arXiv:2611.29576v1 [physics.soc-ph] 22 Nov 2026

政治倾向与预测市场参与：大额押注趋同，小额押注分化

AliHosseinzadeh[1],Ali Hosseiny[1],Reza Jafari[1伊1]
朗德黑兰埃温沙希德·贝赫什提大学物理系，邮编
1983969411 Ali Hosseinzadeh,
ali_hosseinzadeh@sbu.ac.irAli. Hosseiny:
al_hosseiny@sbu.ac.irReza Jafari:
g_jafari@sbu.ac.ir

摘要

预测市场为研究集体预期和信念提供了一种途径：人们表达对各种事件的看法，并通过下注支持自己的预测，承担判断失误的风险。因此，投资规模能够揭示个人行为和信念的更多维度。本研究采用复杂网络方法，基于 Polymarket 平台约 11.4 亿条交易记录，考察用户参与的结构；这些记录涵盖逾 310 万个钱包地址和 76.4 万个预测事件。为此，我们构建了用户与事件的二部网络，并分析投资集中度模式、主题之间的关联以及网络的社区结构。研究发现，金融参与的分布高度偏斜：1% 的事件占总交易量的 64%，而 1% 的钱包占总金融参与的 79.2%。我们还考察了投资规模和政治倾向如何影响用户参与模式。聚焦于2024年美国总统选举，我们比较了只投资特朗普获胜或只投资哈里斯获胜的用户，并按购买规模而非散户或机构身份对钱包分组。结果显示，在大额投资者中，两个政治群体在不同主题上的资本配置模式更为相似；相较之下，小额投资者的这些模式差异更大、波动也更明显。这一结果在六种不同的网络表征与比较方法下均成立，其中四项差异还具有统计学支持。

1 引言

预测市场将人们对未来事件的立场转化为市场价格。这为我们汇总信息、了解集体预期如何形成提供了一种方式。与民意调查或自由表达的观点不同，参与者在这里用资金支持自己的预测，并承担金融风险。将资金置于风险之中，是交易区别于陈述观点的关键。预测市场能够汇总信息，但参与者不必出于相同的原因进行交易，也不必遵循相同的策略。

Polymarket 是最大的去中心化预测平台之一，允许化名地址就从政治、宏观经济学到体育和加密货币等各种主题的合约进行交易。在这些平台上，执行交易不同于铸造、转换或结算合约 [3, 4]；我们的交易量指标统计的是已执行的交易。近期研究探讨了 Polymarket 上的套利、选举市场核算、市场微观结构、政治冲击、利润以及政治倾向的衡量方法 [5, 3, 6, 7, 8, 9]。其他研究则考察预测市场的准确性究竟反映了群体智慧，还是少数熟练账户的表现 [10, 11]。这些发现

促使研究者检验集中度和网络组织结构，但这些发现本身并不足以证明可以根据网络中心性推断技能、私人信息或政治身份。

在本研究中，我们采用复杂网络方法考察整个平台上的用户参与结构。基于约11.4亿条交易记录、超过310万个钱包地址和76.4万个预测事件，我们构建了一个二部网络，其中每个钱包都与其参与交易的事件相连，每条连接的权重由该钱包在相应事件中的总交易金额决定。这种表示方式使我们能够分析投资集中度模式、主题之间的关联，以及网络的社区结构。核心问题是：活动在各个主题和合约中集中于何处？哪些主题共享钱包？与相应的低购买群组相比，由候选人关联的高购买群组在后续主题配置上的相似程度是否更高？

我们最初的预期是，在政治事件中，政治倾向会影响人们理解事件的方式，进而影响他们的投资模式。然而，在金融市场中并不存在唯一的“制胜”策略：一个群体或公司研究某只股票后决定采取卖出策略，而另一个群体或公司则对同一只股票采取买入策略；正是这种策略差异使交易成为可能。在市场生态模型中，交易策略的回报取决于资本如何分配到相互竞争的策略上[12, 13]。然而，在大额投资规模下，我们观察到某种趋势形成；可以说，与小额投资相比，投资策略的多样性较低。基于此，本研究的主要假设是：大额投资者在选举市场之外的资本配置趋同，形成某种共识和趋势；而小额投资者的策略更加多样，行动也更多地受到自身政治偏见的影响。该假设与“大规模下高成本信念的趋同”这一观点一致：涉及金额越大，偏见的代价越高，因此配置会转向更有效、也更相似的模式。

为检验这一假设，我们量化事件层面和钱包层面的集中度，并按照 Clauset、Shalizi和 Newman [22] 的程序，评估货币分布和度分布的尾部是否符合幂律。对每个事件钱包数量的上尾进行幂律拟合后，得到的指数约为 2.4（补充表 S5）。随后，我们在货币权重和钱包归一化权重下分析完整的钱包—事件图，并将共享主题参与情况与度约束二部零模型进行比较。最后，总统选举获胜者市场中2024年9月的购买行为定义了与候选人关联的购买规模分层，并在10月期间测量其非焦点主题配置。本案例研究探讨：在不同候选人结果下，高购买额钱包与低购买额钱包的跨主题配置是否存在差异。本研究的贡献在于结合了三个层面：对整个平台的集中度和幂律分布进行分析；在完整的钱包–事件网络中开展社区检测；以及在真实选举事件背景下，对“大投资者趋同”与“小投资者分化和偏向”假设进行实证检验。

2 从交易到钱包–事件网络

2.1 交易记录与事件映射

在 Polymarket 上，一个事件可以包含多个市场。例如，哈里斯获胜和特朗普获胜的市场都属于同一个 2024年美国总统选举获胜者事件。在两个市场中均进行交易的钱包与该事件之间只有一条连接；在任一市场中的交易都会计入其权重。

已执行交易（成交记录）来自 TimeSeventeen 在 Hugging Face [7, 18, 19] 上提供的、经过清理且基于区块链数据的 Polymarket-v1 和 Polymarket-v2 数据集。v1 记录涵盖 2022年11月21日–2026年4月28日，v2 记录涵盖 2026年4月3日–2026年8月15日；在两代交易所的日历时间重叠期间，我们保留了两者的数据

。截至 2026年9月1日冻结的快照包含 1,143,528,694 条成交记录，总额约为 413.09 亿 USDC。补充材料第 S1 节和表 S1 列明了交易文件及其公开版本。

我们使用 Polymarket 的 Gamma API 中截至 2026年9月1日的开放和已结束事件快照，将每个市场的条件标识符与其所属的上级事件进行匹配。 [20]。[1] 标题和标签反映的是该快照中的信息，不一定与交易发生时可见的元数据一致。我们根据匹配的成交记录计算事件交易量，而未采用平台公布的累计交易量数据。

仅针对排行榜分析，我们于 2026年9月4日从 Polymarket 数据 API 获取了两个历来排名前 1,000 的名单：一个按报告交易量排名，另一个按报告利润排名[21]。分页期间排名可能发生变化，而且这些排名并未定义用于构建网络的钱包群体。

