劳动者的共同在场与企业间网络的演变有关

Ketian Chen[1,2], Hiroyasu Inoue[3,4], YoshihideSekimoto[1,2],Yoshiki Ogawa[2,5],和 Takahiro Yabe[6,7,*]

[1]东京大学工学研究生院，日本东京[2]东京大学空间信息科学中心，日本东京[3]兵库县立大学信息科学研究生院，日本神户[4]理化学研究所计算科学中心，日本神户[5]一桥大学社会数据科学研究生院，日本东京[6]纽约大学坦登工学院城市科学与发展中心，美国纽约[7]纽约大学坦登工学院技术管理与创新系，美国纽约[*]takahiroyabe@nyu.edu

arXiv:2611.29191v1 [physics.soc-ph] 22 Nov 2026

摘要

企业间交易网络是现代经济的基础，对经济增长和韧性至关重要。以往研究大多通过空间邻近性解释企业间关系的形成。然而，人们经常在不同地点之间流动，从而创造超越地理邻近性的接触机会。这些机会与企业间交易关系有何关联，仍未得到充分理解。在此，我们结合大规模手机移动数据与东京196,317家企业多年期交易面板数据，量化劳动者之间的接触机会，并考察其与企业间交易关系的关联。我们发现，彼此之间拥有更多偶遇机会的区域配对，更有可能存在交易联系，也更有可能促成此前彼此未建立联系的企业形成新的链接。这种关联在信息密集型和服务密集型行业中尤为明显，反映出接触机会对于涉及无形知识和服务的交换十分重要。通过相遇网络中的共同邻居形成的间接联系，也与跨行业后续交易形成呈正相关，但在行业内部并非如此。这些发现表明，塑造人类流动所形成连通性的城市政策，可能对经济网络的发展和韧性发挥重要作用。

Main

企业间交易网络是现代经济组织的基础。商品或服务的生产依赖于由上游供应商和下游买方组成的链条，而这些链条的结构塑造着资源
在经济中的配置和冲击传播[1–3]。对单个企业而言，贸易关系会影响企业绩效，以及其暴露于通过生产网络传递的冲击的程度[4–7]。因此，理解这些企业间关系如何形成、持续和消解，对于解释实体经济的组织与动态至关重要。

从理论角度看，内生生产网络模型将合作伙伴选择视为一个优化问题：企业通过权衡生产成本、技术和风险来选择交易伙伴[8–11]。然而在实践中，由于存在摩擦，企业无法毫无成本地识别并与合适的伙伴建立关系。有关生产网络和国际贸易的研究强调了合作伙伴选择过程中涉及的摩擦，包括搜索成本[12,13]、、信任建立[14]、、获取有关市场机会的信息[15,16]、，以及地理距离和出行成本[1]。然而，关于企业如何克服这些摩擦，以及这些过程如何将潜在匹配转化为实际经济联系，直接实证证据仍然有限。要回答这一问题，就需要将注意力从现有关系的企业层面特征，转向企业如何通过克服摩擦而建立联系的过程。

这一过程中的一个重要部分可能通过企业劳动者之间的互动实现。经济关系嵌入于人际网络之中，信息、信任和声誉正是在这些网络中发展起来的[17]。因此，建立和维持交易关系依赖于人与人之间的沟通；他们必须交流信息、评估彼此的可靠性，并协调各自的活动，以减轻摩擦。尽管数字渠道支持跨距离沟通，但当沟通涉及复杂的隐性知识[18,19]或需要紧密合作[20–22]时，面对面互动仍然尤其有价值。与这一论点一致，对商务旅行的研究表明，当面接触有助于贸易中的信息交流和关系建立[23,24]。近期研究表明

来自不同组织的个人共同地点这一现象能够促进跨组织协作[25]并推动企业之间的知识流动[26]。这些证据促使我们考察，劳动者之间的接触机会是否也与其所属企业之间交易关系的形成相关。

大规模人类移动数据的最新进展，使我们能够以前所未有的规模和分辨率观测人类移动[27]。这些数据揭示了人们在哪里工作、消费和出行[28,29]，为了解经济活动的空间组织提供了新见解。以往研究表明，出行模式能够紧密追踪包括销售额和生产[30,31]，、企业韧性[32]，、区域收入[29,33]和房地产价值[34]在内的经济指标。另一方面，流动性数据能够衡量不同社会群体之间的接触机会，以及人们在日常生活中经历的隔离[35–37]。这些方法分别将移动性与经济活动、人际接触联系起来。将两者结合起来，便可以探究劳动者的移动是否揭示了与组织间经济联系相关的接触机会。

本研究考察劳动者之间基于流动性的相遇是否与企业间贸易关系的形成相关。我们使用包含30多万名用户每日位置记录的大规模匿名手机数据，推断劳动者的工作地点，并根据其时空共现构建一个企业所在地理网格之间的相遇网络；随后将该网络与交易关系数据相连，这些数据汇总自涵盖日本东京19万多家企业的记录。会面机会增加一倍，与交易联系存在的几率提高1.4%相关；对于此前没有贸易关系的企业，会面机会增加一倍还与之后链接形成的几率提高1.9%相关。我们还考察了这些关联如何随时间、地理距离和产业结构而变化。除了成对的接触机会，我们还考察了通过相遇网络中的共同邻居形成的间接联系（即两跳联系）。这些间接联系与行业之间后续交易形成呈正相关，但在行业内部并非如此。这些发现表明，工人接触构成了一种可观测的微观层面过程，将城市移动与企业间经济联系的形成联系起来。

结果

大规模测量劳动者之间的接触机会

我们的核心假设是，代表不同企业的人之间的互动会促进企业间交易。这类互动可能帮助企业克服识别、评估潜在合作伙伴以及与其协调时所面临的摩擦。如果确实如此，那么在交易关系形成之前，就应当能够观察到可能促成此类互动的偶遇机会。我们将城市中不同企业劳动者之间的实际接触作为衡量这些机会的可观测代理指标。我们利用两个数据来源，构建了东京大规模的跨企业相遇网络测度。第一个数据集是由 Blogwatcher, Inc. 提供的匿名化智能手机 GPS 数据集。该数据集包含在日本收集的 GPS 位置记录，来自选择共享位置的用户所使用的 140 多款智能手机应用。我们将 GPS 样本限定为2019年7月至9月期间在东京被观测到、且位置记录覆盖超过90个不同日期的用户，最终得到329,149名活跃用户，约占东京劳动年龄人口（20至69岁）的3.5%。我们将原始轨迹转换为停留点，仅保留超过10分钟的停留[38]。我们进一步考察了各市镇的 GPS 采样率，发现各地区的覆盖情况较为一致，平均采样率约为2.2%（补充表 S1）。

我们的第二个数据来源是帝国数据银行提供的企业间交易数据集。帝国数据银行是一家私营企业信用调查公司，维护着日本最全面的企业信息库之一。其企业数据库覆盖约 70% 的日本企业，并记录每家企业的总部所在地、行业分类和唯一企业代码。该交易数据集按年记录每笔报告交易中买方和供应商的企业代码，使我们能够观察企业间经济联系随时间的形成与延续。因此，我们将每一项报告的买卖双方关系表示为二元连接，以指示两家企业在某一年是否存在交易关系。利用这些记录，我们构建了一个面板数据集，涵盖 196,317 家总部位于东京的企业，以及 2016 年至 2022 年间 3,293,848 条年度买卖双方关系观测值。

为将个体移动轨迹转化为跨企业相遇网络，我们首先根据手机位置记录识别通勤员工。由于东京建筑密度极高且建筑垂直延伸明显，基于建筑物层级推断工作地点极易受到GPS定位误差的影响，因此我们使用分辨率为10的H3网格单元将企业位置汇总至六边形网格（平均边长约为76米）。我们将至少包含一家注册企业总部的网格单元定义为企业单元。我们将用户在企业单元内停留至少5小时的工作日定义为工作日。若用户在其主要工作地点至少有40个工作日，且这些工作日分布于三个月观测期内至少12个不同的周，则将其归类为劳动者（方法）。改变劳动者识别所需的每日最短工作地点停留时长、工作日数量和不同周数，并未明显影响识别出的劳动者群体（补充图S4）。为评估所识别工人样本的空间代表性，我们将估计的劳动者人数与日本2020年人口普查进行比较[39]，，发现二者高度相关（$r = 0.80$；补充图S5）。

推断出的劳动者在工作日早晨到达工作场所的时间也较为稳定，这与日本规律的全职通勤模式一致（补充图S6）。

为了量化来自单元格对 (i, j) 的劳动者之间的接触机会，我们使用观测到的时空共现作为代理指标[26]。由于实际接触要求两个人在同一地点和同一时间出现，共同在场为接触提供了必要条件。我们首先将共同在场定义为 $M_ij = \sum_{c,t} n_{ict} n_{jct}$，其中 c 表示 H3网格单元，t 表示预先设定的 30分钟时间间隔，而 n_{ict} 表示来自企业单元 i 的劳动者人数：这些劳动者在时间间隔 t 内至少有 10 分钟的停留时间与单元 c 重叠。因此，M_{ij} 统计了企业单元 i 和 j 之间的工人对共现次数。由于不同企业在全天同时活动的劳动者数量存在显著差异，我们根据其时间活动情况，将观测到的共同在场次数按可能发生会面的劳动者配对最大数量进行标准化。随后，我们将潜在最大会面次数定义为 $PMM_ij = \sum_t (\sum_c n_{ict}) (\sum_c n jct)$，该指标统计数据集中在同一时间 t 活动的劳动者配对数量，而不考虑其所在地点。最后，我们定义会面概率 $MP_{ij} = M_{ij}/PMM_{ij}$，其中，同一地点和时间窗口内的共同在场构成直接接触的必要条件。所得会面概率 MP_{ij} 表示：随机抽取一名来自单元 i 的劳动者和一名来自单元 j 的劳动者后，观察到他们在同一地点、同一时间出现的概率（概念示意图见图1a）。为评估观测到的会面概率（MP）是否易受潜在测量误差影响，我们构建了三种替代会面概率测量。我们降低拥挤地点或广受访问地点中偶遇的权重，并排除仅观测到一次的偶遇。补充图 S8 中的结果表明，这些替代会面概率测量保留了观测结果的主要特征，说明观测到的会面概率的总体结构对于不同设定具有稳健性。详细定义和结果见补充说明 2.3。

为了评估观测到的会面概率是否反映了超出共享空间移动模式的时间对齐，我们构建了一个时间平移零模型，在打乱访问时间对齐的同时保留个体移动模式。具体而言，我们将每位劳动者的完整日常移动轨迹平移到另一个日历日期，并在 100 次随机化实现中重新计算会面概率。此过程保留了访问地点和日内移动模式，同时打乱了劳动者之间偶遇的原始时间安排。例如，两位劳动者可能约好在某天中午于一家咖啡馆碰面。将他们的日常轨迹平移到不同日期后，这次约定的碰面便不复存在，尽管两位劳动者仍会在中午前往同一家咖啡馆。补充图 S9 显示，观测到的会面概率分布系统性地高于零分布。在各单元格对中，观测到的会面概率与零模型MP之比的中位数为 1.61，表明观测到的会面概率通常比零模型预期高 61%。这些结果表明，除共享移动模式本身所能解释的程度外，时间对齐也有助于形成观测到的接触机会。详细结果见补充说明2.4。

接下来，我们考察基于流动性的接触机会的空间组织，并将其与企业间交易联系的空间组织进行比较。图 1b 和图 1c 分别展示了交易联系和流动联系。每个节点代表一个企业单元。交易联系连接的是包含在2020年有可观测交易关系的企业的单元，而流动联系连接的是具有正会面概率（会面概率 > 0）的单元格对。这两个网络呈现出大体相似的空间模式，密集的联系集中在多个主要商业中心周边，与东京的多中心组织结构相符。在确定流动联系的分布位置后，接下来我们考察这些联系的强度。图 1d 展示了企业单元之间会面概率的分布。该分布以 $10^{-4.5}$ 为中心，但右侧尾部明显较长，表明相连单元格对之间的接触强度存在显著差异。这种差异在节点层面也十分明显。企业单元的度（定义为通过正会面概率与之相连的其他单元数量）和节点强度（定义为这些联系中会面概率之和）存在显著差异（补充图S7）。总而言之，这些分布表明，接触机会远非均匀，在强度和连通性方面均相差数个数量级。最后，我们考察流动联系强度如何随两个企业单元之间的地理距离而变化。对于每一对单元 i 和 j，我们计算哈弗辛距离 d_{ij}，并估计图 会面概率$_{ij}$ 与距离之间的关系，结果见图 1e。会面概率随单元间距离增加而下降（皮尔逊相关系数 $r = -0.156$，$p < 0.001$）。估计得到的双对数回归系数为 -0.282，这表明距离加倍时，会面概率约下降 18%。

劳动者之间的实际接触与经济联系相关

我们考察工人接触是否与不同的企业间交易结果相关。具体而言，我们将根据2019年流动数据测得的会面概率，与不同时间滞后和超前期观察到的交易结果联系起来。因变量是一个二元指标，用于表示给定年份两个企业单元之间是否存在交易关系。我们的主要自变量是会面概率（MP）的对数。我们依据方法中的方程 1估计逻辑回归模型，并控制相关协变量和企业单元固定效应。

我们使用三种二元交易结果来区分企业间交易关系的不同阶段（图2a）。第一个结果变量反映连边存在，表示某个单元对在某一年是否至少存在一组买卖双方关系

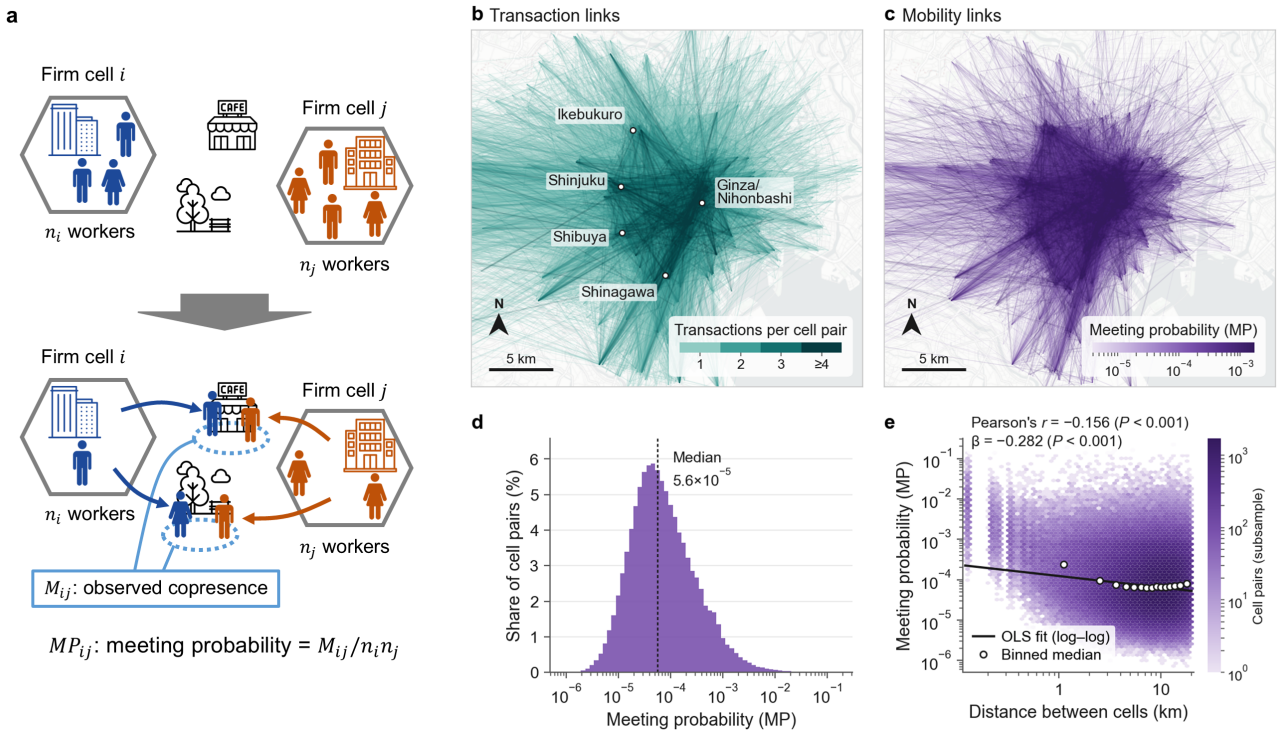


图1. 大规模测量劳动者之间的偶遇。a. 会面概率测量方法的概念示意图。**b.**东京企业单元之间的交易联系。线条连接2020年存在观测到的企业间交易关系的单元。**c.**东京企业位置之间的流动联系。线条连接会面概率为正的单元配对（会面概率 >0）。**d.**会面概率（会面概率）的分布。x轴采用对数刻度。**e.** 会面概率与单元间地理距离的散点图。会面概率随距离增加而降低，二者呈负相关（皮尔逊$r=-0.156$，$p<0.001$）。

。链接形成是指数据所涵盖的前一年均不存在交易联系的一对企业首次出现交易联系。例如，若某一配对在$t=1$时被归类为链接形成，则该配对在2016年至2019年间没有观察到交易，但在2020年建立了交易联系。第三个结果反映交易关系消失，定义为交易联系首次消失。例如，$t=1$时的关系解体意味着，2019年观察到的交易联系在2020年消失。

为评估观测到的会面概率是否包含超出空间距离和网格单元规模所能解释的信息，我们使用广义引力定律构建基于引力的基准：$MP^{gravity}_{ij}=An_i n_j/(d_0+d_{ij})^{\gamma}$，其中$n_i$和$n_j$分别是单元 i 和 j 中的劳动者人数；$d_{ij}$是两个单元之间的距离；$d_0$是固定的距离偏移量；$\gamma$是指数参数。我们使用泊松伪最大似然（PPML）拟合引力模型，在$d_0=0.1$ km时得到$\gamma=1.354$（详见补充说明2.5）。基于引力的基准保留了观测到的会面概率所采用的工人数和归一化方式，但以引力模型预测值替代观测到的共同在场计数。随后，我们分别使用观测到的会面概率和引力 MP 估计交易模型。通过这一比较，可以评估基于空间邻近性和网格单元规模的基准，是否能够重现实际工人接触所呈现的关联（图2b）。

对于连边存在，所有观测年份中，观测到的会面概率均与交易联系的存在呈显著正相关，系数稳定在约0.020（$p<0.001$）。这意味着，会面概率翻倍与企业间交易几率增加约 1.4% 相关。以引力 MP 替代观测到的会面概率后，2017–2022 年的系数均为负，范围从 -0.137 到 -0.083。因此，拟合的基于引力的基准无法再现偶遇与交易连通性之间的正向关联。使用观测到的会面概率的模型，其赤池信息准则值始终低于使用引力 MP 的模型，差异范围为 95 至 118。这种模型拟合度的提升表明，观测到的会面概率捕捉到了企业间连通性的信息，而这些信息并未被基于引力的基准充分体现。

在链接形成方面，所有六年中，观测到的会面概率均与新交易关系的形成呈显著正相关（$p<0.05$），且效应在流动性测量后紧接着的2020年最大（$\beta=0.027$,$p<0.001$）。这意味着，会面概率翻倍与链接形成几率增加约1.9%相关

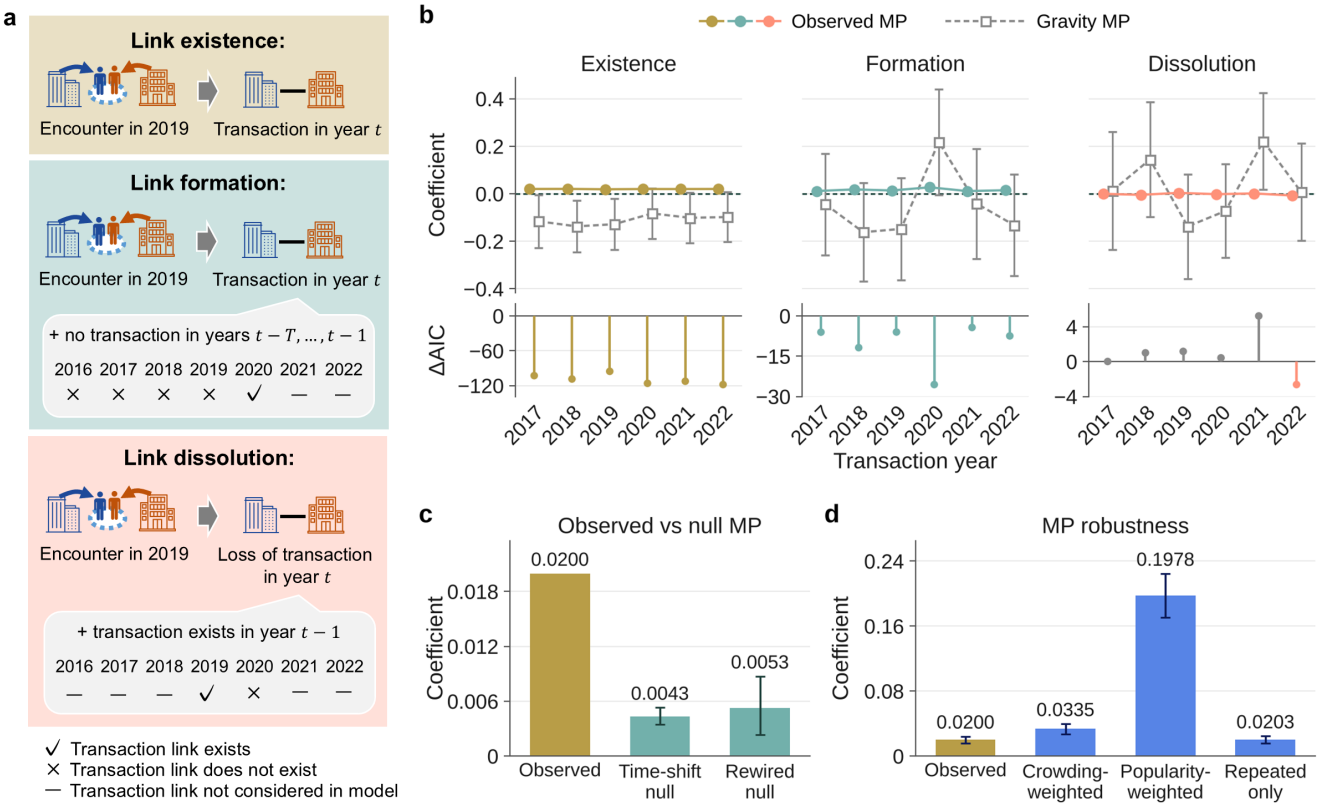


图 2. 会面对企业间交易关系的影响。a. 交易结果的概念定义。**b.**使用观测到的会面概率和引力 MP，考察会面概率与企业间交易关系之间的关联。上方面板展示了逻辑回归中跨年份的连边存在、链接形成和关系解体回归系数。点表示系数估计值，误差条表示95%置信区间。下方面板展示赤池信息准则的差值（$\Delta AIC = AIC_{observed} - AIC^{gravity}$），负值表示观测 MP 模型更优。**c.** 比较2020年连边存在的观测 MP 与两个零模型（时间平移和重新连边）的回归系数。柱形表示观测到的系数，以及每个零模型进行100次随机化所得的平均系数。误差条表示零模型的第2.5至第97.5百分位范围。**d.** 考察2020年 MP 与连边存在之间关联的稳健性，比较四种 MP 度量的回归系数：观测到的 MP、拥挤度加权 MP、受欢迎度加权 MP 和仅重复相遇 MP。所有回归均包含相同的控制变量。误差条表示基于双向聚类标准误的95%置信区间。

