利用时序图学习检测国会议员交易中的信息渠道

Benjamin PhamRoodman*台湾大学R12922210@ntu.edu.tw

Eugene Sy*台湾大学d13949006@ntu.edu.tw

J. Xavier Atero V´azquez台湾大学d11949007@ntu.edu.tw

Yu-Shiang Huang National Taiwan University台湾大学f09946004@ntu.edu.tw

Che Lin台湾大学chelin@ntu.edu.tw

Chuan-Ju Wang中央研究院cjwang@citi.sinica.edu.tw

arXiv:2602.05514v1 [cs.CE] 5 Feb 2026

摘要

国会议员股票交易引发了人们对金融市场中潜在信息不对称和利益冲突的担忧。我们提出一种时间图网络（TGN）框架，用于识别国会议员在交易公司股票时可能获取有利信息的渠道。我们构建了一个多模态动态图，整合多个公开数据集，涵盖国会议员股票交易、游说关系、竞选资金捐款，以及立法者与企业之间的地理关联。我们将检测问题表述为动态边分类任务，以识别在较长时间跨度内表现显著优于标准普尔500指数的交易。为处理这些关系的时间特性，我们设计了一种两步滚动前向验证架构，遵守信息可用性约束并避免前视偏差。我们基于风险调整后收益评估了多种标注策略，并证明 TGN 能够成功捕捉国会与企业互动和后续交易表现之间复杂的时间依赖关系。

1 引言

金融市场的诚信在很大程度上依赖于政府官员与企业实体之间交互的透明度。然而，公众长期以来一直怀疑国会议员会利用特权信息谋取私利（Ziobrowski 等人，2004年；Karadas 等人，2021年）。在多起备受关注的调查指称立法者利用非公开情报进行交易之后（Schweizer，2011年；CBS新闻，2011年），《2012年国会知识交易法》（STOCK Act）得以通过。该法旨在重申内幕交易禁令，并要求及时进行财务披露（Nagy，2011年；Belmont 等人，2022年），但人们仍对这些监管防线的有效性存有疑问。

尽管存在这些监管规定，立法者对其所监管的行业仍保有显著的结构性优势。这一地位使其能够进行信息预判，即议员可在国防合同或税法变更等足以影响市场的事件公开数周之前便对其作出估计（Williams，2024年）。此类预见往往得益于特定的委员会任职，这些任职使其能够提前获取可据以行动的情报（Hanousek Jr 等人，2022年）。市场日益认识到这一持续存在的优势，其证据是出现了以STOCK法案披露信息为信号的投资工具。NANC和KRUZ等交易所交易基金如今明确跟踪在任国会议员的投资组合，从而实质上使追求国会议员超额收益的做法制度化（Tidal Financial Group，2024年）。

*共同贡献

然而，若仅将这些交易视为孤立的时间序列信号，便会忽略其发生时所处的丰富结构语境——例如往往促成一笔交易的特定委员会关系或游说关系。虽然将这一生态系统建模为动态图提供了一种有前景的解决方案，但将标准时序图网络（TGN）应用于金融领域会带来一项关键挑战：信息陈旧。与反馈往往即时发生的社交网络不同，金融市场在观察到一笔交易与其绩效标签兑现（例如18个月收益）之间存在显著的结果兑现延迟。需要即时监督的标准模型通常会在这一尚未兑现的间隔期间丢弃交互。这迫使模型在预测时依赖过时的节点表示，从而舍弃最新、且可能最具信息量的结构信号。

为此，我们提出 **GAP-TGN** （***G***门控 异步 传播 TGN）。通过将国会交易生态系统形式化为动态二分图，并引入结合异步传播的门控融合机制，GAP-TGN 即使在标签兑现间隔期间处理交互，也能保持信息状态的实时更新。我们在这一延迟感知框架下，将显著交易的识别视为一项监督式边分类任务。
我们的贡献有三点：

1. 我们推出 **Capitol Gains**，这是一个整合数据集，将立法披露、游说记录和竞选财务统一为连贯的动态图结构。2. 我们提出 **GAP-TGN**，这是一种专为金融交互设计的新型时序图架构。我们引入门控多模态融合 层，用于结合动态记忆与静态属性，并提出异步传播 策略，即使没有即时标签，也能保持节点嵌入的时效性。3. 初步实验表明，GAP-TGN 在长期期限上的表现优于纯表格基线。基线会出现模式坍缩，而 GAP-TGN 则能保持稳健的性能，这为关系位置是政治阿尔法的重要领先指标提供了概念验证。

除这些初步发现之外，我们将本研究定位为未来研究的催化剂。我们指出，立法-企业图是金融图学习中一个尚未得到充分探索的前沿领域。最后，我们讨论整合复杂高阶交互类型的潜力——例如法案赞助周期和游说强度——为解码政治阿尔法的高维运作机制奠定基础。

2 数据构建

为严格建模国会山与华尔街之间的信息不对称，我们引入**CapitolGains**，这是一个多模态时序图数据集。不同于以往孤立分析国会议员交易的研究，我们旨在将这些金融交易置于更广泛的立法影响力网络中加以理解。该数据集将三个不同领域——立法活动、企业游说和金融市场——统一到一个动态图结构中，并按照每日变化从2013年1月1日持续至2025年12月31日。

该数据集的一项关键设计约束是严格防止前视偏差。我们执行**时点（PIT）**构建协议，使数据可用性与公开披露严格保持一致。在这一机制下，国会议员交易仅在其申报日期，即市场获知该交易的时刻，才会被纳入图中，而不是在私人执行日期纳入。我们将这种时间纪律延伸至宏观经济指标和企业基本面，按照其精确的发布时间戳为所有特征建立索引。这种同步确保模型在任意时间 t 所处的信息状态，都与市场参与者可获得的信息集完全对应，从而保持严格的因果性。

