



富人和穷人的故事

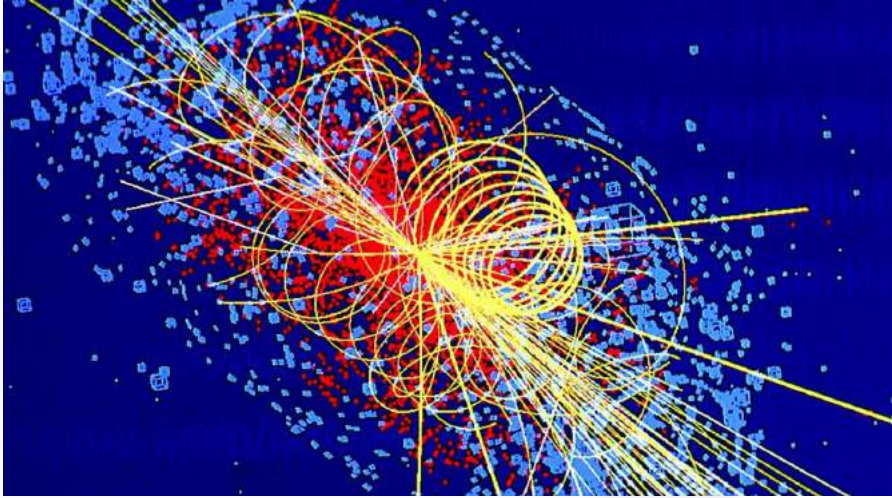
从手机数据看社交网络

主讲人：潘巍



社会科学 The Science of Society

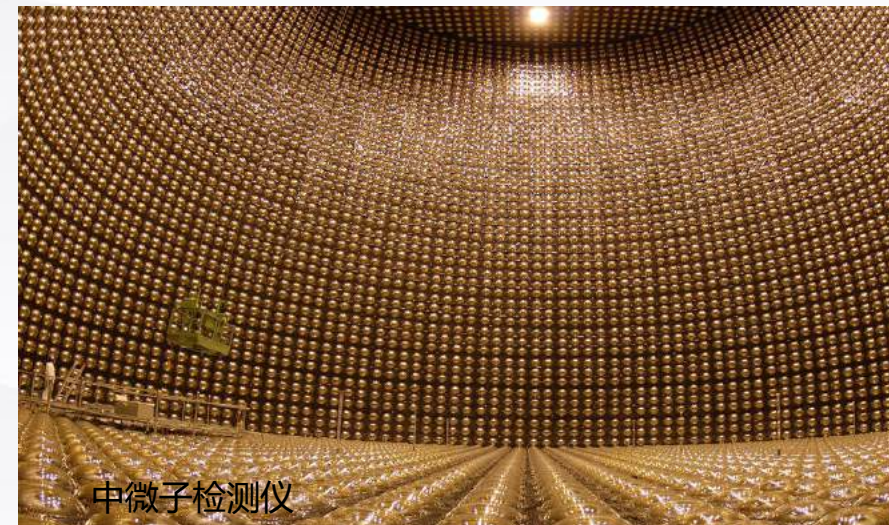
物理 – 理论由大数据采集开始



大型强子加速器



重力波检测仪器



中微子检测仪

“大规模”的手机数据采集 2010-2011

- 我们将安装数据收集软件的智能安卓手机发给了大约120名Westgate的成员。
- Westgate是MIT博士生和他们家庭成员的宿舍楼。
- 许多不同的人(种族, 信仰, 工作, 语言等)。