，且这一形成发生在下一年。对于要求形成前一至四年内均无交易联系的其他定义，这种正向关联依然存在（补充说明 3.6）。相比之下，引力 MP 系数的符号不一，且每一年在统计上均不显著。采用观测到的会面概率的模型始终具有较低的赤池信息准则值。

对于交易关系消失，观测到的会面概率的系数在整个观测期内均较小、符号不一，且在统计上不显著（$p > 0.05$）。引力 MP 同样未显示出始终显著的关联。两种模型的赤池信息准则差异相对较小，也未显示出任何一致的优势。总体而言，这些结果表明，会面概率与新交易联系的形成之间的关联，比其与现有交易联系消失之间的关联更具系统性。对于链接形成和关系解体，采用从一年到四年不等的替代交易历史窗口时，结果仍然一致（补充图 S12）。

前述分析确立了观测到的会面概率与交易联系的存在和形成之间一致的正向关联。接下来，我们将考察这些关联是否受到潜在空间结构和网络结构的影响，并检验其稳健性。首先，我们评估该关联对会面概率度量方式是否敏感。在三种替代测量指标下，对拥挤或经常有人到访地点中的偶遇进行降权，并排除一次性相遇后，估计的会面概率系数仍为正且显著（图 2c）。这些替代测量指标的详细定义和构建过程见补充说明 2.3 和补充表 S6。

其次，为了评估观测到的关联是否反映了时间上同步的偶遇，而非劳动者只是

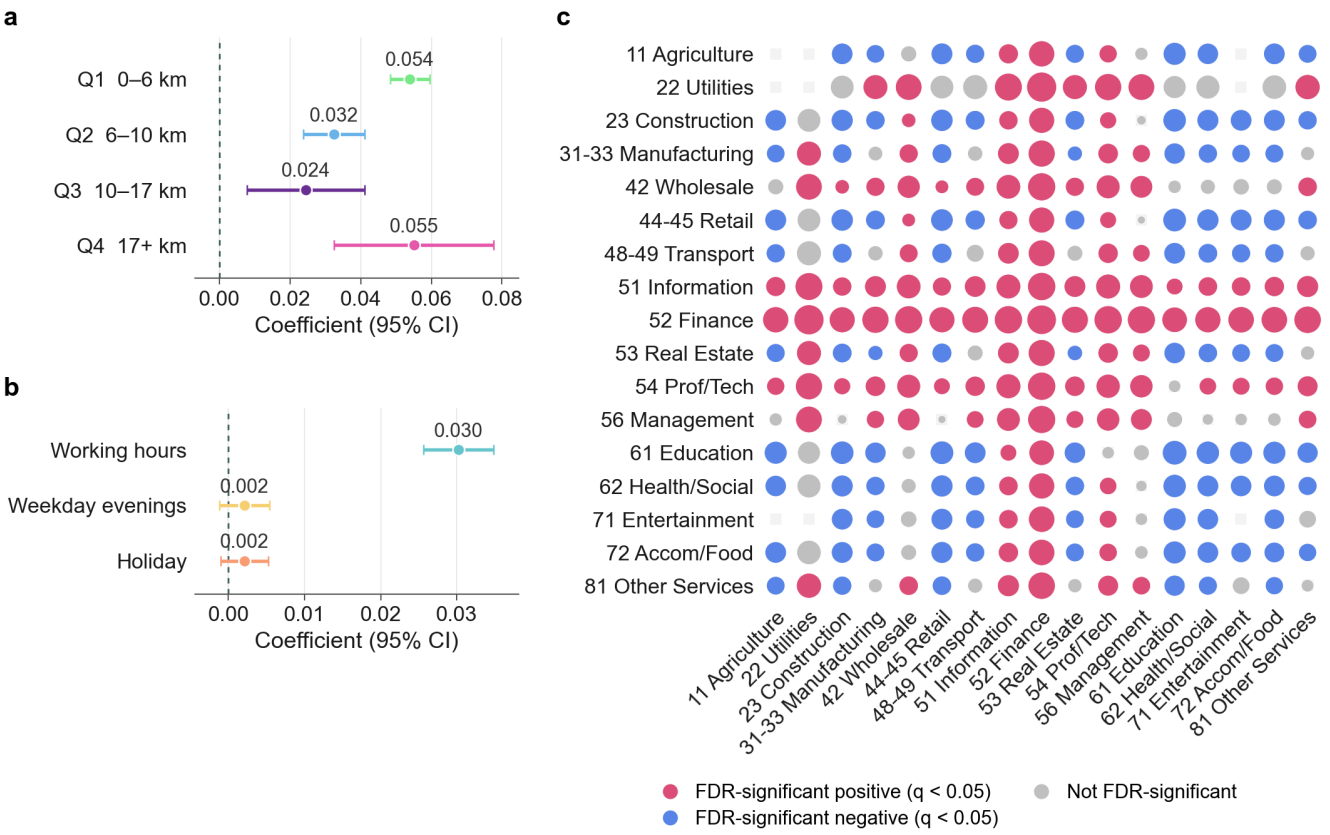


图 3. 异质性分析。a. 基准模型中，单元间距离各四分位组的会面概率回归系数（$t = 1$）。所有距离组的关联均为正，且最近和最远四分位组的系数较大。**b.** 基准模型中工作时间、工作日晚上和节假日期间会面概率的回归系数（$t = 1$）。工作时间的会面概率与连边存在呈现出最强的正关联。误差条表示 95%置信区间。**c.** 点阵图展示了各行业配对的会面概率估计系数。点的大小与系数绝对值成正比。红点和蓝点分别表示经 FDR校正（$q < 0.05$）后仍显著的正系数和负系数；灰点表示估计值不显著。

前往相似地点，我们用由时间平移模型生成的零模型MP替代观测到的会面概率（补充说明2.4和3.4）。零模型保留每位劳动者观测到的轨迹，但随机平移其访问日期，从而保留劳动者前往的地点，同时打乱实际的时间对齐。图2d 中的左侧和中间柱形显示，用零模型MP替代观测到的会面概率后，估计系数降低了约79%。这一降幅表明，仅凭共享访问模式无法重现时间对齐的偶遇所呈现的关联。此外，还有一个问题需要考虑：观测到的关联可能仅仅反映了企业单元整体连通性的差异。连通性较高的单元可能既有更多交易联系，也与其他单元有更高的会面概率。因此，我们在保留每个单元关联的联系数量的同时，对交易伙伴进行随机化（补充说明3.5）。在这些度保持重连网络中，与同一模型下观测到的会面概率相比，会面概率估计值下降了74%（图2c 右侧的柱形）。这表明，会面概率不仅与单元的连通程度有关，也与企业和谁进行交易有关。

时间、距离和行业异质性

地理距离可能会影响接触机会与企业间交易关系之间的关联。如果会面概率仅仅反映空间邻近性，那么随着企业间距离增加，它与交易关系的关联应当减弱。因此，我们将单元格对划分为距离四分位组，并纳入会面概率与距离四分位组指标之间的交互项。沃尔德检验拒绝了四个距离四分位组的系数相等这一零假设（$\chi^2(3) = 26.99$，$p < 0.001$），这表明会面概率与交易关系之间的关联存在显著异质性。图 3a 显示，会面概率的效应在所有距离四分位组中均为正。最近单元对的估计系数相对较大（$\beta = 0.054$，$p < 0.001$），在中等距离时变小

距离（$\beta = 0.032$ 和 0.024，$p < 0.001$），而在距离最远的配对中再次增大（$\beta = 0.055$，$p < 0.001$）。这表明实际接触对于地理上相距较远的企业可能尤其重要，或许有助于它们克服空间摩擦，建立交易关系。完整的回归结果见补充表 S9。补充表 S10 和 S11 显示，在采用其他距离分组方式（五分位数和十分位数）以及剔除距离最远的 1% 或 5% 单元格对后，结果仍然一致（更多详情见补充说明 3.7）。

为考察这种关联是否因偶遇发生的时间而异，我们将会面概率划分为三个特定时段的指标：工作时间（周一至周五，上午9:00至下午6:00）、工作日傍晚（周一至周五，下午6:00至晚上10:00）和节假日（全天）。这种区分使我们能够评估偶遇效应是否集中在与常规经济活动联系更紧密的时段。工作时间内的偶遇更可能反映与工作相关的活动，因此可能为经济互动提供更相关的机会；相比之下，工作日傍晚和周末的偶遇则可能涵盖范围更广的非工作活动。我们在基准逻辑回归模型中同时纳入了这三个特定时段的会面概率指标，结果见图3b。工作时间内的会面概率与连边存在呈正向关联（$\beta = 0.030$，$p < 0.001$）。相比之下，工作日夜间（$\beta = 0.002$，$p = 0.206$）和周末（$\beta = 0.002$，$p = 0.179$）的系数则小得多。这些结果表明，这种关联集中在常规工作时间内发生的偶遇上，而非发生在通常工作时段之外的共同在场。

接下来，我们考察行业间异质性。各行业的知识基础、技术以及组织间互动模式各不相同[40,41]。当信息难以编码或需要快速反馈时，面对面交流尤为重要[19]。这些差异表明，接触机会与企业间交易关系之间的关联可能因行业而异。我们首先根据每个单元中企业数量最多的 NAICS 两位数行业部门，将其主导行业分配给该单元。随后，我们将 $\ln(\text{会面概率}_{ij})$ 与由两个单元的主导行业定义的行业配对指标进行交互。由此得到的系数反映了 2020 年会面概率与连边存在之间因行业配对而异的关联。图 3c 显示，接触机会与交易联系之间的关联在方向和幅度上均存在异质性。涉及信息业或金融与保险业的单元对，在全部 17 个行业类别中均呈现经错误发现率校正后显著的正系数。专业、科学与技术服务也呈现出类似的广泛正向关联模式。相比之下，涉及建筑业、零售业、运输业、教育业、医疗保健业以及住宿与餐饮服务的配对则普遍出现负系数。完整的回归表见补充表 S12。因此，这种模式并非简单地将服务业与非服务业区分开来；正向关联也跨越了行业边界，并不限于行业内配对。作为替代方案，我们根据各行业的员工数量而非企业数量来定义每个单元的主导行业。结果发现，在这两种定义下，83.8% 的行业对系数符号保持一致（补充图 S13）。

间接接触效应对经济联系形成的影响

前面的分析考察了直接接触与交易形成之间的关联。不过，企业也可能通过涉及共同第三方[42–44]的间接联系建立新的关系。基于这一视角，我们考察在控制直接暴露后，基于流动性的相遇网络中，通过共享邻近单元产生的间接暴露是否与交易形成相关。

我们使用工人接触网络，考察当两个企业单元的劳动者与第三个企业单元存在相遇连接时，这两个企业单元是否更有可能形成交易关系。考虑两个企业单元 a 和 c；当二者都与同一个中间单元 b 存在接触机会时，它们便形成间接联系$_k$。形式上，当 $\mathrm{MP}_{ab_k} > 0$ 和 $\mathrm{MP}_{b_k c} > 0$时，存在一条两跳路径 $a \to b_k \to c$。由此，中间单元通过两段连续接触关系连接 a 和 c（概念示意图见图4a）。我们将样本限制为在流动性测量期之前未观察到交易关系的单元对。这一限制使我们能够考察相遇网络中的间接暴露是否与后续链接形成相关。我们构建一项两跳暴露指标，记为 TwoHop_{ac}，定义为 $\sum_k \mathrm{MP}^{ab_k} \times \mathrm{MP}_{b_k c}$。该指标汇总了通过共享接触邻居 b_k连接企业单元 a 和 c 的所有两跳路径。如果某个共享邻居与 a 和 c 的会面概率都较高，那么它对该指标的贡献就更大。我们分别估计流动性测量期后一年、两年和三年内交易联系形成的逻辑回归模型。因变量 $Y_{ac,t}$在第 t 年观察到企业单元 a 和 c 之间形成新的交易关系时取值为 1，否则取值为 0；其中 $t \in 1,2,3$ 表示流动性测量期后的年数。系数 θ 反映了两跳接触暴露与后续交易联系形成之间的关联，同时控制 a 和 c 之间的直接会面概率、控制变量 X_{ac}以及单元固定效应 δ_a 和δ_c。

图4b 展示了间接相遇对后续交易联系形成的估计影响。这种关联主要集中于连接不同行业企业的单元对。对于这些跨行业单元对，较高的间接

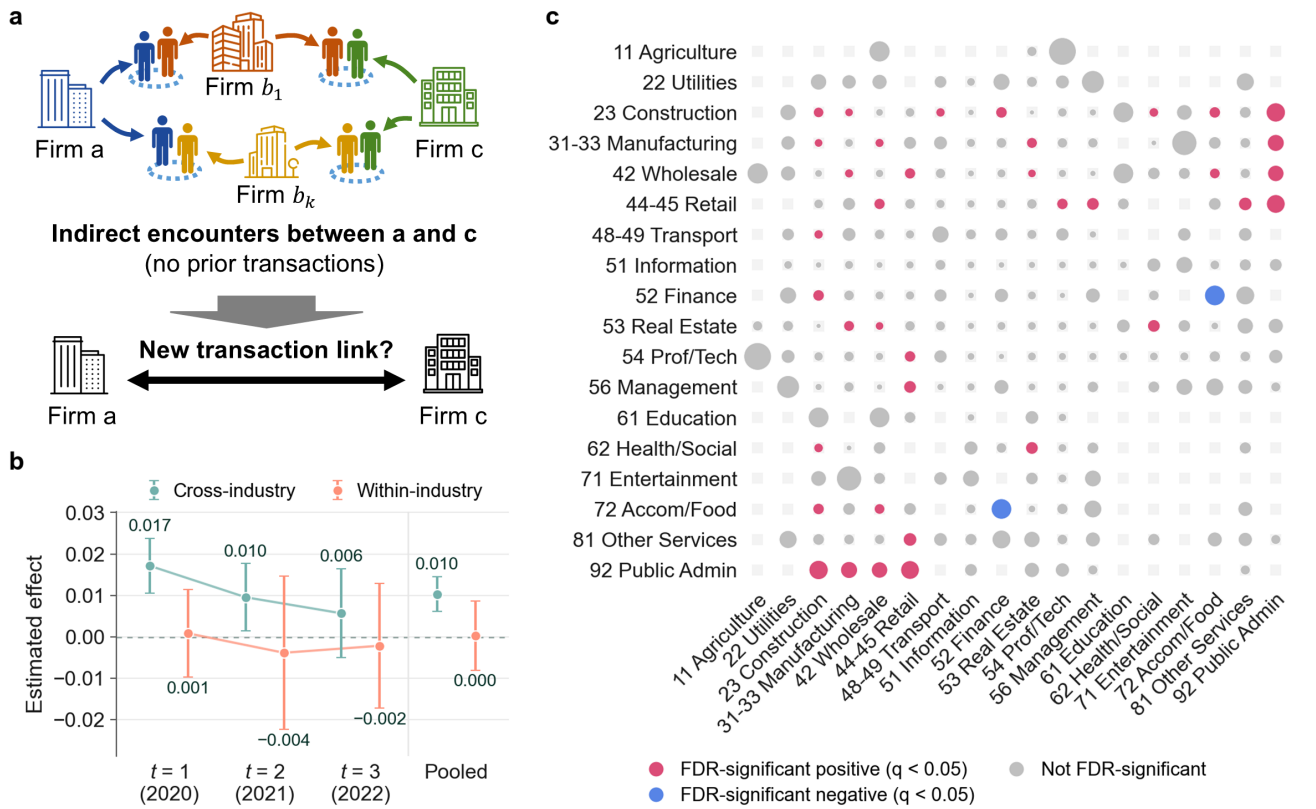


图4. 间接接触效应对经济联系形成的影响。a. 间接相遇暴露示意图。企业单元 A 和 C 未直接相连，但共享一个流动性邻居 B。间接相遇暴露通过汇总连接这两个企业单元的所有两跳路径上的相遇概率乘积计算得出。**b.** 逻辑回归模型中，行业内和跨行业单元对的间接暴露估计系数。间接相遇暴露仅与跨行业单元对后续的链接形成呈正相关，而行业内配对未观察到显著关联。**c.** 展示行业配对间接接触效应估计系数的点阵图。每个点代表一个行业配对的估计系数。点的大小与系数大小成正比，颜色表示估计值的正负及统计显著性。

相遇暴露与观测后第一年形成后续联系的可能性较高相关（$\beta = 0.017$，$p < 0.001$），第二年亦然（$\beta = 0.010$，$p < 0.05$）。估计的关联随时间推移而减弱，到第三年时已不再具有统计显著性。相比之下，间接相遇暴露与连接同一行业企业的配对之联系形成并无显著关联。所有设定下，相应系数均较小且在统计上不显著。合并估计值呈现出相似模式（完整结果见补充表S13）。这些结果在控制网络度，并采用允许共享任一单元的二元组之间存在相关性的二元聚类稳健标准误后，依然稳健。完整的稳健性结果见补充说明3.10。我们的研究结果表明，间接暴露与交易联系形成之间的关联仅存在于跨行业配对中。

图4c 考察了间接相遇暴露与后续交易联系形成之间的关联如何因行业配对而异。每个点代表特定行业配对的估计系数。显著的正向关联主要集中在涉及公共管理、建筑业、制造业、批发贸易和零售贸易的配对中。这些行业与生产、分销和制度协调密切相关。这一模式表明，在以协调密集型或制度中介型经济关系为特征的行业中，间接相遇连接可能尤为重要。

讨论

通过将东京的工人流动数据与年度企业间交易记录相结合，我们发现，工作场所单元之间的相遇机会与交易联系的存在以及这些联系随后首次出现在记录中均呈正相关。观测到的共同在场提供了超出基于地理距离和工人数的基准模型的信息，这表明劳动者在何处、何时停留，有助于刻画与贸易相关的邻近性。交易联系存在与否的关联在不同距离组中均为正向，在工作时间内最为明显，并且因行业配对而有显著差异。通过共享的邻近网格产生的间接暴露，也与跨行业配对中后续交易联系的形成呈正相关。相比之下，我们没有发现其与交易关系消失之间存在一致的关联。这些结果将日常劳动者流动的组织方式与企业间经济关系的形成联系起来。

这些发现补充了将生产网络视为高成本合作伙伴搜寻与匹配结果的研究。[1]表明，高速铁路扩张后客运成本的下降有助于建立新的买方—供应商联系，而[45] 发现，在其他供应商的地理分布更为集中的地方，企业与供应商重新匹配的速度更快。实验性证据还表明，企业管理者定期会面有助于引荐并促成直接合作关系[46]。在这些情境中，能否接触潜在合作伙伴取决于发现并评估他们的机会。本研究则补充了有关日常城市活动的证据。即使考虑了距离和地方特征，工作场所人群之间的共同在场仍与哪些地点随后建立新的交易联系相关。这为基于交通基础设施、市场厚度和有意组织的商务会议的解释，提供了行为层面的补充。

这一视角也有助于将生产网络的形成与城市集聚理论联系起来。共享、匹配和学习分别为经济活动在空间上的集聚提供了不同的解释[47], ，而有关产业共同集聚的证据则表明，投入产出联系、劳动力共享和知识流动都发挥着作用[48]。即使许多企业集中在同一地区，各企业获得接触机会的情况也可能不同。我们的距离、引力和时间移位比较表明，工作场所的位置、劳动者到访的地点以及到访时间，都包含有关经济连通性的不同信息。不同距离组之间的正向关联表明，共同在场与交易关系之间的联系并不局限于周边地区。对于集聚研究而言，这意味着研究潜在合作伙伴如何变得触手可及时，应将企业的集聚情况与连接这些企业的活动模式结合起来考察。

行业层面的结果表明，这些联系的重要性取决于所涉及的活动。在图 3c 中，会面概率与连边存在呈正相关，适用于涉及信息行业或金融与保险行业的所有行业配对，也适用于涉及专业、科学与技术服务业的许多配对。其他一些行业配对则呈现负相关。这种异质性与以下观点一致：当信息复杂、变化迅速或难以编码时，面对面交流具有重要价值[19]。关于知识溢出的移动性研究同样将劳动者交互的机会与专利引用联系起来[26]。然而，制造业中也存在以沟通为密集特征的交流，其中商务旅行与差异化商品和研发密集型商品贸易之间的关联尤其强[23]。更广泛的含义是，信息与协调需求可能有助于解释行业配对之间的差异。