最终形成的图模式如图1所示，将异构数据集整合为两个主要节点集：国会议员和公司，它们通过下文定义的动态交互相连接。数据集的统计信息详见附录A中的表3。

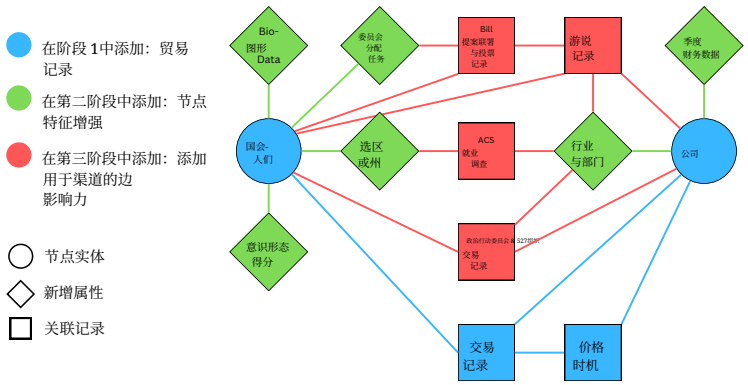

图1：我们数据的关系图。

2.1 国会议员信息

我们使用传记、地理和意识形态属性来描述立法节点，以捕捉其获取私人信息的潜在能力。

立法者档案：我们使用 congress-legislators[1] 仓库（United States Congress Project, 2025年）获取核心传记数据、任期和党派归属。为捕捉地方经济激励因素，我们利用**美国社区调查（ACS）**的就业数据，将立法者与其选区的产业构成对应起来，并对标准化的选区层级报告实施之前的年份进行邮政编码交叉映射。

动态意识形态与委员会：立法者的立场并非静态。我们利用根据唱名表决得出的每日 W-NOMINATE 得分（Lewis 等人，2025年），追踪其不断变化的政治立场，并以头两个维度表示经济立场和社会立场（采用两年滚动回溯更新）。此外，委员会任职通常是获取特权监管信息的来源，我们根据涵盖所有常设委员会和特设委员会的固定词汇表，采用多热编码方案对其进行嵌入。

2.2 交易与影响力记录

在刻画立法者主体及其属性后，我们现在描述将其与企业部门联系起来的动态机制。我们区分两类交互：由直接股票交易产生的明确**金融信号**，以及通过游说活动和竞选捐款建立的、更广泛的**关系背景**。这些边共同描绘了资本与政治影响力的流动。

国会议员交易记录：我们的主要目标边是来源于**Quiver Quantitative**（Quiver Quantitative，2025年）的股票交易。我们对原始数据进行筛选，以分离出股票类信号，并剔除非股票资产。每条边包含交易方向（买入/卖出）、时间元数据（交易与申报日期之间的时滞）以及交易量。依据STOCK法案，交易金额以区间档位报告（例如，$1K–$15K）；我们采用这些区间档位的下限，以便对交易量进行保守估计。

游说与竞选财务记录：我们描绘了企业影响力的两条渠道。首先，我们整合**LobbyView**（Kim, 2018年），将企业客户与具体法案关联起来。随后，我们利用**VoteView**数据，将这些法案与立法者联系起来。如果立法者发起了受游说影响的法案，我们便定义一条“强关联”边；如果他们只是投了“赞成”票，则定义一条“弱关联”边。其次，我们汇总**OpenSecrets**中的竞选捐款，将企业政治行动委员会（PAC）和527组织与立法者关联起来。这些边量化了国会议员从其监管行业获得的直接资金支持。

[1]https://github.com/unitedstates/congress-legislators

2.3 金融与市场数据

我们构建了一个包含每日价格和成交量指标的数据集。所有股票的每日开盘价、最高价、最低价、收盘价和成交量（OHLCV）数据均通过**Massive API**（原 Poly-gon.io）私下获取（Massive Data，2025年）。我们在标准价格数据的基础上，增加了暗池交易量，该指标通过汇总各股票代码对应的场外交易报告得出。

2.4 经济与公司背景

XT

为捕捉更广泛的经济环境，我们纳入了 64 项宏观经济指标，这些指标来源于圣路易斯联邦储备银行的联邦储备经济数据档案（ALFRED）API；此外，我们还纳入了从美国证券交易委员会的电子数据收集、分析与检索系统（EDGAR）API 检索的季度 10-Q 定期报告中的公司财务基本面数据（美国证券交易委员会，2025年；圣路易斯联邦储备银行，2025年）。这些特征分别详见附录 D 和 E，并严格依据其公开发布日期或申报日期更新，而非依据报告期更新，从而确保模型的信息状态仅反映在任何给定交易发生时市场参与者能够获取的数据。

3 方法

3.1 问题定义

我们将国会议员交易中的信息渠道检测形式化为连续时间异构图上的动态边分类任务。

图的形式化定义。我们将国会生态系统建模为连续时间异构图 $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$。节点集 $\mathcal{V} = \mathcal{V}_L \cup \mathcal{V}_C$ 将实体划分为政治人物（立法者）和公司（发行人），每个实体都关联着随时间变化的特征向量 $\mathbf{x}_v(t)$，用于 $v \in \mathcal{V}$。图通过时序事件流 $\mathcal{S} = \{e_1, e_2, \ldots, e_N\}$演化。我们将历史上下文 $\mathcal{H}(t) = \{e_j \in \mathcal{S} \mid t_j < t\}$ 正式定义为时刻t可用的所有先前事件的集合。每个交互事件涉及两个节点或单个节点；涉及两个节点的事件均定义为一个元组 $e = (u, v, t, k, \mathbf{x})$，其中 $u \in \mathcal{V}_L$ 和 $v \in \mathcal{V}_C$ 是时刻 t发生交互的节点，$k \in \mathcal{K}$ 表示交互类型，而 $\mathbf{x}_k(t) \in \mathbb{R}^{d_k}$ 则刻画事件属性。更新单个节点属性的交互事件定义为 $e = (u, t, x)$，其中 $u \in \mathcal{V}$ 和 t保持不变。此外，除了事件流和节点属性之外，我们还纳入系统性市场特征向量 $\mathbf{x}_{\text{mkt}}$。

