

提示：采用手机 safari 微软翻译技术

## 1. 多样性组合对大规模 mimo isi 的影响

作者: [arkady molev shteiman](#), [Stefano galli](#), [laurence mailaasender](#), [xiao feng qi](#)

**抽象:** 我们分析了在基站 (bs) 中采用多样性组合技术时, 大规模多输入多输出 (m-mimo) 系统对符号间干扰 (isi) 的弹性。我们证明, 仅最大比率组合 (mrc) 就可以在天线数量无限制增长的情况下均衡 isi 通道。对于通道的性质, 必须根据到达角 (aoa) 的信息是否在接收方被利用, 假定存在其他限制。有趣的是, 更简单的相等增益组合器 (egc) 接收器也能够随着天线数量的增长而均衡通道, 但在这种情况下, 至少有一个通道必须是 ricean 褪色的。这些发现通过在 wssus 通道和使用光线跟踪引擎模拟在香港市中心和上海的实际 bs 部署生成的通道的模拟得到证实。最后, 观测到的缩放定律表明, 归一化 isi 功率降低  $n$ -折叠为每  $n$ -在 bs 的天线数量的两倍增加。少

2018 年 11 月 1 日提交;最初宣布 2018 年 11 月。

## 2. 运动模式的周期性塑造城市环境下的流行病风险

作者: [du zhanwei](#), [spencer j fox](#), [petter holme](#), [jying liu](#), [alison p.galvani](#), [lauren ancel meyers](#)

**摘要:** 人类流动性的日常变化调节了新出现疫情的速度和严重程度, 但大多数流行病学研究都假定存在静态接触模式。**中国上海**的人口流动性超过 2400 万, 是传染病进口和随后全球传播的高风险交通枢纽。在这里, 我们使用**基于上海**每小时过境数据的动态亚群模型来估计数千个疫情情景中的流行病风险。我们发现, 根据疾病传入的时间和邻里情况, 最初的流行病增长率变化超过 20 倍。风险最大的介绍是那些发生在市中心附近和周五-这桥梁工作日和周末交通模式, 从而连接否则断开的部分人口。确定这些时空热点可以为更有效的哨点监测目标和缓解传播的策略提供信息。少

**2018 年 9 月 13 日提交;最初宣布** 2018 年 9 月。

### 3. 维基百科网络对大学的世界影响和互动

**作者:**[célestin coquidé](#), [josélages](#), [dia l. shepelyansky](#)

**摘要:** 基于对 2017 年 5 月收集的 24 个维基百科版本网络的分析, 我们提供维基百科世界大学排名 (wrwu)。通过 pargrank 和 cheirank 算法, 我们确定了大学在这些版本的文化观点上的平均排名。与**上海**排名相比, 百强大学的重叠率为 6 0%, 表明 wrwu 对其历史发展给予了更多的重视。我们表明, 新的减少的谷歌矩阵算法允许在十个世纪的范围确定领先大学之间的互动。

这种做法还决定了特定大学对世界各国的影响。我们还比较了维基百科版本对大学意义和影响的不同文化观点。少

2018 年 9 月 2 日提交;最初宣布 2018 年 9 月。

#### 4. 关系提取复发胶囊网络在冠心病严重程度分类中的实际应用

作者:王琪,邱家辉,扬州\*,童阮\*,高大奇,朱高

**抽象:** 冠状动脉疾病 (cad) 是心血管疾病死亡的主要原因之一。cad 病情进展迅速, 如果不在早期诊断和治疗, 最终可能导致心脏肌肉死亡的不可逆转状态。侵入性冠状动脉造影是 cad 诊断的金标准技术。冠状动脉造影文本描述了哪个部分有狭窄, 有多少狭窄是在细节。对 cad 进行严重程度的分类是至关重要的。本文提出了一种复发性胶囊网络 (rcn), 用于提取中国冠状动脉造影文本中临床命名实体之间的语义关系, 通过该网络, 我们可以自动找出每个腔的最大狭窄, 从而推断如何推断严重的 cad 是根据根西尼的改进方法进行的。从上海曙光医院采集到的语料库实验结果表明, 我们提出的 rcn 模型达到了 f1-评分为 0.9641 的关系提取, 优于基线方法。少

