Modeling Shock Propagation on a Corporate Network: A Feasibility Study and Research Blueprint

I. The Corporate Supply Web as a Complex System: Theoretical Foundations for Shock Propagation

The intricate web of customer-supplier relationships that underpins the modern economy can be conceptualized not as a random assortment of firms, but as a complex, structured network. This perspective, rooted in network science and econophysics, provides a powerful framework for understanding how localized economic shocks can propagate, creating systemic effects that may not be immediately apparent to market participants. This section establishes the theoretical foundations for a quantitative research project aimed at modeling and exploiting these propagation dynamics. It synthesizes literature from financial economics, network science, and epidemiology to argue that the corporate supply web is a medium for the gradual diffusion of information, where fundamental shocks can create predictable, albeit temporary, market inefficiencies.

1.1 Information Inefficiency and Predictable Returns in Economic Networks

The strongest form of the Efficient Market Hypothesis (EMH) posits that asset prices instantaneously and fully reflect all available information. However, a substantial body of research suggests that this is an idealized condition. In reality, information diffuses gradually through the market, particularly when its implications are complex or indirect.[1] This gradual diffusion process creates predictable lead-lag effects, where the price movements of certain assets can forecast the future price movements of others.

This phenomenon is not random; it is often structured by underlying economic and social networks. For instance, studies of broker-investor networks have shown that central brokers, by virtue of their position, gather information from informed trades and subsequently leak this information to their preferred clients, who can then trade on it before it is fully priced in.[2] Similarly, intra-industry studies reveal that value-relevant news about large, highly-visible firms tends to diffuse slowly to smaller, less-analyzed firms within the same sector, creating a

predictable return pattern.[1]
The central hypothesis of this research project is that the customer-supplier network represents a primary and powerful channel for this type of gradual information diffusion. A significant corporate event, such as a large earnings surprise, constitutes new, value-relevant information not only for the announcing firm but also for its direct economic partners. A positive earnings surprise for a major customer may signal increased future demand for its suppliers' products, while a negative surprise may signal the opposite. The market, however, may be slow to fully process these second-order implications, creating a window of opportunity. This concept is empirically supported by research demonstrating the existence of "supply chain alpha," where the performance of a company's customers and suppliers is shown to be predictive of its own stock returns and fundamentals.[3]

1.2 Mechanisms of Shock Propagation: Amplification, Persistence, and Decay

The academic literature provides robust evidence that economic networks act as conduits for shocks, which can propagate, persist, and even amplify as they traverse the system. Studies of major exogenous events, such as natural disasters, have shown that the indirect economic effects propagated through supply chains are often substantially larger than the direct physical damage itself.[4] This highlights the network's role as an amplification mechanism.
The specific dynamics of this propagation are moderated by both the network's topology and the characteristics of the individual firms (nodes) within it. Several key factors have been identified:

- **Network Centrality:** Firms or industrial sectors that are more central to the network—meaning they have more connections or are critical bridges between other parts of the network—tend to have a larger systemic impact when they are hit by a shock.[6] A shock originating at a hub can spread more widely and rapidly than one originating at the periphery.
- **Input Substitutability:** The propagation of shocks is more pronounced and persistent when the inputs a firm uses are highly specific and difficult to substitute. If a customer cannot easily switch to an alternative supplier, it is more vulnerable to disruptions affecting its current one.[4]
- **Financial Constraints:** The financial health of a firm dictates its response to a shock. Financially unconstrained firms may have the capacity (e.g., through cash reserves or access to credit) to absorb a shock, acting as a "firewall" in the network. In contrast, liquidity-constrained firms are more likely to transmit the shock downstream to their customers, for instance by tightening trade credit.[8]
- **Adjustment Frictions:** The economic relationships between firms are not frictionless. Forming new supply links or dissolving old ones involves search costs, negotiation, and time. These "adjustment frictions" can make the effects of shocks longer-lived, as firms cannot instantaneously reconfigure their network connections in response to a

disruption.[9]

The physics-based analogy of a shock diffusing through a lattice with a certain decay rate provides a useful, albeit simplified, mental model. It captures the core intuition that the "energy" of a shock dissipates as it travels further from its source. However, unlike a passive physical medium, the economic network is composed of active, strategic agents. Firms react to shocks based on their individual circumstances, leading to complex, non-linear propagation dynamics. This suggests that a simple linear diffusion model might be a first approximation, but a more sophisticated model should account for these moderating firm-level characteristics.

1.3 Analogous Frameworks: Information Diffusion and Financial Contagion

The propagation of an earnings surprise through the supply chain network can be formally modeled using frameworks adapted from other scientific domains that study contagion and diffusion phenomena.

- **Epidemiological Models:** The spread of an economic shock is frequently compared to the spread of an infectious disease, leading to the application of compartmental models from epidemiology, such as the Susceptible-Infected-Recovered (SIR) model.[6] In this analogy, firms in the network can be classified into states:
    - **Susceptible (S):** A firm that is economically linked to the shocked firm but has not yet shown a market reaction.
    - **Infected (I):** A firm that is currently exhibiting statistically significant abnormal returns as a result of the shock's propagation.
    - Recovered (R): A firm whose stock price has fully incorporated the new information and whose abnormal returns have reverted to zero.
      The power of this framework lies in its explicit modeling of a transmission rate ($ \beta)andarecoveryrate(\gamma $). In a financial context, $ \beta $ could be conceptualized as a function of the economic linkage strength between two firms, while $ \gamma $ could represent the speed of market efficiency in pricing the new information for a given stock.10 This approach moves beyond simple diffusion by modeling the probability and duration of the impact.
- **Financial Contagion Models:** A rich parallel exists in the literature on systemic risk, which studies how the failure of one financial institution can trigger a cascade of failures through a network of interbank liabilities or overlapping portfolios.[13] This field provides critical insights into the topological features that make a network fragile, such as the presence of highly connected hubs (heavy-tailed degree distributions), high clustering, and the trade-off between risk diversification and contagion channel creation.[15] An earnings surprise can be viewed as a "micro-shock" that propagates through similar network mechanisms, albeit with price fluctuations as the outcome rather than outright defaults.

