

应用课堂

如影随形的时空大数据应用



童咏昕

计算机学院 软件开发环境国家重点实验室

yxtong@buaa.edu.cn

汇报提纲

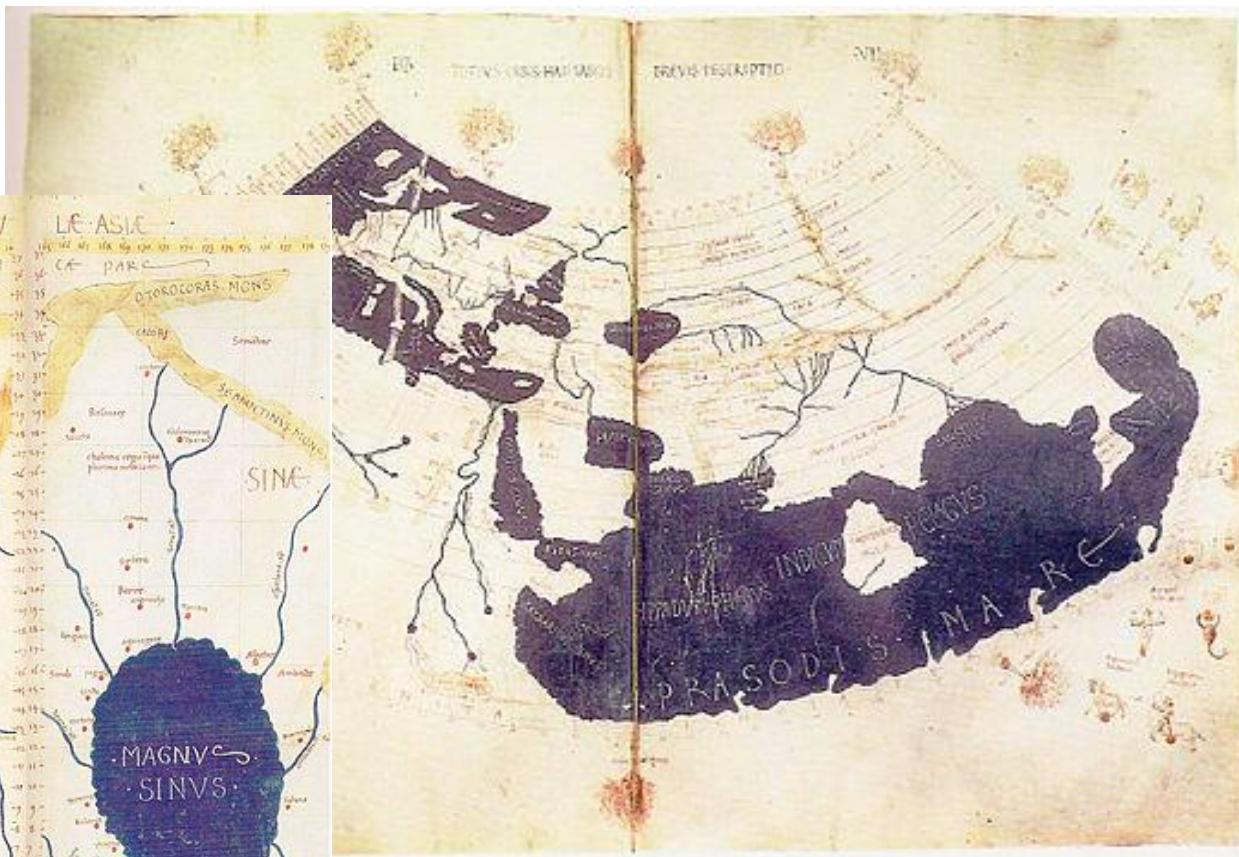
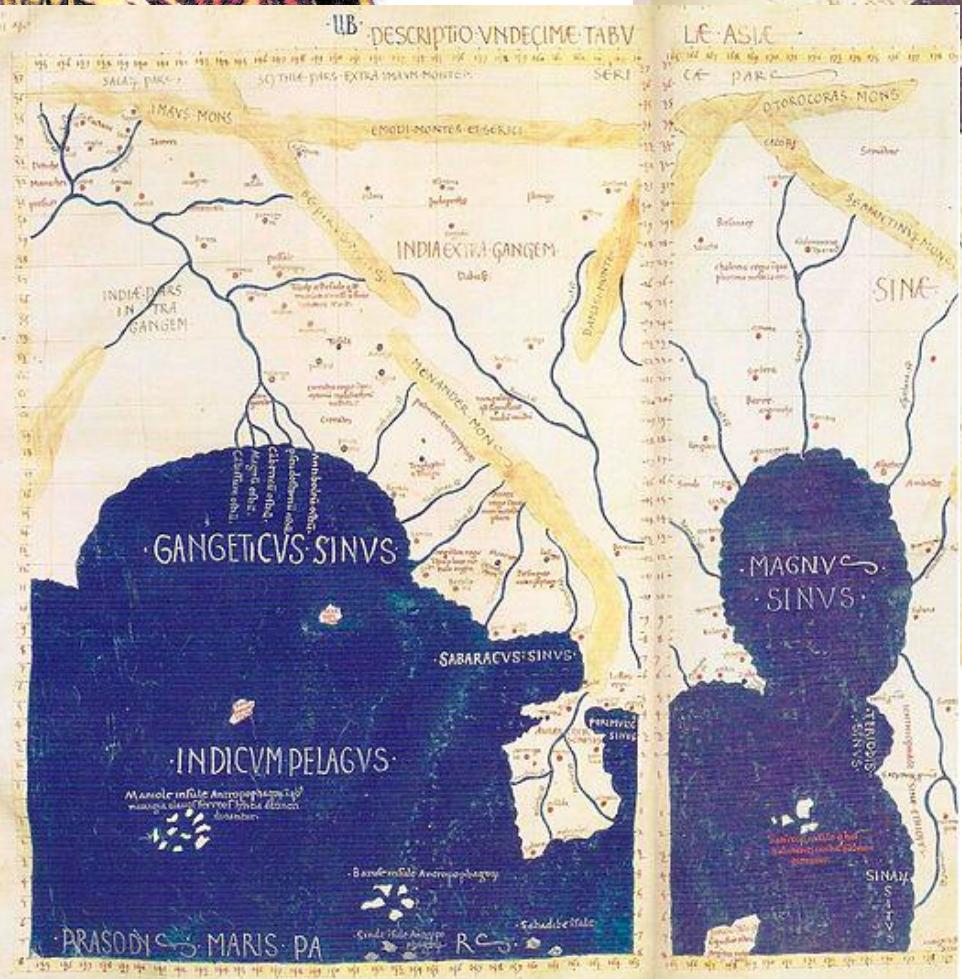
- 历史回顾
- 基础操作
- 典型应用
- 总结展望

汇报提纲

- 历史回顾
- 基础操作
- 典型应用
- 总结展望



克劳狄乌斯·托勒密(AD90–AD168)

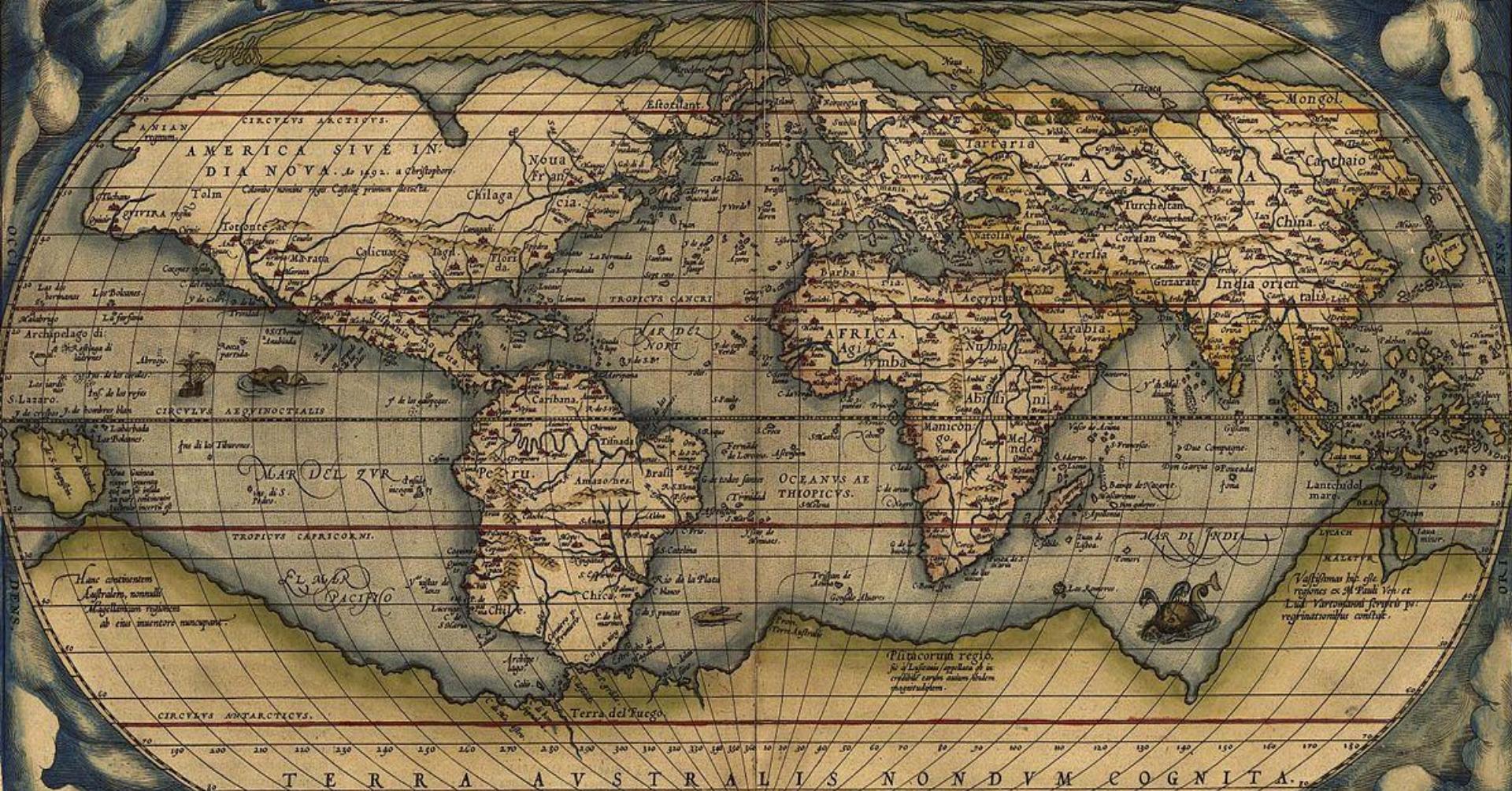


穆罕默德·伊德里西(1099-1165)



TYPVS ORBIS TERRARVM.