2018 年 7 月 17 日提交;最初宣布 2018 年 7 月。

#### 5. 注意人群统计的头部位置

作者:张友梅,周春兰,张法良, alex c. kot

**文摘:** 遮挡、复杂背景、比例变化和非均匀分布对实际应用中的人群计数提出了巨大的挑战。本文提出了一种利用注意模型开发头部位置的新方法, 这是人群计数最重要的线索。注意模型估计了一个概率图, 其中高概率表示头部可能存在的位置。利用估计概率图从卷积神经网络的几个多尺度特征提取分支中抑制特征图中的非头部区域, 以进行人群密度估计, 使我们的方法对复杂的背景、尺度具有鲁棒性变化和不均匀分布。此外, 我们还引入了相对偏差损失来补偿常用的训练损失欧几里得距离, 以提高稀疏人群密度估计的准确性。在**上海科技**、UCF\_CC\_50 和世界博览会 10 数据集上的实验证明了我们的方法的有效性。少

**2018 年 6 月 26 日提交;最初宣布** 2018 年 6 月。

## 6. 上海城市电网数据揭示了什么

**作者:**王永坤,金耀辉,范波

**摘要:** 随着近十年经济的快速增长, 中国城市经历了巨大的变化, 同时, 城市电网管理数据的数量也出现了巨大。到目前为止, 对城市电网管理的研究并不常见。然而, 这种研究很重要, 因为城市网格数据描述了社区的个人行为 and 详细问题, 揭示了政策和社会关系变化的动态。本文对**上海城市电网管理数据**进行了初步研究, 探讨了这样一个快速发展的大都市地方政府与市民互动的主要特点。我们的调查说明了经济和生活环境之间的协同演变动态。我

他们还开发了数学模型，定量地发现数据中发现的事件之间的空间和时间内部关系，为地方政府提供洞察，以微调资源分配政策，并提供适当的激励措施来推动共同发展到最佳状态，从而实现善治。少

**2018 年 5 月 22 日提交;最初宣布** 2018 年 5 月。

## 7. 留还是离开：城市移民初期的变迁预测

**作者:**杨阳,刘宗涛, 谭晨浩,吴飞,岳亭庄,李亚丰

**摘要:** 在中国, 2.6 亿人迁移到城市, 以实现他们的城市梦想。尽管这些移民在快速城市化进程中发挥着重要作用, 但他们中的许多人未能安顿下来, 最终离开了城市。因此, 移徙者的融合进程给学者和决策者提出了一个重要问题。本文以上海为例, 调查了移民头几周的行为, 特别是他们的行为与提前离开的关系。我们的数据集包括一个为期一个月的完整数据集, 其中包括 5400 万用户之间的 698 个电信日志, 以及上海 18k 房地产的新的和公开的房价数据。我们发现, 最终提前离开的移民往往在最初几周既不发展多样化的联系, 也不在城市中流动。他们的活跃地区的房价也高于留在当地的移民。我们制定了一个搅乱预测问题, 以确定移民是否会根据她最初几天的行为离开。随着我们包含更多天的数据, 预测性能得到改善。有趣的是, 在使用相同的功能时, 仅从最

初几天训练的分类器就已经和使用完整数据训练的分类器一样好, 这表明性能差异主要在于功能之间的差异。少

**2018 年 3 月 18 日提交;**v1 于 2018 年 2 月 27 日提交;**最初宣布** 2018 年 2 月。

## 8. 用于保护位置隐私的人为冒名顶替者

**作者:**[王志阳](#),[谢志阳](#)

**文摘:** 基于位置的服务的进展导致了对位置隐私泄露的严重关切。为了有效和高效地保护位置隐私 (lpp), 现有的方法仍然不能完全胜任。他们往往在带有侧面信息的识别攻击下很脆弱, 或者由于计算复杂性高而难以实现。本文同时追求的是高保护效能和低计算复杂度。我们提出了一种基于假冒位置范式的可扩展 lpp 方法。为了使假地点非常合理, 我们通过合成人造冒名顶替者 (ai) 来伪造这些位置。ai 指的是与实际跟踪具有相似语义特征的合成跟踪, 并且不包含任何目标位置。设计了两种专用技术: 基于采样的合成方法和群体级语义模型。它们在综合认可机构的两个关键步骤中发挥着重要作用。我们在两个城市 (**中国上海**和西班牙阿斯图里亚斯) 对真实数据集进行了实验, 以验证该方法的高效度和可扩展性。在这两个数据集中, 实验结果表明, 我们的方法达到了 9765% 和 9612%, 它的运行时间, 建设发电机只 230.47 和

