

公交江湖 | 公共交通与小汽车出行时长的差异

简介

小汽车使用的增加会给环境带来许多负面影响，包括交通拥堵、停车矛盾、空气污染加剧和增加温室气体排放。而公共交通则是一种低成本、高效率、低污染并推进社会公平的出行方式。全球许多地方的政府已认识到在城市推动从小汽车向公共交通出行转移的重要性，这将是解决负面环境影响、增加公平和应对气候变化的政府决策。例如，德国柏林承诺从 2019 年起投入 280 亿欧元来改善城市公共交通。

然而，抑制小汽车出行、鼓励使用公共交通并不是那么简单，尽管投入巨大，但是在美国和其他地方，公共交通出行分担率一直在持续下降。公交客运量增长会受到许多因素的制约，如线路与运营时刻固定、覆盖人口密度较低以及出行者的基本态度。在这些制约条件中，出行时间和建成环境被认为是影响交通方式选择最关键因素。公交客流量增长的一个主要驱动力是乘客感知的边界成本（包括出行时间）。出行时间是量化公共交通服务质量的关键指标。在一篇综述论文中，Redman 等人总结了针对不同方面提升公共交通服务的策略，发现大多数研究都将速度作为提高公交客流量的关键因素。在纽约的一项研究中表明，通勤时间缩短 15 分钟意味着轨道交通出行客流量将增加约 25%。一项针对小汽车用户出行后的问卷调查表明，更短的出行时间是让公共交通更有吸引力的关键因素之一。

虽然有强有力的证据表明小汽车出行往往比公共交通更快，但这两种出行模式的时间差的时空变化趋势也许会有助于城市规划及政策制定，以倡导小汽车向公交出行转移。最近的研究表明，当越来越多的研究开始使用 GPS 和移动手机来追踪出行速度，研究人员发现公共交通出行时间会依据时段、路线有很大的不同，同时城市交通整体运行状况也会随一天中不同的交通拥挤情况有所差异。传统上关于出行时间的计算在很大程度上依赖于静态、简单的假设：使用恒定的车辆速度，而忽略了一天中出行速度的变化。

当前，越来越多的文献开始使用详细的空间数据和随时间变化的交通运行数据来掌握小汽车和公共交通在出行时间上的时空差异，这为本研究中对不同出行方式出行时间的差异进行更现实的评估提供了机会。一些新兴的数据源和地理信息系统(GIS)极大地提高了挖掘城市交通系统运行特征数据的可行性，可用数据包括交通速度数据、出租车 GPS 数据和公交智能卡数据。实时交通速度数据的为出行者提供了更先进的出行信息，以用来进行路径选择与交通规划。新的数据源如 HERE Traffic，到目前为止已覆盖了 83 个国家的城市，可以收集和提供关于实时道路速度、事故和事故的信息。可用的数据量以及数据的时空高分辨率为我们客观估计出行时间提供了良好的基础。同时开放数据的兴起也支持计算更准

确的出行时间。开放数据标准，如通用交通供给规范（GTFS）和众包倡议，如开放街道地图（OSM），他们提供的书籍可用于估计公共交通的出行时间和实时调度信息。

在可达性研究中，小汽车和公共交通的出行时间通常是依据假定目的地来估计的，或者局限于具有特定功能的地点，如工作场所或医院。在可达性分析中，地方之间的实际出行需求并不是主要关注的焦点，而研究往往仅关注从一个固定地点和一个地区出行的“可达性潜力”。然而，当考虑到运输需求和运输服务如何随着时间和地点的变化而变化，两地之间的人口流动需求对于评估实际需求下的交通状况是至关重要的。尽管关于小汽车与公共交出行时间差异的研究文献越来越多，但考虑实际出行需求时，还有待进一步研究这种时间差异是如何变化的。充分了解这两种交通方式在出行时间上的差异有助于确定公共交通在何时何地与小汽车相比存在更大的竞争优势，同时也并有助于了解必须依赖公共交通的社区成员在交通出行方面的相对劣势。大规模、有代表性的动态出行需求数据对于更真实地评估这一时间差异是至关重要的。

与其他传统/新数据源（如出行调查、交通流量数据和 OD 出行矩阵）相比，带有地理标记的社交媒体数据（如带有指定位置的 GPS 坐标的 tweet)已被发现能较好的反映出行需求。不像 OD 矩阵通常只能捕获每日平均值，以 Twitter 数据为代表的社交媒体数据能记录不同地理做坐标下的发文密度，并合理地捕获人们在何时何地从事何种活动的精确信息，因此它能较低成本地获取城市动态出行需求数据。

本研究利用多个大尺度数据源，以精细的分辨率捕捉四个不同城市（巴西圣保罗、瑞典斯德哥尔摩、澳大利亚悉尼和荷兰的阿姆斯特丹）小汽车和公共交通出行时间的时空分布特征。本研究计算了工作日中两种交通方式平均出行时间详细的时空变化，从而帮助我们更精确地掌握公共交通和小汽车之间的出行时间差异。我们将多个数据源整合在一起使用，包括用 HERE Traffic 数据获取一年来道路运行速度、过去 9 年 Twitter 数据、最新的 GTFS 公共交通数据库和 OpenStreetMap 的道路网络。每个城市被划分成由 500m 六边形单元构成的网格系统，系统内的每个单元不同方式的出行时间是根据该单元到任何高访问量单元(目的地)门到门的出行时间取算数平均值所得，而这些高访问量单元则是基于带有地理标记的 Tweet 来确定的。具体而言，在选定的时段内(例如，8:10 am 到 8:25 am)，某一起点的平均出行时间被定义为从该起点到多个目的地的出行时间的平均值，这些目的地的地理标记 tweets 量则作为权重以计算平均值。为了量化出行时间的差异，我们使用了出行时间比率 R（R 定义为给定出发时间时段、特定 OD 对小汽车出行时间除以公共交通出行时间）。最后，我们对结果进行可视化分析，以展示在所有研究城市中小汽车和公共交通出行时间是如何在时空上变化的，并对这些城市研究结果展开系统性的跨区域比较。