SEPTEN. RIO.



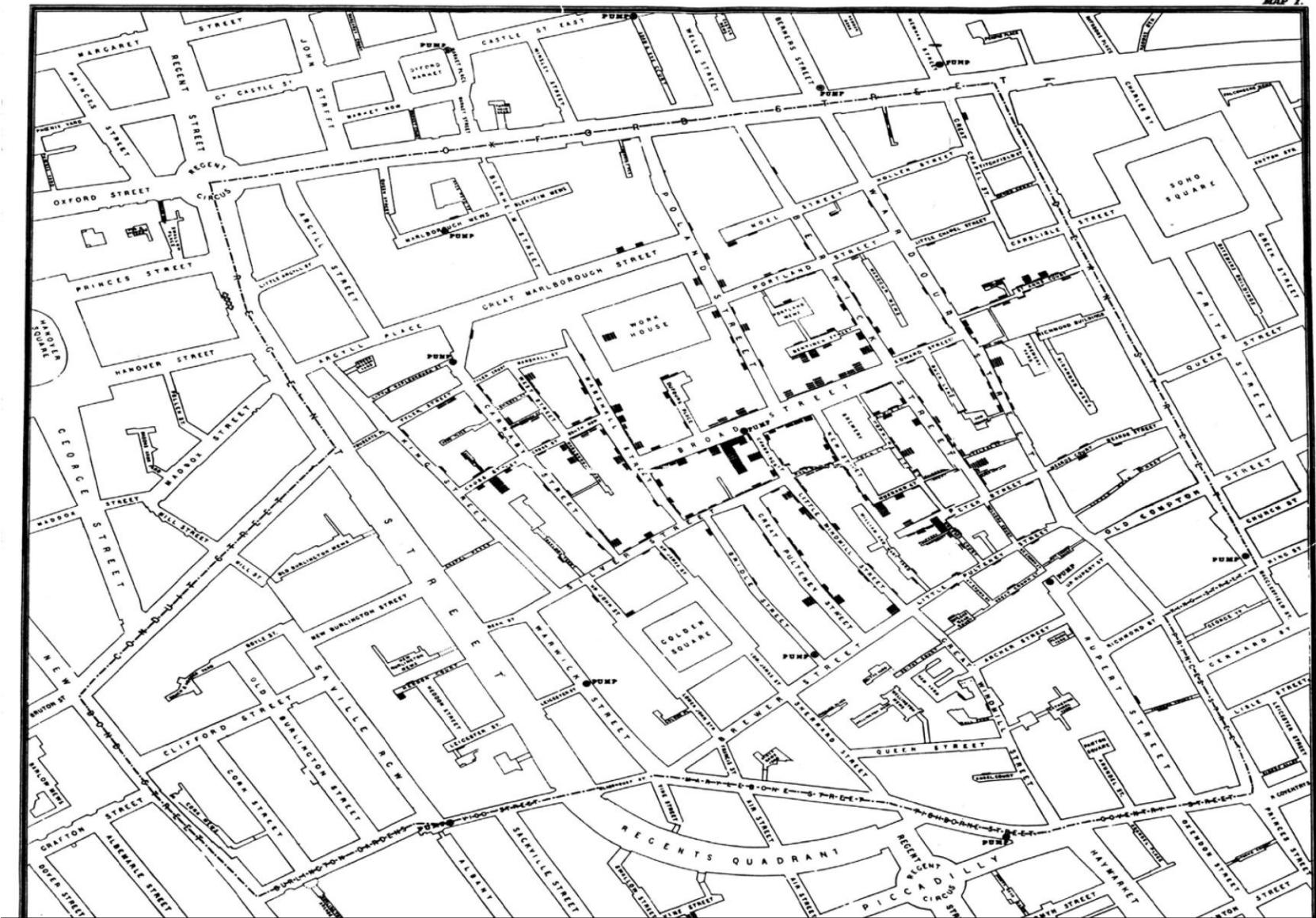
MERIDIES.

QVID EI POTEST VIDERI MAGNUM IN REBUS HUMANIS, CVI AETERNITAS
OMNIS, TOTIVSQUE MVNDI NOTA SIT MAGNITVDO. CICERO:

ARGONAVTICA.

