European Research Studies Journal
Volume XXVII, Special Issue 2, 2024

pp. 25-37

Supporting Supply Chain Risk Management: An Innovative Approach Using Graph Theory and Forecasting Algorithms

Submitted 18/02/24, 1st revision 16/03/24, 2nd revision 20/04/24, accepted 16/05/24

Edmund Wąsik[1], Tomasz Sidor[2], Tomasz Wołowiec[3], Jacek Piwkowski[4], Michał Jasieński[5]

Abstract:

Purpose: *The article aims to develop a tool supporting risk management in the supply chain.*
Design/Methodology/Approach: *Graph theories were employed for data analysis. The study introduces a method for forecasting the behavior of sales risk values over a specific time horizon using a time series approach. Several metrics were utilized, starting with degree centrality, which assumes that crucial graph nodes have numerous connections. Betweenness centrality was also considered, assuming that key nodes link other nodes and measuring how often a vertex lies on paths between other vertices. Additionally, the PageRank algorithm, developed by Google, was applied.*
Findings: *The study produced an analysis of supply chain demand forecasting using the Monte Carlo method. Preparing such a forecast allows you to check many possible outcomes of the decision-making process. It can be used to assess the impact of risk, which in turn allows for better decision-making in conditions of uncertainty.*
Practical Implications: *The proposed method for forecasting sales values over a particular time horizon enables consideration of demand prediction within the supply chain, including associated risks.*
Originality/Value: *What is new is the use of graph theory to review supply chain risks and algorithms for applications supporting supply chain management.*

Keywords: *Supply chain management, forecasting, graph theories, financial risk.*

JEL codes: *C44, C61, D81, E27, L14.*

Paper type: *Research article.*

[1]*Corresponding Author: WSEI University, Lublin, Poland, e-mail: Edmund.Wasik@wsei.lublin.pl;*
[2]*WSEI University, Lublin, Poland, e-mail: Tomasz.Sidor@wsei.lublin.pl;*
[3]*WSEI University, Lublin, Poland, e-mail: Tomasz.Wołowiec@wsei.lublin.pl;*
[4]*WSEI University, Lublin, Poland, e-mail: Jacek.Piwkowski@wsei.lublin.pl;*
[5]*Wyższa Szkoła Biznesu - National Louis University, e-mail: pstalinski@wsb-nlu.edu.pl;*

1. Introduction

Supply chain risk management (SCRM) involves identifying, evaluating, and addressing risks in a company's supply chain (Supply Chain Risk ..., 2020). Global management strategies enable businesses to function more efficiently, cut costs, and enhance customer service.

Supply chain management focuses on how organizations handle the movement of their goods, covering all processes involved in converting raw materials into the final products or services the organization offers. It encompasses the planning and managing of procurement, distribution, and logistics activities.

A primary motivation for companies to adopt global supply chain management strategies is to enhance their competitive advantage. However, the benefits of lengthening supply chains can increase risks to quality, security, business continuity, reputation, and more. Therefore, regardless of industry, every company is exposed to internal and external risks associated with supply chain disruption (Zsidisin, 2009).

Risks in the internal supply chain include those caused by:

- Disruption to internal operations;
- Changes in key management, personnel, and business processes;
- Failure to implement contingencies in case something goes wrong;
- Failure to implement appropriate cyber security policies and controls to protect against cyber-attacks and data breaches;
- Failure to comply with environmental or labor laws;
- Failure to have merchandise that meets the needs of customers;
- And many more.

External supply chain risks include those caused by:

- Unpredictable or misunderstood customer demand;
- Interruptions in the flow of products, including raw materials, parts, and finished goods;
- Social, governmental, and economic factors, including the threat of terrorism;
- Supplier risk management, including concerns about the supplier's physical facility and regulatory compliance;
- Natural disasters, including earthquakes, hurricanes and tornadoes;
- And many others.

An organization can also suffer financial risk in the supply chain if something such as bankruptcy threatens a supplier's economic health. In addition, a company can face reputational risk if a supplier engages in unethical behavior, such as bribery, working against the code, or anything that could negatively affect the company's brand. A supplier's social media presence can also damage your brand. Mitigating these risks is critical, and many companies deal with this by diversifying their suppliers (Kadlubek *et al.*, 2022a; 2022b).

The risks mentioned above are only a tiny part of the real risks affecting directly or indirectly the smooth operation and existence of the supply chain.