结果

出行时间的时空变化

图一中显示了几个到访人数较多地区的平均出行时间随时间的变化。两种交通模式也显示出截然不同的时间变化趋势。对于小汽车，Sydney 和 Stockholm 都有早晚高峰小时，但各不相同。São Paulo 和 Amsterdam 在白天出行时间随时间不断增加，且没有明显的高峰点。对于公共交通，从午夜到凌晨时段，公交服务暂停或低频次的时候，出行时间通常较长。在 São Paulo，公交线路可以覆盖整个城市路网的 80%，而 Stockholm，Sydney 和 Amsterdam 的覆盖率相对较低。根据某一个工作日对不同道路网格到达某一地点的可达性分析，可以看出 Sydney 的覆盖率随时间的波动幅度最大。图二中显示了各城市各起点的出行时间比（R）的时空变化。在高峰小时，R 在空间上的变化相比一天内的其他时间较小。在 Amsterdam，可到达某一种点的道路网格约占 40%，是所有城市中最小的（图 1），而这是因为其多城市中心的城市结构以及公交线路多集中在 Amsterdam 的交通走廊（图 2）。

通过对比公交和小汽车使用比和人口密度分布，出行时间比 R 在图 3 中显示了每天的平均出行时间在空间上的分布。对所有研究地区，每个交通网格中，公共交通使用率约是小汽车使用率的两倍。在这四个地区中，R 的空间分布反映了这些城市的空间分布，因此在城市中心区域，人口和活动密度较大的地区，以及靠近公交基础设施的地区小汽车比公交的优势较小。

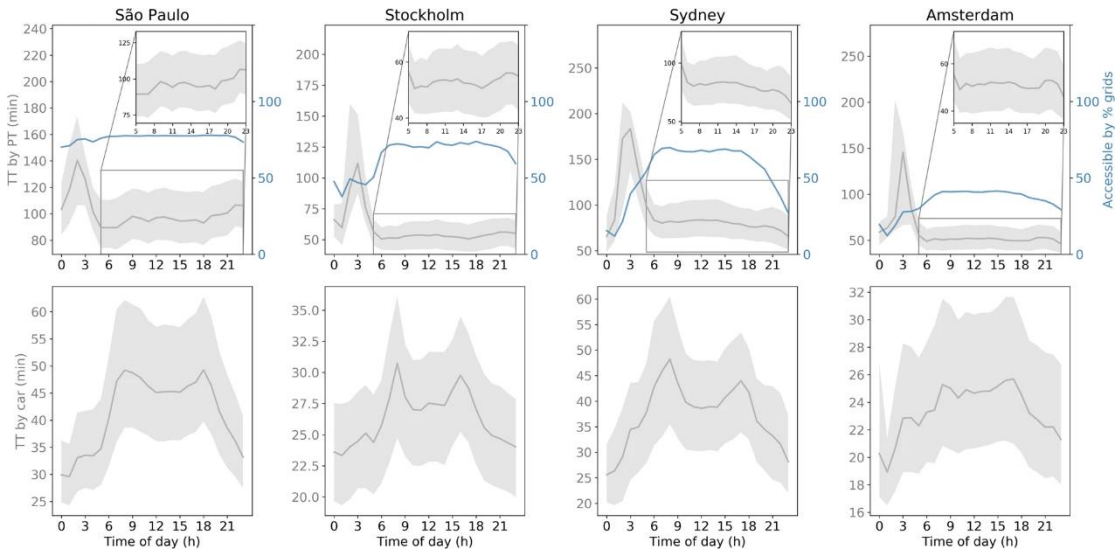


图 1. 在一个工作日中，公共交通（第一行）和小汽车（第二行）的出行时间的时间变化。这里的出行时间是根据全程不同的起点，根据人口密度作为权重计算出的平均出行时间。阴影区域指出了出行时间的 25 分位数和 75 分位数，还显示了白天公交可达的交通网格覆盖率。插图中放大了从 5 时到 23 时的数据，以更好的展示公交出行时间的变化。