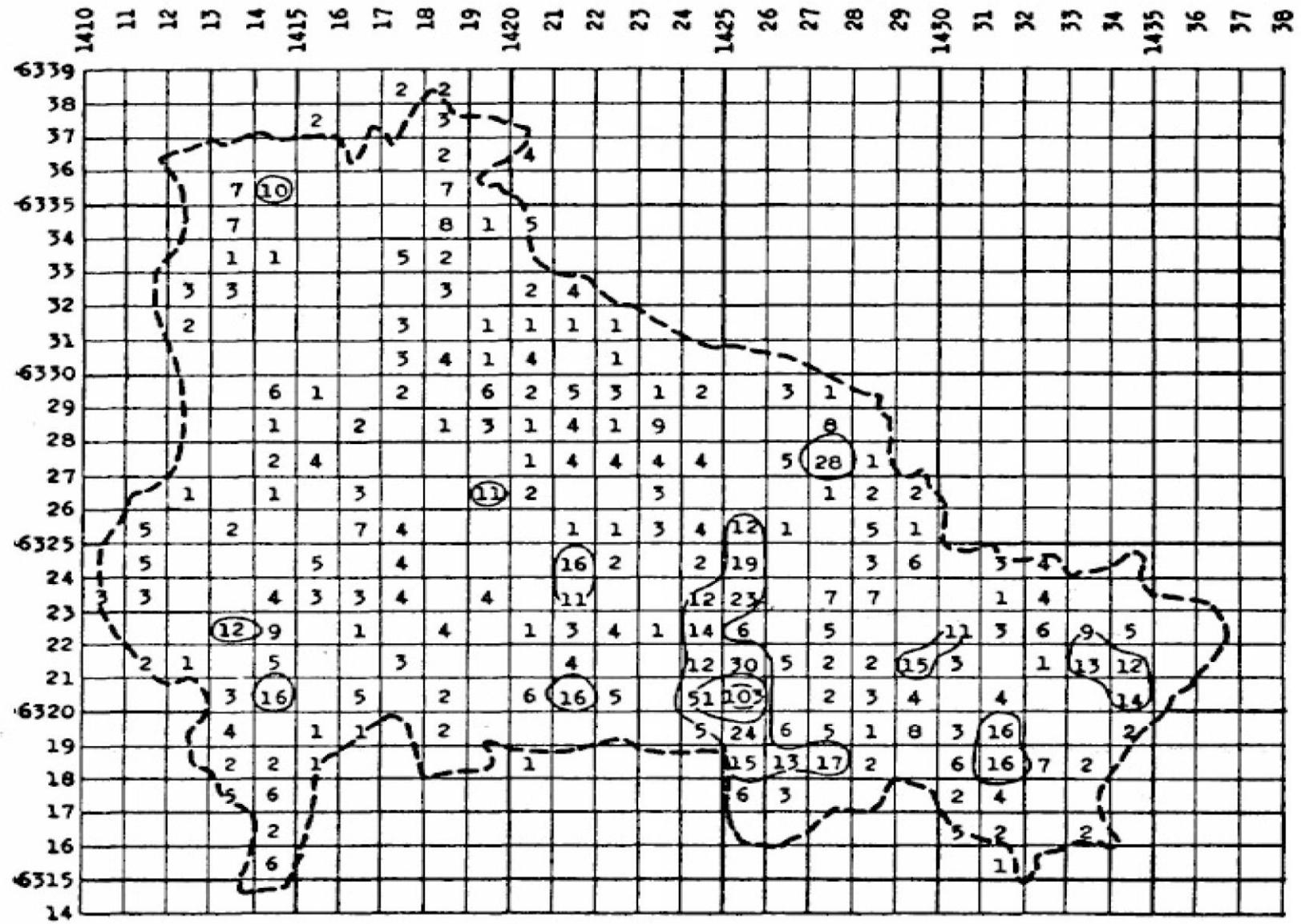


历史回顾



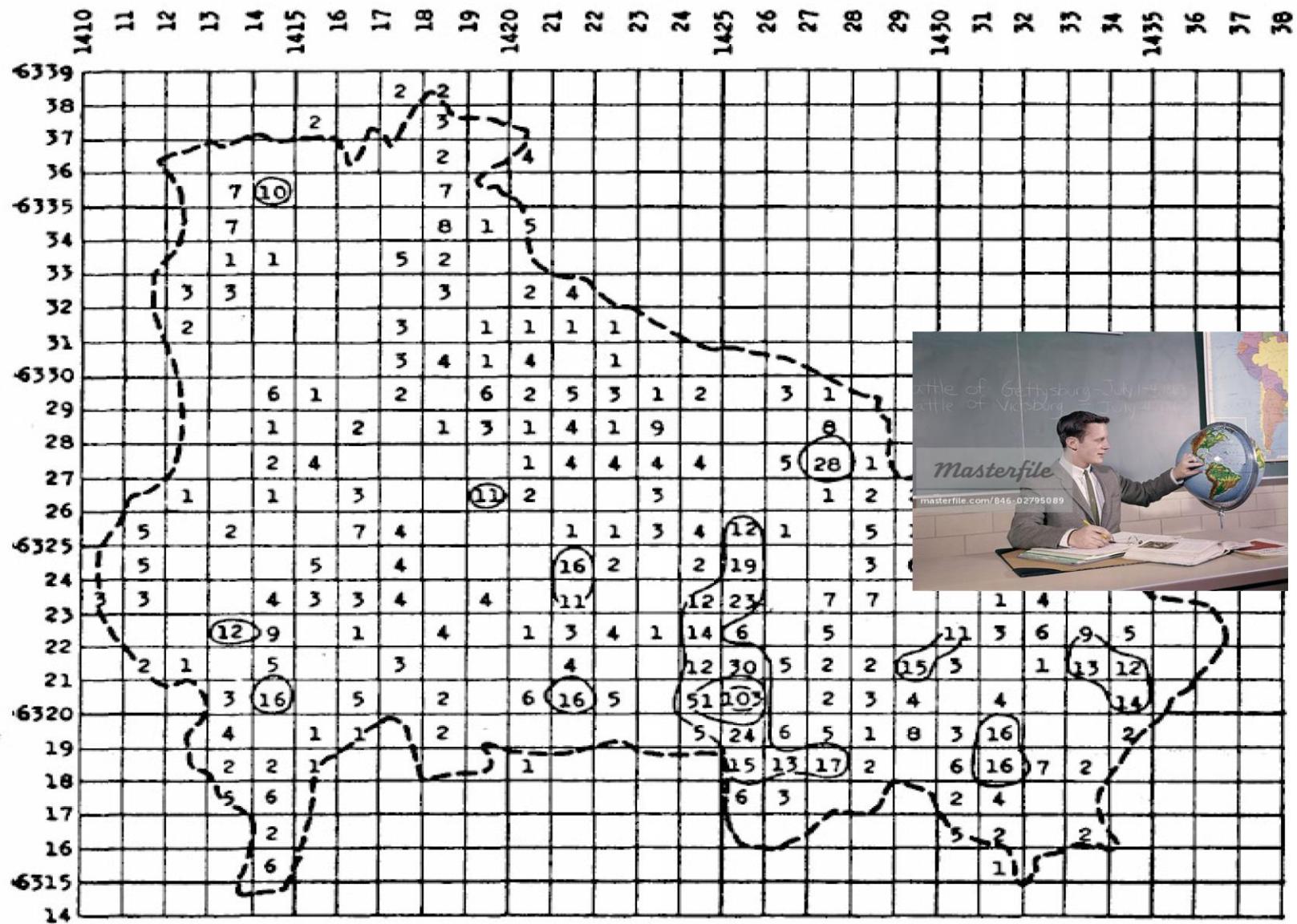
1854年伦敦霍乱传染情况

历史回顾



1940年瑞典穆海达市教区15岁以下儿童分布

历史回顾



1940年瑞典穆海达市教区15岁以下儿童分布



计算机技术虽然酷，
但是可以把它应用到
我的领域吗？



但是那不是为我
量身定做的，不能
够直接使用。



第一台计算机ENIAC
诞生了！

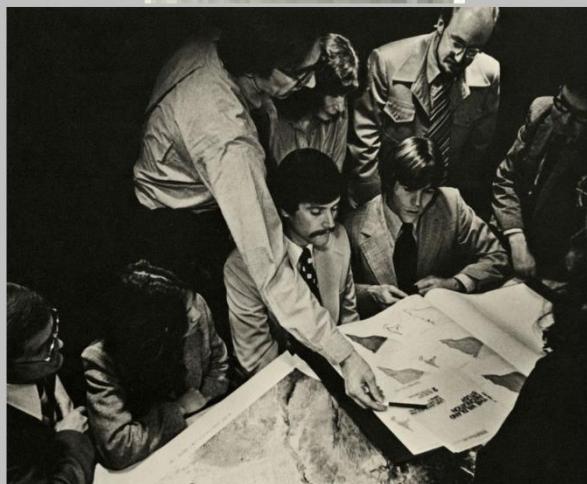




让我来了解你的
需求。



1969



让我来学习你拥有的
技术。



美国环境系统
研究所



我有大量数据，你的技术对我没有帮助。



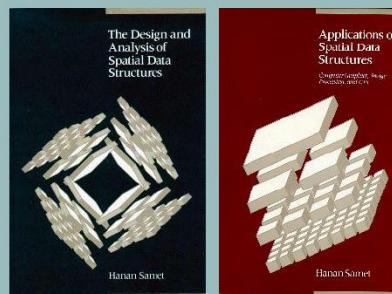




让我来了解你的
需求。



让我来学习你拥有的
技术。

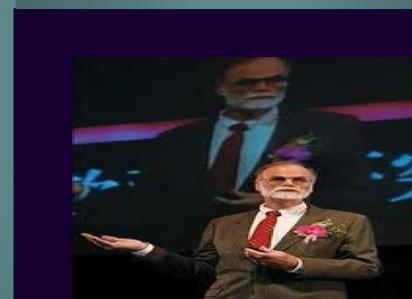
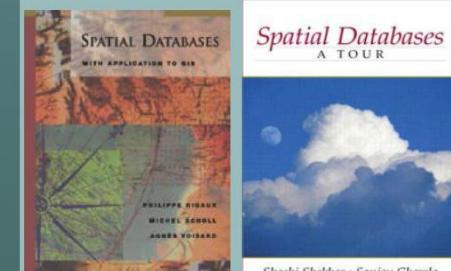
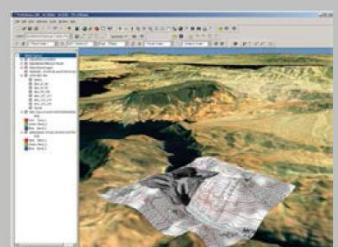




再一次寻求帮助！
我有大量数据，但
你们的技术无法帮
助我。



看起来数据库管理
技术无法适应继续
扩展。





让我来学习你拥有的技术。

facebook

Map Reduce

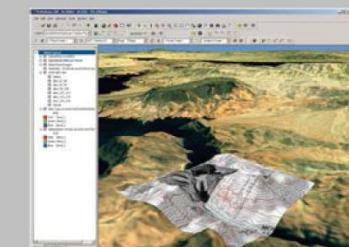
hadoop

HIVE
Spark

amazon
web services™



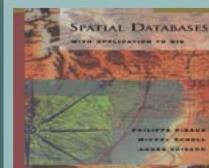
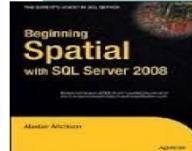
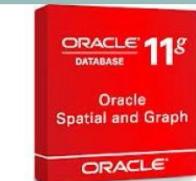
Arc
ESRI GIS



idrisi Q A



PostGIS
Geospatial Objects for PostgreSQL



ORACLE®

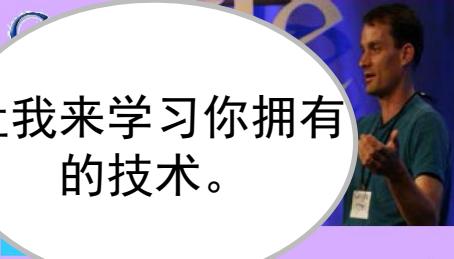
IBM DB2®

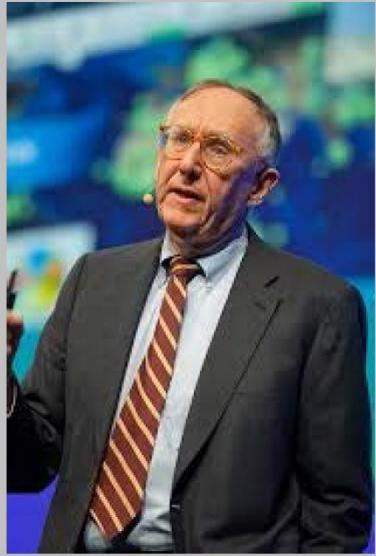
Microsoft SQL Server®

PostgreSQL
The world's most advanced open source database

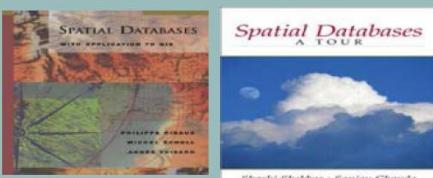
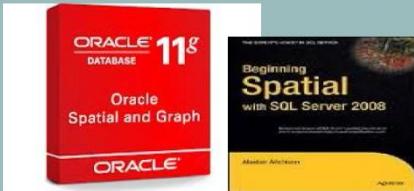
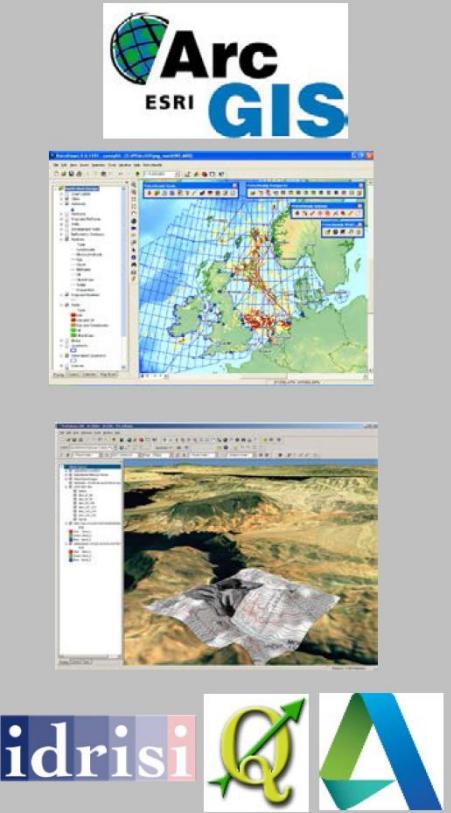
MySQL