215.92 秒, 分别。这项研究将为研究界提供新的见解, 通过假冒地点提高最先进的 lpp 范式的实用性。少

**2018 年 7 月 29 日提交;**v1 于 2018 年 1 月 21 日提交;**最初宣布** 2018 年 1 月。

## 9. 使用兴趣点的智能手机应用程序使用情况预测

**作者:**[余东汉](#),[李勇](#),[徐凤丽](#),[张鹏宇](#),[瓦西里斯·科斯塔科斯](#)

**摘要:** 本文首次对智能手机上的应用使用进行了人口级的城市分析。通过网络运营商层面进行深度数据包检测, 我们获得了一个地理标记数据集, 拥有超过 600 万台独特设备, 在上海市各地推出了超过 10, 000 个独特的应用程序, 为期一周。我们开发了一种技术, 利用转移学习来预测哪些应用程序最受欢迎, 并根据该特定位置的兴趣点 (poi) 信息估计整个使用分布。我们证明, 我们的技术在成功识别五大热门应用方面具有 83.0% 的命中率, 在仅使用 10% 的采样稀疏数据估计使用情况时具有 0.15 mse。与现有的最先进方法相比, 它的性能高出约 25.7%。我们的发现为预测哪些应用与用户相关, 考虑到它们目前的位置, 哪些应用程序在哪里很受欢迎。我们的发现的含义是广泛的: 它使一系列系统能够从这种及时的预测中受益, 包括操作系统、网络运营商、应用商店、广告商和服务提供商。少

**2017 年 11 月 26 日提交;**最初宣布 2017 年 11 月。

## 10. 联合国数字蓝盔部队作为网络维持和平的起点

作者: [nikk 莱 akatyev](#), [joshua i. james](#)

**摘要:** 此前的作品, 如《塔林适用于网络战的国际法手册》, 重点关注网络战的情况。许多组织正在考虑如何进行网络战, 但很少有人讨论减少甚至预防网络冲突的方法。最近的一系列出版物开始制定网络维持和平框架及其法律要求。这些工作评估了国际电联 impact、北约 ccdcoe 和 **上海** 合作组织等组织的现状, 发现它们不符合有效主办 cpk 活动的要求。对目前在与 cpk 有关的领域开展工作的组织进行的评估发现, 联合国的任务和组织结构似乎与 cpk 的需要有些重叠。然而, 联合国目前的维持和平方针不能直接映射到网络空间。在这项研究中, 我们分析了联合国传统维持和平的发展, 以及目前在网络空间的举措。具体而言, 我们将把拟议的 cpk 框架与联合国最近提出的名为 "数字蓝盔" 的倡议以及联合国中有助于预测和缓解冲突的其他项目进行比较。我们的目标是为在联合国实施 cpk 框架找到切实可行的建议, 并研究 cpk 框架中界定的责任如何与 "数字蓝盔" 和全球脉动计划的责任重叠。少

2017 年 11 月 13 日提交;最初宣布 2017 年 11 月。

## 11. 利用 lstm 预测房价

作者: [陈晓晨](#), [赖伟](#), [徐嘉新](#)



**摘要:** 本文利用 2004 年 1 月至 2016 年 10 月的房价数据, 预测了北京、上海、广州和深圳各区 2016 年 11 月和 12 月的平均房价。我们应用自回归综合移动平均线模型生成基线, 而 lstm 网络建立预测模型。这些算法是根据平均平方误差进行比较的。结果表明, lstm 模型在预测时间序列方面具有良好的性能。并利用有状态的 lstm 网络和堆栈 lstm 网络进一步研究了房屋预测模型精度的提高。少

2017 年 9 月 25 日提交;最初宣布 2017 年 9 月。

## 12. 一种用于城市结构检测的概率嵌入聚类方法

**作者:**辛林,李海峰,张燕,高磊, 赵玲, 邓敏

**摘要:** 城市结构检测是城市地理学的一项基本任务。聚类是检测城市空间结构、城市功能区域等格局的核心技术。在大数据时代, 记录人类行为和人类社会活动等信息的各种城市传感数据集, 在高维、高噪声方面存在复杂性。遗憾的是, 最先进的聚类方法不能同时处理高维、高噪声问题。本文提出了一种概率嵌入聚类方法。首先, 我们提出了一个概率嵌入模型 (pem), 通过概率模型学习从高维城市传感数据中寻找潜在特征。通过潜在特征, 我们可以捕捉隐藏在高维数据中的基本特征 (称为模式); 利用概率模型, 还可以降低高噪声带来的不确定性。其次, 通过参数的调整, 我们的模型可以发现两种城市结构, 即同构同构和结构等价, 这意味

着社区在城市结构中具有密切的相互作用或相同的角色。我们通过对真实数据的实验和在上海(中国)的真实数据实验来评价了模型的性能,证明了我们的方法可以发现两种城市结构,即同构和结构等价,这意味着在城市空间中,以密集的互动或在相同的角色下聚集社区。少