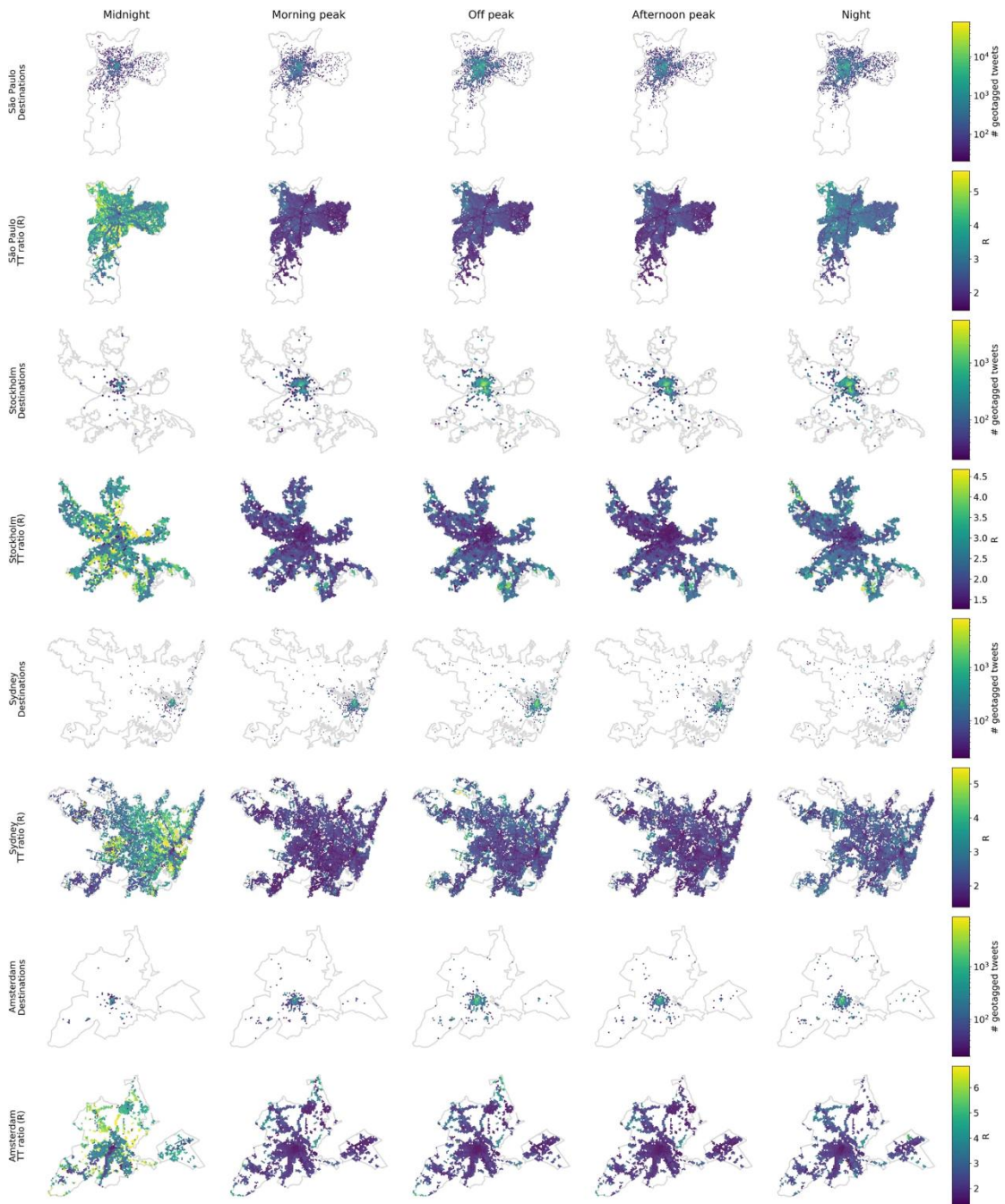


图 2. 一个工作日中，出行时间比 (R) 的时空变化。出行时间比 R 是平均公交出行时间和平均小汽车出行时间之比。对每个城市，上面一行显示了以该位置推特数为权重计算出的出行终点，而下面一行显示了在每个时间段内，到达这些地点的加权平均出行时间。凌晨：0:00-7:00，早高峰：7:00-10:00，非高峰：10:00-16:00，晚高峰

16:00-19:00, 晚间: 19:00-0:00。为表达清晰, 这些地图的空间比例不统一。

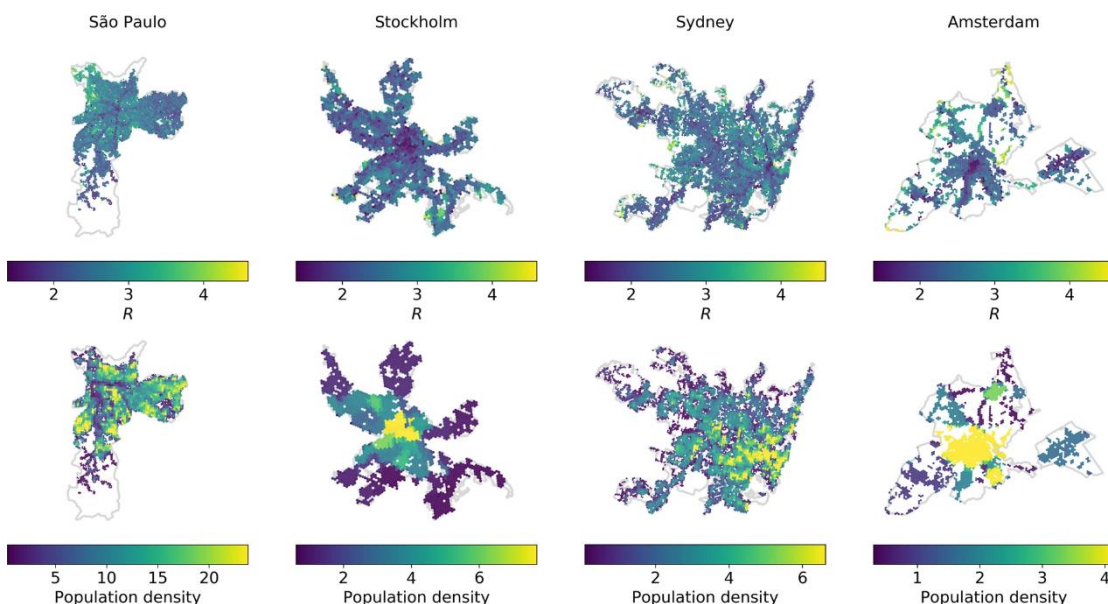


图 3. 到访人数较多的城市的出行时间比 (R) (第一行) 和人口密度 (第二行) 的空间分布。R 值是以每一个网格作为起点, 从 5 时到 23 时公共交通和小汽车的 5 分位到 95 分位的出行时间的加权均值, 权重为终点网格中显示位置的推特数 (下行的权重同上行)。颜色越暖 (黄), 小汽车比公交的优势越大。人口密度的单位为 1000 人每平方千米。