The choice between these modeling frameworks is not merely technical; it reflects a deeper assumption about the nature of the propagation process. A physics-based diffusion model implies a more deterministic, collective process of information averaging across the network. An epidemiological or contagion model, in contrast, suggests a more stochastic, agent-based process where individual firms "catch" the information shock based on their exposure and characteristics. A comprehensive research project could frame this as a key question: Does the propagation of earnings surprises behave more like a physical diffusion process or a biological contagion process? Answering this would provide deeper insights into the microfoundations of this market inefficiency.

II. Architecting the Network: Data and Methodologies for Graph Construction

The foundation of any network-based analysis is the graph itself. The construction of a robust, point-in-time accurate representation of the US corporate supply web is the most critical and technically demanding phase of this project. This section provides a detailed blueprint for acquiring the necessary data and applying rigorous methodologies for graph construction, with a particular focus on the challenges of handling temporal dynamics and the strategic importance of selecting an appropriate edge weighting scheme.

2.1 Data Triad: Sourcing Network, Event, and Market Data

A successful implementation requires the integration of three distinct types of high-quality, research-grade data:

- **Network Data (e.g., customrelation_us):** The project's success hinges on a comprehensive dataset of customer-supplier relationships. While the query mentions a specific dataset name, the industry standards for academic and professional research in this area are:
  - **FactSet Revere Supply Chain Relationships:** This dataset provides a detailed mapping of business interconnections among global companies, sourced from primary documents like annual filings, investor presentations, and press releases. It offers the significant advantage of normalizing relationships into distinct categories (customer, supplier, partner, competitor) and providing relevance rankings. The data is updated weekly, allowing for a relatively high-frequency view of the network's evolution.[16]
  - **Bloomberg Supply Chain Analysis (SPLC):** This is another premier source, offering extensive coverage of both public and private companies globally, with historical relationship data dating back to 2006. A key feature of the SPLC dataset is the inclusion of proprietary, analyst-estimated relationship values, which can

provide a quantitative measure of link strength even when explicit revenue dependency figures are not disclosed.[20]

- **Event Data (IBES):** To identify the initial "shocks," it is necessary to quantify the surprise component of corporate earnings announcements. The Institutional Brokers' Estimate System (IBES) is the canonical source for historical consensus analyst earnings forecasts. This data allows for the calculation of Standardized Unexpected Earnings (SUE), the project's chosen measure for shock magnitude.[24]
- **Market Data (CRSP):** To measure the "shockwaves" or the market's reaction to the propagating information, comprehensive and accurate stock market data is required. The Center for Research in Security Prices (CRSP) is the gold standard for US academic research, providing daily and monthly security price, return, and volume data for all common stocks listed on the NYSE, AMEX, and NASDAQ exchanges. The CRSP database is meticulously maintained to be free of survivorship bias, including data for delisted firms, which is essential for rigorous backtesting. The CRSP/Compustat Merged Database is particularly powerful, as it provides a clean link between CRSP's market data and Compustat's fundamental company data, which can be used to analyze firm-level characteristics.[28]

2.2 Graph Construction: Defining Nodes, Edges, and Temporal Dynamics

The corporate network will be modeled as a directed graph, $G = (V, E)$, where the set of vertices V represents publicly traded US companies, and the set of directed edges E represents customer-supplier relationships. A directed edge from firm i to firm j, denoted $(i, j) \in E$, signifies that firm i is a supplier to firm j.

A critical methodological point is that this network is not static. Relationships are formed, evolve, and are terminated over time. Therefore, the set of edges is a function of time, $E(t)$. Constructing a single, static graph aggregated over the entire sample period would introduce severe lookahead bias, as the analysis would incorrectly assume that future relationships were known in the past. This would render any backtest of a trading strategy invalid.[30]

The only methodologically sound approach is to construct the network as a sequence of temporal snapshots, G_t. For any given event date t, the graph G_t must be constructed using only relationship data known to exist on or before that date. Datasets from FactSet and Bloomberg typically provide start and end dates for disclosed relationships, which enables this historical reconstruction.[17] This process of building a point-in-time correct graph for every day in a multi-year backtesting period is a significant data engineering challenge. It requires an efficient pipeline to process the raw relationship data and generate the appropriate adjacency matrix for each event, and it is likely to be the primary technical bottleneck in a short-term project.[31]

2.3 The Criticality of Connections: A Comparative Analysis of Edge Weighting Schemes

The transmission probability or decay rate in a diffusion model is fundamentally dependent on the strength of the connection between two firms. Treating all connections as equal (i.e., using an unweighted graph) discards a vast amount of information. The choice of an edge weighting scheme is therefore a primary modeling decision that encapsulates a hypothesis about what drives the strength of economic influence. The literature and data sources suggest several distinct approaches, each with a trade-off between economic intuition and data availability.

A systematic comparison is essential for making an informed choice. The following table outlines the primary candidate schemes.

Scheme Name	Description	Mathematical Formulation	Data Requirements	Pros	Cons	Relevant Literature
Binary (Unweighted)	A simple binary indicator of a relationship's existence.	$ W_{ij} = 1 $ if link exists, $ 0 $ otherwise.	Basic relationship data (supplier, customer).	Simple, maximum data coverage.	Ignores link strength, least economically nuanced.	Baseline models
Revenue Dependency	Weight is the percentage of the supplier's revenue derived from the customer.	$ W_{ij} = \frac{\text{Sales from } i \text{ to } j}{\text{Total Revenue of } i} $	Disclosed revenue dependency figures.	Most economically intuitive measure of supplier dependence.	Data is sparsely disclosed, severely limiting network coverage.	[34]
Cost of Goods Sold (COGS) Dependency	Weight is the percentage of the customer's COGS attributable to the supplier.	$ W_{ij} = \frac{\text{Purchases by } j \text{ from } i}{\text{Total COGS of } j} $	Disclosed COGS dependency figures.	Economically intuitive measure of customer dependence.	Data is also sparsely disclosed.	[34]
Network Centrality (Edge Betweenness	Weight is the number of shortest paths	$ W_{ij} = \sum_{s \neq v \neq t} \frac{\sigma_	Full network topology.	Captures the edge's structural importance	Computationally intensive for large, dense	[34]

s)	between all node pairs in the network that pass through the edge (i, j).	{st}(i, j)}{\sigma_{st}} $		or "chokepoint" status. Can act as a proxy for importance when financial data is missing.	networks. Less direct economic meaning.	
PageRank-based Importance	Inspired by Google's PageRank, this scheme assigns an importance score to each node, and this importance flows through the edges.	Iterative calculation based on the importance of connected nodes.	Full network topology.	Captures systemic importance within the entire supply web.	Complex to implement; interpretation can be less direct than financial weights.	[3]
Analyst Coverage Overlap	Weight is the number of equity research analysts that cover both the supplier and the customer firm.	$ W_{ij} =	\text{Analysts}(i) \cap \text{Analysts}(j)	$	Analyst coverage data (e.g., from IBES).	Identifies firms that are perceived as economically linked by market experts.