2017 年 7 月 12 日提交;最初宣布 2017 年 7 月。

### 13. 流动人口的城市梦想--以上海市移民融合为例

作者:杨阳,谭晨浩,刘宗涛,吴飞,庄月亭

**摘要:** 前所未有的人类流动推动了世界各地的快速城市化。在中国, 1978 年至 2012 年期间, 城市人口比例从 17.9% 增加到 52.6。这种大规模的迁徙给决策者带来了挑战, 也给研究人员带来了重要问题。为了调查移民融合的过程, 我们在上海使用了为期一个月的电信元数据集, 拥有 5400 万用户和 6.98 亿通话记录。我们发现当地人和移民在他们的移动通信网络和地理位置之间存在系统性差异。例如, 移民在定居后, 接触的人比较多样化, 在城市周围的范围比当地人大。通过区分新移民(最近移居上海)和定居移民(在上海已经呆了一段时间), 我们展示了新移民在最初三周的融合过程。此外, 我们还制定了分类问题来预测一个人是否为移民。我们的分类器能够在区分定居移民和当地人时获得 0.82 分的 f1 分数, 但由于阶级不平衡, 确定新移民仍然具有挑战性。这种

分类设置为确定将成功融入当地人（被错误归类为本地人的新移民）的新移民提供了希望。少

2017 年 11 月 22 日提交;v1 于 2017 年 6 月 2 日提交;最初宣布 2017 年 6 月。

#### 14. 韵律: 言语的节奏与旋律

作者:[dafydd gibbon](#)

**摘要:** 本资料是关于韵律、节奏和语音旋律的部分教程, 以 2016 年 7 月在中国上海同济大学举行的当代语音和音系学暑期班开设的同名课程为基础。本教程并不是作为实验方法的介绍, 也不是作为关于该主题的文献的概述, 而是在计算的帮助下概述基本频率和时序模式的可观测访问方面可视化, 位于符号等级和解释的符号学框架。在对韵律的基本概念进行了非正式的介绍和对韵律在语言结构中的地位的讨论之后, 选择了音位和重音韵律、单词韵律、短语韵律的声学语音话题讨论了语篇韵律, 介绍了一种直观方面的文体方法。例如, 从一些不同类型的语言: anyi/agni (niger-congo & gt; kwa, 象牙海岸)、英语、kuki-thadou (中藏、印度东北部和缅甸)、普通话、tem (Niger-Congo & gt; gur, 多哥) 和波斯语。主要关注的是基本频率模式, 但也讨论了时间和节奏问题。最后一节概述了进一步的阅读和今后可能的研究方向。少

2017 年 4 月 27 日提交;v1 于 2017 年 4 月 9 日提交;最初宣布 2017 年 4 月。

## 15. 通过融合数据源产生的多个用户上下文推断

作者:徐金亮,王尚光,杨方春, 唐杰

**摘要:** 用户上下文信息的推断, 包括用户的性别、年龄、婚姻状况、位置等, 已被证明对构建上下文意识推荐系统很有价值。然而, 现有的关于用户上下文推断的现有研究有两个短篇: 1. 只关注一个数据源 (例如互联网浏览日志或移动电话记录), 2. 忽略多个用户环境的相互依存性 (例如年龄和婚姻状况之间的相互依存关系), 这导致推理效果不佳。为了解决这一问题, 本文首先利用张量外部乘积将特征空间中的多个数据源融合在一起, 以获得扩展用户特征表示。接下来, 通过以这种扩展用户特征表示为输入, 我们提出了一个名为 mulaprom 的多属性概率模型, 以推断可以利用它们之间相互依存关系的用户上下文。我们的研究基于来自中国上海当地移动运营商的大型电信数据集, 由两个数据源、460 万条通话明细记录和 8000 名移动用户的 750 万条数据流量记录组成。6 个月的课程。实验结果表明, 我们的模型在 \声援 {rec1}、\声援 {足额} 和 \指 f1-计量} 方面的性能可以优于其他模型。少

2017 年 3 月 16 日提交;v1 于 2017 年 3 月 12 日提交;最初宣布 2017 年 3 月。

## 16. 了解地铁交通流的时空活动模式

作者: [杜占伟](#), [杨博明](#), [刘继明](#)

