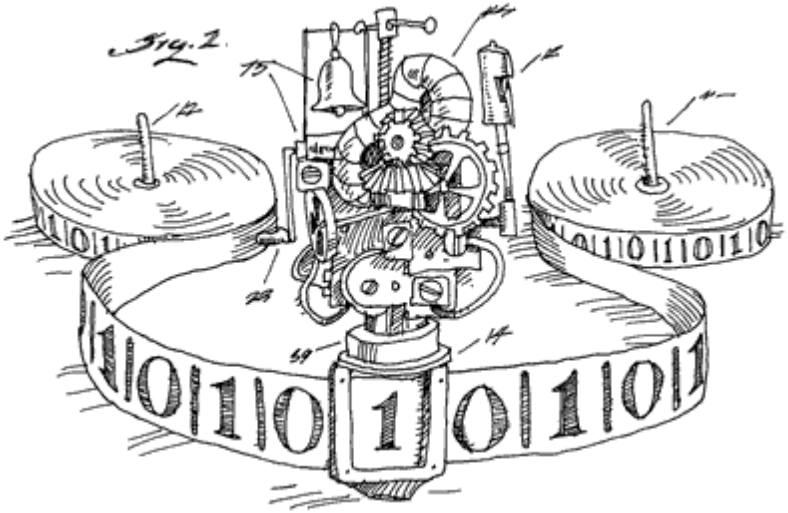
- Γ: 非空有穷纸带字母表
- $\square \in \Gamma$: 空白符
- $\Sigma \subseteq \Gamma \backslash \{ \square \}$: 非空有穷输入字母表
- Q: 非空有穷状态集合
- $q_0 \in Q$: 初始状态
- $F \subseteq Q$: 终止状态集合
- $\delta : (Q \backslash F) \times \Gamma \to Q \times \Gamma \times \{ \triangleleft, \triangleright \}$
  转移规则函数

Turing Machine

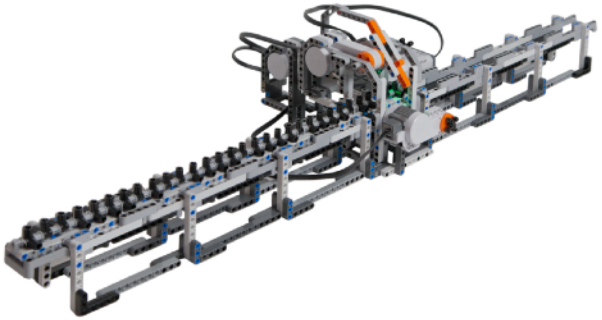


图灵机
Turing machine

Turing Machine



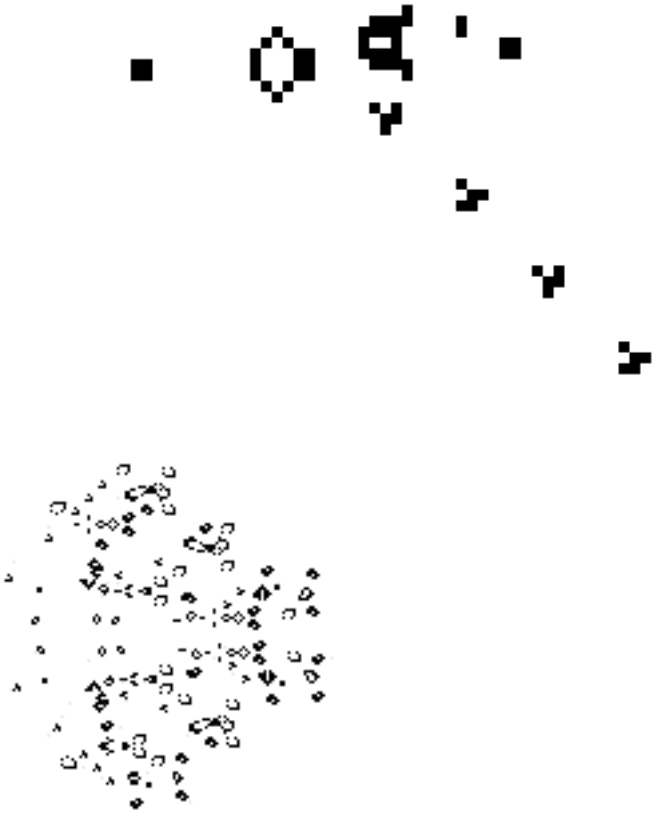
图灵机
Turing machine

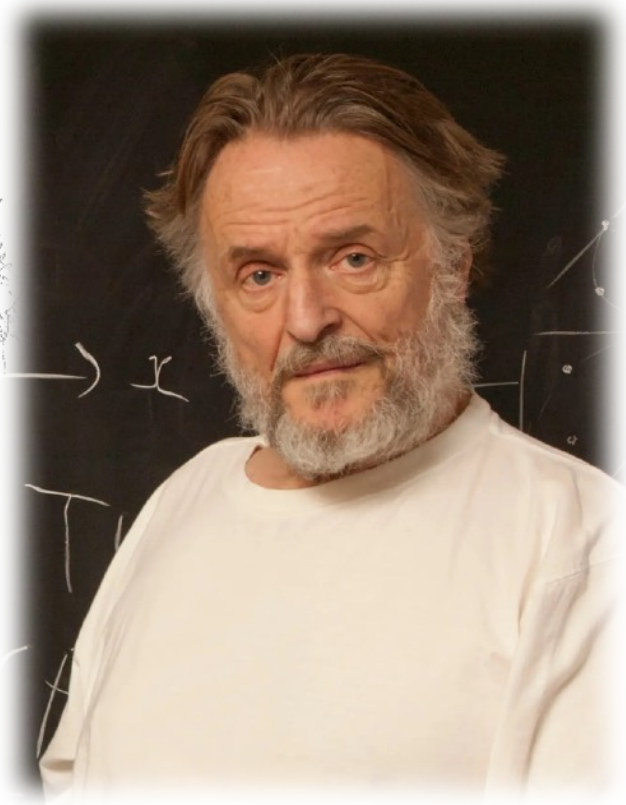


LEGO[tm] Turing Machine

MINECRAFT

Conway's Game of Life





Game of Life
元胞自动机

John Horton Conway
(1937-2020)

Turing Machine

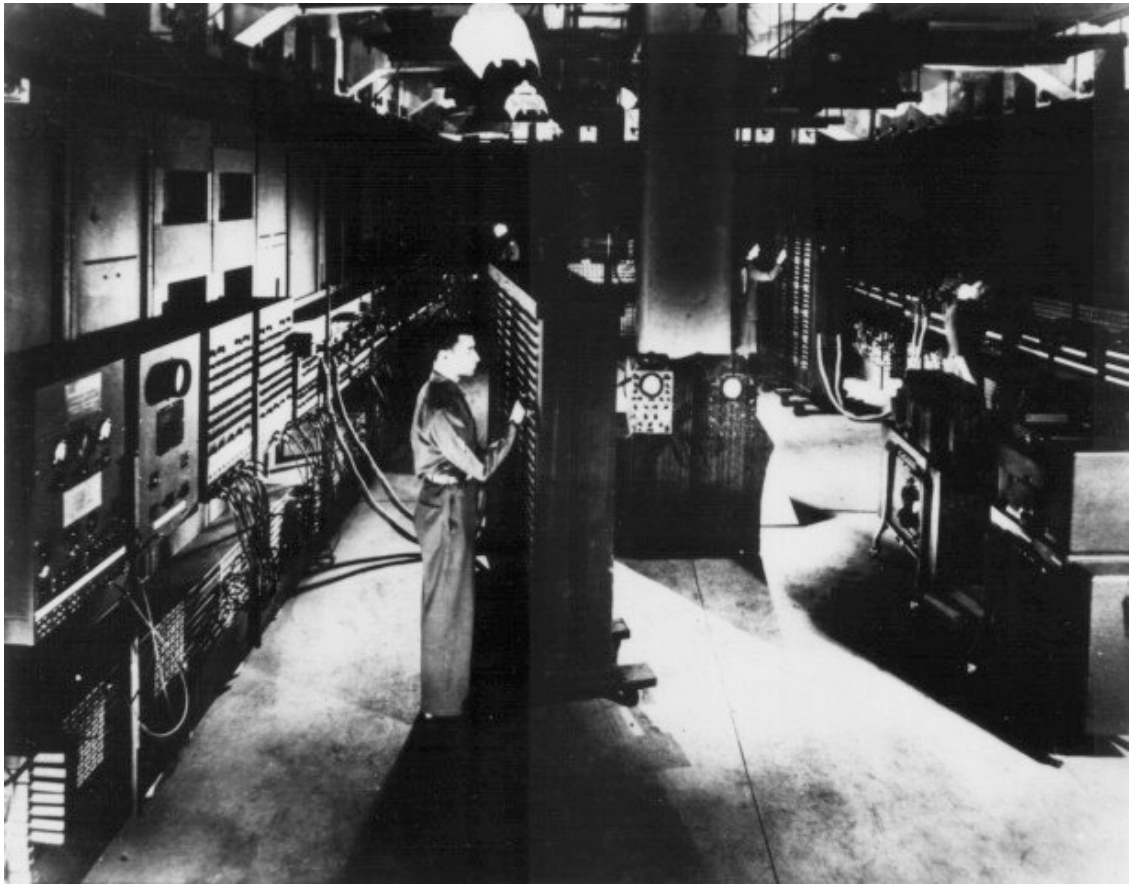
ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO THE ENTSCHEIDUNGSPROBLEM

By A. M. TURING.