目标定义。为构建学习任务，我们将一种特定的交互类型指定为目标关系($k_{\text{target}} \in \mathcal{K}$)，同时将所有其他交互类型视为关系上下文 ($k \in \mathcal{K} \setminus \{k_{\text{target}}\}$)。该框架使模型能够利用上下文边所提供的潜在影响结构来解读显式信号。严格聚焦于 $k_{\text{target}} =$ TRADE 的情形，考虑一笔目标交易 $e_i = (u, v, t, \text{TRADE}, \mathbf{x})$。我们根据其在未来时间跨度 Δt内的超额收益定义一个二元标签 y_i。令 $R_v(t, \Delta t)$ 和 $R(t, \Delta t)$ 分别表示资产和基准的累计收益，我们定义：

$$y_i = \mathbb{I}\left(\left(R_v(t, \Delta t) - \bar{R}(t, \Delta t)\right) > \tau\right) \tag{1}$$

其中，τ是性能阈值。目标是学习一个映射 $f(\mathcal{H}(t), \mathbf{X}(t)\ e_i) \to \hat{} y_i$，基于历史上下文 $\mathcal{H}(t)$ 和演化节点状态$\mathbf{X}(t)$预测这些高阿尔法机会。

3.2 门控异步传播时序图网络

我们提出了**GAP-TGN**（***G**ated**A**synchronous **P**ropagation* 时序图网络），该模型将时序图网络（TGN）架构（Rossi 等，2020）扩展至连续时间动态图（CTDG）。不同于为高频、即时反馈环境（如社交网络）设计的标准 TGN，GAP-TGN 专为应对长时域预测窗口（Δt）导致的**信息陈旧** 问题，以及*CapitolGains*数据集固有的多模态异质性而设计。

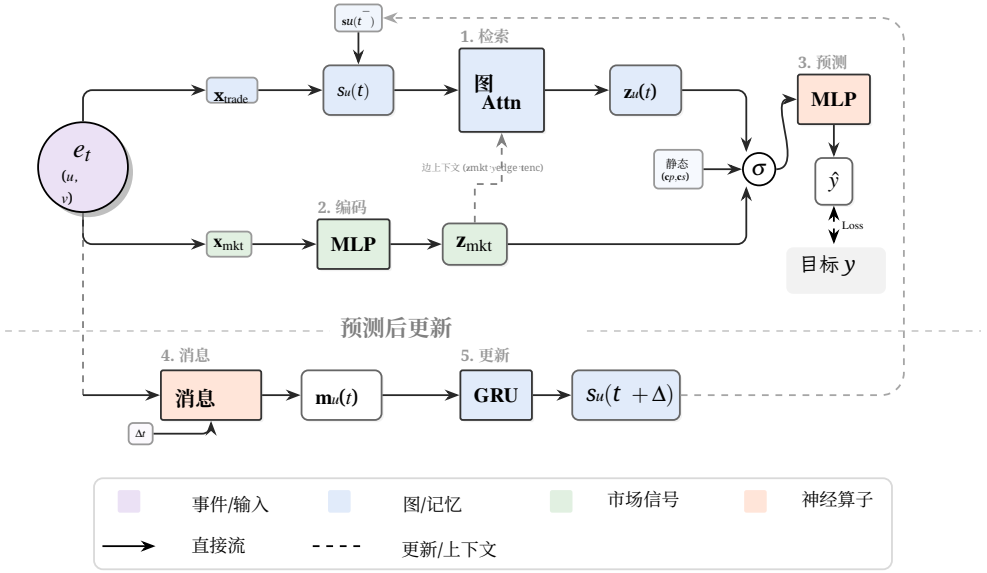

图 2：**GAP-TGN 架构** 处理事件 $e_t(u,\ v)$ 涉及：(1) 对交易和市场特征进行编码；(2) 通过受边上下文调制的图注意力检索记忆 $s_u(t)$；(3) 将动态嵌入与静态属性 ($\mathbf{c}_p,\ \mathbf{c}_s$) 融合，以预测 $\hat{}y$；(4) 在预测后根据事件信号和 Δt 构建消息；(5) 将更新传播至记忆状态$s_u(t+\Delta t)$，严格区分预测时刻与状态演化。

3.2.1 模型输入与特征编码（步骤 0–2）

我们首先定义驱动下游组件的原始输入信号及其潜在编码（步骤 2）。

特征输入。 令$\mathbf{x}_{\text{trade}}$ 表示交易属性向量，令$\mathbf{x}_{\text{mkt}}$ 表示系统性市场特征向量（详见第 4.1 节）。我们还定义谐波时间编码$\mathbf{t}_{\text{enc}}(\Delta t)$，以捕捉时间上的相对位置。此外，每个立法者节点 u 都关联着可学习的静态嵌入$\mathbf{c}_p$ （政党）和$\mathbf{c}_s$ （州），以捕捉不随时间变化的隶属关系。

传播监督信号(y_{edge})。为缓解结果兑现延迟 ΔT 导致的信息陈旧问题，我们在标签集中引入一个潜在状态 ($y = 0.5$)：

$$y_{\text{edge}} = \begin{cases} 1 & \text{Outperformed (Resolved)} \\ 0 & \text{Underperformed (Resolved)} \\ 0.5 & \text{Pending (Latent)} \end{cases} \tag{2}$$