区分链接形成与关系解体同样十分重要。在一段关系尚未建立之前，接触陌生合作伙伴的机会可能尤为重要；而已经建立关系的合作伙伴则可以利用积累的信息和现有的组织安排。科学合作领域的证据也提供了类似的例证。共同地点有助于联合研究的启动，但此前处于共同地点的实验室分离后，并不会降低其开展合作的倾向[49]。在企业间知识网络中也存在类似区别：地理邻近性与关系纽带形成相关，但与关系纽带持续性没有可检测的关联[50]。共同在场与链接形成之间的正相关，以及共同在场与关系解体之间缺乏一致关联，均与启动关系和维持关系之间的这一区别相一致。

间接暴露的结果将合作伙伴形成置于更广泛的网络背景中。关于战略联盟的研究表明，共同第三方与新的组织间关系的形成相关[42], ，而贸易模型则表明，企业可以通过现有联系人发现新的合作伙伴[51]。我们的结果将这一视角拓展至基于流动性的接触。通过共同相邻工作场所单元形成的联系，与后续跨行业关系形成相关，且这一关联是在控制直接暴露后观察到的。一种可能的解释是，这些联系为企业提供了机会，使其能够了解熟悉的行业圈子以外的企业。不过，两跳测量并未追踪信息传递或引荐，而且一条路径的两段可能涉及中间单元中的不同个体。

然而，接触机会只是经济关系形成过程的一部分。在一项近期的现场实验中，[52] 发现，当引荐与开展初始交易的激励相结合时，会促成持久的供应商—客户伙伴关系，而仅提供信息的引荐则几乎没有影响。这一区分凸显了通过经验了解匹配质量的重要性。有关企业间关系的研究还记录了嵌入性的协调优势，以及过度嵌入所伴随的脆弱性[43]。这些研究表明，

更高连通性的后果取决于所交换的信息、所涉及的合作伙伴以及更广泛的网络结构。因此，理解移动性的经济意义，需要考察接触机会如何转化为特定的关系和结果。

这些解释也受到若干局限的约束。首先，本分析是观察性的。尽管在流动性测量之后记录的结果与链接形成的关联更强，但 2019 年的会面概率也与 2019 年之前记录的链接形成呈正相关。既有关系可能促使人们出行，而单元对的持久特征可能同时影响流动性和交易。单元固定效应、观测到的控制变量和随机化基准并不能排除这些可能性。因此，仅凭时间顺序不足以确立偶遇对交易形成的因果效应。重复进行流动性测量，并利用可能外生的接触机会变化，将有助于区分这些解释。

其次，共同在场衡量的是偶遇的机会，而非实际观察到的交谈或商务会面。我们无法确定任何互动的目的或内容，也无法确定参与商业决策的个人所担任的角色。将流动性数据汇总至工作场所单元，也使我们无法确定流动性测量所代表的个人是否受雇于参与已记录交易的企业。因此，结果描述的是工作场所人群之间的共同在场与单元层面的交易连通性之间的关联。此外，我们只观察了 2019 年 7 月至 9 月的流动性，且仅涵盖定期前往工作场所的用户。因此，数据可能无法充分反映季节性变化、不频繁的偶遇以及不那么规律的工作形式。更长的观测窗口和重复进行的流动性测量将有助于评估这些模式在不同季节和年份中的稳定性。后续交易结果还涵盖新冠肺炎疫情期间；在此期间，东京的流动性和接触模式发生了显著变化[53]。

最后，交易数据主要记录重要合作伙伴，并在总部层级识别关系。某条关系首次出现在记录中或从记录中消失，可能反映了报告方式或合作伙伴重要性的变化，也可能意味着关系的建立或终止。本研究还仅涵盖一个密集的大都市区，因此研究结果难以推广至城市结构或商业环境不同的城市。我们无法观测交易额、匹配质量、生产率提升或福利效应。因此，仅凭记录到的关系数量增加，并不足以证明经济状况有所改善。若能在更长时期内、跨多个城市，将接触机会与这些结果联系起来，将有助于评估本文所记录模式更广泛的经济后果。

方法

流动性数据

我们使用由商业移动位置数据服务提供商 Blogwatcher Inc. 提供的匿名手机位置数据集。位置记录通过日本的 140 多款手机应用收集，且已征得所有用户的知情同意。每条记录都包含唯一的匿名用户 ID、纬度、经度和时间戳。为保护用户隐私，所有记录在提供给作者之前均已匿名化，确保无法通过位置轨迹识别个别用户。我们的分析使用了 2019年7月至9月期间在东京观测到的全球定位系统用户。我们将样本限制为在超过 90 个不同日期均有位置记录的活跃用户，以确保个人移动性数据覆盖充分，从而支持后续的工作者识别算法。筛选后，最终样本共包含 329,149 名用户。根据东京都政府的基本住民登记，2020 年 1 月东京总人口为 13,740,732 人。排除 0–19 岁和 70 岁及以上、较不可能从事全职工作的群体后，劳动年龄人口约为 900 万。因此，我们的移动性数据样本约占东京劳动年龄人口的 3.5%。补充表 S1 和补充图 S1 显示，移动性样本在空间上具有代表性，各市镇的用户数量与相应人口分布高度吻合（$r = 0.97$）。我们使用[38], 提出的算法，将原始轨迹记录转换为停留点：在 100 米半径范围内对连续的全球定位系统记录进行聚类。随后，我们仅保留持续时间超过 10 分钟的停留点（更多详情见补充说明 1.1）。

交易数据

我们使用帝国数据银行有限公司提供的企业概况信息和企业间交易数据。帝国数据银行有限公司是一家私营企业信用调查公司，维护着日本最全面的企业信息数据库之一。这些关系通过TDB企业信用调查收集，在调查中，调查人员访谈企业并确定其交易伙伴。数据可分为两部分：企业概况数据集和企业间交易数据集（补充图S2）。两个数据集可以通过唯一企业代码进行关联。企业概况数据集涵盖全国约146万家企业总部。数据包括唯一企业代码、总部地址和行业分类。企业间交易数据提供买方—供应商关系的年度记录，并包括买方和供应商双方的企业代码。该数据约占日本经济总销售额的90%。然而，由于交易数据是在企业总部层级记录的，

无法观测单个经营场所之间的交易。在本研究中，我们聚焦于总部位于东京的196,317家企业，排除了位于伊豆-小笠原群岛的企业，因为这些岛屿在地理上与本研究所考察的东京城市区域相隔离。我们使用涵盖2016年至2022年期间的企业间交易数据。通过共享的企业代码将这些交易记录与企业概况数据关联起来，即可构建纵向企业间交易网络。

估计劳动者之间的接触机会

我们将大规模流动性数据与企业总部所在地数据相结合，量化城市中来自不同企业的劳动者彼此接触的机会。由于东京极端的空间密度和垂直复杂性使建筑层面的工作地点推断极易受到识别误差的影响，我们使用 Uber H3 网格系统第10级分辨率（单元面积约为0.015平方公里），将企业位置聚合为六边形网格。至少包含一个已登记企业位置的单元被指定为企业单元（更多详情请参阅补充说明1.3）。

对于每位用户，我们考察其在工作日的停留点，并排除日本国定假日。如果用户在工作日的白天时间窗口（上午9:00–下午6:00）内，在某个企业单元停留至少5小时（不包括日本国定假日），则将该日归类为工作日。我们将每位用户工作日数量最多的企业单元确定为其主要工作地点。如果用户在其主要工作地点至少有40个工作日，且观测期内至少涵盖12个不同周，则将其归类为劳动者。我们进一步在补充图S4中检验已识别工人样本对每日停留时间、最低工作日数和最低活跃周数等替代阈值的稳健性。为验证工人样本的空间代表性，我们将估计的劳动者人口与东京2020年人口普查进行比较，发现两者在基本单位区块层级和町–字层级均具有高度相关性（$r = 0.80$ 和 $r = 0.90$；补充图S5）。补充图S6显示，已识别的劳动者还呈现出稳定的通勤模式，能够规律地到达其工作地点，进一步支持了该样本的可靠性。

Following the approach of[26], 我们基于劳动者的空间和时间共同在场情况，估计潜在的接触机会。令n_{ict} 和n_{jct} 分别表示在30分钟时间窗口t内，来自企业单元i 和企业单元j 、位于空间网格单元c中的劳动者数量。我们将企业单元i和j之间的会面总数定义为$\mathrm{M}_{ij} = \sum_{c,t} n_{ict} n_{jct}$，该数值统计了劳动者对在同一地点和时间窗口内共同在场的次数。进一步地，我们将潜在的最大会面数定义为$\mathrm{PMM}_{ij} = \sum_t (\sum_c n_{ict})\,(\sum_c n_{jct})$，该指标反映所有可能的共同在场机会总数，即在同一时间窗口内，来自两个企业单元的劳动者在数据集中同时处于活跃状态的次数，而不考虑地点。随后，将会面概率定义为$\mathrm{MP}^{ij} = \mathrm{M}_{ij}\mathrm{PMM}_{ij}$。按照定义，$\mathrm{MP}_{ij}$衡量的是：从企业单元$i$中随机选取一名劳动者、从企业单元$j$中随机选取另一名劳动者时，观察到二者在同一时间位于同一地点的概率。该方法将同时发生的空间和时间重叠作为面对面接触的必要条件，但不直接观测互动本身。

工人接触对企业间交易影响的建模

为考察工人的相遇机会是否与企业间经济联系相关，我们估计了逻辑回归模型（补充说明3.1）。每条观测对应一对企业单元，即单元i 和单元j。因变量$Y_{ij,\ t}$ 表示流动性观测期之后或之前t 年的两个单元中企业之间的交易结果。此处，t 表示与流动性测量年份之间的时间间隔（例如，$t = 1$ 对应于下一年的结果）。主要解释变量为会面概率的对数，即 $\ln(\mathrm{MP}_{ij})$，用于衡量流动性观测期内来自两个不同企业单元的劳动者之间的接触机会。

我们估计以下设定：

$$\mathrm{logit}[\Pr(Y_{ij,t} = 1)] = \beta \ln(\mathrm{MP}_{ij}) + \mathbf{X}'_{ij}\gamma + \delta_i + \delta_j. \tag{1}$$

其中，我们纳入单元固定效应 δ_i和 δ_j，以吸收各单元的不可观测异质性；$\mathbf{X}_{ij}$为一组二元组层面控制变量：我们控制单元质心之间的地理距离以及行业重叠度，后者定义为两个单元内企业的行业份额分布的内积。我们还控制地方经济和劳动力特征的差异，包括就业规模、企业密度和劳动力性别构成。这些变量控制了可能同时影响工人接触机会和企业间交易概率的空间、行业及社会经济相似性。标准误按单元i 和j进行双向聚类。

数据可用性

本研究分析的流动性数据经授权后从 Blogwatcher Inc. 获得。根据协议条款，原始数据的访问受到限制，因此无法公开。有关如何申请访问数据及其相关条件和限制的信息，请参阅 https://www.blogwatcher.co.jp/。

企业交易数据经授权后从帝国数据库有限公司获得，且不对外公开。企业实体可通过 https://www.tdb.co.jp/ 向帝国数据库有限公司购买这些数据。

代码可用性

分析使用 Python 完成。用于根据汇总数据复现图中主要结果的代码已公开发布在 GitHub https://github.com/chenket/mobility-transaction-network。

参考文献

1. Bernard, A. B., Moxnes, A. & Saito, Y. U. 生产网络、地理与企业绩效。《政治经济学杂志》127, 639–688 (2019)。2. Acemoglu, D., Carvalho, V. M., Ozdaglar, A. & Tahbaz-Salehi, A. 总量波动的网络起源。《计量经济学》80, 1977–2016 (2012)。3. Inoue, H. & Todo, Y. 冲击通过供应链网络在企业层面的传导。《自然·可持续发展》2, 841–847 (2019)。4. Bernard, A. B., Dhyne, E., Magerman, G., Manova, K. & Moxnes, A. 企业异质性的起源：一种生产网络方法。《政治经济学杂志》130, 1765–1804 (2022)。5. Barrot, J.-N. & Sauvagnat, J. 投入专用性与特质冲击在生产网络中的传导。《经济学季刊》131, 1543–1592 (2016)。6. Carvalho, V. M., Nirei, M., Saito, Y. U. & Tahbaz-Salehi, A. 供应链中断：东日本大地震的证据。《经济学季刊》136, 1255–1321 (2021)。7. Yang, S., Ogawa, Y., Ikeuchi, K., Shibasaki, R. & Okuma, Y. 灾害后供应链中断：使用图神经网络预测企业层面的销售额。《国际灾害风险削减杂志》110, 104664 (2024)。8. Dhyne, E., Kikkawa, A. K., Kong, X., Mogstad, M. & Tintelnot, F. 存在固定成本的内生生产网络。《国际经济学杂志》145, 103841 (2023)。9. Elliott, M., Golub, B. & Leduc, M. V. 供应网络形成与脆弱性。《美国经济评论》112, 2701–2747 (2022)。10. Oberfield, E. 投入产出结构理论。《计量经济学》86, 559–589 (2018)。11. Taschereau-Dumouchel, M. 内生生产网络经济中的级联与波动。《经济研究评论》93, 1354–1392 (2026)。12. Rauch, J. E. 国际贸易中的商业与社会网络。《经济文献杂志》39, 1177–1203 (2001)。13. Atkin, D., Khandelwal, A. K. & Osman, A. 出口与企业绩效：一项随机实验的证据。《经济学季刊》132, 551–615，DOI：10.1093/qje/qjx002 (2017)。14. Macchiavello, R. & Morjaria, A. 关系的价值：肯尼亚玫瑰出口遭遇供应冲击的证据。《美国经济评论》105, 2911–2945 (2015)。15. Allen, T. 贸易中的信息摩擦。《计量经济学》82, 2041–2083 (2014)。16. Steinwender, C. 信息摩擦的实际影响：美国与英国合并之时。《美国经济评论》108, 657–696，DOI：10.1257/aer.20150681 (2018)。17. Granovetter, M. 经济行动与社会结构：嵌入性问题。《美国社会学杂志》91, 481–510 (1985)。18. Leamer, E. E. 一个平坦的世界、一个公平的竞争环境、一个终究很小的世界，还是以上皆非？评Thomas L. Friedman《世界是平的》。《经济文献杂志》45, 83–126，DOI：10.1257/jel.45.1.83 (2007)。19. Storper, M. & Venables, A. J. 喧嚣：面对面接触与城市经济。《经济地理学杂志》4, 351–370 (2004)。20. Battiston, D., Blanes i Vidal, J. & Kirchmaier, T. 组织中的面对面交流。《经济研究评论》88, 574–609 (2021)。21. Emanuel, N., Harrington, E. & Pallais, A. 与同事邻近的力量：为明天而培训，还是提升今天的生产率？技术报告，美国国家经济研究局 (2023)。22. Inoue, H., Nakajima, K., Okazaki, T. & Saito, Y. U. 人际互动在创新中的作用：来自日本1918年流感大流行的证据。《日本经济评论》76, 795–820 (2025)。23. Cristea, A. D. 国际贸易中的买卖双方关系：来自美国各州出口与商务舱旅行的证据。《国际经济学杂志》84, 207–220 (2011)。

24. Söderlund, B. 商务旅行对贸易的重要性：来自苏联领空开放的证据。《国际经济学杂志》145, 103812 (2023)。25. Chai, S. & Freeman, R. B. 临时共址与协作发现：谁在会议上交流？《战略管理杂志》40, 2138–2164，DOI：10.1002/smj.3062 (2019)。26. Atkin, D., Chen, M. K. & Popov, A. 面对面互动的回报：硅谷的知识溢出。美国国家经济研究局工作论文系列 (2022)。27. Xu, F. 等。利用人类流动数据量化人们经历的城市不平等。《自然·人类行为》9, 654–664 (2025)。28. Miyauchi, Y., Nakajima, K. & Redding, S. J. 空间流动的经济学：理论与基于智能手机数据的证据。《经济学季刊》140, 2507–2570，DOI：10.1093/qje/qjaf038 (2025)。29. Kreindler, G. E. & Miyauchi, Y. 利用手机记录衡量城市内部的通勤与经济活动。《经济与统计评论》105, 899–909 (2023)。30. Matsumura, K., Oh, Y., Sugo, T. & Takahashi, K. 利用流动性数据即时预测经济活动。《日本国际经济学杂志》73, 101327 (2024)。31. Yanotti, M. B., Navon, Y., de Silva, A., Angelopoulos, S. & Sinclair, S. 步行经济：作为区域经济活动实时指标的行人计数。《经济分析与政策》 (2026)。32. Yabe, T., García Bulle Bueno, B., Frank, M. R., Pentland, A. & Moro, E. 地点间基于行为的依赖网络塑造城市经济韧性。《自然·人类行为》9, 496–506 (2025)。33. Blumenstock, J., Cadamuro, G. & On, R. 根据手机元数据预测贫困与财富。《科学》350, 1073–1076 (2015)。34. Wang, S. 等。低频活动可预测美国主要城市的经济结果。《自然·城市》1, 305–314 (2024)。35. Nilforoshan, H. 等。人类流动网络揭示大城市隔离加剧。《自然》624, 586–592 (2023)。36. Moro, E., Calacci, D., Dong, X. & Pentland, A. 流动模式与人们在美国大城市中经历的收入隔离相关。《自然·通讯》12, 4633 (2021)。37. Yabe, T., Bueno, B. G. B., Dong, X., Pentland, A. & Moro, E. 新冠肺炎疫情期间的行为变化降低了城市交往中的收入多样性。《自然·通讯》14, 2310 (2023)。38. Hariharan, R. & Toyama, K. Project lachesis：解析并建模位置历史记录。《国际地理信息科学会议》106–124 (2004)。39. 日本统计局。2020年人口普查：劳动力基本统计 (2020)。40. Pavitt, K. 技术变革的部门模式：迈向分类法与理论。《研究政策》13, 343–373 (1984)。41. Malerba, F. 部门创新与生产体系。《研究政策》31, 247–264，DOI：10.1016/S0048-7333(01)00139-1 (2002)。42. Gulati, R. & Gargiulo, M. 组织间网络从何而来？《美国社会学杂志》104, 1439–1493 (1999)。43. Uzzi, B. 企业间网络中的社会结构与竞争：嵌入悖论。《行政科学季刊》42, 35–67，DOI：10.2307/2393808 (1997)。44. Gulati, R. 社会结构与联盟形成模式：纵向分析。《行政科学季刊》619–652 (1995)。45. Miyauchi, Y. 匹配与集聚：来自日本企业间贸易的理论与证据。《计量经济学》92, 1869–1905 (2024)。46. Cai, J. & Szeidl, A. 企业间关系与企业绩效。《经济学季刊》133, 1229–1282 (2018)。47. Duranton, G. & Puga, D. 城市集聚经济的微观基础。载于 Henderson, J. V. & Thisse, J.-F. (编)《区域与城市经济学手册》，第4卷，2063–2117页 (Elsevier, 2004)。48. Ellison, G., Glaeser, E. L. & Kerr, W. R. 什么导致产业集聚？来自共同集聚模式的证据。《美国经济评论》100, 1195–1213 (2010)。49. Catalini, C. 微观地理与发明活动的方向。《管理科学》64, 4348–4364 (2018)。50. Juhász, S. & Lengyel, B. 集群知识网络中关系的建立与延续。《经济地理学杂志》18, 1203–1226 (2018)。51. Chaney, T. 国际贸易的网络结构。《美国经济评论》104, 3600–3634 (2014)。

52. Cai, J.、Lin, W. 和 Szeidl, A.《生产网络中的企业间访问》（2026年）。工作论文，1月26日修订版。
53. Yabe, T. 等。 COVID-19疫情期间，非强制性措施已足以降低东京的人类流动。《科学报告》 **10**，18053（2020年）。

致谢

我们谨感谢帝国数据库有限公司和Blogwatcher Inc.慷慨提供数据。本研究是与帝国数据库有限公司开展联合研究的一部分。K.C.感谢JST SPRING（资助编号：JPMJSP2108）的支持。H.I.感谢日本学术振兴会科学研究费助成事业（资助编号：JP23K25520、JP24K00247、JP24K01108、JP25K01454、JP26K00349和JP26K22092）的支持。Y.O.感谢日本学术振兴会科学研究费助成事业（资助编号：JP24K01000）的支持。T.Y.感谢日本学术振兴会科学研究费助成事业（资助编号：24K01108）的支持。本研究还部分包含跨部门战略创新促进计划（SIP）第三期“智能移动平台开发”的成果。该计划由内阁府科学技术与创新委员会推进（项目管理机构：新能源与产业技术综合开发机构（NEDO）；项目代码：JPNP23023）。

作者贡献声明

K.C. 设计了算法，进行了分析，并开发了模型和模拟。H.I. 和 T.Y. 部分参与了模型和模拟的开发。H.I.、Y.S.、Y.O. 和 T.Y. 负责监督研究。所有作者均撰写了论文。公司数据由 K.C. 和 Y.O. 处理。所有作者均审阅了稿件。

利益冲突声明

作者声明不存在利益冲突。

《劳动者共同在场与企业间网络演化相关联》的补充说明

Ketian Chen[1,2], Hiroyasu Inoue[3,4],Yoshihide Sekimoto[1,2], YoshikiOgawa[2,5],和 Takahiro Yabe[6,7,*]

[1]东京大学工学研究生院，日本东京[2]东京大学空间信息科学中心，日本东京[3]兵库县立大学信息科学研究生院，日本神户[4]理化学研究所计算科学中心，日本神户[5]一桥大学社会数据科学研究生院，日本东京[6]纽约大学坦登工学院城市科学与发展中心，美国纽约[7]纽约大学坦登工学院技术管理与创新系，美国纽约[*]通讯作者：takahiroyabe@nyu.edu

2026年11月

目录

1 数据来源与预处理 3

1.1 流动性数据.3 1.2 交易数据.3 1.3 企业单元构建.7 1.4 工人识别.7

2 会面概率的构建与验证 12

2.1 会面概率（MP）的测量.12 2.2 会面概率的网络结构.13 2.3 替代会面概率测量.13 2.4 时间平移零假设MP.15 2.5 基于引力模型的零假设会面概率.17