This comparison reveals a critical trade-off. Revenue dependency is the most theoretically sound measure of a supplier's vulnerability to a customer shock, but its sparse availability makes it difficult to use for a broad, systematic strategy. Conversely, network centrality measures have full coverage (as they only require the network topology) and have been shown to be effective proxies for economic importance.[34] A robust research design could involve a multi-stage approach: first, validate the shock propagation phenomenon on the small subset of links where revenue data is available. Second, test whether centrality measures are correlated with these revenue dependencies. Finally, use the most predictive and scalable weighting scheme (likely a centrality-based one) for the full-universe backtest, with the initial analysis providing justification for its use as a valid proxy.

III. Modeling the Shockwave: A Diffusion Framework for Earnings Surprises

With the network constructed and the data sources identified, the next step is to develop a quantitative model that formally links the initial shock to its propagated response. This section details the process of defining the impulse and measuring the response, and then presents two distinct but related mathematical frameworks—one from physics and one from epidemiology—to model the propagation dynamics.

3.1 The Impulse: Quantifying Earnings Shocks with Standardized Unexpected Earnings (SUE)

The initial "impulse" that perturbs the system is a significant corporate earnings announcement. An earnings surprise occurs when a company's reported earnings per share (EPS) deviates meaningfully from the market's expectation, as captured by the consensus forecast of financial analysts.[24]

To create a standardized and comparable measure of this surprise across different companies and time periods, the Standardized Unexpected Earnings (SUE) score is the most appropriate metric. The SUE score quantifies the magnitude of the surprise in terms of the number of standard deviations it represents relative to the dispersion of analysts' forecasts. This accounts for the fact that a $0.05 surprise for a stable utility company is more significant than the same surprise for a volatile technology startup.[24]

The formula for SUE for firm i at the time of its quarter Q announcement is:

SUEi,Q=SDi,QAi,Q−Fi,Q

where:

- $A_{i,Q}$ is the actual EPS reported by the firm for quarter Q.
- $F_{i,Q}$ is the consensus (e.g., mean or median) EPS forecasted by analysts.
- $SD_{i,Q}$ is the standard deviation of the analysts' individual forecasts for quarter Q.

Data for $A_{i,Q}$, $F_{i,Q}$, and $SD_{i,Q}$ are sourced from the IBES database. A large absolute SUE value (e.g., $|SUE| > 3$) for a firm constitutes the event trigger—the initial energy impulse applied to a node in the corporate network.

3.2 The Response: Measuring Propagation via Event-Time Abnormal

Returns

To measure the "shockwave" as it propagates through the network, a standard event study methodology is employed. This technique is designed to isolate the component of a stock's return that is attributable to a specific event, filtering out general market movements.[37] The analysis is performed not only for the firm announcing the earnings surprise (the source node) but also for its suppliers and customers (the neighbor nodes) at varying network distances.
The key steps are as follows:

1. **Define the Event and Event Window:** The event is the earnings announcement date for the source firm, designated as day $t=0$. The event window is the period over which the returns are measured, for example, from $t=0$ to $t=+5$ trading days.
2. Calculate Normal (Expected) Returns: The expected return for each stock i on each day t in the event window, E, must be estimated. This is the return that would have been expected in the absence of the event. A common approach is the market model, which regresses the stock's historical returns against the returns of a broad market index (e.g., the S&P 500) over an estimation window (e.g., $t=-120$ to $t=-21$).

   $$R_{i,t}=\alpha_i+\beta_i R_{m,t}+\epsilon_{i,t}$$

   The normal return is then $E = \hat{\alpha}_i + \hat{\beta}_i R_{m,t}$. For greater robustness, a multi-factor model like the Fama-French Three-Factor Model can be used, which controls for size and value effects in addition to market risk.38
3. Calculate Abnormal Returns (AR): The abnormal return is the difference between the actual return and the expected return for each day in the event window.

   $$AR_{i,t}=R_{i,t}-E$$
4. Calculate Cumulative Abnormal Returns (CAR): To measure the total impact of the event over the window, the daily abnormal returns are summed. The CAR for firm i from day t_1 to t_2 is:

   $$CAR_i(t_1,t_2)=\sum_{t=t_1}^{t_2} AR_{i,t}$$

   This CAR value is the primary measure of the shockwave's magnitude at any given node in the network. By plotting the average CAR for nodes at network distance 1, 2, 3, etc., from the source, one can visualize the propagation and decay of the shock.

3.3 A Physics-Based Formalism: The Heat Equation on the Corporate Graph

The analogy of heat diffusion can be formalized using the mathematical framework of the heat equation on a graph. This model describes how a scalar quantity distributed across the nodes of a graph (in this case, the "shock energy" represented by CAR) evolves over time as it flows along the edges.[39]

The discrete form of the heat equation on a graph is a system of ordinary differential equations:

$$\frac{\partial \mathbf{u}(t)}{\partial t} = -\kappa \mathbf{L} \mathbf{u}(t)$$

where:

- $\mathbf{u}(t)$ is a vector of size $|V| \times 1$, where $u_i(t)$ is the CAR of firm i at time t after the initial shock. The initial condition is $\mathbf{u}(0)$, where only the source node has a non-zero CAR.
- κ is a positive scalar representing the thermal diffusivity or the overall speed of propagation through the network.
- $\mathbf{L}$ is the Graph Laplacian matrix, a fundamental object in spectral graph theory that captures the connectivity of the graph. It is defined as:

  $$\mathbf{L} = \mathbf{D} - \mathbf{W}$$

  where $\mathbf{W}$ is the weighted adjacency matrix of the network (with weights determined by one of the schemes in Section 2.3), and $\mathbf{D}$ is the diagonal degree matrix, where $D_{ii} = \sum_j W_{ij}$ (the total weight of outgoing edges from node i).