Android手机传感采集系统

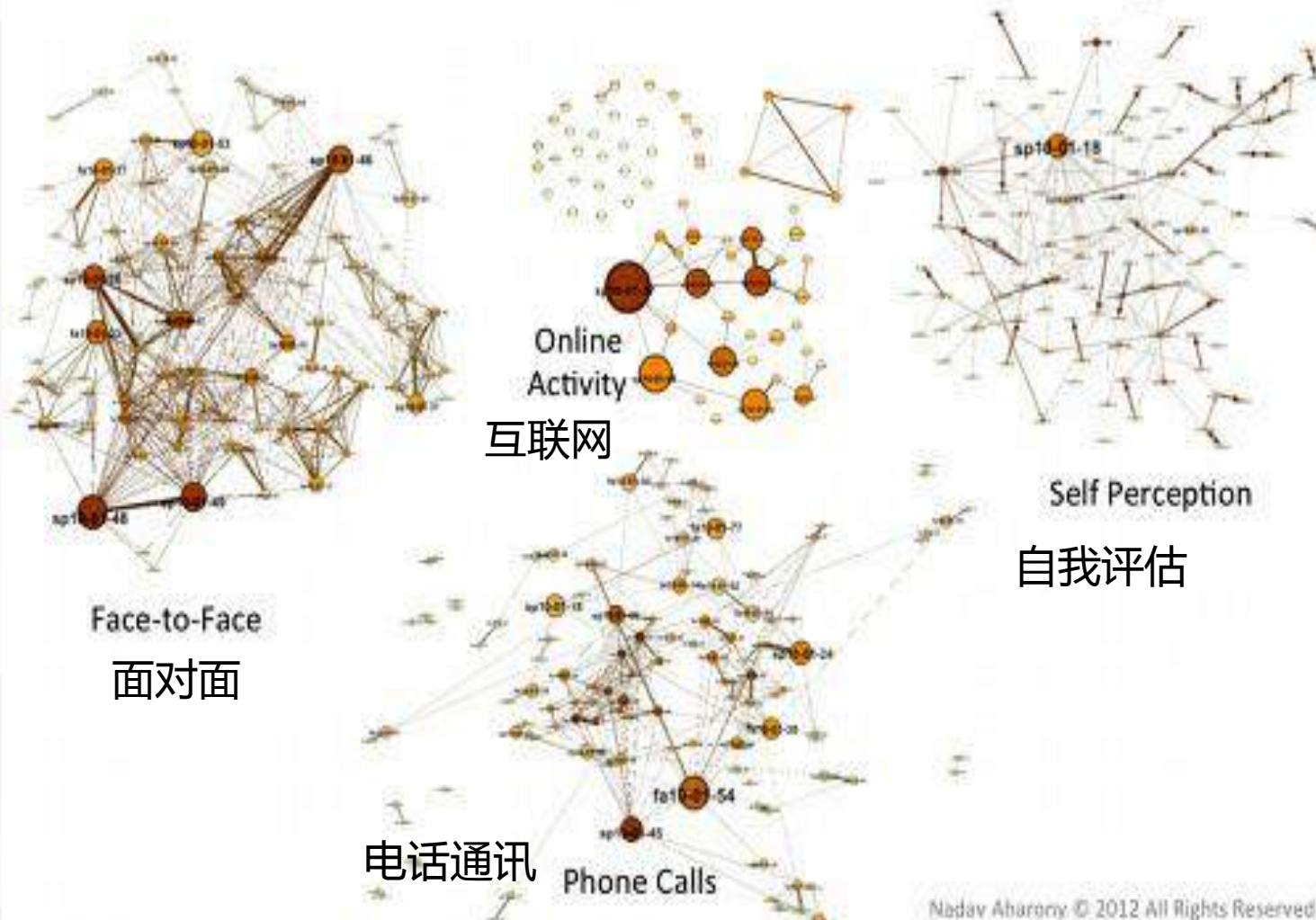
- 开源：funf.media.mit.edu





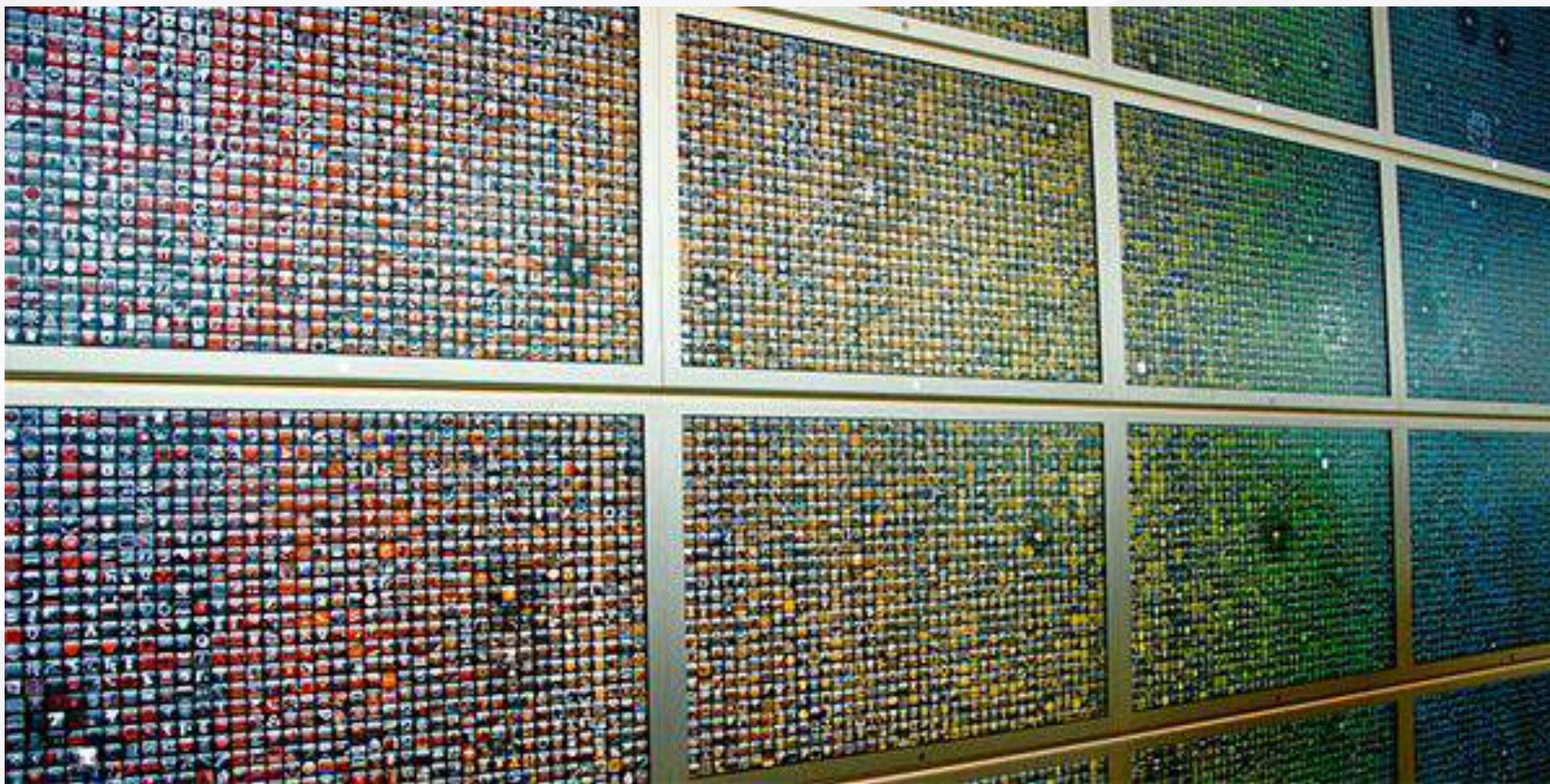
来自手机的社交网络更加丰富

One Community, Many Networks

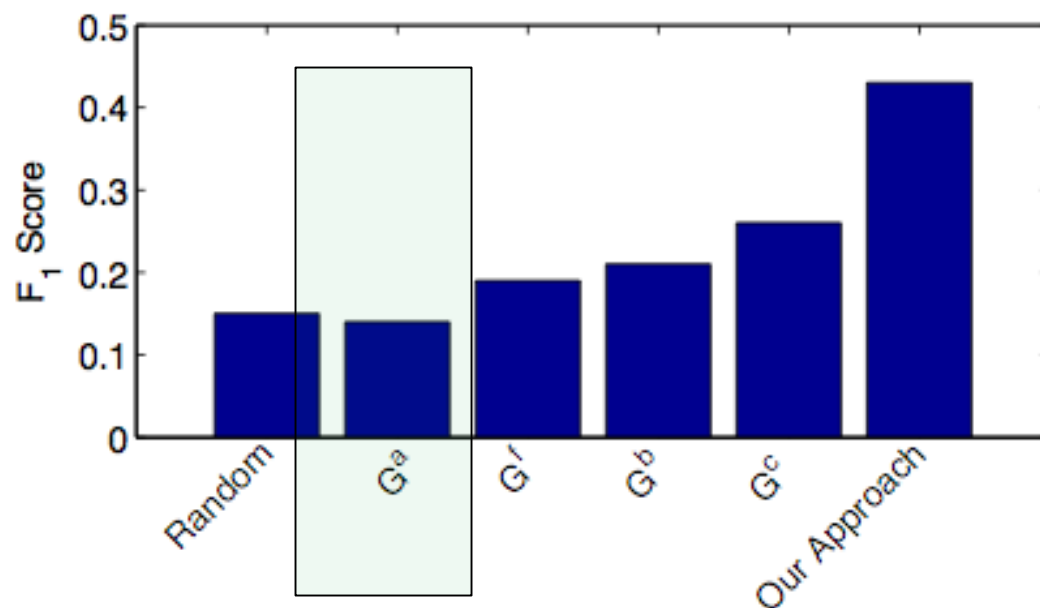


手机会给我们对于社交网络有全新的认识。

影响：手机用户安装应用的预测

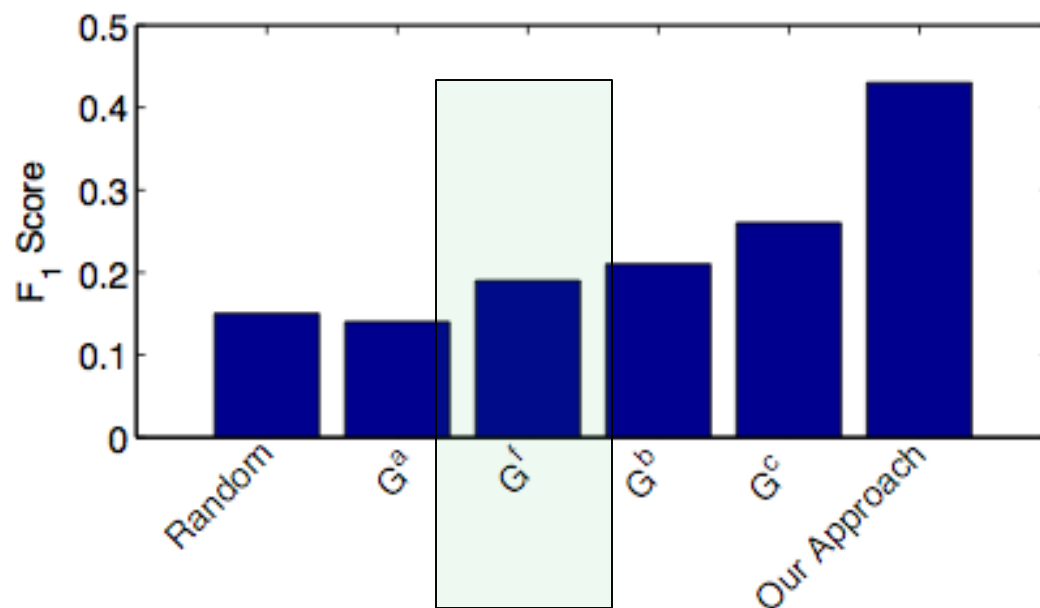


预测用户会安装的运用



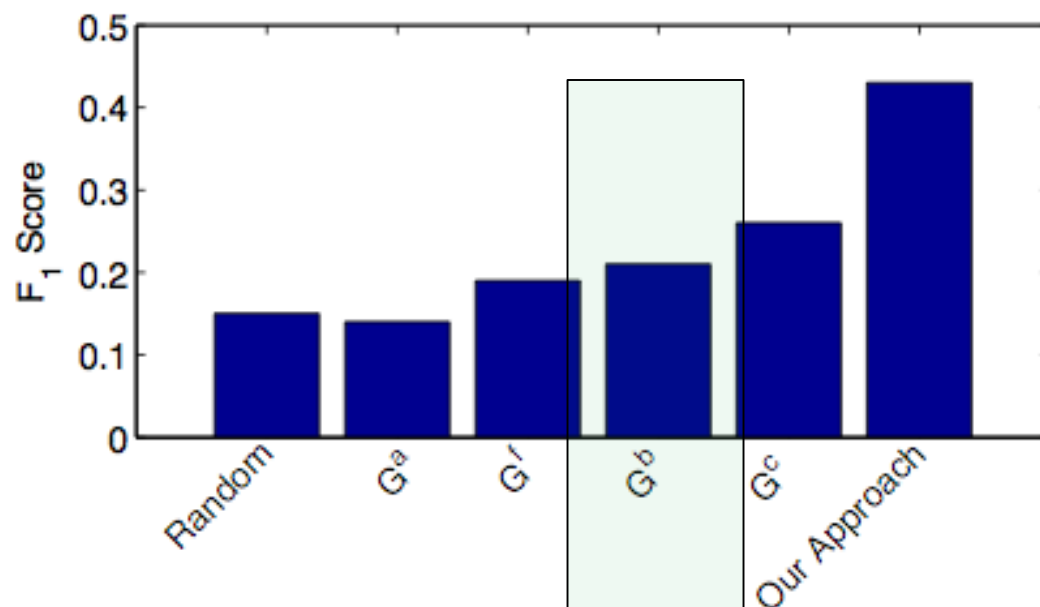
用手机位置计算的网路预测

预测用户会安装的运用