SYBASE® An SAP Company





时空大 数据的 时代

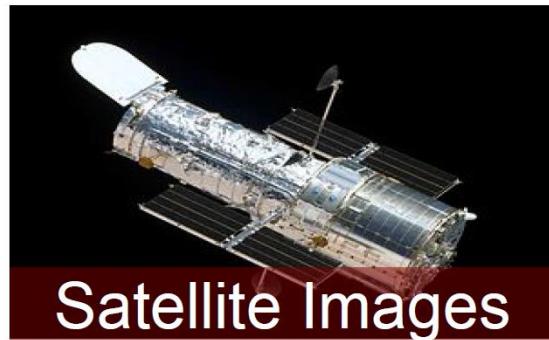


历史回顾

● 时空大数据时代的到来



Smart phones



Satellite Images



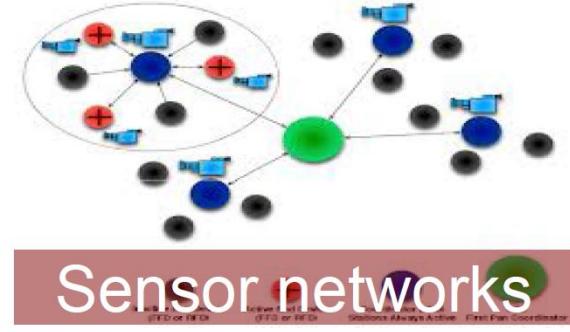
Traffic data



Medical data



VGI



Sensor networks



Geotagged Microblogs

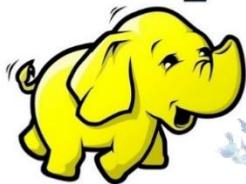


Geotagged pictures

历史回顾

- ~~时空大数据系统~~

hadoop



A P A C H E
HBASE

Spark



oozie



APACHE
STORM™



TEZ

Hortonworks

