

スニペット駆動型マイニングによる中国の「影のサプライチェーン」の可視化

## Visualizing China's "Shadow Supply Chain" through Snippet-Driven Mining

水野貴之\*1 深田大登\*2

Takayuki Mizuno Hiroto Fukada

\*1 国立情報学研究所 National Institute of Informatics

\*2 総合研究大学院大学 Graduate Institute for Advanced Studies, SOKENDAI

**要旨:** 中国国内のサプライチェーンは「世界の工場」を支える基盤であるが、公式の企業情報では取引先の約83%が匿名化されており、その構造は不透明なままである。本研究では、この情報の非対称性を克服するため、Qwen 3 と DSPy 最適化プロンプトを用いた OSINT パイプラインを開発し、上場企業4,331社および大手企業2万社の取引関係を復元した。企業間取引ネットワークは、公式データの55倍の取引関係を捕捉し、スケールフリー性を再現している。ボウタイ構造解析の結果、全体の約56%を占める巨大な強連結核 (SCC) を中心とする高度に集中した構造を有することが明らかとなった。レジリエンス解析では、SCC を含む高次数ハブを標的とした攻撃が、ランダムな障害に比べて著しく効率的にネットワークを分断することが示された。これは、中国のサプライチェーンのレジリエンスが個々の企業規模ではなく、取引ネットワークにおける中核的ハブ構造への依存によって規定されていることを示唆している。

**キーワード:** 企業間取引ネットワーク、サプライチェーン・レジリエンス、OSINT

**Abstract:** China's domestic supply chain is the foundation of the "world's factory," yet its structure remains opaque because approximately 83% of transaction partners in official corporate records are anonymized. To overcome this information asymmetry, this study develops an OSINT pipeline using Qwen 3 with DSPy-optimized prompting to reconstruct transaction relationships among 4,331 listed firms and 20,000 large enterprises. The resulting inter-firm transaction network captures 55 times more transaction relationships than official data and reproduces the scale-free topology characteristic of real-world networks. Bow-tie decomposition reveals a highly concentrated structure centered on a giant strongly connected core (SCC) comprising approximately 56% of the entire network. Resilience analysis further shows that targeted attacks on high-degree hub firms, including those within the SCC, fragment the network far more efficiently than random failures. These findings suggest that the resilience of China's supply chain is governed not by the scale of individual firms, but by its structural dependence on a small set of central hub nodes within the transaction network.

**Keywords:** Inter-firm Transaction Networks, Supply Chain Resilience, OSINT

### 1. はじめに

「世界の工場」の屋台骨と称される中国国内の企業間取引ネットワークは、近年の地政学的な情勢から世界から戦略的な注目を浴びているが、その実態は依然として極めて不透明である。特に、2025年12月に明らかになった中国の半導体自給自足構想や、電気自動車 (EV) 分野における BYD 等の国内企業の急速な台頭は、このネットワークへの関心を一層加速させている。

しかし、中国のサプライチェーンに関する従来の分析は、深刻な情報の非対称性という制約に直面している。中国証券監督管理委員会 (CSRC) が定める中国の上場企業に対する情報開示制度では、主要な仕入先・販売先に関する情報開示が求められているものの、実務上は Top5 の取引先に関する取引額や集中度の開示にとどまるケースが一般的であり、取引先名称については特定されない形で記載される例も少なくない[1]。実際、本研究で利用する上場企業の開示資料を収集・整理した一

次データベースである CSMAR[2] を用いた集計では、2022 年時点で約 83% の取引先名が匿名化されており、戦略的に重要な企業間関係の多くが「影のサプライチェーン (Shadow Supply Chain)」としてブラックボックス化している。

この情報の欠落は、学術的にも深刻な制約をもたらしている。この CSMAR 等の公式データベースに基づく従来のネットワークは、ノード(企業)数に対してエッジ(取引関係)数が下回る(平均次数が 1 未満)など、極めて疎で断片的な「孤立した島の集合体」に過ぎない。そのため、これまでの研究は、断片的な情報を繋ぎ合わせた局所的な事例研究 (Case Study) [3]に留まらざるを得ず、システム全体の構造やレジリエンスを数理的・統計的に俯瞰する「ネットワーク科学」的アプローチは事実上不可能であった。