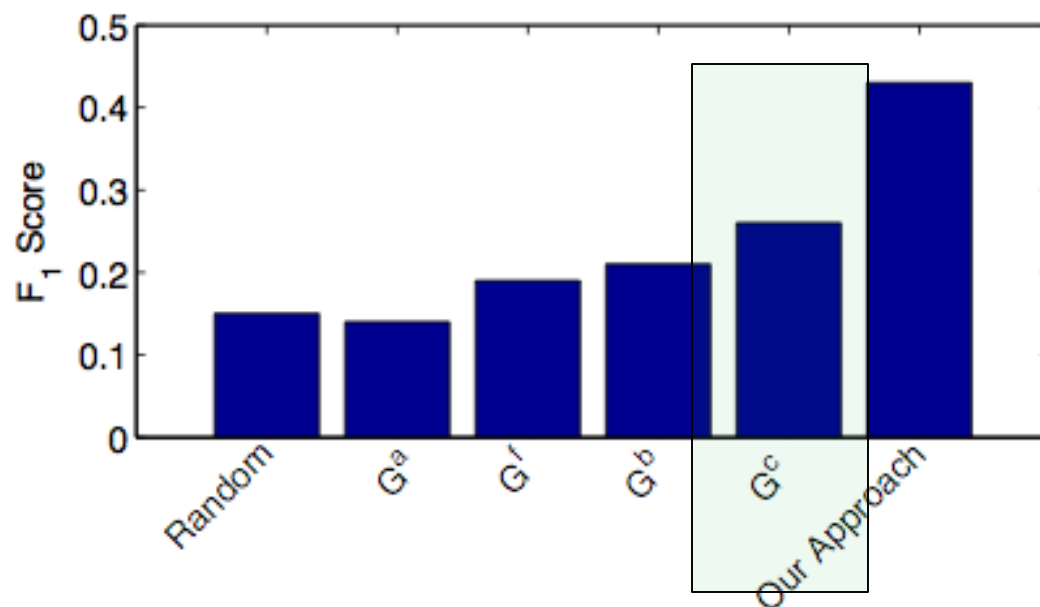
用Facebook/人人网的网络预测

预测用户会安装的运用



用物理见面的网络预测

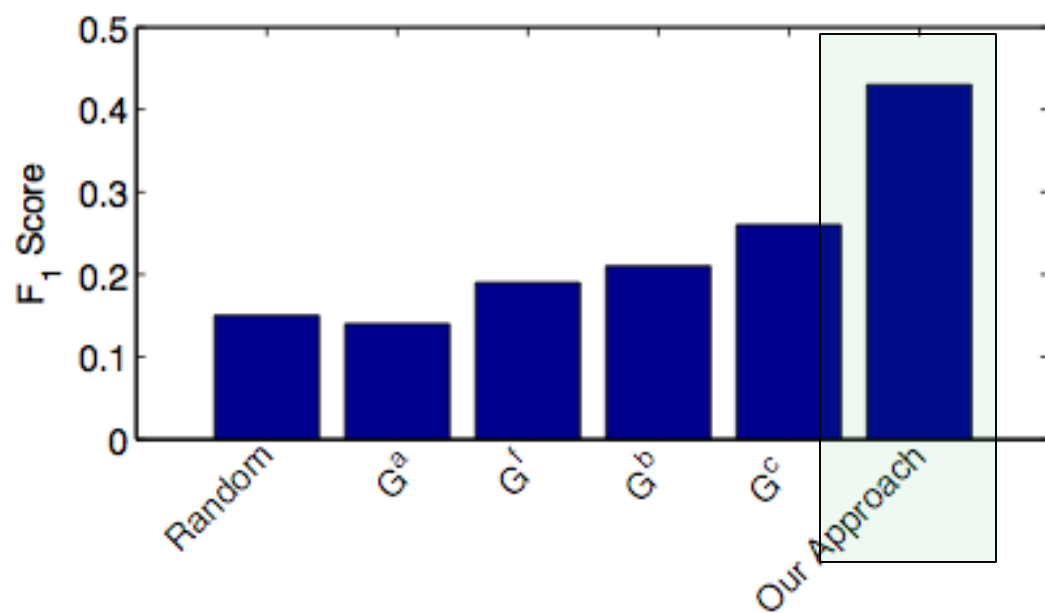
预测用户会安装的运用



用手机相互打电话的网络预测

预测用户会安装的运用

- 50%预测准确率



我开发的算法(使用所有的网络)

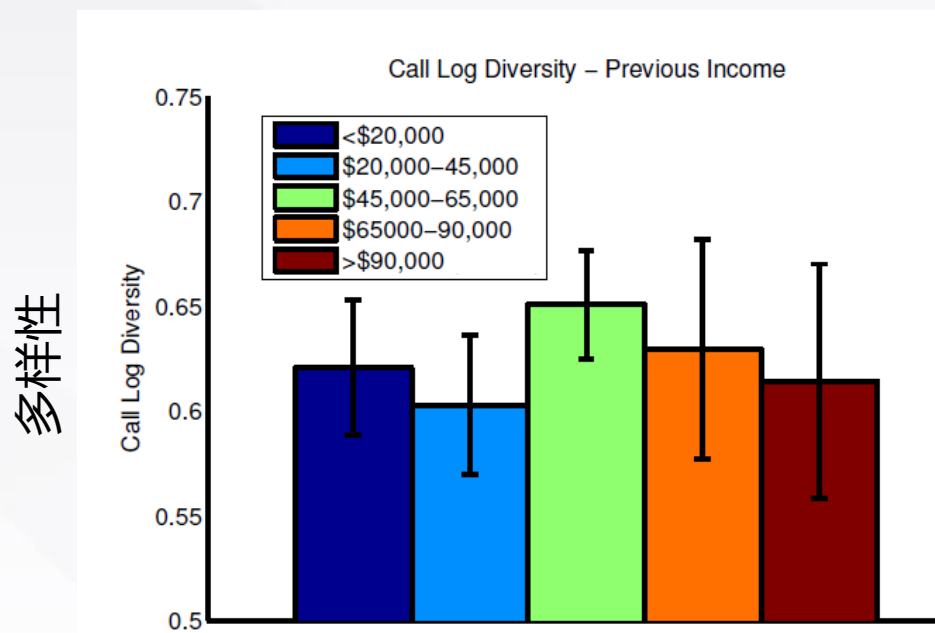
穷人和富人的故事：



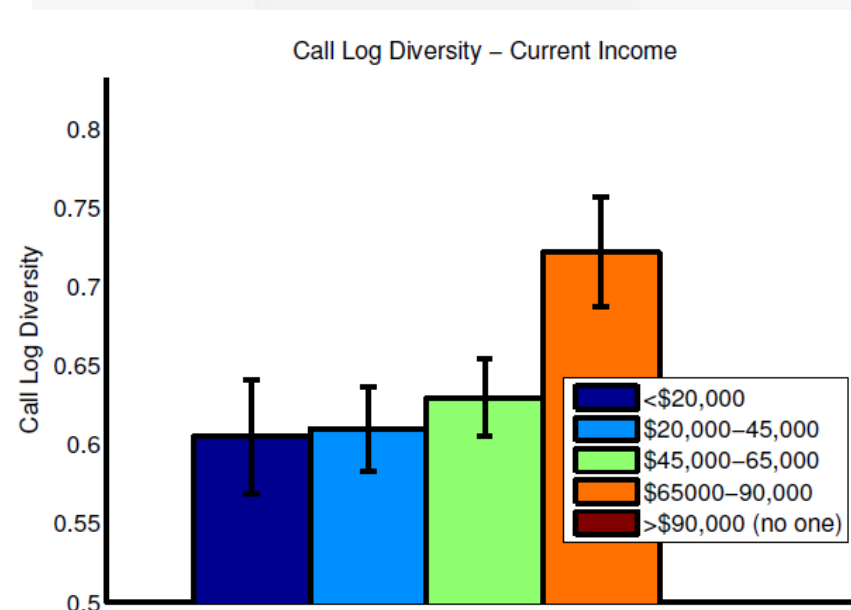
我的发现：

$$D(i) = \frac{-\sum_{j=1}^k p_{ij} \log p_{ij}}{\log k}$$

- 手机通话记录的**多样性**不一样：



过去的收入
($r=0.003$, $p>0.80$)