R 的时间变化如图 4 (a) 所示, 可以看出白天绝大多数时间平均出行时间比在 2 左右, 在凌晨 (午夜到天亮前), 公交服务在减少或根本不运营的时间段中, 两种出行模式之间的出行时间差别最大。在 São Paulo, Sydney, Stockholm 和 Amsterdam 四个城市中, 相比小汽车更喜欢使用公共交通的地区覆盖率分别为: 0.62%, 0.44%, 1.10%和 1.16% (日平均); 0.65%, 0.48%, 1.22%和 1.19% (高峰小时)。图 4 (b) 显示了随着小汽车出行时间的增加, 出行时间比是如何变化的。可以看出对于短程出行, 公共交通更有优势。而对于小汽车出行时间超过 18 分钟的出行, 小汽车的优势就出现了, 并且对于更远距离的出行, 两种出行模式之间的区别就随距离增大而缓慢减小。

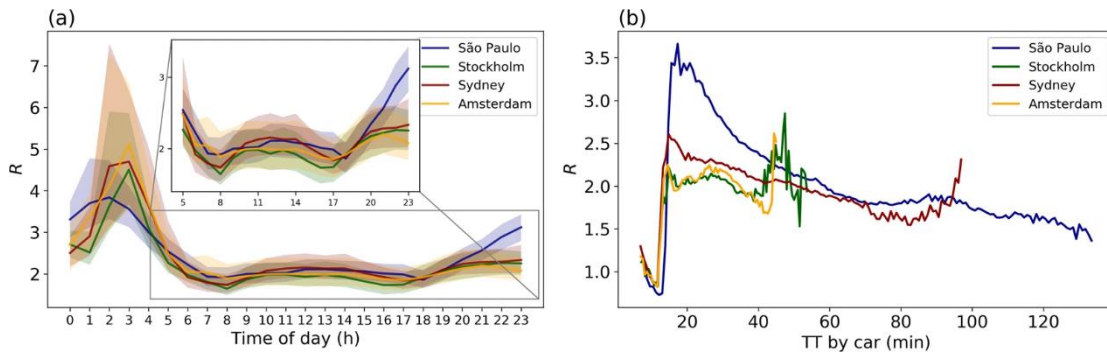


图 4. (a) 全城的平均出行时间比 (R) 的时间变化。阴影区域代表上下四分位数。插入的放大图显示了 5 时到 23 时的数据，以更清楚的显示 R 的时间变化。 (b) 出行时间比 (R) 随小汽车出行时间的变化。

图 5 (a) 显示了小汽车和公共交通的出行时间随出行距离变化的变化（以小汽车为参照）。在相同的出行距离情况下，这些城市的小汽车出行时间各不相同。São Paulo 因其交通拥堵严重，在相同的出行距离下，其小汽车出行时间最长。在出行距离在 0-40km 时，其他三个城市的小汽车出行时间相似，其中 Amsterdam 的最短。对于 São Paulo 和 Sydney，虽然城市大小相似，但随着出行距离的增加，这两座城市的出行时间表现出不同的变化趋势，即随着出行距离的增加，São Paulo 的出行时间增速更快，推测是因为其拥堵更严重。São Paulo 的公共交通出行时间最长，但当出行距离相同时，其他三个城市的公交出行时间相似。总之，对短距离出行（0-20km），出行时间随出行距离增长而快速增长；对中距离出行（20-40km），出行距离的增速较低；对于长距离出行（在 São Paulo > 30km，在 Sydney > 50km），出行时间对于出行距离的增加再次变得敏感。

图 5 (b) 显示了研究地区到达这几个城市，乘坐小汽车比乘坐公交要快的人数，随着人数的增加，出行时间迅速上升且 São Paulo 和 Sydney 的数据有长尾（例如：离终点距离最远的 10% 的人，他们的平均出行时间大约是离终点距离最近的 10% 的人的三倍）。在图 5 (c) 中，当出行距离 < 3km 时， R 会小于 1（乘公交比乘小汽车快），但随着距离的增加，公交很快就失去了其优势。除了 Stockholm，随着出行距离的增加，其他几个城市的 R 变化趋势很相似：即在出行距离小于 15km 时，公交和小汽车出行时间的差距不断增加，但出行距离大于 15km 时，这两者之间的差开始减小。这是因为公交出行时间的增速较小汽车慢，此外，如图 5 (d) 所示，人口密度和 R 值是相关的，即人口密度越大，公交和小汽车的出行时间越相似。