The following will provide examples of applications that allow the user to gain insight into the risks affecting supply chain operations and perform analyses in selected areas of company operations to reduce certain risks. The data analyses presented below are performed using Python (Welcome to Python.org), while Dash (Dash Overview) is used for visualization and integration with the user.

2. A Review of Risk in Graph Theory

In this part, based on sample data regarding risks that may interfere with the operation of the supply chain, an application will be presented to support the user in interpreting risks and their relationships using graph theory.

Traditional risk assessment may be insufficient without considering the hazards that trigger other hazards (infectious risks) and the extent to which they do so. Two-dimensional risk management methodologies focusing on single risk points by measuring high probability and severity can provide a limited view of increasingly complex and globally interconnected risks (Zampeta and Chondrokoukis, 2023). Instead, we need to consider additional dimensions of risk, asking questions about how risks are related, the strength of these relationships, and the potential cumulative impact of such groups of risks.

Hazards such as natural disasters, infectious diseases, critical infrastructure failures, and climate change are likely to be considered low-probability hazards because they are doubtful they will occur more broadly, especially if we only look at specific regions of the world.

However, if they did happen, they could disrupt the organization in the short and long term. The effects can be even more devastating if they trigger other critical high-risk areas. Therefore, organizations must be adequately equipped to respond to these threats.

For example, the COVID-19 pandemic has sparked some unforeseen risks organizations are trying to mitigate. The outbreak of such infectious diseases can trigger operational risk, sourcing risk, working capital risk, etc., in a way the organization could not have foreseen. However, understanding what risks a sudden outbreak can cause and to what extent it can ruin operations will enable organizations to better plan for eventualities.

Traditional analytical techniques are not sophisticated enough, and graph networks are worth considering for performing a multivariate analysis of relationships.

A graph is a collection of interconnected nodes. Nodes represent entities; entity properties are embedded in those nodes, while connections to other nodes represent relationships. For risk assessment purposes, we can define risk as a node in the network. Each such node (risk) will have properties such as probability (node size) and impact (node color).

The interconnectedness of risks will be depicted by creating connections between risks that mutually influence/trigger each other. In contrast, the thickness of the connection will represent the strength of this connection.

We can then implement algorithms to analyze graph networks, gain valuable insights into their properties, and rank risks based on their interrelationships as a more analytical approach to determining the main risks that an organization should focus its limited resources on to mitigate them.

In addition, we can use advanced algorithms to create risk clusters in our network that provide information about how risks relate to each other and which risks can trigger each other. At this stage, we consider the previously calculated and defined probabilities of individual risks, their impact on company operations, and the strength of the links between them.

3. Dash Application

As mentioned, sample data was used to introduce the concept and illustrate the relationships between the different risks, which, through graphs, gives a more accessible view of the existing risks. The data has 33 nodes, which is the total number of unique risks in the graph. There are 85 edges between the 33 nodes, which tells us the total connections. In addition, the sizes of the nodes indicate the probability of the risks occurring, while the colors assess their impact on the supply chain.

In addition to such an interactive graph, using some measures from graph theory to better understand the dependencies and importance of individual risks across the network is beneficial. Three centrality measures are used in the proposed solution, and the results are tabulated. Network centrality measures allow us to identify the most critical nodes in a graph. Many such measures can be used to study a wide range of characteristics of our network.

In our analysis, however, we will only deal with those that can help us gain valuable information to assist in the risk assessment process. Looking at the network graph above, the question immediately arises: how do we understand this? How do we identify which nodes are the most influential in our network? Which nodes have the most connections? With this information, we can locate the most contagious risks in our network.

4. Degree of Centrality

The first measure used for the analysis is the degree of centrality. It is based on the assumption that the critical nodes of a graph have many connections, and in our case, this is followed by the statement that the threats associated with the most significant number of connections are the most contagious.

Using the NetworkX library for graph solutions, the degree centrality of a node v is the fraction of nodes it is connected to and is normalized by dividing by the maximum possible degree in a simple graph n-1, where n is the number of nodes in the graph G (networks.algorithms.centrality.degree_centrality). Using such a statistic, the application created a table containing the degrees of centrality of each risk, ordered in descending order.