The term $-\mathbf{L} \mathbf{u}(t)$ calculates, for each node, the weighted difference between its neighbors' CARs and its own CAR. The model thus predicts that the CAR at a node will change in proportion to this difference, leading to a "smoothing" of the initial shock across the network over time. The research would involve fitting the empirically observed CAR propagation patterns to this differential equation to estimate the diffusion coefficient κ and test the model's goodness-of-fit.

3.4 Alternative Perspectives: Epidemiological Models of Financial Contagion

A simple diffusion model assumes a smooth, continuous flow of information. An alternative, and potentially more realistic, view is that information propagates in a more discrete, probabilistic manner, akin to a contagion process. The SIR model from epidemiology provides a powerful framework for this perspective.[11]

The financial analogy is as follows:

- **Susceptible (S):** A firm in the network that is economically linked to the source of the shock but has not yet exhibited a significant market reaction.

- **Infected (I):** A firm that is currently experiencing statistically significant abnormal returns. A firm transitions from S to I with a probability β, which can be modeled as a function of its connection strength to already infected neighbors.
- **Recovered (R):** A firm whose abnormal returns have subsided, indicating the market has priced in the new information. A firm transitions from I to R with a probability γ, which can be thought of as an inverse measure of the duration of the market inefficiency.

This model is described by a system of differential equations governing the proportion of firms in each state. Unlike the heat equation, the SIR model can capture non-linear dynamics, such as the initial exponential growth of an "epidemic" of abnormal returns if the "basic reproduction number" ($R_0 = \beta / \gamma$) is greater than one. This framework is particularly well-suited to test for heterogeneity in firm responses; for example, some firms might be modeled as "immune" to the shock due to strong financial health or low analyst coverage.

A crucial aspect to investigate is the asymmetry of propagation. Research on supplier disclosures indicates that bad news has a much stronger and more significant impact on customer stock prices than good news.[43] Similarly, macroeconomic literature suggests that demand-side shocks (relevant to suppliers) propagate upstream, while supply-side shocks (relevant to customers) propagate downstream.[8] An earnings surprise contains elements of both. Therefore, the modeling and subsequent analysis must explicitly test for these asymmetries by running separate analyses for positive versus negative SUE shocks and for propagation to suppliers versus customers. This will reveal the dominant channels and non-linearities in the shock propagation process.

IV. From Propagation Dynamics to Market Alpha: A Quantitative Trading Strategy

The ultimate goal of this research project is to determine if the observed shock propagation dynamics are sufficiently predictable and robust to be exploited for profit. This section translates the descriptive models from the previous section into a prescriptive, fully specified quantitative trading strategy, covering signal generation, portfolio construction, and execution logic.

4.1 Signal Generation: Translating Network Proximity and Shock Magnitude into Actionable Signals

The trading signal is the core of the strategy, converting the observation of a shock event into a specific trading decision.

- Event Trigger: The process begins with the identification of a significant earnings surprise. A trigger is fired for firm i at time t if its SUE score exceeds a predefined threshold.

  $$|SUE_{i,t}| > Threshold (e.g., Threshold=3.0)$$
- **Baseline Signal Logic:** The fundamental hypothesis is that the shock will propagate to the firm's suppliers.
  - If $SUE_{i,t} > 3.0$ (a large positive surprise), this signals potentially higher future demand for its suppliers. A **BUY** signal is generated for the set of firms that are direct suppliers to firm i.
  - If $SUE_{i,t} < -3.0$ (a large negative surprise), this signals potentially lower future demand. A **SELL/SHORT** signal is generated for the set of firm i's suppliers.
- **Signal Refinements and Testable Hypotheses:** The baseline signal can be refined with several layers of sophistication to potentially enhance its efficacy.
  - **Network Distance Decay:** The shock's influence is expected to weaken with network distance. The strategy should be tested by trading only 1st-degree neighbors (direct suppliers) versus a more complex version that includes 2nd-degree neighbors, perhaps with a reduced signal weight.
  - **Customer vs. Supplier Propagation:** As discussed in Section III, the direction of propagation is a key research question. The strategy should be tested in three configurations: trading suppliers only, trading customers only, and trading both, to empirically determine which linkage type provides the most reliable signal.
  - Signal Strength Formulation: A binary buy/sell signal is simplistic. A more nuanced signal strength for a neighboring firm j can be formulated as a function of the initial shock's magnitude and the strength of the economic link. A potential formulation is:

    $$Signal_j = f(SUE_{i,t}, W_{ij})$$

    where W_{ij} is the weight of the edge from the shocked firm i to the neighbor j.
  - **The Concentration Hypothesis:** A particularly promising refinement, as suggested by the user query, is that the effect should be stronger for suppliers that are more dependent on the shocked customer. This can be operationalized by making the signal strength inversely proportional to the supplier's customer base (its out-degree in the network). A supplier with few other customers is more exposed to news from a major client.[35] The signal could be modified as:
    $$Signal_j = Out\text{-}Degree_j f(SUE_{i,t}, W_{ij})$$

    This moves the strategy from simply "buy the supplier" to "preferentially buy the most dependent suppliers," a critical nuance that could be a significant source of alpha.

4.2 Portfolio Construction and Weighting Tactics

On any given day, multiple earnings surprises may occur, generating a set of buy and sell signals. These signals must be systematically combined into a tradable portfolio.