[Received 28 May, 1936.—Read 12 November, 1936.]

The "computable" numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means. Although the subject of this paper is ostensibly the computable *numbers*, it is almost equally easy to define and investigate computable functions of an integral variable or a real or computable variable, computable

ENIAC



第一台通用电子计算机 ENIAC（1947年）

David Hilbert

- “希尔伯特问题”（1900）
- 23个数学难题：
- 第2问题：

  为算数建立完备的公理体系
- 第10问题：

  给出解整数多项式方程的算法

大卫·希尔伯特
David Hilbert (1862-1943)

Euclid (300 BC)

欧几里德

Bertrand Russell

罗素悖论(Russell Paradox) 1901：

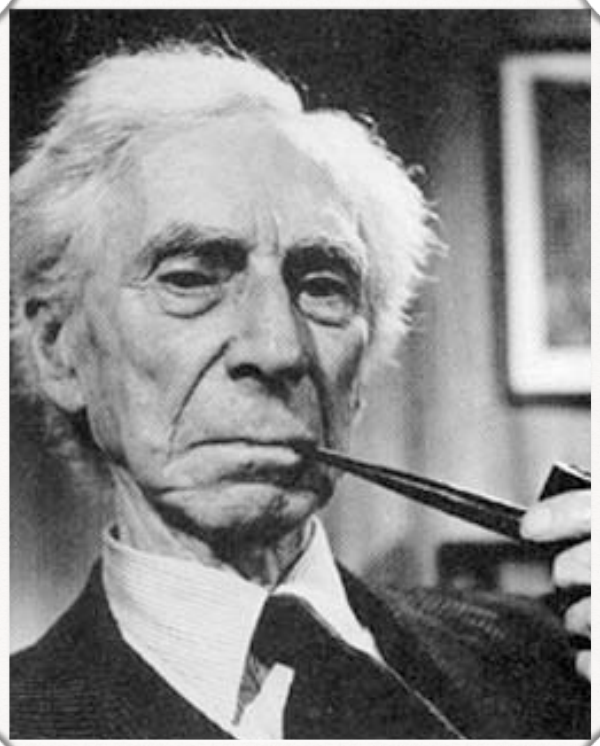
“本命题是假的”

是真是假?

定义集合A为：
所有“自己不属于自己的集合”的集合

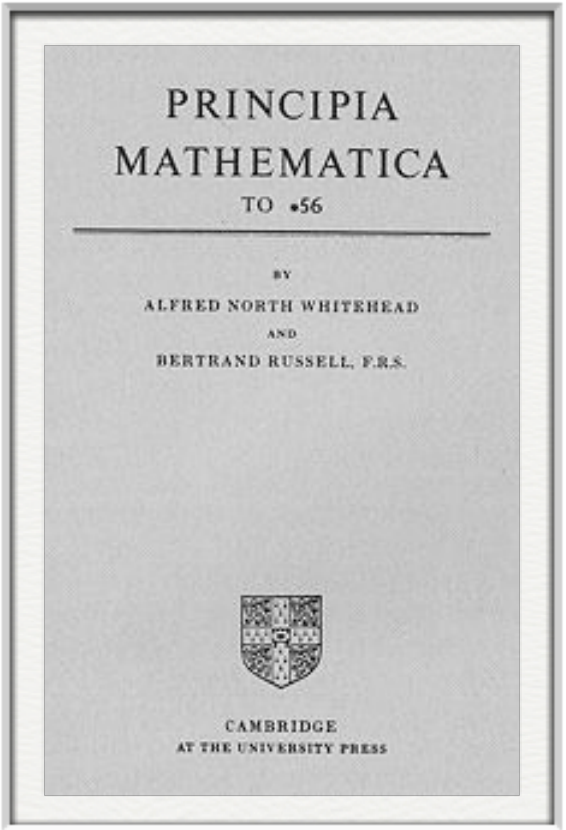
$$A = \{S \mid S \notin S\}$$

A 是否属于它自己?

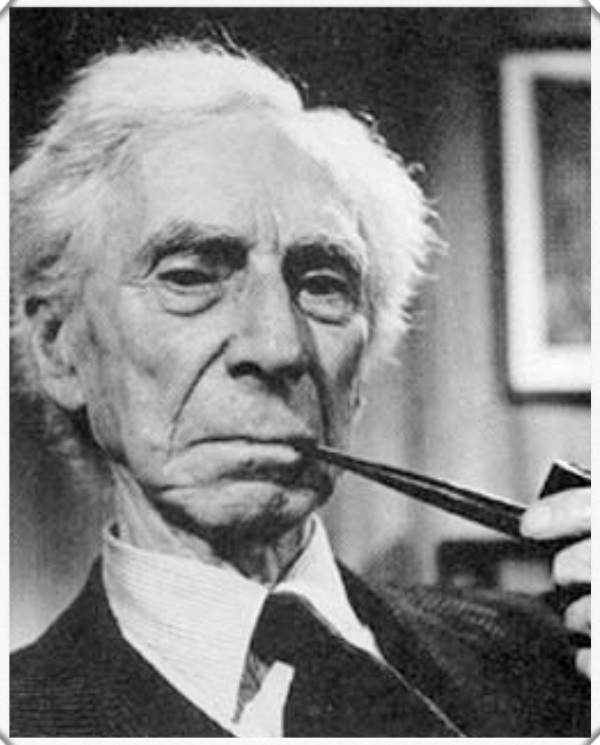


伯特兰 · 罗素
Bertrand Russell
(1872-1970)

Bertrand Russell



《数学原理》
Principia Mathematica
(1910, 1912, 1913)



伯特兰·罗素
Bertrand Russell
(1872-1970)

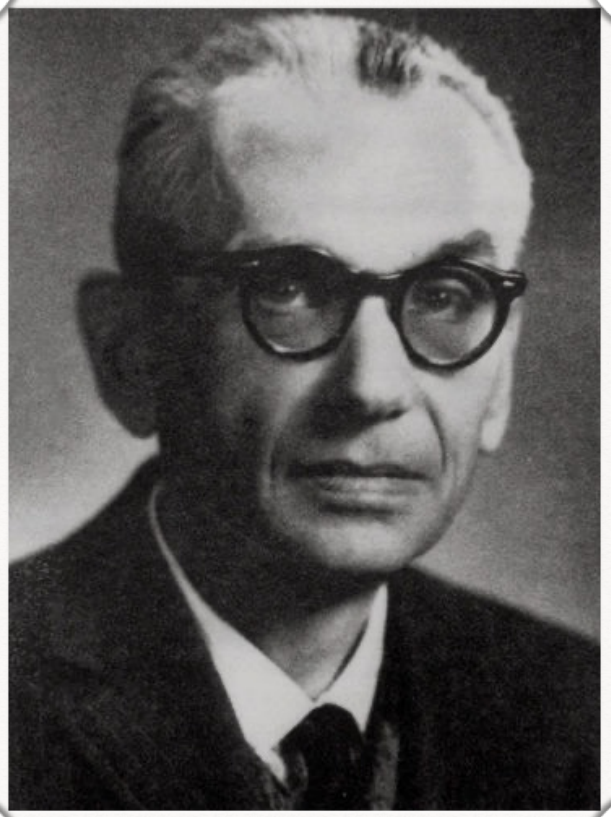
Kurt Gödel

"本命题不能被证明"

可以被证明吗?