そこで本研究では、この構造的制約を突破するため、ウェブ上に散在する公開情報から取引関係を復元する OSINT パイプラインを開発した。検索エンジンの要約情報 (スニペット) や対象ウェブサイトの情報と、DSPy によるプロンプト最適化[4]を施した大規模言語モデル (LLM) を統合することで、大規模な実名特定と取引関係抽出を実現した。本研究の目的は、不透明な中国市場において初めてネットワーク科学的解析を可能にする「データ基盤」を提示し、実社会のトポロジーに基づいたサプライチェーンの強靱性を定量的に評価することにある。

## 2. データセット構築

公的開示情報の不透明性を克服するため、本研究ではウェブ上に散在する取引関係の断片的な証拠を大規模かつ効率的に集約する「スニペット駆動型 OSINT パイプライン」を開発した。本解析の対象は、中国の上場企業 4,331 社、および営業収益が 1.3 億ドルを超える大手企業 20,000 社である。

本手法の工学的優位性は、検索エンジンの要約情報 (スニペット) を主体とした情報抽出を行う点にある。スニペットは検索アルゴリズムによって既に情報が高度に凝縮されているため、対象ウェブサイトのテキスト全文のみを対象とした大規模言語モデル LLM による情報抽出 (RAG 等) と比較して、抽出対象のデータボリュームを約 1,000 分の 1、入力トークン量を数百分の 1 へと大幅に圧縮することができる。これにより、数万社規模

のサプライチェーンを網羅的に復元することが、極めて高い計算効率で可能となった。

LLM を用いた情報抽出において課題となる出力の不安定性や再現性の欠如[5]に対しては、DSPy (Declarative Self-improving Language Programs) [4]を採用した。DSPy は、手動のプロンプト調整に依存せず、最適な「指示 (Instruction)」と「事例 (Few-shot)」の組み合わせを自動的に探索・コンパイルする手法であり、本研究では、Qwen3-Next-80B-A3B-Instruct を用いながら高度な抽出精度と再現性を両立させている。

さらに、抽出データの信頼性を担保するため、Gemini 3 Pro 等の強力なモデルを用いた「LLM-as-a-Judge」による妥当性検証プロセスを導入した。中国証券監督管理委員会 (CSRC) 等の当局公式サイトを検索対象に含めた検証の結果、既存の公式 DB (CSMAR) に記録されている取引関係の 84% を本手法のみで再現できることを確認した。本パイプラインは、この高い再現率を維持しつつ、相手側企業が開示する IR 資料やメディア報道等の外部情報を統合する「トライアングレーション (三角測量)」の手法により、公式データでは不可視であった数万件の潜在的な取引リンクの実名特定を可能にした。

## 3. ネットワーク統計量

本研究によって復元された中国国内の企業間取引ネットワークは、同国の内部サプライチェーン構造について、これまでにない詳細な俯瞰図を提供する。定量的には、政府開示などの公式資料に基づく従来のデータセット (CSMAR 等) と比較して、2024 年の取引関係を対象とした情報収集では、ユニーク企業数で約 26 倍、取引関係数 (リンク数) で約 55 倍という圧倒的なデータ量を捕捉することに成功した。

図 1 は、本手法で構築したネットワークと公式データにおける次数分布 (累積分布関数: CDF) の比較である。公式記録に基づくデータ (赤線) では、主要な仕入先・販売先に関する開示が Top5 の範囲に限定され、かつ取引先名称が匿名化されるという制度的・実務的制約を反映して、取引先数が仕入先数と販売先数を合計した約 10 社付近で分布が不自然に切断されている。この切断は、ネットワークの過度な疎化を招き、構造的特徴の推定や全体構造の理解を著しく困難にしてきた。これに対し、本 OSINT パイプラインで再構築された

ネットワークでは、様々な国のサプライチェーンネットワークに普遍的に観察されるべき乗則 (power-law) に従う次数分布が鮮明に再現されている[6,7]。その分布は  $\beta = -3$  に近い傾きを持つスケールフリー性を示しており、少数の巨大ハブ企業がネットワーク全体を牽引する実態を正確に捉えている。

特筆すべきは、ハブ企業の顔ぶれにおける「劇的な再定義」である。表 1 のように、公式データでは情報開示が比較的整っている国家电网 (State Grid: 次数 103) 等のインフラ系国有企業がハブの中心に位置するが、本手法では、表 2 のように華為 (Huawei: 次数 453) や比亞迪 (BYD: 次数 226) といった、現代のハイテク・製造業を牽引する民間巨人が真のハブとして浮上した。これは、公式統計が捕捉できていなかった「戦略的重要企業」の実態を、ウェブ上の断片的な情報を統合することで初めてネットワークとして可視化できたことを意味する。

