

# Numinor SAM Amplifier 构建数据 v1.0

English: Numinor SAM Amplifier Construct Data 中文: Numinor SAM Amplifier 构建数据

---

## 1. 产品标识

---

产品名称: SAM Amplifier 构建数据 (Construct Data) 版本: v1.0 方法论参考: *Numinor SAM Amplifier 白皮书 v1.6* (2026 年 1 月) 参考实现: [github.com/Numinor-Systems/sam-amplifier-construct-reference](https://github.com/Numinor-Systems/sam-amplifier-construct-reference) — MIT 许可证

### 本产品是什么

一个预构建的股票级“同业 / 上游 / 下游关系图谱”，源自聚源 (ChinaScope) 的 SAM 产品分类体系与供应链披露数据。每个交易日，针对每只 A 股上市公司，Numinor 发布其同业 / 上游 / 下游对手方集合，以及按营收占比派生的边权重。

买方将此图谱作为聚合算子应用于自有因子模型。输出即买方现有因子经 SAM 网络“放大”后的版本——这是一种已被论证有效、可为多因子组合添加横截面信息的方法论。

### 本产品不是什么

- 不是 **alpha** 信号。这是经过重构的数据；alpha 取决于买方自身的因子模型。
- 不是聚源原始数据。原始数据由聚源单独销售。本产品使用该数据作为输入并加以重构。
- 不是预放大的因子文件。我们不发布“特定因子的放大版”——买方将图谱应用于自有因子文件。
- 不是基于 Numinor 示意性因子残差化后的版本。图谱与因子模型无关。

### 目标用户

在中国 A 股市场运作、已使用多因子模型 (Barra、MSCI、FactSet、CIQ 或自有因子) 的量化股票买方，希望在其现有因子敞口之上添加网络聚合通道。

---

## 2. Schema 架构

### 主架构 (parquet 与 CSV — 列结构完全一致)

列名	类型	描述
trade_date	date32	该图谱快照适用的交易日（亚洲/上海时区）。
focal_ts	string	焦点股票代码，格式 NNNNNN.XX（如 000001.SZ、600000.SH）。
relation_type	string	取值之一: "peer"、"upstream"、"downstream"。
counterparty_ts	string	对手方股票代码（格式同 focal_ts）。
weight	float64	营收占比派生的边权重，范围 [0.0, 1.0]。
source_basis	string	该边的聚源数据源依据: "sam_product_overlap"（同业）或 "sam_supply_chain_v4_corein"（上下游）。
source_rpt_date	date32	源数据生效日——计算该边所用的 SAM 报告日期（时点正确）。

### 示例数据

trade_date	focal_ts	relation_type	counterparty_ts	weight	source_basis	source_rpt_date
2026-05-30	000001.SZ	peer	600036.SH	0.034	sam_product_overlap	2025-12-31
2026-05-30	000001.SZ	peer	601398.SH	0.018	sam_product_overlap	2025-12-31
2026-05-30	000001.SZ	upstream	002475.SZ	0.022	sam_supply_chain_v4_corein	2025-12-31
2026-05-30	000001.SZ	downstream	600519.SH	0.015	sam_supply_chain_v4_corein	2025-12-31
2026-05-30	000001.SZ	peer	000063.SZ	0.011	sam_product_overlap	2025-12-31

### 类型 / 取值规则

- trade\_date: ISO 8601 日期格式 (YYYY-MM-DD)。对齐上海/深圳证券交易所交易日历；非交易日不发布。
- focal\_ts、counterparty\_ts: 仅限 A 股上市公司。格式 NNNNNN.SZ（深圳）或 NNNNNN.SH（上海）。非上市公司主体因 SAM 源数据本身就在上市公司层面构建，故不在本产品中。
- relation\_type: 仅小写。三个取值之一。
- weight: 6 位小数精度。永远为正（无负权重）。同业边在同一 focal\_ts 内权重之和为约 1.0（营收占比分配）。上下游边权重之和等于焦点公司的经济敞口占比（若部分对手方为非上市公司，则可能 < 1.0）。
- source\_basis: 字符串字面量。未来版本可能新增取值（如 "sam\_supply\_chain\_v4\_bizsup"），但既有取值稳定。

- `source_rpt_date` : SAM 源数据生效日。始终满足  $\leq \text{trade\_date} - \text{publish\_lag\_days}$  (详见 §4 时点纪律)。