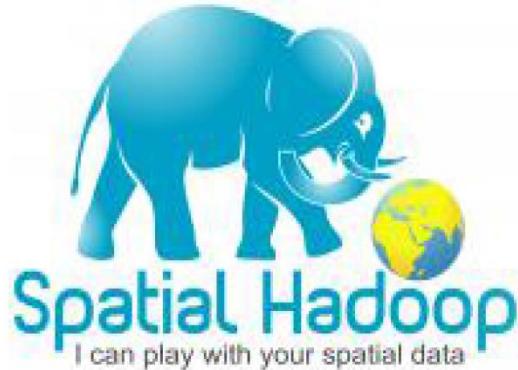


Asterix^{*}**DB**

oozie

历史回顾

• 时空大数据系统



Hadoop-GIS
Spatial Big Data Solutions



GeoSpark

汇报提纲

- 历史回顾
- 基础操作
- 典型应用
- 总结展望

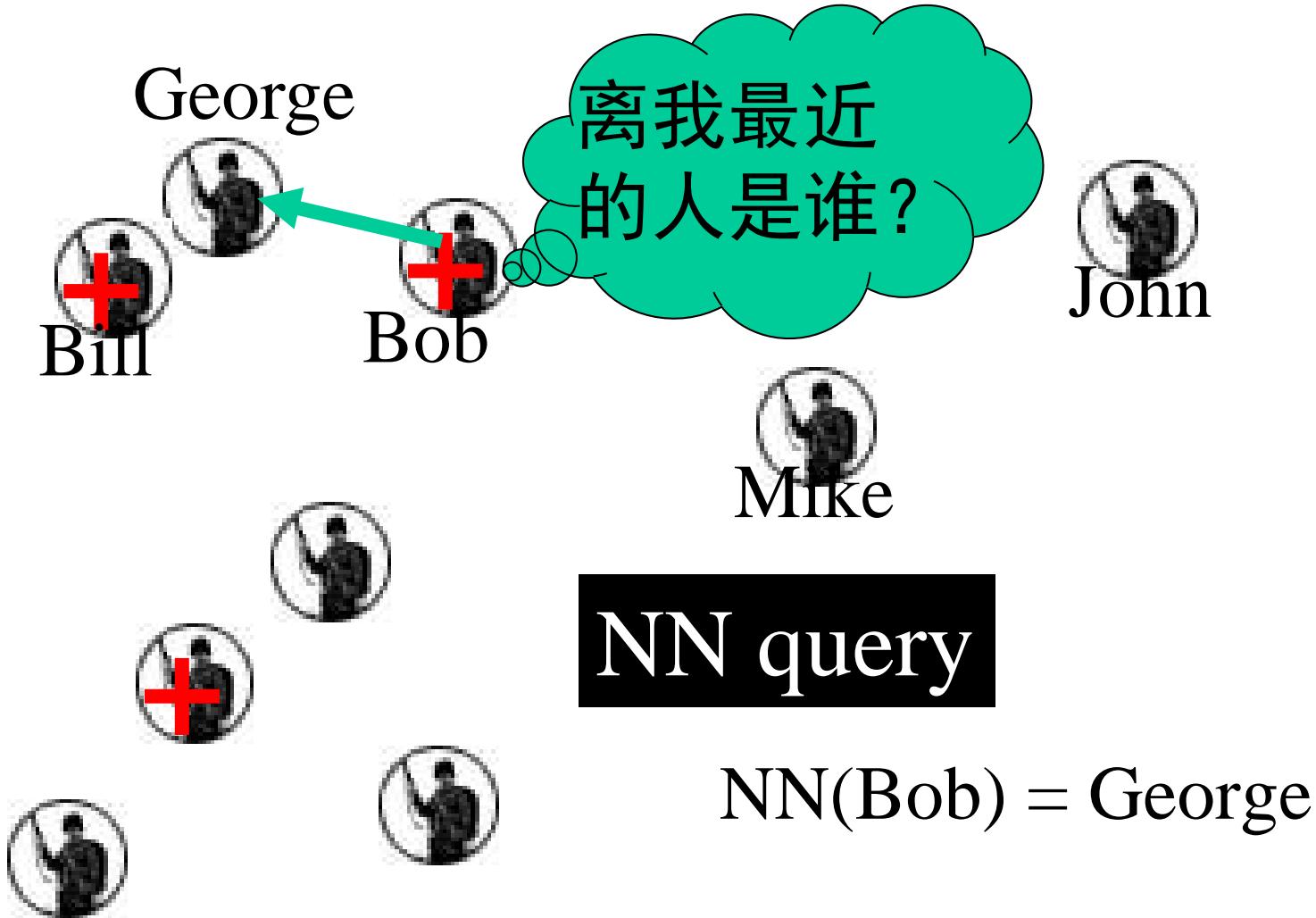
基础操作

- 最近邻居查询
- 反转最近邻居查询
- 范围查询
- 空间集成查询
- 最短路径查询
- 最优选址查询

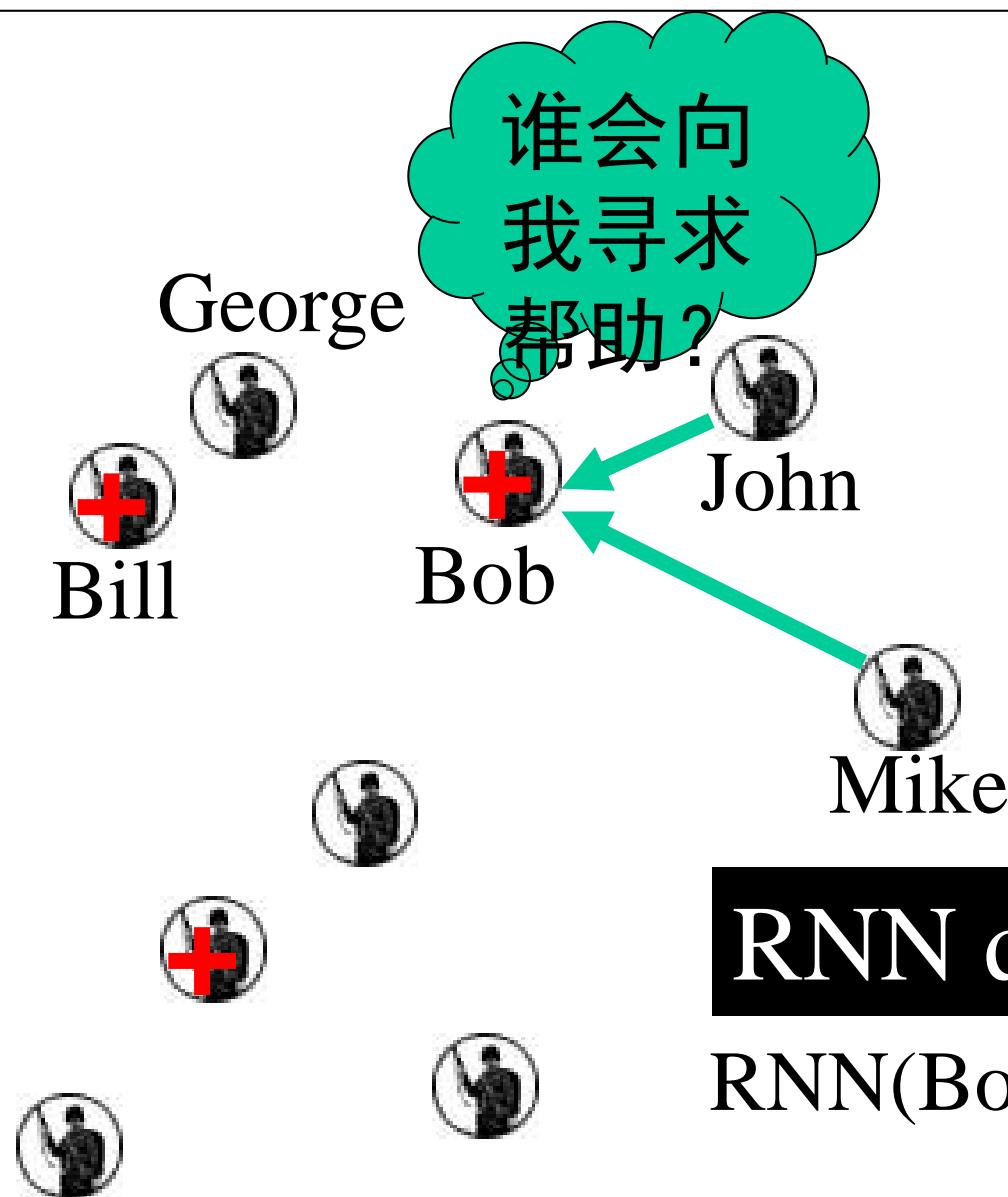
汇报提纲

- 历史回顾
- 基础操作
- 典型应用
- 总结展望

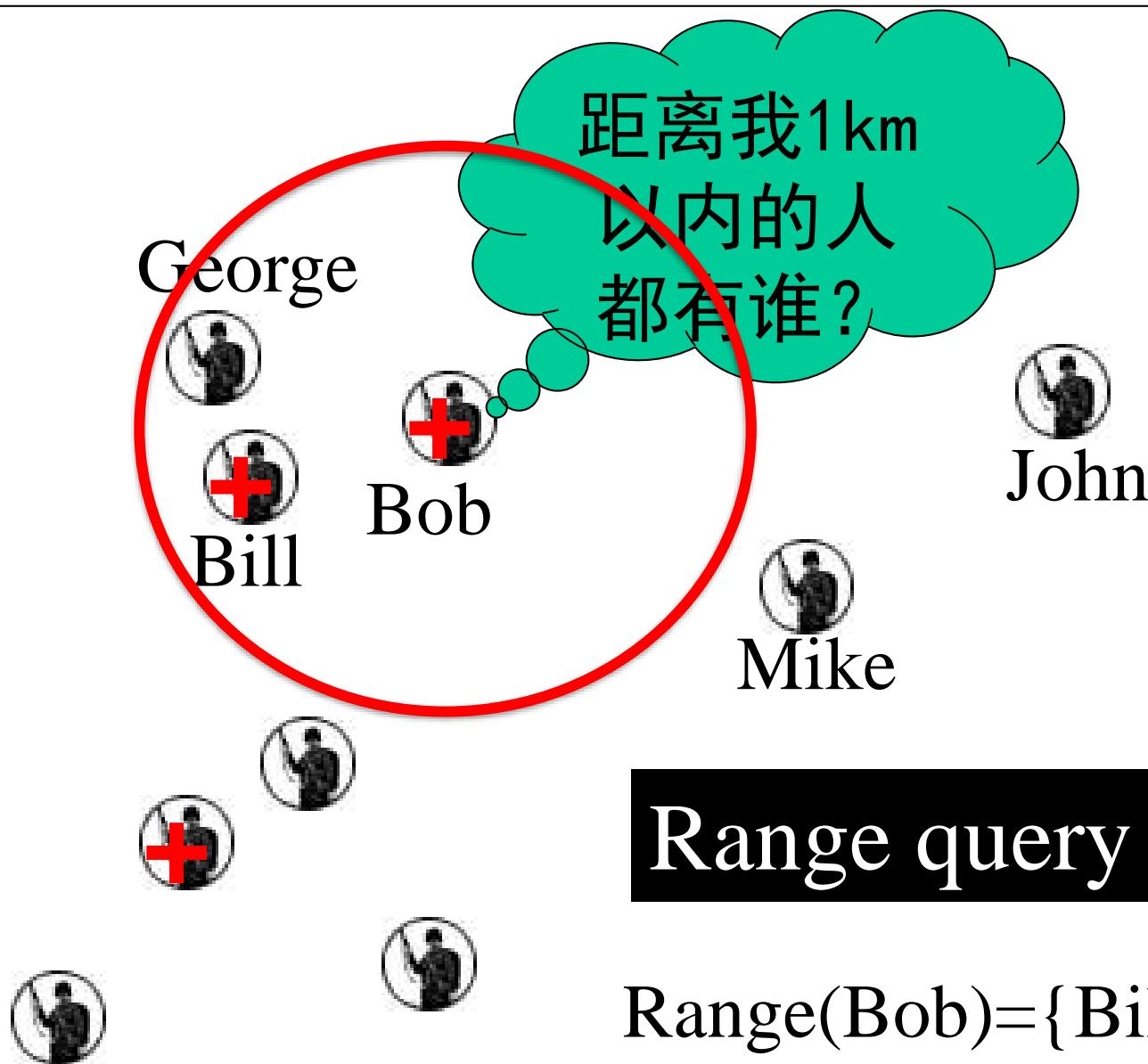
最近邻查询 (NN)



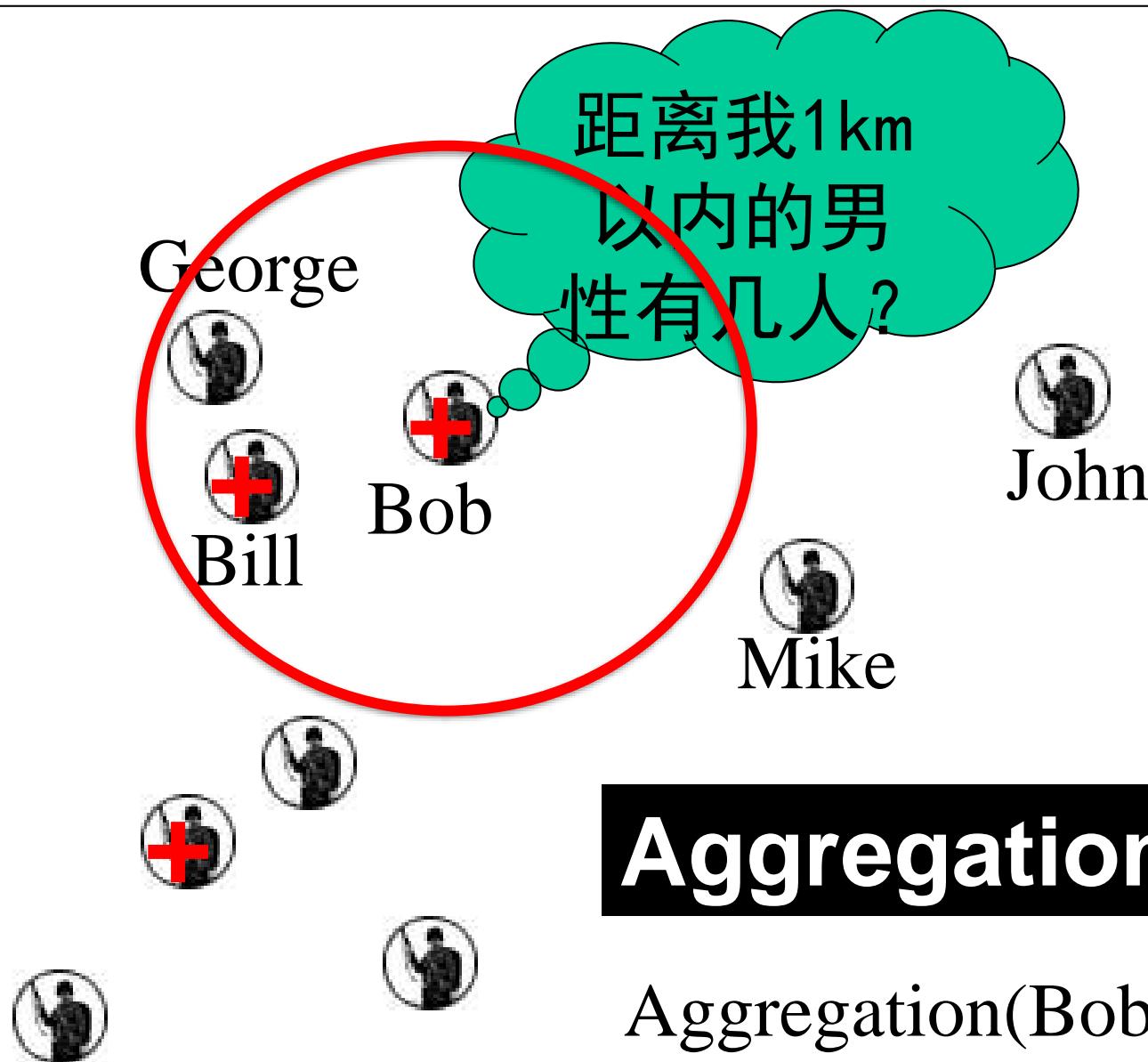
翻转最近邻居查询 (RNN)