潜在市场编码（步骤 2）。原始市场特征经由多层感知机（MLP）投影到潜在空间，以形成市场信号：

$$\mathbf{z}_{\text{mkt}}(t) = \texttt{MLP}(\mathbf{x}_{\text{mkt}}(t)) \tag{3}$$

该潜在信号 $\mathbf{z}_{\text{mkt}}$ 用于两个不同的下游组件：作为历史检索（步骤 1）的上下文，以及作为门控融合（步骤 3）的全局信号。

3.2.2 时间嵌入与上下文检索（步骤 1）

为对时间 t 的目标交互进行分类，模型首先查询节点当前的记忆状态$\mathbf{s}_u(t)$。它检索局部邻域 $\mathcal{N}_u(t)$——该邻域定义为 u 在 t 之前涉及节点的最近 M 次交互序列。我们利用图注意力机制，将这些历史信号聚合为时间嵌入 $\mathbf{z}_u(t)$：

$$\mathbf{z}_u(t) = \text{MultiHeadAttention}\left(\mathbf{q}_u, \mathbf{K}_u, \mathbf{V}_u\right) \tag{4}$$

具体而言，对于每个注意力头 h，我们定义查询 $\mathbf{q}_u^{(h)}$、键 $\mathbf{k}_j^{(h)}$ 和值 $\mathbf{v}_j^{(h)}$ 向量。查询表示焦点节点的当前状态，而键和值则编码

历史交互上下文：

$$\mathbf{q}_u^{(h)} = \mathbf{W}_q^{(h)} \mathbf{s}_u(t) \tag{5}$$

$$\mathbf{k}_j^{(h)} = \mathbf{W}_k^{(h)} \mathbf{s}_j(t^-) + \mathbf{W}_{k,e}^{(h)} \phi_j \tag{6}$$

$$\mathbf{v}_j^{(h)} = \mathbf{W}_v^{(h)} \mathbf{s}_j(t^-) + \mathbf{W}_{v,e}^{(h)} \phi_j \tag{7}$$

其中，t^- 表示紧邻 t 之前的时间戳，$\mathbf{W}_q$，$\mathbf{W}_k$，$\mathbf{W}_v$ 为可学习权重矩阵，$\boldsymbol{\phi}_j$为边特征向量。我们使用缩放点积注意力计算注意力系数 $\alpha_{u,\,j}^{(h)}$，以确定每个历史邻居 $j \in \mathcal{N}_u(t)$ 与当前预测任务的相关性：

$$\alpha_{u,j}^{(h)} = \text{softmax}_j \left(\frac{(\mathbf{q}_u^{(h)})^\top \mathbf{k}_j^{(h)}}{\sqrt{d_k}} \right) \tag{8}$$

最终嵌入 $\mathbf{z}_u(t)$ 通过拼接所有 H 个头的加权和形成：

$$\mathbf{z}_u(t) = \Big\|_{h=1}^{H} \left(\sum_{j \in \mathcal{N}_u(t)} \alpha_{u,j}^{(h)} \mathbf{v}_j^{(h)} \right) \tag{9}$$

该机制使模型能够有效查询其历史上下文。例如，如果当前查询$\mathbf{q}_u$ 编码了一个科技行业购买，注意力权重 $\alpha_{u,\,j}$ 将动态集中于过去的邻居，其中 $\boldsymbol{\phi}_j$ 表示相似的结构模式或高 alpha 交易。

注意力分数由**边上下文** ϕ_j决定。不同于记忆更新，该注意力机制明确纳入了历史邻居交互的结果标签（y_{edge}）。
边特征向量 ϕ_j 由预定义输入组成：

$$\boldsymbol{\phi}_j = [\mathbf{x}_{\text{trade}}(t) \parallel \mathbf{z}_{\text{mkt}}(t) \parallel \mathbf{t}_{\text{enc}}(\Delta t) \parallel y_{\text{edge}}] \tag{10}$$

加入 y_{edge} 后，注意力头能够区分与正阿尔法、负阿尔法或待定结果相关的邻居。

3.2.3 门控多模态融合（步骤 2–3）

该领域的一项关键挑战，是区分特质性内部人阿尔法与系统性市场动量。为此，
GAP-TGN 架构引入一个**门控融合**层（步骤 3），自适应地为结构信号和市场信号分配权重。我们定义了两种不同的潜在表示：

结构信号 ($\mathbf{z}^{\text{graph}}$)：时序图嵌入与可学习的静态嵌入的拼接，后者表示立法者所属的政党和州归属：

$$\mathbf{z}_{\text{graph}} = [\mathbf{z}_u(t) \parallel \mathbf{c}_p \parallel \mathbf{c}_s]$$

市场信号 ($\mathbf{z}_{\text{mkt}}$)：如第 3.2.1 节所定义，这是技术指标的潜在编码。

可学习标量门控 $g \in (0,1)$ 用于协调各模态的贡献：

$$g = \sigma\left(\mathbf{W}_g [\mathbf{z}_{\text{graph}} \parallel \mathbf{z}_{\text{mkt}}] + b_g\right) \tag{11}$$

$$\hat{y} = \texttt{MLP}\left([g \cdot \mathbf{z}_{\text{graph}} \parallel (1-g) \cdot \mathbf{z}_{\text{mkt}}]\right) \tag{12}$$

其中，σ 表示 Sigmoid 激活函数。
这种表述使模型能够动态调整其决策边界：当交易行为偏离市场基本面时，优先考虑结构图模式（高 g）；当结构信号较弱时，则依赖动量指标（低 g）。