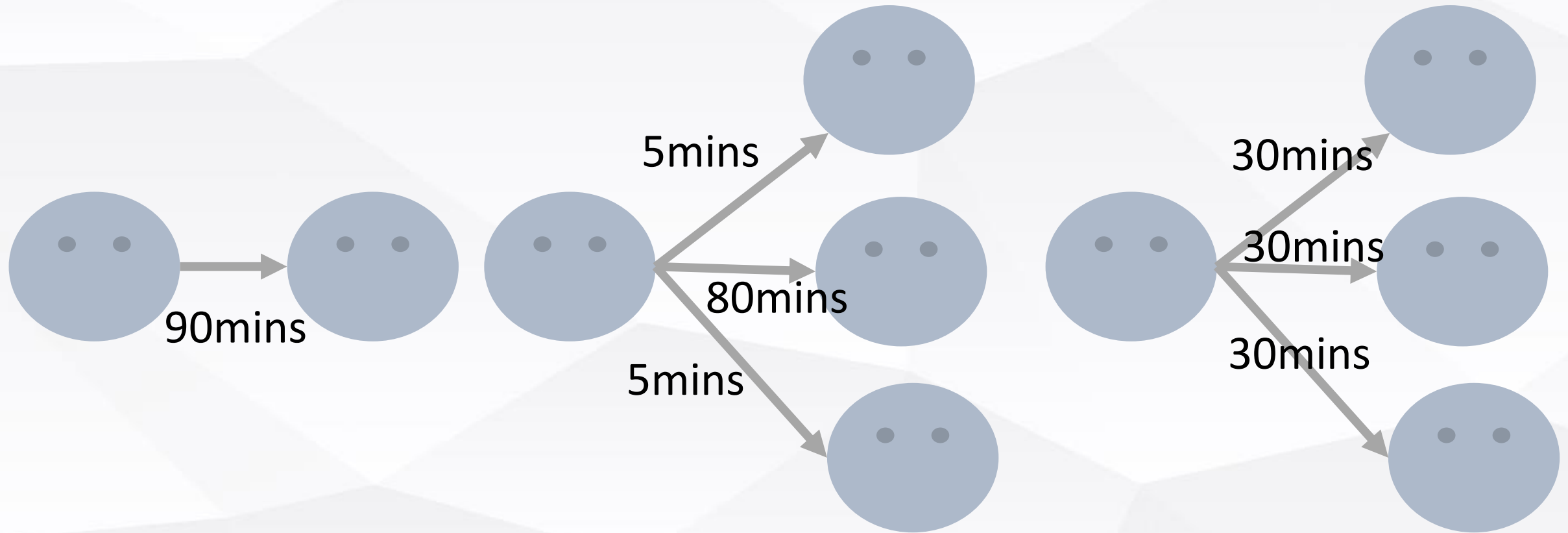


现在的收入
($r=0.28$, $p=0.08$)

均化多样性 (Normalized Diversity)

$$D(i) = \frac{-\sum_{j=1}^k p_{ij} \log p_{ij}}{\log k}$$

均化多样性 (Normalized Diversity)