范围查询



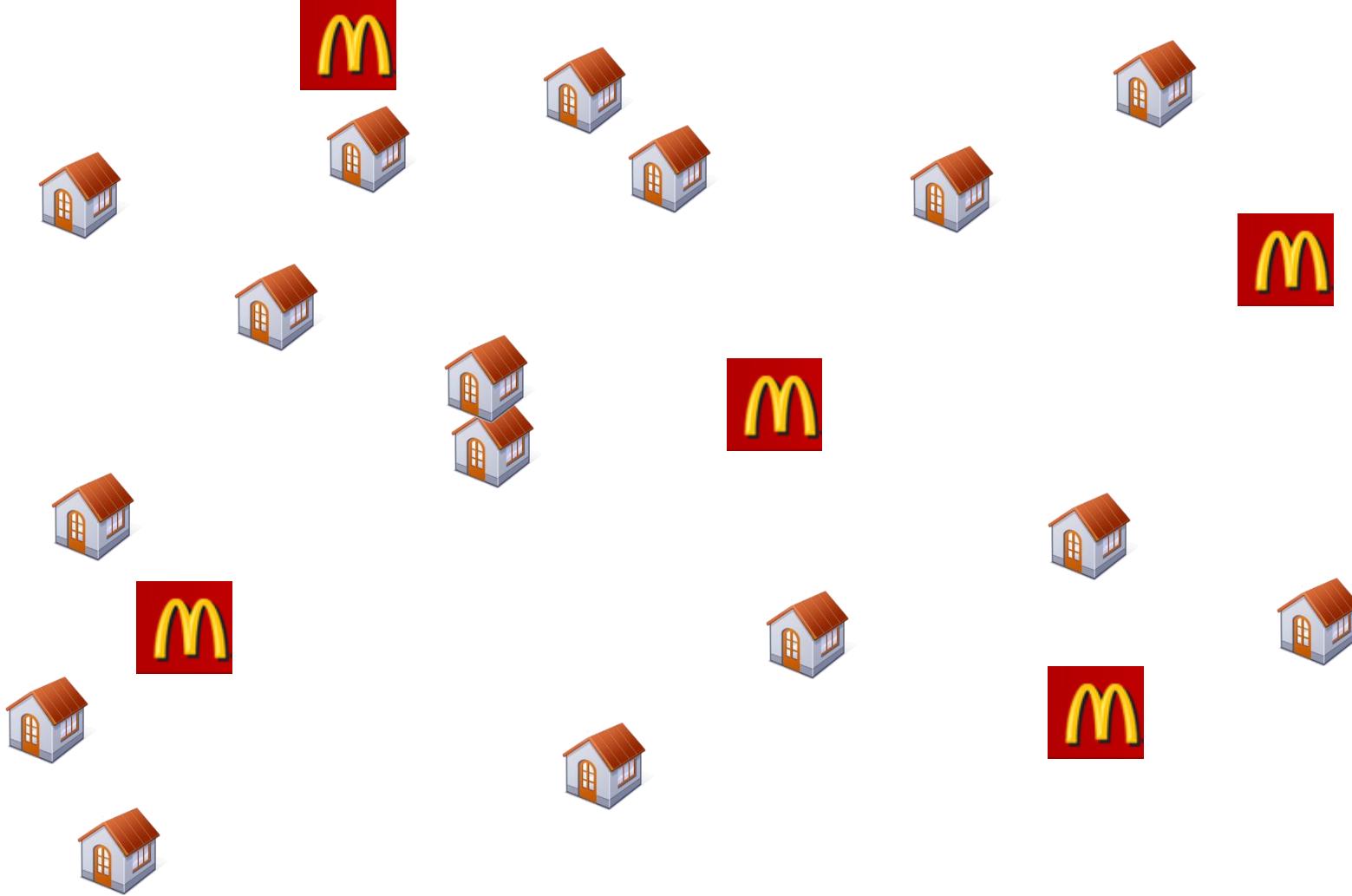
空间集成查询



最短路径查询



最优选址查询

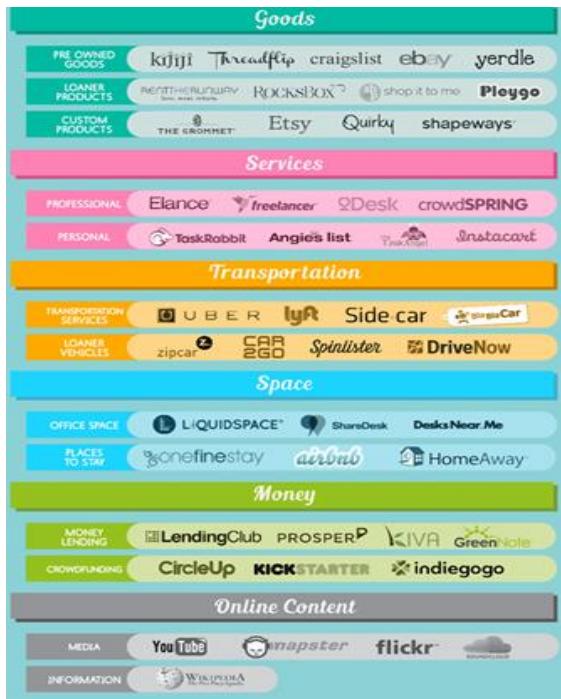


汇报提纲

- 历史回顾
- 基础操作
- 典型应用
 - 实时的专车调度系统
 - 实时的外卖派送系统
 - 时空数据与公共卫生
- 总结展望

典型应用

- 时空大数据服务遍布百姓“衣食住行”的各领域
 - 共享出行：Uber与滴滴出行
 - 共享住宿：Airbnb
 - 共享餐饮：Eatwith与百度外卖等



国外应用



国内应用

汇报提纲

- 历史回顾
- 基础操作
- 典型应用
 - 实时的专车调度系统
 - 实时的外卖派送系统
 - 时空数据与公共卫生
- 总结展望

实时专车调度系统

- 实时专车系统(滴滴专车或神州专车)通常采用“最近邻查询”这一贪心策略进行动态任务匹配
- 一个叫车请求一旦出现，它应该立即被分配处理。



实时专车调度系统

- 实时专车系统(滴滴专车或神州专车)通常采用“最近邻居”这一贪心策略进行动态任务匹配
 - 一个叫车请求一旦出现，它应该立即被分配处理。
 - 根据最差情况分析贪心策略的总距离将比已知全部信息的离线最优解大指数倍！在现实应用中所广泛采用的贪心策略真的错了吗？



实时专车调度系统: 竞争比分析

- 竞争比 (Competitive Ratio: CR)

- $CR = \frac{\text{在线算法成本}}{\text{离线算法成本}}$

- 输入模型

- 对手模型(最差情况分析)

- $CR_A = \max_{\forall G(T,W,U) \text{ and } \forall v \in V} \frac{\text{MinSum}(M)}{\text{MinSum}(OPT)}$

最差的空间位置结构

最差的出现顺序

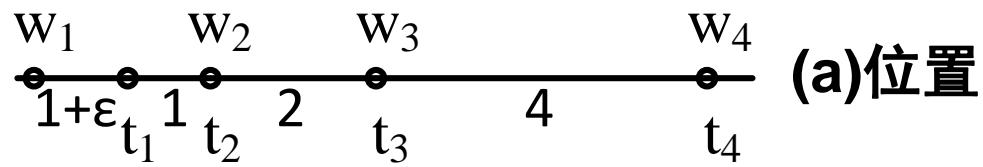
- 随机顺序模型(平均情况分析)

- $CR_{RO} = \max_{\forall G(T,W,U)} \frac{\mathbb{E}[\text{MinSum}(M)]}{\text{MinSum}(OPT)}$

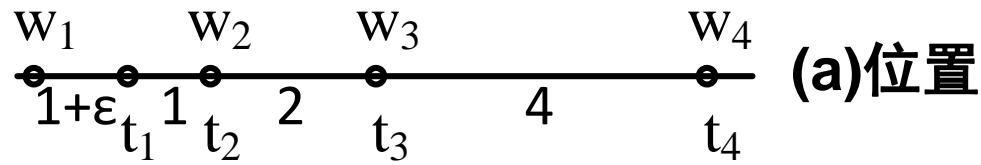
最差的空间位置结构

全部可能出现顺序
的数学期望

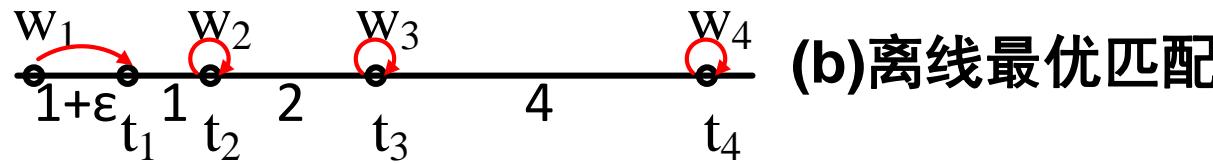
实时专车调度系统: 贪心策略分析



实时专车调度系统: 贪心策略分析

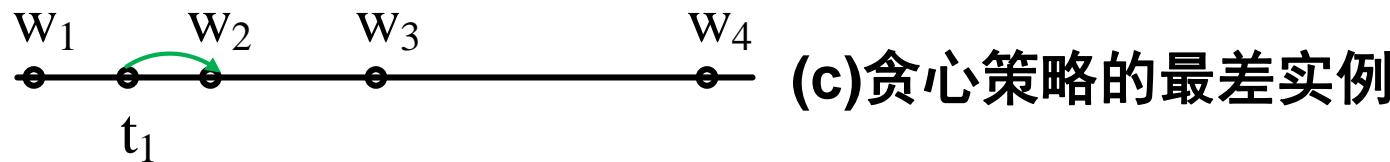
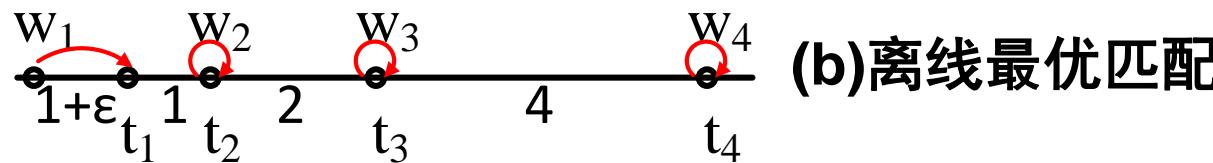
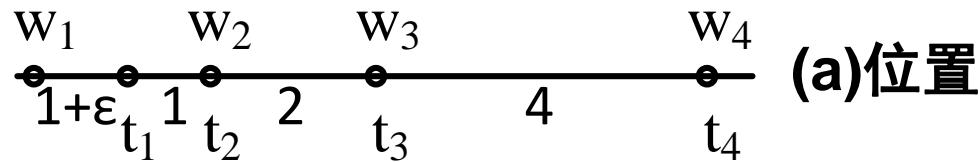