- **Portfolio Structure:** A long-short, market-neutral portfolio is the standard approach for testing a specific alpha factor. This structure aims to isolate the strategy's returns from broad market movements. The long book will consist of all securities with active BUY signals, and the short book will consist of all securities with active SELL signals. The total market value of the long and short positions should be kept roughly equal.
- **Position Weighting:** Within the long and short books, several weighting schemes should be tested:
  - **Equal Weighting:** The simplest method, where each position receives an equal capital allocation.
  - **Signal Weighting:** A more sophisticated approach where the size of each position is proportional to the calculated signal strength ($ \text{Signal}_j $). This allocates more capital to the most promising opportunities.
  - **Risk Parity:** This technique allocates capital based on risk contribution rather than dollar value, preventing the portfolio from being dominated by a few highly volatile stocks.[44]
  - **Factor Neutralization:** To ensure the captured returns are genuinely from the supply chain effect and not from exposure to known risk factors, the portfolio's beta to the market, and its exposure to factors like size (SMB), value (HML), and momentum (UMD), should be neutralized through optimization.[38]

4.3 Execution Logic: Holding Periods, Rebalancing, and Risk Overlays

The final component of the strategy specification deals with the practicalities of execution and risk management.

- **Holding Period:** The optimal holding period for each position is an empirical question that can be answered by the event study analysis (Section 3.2). If the analysis shows that the Cumulative Abnormal Return for neighbor firms tends to peak and then revert after, for example, 3-5 trading days, this would suggest a short-term holding period of the same duration.
- **Rebalancing:** The portfolio is dynamic. It should be rebalanced on a daily basis to incorporate new signals from that day's earnings announcements and to exit positions that have reached the end of their predetermined holding period.
- **Risk Management:** Robust risk management is essential for any quantitative strategy.
  - **Position Sizing:** No single position should exceed a small percentage of the total portfolio value to ensure diversification.

    - **Stop-Losses:** While potentially interfering with the pure signal, stop-loss orders on individual positions can be implemented to cap the maximum loss on any single trade.
    - **Exposure Limits:** Strict limits should be maintained on the portfolio's gross exposure (total value of long and short positions) and net exposure (difference between long and short value, which should be close to zero for a market-neutral strategy).

A critical consideration for this type of short-term, event-driven strategy is its **capacity**. The alpha is derived from a temporary market inefficiency that will likely be arbitraged away as more capital is deployed to trade it. The strategy will have high turnover and be sensitive to transaction costs like commissions and slippage.[44] Therefore, any realistic backtest must rigorously model these costs. The analysis should also include an estimation of the strategy's capacity by examining the typical daily trading volume of the target securities to understand how large a position could be taken without significant market impact.

V. Empirical Validation and Methodological Rigor

A profitable backtest is a necessary but not sufficient condition for a viable quantitative strategy. The history of quantitative finance is littered with strategies that performed spectacularly on historical data but failed in live trading. This failure is almost always due to methodological flaws, such as overfitting or hidden biases in the testing process. This section outlines a framework for rigorous empirical validation, with a strong emphasis on identifying and mitigating the most common statistical traps.

5.1 A Framework for Robust Backtesting

Backtesting is the process of simulating a trading strategy on historical data to estimate how it would have performed.[45] A robust backtesting framework requires several key components:

- **Point-in-Time Data:** The simulation must be conducted with strict adherence to point-in-time data availability. This means using the temporally reconstructed network graphs from Section 2.2, ensuring that for any trade decision on day t, only information that was actually available to the public on or before day t is used.
- **Realistic Cost Modeling:** The simulation must account for real-world trading frictions. This includes:
    - **Commissions:** A reasonable estimate of per-trade commission costs.
    - **Slippage:** The difference between the expected execution price and the actual execution price. For a short-term strategy trading on news, slippage can be a significant cost, especially in less liquid stocks. It can be modeled as a percentage of the bid-ask spread or based on historical trade volume.
- **In-Sample vs. Out-of-Sample Testing:** This is the most critical principle for avoiding

overfitting. The historical data should be split into at least two distinct periods:
    - **In-Sample Period (e.g., 2006-2018):** This period is used for model development, parameter tuning, and hypothesis testing (e.g., determining the optimal SUE threshold or holding period).
    - **Out-of-Sample Period (e.g., 2019-2023):** This period is held out and used only once, at the very end of the research process, to validate the final, chosen strategy. A strategy that performs well in-sample but fails out-of-sample is likely overfit to historical noise and has no genuine predictive power.[47]

5.2 Navigating Statistical Traps: Mitigating Lookahead Bias, Data Snooping, and Survivorship Bias

Quantitative research is highly susceptible to subtle statistical biases that can produce deceptively positive results. Proactively identifying and mitigating these is paramount.

- **Lookahead Bias:** This is the error of using information in a simulation that would not have been available at the time of the decision.[30]
    - **Network Structure:** The most obvious source in this project is using a future version of the supply chain network. This is mitigated by the rigorous temporal network construction outlined in Section 2.2.
    - **Event Data:** A subtle form can arise from mishandling announcement times. An earnings release after market close on day $ t $ should only be tradable on day $ t+1 $.
    - **Fundamental Data:** If the strategy is refined using firm characteristics (e.g., leverage, inventory levels from Compustat), it is crucial to account for the significant reporting lag of this data. A company's Q2 financial statements may not be publicly filed until mid-August. Using this data to inform a trade in July is a form of lookahead bias. All fundamental data must be lagged to reflect its true public availability date.
- **Data Snooping (p-hacking):** This bias results from testing too many different hypotheses on the same dataset. By pure chance, some variations of a strategy will appear to be statistically significant. The researcher may then be tempted to report only the best-performing variation, creating a false impression of a robust effect.[50]
    - **Mitigation:** The best defense is a strong, pre-defined primary hypothesis and strict out-of-sample validation. If a strategy's parameters are heavily tuned to work in-sample, it is almost guaranteed to fail out-of-sample. If many strategy variations are formally tested (e.g., 10 different edge weighting schemes), statistical corrections for multiple testing, such as the Bonferroni correction, should be applied to significance tests.
- **Survivorship Bias:** This bias occurs if the historical dataset used for testing excludes firms that have been delisted due to bankruptcy or acquisition. Since failing firms are systematically excluded, the average performance of the remaining firms will be

artificially inflated.