### 架构中故意不包含的内容

- 自环边 (`focal_ts == counterparty_ts`) 按构造排除。
- 非上市公司主体 (无论焦点或对手方) 排除在外。
- 零权重边排除。
- 无“原始”公司标识符 (如 CSF-id 等) ——仅含股票代码。买方无需处理非上市公司主体。

## 3. 文件布局与交付

### S3 前缀结构 (两个层级, 同一 schema)

API 层与 Sandbox 层从两个互不相通的 S3 前缀消费形状相同的数据——唯一区别在于数据时效。Sandbox 层落后 API 层 30 个日历日 (T-30d)。该限制在存储层强制执行: Sandbox 的 API 密钥无法读取 API 层前缀, 反之亦然。

Parquet (主格式) :

```
s3://numinor-construct-data/sam-amplifier/api/parquet/
├─ historical/sam_amplifier_2016_<latest_year>.parquet      (单一拼接文件)
└─ year=YYYY/month=MM/day=DD/data.parquet                 (Hive 分区, 每日 T+1)

s3://numinor-construct-data/sam-amplifier/sandbox/parquet/
├─ historical/sam_amplifier_2016_<latest_year_minus_1mo>.parquet
└─ year=YYYY/month=MM/day=DD/data.parquet                 (Hive 分区, 每日 T+1, 数据为 T-30d)
```

CSV (次格式, 与聚源管道兼容) :

```
s3://numinor-construct-data/sam-amplifier/api/csv/          (结构同上, 文件为 .csv.zip)
s3://numinor-construct-data/sam-amplifier/sandbox/csv/     (结构同上, 文件为 .csv.zip)
```

两种格式数据完全一致。Parquet 使用 snappy 压缩; CSV 使用 ZIP 压缩 (聚源惯例)。Hive 分区 (`year=YYYY/month=MM/day=DD/`) 对 `pyarrow.dataset`、Spark、DuckDB 和 Athena 的范围查询友好。

### 历史批量文件 vs. 每日分区

- 历史批量文件 (`historical/sam_amplifier_2016_<latest_year>.parquet`) 是 2016-01-04 至切割日的全部交易日单一拼接 parquet 文件。为方便买方接入提供——一次下载即可, 约 12 GB, 无需自行拼接。
- 每日分区 (`year=YYYY/month=MM/day=DD/data.parquet`) 是持续交付通道。每日刷新写入一个新分区。

历史批量每季度重建一次; 两次重建之间, 买方通过分区文件消费切割日与今日之间的差额。

## 文件大小（近似值）

文件	Parquet	CSV.zip
历史批量文件（2016-2025，约 3.75 亿行）	约 12 GB	约 30 GB
典型每日分区（约 15 万行）	约 3-5 MB	约 8-12 MB
半年度更新日分区（SAM 数据更新后，约 50 万行）	约 15-25 MB	约 40-70 MB

## 更新节奏与 SLA

- **API 层每日刷新：**新分区 `year=YYYY/month=MM/day=DD/data.parquet` 在 亚洲/上海 **06:00** 之前发布，针对交易日 T，写入时间为 T+1。
- **Sandbox 层每日刷新：**每夜将 API 层 今日 - 30 天 的分区文件复制到 Sandbox 前缀对应路径。因此 Sandbox 层始终比 API 层晚 30 个日历日。
- A 股市场休市日不发布。
- **历史批量文件：**每季度重建一次。买方可选择拉取新的历史批量，或在两次重建之间使用分区文件。
- **方法论稳定性：**Numinor 承诺在版本内不变更方法论。方法论更新触发版本号升级（如 v1.1），并提前 60 天通知。

## 4. 时点（PIT）纪律

本产品时点正确：标记为 `trade_date = D` 的图谱快照仅使用截至 D 日已公开可得的聚源源数据。

### 发布滞后规则

对 v1.0 中使用的所有数据源，统一规则：

```
source_publish_date + 30 个日历日 ≤ trade_date
```

该缓冲考虑了真实的供应商交付时间（聚源 T+1）+ 机构买方的接入 / 重新计算 / 部署延迟（通常合计 3-4 周）+ 适度的保守余量。该规则与 Numinor SAM PM 构建数据所用规则一致，确保整个产品目录的滞后规则保持一致。

该规则由管道强制执行——`trade_date = D` 的快照中绝不会出现源数据在 `D - 30` 个日历日 之前尚未公开的行。

### 这对买方意味着什么

- 买方绝不会收到相对其 `trade_date` “前瞻”的边。
- 使用本数据进行回测时，时点纪律自动继承——买方端无需额外工作。
- 为审计 / 透明度，买方可对任一行检查 `source_rpt_date + 30 ≤ trade_date`（因为 `source_rpt_date` 包含在 schema 中）。