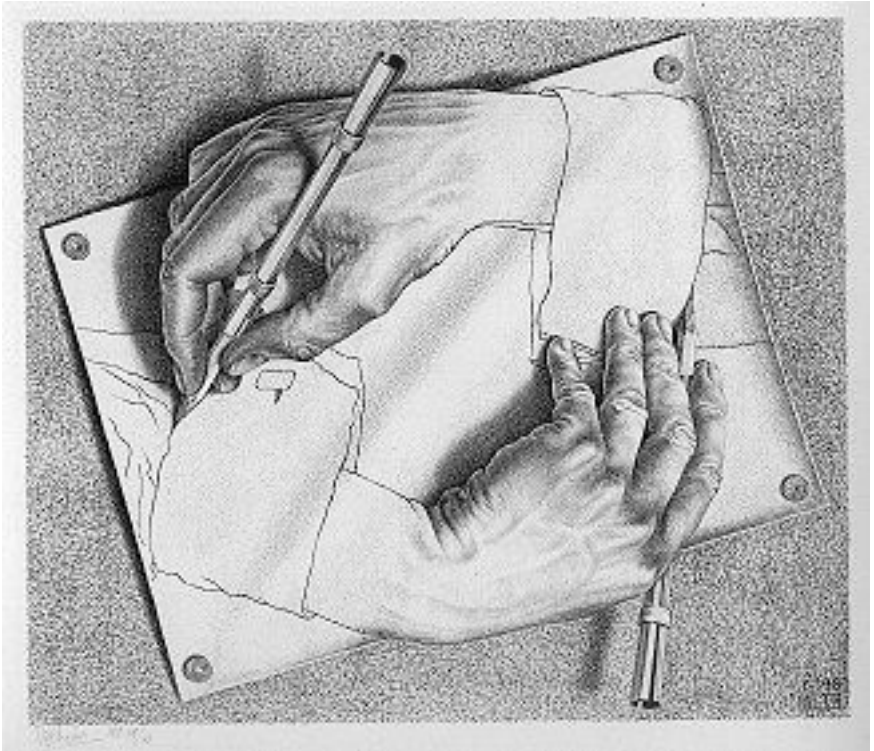
哥德尔不完备定理 (1931)
Incompleteness Theorem：

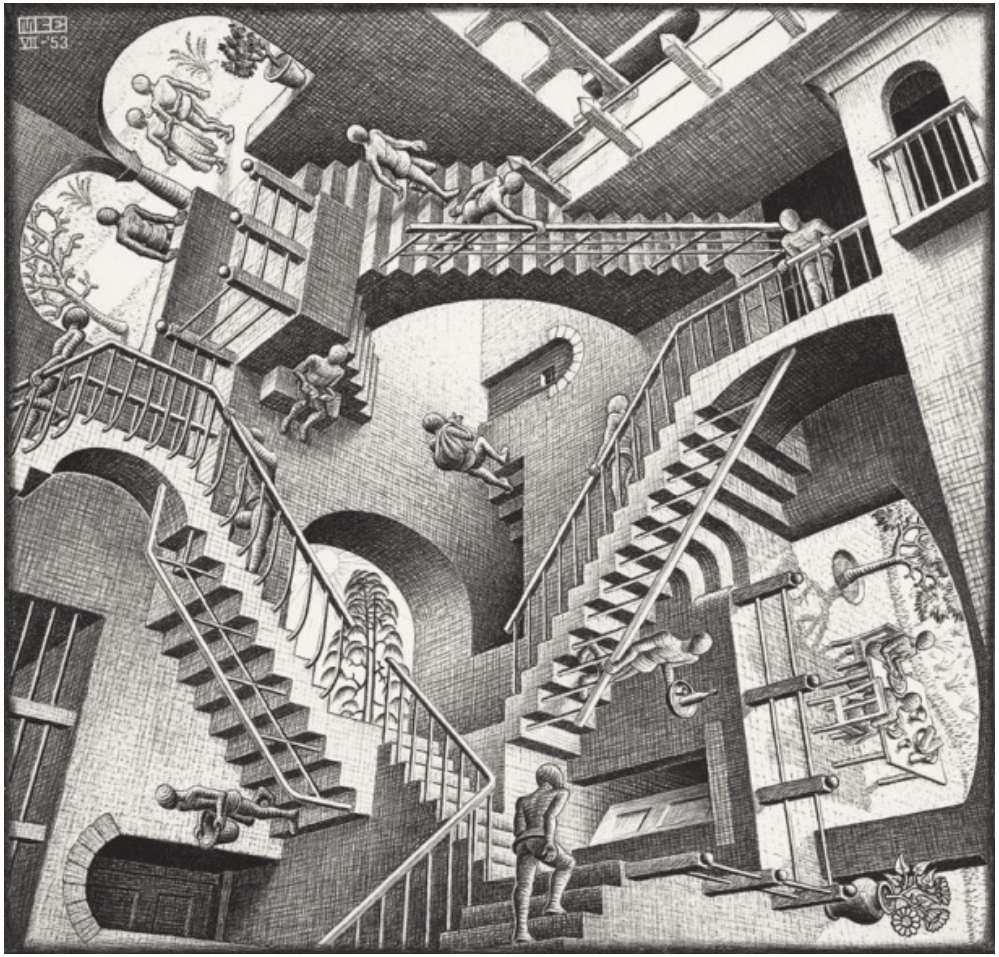
（不自相矛盾的）数学是不完备的



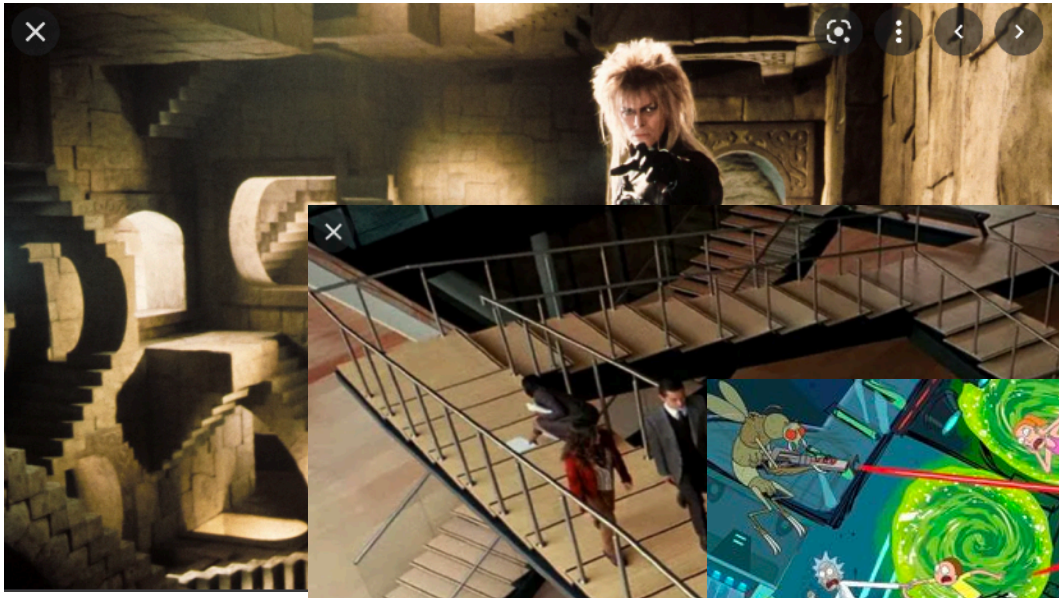
库尔特 · 哥德尔
Kurt Gödel (1906-1978)

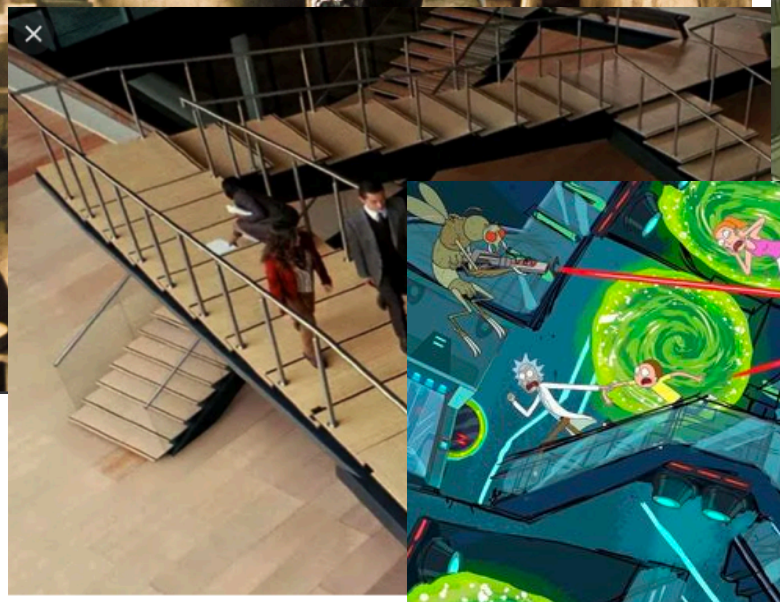
埃舍尔的画

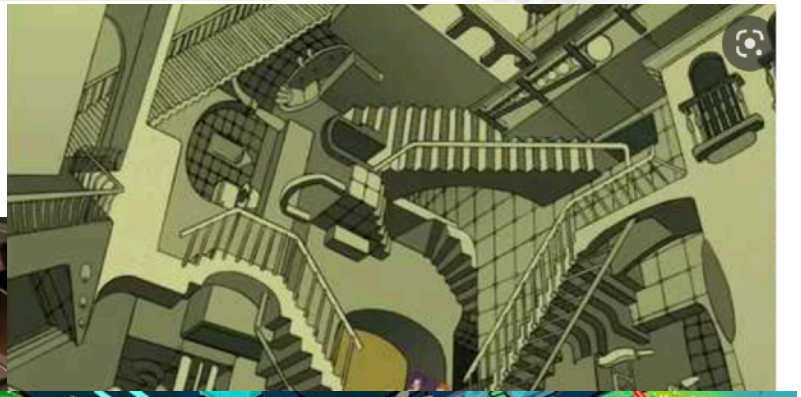




埃舍尔的画







停机问题 (1936)

程序：输入→输出　或永远不结束

“判断程序在某个输入上能否结束”

如果有这样的程序：

就能编个“捣乱程序”！

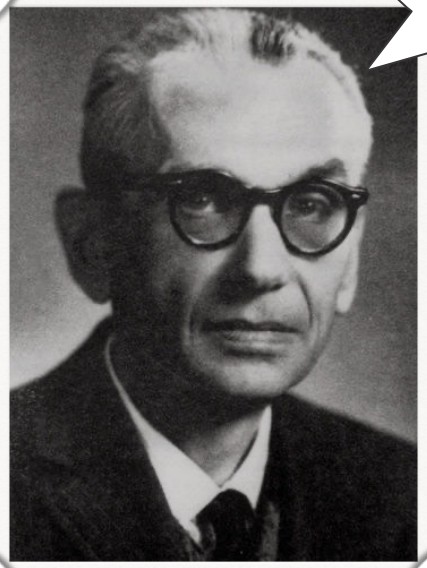
捣乱程序 输入 (程序A)：

如果 程序A 输入 (程序A) 不结束 则结束；

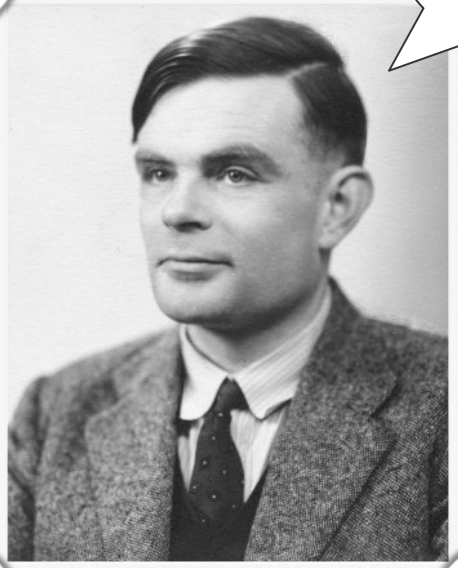
如果 程序A 输入 (程序A) 结束 则死循环。

捣乱程序 输入 (捣乱程序) 是否结束?