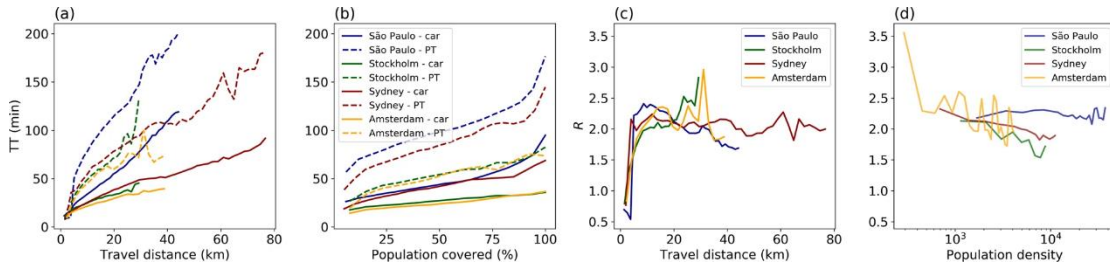


图 5. 出行时间和出行时间比 vs. 出行距离和人口密度。出行时间随 (a) 出行距离和 (b) 可达人口覆盖率的变化。出行时间比 (R) 和 (c) 出行距离和 (d) 人口密度之间的关系。人口密度的单位为 1 人/平方千米。

站在城市角度，表 1 总结了公交和小汽车的各种参数，包括出行时间比和总行程速度。研究地区的出行时间比（未加权）均值分别为 2.2（São Paulo），2.0（Stockholm），2.2（Sydney）和 2.2（Amsterdam），各城市中两种交通模式差别最小的就是 Stockholm。各城市的出行时间比以人口密度为权重的值分别为 1.4，2.6，2.3，和 2.1（São Paulo，Stockholm，Sydney 和 Amsterdam），这说明 São Paulo 和 Amsterdam 的公交服务和其居民居住地很好的配合，然而 Stockholm 和 Sydney 更注重公交服务的空间覆盖率。即使 Stockholm 的公交和小汽车差别最小，但当平均出行时间比经过人口密度加权后，其成为两种出行模式差别最大的城市。而 São Paulo 恰好相反，在人口密度加权前后，它成为公交和小汽车差别最小的城市。而人口密度加权计算对公交的速度影响很小。通常行驶速度低就意味着交通拥堵的严重，而这就解释了为什么 São Paulo 的公交和小汽车的出行时间差距最小了。

城市	出行时间比	出行时间比 _{pop} ^a	速度 (km/h)		速度 _{pop} (km/h) ^b	
			小汽车	公共交通	小汽车	公共交通
São Paulo	2.2	1.4	19.4	9.2	19.9	14.3
Stockholm	2.0	2.6	25.7	12.9	37.6	14.9
Sydney	2.2	2.3	33.8	16.6	30.9	13.9
Amsterdam	2.2	2.1	31.5	15.0	27.6	13.7

表 1. 城市层级的出行时间比。a、b 表示在每个交通网格的人口密度加权的均值。城市层级的出行时间比是基于以显示了终点定位的推特数为权重的所有交通网格全天数据的均值。

讨论

文献中列举了许多可达性指标，用于评估在给定时间内，以限定交通方式到达不同目的地的难易程度。其中，出行时间是最核心的指标。然而，对比理论预测，城市中的不同出行需求可能会显著改变出行时间。在这项研究中，我们从一个不同的角度出发去评估交通系统在实际需求下的表现。利用 Twitter 数据随时间变化的出行需求，我们的方法将交通运输服务需求和运营联系起来，从而可以在城市中的两种模式（小汽车和公共交通）之间进行更现实的比较。该方法可以揭示最大的差距所在，并向规划人员和政策制定者告知，在当今的需求下，改进将产生最大影响的领域。该方法可以揭示最大的差距所在，并告知规划人员和政策制定者，在现在的需求下，改进将产生最大影响的区域。

我们提出了一个可重复且可扩展的数据融合框架，我们结合了新兴数据源，包括 Twitter，HERE Traffic，OSM（openstreetmap 数据）和 GTFS（公共交通通用数据标准），以对小汽车和公共交通的出行时间进行建模。HERE Traffic 提供了范围广泛的所研究城市真实驾驶速度的高频数据。我们将 GTFS 数据和行程规划一起作为“高级”方法，因为 GTFS 数据已经在路线规划中考虑了与拥堵相关的延误。除了实时道路速度和 GTFS 数据外，我们还引入了社交媒体数据，在本例中为 Twitter，以表示平均工作日每天的实际需求位置，我们使用 Twitter 代表实际出行需求的有效性。通过揭示高时空颗粒度下出行时间的差异以及这些模式在不同城市之间的变化，证明了该框架的有用性。针对四个研究区域，图 1 和图 2 详细描述了小汽车和公共交通出行时间的时空变化。它们显示了每个模式的出行时间在平均工作日的一天中如何变化，以及出行时间在不同城市中是如何发生空间变化的。未来的研究可以专注于并覆盖基础设施信息，从而在局部获得更详细的见解。这使人们能够更好地了解城市规划政策，尤其是在鼓励从小汽车出行到使用公共交通的模式转变方面。虽然公共交通出行的平均时间是小汽车乘车时间的两倍，但这种差异随位置（图 2）和一天中的时间（图 5）的不同而有很大差异。出行时间差异与以前的研究一致，但是现在得到了更好，更详细的时空数据和实际出行需求的支持。

我们介绍不同城市之间出行时间在时空上如何变化，并进一步探讨了出行距离和人口密度对城市间出行时间的影响（图 5（a，b））。当行驶距离增加时，小汽车和公共交通的出行时间以不同的方式增加（图 5a）。与其他城市相比，São Paulo 的小汽车和公共交通出行时间明显更长。为了覆盖总人口的相同百分比，乘坐公共交通比开车要花费更长的

时间（大约 2 倍，见图 5b）。而且，São Paulo 和 Sydney 这两个大城市的公共交通对居住在远离目的地的最后 10% 的人口平均出行时间约为前 10% 的 3 倍（图 5b）。