### 与白皮书的关系

已发表的 **SAM Amplifier** 白皮书 v1.6 在其实证分析中使用了更保守的 60 天滞后。v1.0 构建数据使用 30 天滞后，为生产应用提供更新鲜的数据。买方在本构建数据上完整运行白皮书方法论，将得到结构相似但数值略有差异的结果

——信号本质相同，但构建数据 PIT 窗口内多包含约 30 天更新近的 SAM 数据。

## 买方能否调整滞后？

- 更紧滞后 (< 30 天)：本产品不提供。需自行授权聚源原始 SAM 数据并构建管道。
- 更松滞后 (> 30 天，更保守)：买方端易于实施——只需在管道中以 `trade_date + extra_days` 消费图谱，无成本，可增加保守度。

30 天的选择刻意偏向宽松，使买方有空间在自己端收紧；若没有原始数据，无法将本产品的滞后放宽至 30 天以下。

## Sandbox 层时效（与 PIT 分开理解）

上文 PIT 纪律 (`source_rpt_date + 30 days ≤ trade_date`) 在 API 与 Sandbox 层完全相同。两者交付时效不同：

层级	可见的最新 <code>trade_date</code>
API	T（即昨日交易日，T+1 在亚洲/上海 06:00 前写入）
Sandbox	T - 30 个日历日

两层在任一给定 `trade_date` 下包含完全相同的行内容——唯一区别是 Sandbox 看不到 30 个日历日之内的最新数据。Sandbox 用于研究、原型设计与产品评估；API 用于生产。两层目标买方明显不同（散户/学生 vs 机构），两层不组合销售。

## 5. 图谱构建——Numinor 的工程价值

聚源（ChinaScope）发布原始 SAM 数据：每个公司 × 每个产品的营收明细表、以及产品层面的供应链分类（v4 版本）。对于股票层面的量化分析，买方需要将这些原始数据组装成股票对股票的图谱，配以合适的权重、时变快照与 PIT 纪律。这一组装工程，由 Numinor 完成。

### 同业边 (`relation_type = "peer"`)

对每一焦点股票 `i` 在 `source_rpt_date`，其披露营收按 SAM 产品节点分配（产品组合）。对每一产品节点 `p`，所有其他在 `p` 上有营收敞口的上市公司为候选同业，权重 = 双方营收占比之积：

$$\text{peer\_weight}[i \rightarrow j \text{ on product } p] = \text{revenue\_share}[i, p] \times \text{revenue\_share}[j, p]$$

跨所有共享产品求和得到同业边权重：

$$\text{peer\_weight}[i \rightarrow j] = (\sum_p \text{revenue\_share}[i, p] \times \text{revenue\_share}[j, p]) / \sum_p \text{revenue\_share}[i, p]$$

强制自我排除 ( $i \neq j$ )。同一 `focal_ts` 内的权重归一化为合计约 1.0。

### 上下游边 (`relation_type = "upstream" 或 "downstream"`)

对每一焦点股票 `i` 在 `source_rpt_date`，其产品组合映射到 SAM 供应链 v4 的 `COREIN`（核心投入）边。焦点每个产品的 `COREIN` 父节点成为上游候选；每个产品的 `COREIN` 子节点成为下游候选。权重由焦点对源产品的营收敞口 × 全市场所有上市公司对对手产品的营收敞口派生。

跨所有相关供应链边求和，按 `focal_ts` 归一化。

## 全集过滤

计算原始边后：1. 删除自环边（`focal_ts == counterparty_ts`） 2. 删除任一端不在 A 股 SH / SZ / KC / CYB 范围的边 3. 删除权重低于阈值（`1e-6`）的边——视为数值噪声 4. 删除违反 PIT 条件（`source_rpt_date + 30 days > trade_date`）的边

## 每日刷新

在每个新的 `trade_date`：1. 检查聚源每日增量文件中，是否有 SAM 源数据新跨过 `publish_date + 30` 天 阈值 2. 重新计算受影响焦点公司的同业 / 上游 / 下游边 3. 输出仅包含与前一交易日相比有变化的行的增量文件