“存在不能被计算的问题！”



哥德尔



图灵

David Hilbert

- “希尔伯特问题”（1900）
- 23个数学难题：
- 第2问题：

  为算数建立完备的公理体系
- 第10问题：

  给出解整数多项式方程的算法

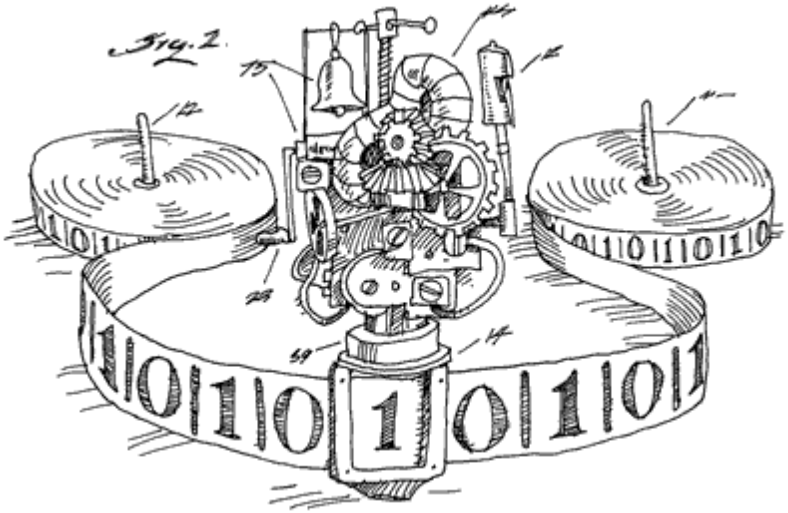
则停机问题可计算！(1970)

Matiyasevich–Robinson–Davis–Putnam定理

大卫·希尔伯特

David Hilbert (1862-1943)

Turing Machine



图灵机
Turing machine

阿兰 · 图灵
Alan Turing (1912-1954)

邱奇-图灵论题

λ演算 (λ-calculus)：

- 变量：x
- 抽象化：$(\lambda x\,.\,M)$
- 应用：$(M\ N)$

- 对于“演算 (calculus)”的抽象
- 与图灵机等价（图灵完备）

阿隆佐 · 邱奇
Alonzo Church (1903-1995)

邱奇-图灵论题

"What is computation?"

邱奇-图灵论题 (Church-Turing thesis)：

"一个问题可被有效计算当且仅当它可以被图灵机计算"

Alan Turing

“计算”正式成为了一个研究的对象

什么是计算？ “邱奇-图灵论题”

计算的界限在哪里?
是什么使一个问题难或者容易?
如何更高效的计算?

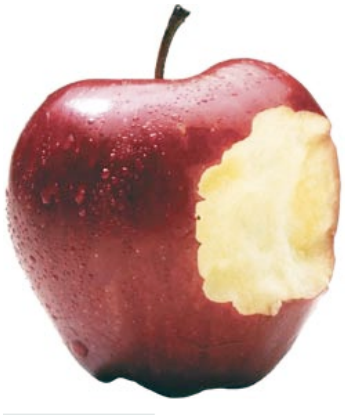
计算机科学:

一门把“问题”和“解决”
作为研究对象的学科

阿兰 · 图灵
Alan Turing (1912-1954)

Alan Turing (1912-1954)

Enigma machine
“迷”



阿兰·图灵
Alan Turing (1912-1954)

ACM
A.M. TURING AWARD

计算的界限

“有些问题是不可被计算的。”

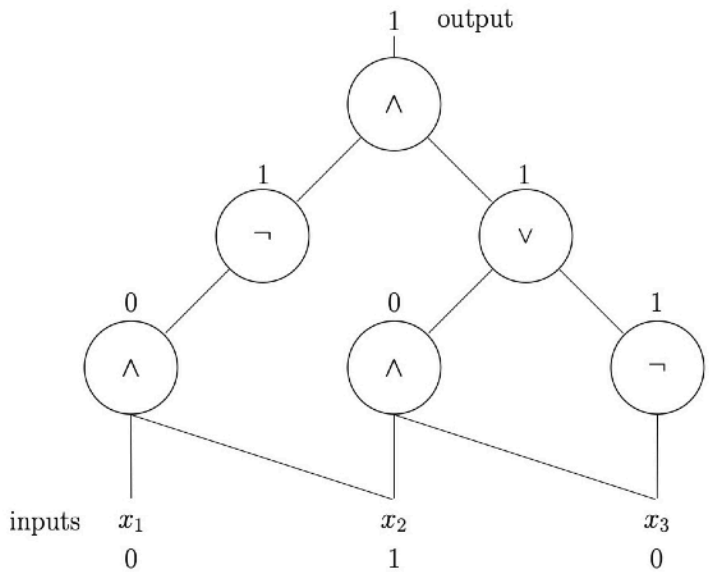
在那些可以被计算的问题当中，有的问题难、有的容易。

容易：有高效的算法　　难：没有高效的算法

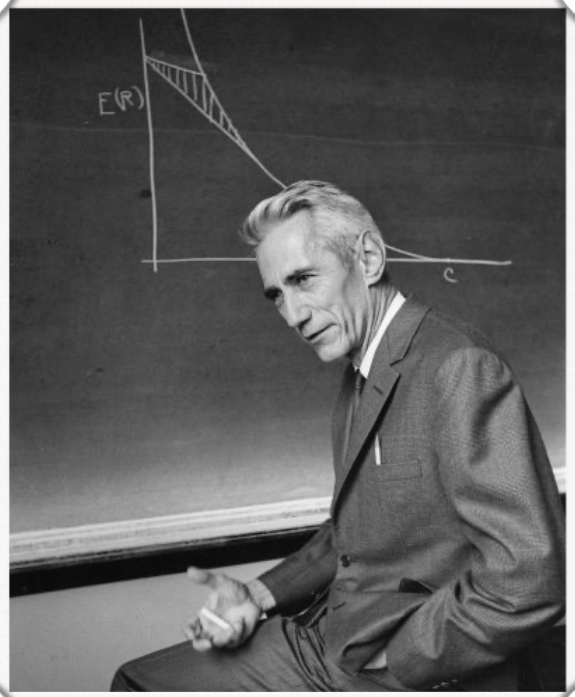
为什么难？为什么容易？

计算的界限

逻辑电路（图灵完备）



- 高效逻辑电路数量 ≪ 判定问题数量（Shannon 1949）
- **几乎所有问题都是计算困难的**



克劳德 · 香农
Claude Shannon
(1916-2001)

计算的界限

“有些问题是不可被计算的。”

在那些可以被计算的问题当中，有的问题难、有的容易。

容易：有高效的算法　　难：没有高效的算法

为什么难？为什么容易？

易与难

容易：判断两个城市之间是否连通

困难：判断能否不走重复路地访问所有省会城市一遍

什么叫容易？

多项式时间

- 多项式时间（polynomial time）：
  - 最坏输入下，（图灵机上）运行时间不超过（输入长度n的）某个多项式：$100n$，$5n\log n$，$8n^2$，n^{100}…
- Why polynomial?
  - 相对于指数级的解空间是“高效”的（n^{100}这种除外……）
  - 不受计算模型的“正常”变化的影响：
    - 纸带→随机访问内存、单带→多带、图灵机→电路、……
  - 对子程序调用“闭合”：
    - 调用一个多项式时间的子程序多项式次，仍是多项式时间

易与难

容易：判断两个城市之间是否连通

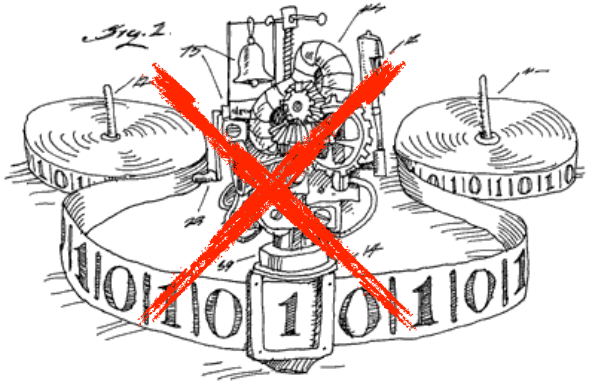
困难：判断能否不走重复路地访问所有省会城市一遍

如何说明这的确是困难的?