3.2.4 记忆动态与状态持久性（步骤 4–5）

预测阶段结束后，模型会更新参与节点的状态，以反映新的交互。每个节点 $v \in \mathcal{V}$ 都维护着一个持久记忆状态 $\mathbf{s}_v(t) \in \mathbb{R}^{d_m}$，用作其历史交互轨迹的压缩表示。

消息计算（步骤 4）。为防止训练过程中发生标签泄漏，我们严格区分记忆更新信号与监督信号。消息函数接收由$\mathbf{x}_{\text{trade}}$定义的、严格可观测的交互属性。消息$\mathbf{m}_u(t)$按以下方式计算：

$$\mathbf{m}_u(t) = \texttt{MLP}\left(\mathbf{s}_u(t^-) \,\|\, \mathbf{s}_v(t^-) \,\|\, \Delta t \,\|\, \mathbf{x}_{\text{trade}}\right) \tag{13}$$

其中，Δt 表示自上一个事件以来经过的时间。

记忆更新（步骤 5）。随后通过一个 GRU 单元更新节点状态：

$$\mathbf{s}_u(t) = \texttt{GRU}(\mathbf{s}_u(t^-), \mathbf{m}_u(t)). \tag{14}$$

3.2.5 训练目标

模型通过端到端训练来最小化加权二元交叉熵损失：

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} w_{\text{pos}} \cdot y_i \log(\sigma(\hat{y}_i)) + (1 - y_i)\log(1 - \sigma(\hat{y}_i)) \tag{15}$$

其中，w_{pos} 会根据训练批次中正例的频率倒数动态调整。

3.2.6 异步传播策略

金融预测任务具有显著的**结果兑现延迟**（ΔT）。由于缺少真实标签，标准方法通常会丢弃间隔 $[t_{now} - \Delta T, t_{now}]$ 内近期发生的交互。

异步传播策略（详见公式 7）可确保记忆状态$\mathbf{s}_u(t)$与最近的实体活动保持同步，即使结果尚未显现。

该传播信号y_{edge} 通过 ϕ_j集成到图注意力层中。这使模型能够利用近期交易在结构上的存在进行情境化处理——例如，识别近期科技产品购买形成的簇，即使这些交易的绩效结果仍未揭晓。这样可确保节点记忆$\mathbf{s}_u(t)$与最近的实体活动保持同步。图 3 展示了我们的间隔利用策略与传统方法的直观对比。

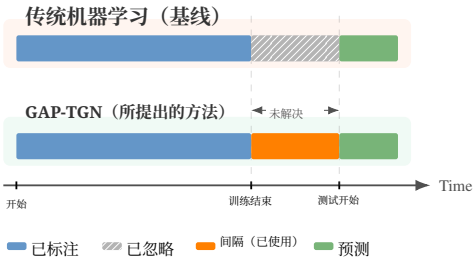

图 3：**间隔利用**。标准模型会在报告间隔期间丢弃交互数据。

4 实验

我们的初步实验评估着重于建立表格基线与所提出的 GAP-TGN 之间的比较。我们报告了基线模型和所提出模型的结果。此外，我们概述了第二阶段和第三阶段的设计，作为捕捉复杂游说互动的扩展方案，并为政治金融图网络的未来研究提供路线图。

4.1 实验设置

我们将 GAP-TGN 与三种表格基线进行比较：逻辑回归、多层感知机 (MLP) 和 XGBoost。这些模型使用所有可用属性的展平特征向量，但忽略拓扑依赖关系。

4.1.1 特征集

交易特征 ($\mathbf{x}_{\text{trade}}$)：我们为每笔交易构建一个三维向量，其中包含：(1) 以美元计的交易金额经对数转换后的值，(2) 二元方向 ($\mathbb{I}_{\text{buy}}$)，以及 (3) 以天为单位的报告延迟经对数转换后的值。

市场特征（$\mathbf{x}_{\text{mkt}}$）：为刻画系统性市场环境，我们构建了一个包含 14 个维度的向量，其中包括：

- **动量：**回看窗口为 $\{1, 5, 10, 20\}$ 天的离散累计收益率。
- **波动率：**20 天历史波动率（v_{20d}）。
- **技术指标：**14 天相对强弱指数（RSI）和成交量比。

4.1.2 目标生成

与公式 1 一致，我们将阈值设为 $\tau = 0\%$，并将任何表现优于标准普尔500指数基准的交易标记为正类（$y = 1$）。

为评估各组件的贡献，我们提出了一份路线图，区分当前实现与计划中的扩展。首先，**交易基线** 构成了当前的结构基础，仅基于带有价格上下文的交易图 $\mathcal{G} = (\mathcal{V},\ \mathcal{E}_{trade})$运行。作为近期工作的组成部分，**节点增强** 将为议员党派和州隶属关系添加可学习的嵌入。最后，计划通过**边增强** 引入多关系上下文，包括游说、捐赠和地理产业关联，以捕捉金融交易之外的软性影响渠道。所有模型均按月度滚动方式进行训练和测试，时间范围为2019年1月至2024年4月。

表 1：**长周期表现**：GAP-TGN 在 18M/24M 下的比较。

	AUROC		F1分数	
模型	18M	24M	18M	24M
逻辑回归	0.486	0.465	0.300	0.195
MLP	0.496	0.498	0.402	0.378
XGBoost	**0.525**	**0.524**	0.334	0.291
GAP-TGN	0.490	0.518	**0.438**	**0.440**

表 2：**消融研究**：组件贡献分析。

	AUROC		F1分数	
模型	18M	24M	18M	24M
GAP-TGN（完整模型）	**0.490**	0.518	**0.438**	**0.440**
原始 TGN	0.478	0.517	0.416	0.393
无门控融合	0.476	0.508	0.431	0.427
无异步传播	0.478	**0.521**	0.424	0.405