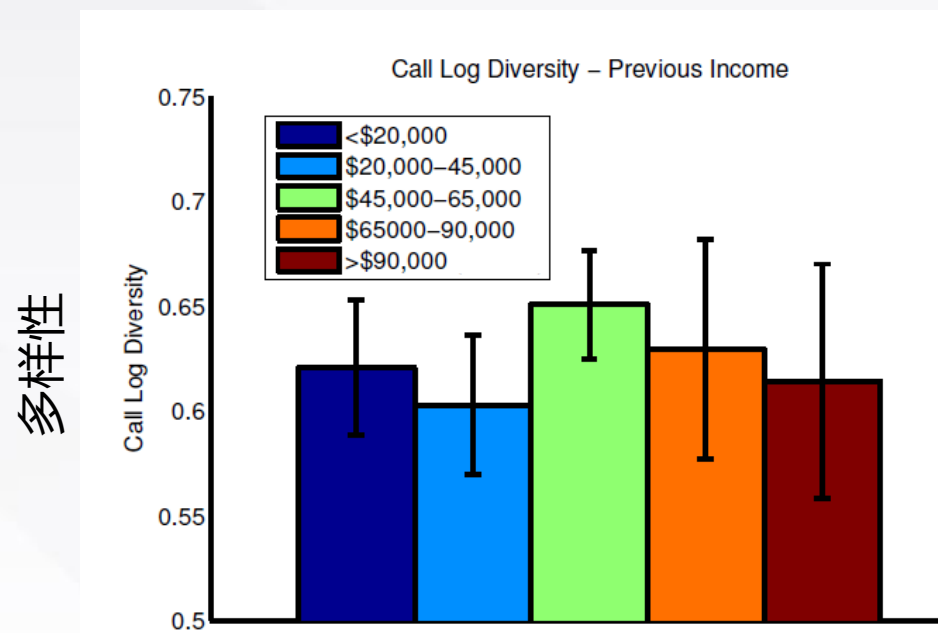


均化多样性

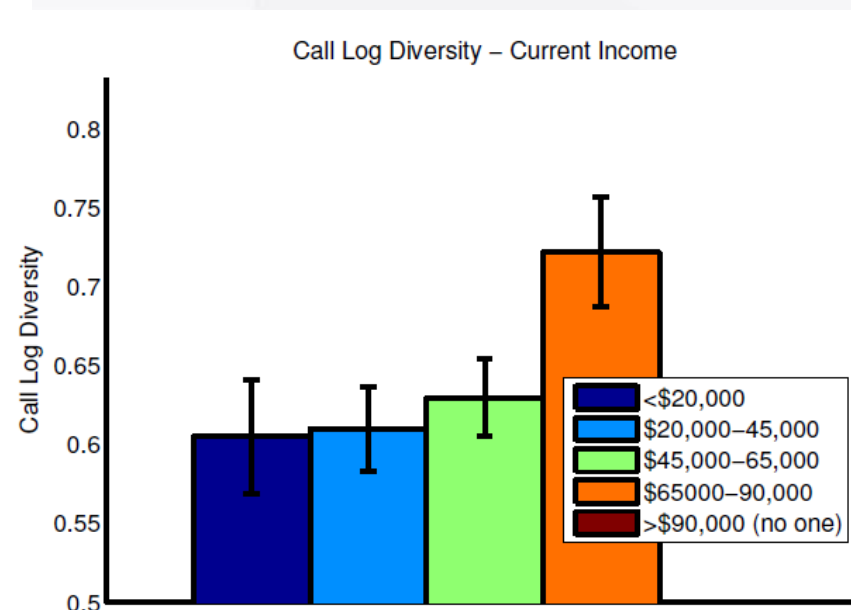
我的发现：

$$D(i) = \frac{-\sum_{j=1}^k p_{ij} \log p_{ij}}{\log k}$$

- 手机通话记录的**多样性**不一样：

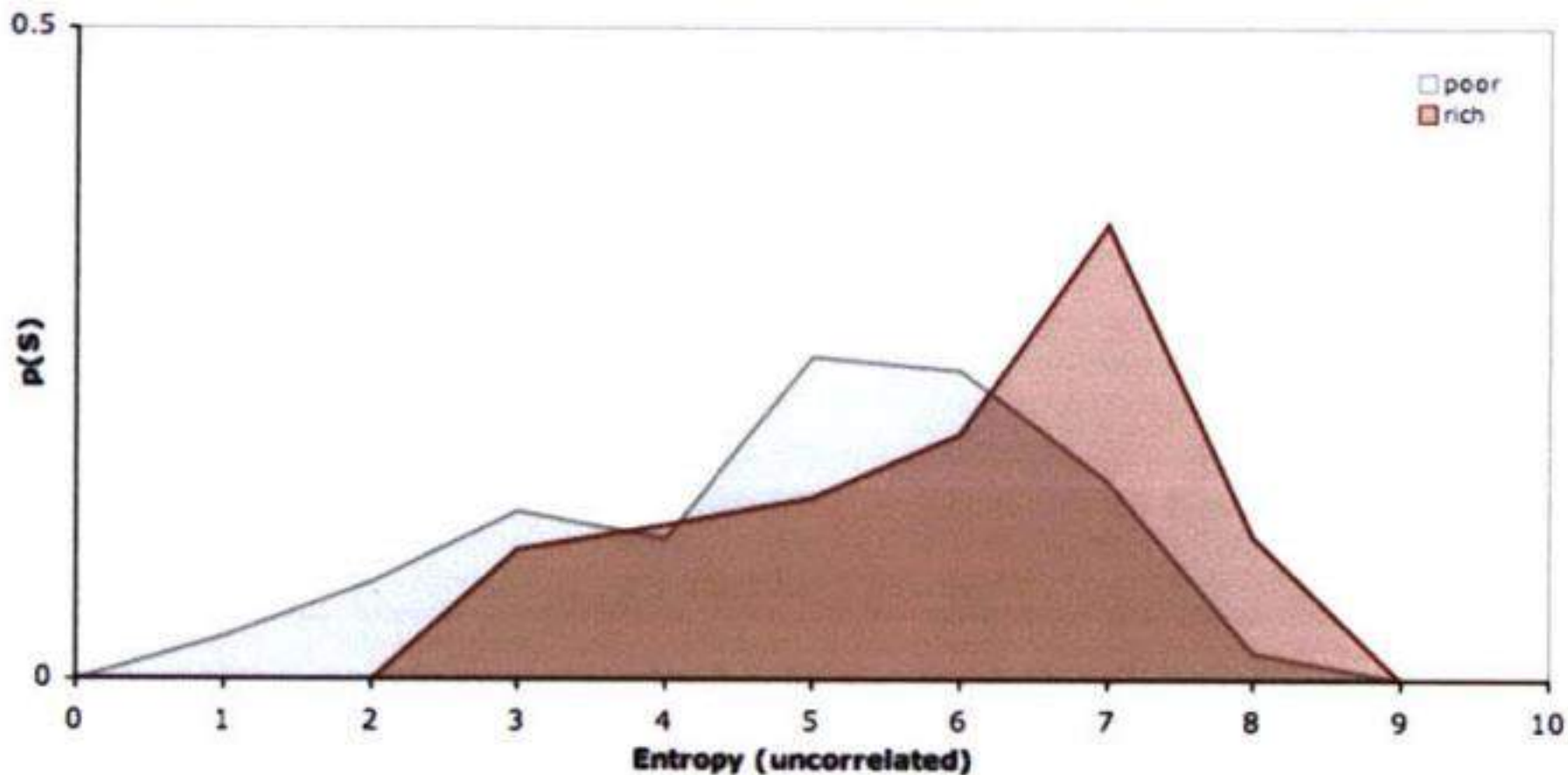


过去的收入
($r=0.003$, $p>0.80$)



现在的收入
($r=0.28$, $p=0.08$)

利用信用卡数据分析消费行为后 得到同样的结果

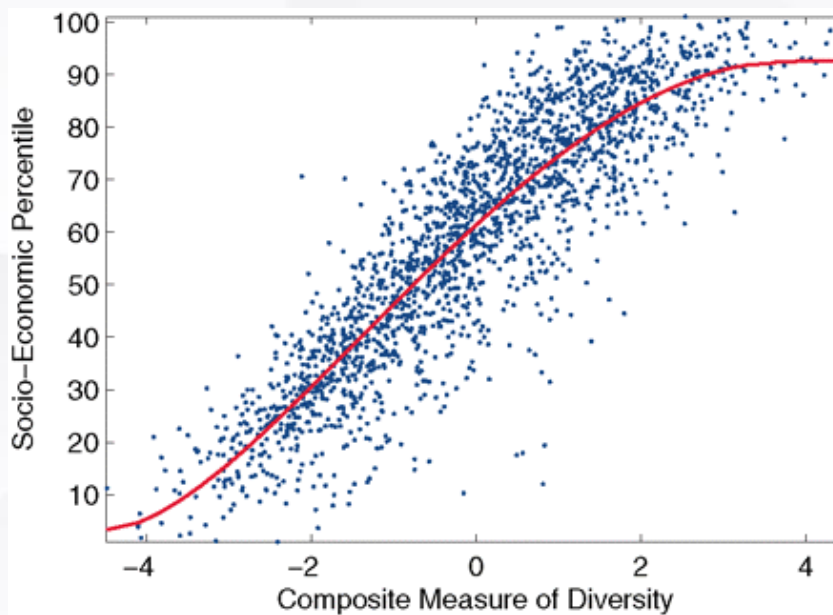
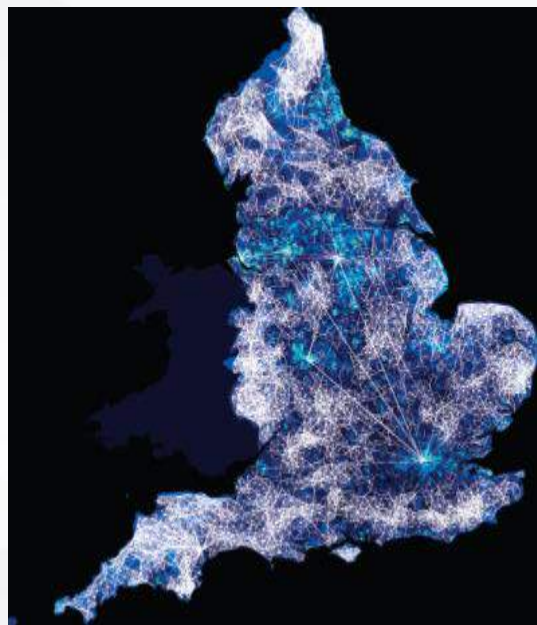


多样性分布

[Krumme, MIT, 201P]

经济与用户行为的关系

- 从电话公司的通讯记录中的神奇发现[Eagle et al. 2010 Science]

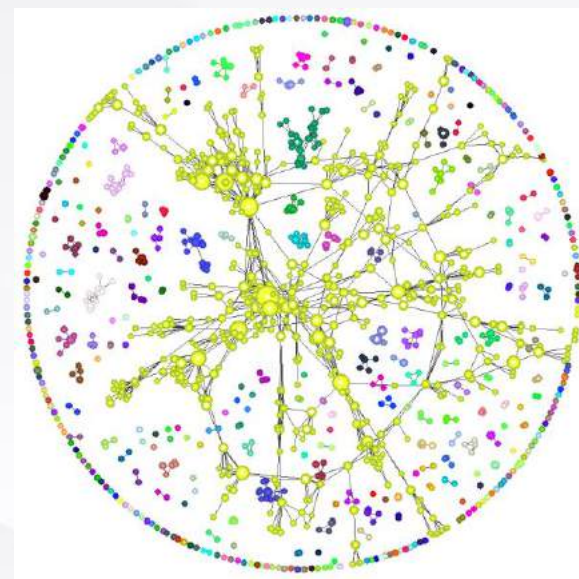
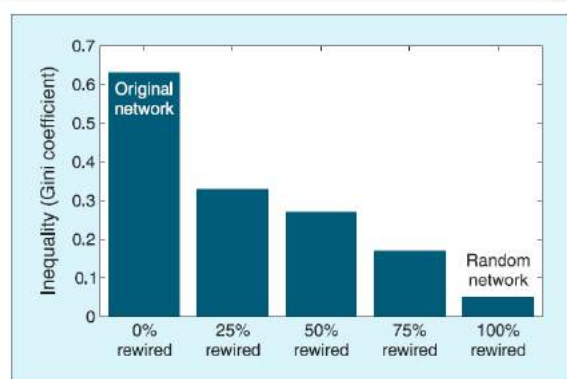


联系人的**多样性**和这个城市的经济指标具有很强的相关性 ($r=0.73$).