  - **Mitigation:** This is primarily a data quality issue. Using high-quality, research-grade databases like CRSP, which explicitly include data for delisted securities, is the standard and necessary solution to this problem.[28]

5.3 Interpreting Results: Performance Metrics and Factor Analysis

The output of the backtest should be evaluated using a comprehensive suite of performance and risk metrics.[47]

- **Standard Performance Metrics:**
  - **Cumulative Return:** The total growth of the portfolio over the backtest period, often visualized as an equity curve.
  - **Annualized Return and Volatility (Standard Deviation):** The geometric average annual return and its variability.
  - **Sharpe Ratio:** The primary measure of risk-adjusted return, calculated as (Annualized Return - Risk-Free Rate) / Annualized Volatility.
  - **Maximum Drawdown:** The largest peak-to-trough decline in the portfolio's value, a key measure of downside risk.
  - **Win Rate and Profit Factor:** The percentage of trades that are profitable and the ratio of gross profits to gross losses.
- Factor Analysis: To demonstrate that the strategy has discovered a new source of alpha, its returns must be shown to be independent of known risk factors. This is done by running a time-series regression of the strategy's daily or monthly returns ($ R_{strat} $) against the returns of standard academic risk factors:
  $$ R_{strat,t} - R_{f,t} = \alpha + \beta_1 (R_{m,t} - R_{f,t}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 UMD_t + \epsilon_t $$
  A positive and statistically significant intercept, $ \alpha $, indicates that the strategy is generating excess returns that are not explained by exposure to the market, size, value, or momentum factors. This is the standard academic definition of true alpha.

To provide the strongest possible evidence that the network structure itself is the source of alpha, the analysis should include a comparison against a null hypothesis benchmark. This involves running the identical trading strategy, but on a set of randomly rewired networks. By creating thousands of randomized graphs that preserve key properties of the original network (like the number of nodes and the distribution of degrees) and backtesting the strategy on each, one can generate a distribution of performance metrics (e.g., Sharpe Ratios) under the null hypothesis that the specific wiring of the network does not matter. If the Sharpe Ratio achieved on the *true* historical network is an extreme outlier in this distribution (e.g., in the 99th percentile), it provides powerful statistical evidence that the strategy is indeed capturing a real phenomenon related to the specific topology of the corporate supply web.

VI. Synthesis and Project Roadmap

This report has established the theoretical soundness, empirical grounding, and methodological blueprint for a quantitative research project on shock propagation in corporate supply networks. The core hypothesis—that large earnings surprises diffuse predictably through customer-supplier linkages, creating a tradable lead-lag effect—is well-supported by parallel streams of academic literature. The primary challenges are not conceptual but technical, centered on the rigorous construction of temporal network data and the avoidance of statistical biases in backtesting. This final section provides a high-level feasibility assessment, a practical plan for a two-week research sprint, and outlines promising avenues for future investigation.

6.1 Feasibility Assessment and Abridged Plan for a Two-Week Sprint

The proposed project is ambitious but feasible within a concentrated two-week timeframe, provided that access to the necessary datasets is secured in advance and the researcher possesses strong data engineering and programming skills. The project's success will depend on efficient execution, particularly in the data-intensive first week.

Week 1: Data Engineering and Network Architecture

The goal of the first week is to build the foundational data asset: a point-in-time correct temporal network graph.

- **Days 1-2: Data Ingestion and Alignment.**
  - Secure access to and begin downloading the required data from FactSet/Bloomberg, IBES, and CRSP.
  - The immediate task is to align company identifiers across the three datasets (e.g., mapping FactSet IDs and CUSIPs to CRSP PERMNOs). This is a non-trivial task that is essential for linking the network structure to event and market data.
- **Days 3-5: Temporal Network Pipeline Construction.**
  - This is the most critical phase of the project. The focus should be exclusively on writing and debugging a robust data pipeline that can take the raw, timestamped relationship data (e.g., "Company A was a supplier to Company B from 2010-01-15 to 2015-08-30") and generate the correct adjacency matrix for any given historical date.
  - The output should be a function or class that, given a date, returns the corresponding network graph. This function will be the cornerstone of the backtesting engine.

Week 2: Modeling, Strategy Implementation, and Analysis

With the data infrastructure in place, the second week is dedicated to analysis and testing.

- **Days 6-8: Shock Propagation Analysis.**
  - Implement the SUE calculation using IBES and CRSP data to identify all historical

shock events.
    - Code the event study methodology to calculate CARs for the shocked firms and their 1st and 2nd-degree neighbors.
    - Generate the key descriptive plots: average CAR vs. time for different network distances. This will provide the first visual evidence of the propagation effect and inform the choice of holding period for the trading strategy.
- **Days 9-11: Strategy Backtesting.**
    - Code the baseline trading strategy as defined in Section IV, including signal generation, portfolio construction (long-short, market-neutral), and execution logic.
    - Run the backtest on the in-sample period, incorporating realistic transaction costs.
- **Days 12-14: Validation and Documentation.**
    - Run the finalized strategy on the out-of-sample data.
    - Perform the factor regression analysis to calculate the strategy's alpha.
    - Analyze the performance metrics and document all findings, methodologies, and assumptions.

6.2 Key Challenges and Areas for Further Investigation

Even with a robust plan, several challenges and limitations must be acknowledged:

- **Data Quality and Coverage:** Supply chain data is inherently incomplete. Disclosures are voluntary and biased towards larger, more significant relationships. The resulting network will be a partial, not complete, representation of the true supply web. The impact of these missing links on the model's conclusions should be considered.
- **Confounding Events:** Earnings announcements are major corporate events and are often accompanied by other material news releases. Disentangling the market's reaction to the supply chain implication from its reaction to other concurrent news is a significant challenge in event study analysis.
- **Model Specification Risk:** The project involves numerous modeling choices: the SUE threshold, the edge weighting scheme, the holding period, the mathematical propagation model (diffusion vs. contagion), etc. Each of these choices is a form of model specification, and the final results may be sensitive to them. Robustness checks, where key parameters are varied, are essential.

6.3 Extending the Research: Future Directions in Network-Based Alpha

This two-week project serves as a foundational exploration. The framework developed can be extended in numerous promising directions to build a more comprehensive research agenda in network-based alpha generation.