**摘要:** 在城市交通系统中, 地铁系统中的流动反映了工作日的时空动态。为了研究移动流的变化, 我们通过一系列的地铁站在连续时期的快照来分析空间社区。以上海为例, 我们发现空间社区快照揭示了动态的乘客活动。采用双视角, 我们分别应用空间和时间模型来探索个人前往哪里、何时进行娱乐。在这两种模式中, 微博主题和餐饮场所、娱乐企业等空间设施被用来描述每个站的空间受欢迎程度和人们的出行感知。在所研究的案例中, 市中心的特点是社会影响较大, 空间模型更好地描述了它。在时间模型中, 较短的旅行距离会促使个人更早开始旅行。有趣的是, 随着起始站附近与食品有关的设施数量的增加, 直到超过 1563 年, 人们的出行速度也在放缓。本研究为模拟社会特征对移动性流量的影响以及预测新建地铁站的时空流动流提供了一种方法。

少

于 2017 年 2 月 19 日提交;<sup>v1</sup> 于 2017 年 2 月 5 日提交;最初宣布 2017 年 2 月。

## 17. 大数据技术加速基因组学精密医学

作者: [郝力](#)

**摘要:** 在基因组学生命科学研究中, 整个基因组学和生命科学算法的数据量越来越大, 计算为 tb、pb 或 eb 等。关键问题将是如何以优化的方式存储和分析数据。本文展示了英特尔大数据技术和体系结构如何帮助促进和加快数据存储和利用中的基因组学生命科学研究。英特尔为不同的调用数据查询定义了高性能 genomicsdb, 并为基因组学数据存储定义了带有层次结构存储管理的 lustre 文件系统。基于这些伟大的技术, 英特尔定义了基因组学知识共享和交换架构, 并在 bgi 中国和上海儿童医院获得了非常积极的反馈。而这些大数据技术肯定可以扩展到世界上更多的基因组学生命科学合作伙伴。少

**2017 年 1 月 29 日提交;最初宣布** 2017 年 1 月。

## 18. 互联电力系统与电气化交通网络的优化运行

**作者:**[m. hadi amini](#), [orkun karabasoglu](#)

**摘要:** 电气化交通和电力系统是相互耦合的网络。本文为相互依存的电力和交通网络建立了一个新的框架。我们的方法是解决一个迭代成本最低的车辆路径过程, 利用电动汽车 (ev) 与竞争充电站的通信来交换数据, 如电价、能源需求和到达时间。利用电动汽车电池管理系统的输入、充电站的电价、动力总成组件效率和交通网络流量, 使用 dijkstra 算法解决了电动汽车路由问题, 最大限度地降低了出行总成本条件。通过电动汽车与竞争充电站的

双向通信,可以更准确地进行电动汽车充电需求估计。然后,解决了电力系统的最优潮流问题,找出了充电站连接的负荷母线的位置边际价格。最后,电价从充电站到电动车,并关闭环路。区位电价是两个优化问题(即最优潮流问题和最优路由问题)之间的共享参数。电价取决于电力需求,这是受电动车充电的影响。另一方面,电动汽车充电站的位置及其不同的定价策略可能会影响电动汽车的路由决策。我们将电气化运输与电力系统运行相结合的新方法,在解决电气化交通问题和降低能源成本方面具有巨大潜力。利用上海交通网络和 IEEE 9 总线测试系统验证了该方法的有效性。结果验证了电力系统和交通网络的成本节约。少

2017 年 1 月 12 日提交;最初宣布 2017 年 1 月。

## 19. 无线分形蜂窝网络

作者:小虎·葛哥、邱叶红、陈家奇、黄美东、徐辉、徐静、张武雄、杨阳、王成祥、约翰·汤普森

摘要: 随着无线蜂窝网络在现代社会的无缝覆盖,考虑无线蜂窝覆盖的形式是一件有趣的事情。形状是一个规则的六边形,一个不规则的多边形,还是另一个复杂的几何形状?基于分形理论,利用从中国上海采集的无线蜂窝数据确定了无线蜂窝覆盖边界的统计特征。测量结果表明,无线蜂窝覆盖边界呈现极不规则的几何形状,也称为统计分形。此外,利用角尺度估计的赫斯特参数值