2.2 钱包参与度与事件主题

每笔已匹配的成交记录均计入一次事件交易量。计算钱包参与度时，我们会将其金额同时计入挂单方和吃单方：因此，钱包在已匹配事件中的总参与金额是其成交量的两倍。这衡量的是地址的总交易额，而非净投资额或持仓量。钱包集中度以全部 3,106,517 个观测到的地址为基础，并包括 52,022 笔无法匹配到事件的成交记录所产生的活动。这些成交记录不计入事件交易量或网络。

这个二部网络通过 255,704,359 个不同的钱包–事件对，将 3,106,514 个钱包与 764,378 个交易金额为正的事件相连，且未设置排名或活跃度截断值。钱包的度表示其交易过的事件数量；事件的度表示交易过该事件的钱包数量。每条连线的权重采用以下两种方式之一：该钱包在该事件中的交易总金额， w_{ue}；或其已映射交易中分配至该事件的比例， $w_{ue}/\sum_{e'} w_{ue'}$。两种情况下连线都相同，但采用第二种权重方式时，每个钱包的权重总和为一。

我们根据事件标题、网址中的简称以及平台标签，为每个事件分配 21 个主题之一。合约形式和结果的判定规则优先于按顺序匹配标签和标题；未能匹配的事件归入“其他”。例如，日内加密货币主题要求同时出现加密货币相关表述和短时间间隔提示。补充材料第 S2 节和表 S3–S4 列出了规则顺序、主题数量以及映射至七个大类的对应关系。分类器随代码一并提供；对分层抽样样本进行独立双人编码的工作仍待完成。

主题标签不参与社区检测。我们单独使用这些标签来构建主题网络：将钱包在同一主题下各事件中的交易量相加，可得到其对该主题的配置；而只记录钱包是否在那里进行过交易，则可以统计两个主题共有的钱包数量。

3 结果

3.1 市场与参与高度集中

按交易量排名，最大的 1% 事件占映射交易总量的 64.0%。钱包之间的交易分布更为集中：按交易金额排名，最大的 1% 钱包占钱包参与总量的 79.2%。事件交易量的基尼系数为 0.933，钱包参与度的基尼系数为 0.962（图 1）。

[1]我们分页了 `GET /events/keyset`，时间从 11:40:20 到 12:54:17 协调世界时。在 2,948,008 条市场记录中，有 1,380 条缺少 ed 一个条件 标识符。

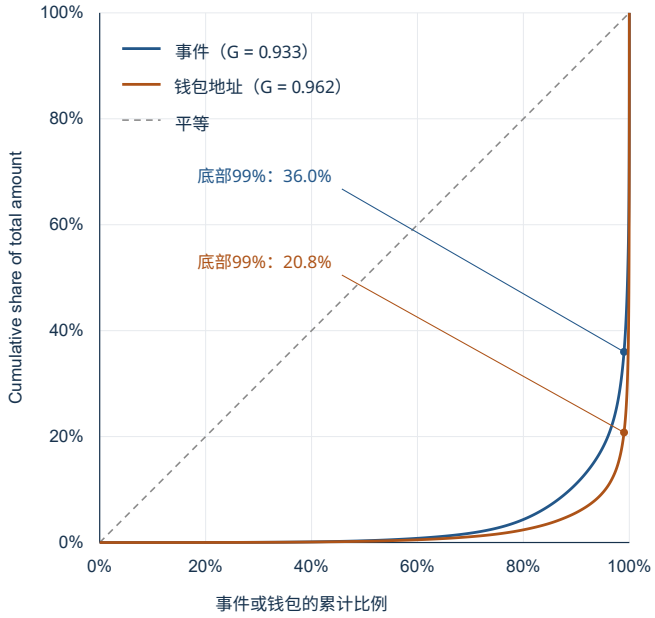


图1：事件和钱包之间的交易集中度。每条洛伦兹曲线都按照交易金额从小到大排列事件或钱包。横轴表示事件或钱包的累计份额；纵轴表示其总金额的累计份额。虚线对角线表示所有事件或钱包的金额均相等，且 G 为基尼系数。标记表示排名最低的99%所占的份额。事件交易量对每笔交易只计算一次；钱包参与度则同时计算挂单方和吃单方的交易金额。

3.2 帕累托标度律仅适用于部分情况，而非普遍适用

重新选定的头部事件和正盈亏排行榜尾部与纯帕累托拟合相兼容，指数约为 2.3，且没有任何经检验的替代模型更受支持（表1；图2）。交易量排行榜也符合绝对拟合优度标准，但在相同的选定观测值上，指数截断幂律更受支持。因此，绝对兼容性和相对模型偏好回答的是不同问题：不拒绝并不能证明数据是由幂律生成的。

表1：按排名筛选的主要货币尾部结果。截断值和指数均按照声明的CSN程序重新估计。数值 $p_{\mathrm{KS}} > 0.10$ 表示拟合的帕累托尾部未被拒绝；它并不代表幂律为真的概率。

分布	n_t	$x_{\min}$(USDC)	$\widehat{\alpha}$ （95%区间）	p_{KS}	相对模型结果
前1,000个事件量	375	10,496,710	2.317 (2.096–2.522)	0.733	无已检验的替代模型更受青睐
交易量排行榜	560	42,551,673	2.362 (2.235–2.685)	0.154	截断幂律倾向于 [a]
正盈亏排行榜	257	792,887	2.322 (2.068–2.486)	0.257	无经检验的替代模型倾向于

[a]在相同选定尾部上比较指数截断幂律与纯帕累托分布：对数似然比 -4.725，数据集内经 Holm 校正的 $p = 0.0063$。区间是 α 的自助法区间；它们并非共同指数检验。

这些结论以观测到的按排名选定的列表为条件。在顶级事件列表中重新估计截断值，会得到相容的尾部；而将其固定为列表最小值则不会。对于完整的事件量样本以及事件度和钱包度样本，所声明的 CSN程序拒绝其截断点搜索所选出的尾部。这并未检验整体分布