仅仅通过测量用户通话记录

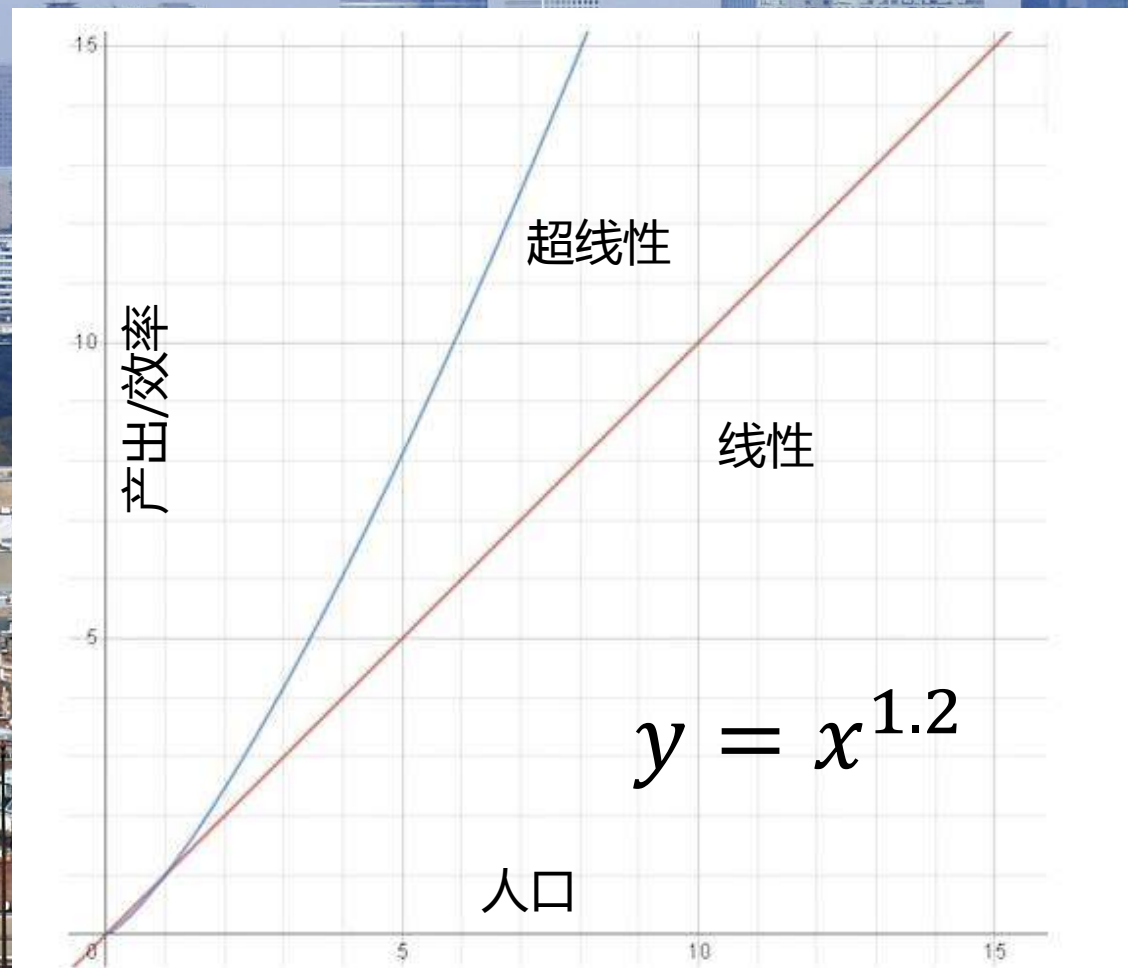
- 我们可以预测用户的：
 - 债务偿还能力 [Björkegren 2015]
 - 信用指数 [Massuia 2015]
 - 找到工作的可能性 [Reis 2015]

多样性会增加知识流(Idea Flow)



很多关于企业管理的研究证明团队之间的关系，知识流通对于团队的产出有很强的关联性。

城市的超线性之谜



[Betterncourt 2009]

城市的超线性之谜

产出：

1. 专利数量
2. 科研经费投入
3. GDP增长
4. 工资

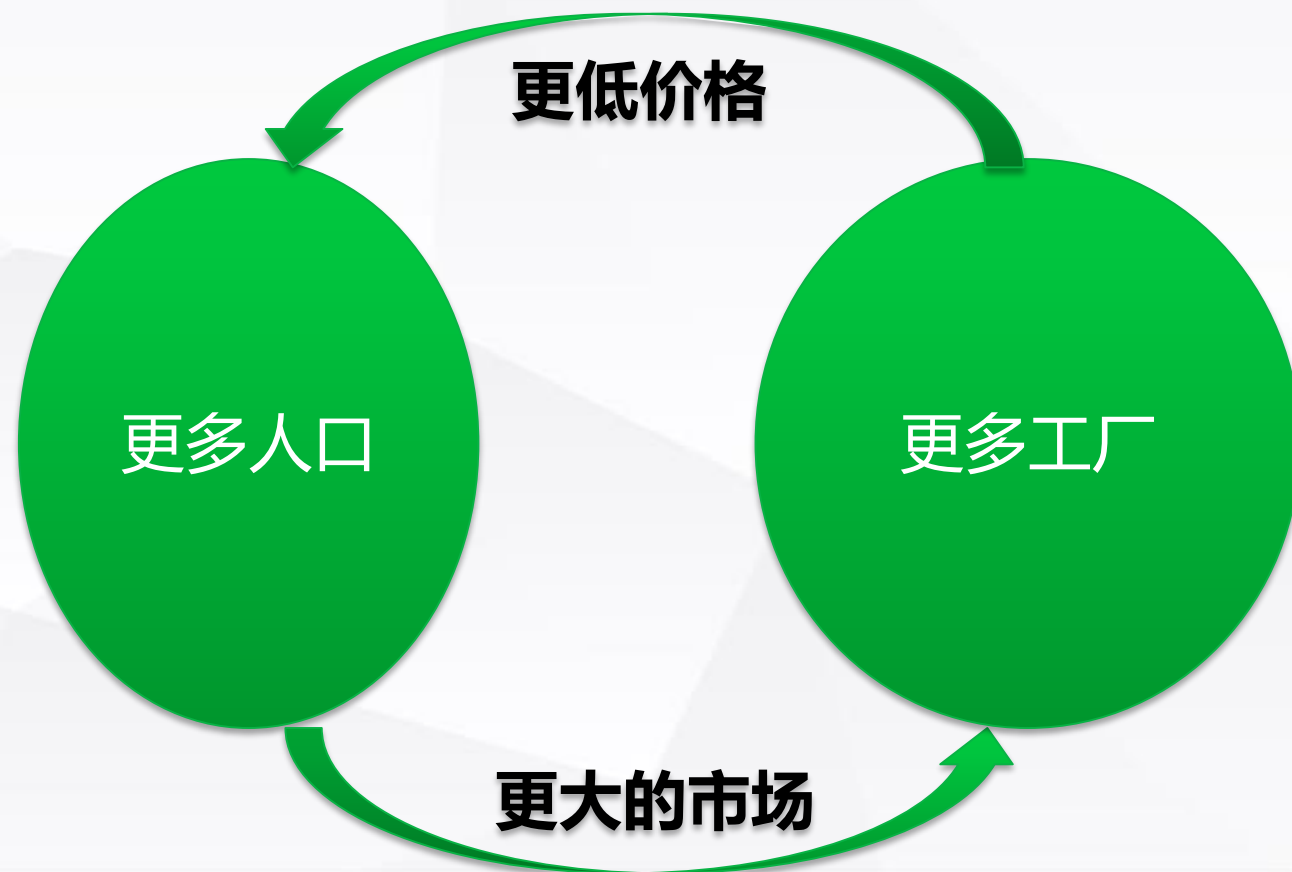
.....