Table 1. *Degree of centrality of individual risks*

No.	Risk	degree
1.	Risk of fraud and theft	0.3438
2.	Regulatory risk	0.3438
3.	Legal risk	0.3125
4.	Capital adequacy	0.2812
5.	Disaster and market crash	0.25
6.	Working capital risk	0.25

Source: *Own creation.*

This shows, among other things, that fraud, theft, and operational risks are associated with the most significant number of risks, demonstrating their mutual influence on risks that can disrupt the supply chain.

5. Degree of Indirectness

This centrality measure is based on the idea that significant nodes link other nodes. It gauges how frequently a vertex is positioned on paths between other vertices. Vertices with high betweenness can significantly impact the network as they regulate the flow of information to other nodes. Their removal would cause considerable disruption in communication between vertices since they are present on many message paths. The degree of indirectness of a node *v* (networkx.algorithms.centrality.betweenness_centrality) is the sum of the fraction of all pairs of shortest paths that pass through *v*:

$$c_{B(v)} = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}, \quad (1)$$

where V is the set of nodes, σ(s,t) is the number of shortest (s,t) paths, and σ(s,t│v) is the number of paths passing through some node v other than s,t. Furthermore, if s=t, then σ(s,t) = 1, and when v∈s,t, then σ(s,t│v)=0.

Assuming this form of statistics, the results for each risk in the application are shown in the following figure.

Table 2. *Degree of indirectness of individual risks*

No.	Risk	degree
1.	Legal risk	0.1937
2.	Risk of fraud and theft	0.1834
3.	Regulatory risk	0.1814
4.	Breakdown of business processes	0.1031
5.	Ryzyko konkurencji	0.0909
6.	Health, safety, environment	0.0831

Source: *Own creation.*

The analysis shows that legal risk is currently the most contagious risk. This is because legal risk is associated with more general risks than just the number of contiguous risks related to, for example, fraud.

6. Page Rank Algorithm

The PageRank algorithm, initially designed by Google to rank websites based on their relevance in search engine results (Amrani, 2020), relies on the concepts of random walks and Markov chains. It evaluates the significance of each graph node by considering the quantity of incoming links and the importance of the source nodes. The fundamental principle is that a node's importance is determined by the significance of the nodes that point to it.

PageRank was introduced in the original Google document as a function that solves the following equation:

$$PR(A) = (1-d) + d\left(\frac{PR(T_1)}{C(T_1)} + \cdots + \frac{PR(T_n)}{C(T_n)}\right) \tag{2}$$

where:
- we assume that node A has nodes from T_1 to T_n pointing to it,
- d is a damping factor in the interval [0,1). It is usually set to 0.85,
- C(A) is the number of edges leaving node A .

This equation is employed to iteratively refine the candidate solution iteratively, providing an approximate resolution to the original equation.

Using this algorithm, our meanings of the individual risks are shown in the table

Table 3. *Importance of individual risks using the PageRank algorithm*

No.	**Risk**	**degree**
1.	Legal risk	0.0613
2.	Regulatory risk	0.0609
3.	Risk of fraud and theft	0.0576
4.	Health, safety, environment	0.0479
5.	Breakdown of business processes	0.0469
6.	Capital adequacy	0.045

Source: *Own creation.*

Implementing the PageRank algorithm is another way to rank risks, taking into account their connections and the connections of their links. This is expected to give a better measure of risk contagion than the degree of centrality, as risks are related to their neighboring risks and the connections of their links.

For example, in practice, an event may trigger a low risk, which may trigger a medium risk, followed by a high risk. Although close to zero, the probability of such an event gives an understanding of how a portfolio of risks is interconnected in such a supply chain.

7. Forecasting Demand Based on Monte Carlo Simulation

This section will present an approach using Monte Carlo simulations to predict the demand of a particular product group over a specific time horizon. In today's world, risk analysis, in the broadest sense, is part of every decision we make, accepting factors of uncertainty or volatility.

Monte Carlo simulation (also known as the Monte Carlo method) allows us to test many possible outcomes of the decision-making process. It can be used to assess the impact of risk, allowing better decision-making under uncertainty. As an introduction, the method itself will be introduced, and then, in line with its ideology, an example of its use in demand forecasting will be presented.

8. Monte Carlo Method

When dealing with high uncertainty in forecasts or estimates, Monte Carlo simulation often offers a better solution by considering multiple values rather than relying on a single average. This technique was initially conceived by Stanislaw Ulam, a Polish mathematician who worked on projects like the Manhattan Project. After World War II, Ulam played countless solitaire games and became intrigued by plotting their outcomes to assess the probability of winning.