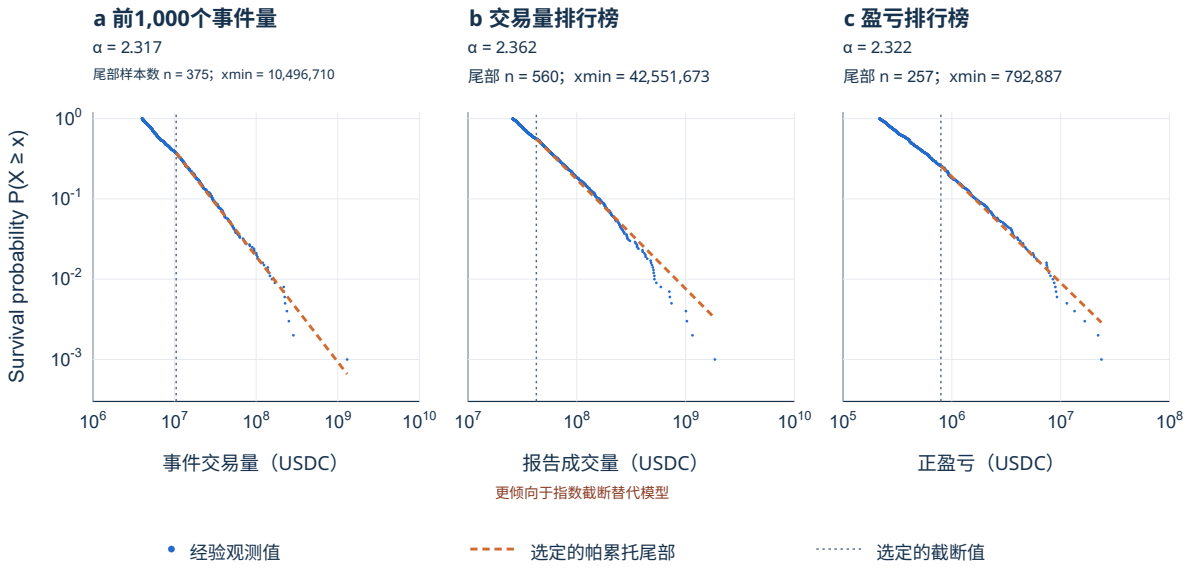


图 2：三个按排名选取的样本中的交易量和利润。 对于每个金额，蓝点表示达到或超过该数值的观测值比例。垂直虚线标示拟合尾部的下限（$x_{\min}$）；橙色虚线表示拟合的帕累托模型。**a**，交易量最大的 1,000 个事件。**b**，历史交易量排行榜上报告的交易量。**c**，历史利润排行榜上的正利润。

，也不能排除所有可能的更窄尾部。补充表 S5 汇总了八项主要拟合结果，并链接至全部 60 项拟合结果，其中包括截断值敏感性和模型比较。因此，我们不将完整网络描述为无标度 [17]。

3.3 参与稀疏且依权重而异的社区结构 e

只有一小部分可能的钱包–事件对建立了关联：二部密度为1.07685×10^{-4}。交易金额最大的100个钱包与参与事件数量最多的100个钱包之间仅有3个地址重合。按金额而非等额对钱包进行加权，也改变了主题总体占比（图3）。

我们通过优化加权二部模块度来识别社群。该指标将钱包和事件群组内的交易情况与根据其总活动量预期的金额进行比较[26, 27]。在分辨率 $\gamma = 1$下，按金额加权的网络产生了149个社群（$Q_b = 0.6729$；其中80.32%的权重位于社群内部），而在每个钱包贡献相等总权重的情况下，则产生了3,527个社群（$Q_b = 0.7814$；83.50%）。在图4中，每个社群的颜色根据其主要详细主题所属的广义类别确定。我们在社群识别后，通过按主题汇总该社群成员事件的全部交易量，并取总量最大的主题来确定该主题；无需达到多数。

3.4 主题覆盖面广，但不能仅用受欢迎程度来解释 e

图 5 中的主题网络回答了两个不同的问题。按数量加权的共同分配视图将 21 个细分类别归并为七个大类，并根据每个钱包的参与总金额赋予权重。所有 21 组大类配对的共同分配均为正值，其中最强的关联集中在体育、政治和加密货币周围。这是一张配置图，而非显著性网络。

第二种视角使用二元钱包–主题矩阵和拟合的二部配置模型（BiCM）。210 对细分主题都共享钱包，但只有 75 对的共享

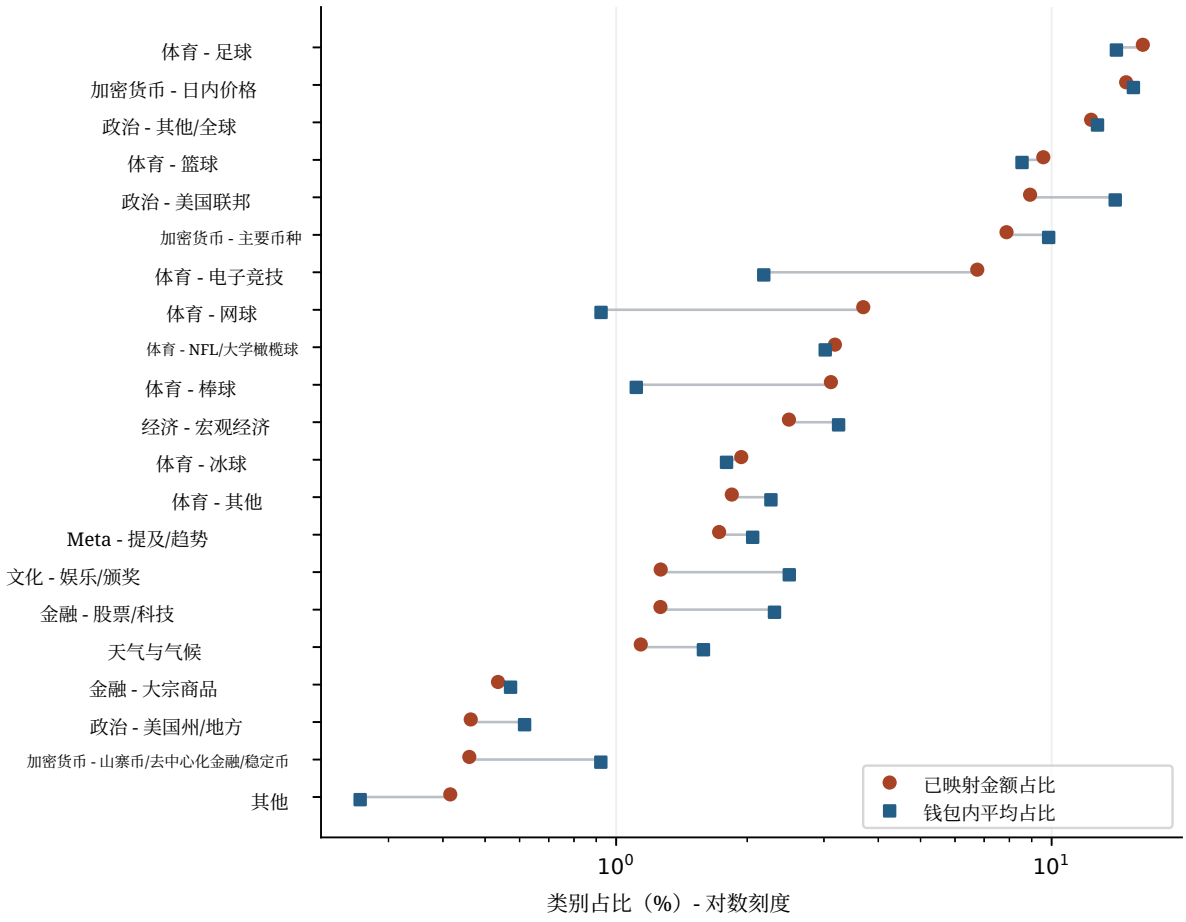


图3：交易金额或钱包获得相同权重时的主题占比。对于21个主题中的每一个，圆形表示所有钱包交易总金额中该主题所占的比例。方形表示先分别对每个钱包计算相同比例，再取平均值，从而使每个活跃钱包获得相同权重。若某主题的圆形位于其方形右侧，意味着该主题在总交易量中所占比例高于在平均钱包交易量中所占比例。横轴采用对数刻度。