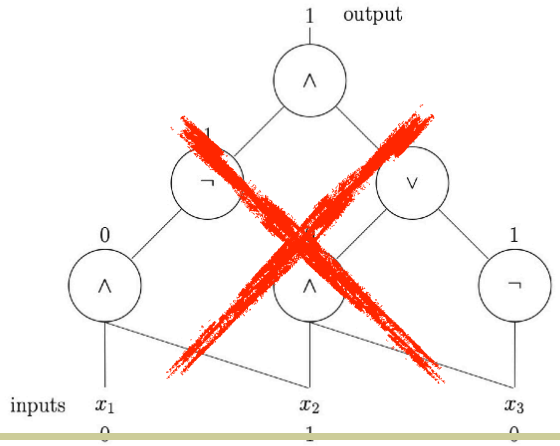
复杂性下界

判断能否不走重复路地访问
所有省会城市一遍

多项式时间图灵机：



多项式大小逻辑电路：



计算 与 验证

“在人群中找出一个人”

“有人帮你指出了那个人，你来确认是不是她”

- 为问题寻找一个解
- 给了一个解，验证它是不是有效的解

- 证明数学题
- 检查一个数学证明的正确性

可以高效计算的问题	可以高效验证的问题
P (Polynomial-time)	**NP** (Nondeterministic Polynomial-time)

易与难

容易：判断两个城市之间是否连通

困难：判断能否不走重复路地访问所有省会城市一遍

CLAY
MATHEMATICS
INSTITUTE

千禧年难题：
Millennium Problems

2000年，克雷数学研究所
列出了七大数学难题，
每道题目悬赏百万美金。

- **P vs NP**
- 霍奇猜想
- 庞加莱猜想
- 黎曼猜想
- 杨－米尔斯存在性和质量缺口
- 内维尔－斯托克斯方程解的存在性与光滑性
- 贝赫和斯维讷通－戴尔猜想

P vs NP 猜想

可以高效计算的问题　　可以高效验证的问题

$$P \neq NP$$

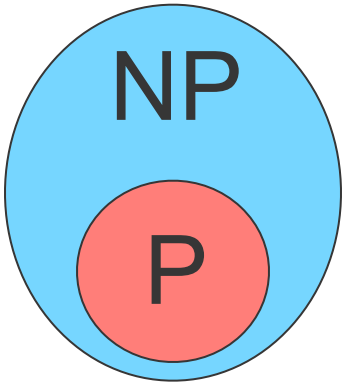
“有一些问题容易验证，计算却很难”

“创造性”是存在的：　不可被机械的过程取代

解决一个数学难题 ≠ 读懂这个解

创作出美妙的音乐 ≠ 欣赏音乐

设计建筑、规划城市 ≠ 赞赏这些设计



NP 完全性

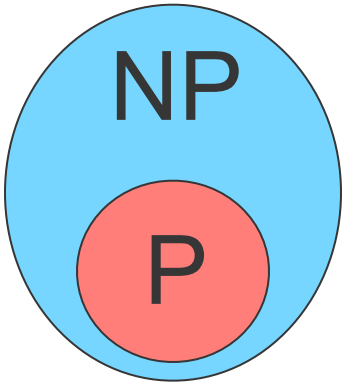
可以高效计算的问题　　可以高效验证的问题

P　　NP

NP-完全（NP-complete）问题：

有一些NP问题，它的解决可以导致所有NP问题的解决

- 判定命题逻辑公式是否可满足；
- 搬家时把杂物装入大小一样的纸箱，用最少的纸箱；
- 把大小不同的货物放到后车厢，放得越多越好；
- 把一些大小不同的砝码放在天平两端，让天平尽量平衡；
- 计划一次旅行，飞遍所有目的地城市，机票钱尽量少；
- 把人们分配到超市的不同结账口，总的结账时间最短；
- 设计医院、超市等设施的选址，尽量方便更多人；
- 玩超级马里奥、大金刚、塞尔达传说……
- ……



“会打超级马里奥，就能做数学家！”

Classic Nintendo Games are (Computationally) Hard

Greg Aloupis* Erik D. Demaine† Alan Guo†‡ Giovanni Viglietta§

February 10, 2015

Abstract

We prove NP-hardness results for five of Nintendo's largest video game franchises: Mario, Donkey Kong, Legend of Zelda, Metroid, and Pokémon. Our results apply to generalized versions of Super Mario Bros. 1–3, The Lost Levels, and Super Mario World; Donkey Kong Country 1–3; all Legend of Zelda games; all Metroid games; and all Pokémon role-playing games. In addition, we prove PSPACE-completeness of the Donkey Kong Country games and several Legend of Zelda games.

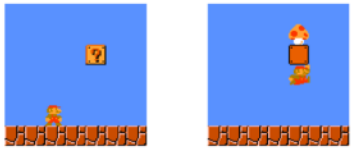


Figure 8: Left: Start gadget for Super Mario Bros. Right: The item block Mushroom

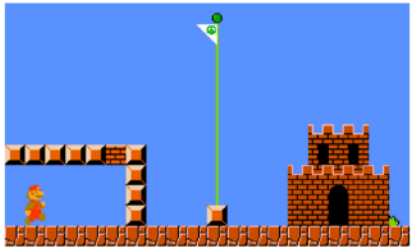


Figure 9: Finish gadget for Super Mario Bros.

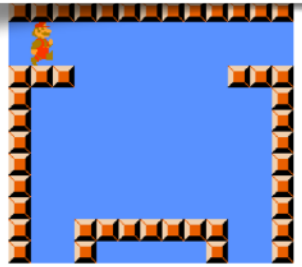


Figure 10: Variable gadget for Super Mario Bros.

ishes until it is collected by Mario.

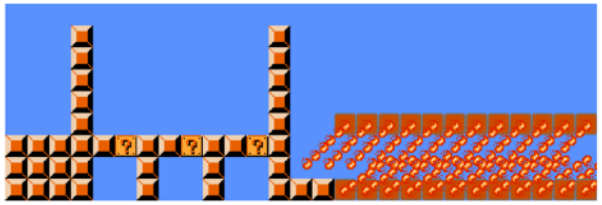
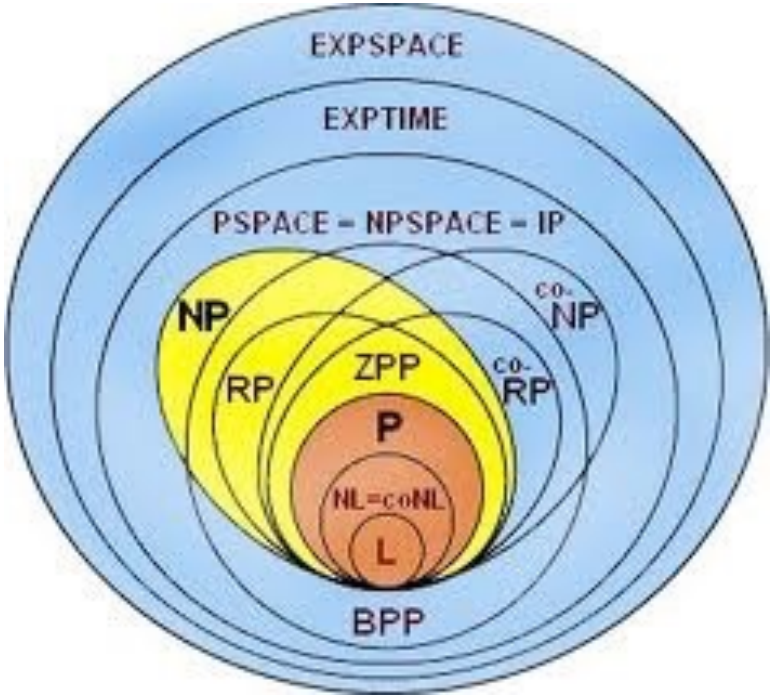


Figure 11: Clause gadget for Super Mario Bros.

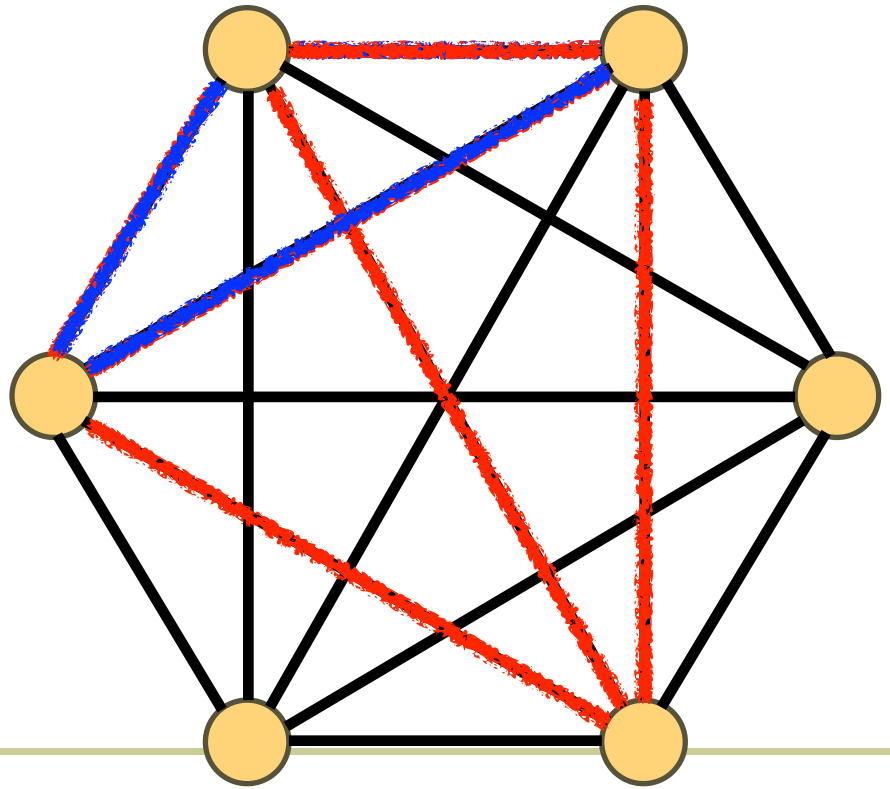
复杂类



复杂类：将问题依据其难度而归类

更难的题

“6个人的聚会，必有3个人相互认识或相互不认识。”