3 回归分析与稳健性检验 18

3.1 会面概率与交易结果的基准关联 18 3.2 对企业数量控制变量的稳健性.21 3.3 对替代会面概率测量的稳健性.21 3.4 使用时间平移零假设MP的零假设检验.21 3.5 保持度数的交易网络零模型.22 3.6 对替代交易历史窗口的稳健性.24 3.7 会面概率与链接形成关联的时间变化.25 3.8 距离异质性.29 3.9 行业异质性.30 3.10 职能部门组之间的行业异质性.37 3.11 两跳流动暴露与链接形成.39

1 数据来源与预处理

1.1 流动性数据

本研究使用了由日本位置数据服务公司 Blogwatcher, Inc. 提供的匿名手机位置数据集。位置记录通过 140 多款手机应用收集，并已获得所有用户的知情同意。每条记录均包含唯一的匿名用户 ID、纬度、经度和时间戳。为保护用户隐私，所有记录在提供给作者之前均已匿名化，确保无法通过位置轨迹识别个别用户。有关该数据集的更多信息，包括数据收集流程、隐私保护措施及其局限性，可在 Blogwatcher 网站（https://www.blogwatcher.co.jp/terms/）上查阅。

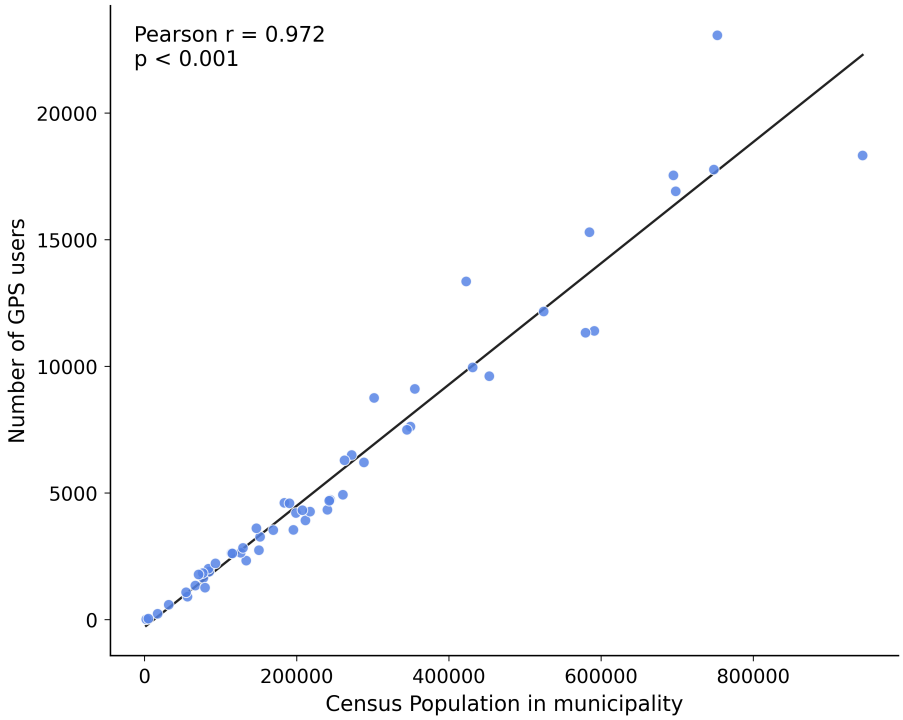
我们的分析聚焦于2019年7月至9月期间在东京被观测到的用户。我们将样本限定为在超过90个不同日期有位置记录的活跃用户，以确保对个人移动情况有足够的覆盖，从而支持后续的工人识别算法。筛选后，最终样本共包含329,149名用户。根据东京都政府的基本住民登记，2020年1月东京总人口为13,740,732人。剔除0至19岁及70岁及以上、较不可能从事全职工作的人员后，劳动年龄人口约为900万。因此，我们的流动性数据样本约占东京劳动年龄人口的3.5%。

为识别有意义的到访，我们按照 [1]的算法，将所有全球定位系统轨迹记录转换为停留点。停留点定义为一组连续位置采样点形成的空间簇，其与质心的最大距离低于距离阈值，且持续时间达到时长阈值。本研究将距离阈值设为100米，最短持续时间设为10分钟。随后，根据停留点推断每位用户的居住地。对于每位用户的每一天，我们将夜间时段（晚上10点至早上6点）累计停留时间最长的停留点确定为候选居住地。在整个样本时间范围内被选中次数最多的位置，则被指定为该用户的居住地。

我们使用用户推断的住所位置评估了移动性样本的空间代表性。日本在行政区划上分为市镇、町、村和特别区。对于每个市镇，我们将观测到的手机用户数量除以2020年人口普查公布的总人口，计算采样率。结果显示，市町村级抽样率平均约为2.2%，各区的分布见补充表1。在对数–对数尺度上，观测到的用户数量与官方人口高度相关（皮尔逊相关系数 $r = 0$。972，$p < 0$。001；补充图1）。这些结果表明，GPS 样本的空间分布总体上反映了东京真实的人口分布。

1.2 交易数据

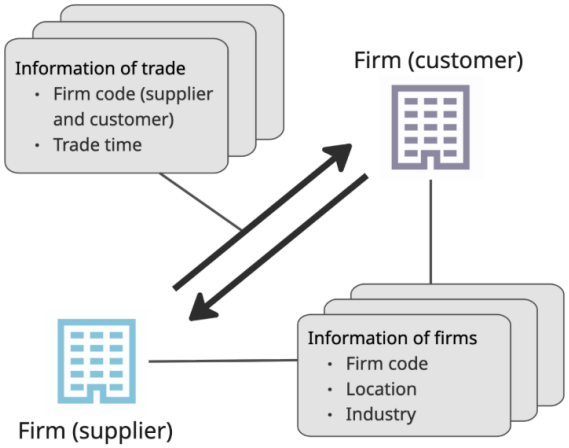
我们使用帝国数据银行（TDB）提供的企业概况信息和企业间交易数据（https://www.tdb.co.jp/）。帝国数据银行（TDB）是一家私人企业信用调研机构，



补充图1：**人类流动数据的空间代表性**。比较各市镇的人口普查人口（横轴）与全球定位系统用户数量（纵轴）。这两项指标呈现强烈的正相关（皮尔逊相关系数 r =0。972， $p < 0$。001）。

补充表1：东京各市镇的 GPS 采样率

Code	市镇	人口普查人口	GPS 用户	采样率（%）
101	千代田区	66,680	1,348	2.02
102	中央区	169,179	3,540	2.09
103	港区	260,486	4,938	1.90
104	新宿区	349,385	7,622	2.18
105	文京区	240,069	4,340	1.81
106	台东区	211,444	3,924	1.86
107	墨田区	272,085	6,503	2.39
108	江东区	524,310	12,171	2.32
109	品川区	422,488	13,357	3.16
110	目黑区	288,088	6,213	2.16
111	大田区	748,081	17,780	2.38
112	世田谷区	943,664	18,336	1.94
113	涩谷区	243,883	4,732	1.94
114	中野区	344,880	7,497	2.17
115	杉并区	591,108	11,403	1.93
116	丰岛区	301,599	8,757	2.90
117	北区	355,213	9,116	2.57
118	荒川区	217,475	4,268	1.96
119	板桥区	584,483	15,299	2.62
120	练马区	752,608	23,080	3.07
121	足立区	695,043	17,551	2.53
122	葛饰区	453,093	9,620	2.12
123	江户川区	697,932	16,918	2.42
201	八王子市	579,355	11,329	1.96
202	立川市	183,581	4,616	2.51
203	武藏野市	150,149	2,747	1.83
204	三鹰市	195,391	3,548	1.82
205	青梅市	133,535	2,331	1.75
206	府中市	262,790	6,293	2.39
207	昭岛市	113,949	2,623	2.30
208	调布市	242,614	4,699	1.94
209	町田市	431,079	9,964	2.31
210	小金井市	126,074	2,640	2.09
211	小平市	198,739	4,215	2.12
212	日野市	190,435	4,593	2.41
213	东村山市	151,815	3,276	2.16
214	国分寺市	129,242	2,832	2.19
215	国立市	77,130	1,648	2.14
218	福生市	56,414	908	1.61
219	狛江市	84,772	1,894	2.23
220	东大和市	83,901	2,016	2.40
221	清濑市	76,208	1,842	2.42
222	东久留米市	115,271	2,620	2.27
223	武藏村山市	70,829	1,781	2.51
224	多摩市	146,951	3,610	2.46
225	稻城市	93,151	2,225	2.39
227	羽村市	54,326	1,082	1.99
228	秋留野市	79,292	1,266	1.60
229	西东京市	207,388	4,324	2.08
303	瑞穗町	31,765	590	1.86
305	日之出町	16,958	236	1.39



补充图2： **交易数据的结构**。

这家搜索公司在日本拥有最全面的企业信息资料库之一。TDB通过企业信用调查收集企业及其交易伙伴的信息，包括现场访谈和利用外部来源进行核实。虽然记录交易伙伴的数量没有正式限制，但每份报告可能无法涵盖企业的所有交易关系。数据库还通过识别其他企业报告的额外交易关系来补充信息。最终形成的数据库记录了企业之间报告的供应商—客户关系。与2021年经济普查进行比较后发现，该数据集涵盖了约70%的日本企业，所记录的销售额占日本经济总销售额的90%以上。该数据集基于在47个都道府县开展的企业调查，因此覆盖全国。

这些数据可分为两部分：企业概况数据集和企业间交易数据集（补充图2）。两个数据集可通过唯一企业ID关联。企业概况数据集包含唯一企业ID、总部地址和行业分类。企业间交易数据提供买方—供应商关系的年度记录，并包含买方和供应商双方的企业ID。交易数据按总部层级记录，因此无法观察单个营业场所之间的交易。本研究聚焦于总部位于东京的196,317家企业，不包括位于伊豆-小笠原群岛的企业，因为这些偏远岛屿在地理上与本研究所指的东京连续城市区域隔离。我们使用涵盖2016至2022财年的企业间交易数据。通过共同的企业ID将这些交易记录与企业档案数据关联，即可构建纵向企业间交易面板。

原始行业信息采用日本行业分类（JSIC）1993年修订版，该分类包含13个宽泛的行业门类。这些门类将经济活动差异显著的领域归并在一起，尤其是在服务业相关行业中，

这限制了我们考察行业异质性的能力。为获得更详细且具有国际可比性的分类，我们将原始JSIC类别与北美行业分类系统（NAICS）进行了协调。具体而言，我们先将JSIC类别映射到《国际经济活动标准行业分类》（ISIC）第4次修订版，再映射到相应的NAICS行业部门。采用这一分类后，我们区分出20个行业部门，从而能够更详细地分析不同行业间企业间接触机会的差异。

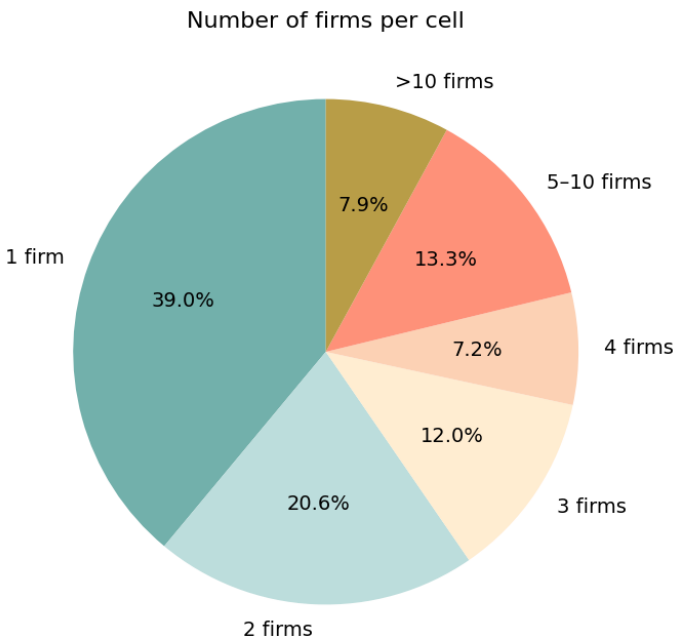
1.3 企业单元的构建

东京的城市环境具有较高的空间密度和垂直复杂性。建筑物彼此距离很近，许多企业办公室位于同一栋建筑内。在这种条件下，基于GPS轨迹在建筑物层面推断工作地点会产生误差，因为即使定位存在很小的不准确，也可能导致点落在正确建筑轮廓之外。因此，我们使用H3六边形网格表示企业位置。企业档案数据中的每个总部所在地都被映射到一个分辨率为10的H3六边形网格，其平均边长约为75米。包含至少一个企业位置的单元被定义为企业单元。企业在企业单元中的分布如补充图3所示。单企业单元占所有企业单元的39.0%，而包含2家企业的单元还占20.6%。总体而言，59.6%的企业单元包含不超过2家企业，78.8%包含不超过4家企业。包含超过10家企业的单元仅占总数的7.9%。

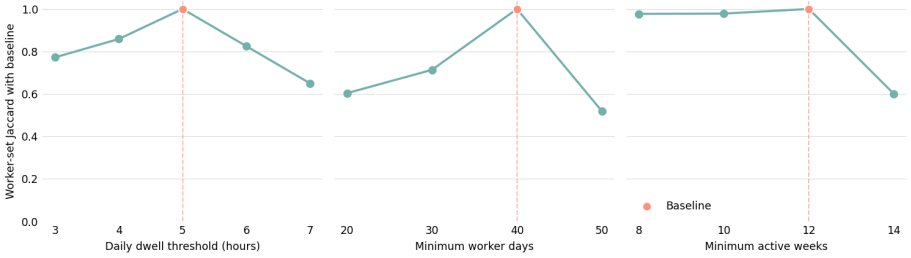
1.4 工人识别

工人识别程序 我们根据劳动者在潜在工作地点反复出现的白天停留情况来识别劳动者。首先，我们保留在工作日（不包括日本国定假日）上午9:00至下午6:00之间观测到的停留点。根据《2019年劳动条件综合调查》，这一时间窗口大致涵盖了日本典型的日间工作时间安排。我们删除位于用户推断的住所位置内的停留点。随后，将每个剩余的停留点分配至相应的企业单元。如果用户在某一天于某企业单元内的总停留时间至少达到5小时，则视为该用户当天到访过该企业单元；这样的日期被归类为工作日。接着，我们将在观测期内此类工作日数量最多的企业单元定义为用户的主要工作地点。为区分常规劳动者与偶尔到访者，我们保留以下用户：其主要工作地点至少在40个工作日被观测到，且在三个月研究期内至少跨越12个不同的周。满足这些条件的用户被归类为与其所识别的主要工作场所单元格相关联的劳动者。该程序识别出52,785名通勤员工，并将每人分配至一个主要工作场所单元格。

对参数选择的敏感性 为评估工人识别对不同参数选择的敏感性，我们比较由此得到的替代工人集合，W_s,



补充图3：**企业在企业单元中的分布**。单企业单元占所有企业单元的39.0%，而59.6%的企业单元包含不超过两家企业，78.8%的企业单元包含四家或更少企业。



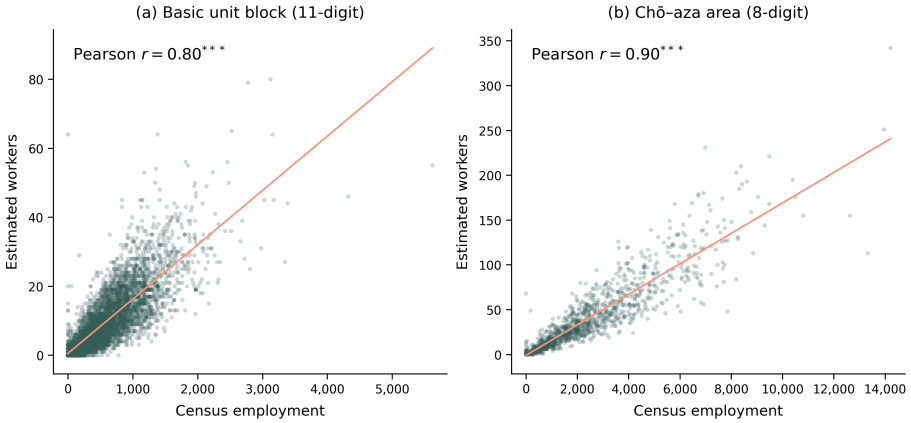
补充图4：**不同参数阈值下工人识别的敏感性**。每个面板报告了在替代参数下识别的工人集合与基准下识别的工人集合之间的Jaccard相似度。每次仅改变一个参数，同时将另外两个参数保持在其基准值。基准设定要求每日工作场所停留时间至少为5小时、工作日不少于40天，且活动至少覆盖12周。

与基准工人集合相比， W_0。工人集合的稳定性通过Jaccard相似度衡量：

$$J(W_0, W_s) = \frac{|W_0 \cap W_s|}{|W_0 \cup W_s|}. \tag{1}$$

数值为1表示基准设定与替代设定识别出完全相同的劳动者，而较低的数值表示样本构成存在更大差异。我们每次改变一个参数，同时将其余两个参数保持在基准值。替代值包括每日停留阈值 [3, 4, 5, 6,7] 小时、最低工作日要求 [20, 30, 40, 50] 天，以及最低活跃周要求 [8, 10, 12, 14] 周。补充图4显示，在一系列替代阈值下，识别出的工人集合与基准集合仍具有较高相似性。总体而言，敏感性分析支持基准参数选择的可行性。当每日停留阈值围绕5小时的基准值进行适度调整时，识别出的工人集合仍相对稳定；在4小时和6小时设定下，Jaccard相似度分别约为0.86和0.82。当最低活跃周要求在8至12周之间变化时，工人集合也高度稳定，这表明在整个观测期内，基准设定下识别出的多数劳动者都表现出持续性的工作模式。对于最低工作日要求，围绕40天基准值设置的替代阈值仍能与基准集合保持较大重叠。综上，这些结果表明，基准设定能够识别出一个稳定的规律通勤者核心群体，且对识别阈值的适度变化不太敏感。

已识别工人的空间代表性我们从两个维度验证工人识别程序：检测到的工人样本相对于日本人口普查的空间代表性，以及

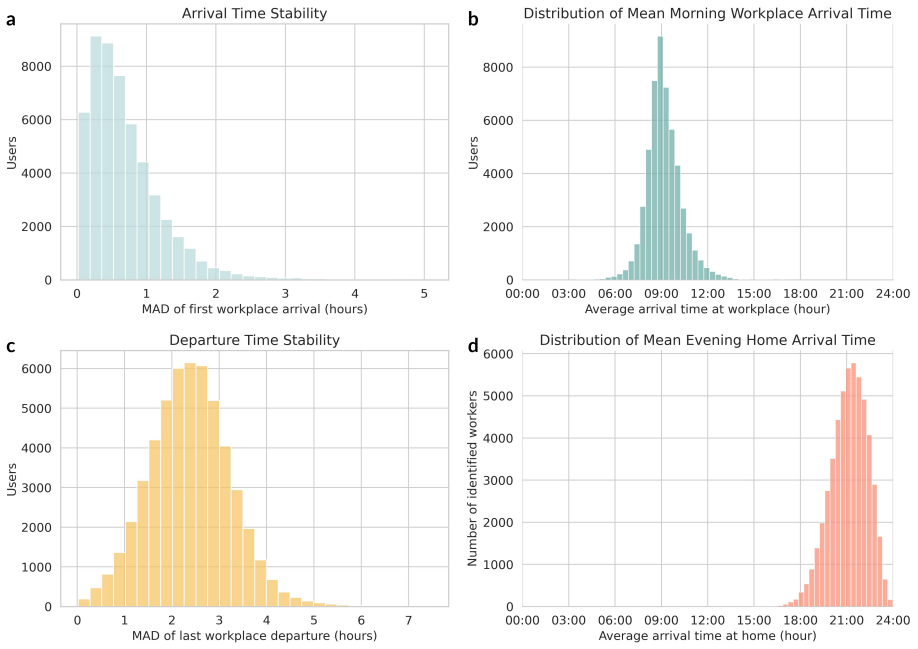


补充图5：**已识别工人样本的空间代表性**。（a）使用完整的11位人口普查区域代码测量的每个基本单位区块中的人口普查就业人数（x轴），与从流动性数据中识别出的相应区域劳动者人数（y轴）之间的比较；皮尔逊相关系数 $r = 0$。80。（b）将基本单位区块依据相应的8位区域代码汇总至更广的ch¯o–aza层级后的相同比较；皮尔逊相关系数 $r = 0$。90。两个空间分辨率下均呈现正相关，表明已识别的劳动者样本总体上能够反映人口普查所报告的办公室就业空间分布。

根据已识别工人观察到的出行模式评估其通勤规律性。

我们首先通过将已识别劳动者样本的空间分布与日本2020年人口普查在两种空间分辨率下的官方就业统计进行比较，评估该样本的代表性。第一种是基本单位区块，这是由完整11位区域代码表示的最细人口普查单元。第二种是由8位区域代码表示的更广泛的ch¯o–aza单元。本研究使用企业总部所在地，不考虑其分支机构、门店或工厂。因此，相关的人口普查基准应反映更可能集中于总部及其他办公地点的职业。我们将管理人员、专业技术人员和文职人员归类为办公室职业（日本2020年人口普查、劳动力状况基本统计表）。补充图5显示了两种不同空间分辨率下人口普查就业人数与已识别劳动者人数之间的空间相关性。在更精细的基本单位区块层级，两个指标高度相关，皮尔逊相关系数为r=0.80。将数据汇总至更广泛的ch¯o–aza层级后，相关性提高至r=0.90。这一结果表明，已识别的劳动者样本能够较好地代表东京办公室就业的空间分布。

通勤规律性的验证 我们进一步验证所识别的劳动者样本，考察这些用户是否呈现出重复性的日常通勤模式。



补充图6：**已识别工人的通勤模式**。(a) 每位工人在推断的工作地点每日首次到达时间的中位绝对偏差（中位绝对偏差）的分布。大多数工人的早晨到达时间较为稳定，中位绝对偏差小于一小时。(b) 工人在推断的工作地点每日首次到达时间的平均值分布。大多数工人于上午9:00左右到达。(c) 每位工人从推断的工作地点每日最后离开时间的中位绝对偏差分布。(d) 工人在推断的住所位置每日到达时间的平均值分布。大多数工人在下午6:00之后回家。