He shared his idea with John von Neumann, who collaborated to develop the Monte Carlo simulation (Eckhardt, 1987).

Monte Carlo simulations model the probability of various outcomes in processes that are difficult to predict due to random variables (Monte Carlo Simulation). The approach involves repeating random sampling to achieve specific results. A variable with inherent uncertainty is assigned a random value, the model is executed, and the result is noted. This process is repeated many times, with random values assigned based on a specific distribution. Once the simulation concludes, the results are averaged to provide an estimate.

9. Identification of Demand Volumes

In the case under consideration, we only have data on the quantity of products sold in the time series. From these, the theoretical distribution that best represents this variable will be estimated, and then, using the adopted distribution, a Monte Carlo simulation will be carried out to determine the total sales volume for the following period.

10. Data Preparation

Due to the constant changes in this area due to any promotions, new proposals, or customer demand, we will consider data from the last three months and try to predict the total market for the next 15 days. It is worth noting here that usually, in such considerations, the dependent variable is calculated with several other variables, the distributions of which we try to predict and then simulate each of them to obtain an appropriate prediction.

In this case, we only show a solution based on the dependent variable whose distribution we are studying (Hastie, 2009). The form of the set under consideration is shown in the following figure:

Figure 1. *Data set under consideration*

	llosc
2019-07-01	436.0
2019-07-02	987.0
2019-07-03	796.0
2019-07-04	683.0
2019-07-05	632.0
...	...
2019-10-11	778.0
2019-10-12	767.0
2019-10-13	756.0
2019-10-14	745.0
2019-10-15	1397.0

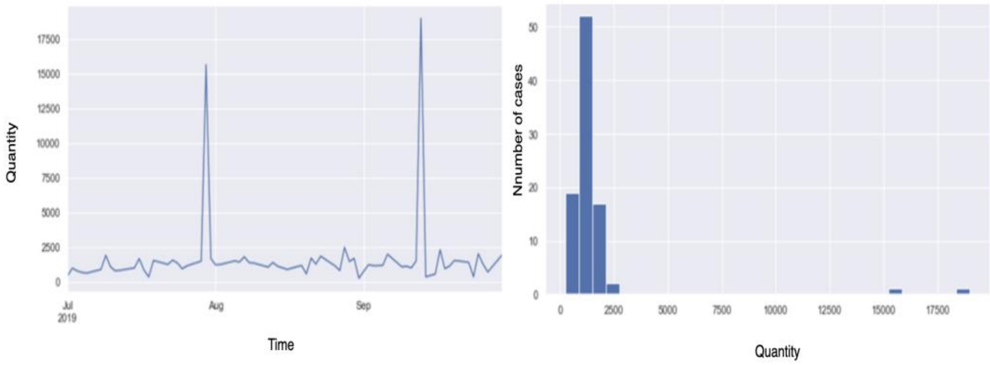
Source: *Own creation.*

As stated earlier, we will use data from July to September 2019 (inclusive) to test the distribution, and we will take the first 15 days of October as test data.

11. Testing and Fitting the Distribution

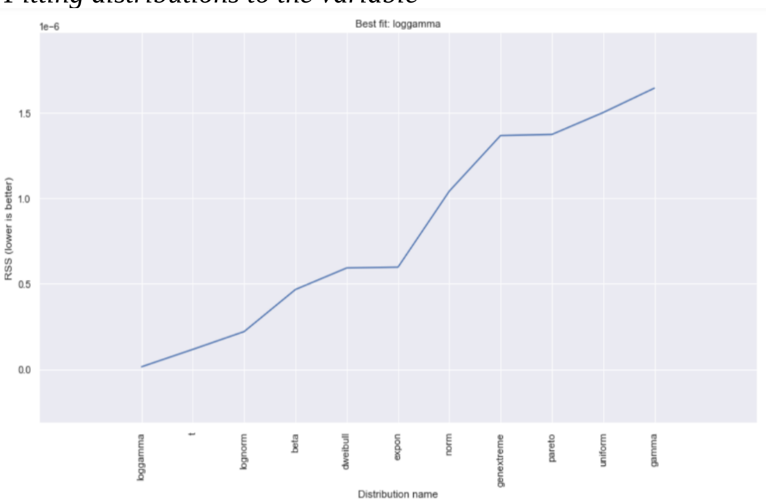
As stated earlier, we will examine the distribution of the variable Quantity over three months. Below is a histogram and a graph of the variable's value over time.