さらに、ボウタイ分解 (Bow-tie decomposition) の結果、中国企業間取引ネットワークは明確なコア-ペリフェリー構造を持つことが確認された。強連結核 (SCC) は 68,775 ノードと全体の 56.2% を占める巨大な中核を形成しており、Huawei や BYD などの高次数ハブ企業が密に結合している。一方、SCC の上流には IN コンポーネント (11.2%)、下流には OUT コンポーネント (15.6%) が配置され、これらは比較的疎な構造を持つ。TUBES や TENDRILS に相当する中間構造はほとんど観測されず、強く結合した中核と周辺の流入・流出部分に二極化したボウタイ型ネットワークが形成されている。

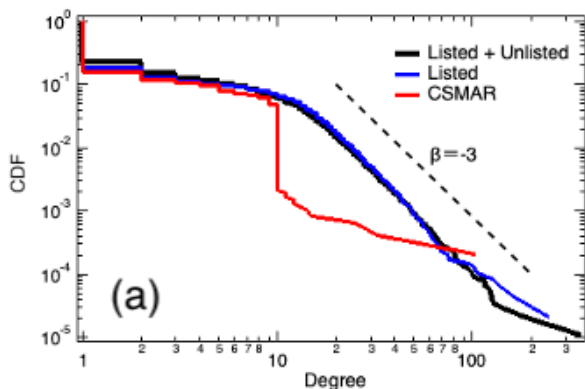


図 1 公式データと OSINT により再構築された企業間取引ネットワークの次数分布の比較 (CDF)

表 1 公式データ (CSMAR) に基づく無方向の取引ネットワークの主要ハブ企業 (上位 5 社)

順位	企業名	次数
1	国家电网 (State Gri)	103
2	中国石化 (Sinopec)	30
3	中国石油 (PetroChina)	25
4	南极人 (Nanji E-Commerce)	15
5	中国移动 (China Mobile)	14

表 2 OSINT により再構築された無方向の取引ネットワークにおける主要ハブ企業 (上位 5 社)

順位	企業名	次数
1	华为 (Huawei)	453
2	比亚迪 (BYD)	226
3	中国电信 (China Telecom)	151
4	深圳市讯商科技 (Shenzhen Xunshang Technology)	149
5	深圳市汇创达科技 (Shenzhen Hui Chuang Da Tech)	140

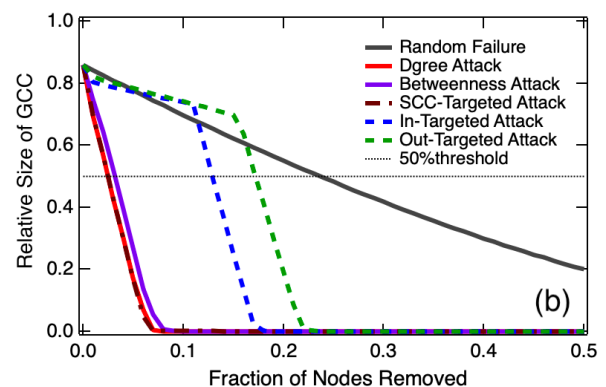


図 2 強連結核 (SCC) への標的型攻撃が中国企業間取引ネットワークの連結性に与える影響

#### 4. 中国サプライチェーンの頑強性

本章では、OSINT によって復元されたネットワークモデルを用い、標的型攻撃およびランダムな故障に対するシミュレーションを通じて、中国国内サプライチェーンの構造的脆弱性を検証する。図 2 は、ノードの除去割合 (Fraction of Nodes Removed) に対する、最大連結成分 (GCC: Giant Connected Component) の相対サイズの推移を示したレジリエンス・プロファイルである。

まず、ランダム故障に対しては、GCC は比較的緩やかに縮小しており、ノードの約 24%が除去された時点で初めて GCC が全体の 50%を下回る。

この結果は、中国サプライチェーンが偶発的な企業退出やランダムな障害に対しては一定の耐性を有していることを示唆している。

一方で、度数に基づく標的型攻撃 (degree attack) は極めて高い破壊力を示す。高次数ノードを優先的に除去した場合、わずか約 3% のノード除去で GCC は急激に崩壊し、ネットワーク全体の連結性がほぼ失われる ( $R = 0.0267$ )。この結果は、中国サプライチェーンが少数のハブ企業に強く依存した構造を持ち、その喪失が全体機能に致命的な影響を与えることを意味する。