关于出行时间的模型差异及其与出行距离和人口密度的关系（图 5c, d），结果与期望相符，对于短距离出行，在城市中心出行以及接近公共交通路线线的情况下公共交通可以超越小汽车。以前的研究在赫尔辛基大区也发现了类似的模式。但是，重要的是要认识到，就出行时间而言，只有极少数网格单元的公共交通优于小汽车。在短距离（<3 公里）内，公共交通比小汽车话费时间更短，尤其是在在 Stockholm 和 Amsterdam 的高峰时段，那时更偏向于公共交通出行的网格单元的比例在 Amsterdam 是 0.8%，在 Stockholm 是 0.6%。在 Sydney 和 São Paulo，这个数字甚至更少。

我们总结了所研究的四个城市的模态差异的城市平均值（表 2）。在城市级别（网格单元由人口密度加权），在 São Paulo 观察到最低的出行时间比率，其次是 Amsterdam，Sydney 和 Stockholm。São Paulo 的交通繁忙解释了这一现象。但是，交通拥堵实际上也会影响公交车的出行时间。在 São Paulo 数据显示出公交车排班与公交车的实际位置之间存在很大的不一致，因此，公交车 GPS 数据已用于解决考虑了拥堵问题。但是，在本研究中，我们没有包括交通拥堵对公共交通的影响。因此，在 São Paulo 我们可能高估了公共交通相对于小汽车出行的优势。对于城市规划者和政策制定者，可以通过实时运行表来识别和解决此类不一致问题。

表 2、四大城市的主要特征和交通方式比例

城市	范围	面积 (km ²)	人口	GPD (美元/ 人)	交通方式比例 (%)			
					小汽车	公共交通	步行	骑行
São Paulo	Municipality	1521	11,967,825	9812	28	31	31	1
Stockholm	Urban Area	414	1,372,565	53253	32	47	14	7
Sydney	UCL	2036	4,321,535	49755	59	25	4	3
Amsterdam	Urban Area	885	1,520,127	45638	20	17	29	32

在选择出行模式时出行时间是最明确的指标，尤其是当公共交通是小汽车的替代出行模式时。在我们的研究中，公共交通在出行时间上可以优于小汽车使用的区域很小，但是在公共交通路线周围还有相当大的区域公共交通出行和小汽车的出行时间差距比其他城市小。在 Amsterdam 对小汽车驾驶员的一项调查显示，当知道的出行时间比率降至 1.6 以下时，参与者愿意乘公共交通。该调查中记录的小汽车出行平均出行时间为 35 分钟。但是，我们的研究表明，对于小汽车行驶时间约为 35 分钟的出行，出行时间比率最高为 2（见图 4b）。当被问及“对于 35 分钟的行程，当比率为 2 时，您还能乘公共交通工具来此行吗？”小汽车驾驶员说“是，但很少这样做”。要让更多的小汽车驾驶员乘坐公共交通，需要进一步缩短公共交通的行驶时间。但是，公共交通的时间弹性大于小汽车，这意味着要增加公共交通的乘车率，出行时间的边际减少要比小汽车小。在鼓励从小汽车到公共交通的模式转换时，仅出行时间并不能说明全部问题，而公共交通像小汽车一样快并不意味着每个人都愿意乘坐。许多其他因素也会影响模式选择，例如舒适度和价格。

总结

为了评估小汽车与公共交通方式出行时间的差异，我们提出了一种计算框架，该框架结合了新型大规模数据集，尤其是用社交媒体数据来表示平均日全天的实际出行需求位置。我们在四个城市中实施了该框架，并对它们进行了系统的比较。该框架通过以高时空粒度揭示了公共交通和小汽车之间的出行时间差异，从而实现详细和局部级别的探索，从而证明了其有用性。平均而言，公共交通出行时间约为小汽车出行时间的两倍，这证实了先前的研究，但具有更详细的实际出行需求数据。在市中心或者公交路线周边差异会小很多。在小于 3km 的短途出行中和在 Stockholm 和 Amsterdam 的高峰时段，公共交通出行要比小汽车出行费时更少。两种出行模式之间的系统性比较指出，各个城市出行时间差异出奇的相似：出行距离小于 3km 时 $R < 1$ ，然后随着距离变长快速增加并稳定到 2。

为了在将来的研究中更好的将小汽车出行与公共交通出行进行比较，可以在地理环境分析中加入更多功能，考虑拥堵对公共交通的影响，通过全天时间，地点和出行距离去不同纬度去讨论环境变化，尤其是温室气体的变化。

方法

图 6 展示了我们的方法框架。我们在研究中结合了四种大数据来源，包括：HERE 交通数据、推特地理签到数据、OpenStreetMap (OSM) 路网数据和 General Transit Feed Specification (GTFS, 公共交通通用数据标准) 数据。每个城市的研究区域按 500 米的对角线长度划分成了六边形网格。所有网格均由其中心的经纬度坐标表示。带有地理标签的推

文数据用于产生每个城市地点的访问频次。一些研究发现，带地理签到的推文数据与交通数据或出行需求数据有很好的相关性。除此之外，带有地理签到的推文还提供了实际活动（即需求）的位置信息，从而可以避免对城市中所有起终点对的出行时间进行估算。我们将经常被访问的位置，也就是那些每个小时内有超过 20 条推文的位置，分类为“目的地”，并计算从城市其他地方到“目的地”的小汽车出行时间，具体方法参考出行时长计算部分。