- **Exploring Alternative Shocks:** The same propagation framework can be applied to other types of localized, value-relevant corporate events. Examples include M&A announcements (how does the acquisition of a major supplier affect its other customers?), announcements of major FDA drug trials, significant product recalls, or dividend cuts.
- **Investigating Alternative Networks:** The customer-supplier network is just one of many inter-firm networks that can mediate the flow of information and economic influence. The same methodology could be used to investigate shock propagation through other network structures, such as:
  - **Patent Citation Networks:** Do technological breakthroughs for one firm predict positive returns for firms that cite its patents?
  - **Board Interlocks:** Do shocks to a firm propagate to other firms that share a common board member?
  - **Institutional Co-ownership Networks:** Does a large sale of a stock by a major institution create price pressure on other stocks that are heavily co-owned by the same set of institutions?
- **Applying Advanced Models:** The physics-based and epidemiological models proposed here impose a specific functional form on the propagation dynamics. A more data-driven approach would be to use advanced machine learning models, particularly Graph Neural Networks (GNNs). A GNN could be trained to learn the complex, non-linear patterns of shock propagation directly from the historical data, potentially capturing more nuanced effects than a pre-specified mathematical model.[32] This represents the state-of-the-art in network-based predictive modeling and is a natural next step for this line of research.

Works cited

1. Information diffusion in the U.S. real estate investment trust market - CentAUR, accessed July 30, 2025, https://centaur.reading.ac.uk/79273/1/InfoDiffusionREIT-2ndrevision.pdf
2. NBER WORKING PAPER SERIES THE RELEVANCE OF BROKER NETWORKS FOR INFORMATION DIFFUSION IN THE STOCK MARKET Marco Di Maggio Frances, accessed July 30, 2025, https://www.nber.org/system/files/working_papers/w23522/w23522.pdf
3. The Logistics of Supply Chain Alpha - Long Finance, accessed July 30, 2025, https://www.longfinance.net/documents/1237/DB_TheLogisticsofSupplyChainAlpha_2015.pdf
4. Propagation of economic shocks through supply chains - CEPR, accessed July 30, 2025, https://cepr.org/voxeu/columns/propagation-economic-shocks-through-supply-chains
5. Using Equity Market Reactions and Network Analysis to Infer Global Supply Chain Interdependencies, accessed July 30, 2025,

https://www.hkma.gov.hk/media/eng/publication-and-research/research/research-memorandums/2020/RM04-2020.pdf
6. Propagation of economic shocks in input-output networks: A cross ..., accessed July 30, 2025, https://link.aps.org/doi/10.1103/PhysRevE.90.062812
7. (PDF) Firm-level propagation of shocks through supply-chain networks, accessed July 30, 2025, https://www.researchgate.net/publication/335133551_Firm-level_propagation_of_shocks_through_supply-chain_networks
8. Financial Constraints and Propagation of Shocks in Production ..., accessed July 30, 2025, https://www.aeaweb.org/conference/2018/preliminary/paper/T36iZ9Gf
9. Production Network Dynamics and the Propagation of Shocks - Department of Economics, accessed July 30, 2025, https://economics.nd.edu/assets/303825/federico_huneeus_compressed.pdf
10. Unlocking Network Insights with SIR Model, accessed July 30, 2025, https://www.numberanalytics.com/blog/unlocking-network-insights-sir-model
11. Using SIRS Model to Study the Risk of Cross ... - Atlantis Press, accessed July 30, 2025, https://www.atlantis-press.com/article/125982715.pdf
12. Understanding SIR Model in Epidemiology - Number Analytics, accessed July 30, 2025, https://www.numberanalytics.com/blog/ultimate-guide-to-sir-model-in-infectious-disease-epidemiology
13. Understanding Financial Contagion: A Complexity Modeling PerspectiveThis article will be a contributed chapter to the SFI edited volume: The Economy as a Complex Evolving System, Part IV - arXiv, accessed July 30, 2025, https://arxiv.org/html/2502.14551v1
14. Network VAR models to Measure Financial Contagion, accessed July 30, 2025, https://economiaemanagement.dip.unipv.it/sites/dip10/files/2022-04/DEMWP0178.pdf
15. Contagion in Financial Networks* - Office of Financial Research (OFR), accessed July 30, 2025, https://www.financialresearch.gov/working-papers/files/OFRwp-2015-21_Contagion-in-Financial-Networks.pdf
16. Supply Chain Data - Climate Action, accessed July 30, 2025, https://engage.climateaction.org/widget/event/sustainable-investment-forum-north-america-2021/product/UHJvZHVjdF80ODY4MTQ=
17. FactSet - WRDS - University of Pennsylvania, accessed July 30, 2025, https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/factset/
18. FactSet Supply Chain Relationships - Databricks Marketplace, accessed July 30, 2025, https://marketplace.databricks.com/details/5172c774-1978-47f1-81b0-43c334c29cff/FactSet_FactSet-Supply-Chain-Relationships
19. FactSet Revere Supply Chain Relationships | Baker Library - Harvard Business School, accessed July 30, 2025, https://www.library.hbs.edu/databases-cases-and-more/datasets/factset-revere-supply-chain-relationships