钱包数量在对每个钱包的主题广度和每个主题的钱包数量进行期望值控制后，仍高于预期，随后进行了霍尔姆校正。另有 22 对在最近四个完整季度中，观测重叠量与预期重叠量之比均大于 1。这项持续性检验记录的是方向，而不是另外四项显著性检验。

一项探索性的体育/非体育分析发现，块内富集程度高于跨块富集程度。在金额投影中，相对于强度乘积基准，28个体育–体育配对中有24个、78个其他块内配对中有63个，以及104个跨块配对中有2个呈现正向提升。相应的 BiCM-Holm 计数分别为16、49和10。这一区分是在注意到该模式后进行检验的；其确切枚举结果属于事后诊断（补充材料第 S5.3 节和图 S1），并非针对预先设定划分的验证性证据。

3.5 高购买量候选队列的主题配置更加相似

总统选举案例研究将群组形成与后续分析区分开来。2024年9月，焦点总统选举获胜者事件中的吃单方 YES 购买用于界定仅哈里斯组和仅特朗普组。购买金额分位数是根据两种指定结果的全部 21,301 名买家计算的，其中包括同时购买两种结果的买家，然后再选出互斥队列。后5%指金额不超过 1.09414 USDC，前5%指金额至少为 2,787.73554 USDC，边界并列者也予以保留。这些

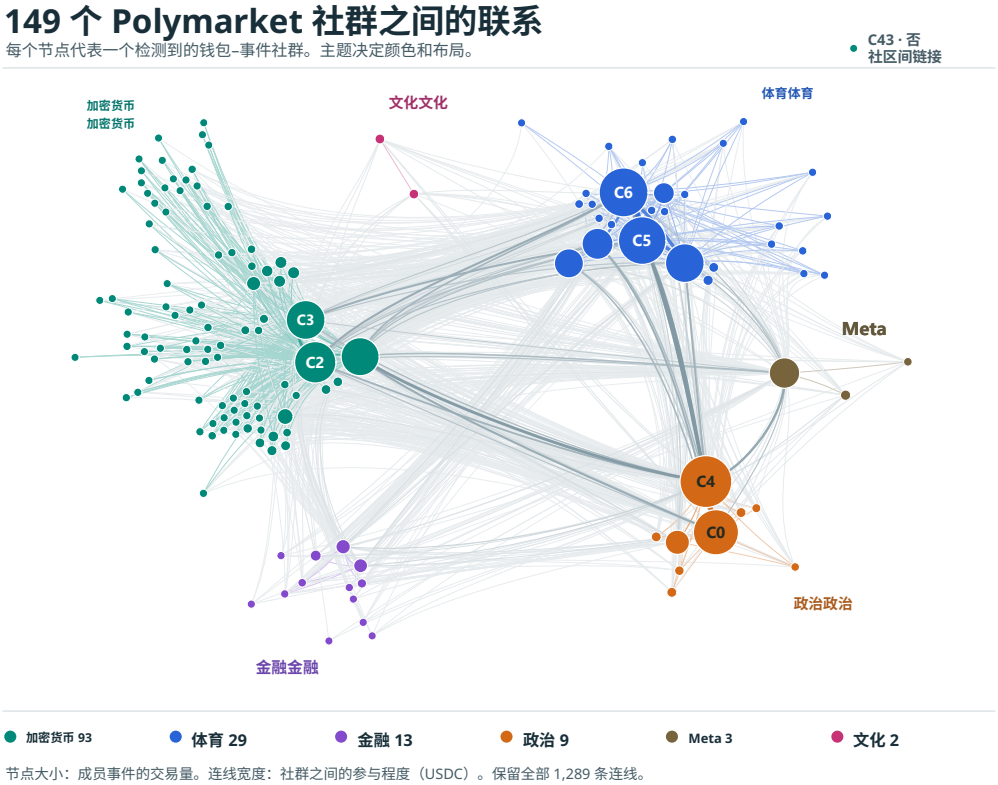


图4：149 个 Polymarket 社群之间的交易关系。每个节点代表一个由钱包和事件构成的社群。节点越大，表示该社群中事件的总交易量越高。完成社区检测后，根据每个节点中交易量最高的细分主题所属的大类为节点着色。显示时，同色节点彼此靠近排列。连线宽度随任一社群中的钱包在另一社群事件中的交易总金额增加而增大。图中显示全部 1,289 条社群间连线；C43 没有此类连线，因此单独绘制。

标签描述的是焦点市场中观察到的购买行为；它们并不衡量个人财富或政治归属。

符合条件的群组规模分别为：哈里斯底部（HB）504人、特朗普底部（TB）541人、哈里斯顶部（HT）334人，以及特朗普顶部（TT）251人。10月观察到的非焦点活动钱包数分别为360、308、168和135个。选择10月作为9月形成后、11月5日选举前的最后一个完整日历月；后续观察中排除焦点事件。这样可以将形成阶段与测量阶段分开，而无需使用不完整的11月时间窗口。比较仍以10月有活动为条件，各群组的活跃比例并不相同，也不能据此认定10月具有其他月份的代表性。

图 6 展示了各群组在相同边预算下最强的22条金额共分配边。所展示的边分别占HB、TB、HT和TT完整非对角边质量的82.0%、80.6%、92.9%和93.2%。底部网络分别需要37条和35条边才能覆盖90%的边质量，而每个顶部网络仅需17条。因此，高购买量网络的活动更集中于较少的一组主题关联。

更集中的顶部网络在不同候选队列之间的配置也更加相似。将每个完整的金额边矩阵归一化，使非对角线元素的总和为 1 后，哈里斯–特朗普顶部组的詹森–香农散度（JSD）低于底部组：分别为 0.243 和 0.457 比特。此比较涵盖全部 231 个主题对，而不仅是图中展示的边（补充表 S8）。