大多数日子增量较小（无新 SAM 文件刚好跨过缓冲）。每年年报 / 半年报披露季（通常 5-6 月与 10-11 月）增量较大。

---

## 6. 全集 (Universe) 规则

### 纳入

- 所有 A 股上市公司，在 `trade_date` 有 SAM 覆盖，沪市 (SH) 或深市 (SZ) 上市。
- `trade_date` 停牌的股票保留（停牌是临时的，其网络位置仍有效）。
- 在 `trade_date` 之前退市的股票从该日起排除。
- 在 `trade_date` 之后 IPO 的股票在挂牌前排除。

### 排除 (按设计)

- 北京证券交易所 (BJ)：排除。Amplifier 方法论在 v1.6 白皮书中仅在 SH+SZ 上验证。
- 科创板 (KC) 与创业板 (CYB)：纳入全集（虽为增长板，但属于沪深主体；曾在 v1.6 白皮书审查中考虑排除，最终保留）。

### 覆盖起点

- 数据集覆盖 `2016-01-04` → 至今。2016 年之前的 SAM 数据深度不足，无法可靠推断同业 / 上下游关系。

---

## 7-13. 方法论概述、API 规范、快速入门、入门清单、版本与变更日志、FAQ、联系方式

注：第 7-13 节内容请参考英文版 [SAM\\_Amplifier\\_Construct\\_Data\\_v1.0.md §7-§13](#)。这些部分包含详细的 API 端点说明、订阅入门检查清单、版本路线图、常见问题 (FAQ) 以及支持联系方式，与英文版完全对应。中文版完整翻译将在 v1.0 稳定后随发布配套。

重要参数 (v1.0 锁定)：

- 定价由网站统一管理 — 请见 [numinor.io/lab/data/sam-amplifier](https://numinor.io/lab/data/sam-amplifier)
- 两层产品：API 层（生产级，T+1 时点数据）+ Sandbox 层（30 天滞后的同一数据，含 Token 计量沙箱环境）

- 无退款政策 (Sandbox 即为验证机制)
- API 域名: `api.numinor.io`
- S3 存储桶: `numinor-construct-data`, 区域 `ap-northeast-2` (首尔)
- 鉴权: 单用户单密钥, SKU 访问由订阅推导

完整 FAQ、API 端点详情与买方入门 workflows 请参阅英文版。

## 附录 A: 架构速查表 (可打印)

Numinor SAM Amplifier 构建数据 v1.0

格式: `parquet (.parquet)` 或 `CSV/ZIP (.csv.zip)`

<code>trade_date</code>	<code>date32</code>	YYYY-MM-DD, 上海交易日
<code>focal_ts</code>	<code>string</code>	NNNNNN.SZ 或 NNNNNN.SH
<code>relation_type</code>	<code>string</code>	"peer"   "upstream"   "downstream"
<code>counterparty_ts</code>	<code>string</code>	NNNNNN.SZ 或 NNNNNN.SH
<code>weight</code>	<code>float64</code>	[0.0, 1.0], 6 位小数精度
<code>source_basis</code>	<code>string</code>	"sam_product_overlap"   "sam_supply_chain_v4_corein"
<code>source_rpt_date</code>	<code>date32</code>	YYYY-MM-DD, PIT ( $source\_rpt\_date + 30 \leq trade\_date$ )

全集: A 股 (SH + SZ + KC + CYB); 不含 BJ

仅上市公司: 非上市主体因 SAM 源数据构造而不在内

覆盖: 2016-01-04 → 至今, 交易日每日发布

PIT 纪律:  $source\_rpt\_date + 30 \text{ 天} \leq trade\_date$

分层时效: API → `trade_date` 至 T (T+1 交付)

Sandbox → `trade_date` 至 (T - 30 个日历日)

S3 布局: `s3://numinor-construct-data/sam-amplifier/{api,sandbox}/{parquet,csv}/historical/sam_amplifier_2016_YYYY.parquet` (批量文件)  
`year=YYYY/month=MM/day=DD/data.parquet` (Hive 分区)

交付: 通过 `https://api.numinor.io/v1/constructs/sam-amplifier/` 签名 URL 派发  
`GET /historical` · `GET /day/{YYYYMMDD}` · `GET /range` · `POST /query`

SAM Amplifier 构建数据 v1.0 规范结束。方法论参考: [Numinor SAM Amplifier 白皮书 v1.6](#)。参考实现: [github.com/Numinor-Systems/sam-amplifier-construct-reference](https://github.com/Numinor-Systems/sam-amplifier-construct-reference) (MIT 许可证)。© 2026 Numinor Systems. 产品定义版权所有; 参考代码 MIT 开源。