人口



传统理论：地理经济学

- 诺贝尔经济学奖Krugman et al: 地理位置的接近带来了生产工厂和消费者之间的更强互动，使得城市更有效率。



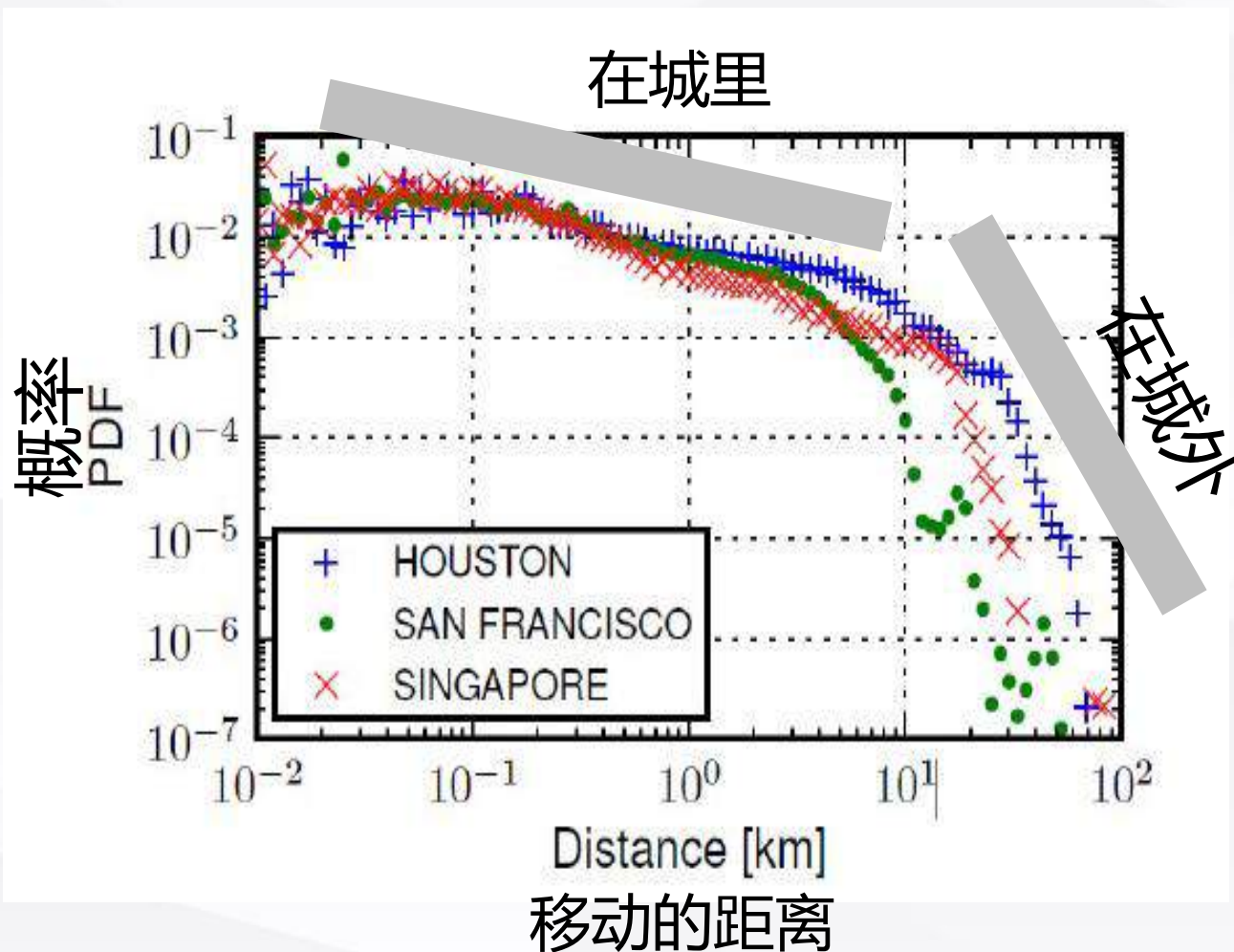
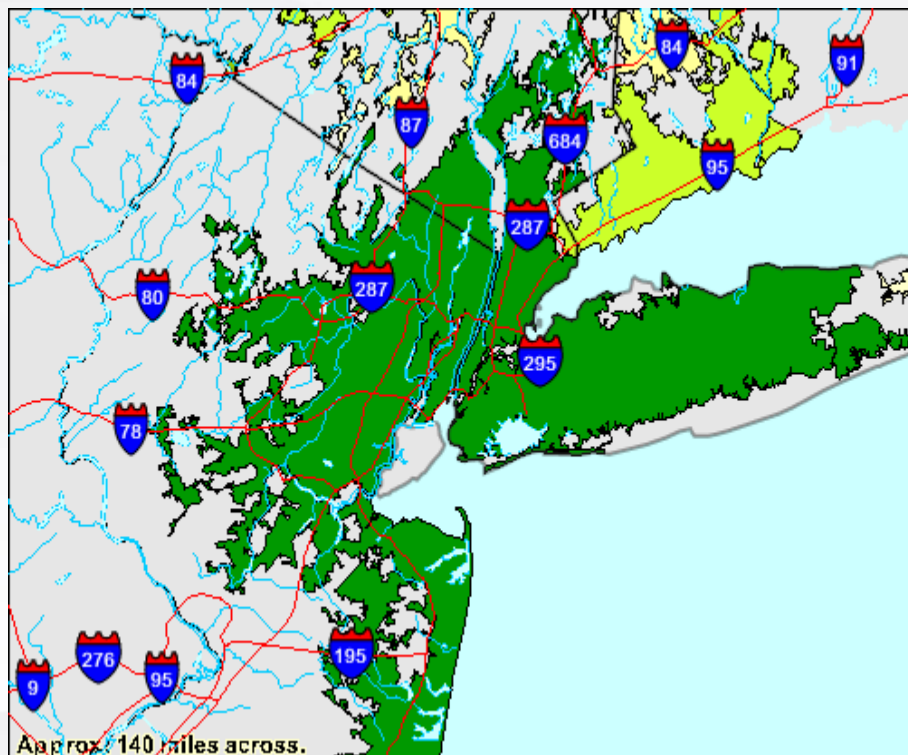
1900年的城市