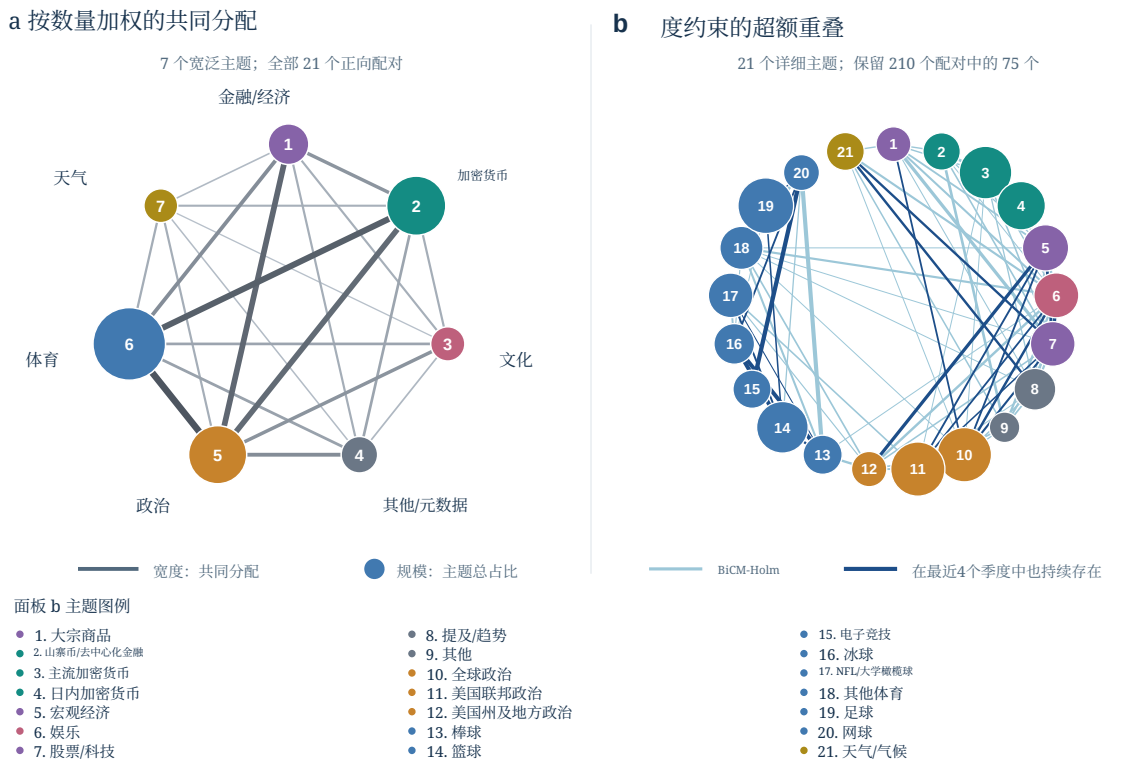


图5：钱包如何在各主题间分配交易，以及哪些主题共享的钱包数超出预期。a，七大主题类别间的交易分配。节点越大，表示总交易量中所占份额越大。当钱包将相当大比例的交易分配给两个类别时，它们之间的连线就越强；交易金额较大的钱包权重也越高。图中展示了全部 21 对类别。**b**，21 个详细主题间的共享钱包。节点越大，表示活跃钱包越多。各连线标示了 210 对主题中共享钱包数超出预期的 75 对；计算时已考虑每个钱包涉及的主题数以及每个主题对应的钱包数（经霍尔姆校正的 $p \leq 0.05$）。连线越粗，表示观测值与期望值之比越大。深色连线标示了 22 对主题，这些主题对在最近四个完整季度中，该比值均大于 1。

在全部六种归一化画像和边表示中，哈里斯–特朗普顶部组都比底部组更接近。金额画像 JSD 对底部组为 0.278 比特，对顶部组为 0.073 比特；流行率画像和二元边对应的底部组/顶部组数值分别为 0.061/0.013 和 0.186/0.074。在全部六种队列配对中，HT–TT 在五种表示中是最接近的一组；例外是金额边视图，其中 HB–TT 更接近。因此，高购买相似性并非仅限于美元加权链接，尽管各项货币比较的结果有所不同（补充表 S8 和 S10）。

队列内钱包自助法得出的逐点 95% 区间显示，金额画像、流行率、二元边以及（幅度较小的）等钱包边中，Bottom 减去 Top 的差异均为正；其余两个区间则包含零（补充表 S8）。另一项独立的固定向量随机化在各购买层级内重新分配候选人标签，同时保留观测到的钱包向量和活跃队列规模。对六项方向性对比进行霍尔姆校正后，等钱包画像、流行率画像、等钱包边和二元边的差异均得到支持（每项校正后的 $p = 0.0006$），但金额画像（$p = 0.3274$）和金额边（$p = 0.3357$）的差异未获支持（补充表 S9）。自助法不确定性涉及从观测到的群组中抽样；而标签随机化则检验在给定各购买层级的钱包向量后，候选人标签排序是否异乎寻常。