弗兰克·拉姆塞
Frank Ramsey
(1903-1930)

更难的题

拉姆塞数 $R(k)$：

“至少 $R(k)$ 个人的聚会，必有 k 个人相互认识或相互不认识。”

R(2)=2

R(3)=6

R(4)=18

R(5)=43~48

R(6)=102~161

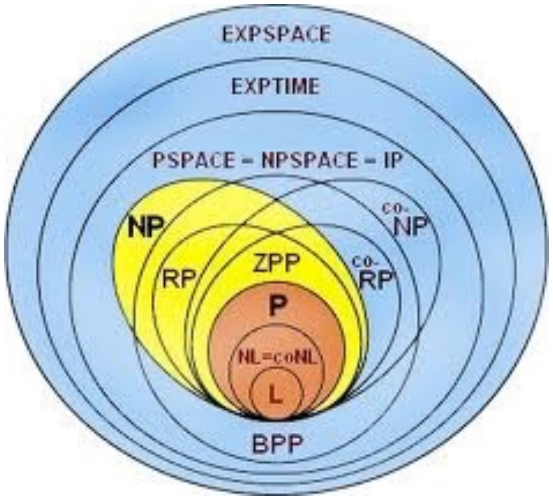
拉姆塞

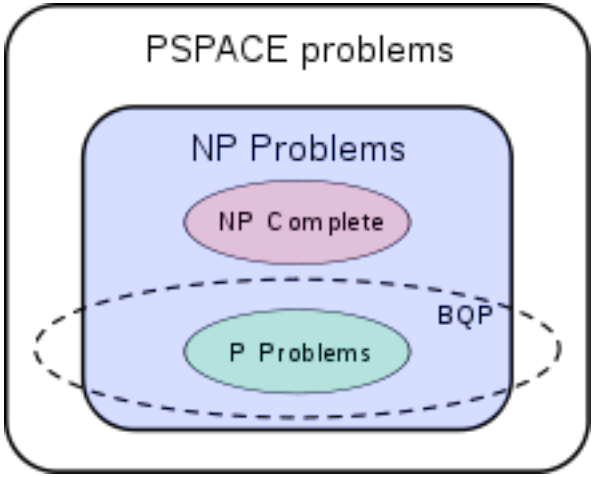
Frank Ramsey (1903-1930)

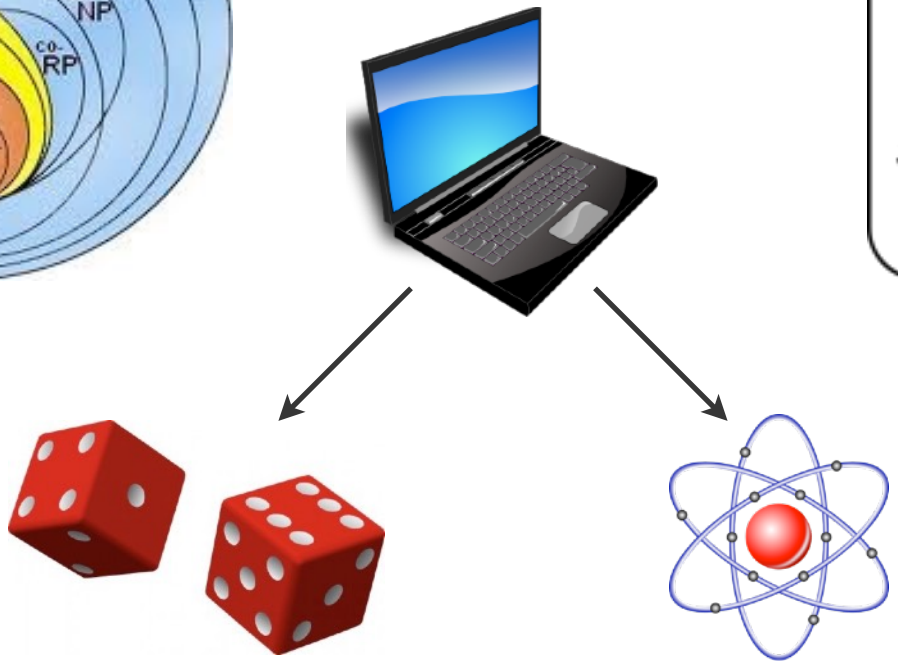
人类也许永远不会知道R(6)真实的值……

Oracle “大神”

更“强”的计算模型


EXPSPACE
EXPTIME
PSPACE = NPSPACE = IP
NP
co-NP
RP
ZPP
co-RP
P
NL=coNL
L
BPP


PSPACE problems
NP Problems
NP Complete
BQP
P Problems



计算难度

可以高效计算的问题　　可以高效验证的问题

P vs NP

“有一些问题容易验证，计算却很难”

密码学、计算机安全：

让好人做得到；

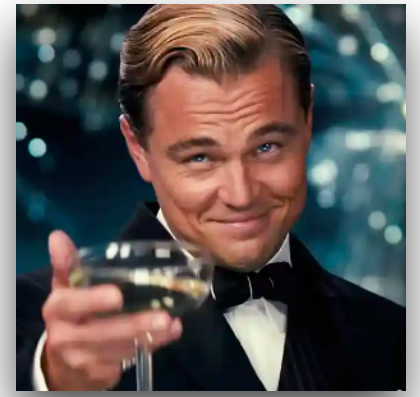
让坏人做不到；

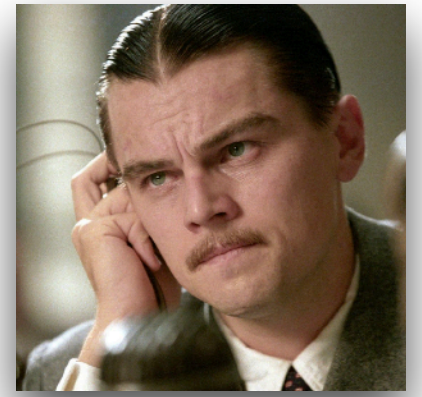
安全：对坏人很难

百万富翁难题

Yao’s Millionaires' problem:

资产 a



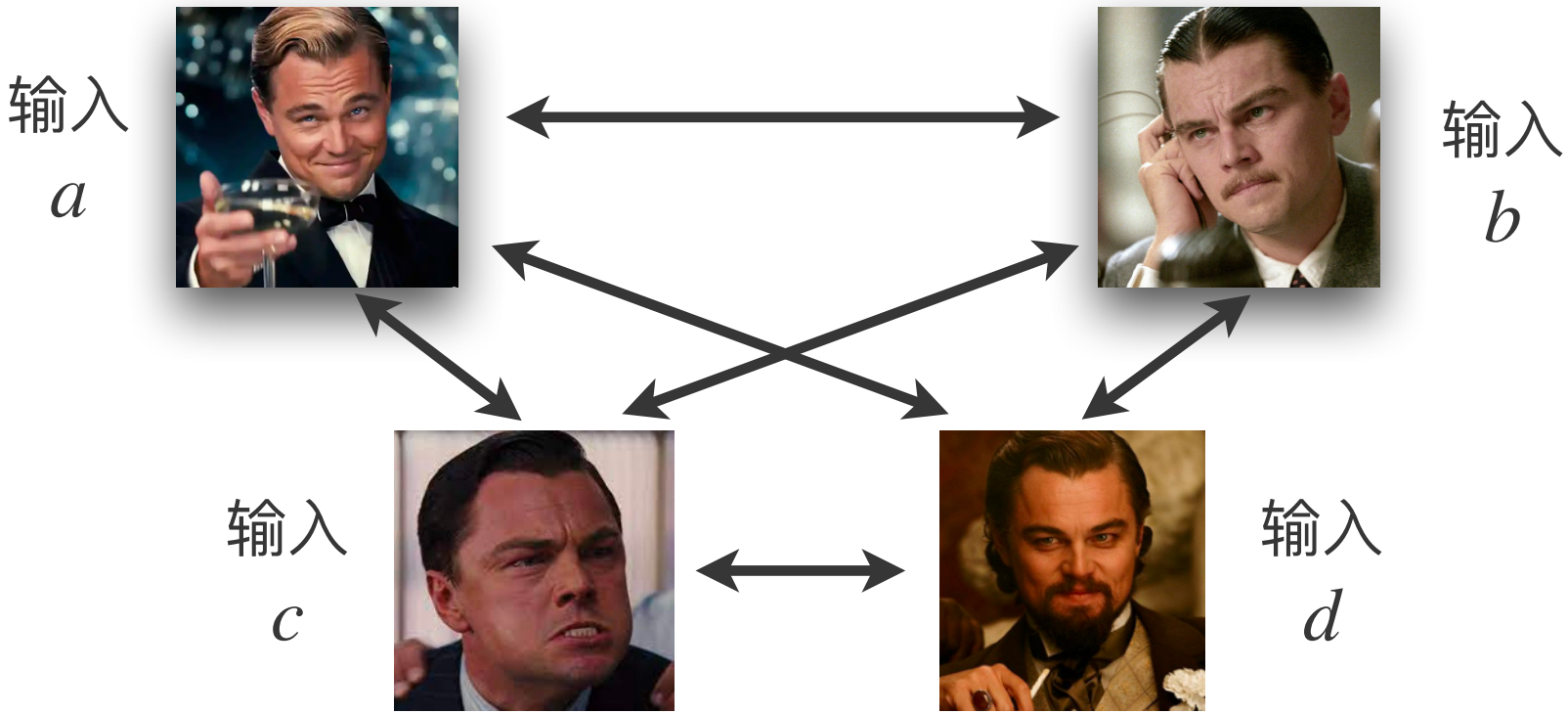


资产 b

不泄露a和b的值，判定是否有 $a > b$？

百万富翁难题

Secure multi-party computation:



不泄露任何任何更多关于输入的信息，计算 $f(a,b,c,d)$？

“零知识”证明

Zero-knowledge proofs:

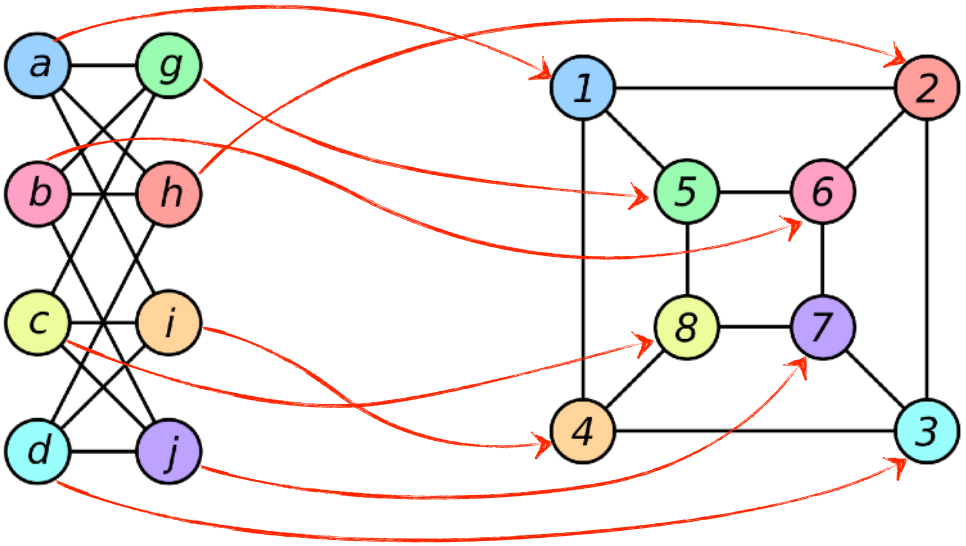
“如何向一个你不信任的人证明一个事实，
除了证明的这个事实之外，不透露任何额外信息。”

“交互式证明”
获第一届哥德尔奖
理论计算机科学最高成果奖

图同构问题

输入：图 G，图 H

输出：G 和 H 是同一个图吗?



如果G和H的确是相同的:
那么这一事实就有一个容易验证的“证明”

图同构问题

输入：图 G，图 H

输出：G 和 H 是同一个图吗?

G, H

随机在G和H挑一个图

并且随机打乱点的顺序

返回G和H中那个正确的图

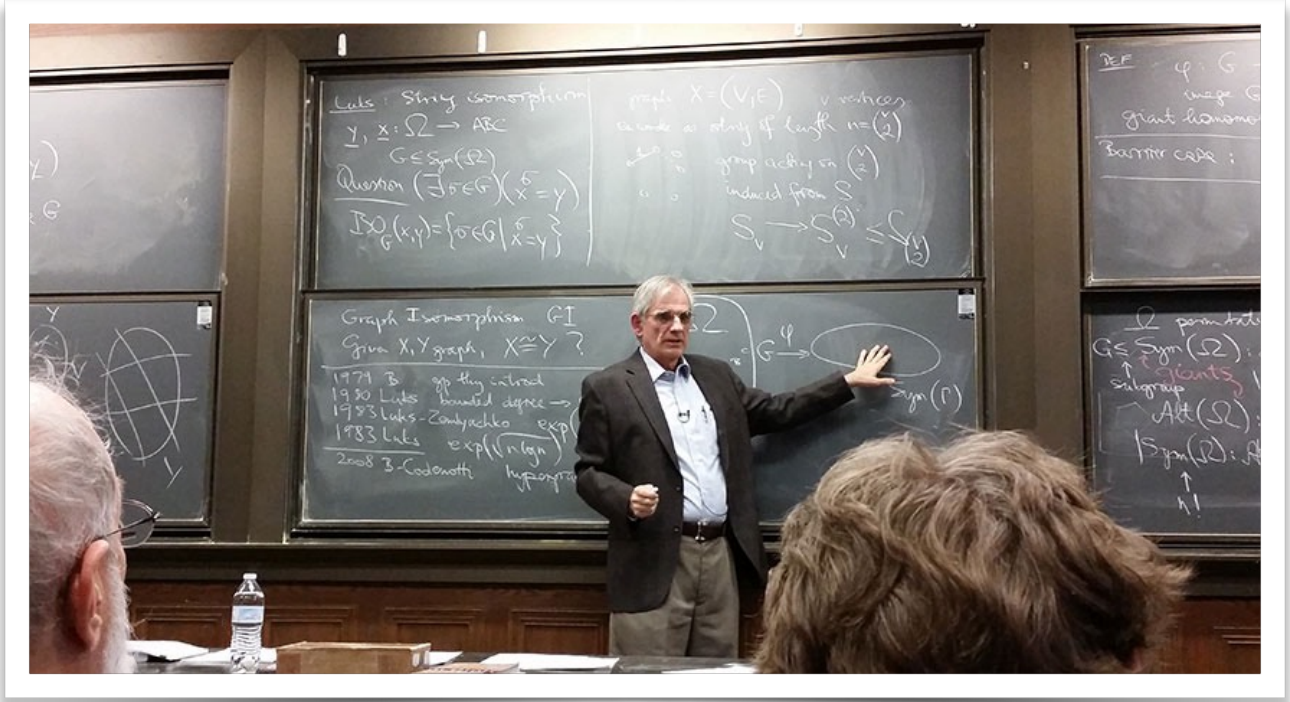
G, H

“亚瑟–梅林 博弈” （Arthur-Merlin game）

（László Babai，1985）

图同构问题

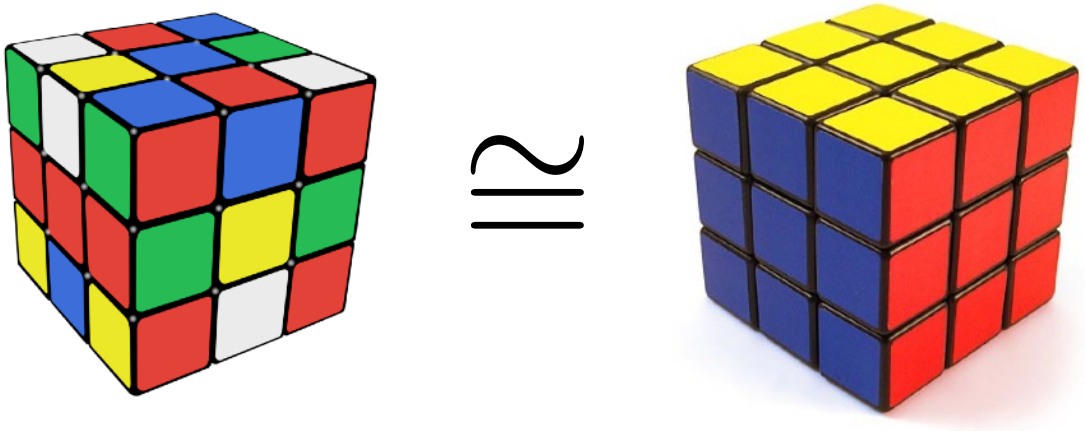
Babai 2015: 图同构问题的准多项式（$n^{\text{polylog}(n)}$）时间算法!



László Babai (born 1950)

同构问题

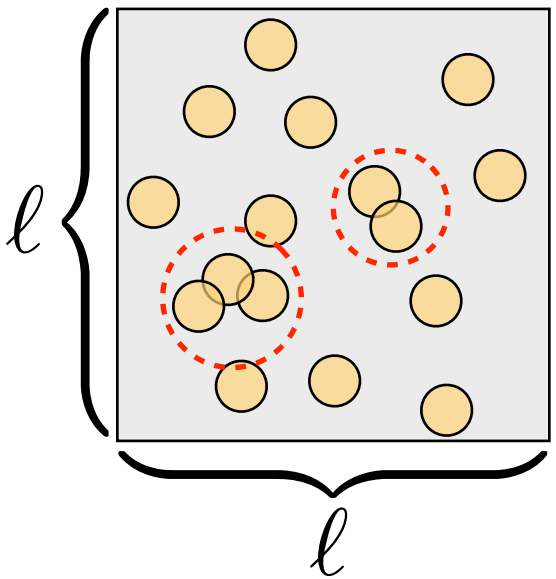
输入：字符串 x 和 y，以及一类对称（“群”）
输出：x 和 y 在这一对称下是相同的吗?



Babai 2017: 该问题有准多项式时间算法!