(a)位置



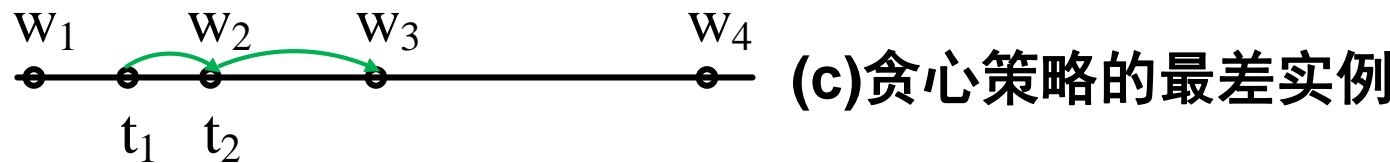
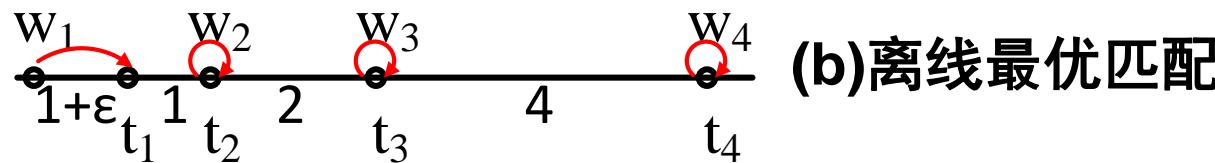
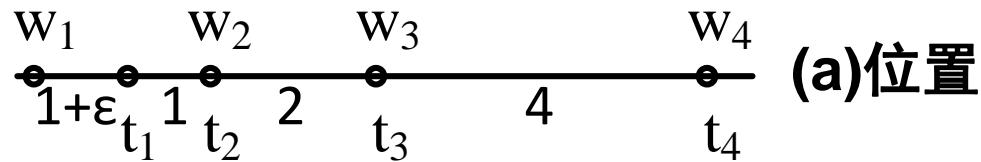
(b)离线最优匹配

实时专车调度系统: 贪心策略分析



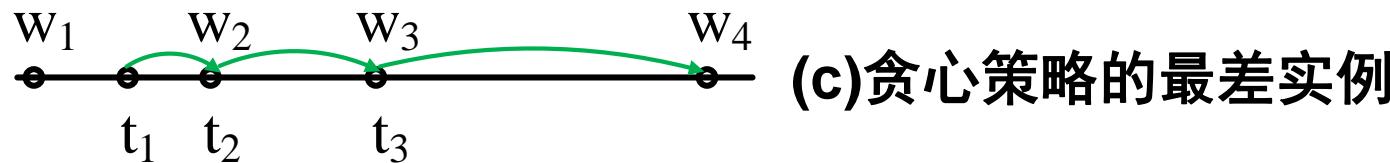
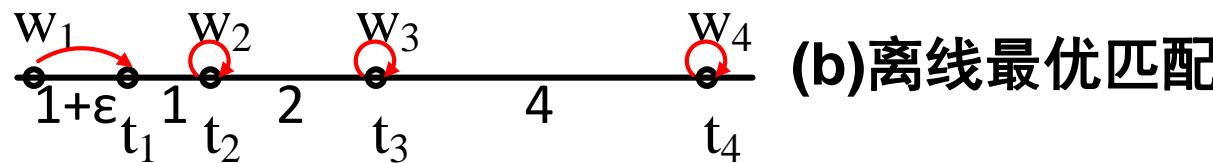
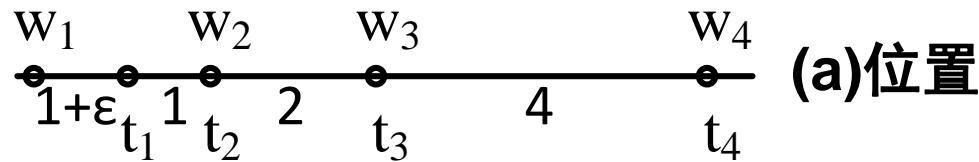
当 t_1 出现时，贪心策略会将 w_2 分配给 t_1 。

实时专车调度系统: 贪心策略分析



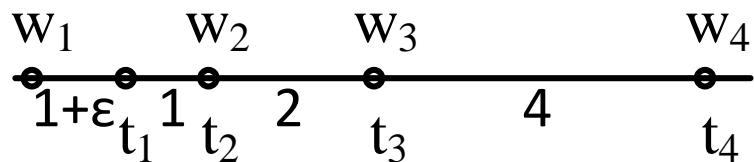
当 t_2 出现时，贪心策略会将 w_3 分配给 t_2 。

实时专车调度系统: 贪心策略分析



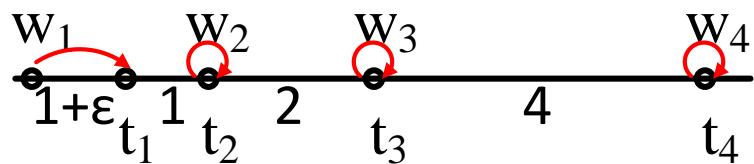
当 t_3 出现时，贪心策略会将 w_4 分配给 t_3 。

实时专车调度系统: 贪心策略分析

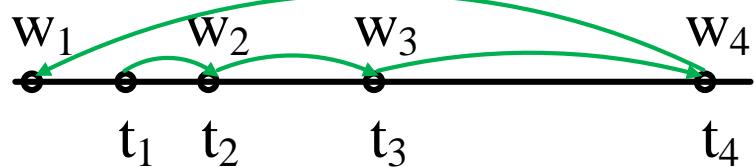


平均情况竞争比分析: $CR_{R0}=3.195$

(a)位置



(b)离线最优匹配



(c)贪心策略的最差实例

当 t_4 出现时，贪心策略只能会将 w_1 分配给 t_4 。

因此，在平均分析的情况下，贪心策略可以获得一个常数阶的竞争比！

实时专车调度系统

- 揭示了“空间数据在线最小匹配”这一大数据与算法领域的经典问题在过去25年研究存在重大误解！
- 在“空间数据在线最小匹配”问题中，贪心算法在平均情况分析中具有常数级近似比而不是最差分析中的指数级近似比，而且也证明了具有指数级近似比的最差情况发生情况为 $1/n!$ (n 为顶点数)。

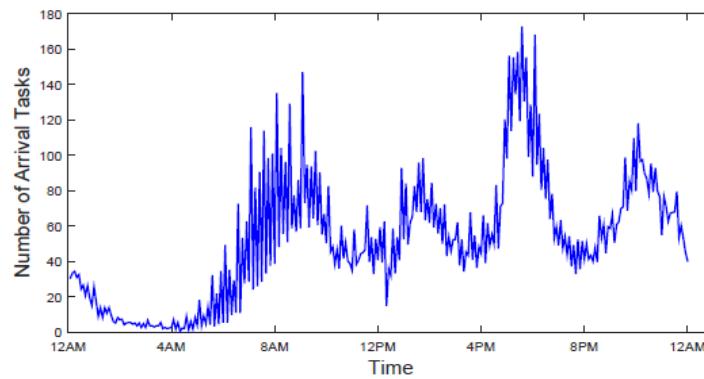


Figure 5: Average number of tasks of taxi-calling per day

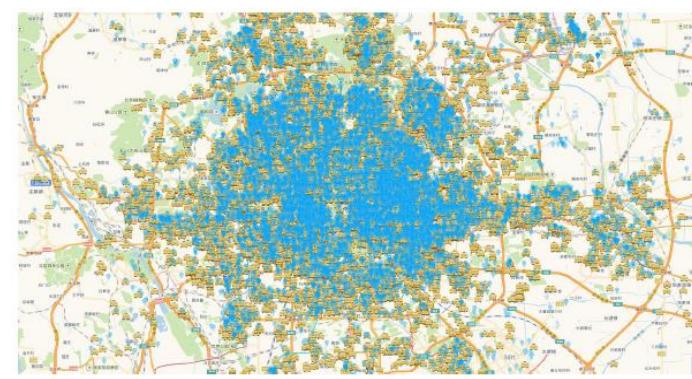


Figure 6: Location distribution of the users and service providers at the ShenZhou taxi-calling platform on one day in Beijing