Figure 2. *Plot of values and histogram of the variable quantity*



Source: *Own creation.*

Figure 3. *Fitting distributions to the variable*



Source: *Own creation.*

It is worth noting that there are two clear spikes in sales over the period considered. To fit the theoretical distribution to the empirical one, we use a python library called distfit. To assess the fit, the Residual Sum of Squares of the differences will be calculated according to the formula:

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2, \tag{3}$$

where y_i is the i-th value of the predicted variable and $f(x_i)$ is the i-th value of the explanatory variable.

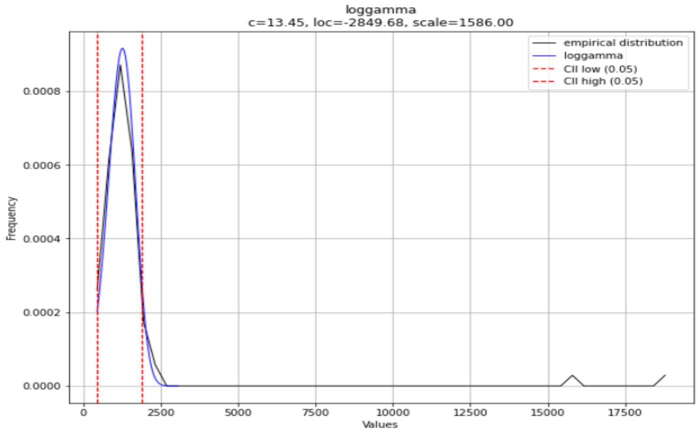
This measure describes the expected deviation from the actual empirical values versus the fitted theoretical values. Accordingly, low RSS values indicate a good fit for the adopted data distribution. Thus, we can use the library above to find distributions closest to our variable's actual distribution.

It follows that the log gamma distribution best represents the actual distribution. According to its definition, a variable has such a distribution if its natural logarithm has a gamma distribution. In addition to the shape and scale parameters derived from the gamma distribution, the parameter c is therefore added so that the density function then takes the following form (scipy.stats.log gamma …):

$$f(x,c) = \frac{\exp\,(cx - \exp(x))}{\Gamma(c)} \tag{4}$$

The estimated parameters c = 13.45, loc = -2849.68 and scale = 1586.00 along with the fit chart of such a distribution to the empirical distribution are presented in the following figure.

Figure 4. *Fitting the log gamma distribution to the data*



Source: *Own creation.*

However, an important step is to check whether the adopted distribution does not differ significantly from the empirical distribution. For this purpose, the Kolmogorov-Smirnov test was performed at a confidence level of 95%, which showed that the difference between the considered distributions is not significant, so we can use the theoretical log gamma distribution to represent our variable in the Monte Carlo simulation.

12. Conducting Simulations

After carefully examining the distribution of the variable under consideration and appropriately adjusting the theoretical distribution, you can start creating simulations. The task is straightforward at this point. Namely, by the obtained results, we should select a variable corresponding to our test fifteen days, assuming the appropriate parameters of the loggamma distribution.

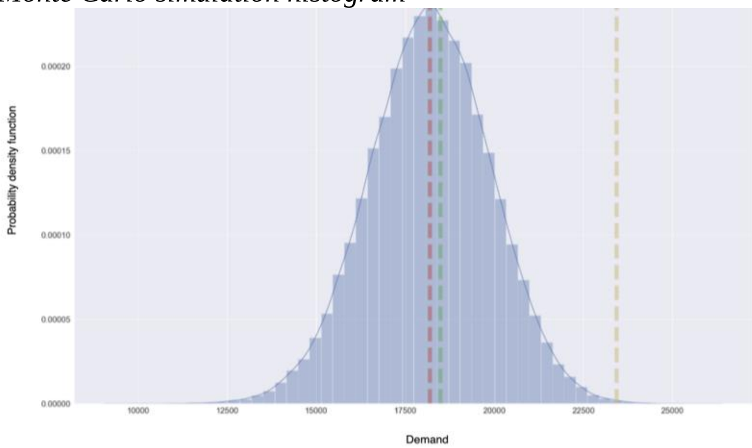
Then - because we want to calculate the cumulative demand value - we sum up the values from the selected 15 days. We carry out this procedure a certain number of times - in our case, we assume 100,000 trials - and then we divide the total demand obtained by this number of trials.