验证了无线蜂窝覆盖边界的统计分形特征。无线蜂窝覆盖边界的统计分形特征可用于评价和设计无线蜂窝网络中移动用户终端的切换方案。少

**2016 年 10 月 18 日提交;最初宣布** 2016 年 10 月。

## **20. 预测 2016 年 10 月 21 日至 2016 年 10 月 11 日期间使用人工神经网络预测未来上海股市价格**

**作者:**[barack wamkaya wanjawa](#)

**摘要:** 预测任何股市的股价，仍然是许多投资者和研究人员的追求。那些在股市交易的人倾向于在预测中使用技术、基本面或时间序列分析。这些方法通常是趋势的指导，而不是确切的可能价格。正因为如此，人工智能系统，如人工神经网络，是具有误差反向传播的前馈多层感知器，可用于此类预测。神经网络应用的一个难点是确定合适的网络参数。作者以前的一项研究已经将网络参数确定为 5:21:1, 80% 的训练数据或 4 年的训练数据，作为足够好的股票预测模型。在预测 2016 年 9 月 21 日至 2016 年 10 月 11 日期间选定的上海证券交易所股票时，该模型已受到考验，大约是在这些预测公布一周后。研究的目的是证实简单的神经网络系统在典型的股市预测中可以相当强大。少

**2016 年 9 月 17 日提交;最初宣布** 2016 年 9 月。



## 21. 基于百度地图查询数据的人群预警：基于上海踩踏的分析

作者:周景波,贝洪斌,吴海山

**摘要:** 如果没有充分的准备和现场管理, 大规模的意外庞大的人群是对公共安全的严重威胁。最近一个令人印象深刻的悲剧是 2014 年上海踩踏, 在上海外滩庆祝除夕, 36 人丧生, 49 人受伤。由于固有的随机和复杂的个人运动, 预测集体集会并不容易, 这可能会导致人群事件。本文利用百度地图上产生的海量数据, 提出了一种新的潜在人群灾害预警方法, 具有深远的公共效益。一个有见地的观察是, 随着移动地图服务的普及和方便, 用户通常会在百度地图上搜索来规划一个常规。因此, 在百度地图上聚合用户的查询数据, 可以获得先验和指示信息, 以便提前估计特定地区未来的人口。我们对百度地图数据的各种事件进行了仔细的分析 and 深入的研究, 也显示了地图查询数量与一个地区定位用户数量之间的强相关性模式。在此基础上, 我们提出了一种利用百度地图上的查询数据, 提前约 1-3 小时对潜在人群事件调用警告的决策方法。然后, 我们还建立了一个具有异构数据 (如查询数据和移动定位数据) 的机器学习模型, 以定量地测量潜在人群灾难的风险。我们评估了我们的方法对百度地图数据的有效性。少

2016 年 3 月 22 日提交;最初宣布 2016 年 3 月。

## 22. 维基百科世界大学排名

作者:[josélages](#), [antoine patt](#) , [dima l. shepelyansky](#)

**摘要:** 我们使用 24 个维基百科语言版本的文章之间的定向网络, 使用页面排名、2drank 和 cheirank 算法制作《世界大学维基百科排名》(wrwu)。这种方法允许利用独立于文化偏好的数学统计分析, 纳入对世界大学的各种文化观点。维基百科对百强大学的排名与上海大学排名重叠了约 60%, 这表明了这种方法的可靠特征。同时, wrwu 整合了 24 个维基百科版本积累的所有知识, 为历史上重要的大学提供了更有力的亮点, 导致对世界各国大学教育效率的不同估计。分析了大学排名在其十个世纪历史中的历史发展。少

**2016 年 2 月 4 日提交;**v1 于 2015 年 11 月 29 日提交;**最初宣布** 2015 年 11 月。

## 23. 车辆网络中数据转发的私有信息共享

作者:[李忠](#),[王成](#), [邵鲁洛](#), [姜长军](#)

**摘要:** 我们讨论了如何有效地转发数据在车辆网络。现有的解决方案没有充分利用附近车辆的轨迹规划或社会属性。机载导航系统的开发为驾驶员提供了一些出行路线信息。我们的方法的主要新颖之处在于设想将部分旅行信息共享到遇到的车辆上, 以便获得更好的服务。我们的数据转发算法利用了这种轻量级的信息下的虚假路径隐私保护以及车辆网络中的社会社区结构。我们假设数

据传输是由车辆和路边单元 (rsus) 传输的, 而蜂窝网络则管理和协调相关的全球信息。近似目标集是经常由目标车辆传递的 rsu 集。通过总结 rsu 在同一连接组件中的遭遇比, 提高了 rsu 的重要性。我们首先定义了时空可接近性的概念, 它来源于共享的局部行驶路线和遭遇信息。它描述了车辆向目的地转发消息的能力。然后, 我们设计了一种新的数据转发算法, 即基于可进性的数据转发算法, 该算法将车载网络中的时空可达性与社会社区属性结合起来。我们根据旧金山出租车发现和上海出租车运动的数据集评估我们基于可接近性的算法。结果表明, 部分共享的出行信息在车载网络数据转发中起着积极的作用。与现有的基于社会的数据转发算法相比, 基于可接近性的数据转发算法在车载网络中取得了更好的性能。少