強連結核 (SCC) への標的型攻撃も同様に高い脆弱性を示しており、GCC は約 3% の除去で急減する ( $R = 0.0268$ )。ただし、レジリエンス指標  $R$  の比較からは、SCC 攻撃は度数攻撃とほぼ同等であるものの、わずかに効果が劣る。すなわち、本ネットワークでは「SCC に属すること」自体よりも、「次数が極めて高いこと」が破壊効果を決定づけている。

これに対し、上流 (IN) および下流 (OUT) コンポーネントへの標的型攻撃では、GCC の崩壊は相対的に緩やかである。IN への攻撃では約 13% ( $R = 0.1079$ )、OUT への攻撃では約 18% ( $R = 0.1397$ ) のノード除去が必要であり、コア部分と比べて局所的な影響にとどまる。このことは、ネットワークの頑強性が周辺部分ではなく、中核ハブの集中度によって規定されていることを示している。

以上より、中国サプライチェーンは、ランダムなショックに対しては比較的頑強である一方、高次数ハブ企業を標的とした意図的・戦略的攻撃に対しては極端に脆弱であることが明らかとなった。この特性は、巨大な強連結核 (全体の約 56%) を有するにもかかわらず、ネットワーク全体の機能が少数の中心的企業に過度に依存しているという、中国サプライチェーンの構造的リスクを浮き彫りにしている。

## 5. まとめと課題

本研究では、極めて情報の非対称性が高い中国市場を対象に、ウェブスニペットと大規模言語モデル (LLM) を統合した OSINT パイプラインを構築し、公式統計では不可視であった「影のサプライチェーン」の実像をネットワーク科学的アプローチによって解明した。

主要な成果は以下の 3 点に集約される。第一に、スニペット駆動型マイニングと DSPy による

プロンプト最適化を組み合わせることで、従来の全文解析手法に比べ計算コストを数百分の 1 に抑制しつつ、公式 DB (CSMAR) の約 26 倍に相当する 122,314 ノード、247,613 エッジからなる大規模な企業間取引ネットワークの復元に成功した。第二に、構築されたネットワークが実世界のサプライチェーンに固有のスケールフリー性を備えていることを立証し、公式データでは隠蔽されていた華為 (Huawei) や比亞迪 (BYD) といったハイテク民間企業が、国家インフラをも凌駕する真のハブとして機能している実態を浮き彫りにした。第三に、ネットワークのレジリエンス解析を通じて、中国のサプライチェーンの脆弱性が個々のハブ企業の規模以上に、巨大な強連結核 (SCC) を含む高次数ハブ構造への依存によって規定されていることを解明した。

今後の課題として、第一に、時系列解析によるサプライチェーンの動態把握が挙げられる。本手法を継続的に適用することで、地政学的ショックや産業政策の変化に伴う企業間取引ネットワークの再編プロセスを観測することが可能となる。第二に、取引財レベルのサプライチェーン推定の導入である。現在は全ての取引関係を同一ネットワークとして統合しているが、実際には財の種類や加工工程に応じて取引経路は大きく異なる。財の加工・変換プロセスを明示的に考慮することで、より精緻なリスク波及シミュレーションや脆弱性評価が可能になると期待される。本研究で構築したデータ基盤は、不透明なグローバルサプライチェーンを対象とした経済安全保障およびレジリエンス分析の高度化に資するものである。

## 謝辞

本研究は、JSPS 科研費 JP25K01458、JP23H00042 および JST A-STEP JPMJTR25RL の支援を受けたものである。

## 文 献

- [1] Wang, X., Shan, Y. G., Song, J. (2023) Customer information disclosure and corporate financing constraints. *International Review of Financial Analysis* 89, 102822.
- [2] CSMAR: China Stock Market & Accounting Research Database. Shenzhen CSMAR Data Technology Co., Ltd. (Research database on Chinese listed firms, financials, and capital markets).
- [3] Ni, J., et al (2023) Customer concentration and financing constraints. *Journal of Corporate Finance* 82, 102432.

- [4] Khattab, O., et al (2023) DSPy: Compiling Declarative Language Model Calls into Self-Improving Pipelines. arXiv:2310.03714.
- [5] Xu, D., et al (2024) Large language models for generative information extraction: a survey. *Frontiers of Computer Science* 18, 186357.
- [6] Mizuno, T., Souma, W., Watanabe, T. (2014) The structure and evolution of buyer-supplier networks. *PLOS ONE* 9(7), e100712.
- [7] Mizuno, T., Ohnishi, T., Watanabe, T. (2026) Structure of global buyer-supplier networks and its implications for conflict minerals regulations. *EPJ Data Science* 5, 2.