20. Bloomberg SPLC Data for Apple (APPL) Quantified Suppliers as of 12/31/2018, accessed July 30, 2025, https://www.researchgate.net/figure/Bloomberg-SPLC-Data-for-Apple-APPL-Quantified-Suppliers-as-of-12-31-2018_tbl2_344841700
21. Discover company interlocking relationships. - Bloomberg ..., accessed July 30, 2025, https://data.bloomberglp.com/professional/sites/10/125966_BBGT_QUANT_SupplyChain_SFCT_DIG-1.pdf
22. Exploring supply chain data on Bloomberg and Workspace - Cranfield University Blogs, accessed July 30, 2025, https://blogs.cranfield.ac.uk/library/supply-chain-bloomberg-workspace/
23. (PDF) Bloomberg Supply Chain Analysis: A Data Source for Investigating the Nature, Size, and Structure of Interorganizational Relationships - ResearchGate, accessed July 30, 2025, https://www.researchgate.net/publication/344841700_Bloomberg_Supply_Chain_Analysis_A_Data_Source_for_Investigating_the_Nature_Size_and_Structure_of_Interorganizational_Relationships
24. Standardized Unexpected Earnings - University of West Georgia, accessed July 30, 2025, https://www.westga.edu/~bquest/2002/unexpected.htm
25. Unexpected Earnings - Overview, SUE Formula, Importance - Corporate Finance Institute, accessed July 30, 2025, https://corporatefinanceinstitute.com/resources/accounting/unexpected-earnings/
26. A New Measure of Earnings surprises and Post-Earnings-Announcement Drift - Brandeis, accessed July 30, 2025, https://peeps.unet.brandeis.edu/~heidifox/ese.pdf
27. corporatefinanceinstitute.com, accessed July 30, 2025, https://corporatefinanceinstitute.com/resources/accounting/unexpected-earnings/#:~:text=%E2%80%9CUnexpected%20earnings%E2%80%9D%20is%20the%20term,as%20an%20%E2%80%9Cearnings%20surprise.%E2%80%9D
28. Center for Research in Security Prices, LLC (CRSP) - WRDS, accessed July 30, 2025, https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/center-for-research-in-security-prices-crsp/
29. CRSP Stock Price Quote | Morningstar, accessed July 30, 2025, https://www.morningstar.com/stocks/xnas/crsp/quote
30. Look-Ahead Bias - Definition and Practical Example, accessed July 30, 2025, https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/look-ahead-bias/
31. (PDF) Temporal efficiency evaluation and small-worldness characterization in temporal networks - ResearchGate, accessed July 30, 2025, https://www.researchgate.net/publication/307950008_Temporal_efficiency_evaluation_and_small-worldness_characterization_in_temporal_networks
32. Optimizing Supply Chain Networks with the Power of Graph Neural Networks - arXiv, accessed July 30, 2025, https://arxiv.org/html/2501.06221v1

33. Stock market as temporal network - IDEAS/RePEc, accessed July 30, 2025, https://ideas.repec.org/p/arx/papers/1712.04863.html
34. Propagating Momentum Information Through Global Supply Chain Networks, accessed July 30, 2025, https://insight.factset.com/propagating-momentum-information-through-global-supply-chain-networks
35. Customer momentum research based on centrality of supply chain network - ResearchGate, accessed July 30, 2025, https://www.researchgate.net/publication/362152700_Customer_momentum_research_based_on_centrality_of_supply_chain_network
36. Extracting Alpha from Financial Analyst Networks - arXiv, accessed July 30, 2025, https://arxiv.org/html/2410.20597v1
37. Event Study: Definition, Methods, Uses in Investing and Economics, accessed July 30, 2025, https://www.investopedia.com/terms/e/eventstudy.asp
38. Fama-French Three-Factor Model - Components, Formula & Uses, accessed July 30, 2025, https://corporatefinanceinstitute.com/resources/valuation/fama-french-three-factor-model/
39. Solve Heat Equation Using Graph Neural Network ... - MathWorks, accessed July 30, 2025, https://www.mathworks.com/help/pde/ug/solve-heat-equation-using-graph-neural-network.html
40. Learning Heat Diffusion Graphs - MIT Media Lab, accessed July 30, 2025, https://web.media.mit.edu/~xdong/paper/tsipn17.pdf
41. Heat equation - Wikipedia, accessed July 30, 2025, https://en.wikipedia.org/wiki/Heat_equation
42. Mastering SIR Model in Epidemiology - Number Analytics, accessed July 30, 2025, https://www.numberanalytics.com/blog/ultimate-guide-sir-model-epidemiology
43. Supplier Disclosures and Customer Performance - Miami Herbert ..., accessed July 30, 2025, https://www.herbert.miami.edu/faculty-research/business-conferences/winter-warmup/krupa_paper_miami.pdf
44. Using Quantitative Investment Strategies - Investopedia, accessed July 30, 2025, https://www.investopedia.com/articles/trading/09/quant-strategies.asp
45. Backtesting Trading Strategies: How To Backtest A Strategy ..., accessed July 30, 2025, https://www.quantifiedstrategies.com/backtesting-trading-strategies/
46. Backtesting: How to Backtest, Analysis, Strategy, and More - Interactive Brokers LLC, accessed July 30, 2025, https://www.interactivebrokers.com/campus/ibkr-quant-news/backtesting-how-to-backtest-analysis-strategy-and-more-2/
47. Backtesting: Definition, Example, How It Works, and Downsides - QuantifiedStrategies.com, accessed July 30, 2025, https://www.quantifiedstrategies.com/backtesting/
48. Guide to Quantitative Trading Strategies and Backtesting - QuantVPS, accessed July 30, 2025,

https://www.quantvps.com/blog/guide-to-quantitative-trading-strategies-and-backtesting
49. Exploring the Dangers of Lookahead Bias in Financial Trading - FasterCapital, accessed July 30, 2025, https://fastercapital.com/content/Exploring-the-Dangers-of-Lookahead-Bias-in-Financial-Trading.html
50. Data snooping - Stanford Data Science, accessed July 30, 2025, https://datascience.stanford.edu/news/data-snooping
51. Understanding Data Snooping: Key Techniques to Prevent Analysis Bias - Number Analytics, accessed July 30, 2025, https://www.numberanalytics.com/blog/understanding-data-snooping-techniques-prevent-analysis-bias
52. Data Snooping Mining Dredging | ASYMMETRY® Observations, accessed July 30, 2025, https://asymmetryobservations.com/definitions/quantitative-research/data-snooping-data-mining-data-dredging/
53. Abstract: DATA-SNOOPING BIASES IN FINANCIAL ANALYSIS - MIT, accessed July 30, 2025, http://web.mit.edu/Alo/www/Papers/lo-94b.html
54. DGDNN: Decoupled Graph Diffusion Neural Network for Stock Movement Prediction - SciTePress, accessed July 30, 2025, https://www.scitepress.org/Papers/2024/124064/124064.pdf
55. Graph Neural Networks in Supply Chain Analytics and Optimization: Concepts, Perspectives, Dataset and Benchmarks - arXiv, accessed July 30, 2025, https://arxiv.org/html/2411.08550v1