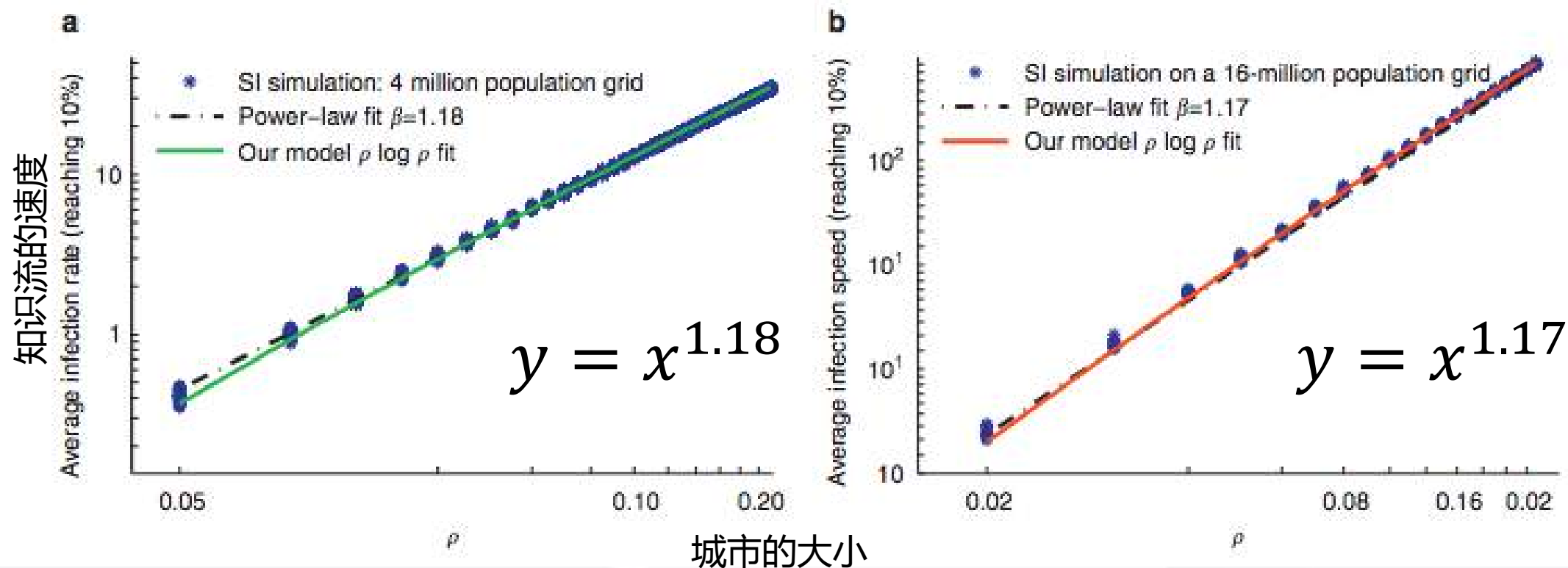
2000年的城市



城市：帮助你超线性增加社交多样性



我们的理论：城市帮助提升知识流

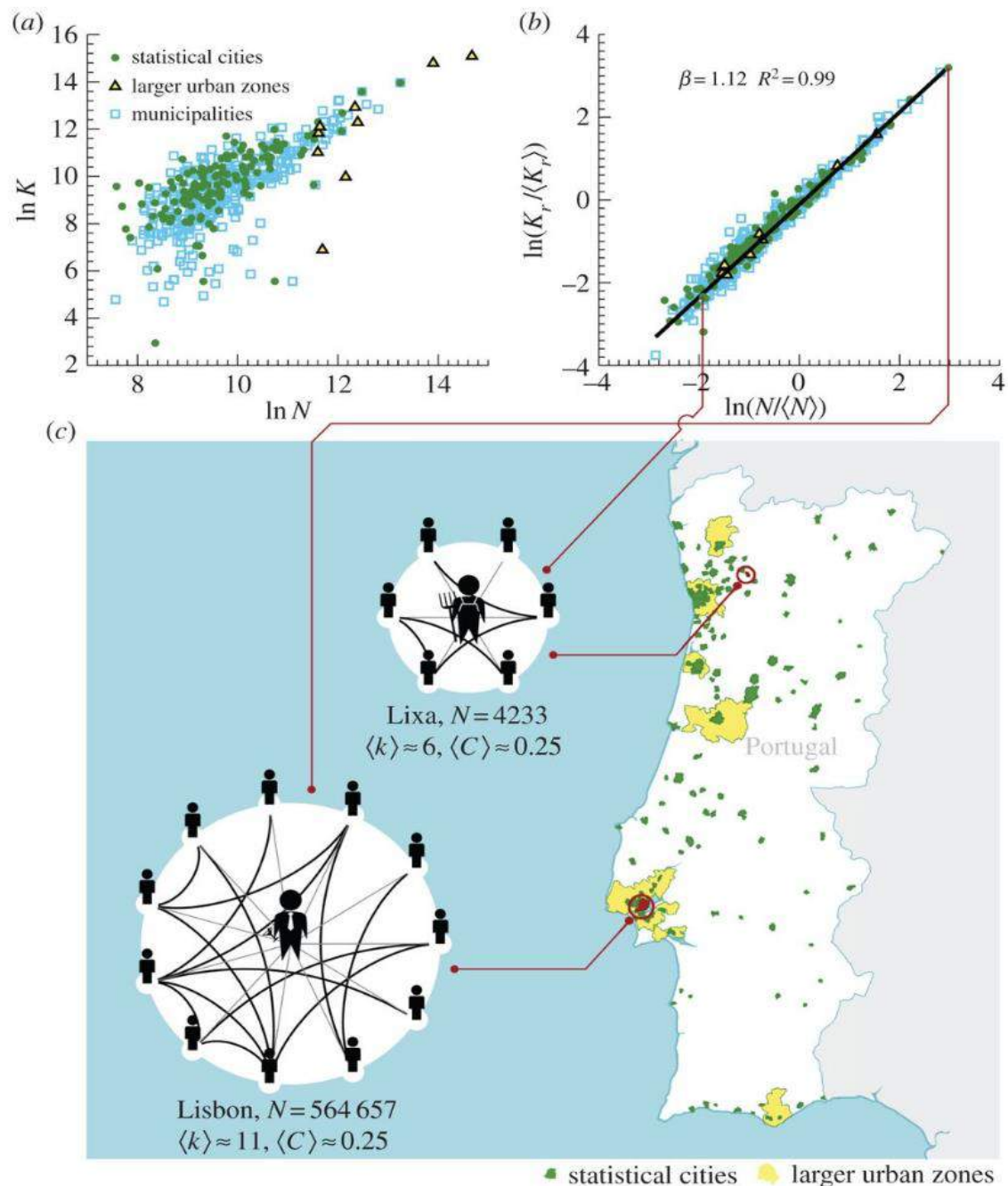


[PAN et al, Nature Comm, 2013]

用大数据来测试我们的理论

我们的论文发表一年后，研究员们利用来自葡萄牙几个重要城市的用户移动手机之间的通话记录，来测试我们关于信息流的理论是否正确：

“这些数字结果验证了城市的知识流速度随城市的大小而增加，符合最近提出的理论的结果。”



运用移动大数据来测量城市的脉搏 – 信息流



如何测量信息流

在城市的尺度测量移动用户的：

1. 各种层次的社交网络（ 通话， 面对面， 等等 ）
2. 移动位置信息



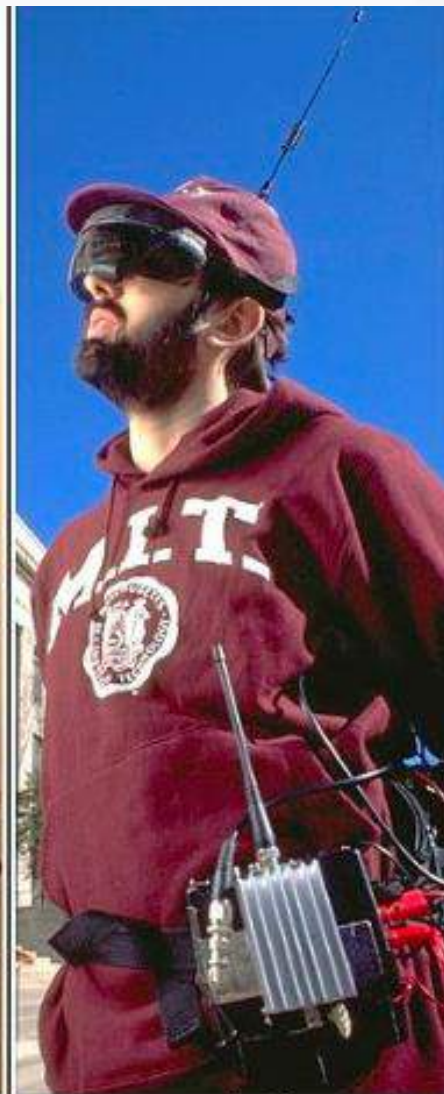
(a)
1980



(b)
Mid 1980s



(c)
Early 1990s

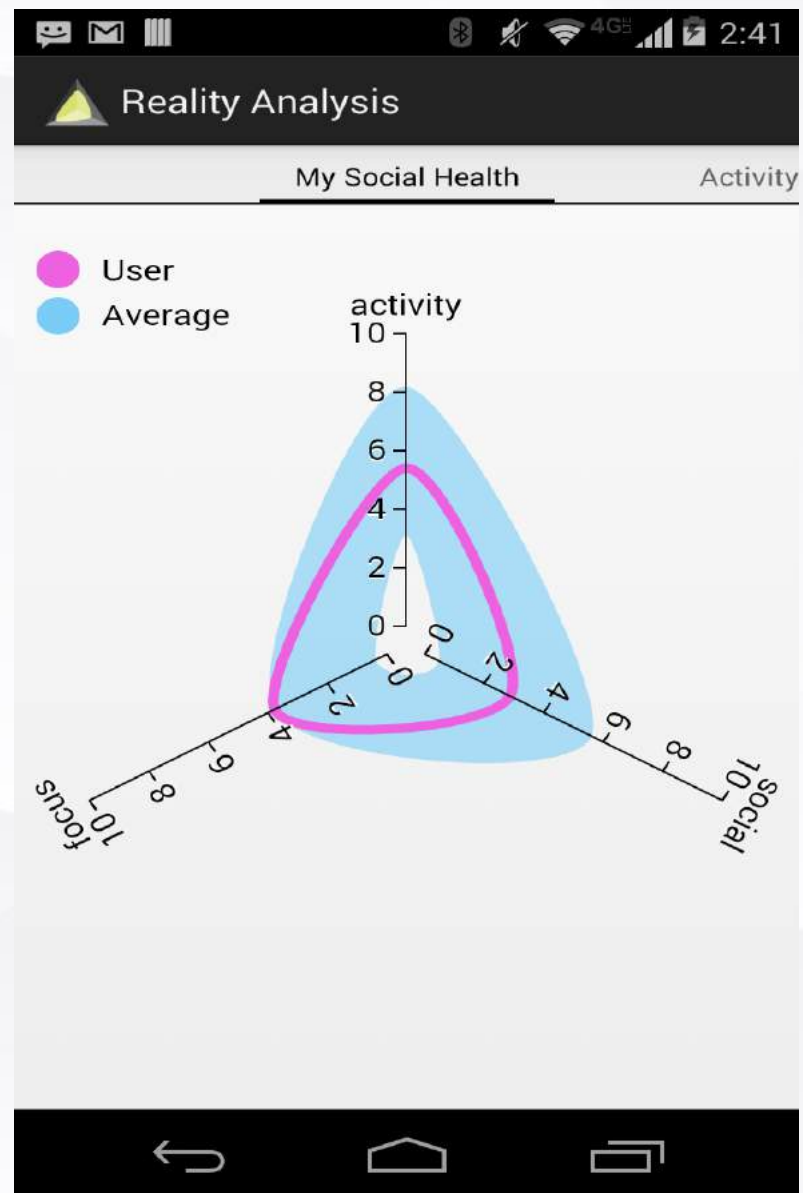


(d)
Mid 1990s



(e)
Late 1990s

通过计算社会学，
也许智能手机能为我们的社交行为提供智慧，
而不仅仅是一个社会的测量工具。





THANKS