前五枢纽杰卡德相似度（枢纽集合交集除以并集）在金额视图和等钱包视图中，顶部配对为 0.667，底部组为 0.250；二元枢纽

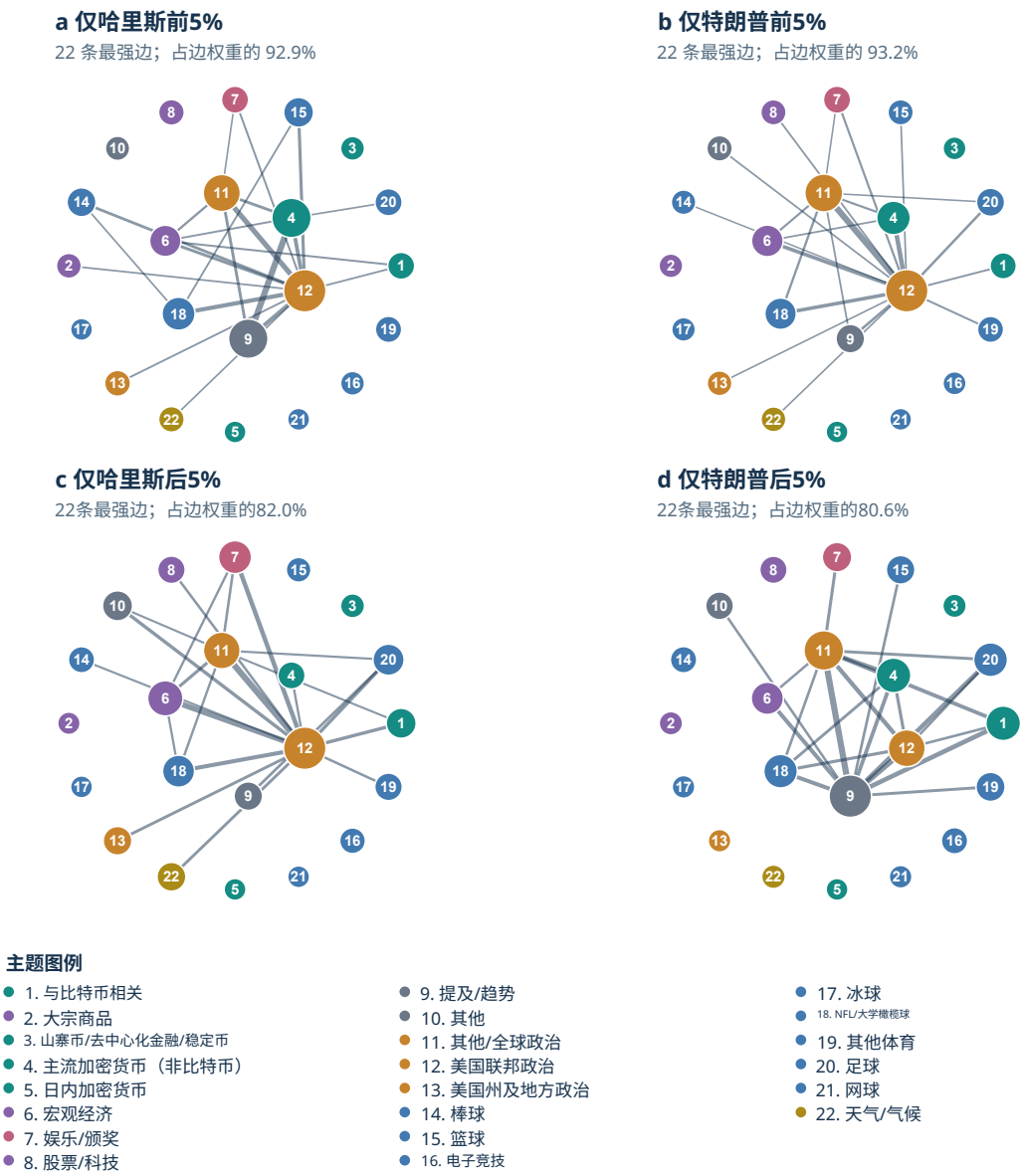


图6：10月在9月支持哈里斯或特朗普的钱包中的各主题交易。 各列区分了购买哈里斯「是」但未购买特朗普「是」的钱包（左侧），以及购买特朗普「是」但未购买哈里斯「是」的钱包（右侧）。各行显示按9月选举购买金额排名的前5%（a,b）和后5%（c,d）。节点代表主题；连线表示钱包在焦点选举事件以外的交易中同时涉及两个主题。节点大小反映某个主题所有连线的权重总和，线宽反映钱包同时向两个主题分配交易的程度。图中仅显示权重最高的22条连线；每个面板上方的百分比表示这些连线占连线权重总和的比例。各面板中的主题位置固定，但节点大小和线宽按面板分别缩放，不用于比较不同群体之间的美元金额。

相似度分别为 1.000 和 0.429。22 条最强金额边在顶部配对中的 Jaccard 相似度为 0.571，在底部组中为 0.333（补充材料第 S6.4 节）。

这些比较涉及群组聚合剖面和网络，而非同一群组内单个钱包之间的相似性。聚合配置相似，也不意味着主题节点社群相匹配：在金额、等权钱包和二元表示下，最佳主题划分的跨群组调整兰德指数都较低。按强度排名的主题核心置换诊断在四个网络中均显著，但最大的富集出现在 TB，而非顶部群组。这些比较揭示了共享剖面、枢纽和主导连接，而非共同的模块划分或钱包富人俱乐部。

4 讨论

在 Polymarket 上交易，并不只是表达观点。用户会为自己的预测投入资金，并承担金融风险。由此，他们展现出一种不同于无代价表达的信念。这种机制可视为一种“高代价信念”，因为用户不仅表达自己的看法，还将资金置于风险之中。不过，承担金融风险并不一定意味着这些用户掌握了更好的信息，或能做出更准确的预测。本研究利用约 11.4 亿条交易记录、超过 310 万个钱包地址和 76.4 万个预测事件，从三个角度分析 Polymarket 的参与结构：金融集中度、网络组织，以及与政治选择相关的参与模式。

我们的结果表明，Polymarket 上的参与高度不平等：事件排名前1%的事件占总交易量的64%，而钱包排名前1%的钱包占金融参与的79.2%。这种集中度也体现在基尼系数上：事件的基尼系数为0.933，钱包的基尼系数为0.962。对分布的分析显示，重尾行为以及与帕累托分布的兼容性仅限于数据的特定部分。在完整的钱包–事件网络中，总交易金额最高的100个钱包中，只有3个同时位列参与不同事件数量最多的100个钱包之中。

这些结果呈现出一幅有趣且有所不同的图景。在 YES 购买量排名前5%的钱包中，围绕焦点选举事件之外主题的参与配置模式，在仅特朗普组与仅哈里斯组之间相似度明显更高。相比之下，购买量处于后5%的钱包表现出更加多样且分歧更大的模式。在研究考察的全部六种表示中，包括金额和参与普及度剖面，以及基于边的表示，都观察到了这一结果。然而，经过多重检验校正的置换检验为其中四种对比提供了统计支持（最小调整后 p值为0.0006）。换言之，尽管两组做出了不同的选举相关选择，但在金融参与度较高的钱包中，它们在其他主题上的配置模式更加相似。相比之下，金融参与度较低的钱包表现出更加多样的模式。

与哈里斯和特朗普相关的群体，在选举相关金额较高时，其跨主题交易配置比金额较低时更为相似。这一模式与“高成本信念趋同”的观点一致：当投入的金额较少时，政治偏好可能为配置差异留下更大的空间。不过，这种比较是在群体之间进行的。高成本时的相似性并不意味着个体交易者意见一致或掌握了更多信息。我们也不能将低成本时更大的差异归因于政治偏见：钱包地址未必对应不同的个人，而一次选举相关购买也无法确定买家的政治归属或动机。

5 方法

5.1 集中度度量

对于正金额 $x_{(1)} \le \cdots \le x_{(N)}$，经验洛伦兹曲线将 (0，0) 与

$$\left(k/N,\ \sum_{i=1}^{k} x_{(i)} / \sum_{i=1}^{N} x_{(i)} \right), \quad k = 1, \ldots, N,$$

以及基尼系数为 $2\sum_i i x_{(i)} / (N \sum_i x_{(i)}) - (N+1)/N$，即曲线与完全平等对角线之间面积的两倍 [14, 15]。顶端份额采用最大的 $\lceil qN \rceil$ 个观测值，且不作帕累托假设。

5.2 幂律分析

对于连续值 $x \ge x_{\min}$，候选概率密度为

$$f(x) = (\alpha - 1) x_{\min}^{\alpha - 1} x^{-\alpha}, \qquad \widehat{\alpha} = 1 + \frac{n_t}{\sum_{x_i \ge x_{\min}} \log(x_i / x_{\min})}. \tag{1}$$

分析遵循 CSN流程：最大似然估计、通过最小化双侧柯尔莫哥洛夫–斯米尔诺夫距离选择截断值，以及在每个合成样本中重新拟合的半参数拟合优度模拟 [22, 16]。尾部至少包含 100 个观测值。三个主要的 $n = 1{,}000$ 个数据集使用穷举搜索，在 901 个可接受的截断值中进行选择；较大的连续数组则结合 500 个线性间隔和 500 个几何间隔的尾部规模目标。该网格是明确声明的计算近似，并非针对最大数组的穷举截断点搜索。整数度采用精确的离散赫尔维茨ζ似然。

每个指定分布均使用2,500次有效拟合优度模拟，但钱包度使用10,000次。合成样本将重抽样的经验主体与生成的帕累托尾部结合起来，使尾部数量能够按二项分布变化，同时保留原始样本量。除明确指定的固定截断值的前1000项敏感性分析外，截断值和指数均重新估计；在该敏感性分析中，仅重新拟合指数。报告的概率采用$(K_{\ge} + 1)/(B + 1)$表示，其中 $K_{\ge}$ 是模拟KS距离至少与观测到的 [23]一样大的次数。当$p_{\mathrm{KS}} \le 0.10$时，纯帕累托候选分布将被拒绝。对于58个连续候选分布，在同一选定尾部上，将拟合的帕累托分布与指数分布、对数正态分布、拉伸指数分布和指数截断幂律分布等备选分布进行比较。非嵌套比较使用有符号似然比 [24]；对于每个候选分布，霍尔姆校正涵盖四个备选分布 [25]。两个离散度候选分布接受绝对拟合检验，但不进行备选分布族比较。图中展示的三个指数区间则使用每个观测数据集的1,000个非参数自助重抽样样本，并重新选择截断值和重新拟合指数。逐点生存曲线图不参与估计。