补充图6a显示，大多数劳动者每天首次到达工作地点的时间变化相对较小，表明他们在各工作日反复于相近时间到达推断的工作地点。补充图6b进一步显示，工作地点的平均到达时间集中在早晨，分布峰值约为上午9:00。正如补充图6c所示，离开工作地点的时间变化更大，这与工作日结束时间存在更大的日间波动相一致。补充图6d显示，到达推断的住所位置的平均时间集中在傍晚，主要在晚上8:00至11:00之间。早晨定时抵达推断的工作地点、傍晚返回推断的住所位置，这一规律与东京的日常通勤情况相符。这些模式与规律性通勤相一致，也支持将这些用户归类为与推断的工作场所单元相关的劳动者。

2 会议概率的构建与验证

2.1 会面概率（MP）的测量

遵循 [2], 的方法，我们根据劳动者的时空共现构建会面概率。设 n_{ict} 和 n_{jct} 分别表示在 30分钟时间窗口 t 内，来自企业单元 i 和企业单元 j、位于单元 c 中的劳动者数量。

我们将企业单元 i 与 j 之间的会面总数定义为

$$M_{ij} = \sum_{c,t} n_{ict}\, n_{jct},$$

该指标统计了来自 i 和 j 的劳动者在同一时间位于同一单元中的共在事件数量。

我们进一步将潜在最大会面次数定义为

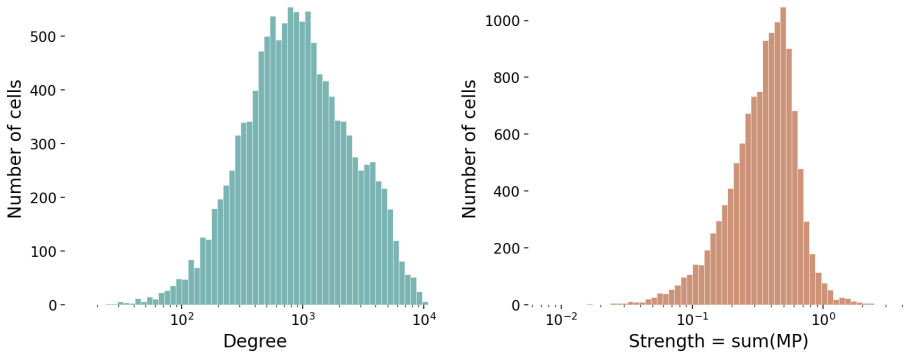
$$PMM_{ij} = \sum_{t} \left(\sum_{c} n_{ict} \right) \left(\sum_{c} n_{jct} \right),$$

该指标反映了可能发生的共同在场机会的最大数量，也就是来自两个企业单元的劳动者在同一时间窗口内同时活跃的次数，即使他们并不在同一地点。此归一化假设，在同一时间窗口内活跃的劳动者无论具体位置如何，都可能发生互动。

会面概率定义如下：

$$MP_{ij} = \frac{M_{ij}}{PMM_{ij}}.$$

根据定义，MP_{ij}衡量从企业单元 i 随机选取的一名劳动者与从企业单元 j随机选取的一名劳动者在时空上同时共处的概率。这种方法将时空上的同时重叠作为发生面对面接触的必要条件，但不会直接观察互动本身。



补充图7：**流动网络中的节点度和强度**。(a) 节点度分布，节点度定义为通过非零会面概率（MP >0）与每个企业单元相连的其他企业单元的数量。(b) 节点强度分布，节点强度定义为每个企业单元所有连接的会面概率之和。

2.2 会面概率的网络结构

我们考察由正会面概率所诱导的网络结构，其中节点代表企业单元，边连接会面概率非零的企业单元配对。互补累积分布（补充图 7）呈现重尾模式。中位数企业单元与另外 867 个单元相连，而连接最广泛的单元则与超过 10,000 个单元相连，显示出全市相遇连接广度存在显著差异。这种连通性的不均衡还伴随着节点强度的显著差异；节点强度定义为与某个企业单元相连的所有关系上的会面概率之和。度和强度分布共同表明，企业单元之间不仅在通过移动性接触的其他单元数量上存在显著差异，在这些偶遇的总体强度上也存在显著差异。

2.3 替代会面概率测量

为考察观测到的会面概率（MP）是否过度受到拥挤或人流密集地点中偶然共同在场的影响，我们构建了四种替代性 MP 度量：（1）拥挤度加权 MP；（2）受欢迎度加权 MP；（3）重复会面概率；（4）排除企业单元的会面概率。这些替代指标用于检验，在考虑以下四种可能的测量偏差来源后，观测到的会面概率是否仍然相近：（1）某地点在特定时段异常拥挤时发生的偶遇；（2）在整个观测期内始终人流密集的地点发生的偶遇；（3）仅观测到一次的偶遇；（4）发生在任一劳动者所属企业单元内的偶遇。

拥挤度加权 MP 有些偶遇发生在特定时段异常拥挤的地点，例如高峰时段的主要车站，或活动开始前后不久的大型场馆。在这些情况下，来自两个企业单元的劳动者可能仅仅因为许多人同时在场而被观测到共同出现。为降低这种影响，我们构建了一种拥挤度调整后的会面概率，对发生在出行记录数量较多的单元-时间单位中的偶遇赋予较低权重。令$N_{ct} = \sum_k n_{kct}$ 表示在时间窗口 t内于单元 c 中观测到的劳动者总数，其中 k 索引所有企业单元。我们将调整后的会面次数计算为$M_{ij}^{crowd} = \sum_{c,\ t} \frac{n_{ict} n_{jct}}{N_{ct} - 1}$，并将相应的会面概率定义为$MP_{ij}^{\text{crowd}} = M_{ij}^{\text{crowd}} / PMM_{ij}$。

受欢迎度加权 MP 在整个观测期内，有些地点会吸引大量不同的访客。例如，主要交通枢纽、购物中心以及其他广泛使用的公共设施。这些地点发生的偶遇可能部分反映了其广泛的访客群体。因此，我们构建了受欢迎度加权 MP，对整个观测期内被许多用户到访的单元中发生的偶遇赋予较低权重。令 U_c 表示在整个观测期内到访单元 c 的不同用户总数。我们将调整后的会面次数计算为 $M_{ij}^{\text{pop}} = \sum_{c,\ t} \frac{n_{ict} n_{jct}}{U_c}$，并将相应的会面概率定义为$MP_{ij}^{\text{pop}} = M_{ij}^{\text{pop}} / PMM_{ij}$。

重复会面概率 在特定地点仅观察到一次共同在场，可能是偶然重叠或测量噪声所致。因此，我们使用劳动者来自企业单元 i 和 j，且在至少两个时间区间内共同观测到的地点来构建重复会面概率。令 R_{ijc} 表示劳动者来自企业单元 i和 j，且在单元 c中共同观测到的时间区间数。我们通过保留满足 $R_{ijc} \geq 2$的单元中的所有共同在场贡献，计算调整后的会面次数，并将所得数值记为 M_{ij}^{repeat}。随后，我们将相应的会面概率定义为$MP_{ij}^{\text{repeat}} = M_{ij}^{\text{repeat}} / PMM_{ij}$。

排除企业单元的会面概率一些偶遇可能是机械性产生的，因为劳动者在包含其自身工作地点的单元格内停留了相当长的时间。因此，我们构建了排除企业单元的会面概率。对于每位用户，我们移除其指定工作地点内观测到的所有停留记录，同时保留其他所有地点的停留记录，包括其他企业的工作场所单元。利用筛选后的停留记录，我们重新计算会面次数 M_{ij}^{ex} 以及相应的机会分母 PMM_{ij}^{ex}。随后，我们将调整后的会面概率定义为 $MP_{ij}^{\text{ex}} = M_{ij}^{\text{ex}} / PMM_{ij}^{\text{ex}}$。这一指标用于考察观测到的会面概率是否主要由劳动者停留在自身工作场所单元内所驱动。

补充图 8 和补充表 2 将替代 MP度量与观测到的会面概率进行了比较。总体而言，替代测量指标仍与观测到的会面概率密切关联。拥挤度加权 MP 与观测到的会面概率高度相关，皮尔逊相关系数和斯皮尔曼相关系数分别为 0.819 和 0.835。受欢迎度加权 MP 的皮尔逊相关系数较低，为 0.446，但仍保持了

相对较高的斯皮尔曼相关系数 0.796，表明降权保留了配对的相对排序。重复会面概率则呈现相反的模式，其皮尔逊相关系数为 0.997，斯皮尔曼相关系数为 0.463。这表明，排除一次性相遇主要影响较弱联系的相对排名，同时总体强度结构基本保持不变。排除任一劳动者的企业单元内发生的偶遇后，与观测到的会面概率之间也呈现中等程度的相关性（皮尔逊相关系数 $= 0.601$；斯皮尔曼相关系数 $= 0.669$）。总的来说，这些比较表明，替代调整会以不同程度改变 MP 值和排名。没有任何单一调整能够消除其与观测到的会面概率之间的关联。

补充表 2：观测到的会面概率与四种替代会面概率度量之间的相关性

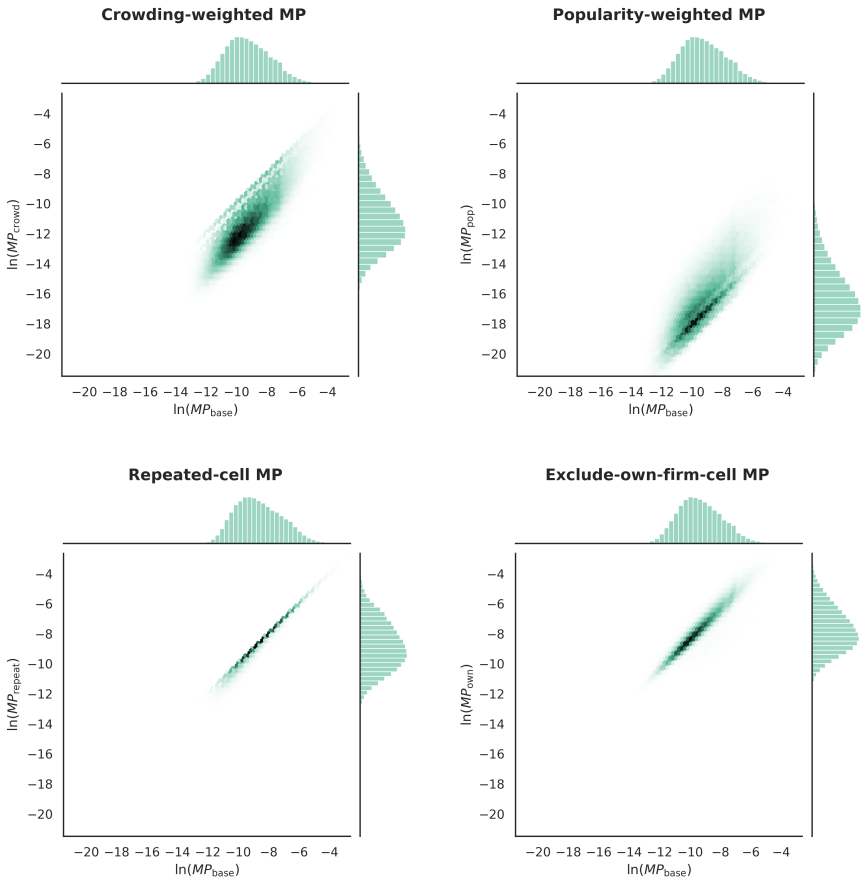
替代度量	皮尔逊相关系数	斯皮尔曼相关系数
拥挤度加权 MP	0.819	0.835
受欢迎度加权 MP	0.446	0.796
重复会面概率	0.997	0.463
排除企业单元的会面概率	0.601	0.669

注：每一行比较一种替代会面概率（MP）度量与基准会面概率在单元格对上的表现。皮尔逊相关系数使用会面概率值计算；斯皮尔曼相关系数使用其秩计算。

2.4 时间平移零模型MP

前几节表明，观测到的会面概率对于能够降低拥挤地点、广受访问地点以及其他特定偶然共处一地来源影响的替代测量指标具有稳健性。接下来，我们考察观测到的会面概率是否反映了时间协调，而不仅仅是访问相同地点。我们构建时间平移零模型，在保留劳动者真实出行模式的同时，打破劳动者之间的时间对齐。我们在观测期内随机将每位劳动者的日常出行轨迹向前或向后平移。为保留工作日与周末出行之间的系统性差异，我们将每位劳动者的出行记录随机重新分配到其他周中相应的日期。工作日记录仅在工作日之间打乱，周末记录仅在周末之间打乱。该过程保留了出行惯例，但破坏了劳动者之间原有的日历日期对齐。在这种情况下，劳动者仍会前往相同的实际地点，但依赖时间对齐的偶遇不再得到系统性保留。例如，两名劳动者可能约定在某天中午于一家咖啡馆见面。将他们的日常轨迹随机平移到不同日期，会破坏这次有计划的会面。

我们独立重复了这一过程，共进行 $B = 100$ 零模型迭代。在每次迭代中，都会生成一组新的针对工人的时间平移，并根据平移后的轨迹重建相遇网络。对于每个单元对（$i,\ j$），我们重新计算实际相遇 $M_{(b)ij}$、潜在相遇 $PMM_{(b)ij}$，以及相应的会面概率，



补充图 8：**会面概率对替代测量指标的稳健性**。每个面板都将观测到的会面概率（MP）与一种替代会面概率度量进行比较：拥挤度加权 MP、受欢迎度加权 MP 、重复会面概率或排除公司单元的 MP。中央六边形图层显示单元格对的局部密度，边缘直方图则显示基准会面概率值和替代会面概率值的分布。

$$MP_{ij}^{(b)} = \frac{M_{ij}^{(b)}}{PMM_{ij}^{(b)}}.$$

$\{MP_{ij}^{(b)}\}_{b=1}^{100}$ 的分布构成了零分布，用于与观测到的会面概率进行比较。

将观测到的企业对相遇概率与经验零分布进行比较

$$\{MP_{ij}^{(1)}, MP_{ij}^{(2)}, \ldots, MP_{ij}^{(100)}\}.$$

上尾经验 p 值的计算方式如下

$$p_{ij}^{\text{upper}} = \frac{1 + \sum_{b=1}^{100} \mathbf{1}\left(MP_{ij}^{(b)} \geq MP_{ij}^{\text{obs}}\right)}{100 + 1}.$$

如果观测到的 MP_{ij}^{obs}系统性地高于时间平移零分布，这表明观测到的共同地点现象不仅是劳动者前往相似地点的结果，也取决于他们的时间同步。

如补充图 9a 所示， MP_{ij}^{obs}的分布相较于时间平移零分布向较高值偏移。在 100 次时间平移实现中，没有一次产生与 MP_{ij}^{obs} 一样高的总体会面概率（ $p < 0。01$）。这表明，共同的空间活动模式只能解释观测到的部分偶遇。这种差异与劳动者之间存在额外的时间对齐相符，但无法据此判断任何共同在场是否涉及计划会面。我们还在补充图 9b 中计算了每个单元格对的观测会面概率与零模型会面概率之比。观测值与零模型值之比的中位数为 1.61，表明 MP_{ij}^{obs} 通常比零模型会面概率高 61%。该比值的第 90 百分位数达到 7.79，说明对于处于上尾 10% 的单元格对，观测到的会面概率几乎是零模型预期值的 8 倍。总之，这些结果表明，时间同步显著影响了 MP_{ij}^{obs}。

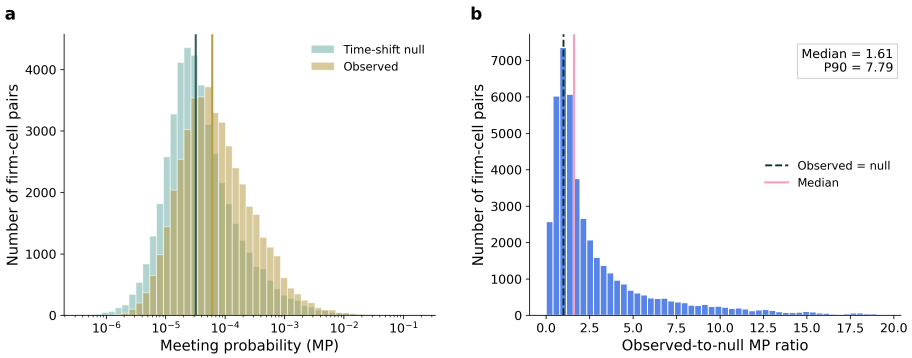
2.5 基于引力的零模型会面概率

地理邻近性或许可以解释企业之间部分观测到的接触机会。来自相邻工作场所的劳动者更有可能造访相同地点，即使他们所属的企业之间没有任何特定关系。根据引力定律，我们将工作场所 i 和 j 之间的预期会面次数建模为

$$\mu_{ij}^{gravity} = \alpha \frac{W_i W_j}{(d_0 + d_{ij})^{\gamma}}, \tag{2}$$

其中， W_i 和 W_j 分别是单元 i 和 j 中的劳动者人数， d_{ij} 是两个单元之间以公里计的地理距离。距离偏移量固定为 $d_0 = 0。1$ 千米， γ 控制距离衰减效应。我们使用泊松伪最大似然（PPML）估计 α 和 γ，并以观测到的会面次数 M_{ij} 作为因变量。估计参数分别为 $\widehat{\alpha} = 0。06365$ 和 $\widehat{\gamma} = 1。35413$。

随后，我们使用与观测到的会面概率相同的潜在会面次数 PMM_{ij} 对预测会面次数进行标准化：



补充图9：**观测到的会面概率和时间平移零模型会面概率**。(a) 观测到的会面概率（MP）以及时间平移零模型100次实现的平均MP的分布。(b) 每个单元对的观测值与零分布会面概率之比，计算方法为将观测到的MP除以其在100次零模型实现中的平均值。虚线表示观测到的MP与零模型MP相等（比值为 = 1），粉色线表示比值的中位数。

$$MP_{ij}^{gravity} = \frac{\widehat{\mu}_{ij}^{gravity}}{PMM_{ij}} = 0.06365 \frac{W_i W_j}{PMM_{ij}(0.1 + d_{ij})^{1.35413}}. \tag{3}$$

3 回归分析和稳健性检验

3.1 MP与交易结果之间的基准关联

为考察工人的相遇机会与后续企业间交易之间的关联，我们使用 2019 年测得的会面概率和 2016 年至 2022 年的交易记录估计逻辑回归模型。令 t 表示 2019 年之后的年份数，因此 $t = 1$ 表示 2020 年；令 $Y_{ij,\ t}$ 表示年份 t单元对（i，j）的二元交易结果。

我们使用三个结果变量：(1) 连边存在；(2) 链接形成；(3) 交易关系消失。(1) 当某个单元对在 t年至少有一笔买卖双方交易时，连边存在取值为1。(2) 当交易联系在此前观测到的任何年份均不存在、而后首次出现时，链接形成取值为1；例如，$t = 1$ 年的关系形成意味着2016—2019年没有观测到交易，而2020年发生了交易。(3) 对于在2019年之后的所有前置年份中持续存在的交易联系，在其首次消失的年份，交易关系消失取值为1。例如，$t = 1$ 年的关系解体意味着2019年观测到的交易联系在2020年不存在。我们使用具有二项分布和逻辑斯蒂连接函数的固定效应逻辑斯蒂模型对这些结果进行建模，并在R中使用fixest程序包的feglm函数进行估计。

主要解释变量是会面概率的对数 $\ln(MP_{ij})$，它反映了来自两个不同企业单元的劳动者之间的接触机会

在流动性观测期内。

我们估计以下模型设定：

$$\operatorname{logit}\left[\Pr\left(Y_{ij,t}=1\right)\right]=\beta\ln(\mathrm{MP}_{ij})+\mathbf{X}_{ij}'\boldsymbol{\gamma}+\delta_i+\delta_j. \tag{4}$$

该模型包括单元固定效应 δ_i 和 δ_j 以及一组配对层面控制变量。我们首先控制网格质心之间的地理距离（距离）。企业并非随机分布于城市空间之中：从事相似活动、需要频繁协调或具有潜在合作可能性的企业，可能会内生地选址于彼此邻近的位置。在这种情况下，较高的会面概率可能部分反映了企业向邻近地点进行的空间排序，从而混淆其与交易结果之间的关联。

对于企业单元 i 和 j，我们将行业相似度指数（行业重叠度，industryoverlap）定义为其行业份额向量的内积，即 $\text{IndustrySimilarity}_{ij}=\sum_{k=1}^{K}s_{ik}s_{jk}$。该指数的取值范围为0至1，可以解释为分别从单元 i 和 j 中独立抽取的两家企业属于同一行业的概率。当两个单元不共享任何行业时，该指数等于0；当两个单元完全由同一行业的企业构成时，该指数达到1。因此，数值越高表示其行业构成的重叠程度越高。

最后，我们利用2021年《经济普查商业活动调查》控制劳动力构成和当地经济结构的差异。具体而言，我们纳入两个企业单元在总就业人数（d_emp）、企业密度（d_est）和劳动力性别构成（d_maleshare）方面的差异。这些控制变量考虑了这样一种可能性：经济和人口特征相似的单元对可能既有重叠的工人流动，也有更高的交易联系发生概率。结合单元固定效应，配对层面控制变量可调整已测量的空间、行业和社会经济差异。交易结果和配对层面协变量的描述性统计见补充表3。所有方差膨胀因子（VIF）均低于2，表明解释变量之间不存在显著的多重共线性。

回归结果见补充表4。

补充表3：描述性统计

变量	N	Mean	SD	Min	Max
$\ln(MP)$	76,106,953	-13.318	1.457	-13.816	-1.187
距离的自然对数	76,106,953	9.174	0.778	4.443	10.917
行业相似度（按企业加权）	76,106,953	0.101	0.133	0.000	1.000
行业相似度（按员工加权）	76,106,953	0.095	0.185	0.000	1.000
d_emp	66,349,440	1.430	1.093	0.000	11.543
d_est	66,349,440	1.041	0.787	0.000	7.591
d_男性占比	66,337,921	0.116	0.088	0.000	0.918