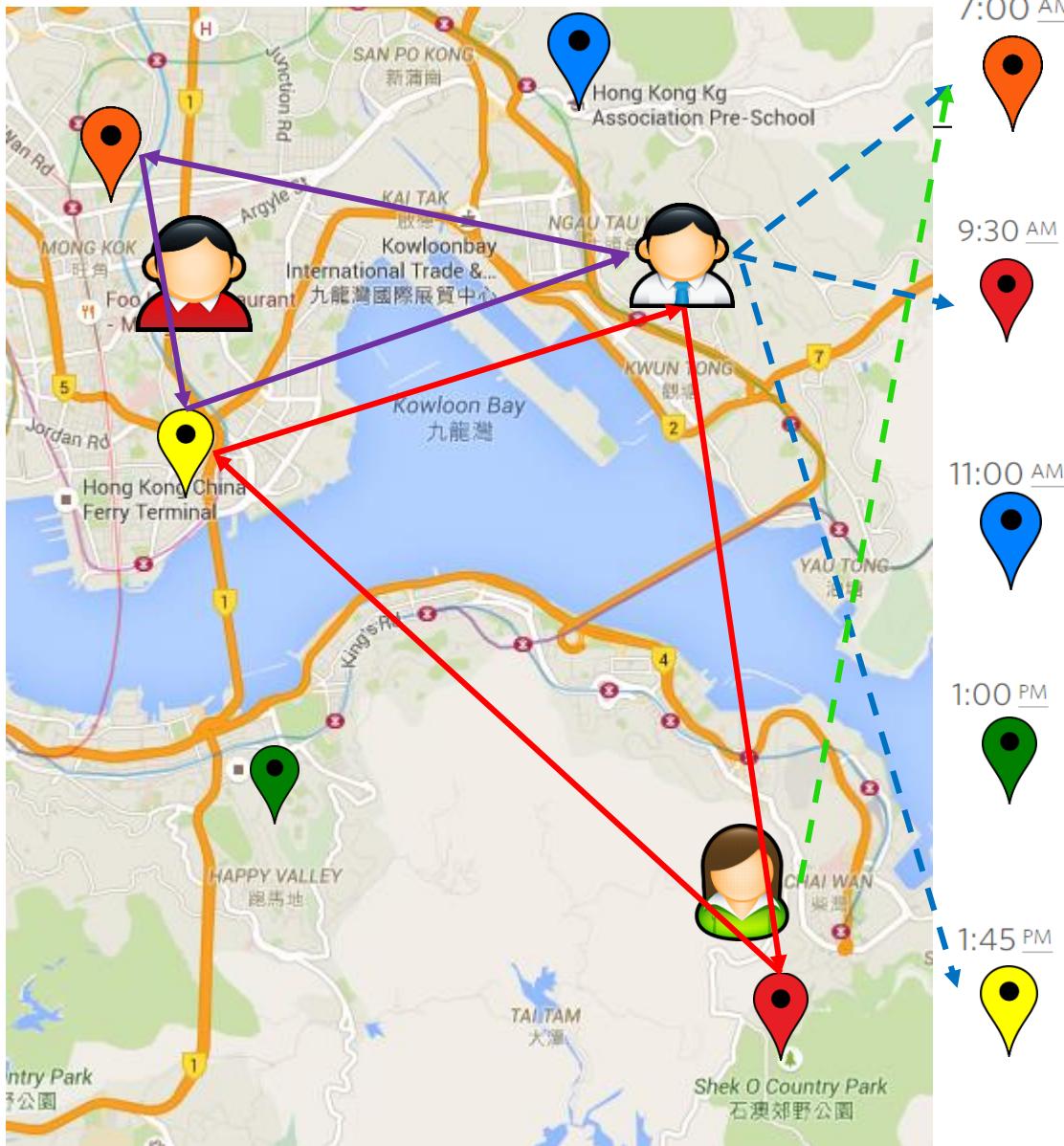
**Yongxin Tong, Jieying She, Bolin Ding, Lei Chen, Tianyu Wo, Ke Xu:
Online Minimum Matching in Real-Time Spatial Data: Experiments and
Analysis. In VLDB, Pages 1053-1064, 2016. (CCF A类会议)**

汇报提纲

- 历史回顾
- 基础操作
- 典型应用
 - 实时的专车调度系统
 - 实时的外卖派送系统
 - 时空数据与公共卫生
- 总结展望

实时的外卖配送系统





7:00 AM

Hong Kong Squash Lovers Club

订单

火腿三明治+早啊100%果蔬汁+软式小面包

9:30 AM

Hong Kong Hiking Meetup 香港远足聚会

订单

福满业香玛奇朵+巧克力星冰乐

11:00 AM

Hong Kong Volunteers 香港义工团

订单

双拼饭(叉烧加烧肉)

1:00 PM

Weekends Badminton, Ballroom, Latin Dance & Ad-hoc Dining

订单

茄牛公仔面+蕃茄蛋通粉

1:45 PM

Hong Kong Dolls, Figures and Toys Collectors Meetup

订单

浓情巧克力蛋糕+法式鸡肉三明治

Jieying She, Yongxin Tong, Lei Chen. Utility-Aware Event-Participant Planning. In SIGMOD, Pages 1629-1643, 2015. (CCF A类会议)

汇报提纲

- 历史回顾
- 基础操作
- 典型应用
 - 实时的专车调度系统
 - 实时的外卖派送系统
 - 时空数据与公共卫生
- 总结展望

时空数据与公共卫生

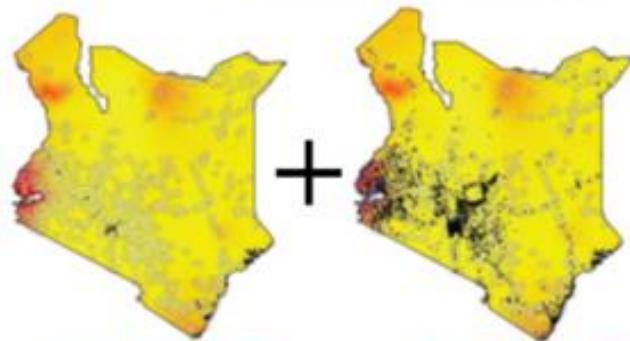
● 城市疾病的流行溯源追踪



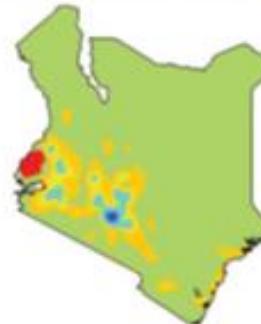
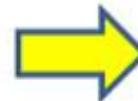
Quantifying the Impact of Human Mobility on Malaria,
《科学》杂志, 2012年10月12日

调整城市功能, 改变人口流动, 控制疾病传播

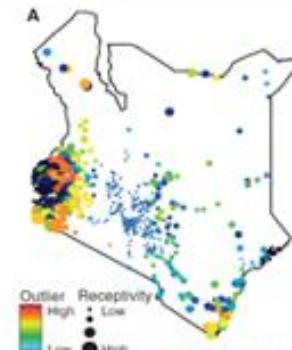
肯尼亚的
疟疾流行
与城市功
能分布相
关!



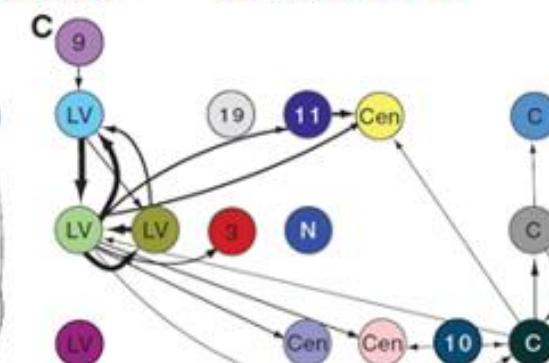
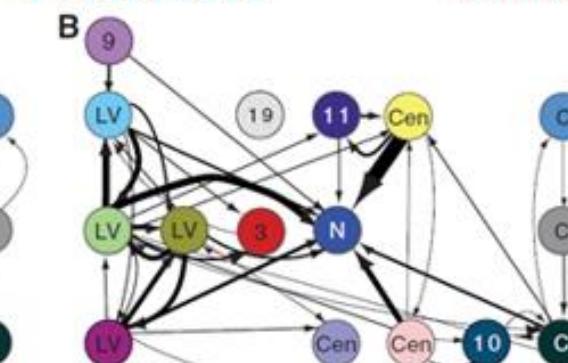
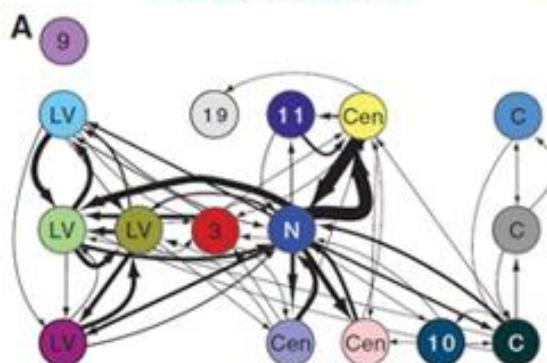
疟疾发病数据



疟疾传播溯源



区域风险评估



疟疾在不同城市间传播的转移图

汇报提纲

- 历史回顾
- 基础操作
- 典型应用
- 总结展望

总结与展望

- 空间大数据在日常生活中如影随形
 - 其遍及大家日常生活中的“衣食住行”各个方面



衣

食

住

行

时空大数据的应用随着智能手机与移动互联网的普及而越发兴盛，与此同时，一种新型的经济形态（共享经济）也被这新一轮技术革命所催生。共享经济又对计算有什么样的影响呢？请听下回：共享经济时代的众包计算

謝謝