相变（phase transition）

n枚硬币被随机扔在方形地毯上



求没有任意两枚硬币相互重叠的概率

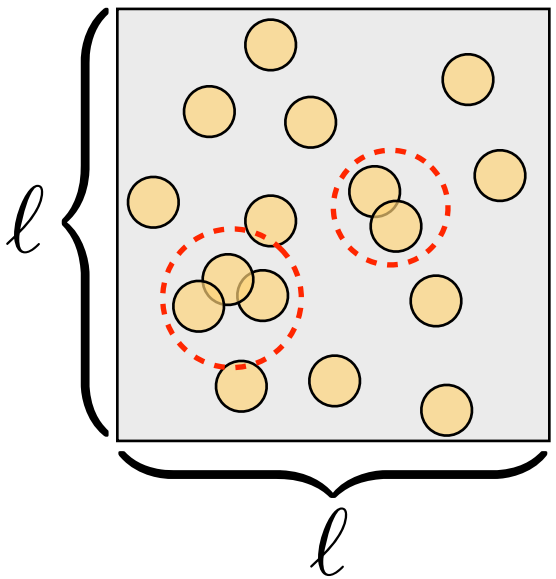
Gian-Carlo Rota
(1932-1999)

- 1维：容易求解
- 2维：Nothing is known about this problem.

(1979)

相变（phase transition）

*n*枚硬币被随机扔在方形地毯上

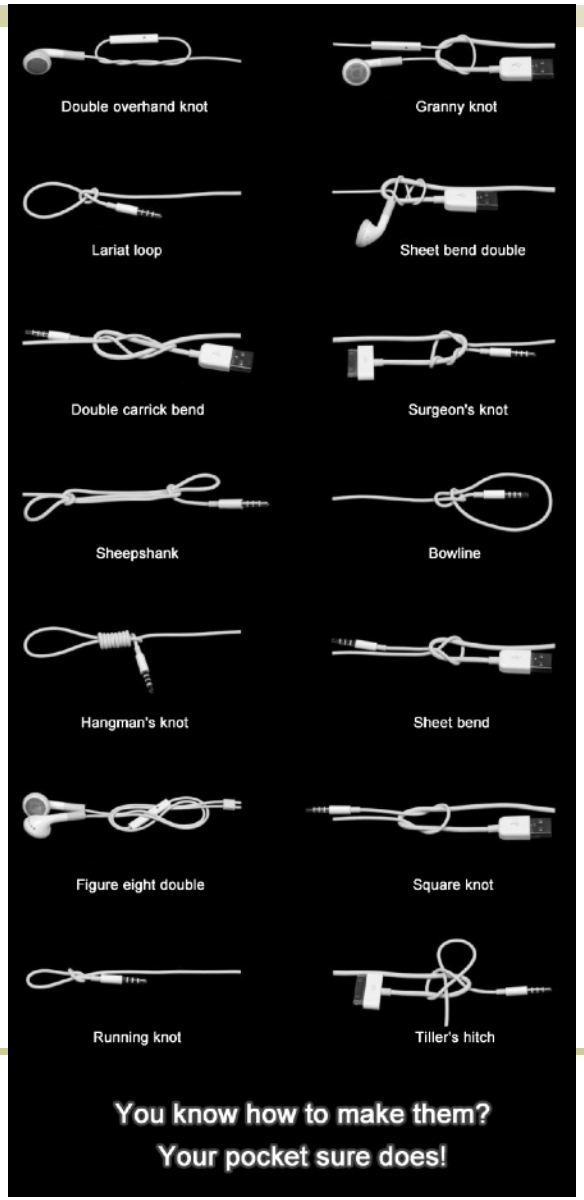


求没有任意两枚硬币相互重叠的概率

Gian-Carlo Rota
(1932-1999)

硬球(hard sphere)模型：This problem is one of the most important problems of statistical mechanics. If we could answer it we would know, for example, why water boils at 100°C, on the basis of purely atomic computations.

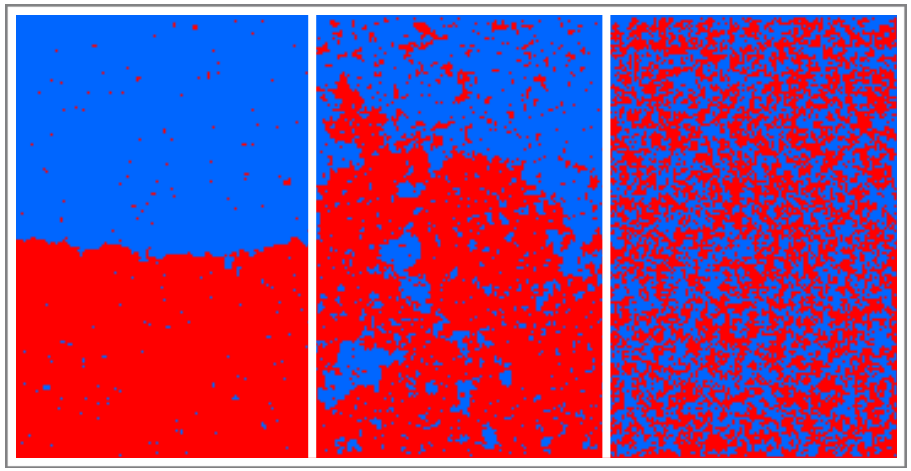
相变（phase transition）



相变（phase transition）

伊辛模型（Ising model）：

<临界温度　　临界温度　　>临界温度



计算	物理
变量	粒子
取值	状态
局部约束	相互作用
解	配置
典型解	吉布斯态
生成典型解	求解模型

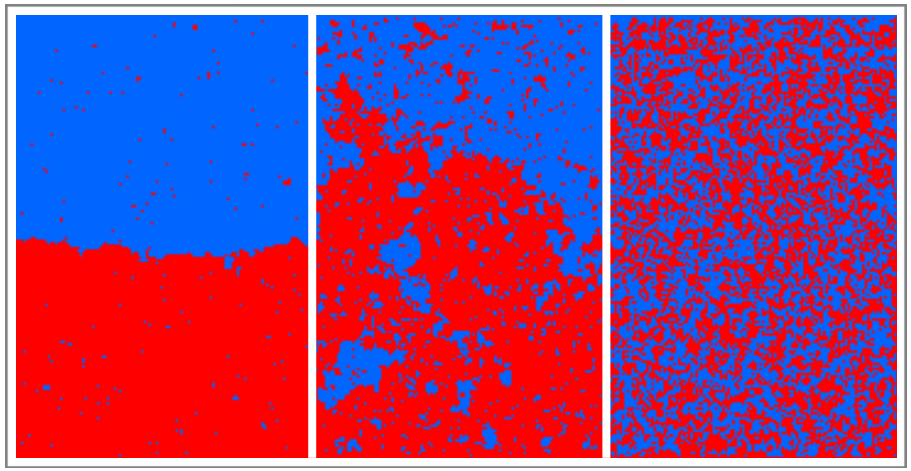
“求解”模型：Ising (1924), Onsager (1944), Lee-Yang (1952)

Kaufman, Onsager, Yang, Kac, Ward, Potts, Montroll, Hurst, Green, Kasteleyn, McCoy, Wu, Vdovichenko, Fisher, Baxter, ...

相变（phase transition）

伊辛模型（Ising model）：

<临界温度　　临界温度　　>临界温度



计算	物理
变量	粒子
取值	状态
局部约束	相互作用
解	配置
典型解	吉布斯态
生成典型解	求解模型

生成典型解：

- <临界温度：NP-难（Sly 2010；Sly-Sun 2012）
- >临界温度：多项式时间易解（Chen-Feng-Yin-Zhang 2021, 2022）

指导教师 (按姓氏字母排序)

博士·清华大学
国家优青（海外）
黄棱潇 副教授

博士·普林斯顿大学
教育部长江学者
栗 师 教授

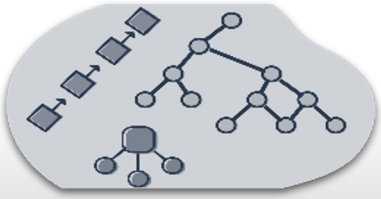
博士·加州大学伯克利分校
国家青年高层次人才
刘景铖 副教授

博士·新加坡国立大学
国家青年高层次人才
姚鹏晖 副教授

博士·耶鲁大学
国家优青·CCF青年科学家
尹一通 教授

博士·新加坡国立大学
德国弗莱堡大学博士后
郑朝栋 助理教授

研究方向 (按英文首字母排序)



算法理论

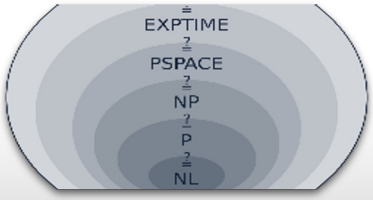
Algorithm

- 近似算法、随机算法、在线算法、组合优化、采样与计数

大数据理论

Big Data

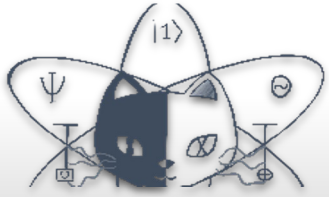
- 数据摘要与分析、数据结构、并行与分布式计算、计算隐私与公平



计算复杂性

Complexity

- 计算复杂性、通信复杂性、计算相变、去随机化、布尔函数分析



量子计算

Quantum

- 量子算法、量子计算复杂性、量子信息论、量子分布式计算

谢谢！