4.2 初步性能分析

表 1 和表 2 展示了对我们交易基线的评估结果。排序能力与分类稳定性之间呈现出明显的权衡。表格基线（如 XGBoost）的 AUROC 略高于 GAP-TGN（24 个月时分别为 0.524 和 0.518），但在必须做出可执行的二元预测时，其信号会严重衰减。具体而言，XGBoost 的 F1分数在 24 个月时间跨度下骤降至 **0.291**。与之形成鲜明对比的是，GAP-TGN 表现出更强的**信号稳定性**，F1分数保持在**0.440**，识别正例的能力提升了超过**51%** 。

消融研究进一步验证了我们的架构设计。移除异步传播策略后，F1分数降至 0.405，这证实弥合尚未解决的间隔对于长期记忆保留至关重要。同样，不进行市场融合的原始 *TGN*性能降至 0.393，凸显了结合市场数据对结构模式进行情境化处理的必要性。总体而言，这些初步结果表明，当前的结构基线能够有效捕捉持久的潜在因子，并取得较高的 F1 分数，但与 AUROC 较低的表格模型相比，它可能仍然缺乏特征。我们认为，计划中的**节点增强**和**边增强**将提供必要的信息密度，在保持模型结构稳健性的同时弥合 AUROC 差距。

5 结论

在本研究中，我们介绍了*Capitol Gains*，这是一个用于建模立法金融生态系统的动态图数据集。GAP-TGN**交易基线**的初步结果表明，

即使没有辅助特征，结构信号的分类稳定性也优于表格基线。为支持进一步探索，我们在附录B中详细提出了**下游金融应用**方案，并在附录C中全面回顾了**相关工作**，展示了该数据集在现实市场环境中的潜在效用。未来的工作将着重于完成路线图，整合计划中的节点和边增强模块。

参考文献

William Belmont、Bruce Sacerdote、Ranjan Sehgal 和 Ian Van Hoek。参议员和众议院议员能否跑赢股市？来自 STOCK法案的证据。《公共经济学杂志》，207:104602，2022年。

Anthony Brabazon 和 Michael O'Neill。《金融中的进化计算》。施普林格，2008。CBS新闻。内幕人士：通往 STOCK法案之路。《60分钟》，2011年。网址 https://www.cbsn ews.com/news/insiders-the-road-to-the-stock-act/。

Nick Cleveland-Stout。今年至少有37名美国立法者交易了价值高达1.13亿美元的军工股。https://truthout.org/articles/at-least-37-us-lawmakers-traded-up-to-113-million-in-arms-stocks-this-year/，2024年12月。

Andrew C. Eggers 和 Jens Hainmueller。《国会损失：国会议员股票投资组合的平庸表现》。《政治学杂志》，75(2):535–551，2013年。

圣路易斯联邦储备银行。ALFRED：联邦储备经济数据档案。https: //alfred.stlouisfed.org/，2025年。访问日期：2025年12月31日。

Richard C. Grinold 和 Ronald N. Kahn。《主动投资组合管理：实现卓越回报并控制风险的定量方法》。麦格劳-希尔，纽约州纽约市，第二版，1999年。

Jan Hanousek Jr、Hoje Jo、Christos Pantzalis 和 Jung Chul Park。政治内幕交易中自利与道德责任的两难。《商业伦理学杂志》，177:1–25，2022年。

Serkan Karadas 和 Minh Tam Tammy Schlosky。什么因素能够解释国会议员的交易行为？基于超过 100,000 笔国会议员股票交易的证据。《国际经济与金融评论》，96:103591，2024年。

Serkan Karadas、Minh Tam Tammy Schlosky 和 Joshua Hall。政治人物是否在股票交易中利用了非公开宏观经济信息？来自2012年STOCK法案的证据。《风险与财务管理杂志》，14(6):256，2021年。

In Song Kim。LobbyView：公司层面的游说与国会法案数据库。工作论文，2018年。网址 https://www.lobbyview.org/data-download/datasets/。

Ilyana Kuziemko、Ina Taneva 和 Atif Mian。国会议员股票交易严重损害公众信任和守法意愿。工作论文，雷迪管理学院，加利福尼亚大学圣迭戈分校，2025年5月。

Jeffrey B. Lewis、Keith Poole、Howard Rosenthal、Adam Boche、Aaron Rudkin 和 Luke Sonnet。VoteView：国会唱名表决数据库，2025年。网址 https://voteview.com/。版本 2022.1。

Marcos L´opez de Prado。《金融机器学习进阶》。约翰·威立父子出版公司，新泽西州霍博肯，2018年。参见第10章：下注规模。

Massive Data。Massive Data：机构级股票数据 API，2025年。网址 https://massiv e.io/。访问日期：2025年12月31日。

Peter Molk 和 Frank Partnoy。国会中的负向交易。《印第安纳法律杂志》，100(3):1117，2025年。

Donna M. Nagy。内幕交易、国会官员与受托义务。《波士顿大学法律评论》，91:1105，2011年。

RobertPardo。《交易策略的评估与优化》。约翰·威立父子出版公司，新泽西州霍博肯，第二版，2008年。

Quiver Quantitative。国会交易数据。专有数据集，2025年。访问日期：2025年12月31日。

Responsible Statecraft。2024年国会顶级国防股票交易者。https://responsibl estatecraft.org/congress-defense-stocks/，2024年12月。

Emanuele Rossi、Ben Chamberlain、Fabrizio Frasca、Davide Eynard、Federico Monti 和 Michael Bronstein。用于动态图深度学习的时序图网络，2020年。网址：https: //arxiv.org/abs/2006.10637。