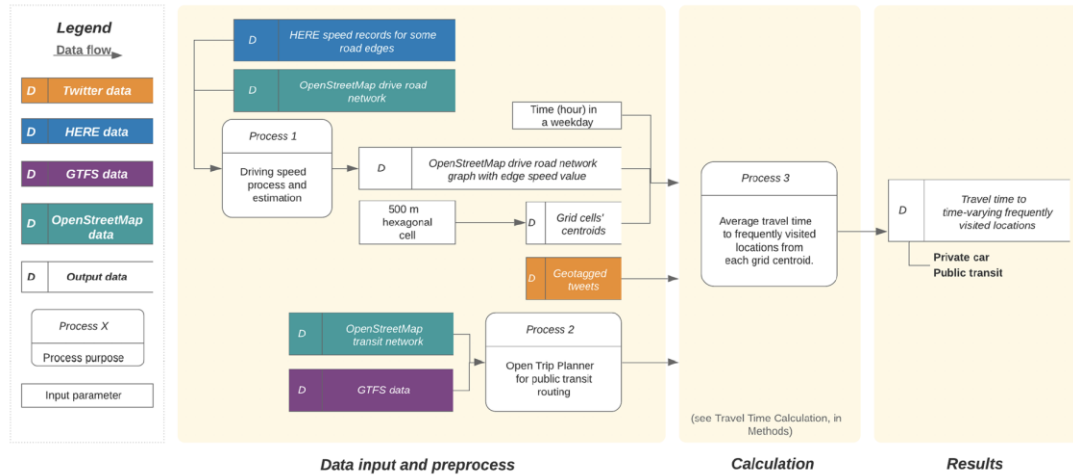


图 6、计算小汽车与公共交通出行时长的分析框架流程图

数据与数据预处理

HERE 交通数据：行驶速度。从 2018 年 1 月 1 日到 2018 年 12 月 31 日，行驶速度的采样频次是每五分钟一次。在本研究中，我们关注的是典型工作日场景。因此，我们剔除了 1 月、2 月、7 月、8 月和 12 月收集的数据（寒暑假），以消除假期的影响。由于不同城市采样时间与网络延迟的差异性，用于建模的数据是按 15 分钟的时间窗格进行采样（也就是每天 96 个时间窗格），每个时间窗格对应的速度是平均速度。道路在地理上由一系列边构成的路段进行表征。我们计算工作日每个路段的小时平均车速。

GTFS 数据：公共交通时间表。GTFS 静态数据集是一组文本文件，包含了所有用来重构公共交通机构时间表所需要的全部信息，包括停靠点的位置以及所有路线和车辆出行的时间。GTFS 数据是从可公开获得的各种来源收集的。巴西圣保罗的数据收集自 SPTrans 提供的 OpenMobilityData，瑞典斯德哥尔摩的数据收集自 Stockholm SL 提供的

OpenMobilityData, 澳大利亚悉尼的数据由新南威尔士交通部提供, 荷兰阿姆斯特丹的数据由开放数据门户网站 9292 提供。

OSM 数据: 用于驾驶的道路网。要计算汽车行驶时间, 需要通过 OSMnx (一个 Python 的开源包) 下载每个研究区域对应的小汽车交通道路网络。然后将道路网络转换成有向图, 然后进一步转换成 igraph 中的对象。道路网络数据同样可以用于计算往返于公共交通站点的步行距离。为了计算公共交通出行时间, 可以从 Geofabrik 下载国家层面的 OSM 数据文件, 进一步提取每个城市的 OSM 数据。

推特数据: 旅行需求随时间变化。为了确定经常发带地理签到推特的非商业地理标签用户, 我们使用了从推特购买的 Gnip 数据库, 时间跨度从 2015 年 12 月 20 日到 2016 年 6 月 20 日。Gnip 是推特的子公司, 大量销售历史推文, 并提供对 Firehose API (这组数据涵盖了所有的推文和相关信息) 的访问权限。如果每年发布至少 50 条带地理签到的推文, 则将这些用户定义为顶级地理签到用户。我们使用用户的时间轴 API 提取了研究区域内这些顶级地理签到用户的历史推文。对获得的带地理签到的推文进行进一步的处理, 以代表城市中经常访问的区域。关于地理签到的推文预处理方法以及其在总体水平需求表征方面有效性更详细的说明, 可以参考我们先前的研究。对于工作日每个小时间隔, 被访问最频繁的地点以每个网格单元的中心来表征, 需要在该小时内至少捕获了 20 条带地理签到的推文。每个城市中带地理签到的推文数量和用户总数在下面的数据描述中进行了汇总。

数据描述。表格 2 里是四个城市的总体统计信息。巴西圣保罗缺少都市区域的数据, 研究区域是自治市的边界。对于其他城市, 研究区域是功能性城市区域。

表 3 是有关四个城市数据集的基本信息。图 7 是带地理签到的推特数据的研究区域和空间分布。此外, 分析中还包括了人口密度数据。人口密度数据用于不同模式出行时间空间分布的可视化对比, 还用于城市级出行时长比例 R 在不同网格单元上权重的初始化。

表 3、本文用到数据集的基本信息

城市	网格数量	HERE (%)	GTFS (日期)	推特		
				用户数	地理标签数	时间范围
São Paulo	7423	55	2019-5-10	21522	2511710	2010-09—2019-06
Stockholm	2543	48	2019-5-15	5344	310609	2010-09—2019-03
Sydney	10564	51	2019-5-14	11707	608699	2010-09—2019-05
Amsterdam	4638	73	2019-5-28	9424	284281	2010-09—2019-04