5.3 二部模块度与社区检测

对于第 2 节中定义的任一种加权方式，令 w_{ue} 表示所选的边权重，$s_u = \sum_e w_{ue}$表示钱包强度，$s_e = \sum_u w_{ue}$ 表示事件强度。加权二部模块度为

$$Q_b = \frac{1}{m} \sum_{u,e} \left(w_{ue} - \gamma \frac{s_u s_e}{m} \right) \mathbf{1}\{g_u = g_e\}, \qquad m = \sum_{u,e} w_{ue}. \tag{2}$$

零模型项在期望意义下保留强度 [26, 27]。内存受限搜索在原始节点移动与对精确聚合商图进行有向Leiden优化之间交替进行，并针对所声明的随机种子和分辨率 [28, 29]重复执行。数值检验验证了配置不会改变目标函数。主题和季度标签仅在优化完成后附加。

5.4 主题共同分配与度约束重叠

设 x_{uc} 为钱包 u在主题 c中的金额， $v_u = \sum_c x_{uc}$ 以及 $p_{uc} = x_{uc}/v_u$。金额共分配矩阵和等钱包权重共分配矩阵为

$$W^A_{cd} = \frac{\sum_u v_u p_{uc} p_{ud}}{\sum_u v_u}, \qquad W^E_{cd} = \frac{1}{n}\sum_u p_{uc} p_{ud}, \quad c \neq d. \tag{3}$$

七类别面板按照固定映射汇总21类矩阵；不筛除任何边。二元共享钱包重叠 $V_{cd} = \sum_u A_{uc}A_{ud}$ 以具有独立概率的二部配置模型进行检验 $P(A_{uc} = 1) = x_u y_c/(1+x_u y_c)$，拟合乘数，使钱包度和主题度在期望上与观测值一致 [30]。对于210个主题对，每对进行19,999次边际共享数模拟，得到模拟重叠至少与观测值一样大的加一概率。霍尔姆校正涵盖这210对构成的检验族；校正后的 $p \leq 0.05$处保留边。逐点 Clopper–Pearson 界限用于量化蒙特卡洛误差，但并非同时置信区间 [31]。投影边表示共同参与，不代表交流或社会关系 [32, 33]。

5.5 选举队列形成与十月测量

2024年9月，在焦点总统选举获胜者事件中的买方 YES 购买记录，用于界定仅购买哈里斯结果而未购买特朗普结果的仅哈里斯钱包，以及模式相反的仅特朗普钱包。分位数参考池包含购买任一结果的全部21,301名买家：7,474个仅哈里斯钱包、9,361个仅特朗普钱包，以及4,466个两种结果均购买的钱包。截断值依据所列结果的 YES 购买总金额确定，并保留边界并列；随后才选出互斥队列。10月测量排除焦点事件，并采用原始分类体系以及经标题审核的21个事件比特币细分，最终形成22个互斥主题。

对于每个十月活跃钱包 u，令 x_{ui} 表示主题 i中的金额， $v_u = \sum_i x_{ui}$ 和 $p_{ui} = x_{ui}/v_u$。三个完整的非对角主题矩阵为

$$W^A_{ij} = \frac{\sum_u v_u p_{ui} p_{uj}}{\sum_u v_u}, \qquad W^E_{ij} = \frac{1}{n}\sum_u p_{ui} p_{uj}, \qquad W^B_{ij} = \frac{1}{n}\sum_u \mathbf{1}(x_{ui} > 0)\mathbf{1}(x_{uj} > 0). \tag{4}$$

相应的节点概况分别为汇总金额占比、钱包内平均占比和主题普及率。比较前，每个边向量都会归一化为总和为 1。

对于归一化向量 P和 Q，以比特计的詹森–香农散度为 $\frac{1}{2}D_{\mathrm{KL}}(P\|M) + \frac{1}{2}D_{\mathrm{KL}}(Q\|M)$，其中 $M = (P+Q)/2$ ，且对数以 2 为底 [34]。六视图相似度等于 1 减去三个概况和三个边散度的未加权平均值。图示在固定节点位置下，显示每个群组中最强的 22 条边；所有报告的散度均使用完整矩阵计算。边集中度以 $D_2^{\mathrm{edge}} = 1/\sum_e q_e^2$概括，即在具有相同集中度 [35]的情况下等权边的数量。

5.6 统计分析

选举队列底部减顶部的JSD不确定性根据 1,999 次群组内钱包自助法重复样本估计 [36]。每次抽样均从符合条件的九月队列钱包中有放回地抽取与原始数量相同的钱包，其中包括十月活动为零的向量。根据抽样得到的十月活动为正的行重建表示，因此后续活动计数可能有所不同。六个 2.5/97.5 百分位区间均为逐点区间，且未对相关视图进行调整。在回顾性敏感性分析中，通过 9,999 次固定向量随机化检验方向性对比 $\Delta = \mathrm{JSD}_{\mathrm{Bottom}} - \mathrm{JSD}_{\mathrm{Top}} > 0$： 在保持活跃群组规模不变的同时，于每个购买层级内独立重新分配候选人标签；每个重复样本中的六种表示均使用相同的分配结果。经加一校正的上尾概率采用 Holm 法对六个视图进行调整。该零模型以观测到的分层钱包向量多重集为条件，其中包括支出、广度和

配置集中度，同时假定候选人标签可交换且钱包行相互独立。仅为考察模块细分的敏感性，在分辨率为 1 的条件下，以 200 个随机种子重复进行加权 Louvain 优化 [37]。体育/非体育划分属于事后分析。补充表 S11 区分了重采样、零模型模拟和优化重复；第 S3–S7 节说明了相应的统计量、总体和设置。

数据可用性

交易记录来自 TimeSeventeen 在 Hugging Face 上公开的 Polymarket-v1 和 Polymarket-v2 数据集 [18, 19]；Qin 和 Yang 描述了 v1 [7]。我们使用了清理后的 Parquet 交易数据层（`daily_aligned` 和 `daily_aligned_multi`），其修订版本见补充表 S1。我们未直接从 Polygon 节点或 Polymarket 的订单簿交易历史中获取成交记录。

本研究从 Gamma API 收集了事件和市场元数据，并从 Data API 收集了排行榜观测值。排行榜请求（`GET /v1/leaderboard`）使用了`category=OVERALL` 和 `timePeriod=ALL`，并分别按 `VOL` 和 `PNL`排序。研究档案包含用于事件映射和分类的元数据字段、两个带日期的排名序列、图表的数值输入以及文件校验和。档案不会重新分发第三方成交文件或完整的钱包—事件边列表。带版本的代码和数据可在研究代码库中获取。

代码可用性

研究档案包含事件分类器、分析代码，以及用于根据数值输入重现图表的脚本。重建完整的钱包–事件网络需要原始成交文件和带日期的条件到事件的映射；汇总图表的输入不包含单条边或节点分配（补充信息第 S7 节）。