补充表4：2017–2022 年会面概率与交易结果之间的基准关联

Year	2017	2018	2019	2020	2021	2022
面板*A*：连边存在						
ln(MP)	0.0202***	0.0206***	0.0190***	0.0200***	0.0199***	0.0203***
	(0.0023)	(0.0023)	(0.0022)	(0.0022)	(0.0022)	(0.0021)
行业相似度	2.6381***	2.7350***	2.7590***	2.7482***	2.7815***	2.8314***
	(0.1828)	(0.1734)	(0.1703)	(0.1642)	(0.1640)	(0.1653)
距离的自然对数	-0.3481***	-0.3468***	-0.3548***	-0.3521***	-0.3508***	-0.3433***
	(0.0114)	(0.0112)	(0.0109)	(0.0108)	(0.0108)	(0.0109)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
单元固定效应（i 和 j）	Yes	Yes	Yes	Yes	Yes	Yes
观测值	25,105,039	25,776,838	26,130,069	26,409,622	26,054,146	25,737,824
伪 R^2	0.3137	0.3165	0.3163	0.3183	0.3183	0.3191
*B*面板：链接形成						
ln(MP)	0.0116**	0.0183***	0.0140***	0.0272***	0.0113**	0.0151***
	(0.0050)	(0.0048)	(0.0051)	(0.0055)	(0.0052)	(0.0052)
行业相似度	3.1580***	3.1238***	3.0327***	2.7347***	3.2137***	3.2694***
	(0.2172)	(0.1876)	(0.2147)	(0.1973)	(0.2289)	(0.2317)
距离的自然对数	-0.3863***	-0.3235***	-0.3602***	-0.3284***	-0.3423***	-0.3165***
	(0.0172)	(0.0175)	(0.0177)	(0.0185)	(0.0192)	(0.0195)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
单元固定效应（i 和 j）	Yes	Yes	Yes	Yes	Yes	Yes
观测值	9,142,301	8,304,440	8,285,680	7,892,753	6,956,230	7,658,881
伪 R^2	0.1656	0.1637	0.1584	0.1606	0.1541	0.1642
面板*C*：交易关系消失						
ln(MP)	-0.0002	-0.0046	0.0038	-0.0017	0.0014	-0.0073
	(0.0059)	(0.0055)	(0.0053)	(0.0050)	(0.0049)	(0.0047)
行业相似度	-0.0643	0.0202	0.3749	-0.0452	0.1731	0.4844**
	(0.2965)	(0.2783)	(0.2868)	(0.2611)	(0.2628)	(0.2457)
距离的自然对数	0.0029	0.0032	-0.0024	0.0009	0.0221	-0.0137
	(0.0175)	(0.0176)	(0.0163)	(0.0150)	(0.0151)	(0.0153)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
单元固定效应（i 和 j）	Yes	Yes	Yes	Yes	Yes	Yes
观测值	161,808	174,936	184,636	196,207	200,815	203,487
伪 R^2	0.1806	0.1790	0.1606	0.1562	0.1699	0.1671

注：所有面板均报告 2017–2022 年单元格对层面的固定效应逻辑回归模型。面板A表示所示年份是否存在交易联系。B面板表示在此前没有交易记录的单元格对中，所示年份是否存在交易联系。面板C表示前一年活跃的企业对关系中是否至少有一项在该单元格对内解除。标准误按端点单元格进行双向聚类 i 和 j，并在括号中报告。 $^{*}p<0。10$， $^{**}p<0。05$， $^{***}p<0。01$。

3.2 对企业数量控制变量的稳健性

包含更多企业的单元格对拥有更多潜在企业组合，因此更可能同时出现工人接触和交易关系。为控制这一点，我们纳入了两个单元中企业数量乘积的十分位组指示变量。补充表5显示，三个年份的会面概率系数都非常稳定，介于0.0199和0.0202之间。

补充表5：对企业数量控制变量的稳健性

	t=1	t=2	t=3
ln(MP_{ij})	0.0200***	0.0199***	0.0202***
	(0.0022)	(0.0022)	(0.0021)
ln(距离ij)	-0.3510***	-0.3495***	-0.3419***
	(0.0108)	(0.0108)	(0.0109)
行业相似度	2.7731***	2.7958***	2.8318***
	(0.1532)	(0.1540)	(0.1551)
观测值	26,409,622	26,054,146	25,737,824
企业数量-产品十分位控制变量	Yes	Yes	Yes
单元 i 和 j 固定效应	Yes	Yes	Yes
配对层面控制变量	Yes	Yes	Yes
双向聚类标准误	Yes	Yes	Yes

注：二元结果是指在所示年份企业单元 i 和 j 之间是否存在连边。根据两个单元中企业数量乘积，将单元对划分为十分位数。所有设定均纳入十分位组指示变量，并省略最低十分位作为参照类别。所有其他基准控制变量均已纳入。所有设定均采用带有 i 和 j 单元固定效应的固定效应逻辑回归模型。标准误按企业单元 i 和 j 进行双向聚类，并列于括号内。$^{*}p<0$。10，$^{**}p<0$。05，$^{***}p<0$。01。

3.3 对替代MP度量的稳健性

替代MP度量（补充说明 2.3）会降低以下情形中偶遇的权重：（1）拥挤的单元格—时间单元；（2）人流量大的地点；（3）仅保留在至少被观测到两次的地点发生的共同在场；（4）排除每位劳动者自己工作场所单元格内的停留。我们在基准回归模型（2020年连边存在）中，使用各项替代度量分别替换观测到的 MP_{ij}，并比较所得估计结果。

在全部四种替代设定中，会面概率的估计系数仍为正且具有统计显著性，但其大小因度量方式而异（补充表6）。这些结果表明，正向关联对不同的相遇定义具有稳健性，但估计值的大小取决于如何为偶遇加权以及纳入哪些地点。

3.4 使用时间平移零假设MP进行零假设检验

在确立观测到的会面概率与连边存在和链接形成之间均存在正向关联后，我们接下来考察了仅凭共享空间活动模式是否能够再现这些关联

补充表6：替代会面概率度量的稳健性

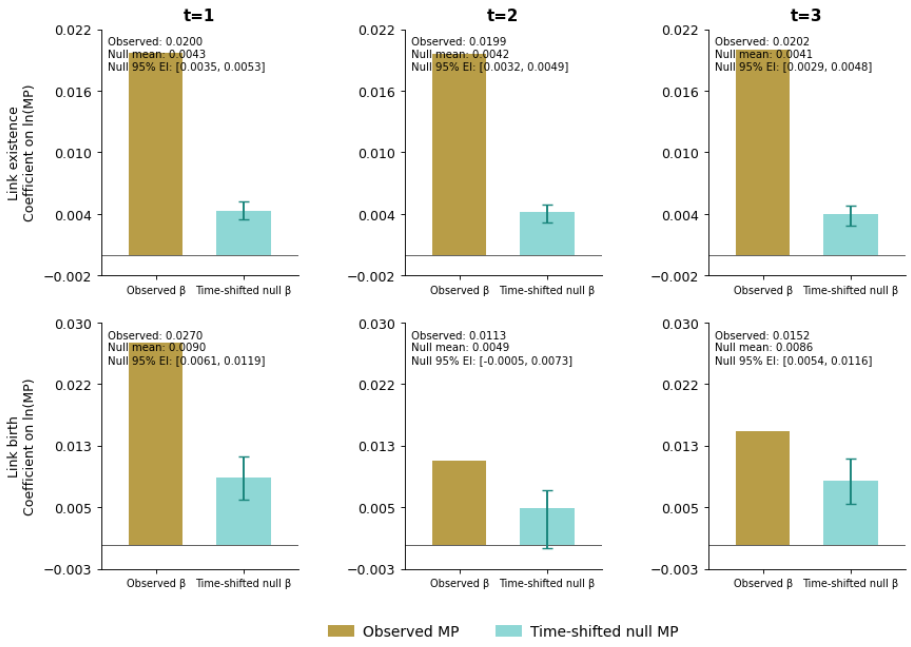
	观测到的会面概率	拥挤度加权 MP	受欢迎度加权 MP	重复会面概率	排除企业单元的会面概率
ln(MP_{ij})	0.0200***	0.0335***	0.1978***	0.0203***	0.0046***
	(0.0022)	(0.0032)	(0.0137)	(0.0022)	(0.0017)
ln(距离$_{ij}$)	-0.3521***	-0.3459***	-0.3310***	-0.3511***	-0.3676***
	(0.0108)	(0.0108)	(0.0111)	(0.0108)	(0.0107)
行业相似度	2.7482***	2.7524***	2.7654***	2.7487***	2.7508***
	(0.1642)	(0.1644)	(0.1653)	(0.1637)	(0.1642)
观测值	26,409,622	26,409,622	26,409,622	26,409,622	26,409,622
麦克法登伪 R^2	0.3183	0.3183	0.3184	0.3183	0.3182
单元 i 和 j 固定效应	Yes	Yes	Yes	Yes	Yes
配对层面控制变量	Yes	Yes	Yes	Yes	Yes
双向聚类标准误	Yes	Yes	Yes	Yes	Yes

注：二元结果是 2020年 i 与 j 之间是否存在连边。每一列均报告一个固定效应逻辑回归模型，使用观测到的会面概率（MP）或四种替代 MP 度量之一：拥挤度加权 MP、受欢迎度加权 MP、重复会面概率，以及排除企业单元的会面概率。所有规格均纳入相同的基准控制变量，以及针对 i 和 j 的单元固定效应。标准误按企业单元 i 和 j 进行双向聚类，并在括号内报告。 $^{*}p<0$。10， $^{**}p<0$。05， $^{***}p<0$。01。

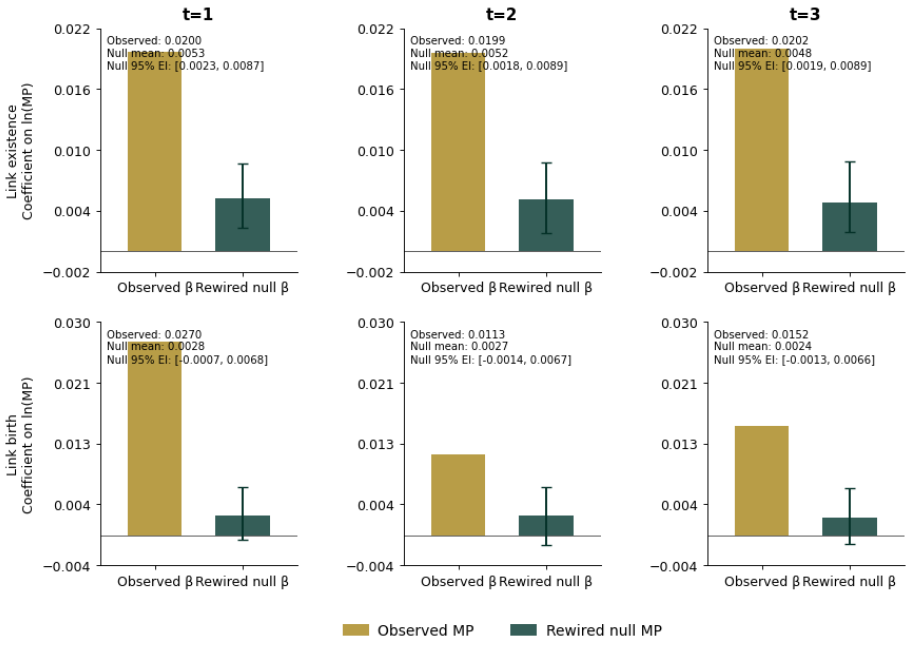
。我们应用了补充说明2.4中所述的时间平移零模型：在保留每位劳动者访问地点和日内移动模式的同时，将其日常移动轨迹重新分配到另一个日期。该过程打乱了劳动者之间原有的日历日期对齐。对于100次随机置换实现中的每一次，我们都重构了相遇网络，重新计算会面概率，并重新估计用于观测到的会面概率的相同回归设定。补充图10比较了观测系数与零模型估计值的均值及经验95%区间。在连边存在和链接形成两种情况下，观测系数在观测后的三个年份中均超过相应零模型区间的上界。对于连边存在，观测系数始终接近0.020，约为零模型均值0.0041–0.0043的4.7–4.9倍。对于链接形成，观测系数分别为0.0270、0.0113和0.0152，分别对应于零模型均值0.0090、0.0049和0.0086的3.0倍、2.3倍和1.8倍。这些下降表明，仅凭共享访问模式无法再现保留劳动者原始时间对齐时所观察到的关联。

3.5 保持度数的交易网络零模型

我们构建了一个基于网络的零模型，以评估 MP 与交易之间的关联是否可由交易网络中的一般连通性解释。对于每一年，我们将观测到的交易关系表示为一个网络，其中节点对应企业单元，边表示该年两个企业单元之间至少存在一笔交易关系。我们通过保持度数的边交换对每年的交易网络进行随机化。在每次交换中，选取两条现有边（a， b）和（c， d），并将其替换为（a， d）和（c， b），前提是拟议的新边不会形成自环或重复边。该过程重复 100 次，以随机改变交易伙伴的身份，同时保持每个节点的度数和网络中的边总数不变。我们为每个网络生成 100 个独立的重连实现。对于每个重连实现，我们重建了相应的二元交易



补充图 10：**观测到的会面概率与时间平移零假设会面概率系数**。每个面板比较观测到的会面概率与时间平移零假设会面概率的估计系数。各行分别对应连边存在模型和链接形成模型。零模型柱状图表示 100 次零模型实现的平均系数，误差条表示经验 95% 区间。



补充图1：**观测到的交易网络与重连后的交易网络中的会面概率系数**。每个面板比较观测到的会面概率与重连后的零模型会面概率的估计系数。各行分别对应链接存在模型和链接形成模型。零模型柱状图显示100次重连实现的平均系数，误差条表示经验95%区间。

结果，并重新估计了主要分析中使用的相同回归设定。补充图1显示了估计系数。对于2020年的链接存在模型，观测到的会面概率系数（0.020）约为交易网络重连下平均会面概率系数（0.0054）的3.8倍。链接形成模型也呈现出类似模式。交易网络重连下较小的会面概率估计值表明，基于流动性的相遇与企业会与哪些对象进行交易有关，而不仅仅是识别出交易频率更高的企业。

3.6 对替代交易历史窗口的稳健性

为评估估计的关联是否取决于近期交易历史的定义方式，我们使用一至四年交易历史窗口构建替代的新生与解散结果。这些设定改变了在识别转变之前所需的既往交易历史数量，从而使我们能够检验结果是否对特定的窗口定义敏感。

对于链接形成，窗口要求此前连续一至四年没有交易。当此前未活跃的企业—单元对

在目标年份开始交易时，即发生新生事件。对于关系解体，窗口涵盖目标年份之前具有连续一至四年交易历史的关系。当配对在目标年份停止交易时，即发生关系解体事件。由于只能观测到2016年及之后的交易历史，较长窗口仅适用于较晚的目标年份。例如，四年窗口最早可用于构建2020年的结果。因此，这些替代设定为以下问题提供了稳健性检验：估计的会面概率与交易关系转变之间的关系是否取决于此前未活跃期或此前交易历史的长度（补充图12）。

在不同的交易历史窗口定义下，会面概率与链接形成之间的关联始终为正，且具有统计显著性。这一模式在一至四年的未活跃期窗口中均保持稳定，表明主要的关系新生结果并不受此前未活跃期具体定义的影响。相比之下，交易关系消失的估计结果在不同年份和窗口定义下稳定性要差得多，且通常不具有统计显著性。这种较弱且一致性较低的模式表明，现有交易关系的终止可能更大程度上取决于其他因素。

3.7 MP与交易关系形成关联的时间变化

为评估会面概率与链接形成之间的关联是否因年份而异，我们合并了 2017 年至 2022 年的单元格对交易观测。如果某项交易联系在该年首次出现，则因变量取值为 1。我们估计以下合并逻辑回归模型：

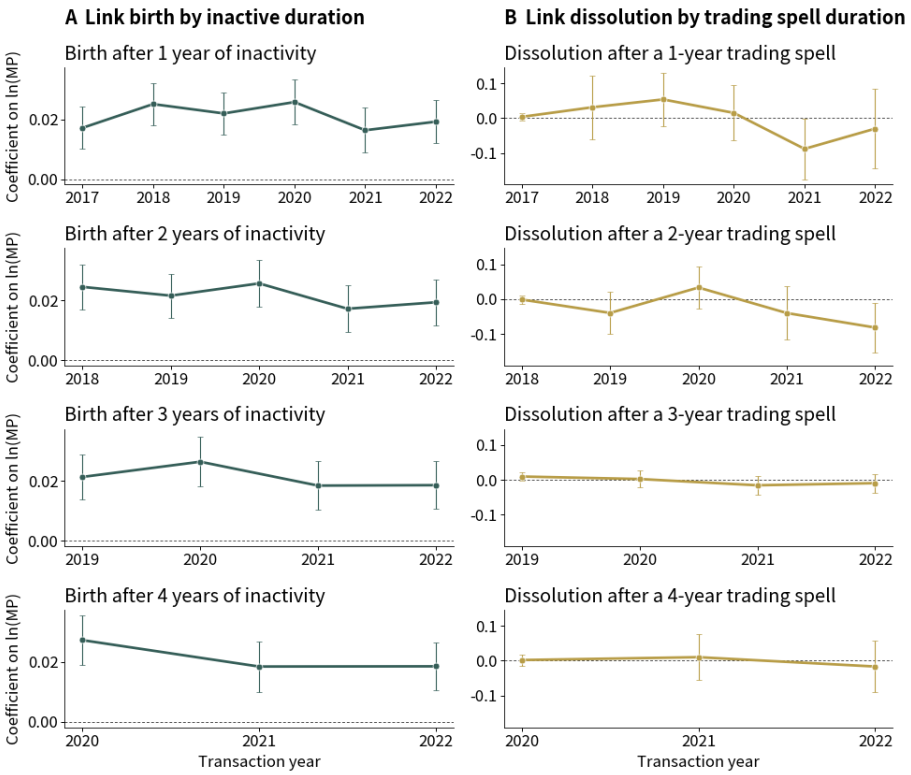
$$\operatorname{logit}\left[P(Y_{ij,t}^{\text{birth}}=1)\right]=\sum_{s=2017}^{2022}\beta_s\left[\ln(\text{MP}_{ij})\times\mathbb{1}(t=s)\right]+\mathbf{X}_{ij}'\boldsymbol{\gamma}+\delta_i+\delta_j+\lambda_t, \quad (5)$$

其中， i 和 j 表示构成一个单元对的两个单元。会面概率$_{ij}$ 根据 2019 年的流动性数据测量，并在各结果年份保持不变。系数β_s 表示会面概率与链接形成之间的逐年关联。模型纳入基准单元格对控制变量$\mathbf{X}_{ij}$、单元固定效应 δ_i和 δ_j，以及年份固定效应 λ_t。我们使用沃尔德检验，评估六个逐年系数是否全都相等这一联合原假设。

为检验这一关联在流动性测量前后两个时期是否存在差异，我们进一步估计了一个模型，分别针对2017–2019年和2020–2022年设定不同的MP系数：

$$\begin{aligned}\operatorname{logit}\left[P(Y_{ij,t}^{\text{birth}}=1)\right]=\;&\beta_{\text{pre}}\ln(\text{MP}_{ij})\times\mathbb{1}(t\le 2019)\\&+\beta_{\text{post}}\ln(\text{MP}_{ij})\times\mathbb{1}(t\ge 2020)\\&+\mathbf{X}_{ij}'\boldsymbol{\gamma}+\delta_i+\delta_j+\lambda_t.\end{aligned} \quad (6)$$

在此， β_{pre} 表示在流动性测量年份之前或当年观察到的结果所对应的关联，而 β_{post} 表示后续结果所对应的关联。我们使用关于 $H_0:\ \beta_{\text{pre}}=\beta_{\text{post}}$的沃尔德检验，检验这两个系数是否存在差异。在两个模型设定中，标准误均按单元 i 和 j进行双向聚类。



补充图12：**对不同交易历史窗口设定的稳健性**。ln(MP) 对一至四年交易历史窗口内链接形成和交易关系消失的估计系数。对于链接形成，各窗口要求此前连续一至四年内没有交易。对于交易关系消失，窗口涵盖在目标年份之前具有相应时长的连续交易期的关系。点表示系数估计值，误差条表示95%置信区间。

按年份估计的会面概率系数在2017–2022年间始终为正，但系数大小和统计显著性因年份而异（补充表7）。最大系数出现在2020年（$\beta = 0。0299$，$p < 0。001$），即流动性测量后的第一年。拒绝各年份系数相等的联合假设（$\chi^2(5) = 16。86$，$p = 0。0048$）。在基于时期的设定中，2020–2022年的系数（$\beta = 0。0205$，$p < 0。001$）大于2017–2019年的系数（$\beta = 0。0115$，$p < 0。001$），且差异具有统计显著性（$\chi^2(1) = 4。63$，$p = 0。0315$；补充表 8）。因此，在流动性测量之后观察到的交易形成与会面概率的关联更强。不过，与较早结果呈正向关联，表明会面概率也可能捕捉到持续存在的配对层面特征或关系。

补充表7：会面概率与链接形成之间的汇总关联：按年份划分的系数

	链接形成
ln(MP) × 2017	0.0097**
	(0.0046)
ln(MP) × 2018	0.0156***
	(0.0048)
ln(MP) × 2019	0.0091*
	(0.0050)
ln(MP) × 2020	0.0299***
	(0.0052)
ln(MP) × 2021	0.0066
	(0.0058)
ln(MP) × 2022	0.0227***
	(0.0059)
控制变量	Yes
年份固定效应	Yes
单元固定效应（i 和 j）	Yes
观测值	154,105,210
伪 R^2	0.2087
沃尔德检验	
$\chi^2(5)$	16.8634
p值	0.0048