图 7、研究城市：上排是研究的城市区域，下排是工作日地理签到推文的 的空间分布

出行时长计算。 起点、终点和出行时长。我们计算从所有网格单元的地理中心到城市中带有地理签到的推文频繁访问的目的地间的最快路线的旅行时间。目标单元根据每天时间变化，每小时汇总一次。在某一时间间隔 t 内（如从早上八点到九点），一个给点的起点单元平均出行时长，可以定义为按照发推频次加权的一到多（频繁访问目的地）的平均出行时长。

$$T_{mode}(i, t) = \frac{\sum_{j=1}^{n(t)} [f(j, t) \cdot T_{mode}(i, j, t)]}{\sum_{j=1}^{n(t)} f(j, t)}$$

这里， $T(i, j, t)$ 代表从起点单元到终点单元到最短出行时长。

典型工作日日平均出行时长定义如下：

$$T_{mode}(i) = \frac{\sum_{t=0}^{24} \sum_{j=1}^{n(t)} [f(j, t) \cdot T_{mode}(i, j, t)]}{\sum_{t=0}^{24} \sum_{j=1}^{n(t)} f(j, t)}$$

在城市层面，平均出行时长计算公式如下（人口、空间、时间加权）：

$$T_{mode} = \frac{\sum_{t=0}^{23} \sum_{i=1}^{N(t)} \sum_{j=1}^{n(t)} [f(j, t) \cdot T_{mode}(i, j, t) \cdot Pop(i)]}{\sum_{t=0}^{23} \sum_{i=1}^{N(t)} \sum_{j=1}^{n(t)} f(j, t) \cdot Pop(i)}$$

这里 $N(t)$ 指的是所有与目的地联通的六边形单元的个数， $Pop(i)$ 指的是起点单元的人口密度。

接下来两个部分将会详细阐述小汽车出行时长和公共交通出行时长的计算方法。

小汽车出行时长计算。我们通过 HERE Traffic 的实时道路速度记录和 OSM 下载的小汽车道路网络中的速度限制来计算小汽车的出行时长。OSM 数据中并非每条道路都被 HERE Traffic 数据覆盖。对于没有覆盖的道路，如果可能的话，根据 HERE Traffic 中覆盖的相同道路类型的数据来估算平均行驶速度。对于那些没有实时记录或对应相同类型的道路，速度将根据 OSM 中的限速值进行估算，其他道路将统一赋以 30 公里每小时的值，这只占到所有道路的 0.05%。

处理完路段速度后，对小汽车道路网络构建有向图，基于平均速度和路段长度的出行时长将被赋值给有向图的边。我们从 0:00-23:00 按小时计算每个典型工作日道路边的行驶时间。起点和终点是距离各个单元中心最近的道路节点，选择路线的阻抗是道路边的行驶时间。从给定始发节点到给定目的地节点到出行时间由 Dijkstra 算法的最短路径的出行时长加上一个 5 到 10 分钟的随机停车时间表示。

公共交通出行时长。公共交通出行时长是通过典型工作日的 GTFS 数据和 OSM 数据计算的。由于研究城市之间数据可用性的差异，我们选择了 GTFS 更新日期后的五月的第一个星期三，如表 3 所示。在给定起点-终点对之间的路径规划是使用开源多模式路径规划引擎 Open Trip Planner (OTP) 进行的。公共交通出行可能包括所有可能的公共交通模式，公交车、电车、火车、地铁等以及步行（这里步行速度取 1.4m/s）。对于每个起讫点对，OTP 会把给定的出发时间和可用的出行方式相结合，找到最快的门到门出行方案。最大的步行距离设置为 800 米，为了更好地平衡误差和计算时间，我们使用采样率为 15 分钟的混合采样方法，这有利于避免可修改的时间单元问题（MTUP，译者注：MTUP 问题是指不同的时间采样频率对结果的影响，可以参考

https://www.researchgate.net/publication/216802584_Modifiable_Temporal_Unit_Problem）

。从随机选择的出发时间开始，将出发时间设置为每 15 分钟一次，直到典型工作日被 96 个均匀间隔的时间采样。如果两个位置之间的最快路径不满足最大步行距离 800 米的限制或者行程时间超过 240 分钟，则将网格单元标记为“公共交通无法到达”。

小汽车与公共交通的相对使用时间。我们提出了一个以网格单元起点到目的地的出行时长比，以量化小汽车出行与公共交通出行两种模式之间的差异。如下所示：

$$R(i, j, t) = T_{PT}(i, j, t) / T_{car}(i, j, t)$$

比值 $R(i, j, t)$ 代表的是同一时刻的起讫点对间公共交通出行时长与小汽车出行时长的比值， R 值越高，代表公共交通相比于小汽车在出行时长方面没有吸引力。

值得注意的是，我们是将实证数据估算的小汽车出行时长与按计划执行的公共交通出行时长进行比较。尽管 GTFS 数据通常会考虑到常发性具体拥堵，但先前的研究表明，计划中的公共交通服务可能会受到交通事故和天气都偶然因素的影响，从而导致实际情况与时间表不符。但是，我们找不到本文分析的公共交通服务的 GPS 数据来补偿实际与计划的偏差。使用已经包含常发性拥堵的 GTFS 数据，并采用我们提出的出发时间混合采样方法，可以最大程度的减少最终的估计偏差。如果考虑这些误差的话，我们估计的汽车和公共交通之间的出行时间差会偏低。

当计算不同时间或网格单元的出行时长比 R 值的平均值时，我们采用公式(1)-(3)中类似的加权方法。