附加信息

本文附有补充信息。

参考文献

[1] Wolfers J, Zitzewitz E. 预测市场。《经济展望杂志》.2004;18(2):107–126.doi:10.1257/0895330041371321。[2] Rothschild DM, Sethi R. 交易策略与市场微观结构：来自预测市场的证据。《预测市场杂志》. 2016;10(1):1–29. doi:10.5750/jpm.v10i1.1179。[3] Tsang KP, Yang Z. 区块链预测市场剖析：2024年美国总统选举中的 Polymarket。2026。预印本, arXiv:2603.03136v3。doi:10.48550/arXiv.2603.03136。[4] Rahman N, Al-Chami J, Clark J. SoK：去中心化预测市场（DePM）的市场微观结构。在可信智能合约研讨会、金融密码学与数据安全会议上发表；2026。arXiv:2510.15612。doi:10.48550/arXiv.2510.15612。[5] Saguillo O, Ghafouri V, Kiffer L, Suarez-Tangil G. 揭开概率森林的奥秘：预测市场中的套利。载于：第7届金融科技进展会议（*AFT 2025*）。LIPIcs。2025;354:27:1–27:24。doi:10.4230/LIPIcs.AFT.2025.27。[6] Tsang KP, Yang Z. 政治冲击与预测市场中的价格发现：来自2024年美国总统选举的证据。2026。预印本, arXiv:2603.03152v4。doi:10.48550/arXiv.2603.03152。

[7] Qin B, Yang R. Polymarket-v1数据库。2026。预印本，arXiv:2606.04217v2。doi:10.48550/arXiv.2606.04217。[8] Akey P, Grégoire V, Harvie N, Martineau C. 谁在预测市场中获胜，谁又在亏损？来自 Polymarket 的证据。CEPR讨论论文 21615；2026。亦可见 SSRN 6443103。doi:10.2139/ssrn.6443103。[9] Chen H, Duan X, El Saddik A, Cai W. Web3博彩中的政治倾向：解读政治动机与盈利动机之间的相互作用。载于：第17届ACM网络科学会议论文集。2025。第96–105页。doi:10.1145/3717867.3717884。[10] Gómez-Cram R, Guo Y, Jensen TI, Kung H. 预测市场的准确性：群体智慧还是知情少数？2026。工作论文。doi:10.2139/ssrn.6617059。[11] Luong KL, Heesen G. 少数人的智慧：高技能交易者与预测市场准确性。2026。工作论文，SSRN 6758662。doi:10.2139/ssrn.6758662。[12] Farmer JD. 市场力量、生态与演化。《产业与公司变革》。2002;11(5):895–953。doi:10.1093/icc/11.5.895。[13] Scholl MP, Calinescu A, Farmer JD. 市场生态如何解释市场失灵。《美国国家科学院院刊》。2021;118(26):e2015574118。doi:10.1073/pnas.2015574118。[14] Lorenz MO. 财富集中度的测量方法。《美国统计协会出版物》。1905;9(70):209–219。doi:10.1080/15225437.1905.10503443。[15] Gastwirth JL. 洛伦兹曲线和基尼指数的估计。《经济学与统计学评论》。1972;54(3):306–316。doi:10.2307/1937992。[16] Newman MEJ. 幂律、帕累托分布与齐普夫定律。《当代物理学》。2005;46(5):323–351。doi:10.1080/00107510500052444。[17] Broido AD, Clauset A. 无标度网络并不常见。Nature Communications。2019;10:1017。doi:10.1038/s41467-019-08746-5。[18] TimeSeventeen. Polymarket-v1 [数据集]。Hugging Face。 https://huggingface.co/datasets/TimeSeventeen/Polymarket-v1/tree/5aa1b9d52316a8b2e789e81c8ae42c7ed532e8aa。特定版本的公开数据集；访问日期：2026年9月1日。[19] TimeSeventeen. Polymarket-v2 [数据集]。Hugging Face。 https://huggingface.co/datasets/TimeSeventeen/Polymarket-v2/tree/2a8dba6472f4c9ca558869e9cc1a65b9e1614264。特定版本的公开数据集；访问日期：2026年9月1日。[20] Polymarket. 列出事件（键集分页） [Gamma API文档]。https://docs.polymarket.com/api-reference/events/list-events-keyset-pagination。访问日期：2026年9月13日；元数据收集日期：2026年9月1日。[21] Polymarket. 获取交易者排行榜排名 [Data API v1文档]。https://docs.polymarket.com/api-reference/core/get-trader-leaderboard-rankings。访问日期：2026年9月13日；排行榜快照日期：2026年9月4日。[22] Clauset A, Shalizi CR, Newman MEJ. 经验数据中的幂律分布。《SIAM评论》。2009;51(4):661–703。doi:10.1137/070710111。[23] Phipson B, Smyth GK. 随机抽取置换时，置换 p值绝不应为零：计算精确的 p值。《统计学在遗传学与分子生物学中的应用》。2010;9(1):第39篇文章。doi:10.2202/1544-6115.1585。[24] Vuong QH. 用于模型选择和非嵌套假设的似然比检验。《计量经济学》。1989;57(2):307–333。doi:10.2307/1912557。[25] Holm S. 一种简单的序贯拒绝式多重检验程序。《斯堪的纳维亚统计学杂志》。1979;6(2):65–70。[26] Barber MJ. 二部网络中的模块度与社区检测。《物理评论E》。2007;76:066102。doi:10.1103/PhysRevE.76.066102。[27] Beckett SJ. 加权二部网络中改进的社区检测。《英国皇家学会开放科学》。2016;3(1):140536。doi:10.1098/rsos.140536。[28] Leicht EA, Newman MEJ. 有向网络中的社区结构。《物理评论快报》。2008;100:118703。doi:10.1103/PhysRevLett.100.118703。[29] Traag VA, Waltman L, van Eck NJ. 从Louvain算法到Leiden算法：确保社区连通良好。Scientific Reports。2019;9:5233。doi:10.1038/s41598-019-41695-z。

[30] Saracco F, Di Clemente R, Gabrielli A, Squartini T. 二部网络随机化：以世界贸易网络为例。Scientific Reports. 2015年;5:10595. doi:10.1038/srep10595.[31] Clopper CJ, Pearson ES. 以二项分布为例说明置信限或枢轴限的使用。Biometrika. 1934年;26(4):404–413. doi:10.1093/biomet/26.4.404.[32] Neal ZP. 何时可以根据观测到的群组推断网络？Network Science. 2024年;12:189–200. doi:10.1017/nws.2024.6.[33] Peel L, Peixoto TP, De Domenico M. 统计推断连接网络科学中的数据与理论。Nature Communications. 2022年;13:6794. doi:10.1038/s41467-022-34267-9.[34] Lin J. 基于香农熵的散度度量。IEEE Transactions on Information Theory. 1991年;37:145–151. doi:10.1109/18.61115.[35] Hill MO. 多样性与均匀度：统一的记法及其影响。Ecology. 1973年;54(2):427–432. doi:10.2307/1934352.[36] Efron B. 自助法：再论刀切法。The Annals of Statistics. 1979年;7(1):1–26. doi:10.1214/aos/1176344552.[37] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. 大型网络中社群的快速展开。Journal of Statistical Mechanics: Theory and Experiment. 2008年;2008(10):P10008. doi:10.1088/1742-5468/2008/10/P10008.