Table 4. *Result of simulation and average calculation*

Method.	**Value**	**Relative error (%)**
Monte Carlo simulation	18188	1.596
Average calculation.	23445	26.846

Source: *Own creation.*

Figure 5. *Monte Carlo simulation histogram*



Source: *Own creation.*

To compare the results, the number of items sold will also be calculated based on the average (average daily value calculated based on three months multiplied by 15)—a prevalent approach. During the fifteen days examined, the total number of products sold was 18,483. The table below compares the simulation performance and the adoption of the average. As you can easily see, the simulation result is better than simply calculating the average.

Of course, the average result is also subject to error due to jumps in the data. Still, the simulation using the Monte Carlo method was able to detect a trend to illustrate the actual demand adequately. The figure below shows a histogram of the simulations performed. The exact demand value is marked in green, the value obtained through simulation is marked in red, and the value calculated based on the average is marked in yellow.

Thus, such a simulation will not consider seasonal or holiday effects, but it is a simple alternative to much more complex solutions - and often better. The Dash application was created based on the theory and the above example. It allows the user to load a data set as a file from a disk and perform a simulation using the selected distribution.

13. Conclusions, Proposals, Recommendations

Most companies and organizations already know that collecting and using data can help their business. Tracking relevant metrics and statistics related to past actions and results can help you better understand where you are and adjust your actions for the future. This is usually called reactive analysis. The analysis would process the results of past events, gather relevant information and propose a strategy for the future.

However, this approach has a severe drawback because it creates a "blind spot" in the business process. When data from a past event has been analyzed and a decision made on a future strategy, sufficient time may have passed to effectively apply that strategy for the next time horizon. Companies and organizations that are more data mature are trying to take the next step and build various statistical or machine learning models that help them predict the likely outcomes of an event - hence the term predictive analytics.

Although more involved in practice, it is a continuous and often automated approach that should always use the latest relevant data and be able to predict events before they occur to implement appropriate actions.

References:

Amrani, A. 2020. PageRank algorithm, fully explained Towards Data Science. https://towardsdatascience.com/pagerank-algorithm-fully-explained-dc794184b4af.

Dash Overview. https://plotly.com/dash/.

Eckhardt, R. 1987. Stan Ulam, John von Neumann, and the Monte Carlo Method, Los Alamos Sci. 15, no. (Special Issue), 131-137.

Hastie, T., Tibshirani, R., Friedman, J. 2009. The Elements of Statistical Learning. New York, NY: Springer.

Kadłubek, M., Thalassinos, E.I., Domagała, J., Grabowska, S., Saniuk, S. 2022a. Intelligent transportation system applications and logistics resources for logistics customer service in road freight transport enterprises. Energies, 15(13), 4668.

Kadłubek, M., Thalassinos, E.I., Noja, G.G., Cristea, M. 2022b. Logistics customer service and sustainability-focused freight transport practices of enterprises: Joint influence of organizational competencies and competitiveness. J. Green Econ. Low-Carbon Dev, 1(1), 2-15.

Monte Carlo Simulation Definition. https://www.investopedia.com/terms/m/montecarlosimulation.asp.

networkx.algorithms.centrality.betweenness_centrality — NetworkX 2.6.2 documentation. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.betweenness_centrality.html#networkx.algorithms.centrality.betweenness_centrality.

networkx.algorithms.centrality.degree_centrality — NetworkX 2.6.2 documentation. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.degree_centrality.html#networkx.algorithms.centrality.degree_centrality.

scipy.stats.loggamma — SciPy v1.7.0 Manual. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.loggamma.html#scipy.stats.loggamma.

Supply Chain Risk Management https://reciprocity.com/resources/what-is-supply-chain-risk-management/.

Welcome to Python.org. https://www.python.org/.

Zampeta, V., Chondrokoukis, G. 2023. A Comprehensive Approach through Robust Regression and Gaussian/Mixed-Markov Graphical Models on the Example of Maritime Transportation Accidents: Evidence from a Listed-in-NYSE Shipping Company. Journal of Risk and Financial Management, 16(3), 183.

Zsidisin, G.A., Ritchie, B. 2009. Supply Chain Risk, vol. 124. Boston, MA: Springer US.