注释：因变量表示在自 2016 年以来此前所有已观测年份均未出现交易联系的条件下，年份 t首次观测到交易联系的形成。模型纳入 ln(MP) 与年份指示变量之间的交互项、年份和单元固定效应，以及未报告其系数的基准控制变量。标准误按单元 i和 j进行双向聚类，并在括号内报告。伪 R^2 表示 McFadden 伪 R^2。沃尔德检验使用完整的双向聚类协方差矩阵，检验所有特定年份的 MP 系数均相等这一联合原假设（$\beta_{2017} = \cdots = \beta_{2022}$）。 $^{*}p < 0。10$,
$^{**}p < 0\ 05$ $^{***}p < 0\ 01$ 。,。 。

补充表 8：会面概率与链接形成之间的汇总关联：前后时期系数

	链接形成
ln(MP) ×前期（2017–2019年）	0.0115***
	（0.0029）
ln(MP) ×后期（2020–2022年）	0.0205***
	(0.0035)
控制变量	Yes
年份固定效应	Yes
单元固定效应（i 和 j）	Yes
观测值	154,105,210
伪 R^2	0.2087
沃尔德检验	
$\chi^2(1)$	4.6259
p-值	0.0315

注释：因变量表示在自 2016 年以来此前所有已观测年份均未出现交易联系的条件下，年份 t首次观测到交易联系的形成。模型纳入 ln(MP) 与前期和后期指标之间的交互项、年份和单元固定效应，以及未报告其系数的基准配对层面控制变量。标准误按单元 i和 j进行双向聚类，并在括号内报告。伪 R^2 表示 McFadden 伪 R^2。沃尔德检验使用完整的双向聚类协方差矩阵，检验前期和后期的 MP 系数相等这一零假设（$\beta_{\text{pre}} = \beta_{\text{post}}$）。 $^{*}p < 0。10$, $^{**}p < 0。05$, $^{***}p < 0。01$。

3.8 距离异质性

为考察会面概率与连边存在之间关联的距离异质性，我们根据网格质心之间的地理距离，将单元格对划分为四分位组。设 Q_{ij}^{q} 当配对（i，j）属于第 q 个四分位组时取1，并以距离最近的四分位组（第一季度）为参照组。我们构建了一个固定效应逻辑回归模型，在会面概率对数与其他三个四分位组之间加入交互项：

$$\operatorname{logit}\left[\Pr\left(Y_{ij,t}=1\right)\right]=\beta_1 \ln(\mathrm{MP}_{ij})+\sum_{q=2}^{4}\beta_q\left[\ln(\mathrm{MP}_{ij})\times Q_{ij}^{q}\right]+\sum_{q=2}^{4}\theta_q Q_{ij}^{q}+\mathbf{X}_{ij}'\boldsymbol{\gamma}+\delta_i+\delta_j. \tag{7}$$

系数 β_1 是第一季度的对数-MP斜率；第 $q>1$ 个四分位组的斜率为 $\beta_1+\beta_q$。补充表9报告了连边存在的交互模型估计结果。沃尔德检验拒绝了四个距离四分位组的对数-MP斜率相等这一零假设（$\chi^2(3)=26.99$，$p<0.001$）。

补充表9：会面概率与连边存在之间关联的距离异质性

	(1) t=1	（2）t=2	（3）t=3
$\ln(MP_{ij})$	0.0539***	0.0536***	0.0537***
	(0.0028)	(0.0027)	(0.0027)
距离Q2	-0.4726***	-0.4796***	-0.4671***
	(0.0630)	(0.0630)	(0.0632)
距离 Q3	-0.7769***	-0.8078***	-0.8345***
	(0.1237)	(0.1234)	(0.1186)
距离 Q4	-0.7825***	-0.7040***	-0.7127***
	(0.1669)	(0.1559)	(0.1516)
$\ln(MP_{ij})$×距离Q2	-0.0215***	-0.0218***	-0.0219***
	(0.0049)	(0.0049)	(0.0050)
$\ln(MP_{ij})$×距离 Q3	-0.0294***	-0.0305***	-0.0354***
	(0.0090)	(0.0090)	(0.0087)
$\ln(MP_{ij})$×距离 Q4	0.0011	0.0062	0.0039
	(0.0119)	(0.0112)	(0.0109)
观测值	26,409,622	26,054,146	25,737,824
单元 i 和 j 固定效应	Yes	Yes	Yes
配对层面控制变量	Yes	Yes	Yes
调整后伪 R^2	0.301	0.301	0.302

注：二元结果是所示年份中企业单元 i 和 j 之间是否存在连边。距离四分位组根据两者之间的地理距离划分；Q1（距离最近的四分位组）为参照组。$\ln(\mathrm{MP}^{ij})$ 的系数是 Q1 的斜率，每个交互项系数表示与该斜率的差值。固定效应逻辑回归模型包含单元固定效应和配对层面控制变量。标准误按企业单元 i 和 j 进行双向聚类，并列于括号内。 $^{*}p<0.10$， $^{**}p<0.05$， $^{***}p<0.01$。

接下来，我们检验这种非单调模式是否取决于距离区间的定义，或距离分布两端的配对。

替代距离区间。 补充表10报告了每个替代距离区间内估计的对数-MP斜率。对于每种分组方式，斜率等于参考区间的系数加上相应的交互项系数。按五分位数和十分位划分时，模式相似：距离最近的配对关联更强，中距离配对关联较弱，而距离最远的配对关联再次增强。这一模式与简单的单调距离衰减解释不一致。

补充表10：对替代距离分箱定义的稳健性

距离分组	Bin	距离中位数（米）	系数	S.E.
五分位数	1	3,570	0.054***	(0.003)
五分位数	2	6,970	0.033***	(0.004)
五分位数	3	10,285	0.025***	(0.007)
五分位数	4	15,045	0.036***	(0.011)
五分位数	5	25,415	0.058***	(0.012)
十分位	1	2,380	0.061***	(0.003)
十分位	2	4,505	0.020***	(0.004)
十分位	3	6,120	0.033***	(0.005)
十分位	4	7,735	0.026***	(0.006)
十分位	5	9,350	0.030***	(0.008)
十分位	6	11,220	0.013	(0.011)
十分位	7	13,515	0.032**	(0.013)
十分位	8	16,915	0.030*	(0.017)
十分位	9	22,185	0.043***	(0.016)
十分位	10	29,495	0.088***	(0.019)

注： 各项报告的是连边存在的会面概率对数，即 $\ln(\mathrm{MP}_{ij})$，的箱内斜率。对于每种五分位数或十分位定义，固定效应逻辑回归模型将会面概率对数与距离箱指标进行交互；每个报告的斜率均为参照区间系数加上相应的交互项系数。所有模型均纳入基准配对层面控制变量和单元固定效应。标准误按单元进行双向聚类 i 和 j，并在括号中报告。 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$。

排除极端距离。 补充表11报告了在排除距离分布两端的观测值后，各距离四分位组的估计结果。仅剔除距离最长的观测值后，观测到的模式基本保持不变：排除距离排名前1%或前5%的观测值后，第一和第四四分位数仍大于中间四分位数。当两端观测值均被剔除时，第一四分位数的估计值变小，表明非常接近的配对促成了较大的短距离效应。剩余观测中距离最远的四分位组仍大于中间距离四分位组，因此非单调模式并未消失。

3.9 行业异质性

完整的回归结果见补充表12。

补充表11：剔除极端距离后的稳健性

样本限制	四分位数	距离中位数（米）	系数	S.E.
全样本	Q1	3,995	0.054***	(0.003)
全样本	Q2	8,160	0.032***	(0.004)
全样本	Q3	12,920	0.024***	(0.009)
全样本	Q4	23,715	0.055***	(0.012)
剔除前1%	Q1	3,995	0.054***	(0.003)
剔除前1%	Q2	8,075	0.033***	(0.004)
剔除前1%	Q3	12,750	0.026***	(0.008)
剔除前1%	Q4	23,205	0.052***	(0.011)
剔除最高5%	Q1	3,910	0.054***	(0.003)
剔除最高5%	Q2	7,820	0.033***	(0.004)
剔除最高5%	Q3	12,155	0.025***	(0.008)
剔除最高5%	Q4	20,995	0.051***	(0.011)
保留 P1–P99	Q1	4,165	0.021***	(0.003)
保留 P1–P99	Q2	8,160	0.026***	(0.004)
保留 P1–P99	Q3	12,835	0.017**	(0.008)
保留 P1–P99	Q4	23,290	0.046***	(0.011)
保留 P5–P95	Q1	4,675	0.016***	(0.003)
保留 P5–P95	Q2	8,330	0.021***	(0.005)
保留 P5–P95	Q3	12,580	0.014	(0.009)
保留 P5–P95	Q4	21,335	0.042***	(0.012)

注： 各项报告的是在每种样本限制下，连边存在的会面概率对数 $\ln(\mathrm{MP}_{ij})$ 的四分位组内斜率。在每个受限样本中，固定效应逻辑回归模型将会面概率对数与距离四分位组指标进行交互；各项斜率均为参考四分位组系数与相应交互项系数之和。所有模型均纳入基准配对层面控制变量和单元固定效应。标准误按单元 i 和 j 进行双向聚类，并在括号内报告。 $^{*}p < 0$。10, $^{**}p < 0\ 05$ $^{***}p < 0\ 01$ 。,。。

补充表 12：行业矩阵中展示的行业对回归结果（公司数量设定）。

Pair	$\hat{\beta}$	SE	95%置信区间	p	BH q	N_{pair}
农业 × 建筑业	**−0.0011**	0.0002	[−0.0015, −0.0007]	< 0.001	< 0.001	60,648
农业 × 制造业	**−0.0006**	0.0002	[−0.0010, −0.0002]	0.006	0.010	31,752
农业 × 批发	0.0003	0.0002	[−0.0001, 0.0007]	0.107	0.124	52,479
农业 × 零售业	**−0.0011**	0.0002	[−0.0015, −0.0007]	< 0.001	< 0.001	14,028
农业 × 运输	**−0.0006**	0.0003	[−0.0012, −0.0001]	0.014	0.021	4,515
农业 × 信息	**0.0008**	0.0002	[0.0003, 0.0012]	< 0.001	0.001	11,676
农业 × 金融	**0.0024**	0.0006	[0.0013, 0.0035]	< 0.001	< 0.001	1,575
农业 × 房地产	**−0.0006**	0.0002	[−0.0010, −0.0002]	0.002	0.004	23,205
农业 × 专业/技术	**0.0005**	0.0002	[0.0001, 0.0010]	0.019	0.027	16,653
农业 × 行政支持	0.0001	0.0002	[−0.0004, 0.0006]	0.607	0.627	5,985
农业 × 教育	**−0.0014**	0.0002	[−0.0018, −0.0010]	< 0.001	< 0.001	1,281
农业 × 健康	**−0.0010**	0.0002	[−0.0015, −0.0006]	< 0.001	< 0.001	8,379
农业 × 住宿/餐饮	**−0.0011**	0.0003	[−0.0016, −0.0006]	< 0.001	< 0.001	1,638
农业 × 其他服务	**−0.0006**	0.0002	[−0.0010, −0.0001]	0.013	0.019	6,342
公用事业 × 建筑业	0.0016	0.0010	[−0.0003, 0.0035]	0.097	0.116	54,872
公用事业 × 制造业	**0.0020**	0.0009	[0.0002, 0.0038]	0.033	0.043	28,728
公用事业 × 批发	**0.0026**	0.0008	[0.0009, 0.0042]	0.002	0.003	47,481
公用事业 × 零售业	0.0016	0.0010	[−0.0003, 0.0035]	0.104	0.123	12,692
公用事业 × 运输	0.0020	0.0010	[0.0000, 0.0039]	0.046	0.058	4,085
公用事业 × 信息	**0.0031**	0.0008	[0.0015, 0.0047]	< 0.001	< 0.001	10,564
公用事业 × 金融	**0.0043**	0.0007	[0.0029, 0.0056]	< 0.001	< 0.001	1,425
公用事业 × 房地产	**0.0020**	0.0009	[0.0002, 0.0038]	0.033	0.043	20,995
公用事业 × 专业/技术	**0.0028**	0.0008	[0.0012, 0.0044]	< 0.001	0.001	15,067
公用事业 × 行政支持	**0.0026**	0.0009	[0.0009, 0.0044]	0.003	0.005	5,415
公用事业 × 教育	0.0014	0.0010	[−0.0006, 0.0033]	0.161	0.179	1,159
公用事业 × 健康	0.0017	0.0010	[−0.0002, 0.0036]	0.088	0.108	7,581
公用事业 × 住宿/餐饮	0.0018	0.0010	[−0.0002, 0.0039]	0.081	0.100	1,482
公用事业 × 其他服务	**0.0021**	0.0009	[0.0002, 0.0039]	0.027	0.036	5,738

续见下页

续接上一页

Pair	$\hat{\beta}$	SE	95%置信区间	p	BH q	N_{pair}
建筑业 × 建筑业	**−0.0012**	0.0001	[−0.0013, −0.0010]	< 0.001	< 0.001	4,168,828
建筑业 × 制造业	**−0.0007**	0.0001	[−0.0008, −0.0005]	< 0.001	< 0.001	4,366,656
建筑业 × 批发	**0.0002**	0.0001	[0.0001, 0.0003]	< 0.001	< 0.001	7,217,112
建筑业 × 零售业	**−0.0012**	0.0001	[−0.0014, −0.0011]	< 0.001	< 0.001	1,929,184
建筑业 × 运输	**−0.0007**	0.0002	[−0.0011, −0.0004]	< 0.001	< 0.001	620,920
建筑业 × 信息	**0.0007**	0.0001	[0.0004, 0.0009]	< 0.001	< 0.001	1,605,728
建筑业 × 金融	**0.0024**	0.0006	[0.0013, 0.0034]	< 0.001	< 0.001	216,600
建筑业 × 房地产	**−0.0007**	0.0001	[−0.0009, −0.0006]	< 0.001	< 0.001	3,191,240
建筑业 × 专业/技术	**0.0004**	0.0001	[0.0002, 0.0006]	< 0.001	< 0.001	2,290,184
建筑业 × 行政支持	0.0000	0.0002	[−0.0003, 0.0003]	0.863	0.868	823,080
建筑业 × 教育	**−0.0015**	0.0001	[−0.0017, −0.0014]	< 0.001	< 0.001	176,168
建筑业 × 健康	**−0.0012**	0.0001	[−0.0014, −0.0009]	< 0.001	< 0.001	1,152,312
建筑业 × 艺术/娱乐	**−0.0011**	0.0002	[−0.0016, −0.0006]	< 0.001	< 0.001	66,424
建筑业 × 住宿/餐饮	**−0.0011**	0.0002	[−0.0015, −0.0007]	< 0.001	< 0.001	225,264
建筑业 × 其他服务	**−0.0007**	0.0001	[−0.0009, −0.0004]	< 0.001	< 0.001	872,176
制造业 × 制造业	−0.0002	0.0001	[−0.0004, 0.0000]	0.074	0.092	1,142,316
制造业 × 批发	**0.0006**	0.0001	[0.0005, 0.0008]	< 0.001	< 0.001	3,778,488
制造业 × 零售业	**−0.0007**	0.0001	[−0.0009, −0.0006]	< 0.001	< 0.001	1,010,016
制造业 × 运输	−0.0002	0.0002	[−0.0006, 0.0001]	0.174	0.192	325,080
制造业 × 信息	**0.0011**	0.0001	[0.0009, 0.0014]	< 0.001	< 0.001	840,672
制造业 × 金融	**0.0028**	0.0005	[0.0017, 0.0038]	< 0.001	< 0.001	113,400
制造业 × 房地产	**−0.0002**	0.0001	[−0.0004, −0.0001]	0.002	0.004	1,670,760
制造业 × 专业/技术	**0.0009**	0.0001	[0.0007, 0.0011]	< 0.001	< 0.001	1,199,016
制造业 × 行政支持	**0.0005**	0.0002	[0.0002, 0.0008]	< 0.001	0.002	430,920
制造业 × 教育	**−0.0010**	0.0001	[−0.0012, −0.0008]	< 0.001	< 0.001	92,232
制造业 × 健康	**−0.0007**	0.0001	[−0.0009, −0.0004]	< 0.001	< 0.001	603,288
制造业 × 艺术/娱乐	**−0.0006**	0.0002	[−0.0010, −0.0001]	0.021	0.029	34,776
制造业 × 住宿/餐饮	**−0.0006**	0.0002	[−0.0010, −0.0002]	0.004	0.006	117,936
制造业 × 其他服务	−0.0002	0.0001	[−0.0004, 0.0001]	0.232	0.249	456,624

续于下一页

续自上一页

Pair	$\hat{\beta}$	SE	95%置信区间	p	BH q	N_{pair}
批发 × 批发	**0。0014**	0。0001	[0。0012, 0.0016]	< 0.001	< 0.001	3,121,251
批发 × 零售业	**0。0002**	0。0001	[0.0000, 0。0003]	0.009	0.014	1,669,332
批发 × 运输	**0.0006**	0.0002	[0。0003, 0.0009]	< 0.001	< 0.001	537,285
批发 × 信息	**0.0019**	0.0001	[0。0016, 0.0021]	< 0.001	< 0.001	1,389,444
批发 × 金融	**0。0032**	0。0005	[0。0023, 0.0041]	< 0.001	< 0.001	187,425
批发 × 房地产	**0。0006**	0。0001	[0.0005, 0。0008]	< 0。001	< 0。001	2,761,395
批发 × 专业/技术	**0.0017**	0.0001	[0。0015, 0.0019]	< 0.001	< 0.001	1,981,707
批发 × 行政支持	**0.0013**	0.0001	[0。0011, 0。0016]	< 0。001	< 0。001	712,215
批发 × 教育	−0.0001	0.0001	[−0。0003, 0。0000]	0.043	0.056	152,439
批发 × 健康	0.0002	0.0001	[0。0000, 0。0005]	0.068	0.086	997,101
批发 × 艺术/娱乐	0.0003	0.0002	[−0.0001, 0。0008]	0.130	0.147	57,477
批发 × 住宿/餐饮	0.0003	0.0002	[0。0000, 0。0006]	0.095	0.115	194,922
批发 × 其他服务	**0.0007**	0.0001	[0.0005, 0。0009]	< 0。001	< 0。001	754,698
零售业 × 零售业	**−0.0013**	0.0001	[−0。0015, −0.0011]	< 0.001	< 0.001	222,778
零售业 × 运输	**−0.0008**	0.0002	[−0。0012, −0.0004]	< 0.001	< 0.001	143,620
零售业 × 信息	**0。0006**	0。0001	[0.0004, 0.0009]	< 0.001	< 0.001	371,408
零售业 × 金融	**0。0024**	0。0006	[0。0013, 0。0035]	< 0。001	< 0。001	50,100
零售业 × 房地产	**−0.0008**	0.0001	[−0.0009, −0.0006]	< 0.001	< 0.001	738,140
零售业 × 专业/技术	**0.0004**	0.0001	[0。0002, 0.0006]	< 0.001	< 0.001	529,724
零售业 × 行政支持	0.0000	0.0002	[−0。0003, 0。0003]	0.940	0.940	190,380
零售业 × 教育	**−0。0016**	0。0001	[−0。0018, −0.0014]	< 0.001	< 0.001	40,748
零售业 × 健康	**−0。0012**	0。0001	[−0.0015, −0.0009]	< 0.001	< 0.001	266,532
零售业 × 艺术/娱乐	**−0。0011**	0。0002	[−0。0016, −0.0006]	< 0.001	< 0.001	15,364
零售业 × 住宿/餐饮	**−0。0012**	0。0002	[−0。0016, −0.0007]	< 0.001	< 0.001	52,104
零售业 × 其他服务	**−0.0007**	0.0001	[−0。0010, −0.0004]	< 0.001	< 0.001	201,736
运输 × 运输	−0.0005	0.0003	[−0。0011, 0。0002]	0.159	0.178	23,005
运输 × 信息	**0.0011**	0.0002	[0.0007, 0。0015]	< 0。001	< 0。001	119,540
运输 × 金融	**0.0027**	0.0006	[0。0016, 0。0038]	< 0。001	< 0。001	16,125
运输 × 房地产	−0.0003	0.0002	[−0.0006, 0。0001]	0.113	0.130	237,575

下页继续

接上一页

Pair	$\hat{\beta}$	SE	95%置信区间	p	BH q	N_{pair}
运输 × 专业/技术	**0.0008**	0.0002	[0.0005, 0.0012]	< 0.001	< 0.001	170,495
运输 × 行政支持	**0.0005**	0.0002	[0.0000, 0.0009]	0.038	0.049	61,275
运输 × 教育	**−0.0011**	0.0002	[−0.0015, −0.0007]	< 0.001	< 0.001	13,115
运输 × 健康	**−0.0007**	0.0002	[−0.0011, −0.0003]	< 0.001	0.001	85,785
运输 × 艺术/娱乐	**−0.0006**	0.0003	[−0.0012, −0.0001]	0.026	0.036	4,945
运输 × 住宿/餐饮	**−0.0007**	0.0003	[−0.0012, −0.0001]	0.014	0.021	16,770
运输 × 其他服务	−0.0002	0.0002	[−0.0007, 0.0002]	0.289	0.305	64,930
信息 × 信息	**0.0019**	0.0002	[0.0016, 0.0023]	< 0.001	< 0.001	154,290
信息 × 金融	**0.0033**	0.0004	[0.0025, 0.0041]	< 0.001	< 0.001	41,700
信息 × 房地产	**0.0011**	0.0001	[0.0009, 0.0014]	< 0.001	< 0.001	614,380
信息 × 专业/技术	**0.0020**	0.0001	[0.0017, 0.0022]	< 0.001	< 0.001	440,908
信息 × 行政支持	**0.0016**	0.0002	[0.0013, 0.0019]	< 0.001	< 0.001	158,460
信息 × 教育	**0.0004**	0.0001	[0.0001, 0.0006]	0.003	0.005	33,916
信息 × 健康	**0.0007**	0.0002	[0.0004, 0.0010]	< 0.001	< 0.001	221,844
信息 × 艺术/娱乐	**0.0007**	0.0002	[0.0003, 0.0011]	< 0.001	< 0.001	12,788
信息 × 住宿/餐饮	**0.0008**	0.0002	[0.0004, 0.0012]	< 0.001	< 0.001	43,368
信息 × 其他服务	**0.0012**	0.0002	[0.0009, 0.0015]	< 0.001	< 0.001	167,912
金融 × 金融	**0.0039**	0.0008	[0.0023, 0.0056]	< 0.001	< 0.001	2,775
金融 × 房地产	**0.0027**	0.0005	[0.0017, 0.0038]	< 0.001	< 0.001	82,875
金融 × 专业/技术	**0.0033**	0.0005	[0.0024, 0.0042]	< 0.001	< 0.001	59,475
金融 × 行政支持	**0.0032**	0.0005	[0.0022, 0.0042]	< 0.001	< 0.001	21,375
金融 × 教育	**0.0022**	0.0006	[0.0011, 0.0033]	< 0.001	< 0.001	4,575
金融 × 健康	**0.0024**	0.0006	[0.0013, 0.0035]	< 0.001	< 0.001	29,925
金融 × 艺术/娱乐	**0.0025**	0.0006	[0.0014, 0.0037]	< 0.001	< 0.001	1,725
金融 × 住宿/餐饮	**0.0025**	0.0006	[0.0013, 0.0037]	< 0.001	< 0.001	5,850
金融 × 其他服务	**0.0029**	0.0006	[0.0018, 0.0040]	< 0.001	< 0.001	22,650
房地产 × 房地产	**−0.0003**	0.0001	[−0.0005, 0.0000]	0.017	0.024	609,960
房地产 × 专业/技术	**0.0009**	0.0001	[0.0006, 0.0011]	< 0.001	< 0.001	876,265
房地产 × 行政支持	**0.0005**	0.0002	[0.0002, 0.0008]	0.002	0.003	314,925