Peter Schweizer。《把他们全部赶出去》。Houghton Mifflin Harcourt，2011年。

The Motley Fool。国会议员股票交易：谁在交易，谁赚得最多。https://www. fool.com/research/congressional-stock-trading-who-trades-and-mak es-the-most/，2026年。数据来源：Unusual Whales 和 Capitol Trades。

Tidal Financial Group。交易所交易基金聚焦：NANC 和 GOP，2024年。网址：https://tidalfinanci algroup.com/etf-spotlight-nanc-and-kruz/。访问日期：2025-01-14。

United States Congress Project。congress-legislators：美国国会议员，1789年至今，2025年。网址：https://github.com/unitedstates/congress-legislato rs。公有领域（CC0 1.0 Universal）。

美国证券交易委员会。EDGAR：电子数据采集、分析与检索系统。https://www.sec.gov/edgar，2025年。

Hong Vo、Quoc-Dat Trinh、Minh Le 和 Thuy-Ngan Nguyen。经济政策不确定性是否会影响投资对同业股票价格的敏感度？《经济分析与政策》，72：685–699，2021年。

Shang-Jin Wei 和 Yifan Zhou。“国会山上的‘船长收益’”。国家经济研究局工作论文第34524号，2025年。

Sarah J. Williams。规范国会议员内幕交易：烂鸡蛋策略。《卡多佐法律评论》，45：1399，2024年。网址：https://larc.cardozo.yu.edu/clr/vol45/iss5/3。

David H. Wolpert。堆叠泛化。《神经网络》，5(2)：241–259，1992年。

Alan J. Ziobrowski、James W. Boyd、Ping Cheng 和 Brigitte J. Ziobrowski。美国参议院普通股投资的异常收益。《金融与定量分析杂志》，39(4):661–676，2004年。

A 数据集统计

表3：数据集统计与规模

数据集	来源	条目
国会议员交易	Quiver Quantitative	32,401
单届任期议员元数据	congress-legislators	45,525
唱名表决	VoteView	26,262,296
意识形态得分（季度）	VoteView	32,740
选区就业	美国社区调查	5,000
游说报告	LobbyView	1,690,362
政治行动委员会交易	OpenSecrets	423,350,920
527组织交易	OpenSecrets	8,714,809
企业季度财务指标	美国证券交易委员会的 EDGAR API	20,309,853
经济指标	FRED 的 ALFRED API	260,778

B 下游金融应用

由于本研究仍在进行中，我们目前的重点是确立 *Capitol Gains* 数据集的质量以及 GAP-TGN 在架构上的有效性。为了在真实的市场环境中全面评估这些结构信号的经济效用，我们提出了一种基于**时间序列堆叠架构**的未来回测环境。

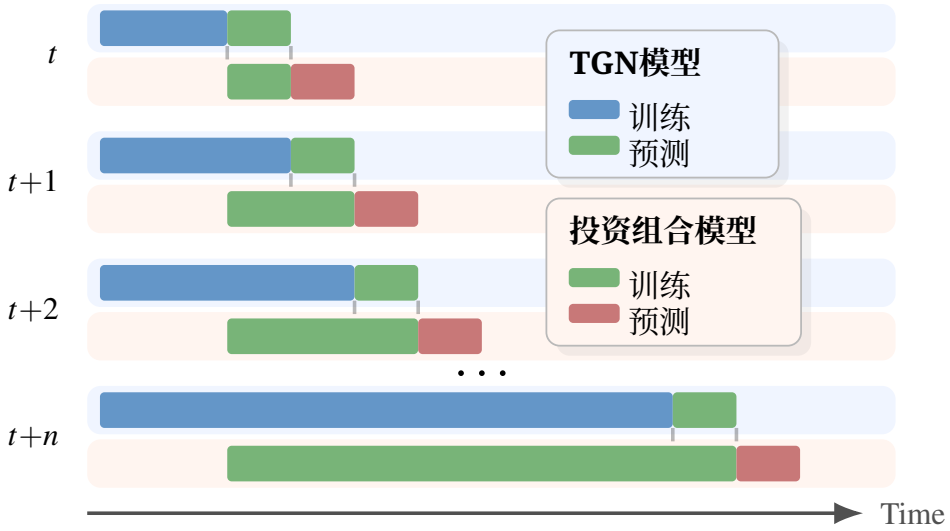

图 4：GAP-TGN 与投资组合配置模型的两级滚动前向训练。

B.1 投资组合配置模型

所提出的投资组合配置模型通过处理与每笔交易相关的预测标签，从申报当日起每日重新确定投资组合配置。每个标签都被视为置信度信号。该配置模型可分为四个步骤：

1. **篮子置信度：** 衡量活跃股票篮子的整体置信度。 2. **个股置信度：** 根据 TGN 的信号强度衡量每只个股的置信度，并采用元标签法确定下注规模（L´opez de Prado，2018年）。 3. **风险调整：** 根据风险承受能力和每只股票的历史波动率，通过逆波动率加权调整权重，以规范风险贡献。 4. **最终加权：** 确定分配给每只股票和标准普尔500指数基准的最终资本权重。

该模型采用信号衰减机制。阿尔法衰减这一概念确保较旧的信号对当前配置的影响逐渐减弱（Grinold & Kahn，1999）。配置函数由七个超参数控制：标普指数与股票权重平衡、置信度敏感性、波动率敏感性、风险敏感性、多空权重、信号衰减率，以及信号激活阈值。这些参数通过遗传算法（Brabazon & O'Neill，2008年）进行优化，该算法在训练窗口内迭代演化参数集。