**2016年10月24日提交,v1** 于 2015年2月19日提交;**最初宣布** 2015年2月。

## 24. 城市时空活动结构: 通过社交媒体签入数据推断城市内功能区域的新方法

**作者:**叶志,刘宇,王少文,邓敏, 高静,李海峰

**摘要:** 大多数现有的文献都集中在人类运动的外部时间节奏上, 以推断一个城市的功能区域, 但它们忽略了功能区域和人类活动之间潜在的相互依存关系, 而这些相互依存关系更详细地揭示了

区域的特征。在这项研究中, 我们提出了一个新的模型, 基于低等级近似 (上帝军), 以检测功能区域使用数据从大约 1, 500 万个值机记录在一个为期一年的时间在上海, **中国**. 我们发现了一系列潜在的结构, 称为城市时空活动结构 (ustas)。在解释这些结构时, 可以在空间活动模式和时间活动模式之间找到一系列突出的基础关联。此外, 我们不仅可以复制具有较低维度代表的观测数据, 还可以同时在同一坐标系中投射空间和时间活动模式。利用 k 均值聚类算法, 可以得到与时间活动的对应组合直接注释的五种重要聚类类型。这清楚地说明了各区域组如何在一天中的不同时间与不同的活动相联系。除了商业和交通的优势区域外, 我们还发现了两种住宅区, 发达的居民区和发展中的居民区。我们从城市形态分析的角度进一步验证了这些集群的空间分布。结果表明, 与同期政府规划高度一致, 表明我们的模型适用于通过社交媒体入住数据推断功能区域, 可从城市规划、公共等广泛领域受益服务和基于位置的推荐系统和其他用途。少

**2015年1月20日提交;**v1 于 2014年12月23日提交;**最初宣布** 2014年12月。

## 25. 大型手机呼叫网络统计特性的比较分析

作者:**李明霞**,**蒋志强**,**谢文杰**, 萨尔瓦托尔·米切奇, **米歇尔·图米内洛**, 周伟兴, **罗萨里奥** [n. mantegna](#)

**摘要:** 手机通话是现代社会应用最广泛的通信方式之一。手机用户之间的通话记录为理解社交网络中嵌入的人类通信模式提供了宝贵的代理。移动电话用户相互呼叫, 形成一个定向呼叫网络。如果只考虑对等呼叫, 我们会获得一个无方向的相互呼叫网络。可以对两个连接用户之间的优先通信行为进行统计测试, 并生成两个具有统计验证边缘的邦费罗尼网络。我们对这四个网络的统计特性进行了比较分析, 这些数据是根据上海 900 多万人在 110 天时间里的通话记录构建的。我们发现, 与以前研究过的大型移动呼叫网络相比, 这些网络具有许多共同的结构特性, 也表现出独特的特征。实证研究结果为我们提供了一个有趣的图片, 一个具有代表性的大型社交网络, 可能会给大型社交网络的建模带来新的启示。少

**2014 年 5 月 6 日提交;v1** 于 2014 年 2 月 25 日提交;**最初宣布** 2014 年 2 月。

## 26. 论上海排名的内在动力

**作者:**[domingo docampo](#), [lawrence cram](#)

**摘要:** 上海交通大学研究人员公布的《世界大学学术排名》已成为高校管理人员、国家官员、学生和广大公众的主要信息来源。最近关于其内部动态的发现允许反转已公布的 arwu 指标分数, 以重建 500 所世界一流大学的原始分数。本文探讨了 arwu 和其他

竞赛中的原始分数,以对比排名驱动和分数驱动表的动态,并解释为什么 arwu 排名是一个分数驱动的过程。我们表明, arwu 指标构成了研究业绩的单一因素核算的子尺度,并提供了 arwu 使用的收益和非线性系统的说明。本文讨论了 arwu 所选择的非线性,认为它们是用来表示测量研究性能的指标的回归特性的。我们建议,可以通过替换基于表现最好的人的原始分数的年度重新缩放的不必要的动态影响来大大提高 arwu 的效用和可用性。少

2013 年 9 月 26 日提交;最初宣布 2013 年 9 月。

## 27.从企业技术、组织和地理属性的协同作用来衡量中国的知识经济

作者:[loet leydesdorff](#), [ping zhou](#)