续见下页

接上页

Pair	$\hat{\beta}$	SE	95%置信区间	p	BH q	N_{pair}
房地产 × 教育	**−0.0010**	0.0001	[−0.0012, −0.0009]	< 0.001	< 0.001	67,405
房地产 × 健康	**−0.0007**	0.0001	[−0.0009, −0.0004]	< 0.001	< 0.001	440,895
房地产 × 艺术/娱乐	**−0.0006**	0.0002	[−0.0010, −0.0001]	0.014	0.021	25,415
房地产 × 住宿/餐饮	**−0.0006**	0.0002	[−0.0010, −0.0002]	0.005	0.007	86,190
房地产 × 其他服务	−0.0002	0.0001	[−0.0004, 0.0001]	0.200	0.218	333,710
专业/技术 × 专业/技术	**0.0018**	0.0002	[0.0014, 0.0021]	< 0.001	< 0.001	314,028
专业/技术 × 行政支持	**0.0015**	0.0002	[0.0012, 0.0018]	< 0.001	< 0.001	226,005
专业/技术 × 教育	0.0001	0.0001	[−0.0001, 0.0003]	0.352	0.368	48,373
专业/技术 × 健康	**0.0004**	0.0001	[0.0001, 0.0007]	0.003	0.005	316,407
专业/技术 × 艺术/娱乐	**0.0005**	0.0002	[0.0001, 0.0009]	0.023	0.031	18,239
专业/技术 × 住宿/餐饮	**0.0005**	0.0002	[0.0001, 0.0009]	0.007	0.011	61,854
专业/技术 × 其他服务	**0.0009**	0.0002	[0.0006, 0.0012]	< 0.001	< 0.001	239,486
行政支持 × 行政支持	**0.0011**	0.0003	[0.0006, 0.0016]	< 0.001	< 0.001	40,470
行政支持 × 教育	−0.0003	0.0002	[−0.0006, 0.0000]	0.073	0.091	17,385
行政支持 × 健康	0.0001	0.0002	[−0.0003, 0.0004]	0.739	0.748	113,715
行政支持 × 艺术/娱乐	0.0001	0.0002	[−0.0004, 0.0006]	0.689	0.702	6,555
行政支持 × 住宿/餐饮	0.0001	0.0002	[−0.0003, 0.0006]	0.569	0.591	22,230
行政支持 × 其他服务	**0.0006**	0.0002	[0.0002, 0.0009]	0.003	0.005	86,070
教育 × 教育	**−0.0019**	0.0001	[−0.0021, −0.0016]	< 0.001	< 0.001	1,830
教育 × 健康	**−0.0015**	0.0001	[−0.0018, −0.0012]	< 0.001	< 0.001	24,339
教育 × 艺术/娱乐	**−0.0014**	0.0003	[−0.0019, −0.0009]	< 0.001	< 0.001	1,403
教育 × 住宿/餐饮	**−0.0014**	0.0002	[−0.0019, −0.0009]	< 0.001	< 0.001	4,758
教育 × 其他服务	**−0.0010**	0.0002	[−0.0013, −0.0007]	< 0.001	< 0.001	18,422
健康 × 健康	**−0.0011**	0.0003	[−0.0016, −0.0006]	< 0.001	< 0.001	79,401
健康 × 艺术/娱乐	**−0.0010**	0.0003	[−0.0016, −0.0005]	< 0.001	< 0.001	9,177
健康 × 住宿/餐饮	**−0.0011**	0.0002	[−0.0016, −0.0006]	< 0.001	< 0.001	31,122
健康 × 其他服务	**−0.0006**	0.0002	[−0.0010, −0.0003]	< 0.001	< 0.001	120,498
艺术/娱乐 × 住宿/餐饮	**−0.0009**	0.0003	[−0.0015, −0.0002]	0.009	0.013	1,794
艺术/娱乐 × 其他服务	−0.0005	0.0003	[−0.0010, 0.0001]	0.103	0.123	6,946

续见下页

接上一页

Pair	$\hat{\beta}$	SE	95%置信区间	p	BH q	N_{pair}
住宿/餐饮 × 住宿/餐饮	**−0。0009** 0。0004		[−0.0017, −0.0001]	0.032	0.043	3,003
住宿/餐饮 × 其他服务	**−0.0005** 0.0002		[−0。0010, −0.0001]	0.028	0.037	23,556
其他服务 × 其他服务	−0。0001 0。0003		[−0.0006, 0。0004]	0.686	0.702	45,451

注： 模型观测总数：66,276,421。粗体系数表示 BH $q < 0$。05。BH 值直接取自完整的160项检验族，未重新计算。95%置信区间为未经调整的正态区间。系数、标准误和置信区间上下限保留四位小数； p 和 BH q 值保留三位小数，低于0.001的值显示为 < 0。001。缩写：Agric. = 农业、林业、渔业和狩猎业；Util. = 公用事业；Constr. = 建筑业；Manuf. = 制造业；Whsl. = 批发贸易；Retail = 零售贸易；Transp. = 运输和仓储业；Info. = 信息；Finance = 金融与保险；Real est. = 房地产、租赁与出租服务；Prof./tech. = 专业、科学与技术服务；Admin. supp. =行政及支持和废物管理服务业；Educ. = 教育服务；Health = 医疗保健与社会援助；Arts/rec. = 艺术、娱乐与休闲；Accom./food = 住宿与餐饮服务；Other serv. = 其他服务；Public admin. = 公共管理。

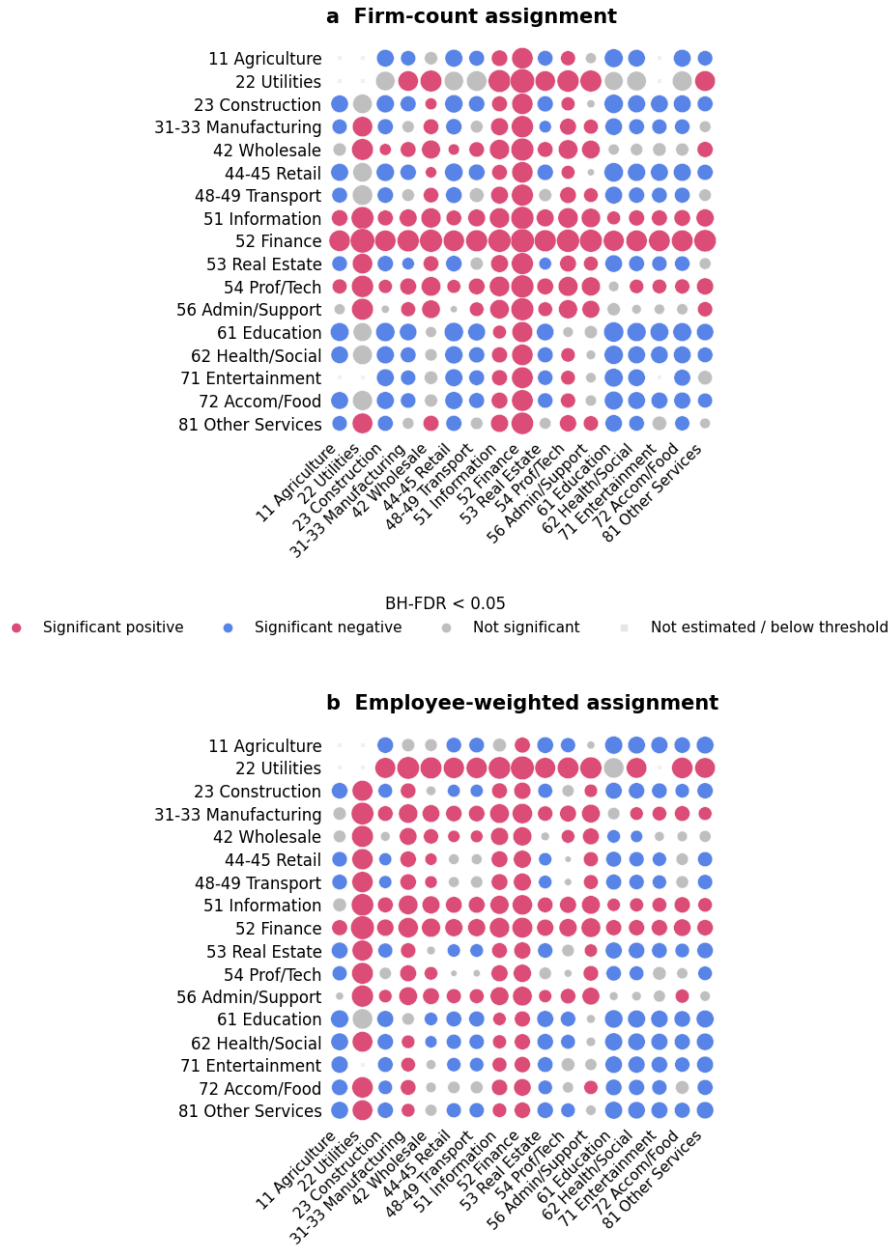
为评估行业异质性结果是否取决于每个单元格的主导行业定义方式，我们构建了两种替代性的单元格层面行业归属。在企业数量归属中，我们统计每个单元格内各NAICS两位数行业部门的企业数量，并将企业数量最多的部门定义为主导行业。在员工加权归属中，我们则汇总每个NAICS两位数行业部门内各企业的员工数量，并将总就业人数最多的部门定义为主导行业。

补充图13显示，在这两种定义下，行业异质性的总体模式高度相似。在两种设定下估计的行业对系数中，皮尔逊相关系数为 $r = 0$。824，斯皮尔曼秩相关系数为 $\rho = 0$。836。此外，83.8%的系数在两种归属方式下保持相同符号。这些结果表明，当使用就业人数而非企业数量定义主导行业时，行业对关联的幅度和相对排序总体上都得以保留。

3.10 跨功能行业组的行业异质性

行业异质性分析表明，会面概率与链接形成之间的关联因行业而异。接下来，我们将考察这种差异是否遵循更广泛的经济功能划分。传统的第一产业—第二产业—第三产业分类将许多不同的服务活动归为一类，因此，我们依据 Browning 和 Singelmann [3, 4] 的框架，将北美行业分类系统的行业门类整合为六个功能组。各行业门类的归属如下。

- *1.* 采掘业：农业、林业、渔业和狩猎业（北美行业分类系统 11）；以及采矿、采石和油气开采业（21）。
- *2.* 转型产业：公用事业（22）；建筑业（23）；以及制造业（31–33）。
- *3.* 分配性服务业：批发贸易（42）；零售贸易（44–45）；运输和仓储业（48–49）；以及信息业（51）。之所以将信息业纳入其中，是因为该行业在生产、传输和



补充图13：**替代行业分配下的行业对异质性**。面板**a** 使用企业数量，面板**b** 则使用总就业人数来确定每个企业单元的主导行业。圆圈大小表示系数大小的绝对值。粉色和蓝色圆圈分别表示经过 BH-FDR 校正后在 $q < 0.05$ 水平上显著为正和显著为负的系数，灰色圆圈则表示不显著。

传播信息，尽管某些活动也可能直接为生产提供投入。

- *4.* 生产者服务：金融与保险（52）；房地产、租赁与出租服务（53）；专业、科学与技术服务（54）；企业管理服务（55）；以及行政与支持服务（56）。
- *5.* 社会服务：教育服务（61）；医疗保健与社会援助（62）；以及公共管理（92）。
- *6.* 个人服务：艺术、娱乐与休闲（71）；住宿与餐饮服务（72）；以及其他服务（81）。

我们将 ln(MP$_{ij}$) 与各功能组配对的指示变量进行交互，同时保留与基准设定相同的控制变量和固定效应。该模型分别估计每种功能组组合下会面概率与交易形成之间的关系。结果见补充图14。每个圆点的大小代表估计系数的绝对值；红色和蓝色分别表示统计显著的正向和负向关联，灰色圆点则表示统计上不显著。正向关联主要集中在涉及分配性服务业的配对中。分配性服务业之间的配对，以及分配性服务业与生产者服务之间的配对，呈现出最大的正系数；此外，分配性服务业与其他若干服务组之间也存在显著的正向关联。涉及分配性服务业的配对表现出更强的关联，这与基于流动性的接触在流通商品、人员和信息的行业中发挥作用相符，但该分析并未确立因果机制。相比之下，涉及采掘业的大多数配对在统计上并不显著。

3.11 两跳流动性暴露与链接形成

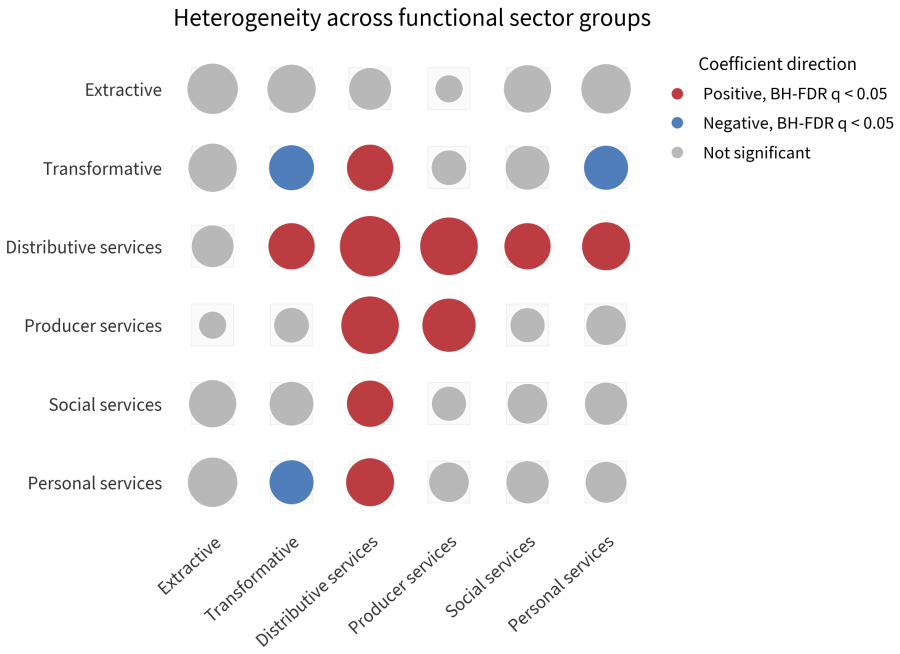
为考察间接流动网络连通性是否能够预测后续经济链接形成，我们考虑通过至少一个共享流动邻居 b_k 相连的企业单元配对 a 和 c，其中 $MP_{ab_k} > 0$ 和 $MP_{b_kc} > 0$ 均成立。我们将两跳连接的强度定义为

$$\mathrm{TwoHop}_{ac} = \sum_{k \in \mathcal{N}_{ac}} MP_{ab_k} \times MP_{b_kc}, \tag{8}$$

其中，N_{ac} 是 a 和 c 的共享流动邻居集合。该测度会随着共享邻居数量的增加，以及构成每条两跳路径的两条相遇概率链接的强度增大而上升。我们估计以下固定效应逻辑回归模型

$$\Pr(Y_{ac,t} = 1) = \Lambda\Big(\beta \ln(MP_{ac}) + \theta_1 Twohop_{ac} + \theta_2 I_{ac}^{\text{same}} + \theta_3 Twohop_{ac} I_{ac}^{\text{same}} + \mathbf{X}_{ac}^{\top}\boldsymbol{\gamma} + \delta_a + \delta_c\Big).$$

结果变量$Y_{ac,\ t}$表示在流动性测量后 t 年首次观测到 a 和 c 之间的新交易联系，其中 $t = 1,\ 2,\ 3$ 对应 2020年，



补充图14：**不同职能部门组之间的异质性**。每个点代表一个功能部门对在 $\ln(\mathrm{MP}_{ij})$ 上的估计系数。点的面积与系数大小的绝对值成正比。珊瑚色和青绿色的点分别表示置信区间不包含零的正系数和负系数，而灰色的点表示统计上不显著的估计值。

分别为 2021 年和 2022 年。当单元属于同一行业时，指示变量 I_{ac}^{same}取值为 1，否则取值为 0。系数 θ_1 反映跨行业配对中两跳暴露与后续交易形成之间的关联，而 $\theta_1 + \theta_3$ 则给出同行业配对中的相应关联。两项估计均以直接会面概率、$\mathbf{X}_{ac}$中的控制变量以及固定效应为条件。

我们通过加入两个配对层面的控制变量来检验结果对网络位置的稳健性。令D_i表示端点 i的度，令 $W_i = \sum_j MP_{ij}$表示其总会面概率。我们控制 $ln（D_a + D_c）$和 $ln（W_a + W_c）$。这些控制变量考虑了这样一种可能性：涉及中心性更高或活动更频繁的单元的配对，其两跳暴露在机械意义上可能更高。度反映与每个单元相关的流动性连接数量，而总会面概率反映这些连接的整体强度。这些控制变量有助于区分特定两跳连通性的关联与一般网络中心性和活动的关联。

表13的第（1）列和第（2）列报告了按 a 和 c进行双向聚类的标准误。第（2）列在基准设定的基础上加入了网络控制变量。第（3）列沿用第（2）列的估计结果，但报告二元聚类稳健标准误。

参考文献

[1] Ramaswamy Hariharan 和 Kentaro Toyama。Project lachesis：解析并建模位置历史记录。地理信息科学国际会议，第106–124页，2004年。[2] David Atkin、M. Keith Chen 和 Anton Popov。面对面互动的回报：硅谷的知识溢出。*NBER*工作论文系列，（w30147），2022 年。[3] Harley L. Browning 和 Joachim Singelmann。服务型社会的兴起：美国劳动力部门转型的人口统计学和社会学方面。技术报告 DLMA-21-48-73-45-1，人口研究中心，德克萨斯大学奥斯汀分校，弗吉尼亚州斯普林菲尔德，1975年6月。ERIC文献编号 ED131193。[4] Harley L Browning 和 Joachim Singelmann。美国劳动力的转型：行业与职业的交互作用。《政治与社会》，8(3-4)：481–509，1978年。

补充表13：两跳暴露与链接形成：对网络控制变量和二元依赖的稳健性

	(1) 基准	(2) 网络控制变量	(3) 二元稳健标准误
面板A：曾经（$t \geq 1$）			
跨行业	0.0103***	0.0100***	0.0100***
	(0.0022)	(0.0021)	(0.0022)
行业内	0.0003	-0.0006	-0.0006
	(0.0043)	(0.0044)	(0.0045)
观测值	18,402,728	18,402,728	18,402,728
伪 R^2	0.1889	0.1890	0.1890
B面：2020年（$t = 1$）			
跨行业	0.0171***	0.0168***	0.0168***
	(0.0034)	(0.0034)	(0.0035)
行业内	0.0008	0.0001	0.0001
	(0.0054)	(0.0055)	(0.0057)
观测值	11,991,291	11,991,291	11,991,291
伪 R^2	0.1355	0.1355	0.1355
面板C：2021（$t = 2$）			
跨行业	0.0095**	0.0090**	0.0090**
	(0.0041)	(0.0044)	(0.0043)
行业内	-0.0039	-0.0045	-0.0045
	(0.0094)	(0.0096)	(0.0096)
观测值	5,443,331	5,443,331	5,443,331
伪 R^2	0.0544	0.0544	0.0544
面板 D：2022 年（$t = 3$）			
跨行业	0.0057	0.0035	0.0035
	(0.0055)	(0.0068)	(0.0069)
行业内	-0.0022	-0.0039	-0.0039
	(0.0077)	(0.0080)	(0.0082)
观测值	5,822,859	5,822,859	5,822,859
伪 R^2	0.0676	0.0678	0.0678
配对层面控制变量	Yes	Yes	Yes
网络控制变量	No	Yes	Yes
单元格 a 和 c 固定效应	Yes	Yes	Yes
标准误	双向 a/c	双向 a/c	二元稳健

注：二元结果是企业单元 a 与 c之间的链接形成。跨行业斜率是两跳暴露的主效应；同业斜率是主效应与其同业指标交互项之和。第（2）列在基准设定中加入了网络控制变量。第（3）列使用与第（2）列相同的设定和样本，但报告二元聚类稳健的 actor-CR1 标准误。第（1）—（2）列报告按企业单元 a 与 c进行双向聚类的标准误。标准误列于括号内。所有设定均为固定效应逻辑回归模型，并包含方法论中所述的配对层面控制变量。 $^{*}p < 0。10$， $^{**}p < 0。05$， $^{***}p < 0。01$。