B.2 时间序列堆叠架构

我们采用了堆叠泛化框架（Wolpert，1992年），并根据 Pardo（2008年）的描述，将其调整以适用于时间数据结构。如图 4 所示，该架构由两个层级组成。第 0 层是前文介绍的 TGN。它接收原始图数据，并为每笔国会议员交易输出一个概率分数。第 1 层是投资组合配置模型。它接收第 0 层生成的样本外概率分数，并将其作为交易信号。

两个模型的训练窗口和测试窗口按照协调一致的前推序列移动。对于任意一个测试月份，Level 0模型均使用24个月的滚动窗口进行训练。验证阶段中TGN生成的样本外预测会拼接起来，构成投资组合配置模型的训练集。

B.3 计划采用的比较指标

完成计划于2016年1月至2025年12月期间进行的金融回测后，我们将比较五种投资组合的表现：被动型标准普尔500指数基准、由基线（XGBoost）模型驱动的投资组合，以及由阶段1至3的TGN模型驱动的三种投资组合。我们将使用以下指标评估这些模型：

- **年化收益率和波动率：**用于衡量原始表现和风险。
- **夏普比率：**用于衡量单位风险所对应的收益效率。
- **最大回撤：**用于评估相对于基准的下行风险。
- **索提诺比率和卡玛比率：**用于进一步揭示下行波动性和恢复潜力。

我们预计将分析累计收益率，以观察其相对于基准的偏离，尤其是在市场波动率上升的时期，此时 TGN 的结构特征可能带来独特的信息优势。此外，我们还将分析滚动夏普比率，以评估风险调整后收益的时间稳定性。

C 相关工作

我们将有关国会议员交易的现有文献分为三类：相对于监管的历史表现、基于市场的结果预测因子，以及归因于特定领导职务或制度趋势的表现。表4总结了这些研究发现。

表4：国会议员交易相关工作的分类

类别	研究	主要发现
历史与监管-制	Ziobrowski 等人（2004）	记录了显著的信息优-势，年度异常收益为 ∼12% ，发生于20世纪90年代。
	Eggers 和 Hainmueller（2013）	修正后的数据集显示，国会议员的投资组合表现逊于被动基准（2004—2008年）。
	Belmont et al. （2022年）；Karadas等（2021）	证实了STOCK之后的市场效率法案（2012）；该法案降低了预测能力re-关于宏观信息。
预测因子与Instru-ments	Karadas与Schlosky（2024年）；Vo等（2021年）	交易量与.....呈正相关经济政策不确定性（EPU）；激进
	Molk & Partnoy（2025年）	交易发生在公开数据不可靠时的 .发现业绩不对称；空头头寸可产生显著的异常收益而多头头寸则不会。
角色与行业	Wei & Zhou（2025年）	领导职位实现 ∼47%的年化异常收益；普通成员的表现则处于正常水平。
	Kuziemko 等人（2025年）	公众曝光高绩效群体会降低对政府的信任和合规度与联邦法律。
	The Motley Fool（2026年）	2024年，民主党人的表现优于共和党人（31–33%对26%），原因是更重的配置于科技行业。
	Cleveland-Stout（2024年）；Re-负责任的治国方略（2024年）	国防委员会成员交易了国防承包商，发生在地缘政治紧张局势 .期间

D 宏观经济指标

我们收集了64项宏观经济指标，作为全局上下文特征。这些指标根据圣路易斯联邦储备银行（FRED/ALFRED）当时报告的最新值每日更新。

表 5：宏观经济指标列表

类别	**指标（股票代码）**
价格与通胀	消费者价格指数（CPIAUCSL），核心消费者价格指数（CPILFESL），个人消费支出价格指数（PCEPI），核心个人消费支出价格指数（PCEPILFE），生产者价格指数（PPIACO），国内生产总值价格平减指数（GDPDEF）
货币政策与利率	联邦基金利率（FEDFUNDS），10年期美国国债利率（DGS10）、2年期国债收益率（DGS2）、M2货币供应量（M2SL）、10年期通胀保值国债（DFII10）
经济产出与活动	实际GDP（GDPC1）、潜在GDP（GDPPOT）、工业生产（INDPRO）、新屋开工（HOUST）
劳动力市场	失业率 (UNRATE), 非农就业人数 (PAYEMS)、初次申请失业救济人数 (ICSA)、职位空缺 (JTSJOL) 非农就业人数
大宗商品与原材料	WTI原油价格 (DCOILWTICO)、布伦特原油价格（DCOILBRENTEU），美国常规汽油价格（GASREGCOVW）
市场情绪与预期	芝加哥期权交易所波动率指数（VIXCLS），消费者信心（UMCSENT）

E 公司财务事实

以下财务数据摘自通过 SEC EDGAR API 获取的季度 10-Q 定期报告。这些特征按申报日期编制索引，以防止前视偏差。

- 流动应付账款
- 流动应收账款净额
- 扣除税项后的其他综合收益（损失）累计净额
- 资本公积
- 无形资产摊销
- 资产
- 流动资产
- 按账面价值计量的现金及现金等价物
- 已宣派的普通股每股股息
- 普通股价值
- 基本每股收益
- 稀释每股收益
- 商誉
- 毛利
- 所得税费用（收益）
- 利息费用
- 负债
- 负债及股东权益
- 流动负债
- 融资活动产生（使用）的现金净额

- 投资活动提供（使用）的现金净额
- 经营活动提供（使用）的现金净额
- 净利润（亏损）
- 营业外收入（费用）
- 营业费用
- 营业利润（亏损）
- 其他非流动资产
- 购置不动产、厂房和设备的付款
- 不动产、厂房和设备净额
- 研发费用
- 留存收益（累计亏损）
- 收入
- 销售、一般及行政费用
- 股东权益
- 稀释后加权平均流通股数
- 基本加权平均流通股数