**摘要:** 利用 orbis 数据库中企业的地理、规模和技术分布之间可能存在的协同作用,我们发现,在中国 31 个省的水平上,不确定性减少幅度最大,在国家一级增加 18.0。一些沿海省份如预期的那样突出,但北京和上海的大都市地区(天连和重庆)在下一个较低的行政级别(339),因为这四个大都市是在这两个级别上都定义了管理定义。以高科技和中等技术制造业为重点,表示向北京和上海转移,协同作用平均得到加强(不出所料;但并非所有省份都有)。遗憾的是, orbis 数据不完整,因为这些数据是为商业目的而不是为行政或政府目的收集的。然而,我们展示了一种方

法, 其他人可以使用这种方法, 这些人可能可以获得更高质量的统计数据测量。少

2013年10月29日提交;v1于2013年5月27日提交;最初宣布2013年5月。

## 28. 对国际背景下国家大学排名重要性的洞察: 西班牙大学 i-ugr 排名案例

作者: nicolás roinson-garcía, daniel torres-salinas, emilio delgado lópez-cózar, francisco herrera

**摘要:** 国际排名在研究政策领域所取得的巨大重要性, 警告这些工具存在的缺陷和缺点所带来的许多威胁后果。其中之一与无法准确代表国家大学系统有关, 因为他们最初的目的只是对世界级大学进行排名。另一个问题是大学纪律简介缺乏代表性, 因为它们通常提供一个独特的表格。尽管一些排名提供了很好的覆盖面, 另一些排名提供了按领域分列的排行榜, 但没有一个国际排名做到两者兼而有之。为了超越这种局限性, 本文从研究政策的角度分析了利用国家排名来补充国际排名的可能性。为此, 我们分析了西班牙大学系统作为一个研究案例, 按领域和子领域介绍了西班牙大学的 i-ugr 排名。然后, 我们将他们的结果与上海排名、qs 排名、莱顿排名和台大排名获得的结果进行比较, 因为它们都有基本的共同点, 可以进行这样的比较。我们的结论是, 最好使用

国家排名来补充国际排名，但我们认为，必须谨慎行事，因为它们所采用的方法以及在实地建设方面存在差异。少

**2014 年 3 月 3 日提交;**于 2013 年 5 月 6 日提交;**最初宣布** 2013 年 5 月。

## 29. 在路网上提供服务保证的大型实时岭处理

**作者:**[严黄](#),[金若明](#),[巴斯塔尼](#),[王晓阳](#)

**文摘:** 在美国，个人车辆旅行的平均占用率只有每辆车 1.6 人。城市交通拥堵是一个熟悉的场景。里沙林有可能解决许多环境、拥堵和能源问题。本文介绍了路网上具有服务保障的大规模实时骑行问题。在满足等待时间和服务时间限制的同时，服务器和行程请求会动态匹配。我们首先提出了两种基本算法：分支绑定算法和整数编程算法。然而，这些算法结构并不能很好地适应骑行问题的动态性质。然后，我们提出了一种能够更好地调度动态请求和动态调整路径的动力学树算法。我们在上海的大型出租车数据集上进行实验。结果表明，动力学树算法在响应时间上比其他算法快。少

**2013 年 2 月 27 日提交;**最初宣布 2013 年 2 月。

## 30. 在机会主义网络中提供信息传播时间的概率保证

**作者:**[yoor kim](#), [kyunghan lee](#), [ness b. shroff](#), [injong rhee](#)



**摘要:** 已经开发了各种数学工具来预测在各种环境中的传播模式, 包括传染病、计算机病毒和向移动代理 (如人类、车辆和移动代理) 广播的紧急信息设备)。这些工具主要侧重于估计点差达到一定比例的平均时间 (例如,  $\alpha\alpha$ ) 的代理, 即所谓的平均完成时间  $e(t\alpha\alpha)$ . 我们声称, 在点差的时间提供概率保证  $t\alpha\alpha$  而不仅仅是其平均值, 它使人们对传播有了更好的了解, 因此可以用来设计预防流行病的改进方法或设计加速的数据分发方法。为了展示其优点, 我们引入了一个新的度量  $G_{\alpha\alpha, \beta}$  表示保证所需的时间  $\alpha\alpha$  完成与概率  $\beta$ , 并制定一个新的框架, 以确定  $t\alpha\alpha$  用于各种扩散参数, 如种子数量、接触速率水平和接触速率的异质性。我们将我们的技术应用于上海出租车的实验移动跟踪, 并表明我们的框架使我们能够分配资源 (即控制扩散参数), 以比最先进的。少

**2013 年 1 月 10 日提交;最初宣布** 2013 年 1 月。