

Predicting Crisis in Global Trade Network: An Enhanced Decision Tree Based Methods

Vashisht Marhwal, Piyush Bamel, Tanay Agarwal

Abstract: *International Trade Relations represent a natural Social Information Network that has been extensively analyzed for various purposes like monitoring the global economy. The aim is to use the Global Trade Network to predict the occurrence of natural disasters or financial crisis based on the fact that the trade relations tax a hit in their patterns. The Global Network comprises of Export-Import Relations between the countries in the form of a Weighted Social Network. Predicting Trade relations help us effectively predict any future crisis and prepare for the same. An analysis of the Global Trade Network would discuss the centrality measures and Degree strengths. Using a list of crises which has occurred in the past and with the help of an efficient Machine Learning Model and Sampling Technique the aim is to improve the accuracy and precision of our prediction and discuss the implications on the network.*

Keywords : *Crisis Prediction, Decision Tree, Global Trade Network, Social Network Analysis.*

I. INTRODUCTION

Analysis of information networks provides immense insight in overall stability of the network as well as impact of a node on overall network. This particular use case takes into account the analysis of a global trade network through which steps are taken to minimise the impact of entire network by reducing dependence on a node. The network consists of countries, represented as nodes and their trade relations in form of links. Due to any disasters such as earthquakes, floods and hurricanes/ typhoons in a country overall network experiences a drastic change in trade volume over a period. Such model can help in finding non-natural anomalies in network as well as node of origin through further research. Expansion to local networks can help identifying problem in supply chain for business use case. Network measures such as link density, number of links, degree of node, average nearest neighbour degree of node, weighted average nearest neighbour degree of node, random walk betweenness centrality of node and weighted degree/strength of node act as quantifying unit for entire network. Trade volume between reporter and partner countries is represented in form of a weighted link in the graph.

Decision Tree based approach for predicting any crises that may have occurred in the network is the best suited

Revised Manuscript Received on December 16, 2019.

* Correspondence Author

Vashisht Marhwal*, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India. Email: vashisht_marhwal2016@vitstudent.ac.in

Piyush Bamel, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India. Email: piyush.bamel2016@vitstudent.ac.in

Tanay Agarwal, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India. Email: tanay.agarwal2016@vitstudent.ac.in

approach due to there being two distinctive class labels. This approach takes into the account the impact of each attribute that is present in the dataset and based on the how the class label acts, forms a tree which is termed as Decision Tree. This approach is highly acclaimed learning model which can effectively predict the outcome with ease. However, presence of an imbalance data might hamper its progress as the tree starts leaning towards the majority class label.

SMOTE is an over sampling technique that is used to balance out the dataset. It makes replicas of the minority instances in the dataset and then adds them so that any learning classifier is not overwhelmed by the majority class. This step is necessary to obtain a correct prediction which is not swayed by the majority class.

II. LITERATURE SURVEY

“Structure and Response in World Trade Network”[1] describes how the Global Trade Network has changed with globalization over the past 50 years. They primarily discuss how recession shocks affect the structural changes in the world trade network. It hasn’t taken into account the specific factors that directly affect trade, such as natural disasters or financial crisis that might occur in a country central to the network. In contrast, this approach targets specific areas that affect the network, analyzes them and, predict any crisis using a machine learning model.

“Social Network Analysis of Online Marketplaces”[2] presents a step towards the visualization of social information networks for online marketplaces such as eBay. A visualization toolkit is used on a processed dataset to study the social interactions that take place among the buyers and sellers of the platform. Their work is not limited to specific regions as these interactions take place worldwide. Their results provide insights into consumer interactions and can thus help improve the current business model of such marketplaces.

In their paper[3], Samgmoon Kim et al. attempt to empirically examine the nature of trade networks to study the changes in globalization patterns in the latter half of the 20th century. Their Social Network approach brings out an exciting inference about the increasing decentralization of global trade and a more recent push towards regionalization. This recent trend, however, does not undermine the spread of globalization, indicating that the two are not mutually exclusive. The author argues that trade regionalization leads to the creation of more trade opportunities but also presses the fact that there is not enough empirical evidence to prove such. This paper helps us identify the role of the regional trade network in the scope of a more extensive and more sophisticated global trade network. Thus, even though decentralization has increased,



its subsequent effects on globalization allow us to use the global trade network as a measure to predict crises.

In “Predicting Crisis in the Global Trade Network”[4] by Christina Kao, Lili Yang, and Ye Yuan did network analysis and focused on different network measures while applying Logistic Regression to produce their results. Even though Logistic Regression is a popular approach Decision tree has a fair advantage in the scope of this research. The decision tree divides the space into various smaller pieces while Logistic Regression divides it only into two pieces and generalize to plane and hyperplanes for higher-dimensional data. Also, the imbalanced data could be handled via different oversampling techniques such a SMOTE (Synthetic Minority Over-sampling), which is being utilized in our approach.

In the paper[5], a fast unfolding of communities in large networks, the authors propose a method to extract community structures in large networks using a heuristic method that is based on modularity optimization. Their proposed method outperforms other community detection techniques and also does not have the common bottleneck of computation power and is instead limited by storage, making it suitable for our work.

III. PROPOSED WORK

This paper proposes a methodology for understanding whether a machine learning model can be built that can

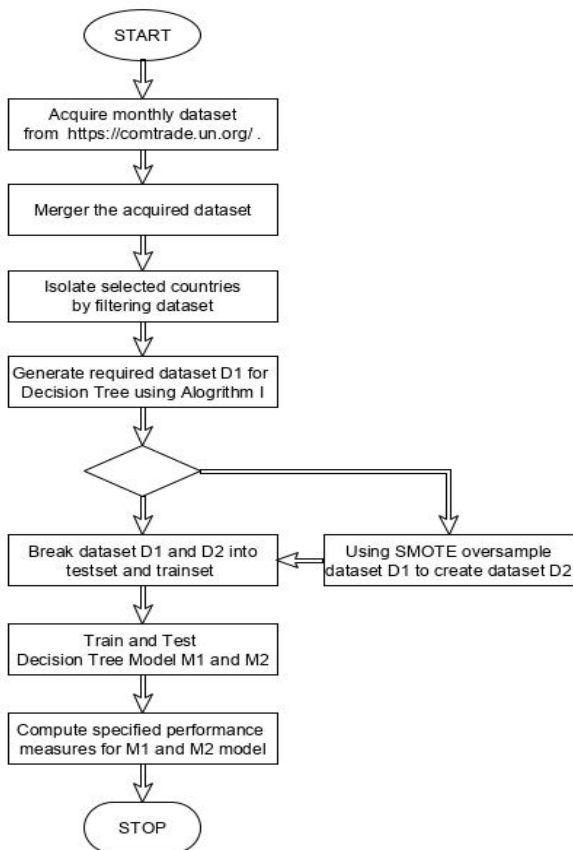


Fig. 1.Process Flow

identify the occurrence of a local crisis by analyzing the global trade network. The use of isolated systems for proving a hypothesis in theoretical physics inspires the approach used, where a set of 16 countries is isolated as a sub-graph of the global trade network based on the availability of disaster data. The sub-graph is analyzed to find different network measures that can be used as suitable parameters in the

dataset on which the machine learning model is trained. On further analysis, the dataset showed properties of an imbalanced dataset, which led to the creation of synthetic data points using SMOTE for model improvement. Hence, two separate models M1, without SMOTE and M2, with SMOTE, were generated.

A. Data Acquisition

Data has been acquired from an online repository managed

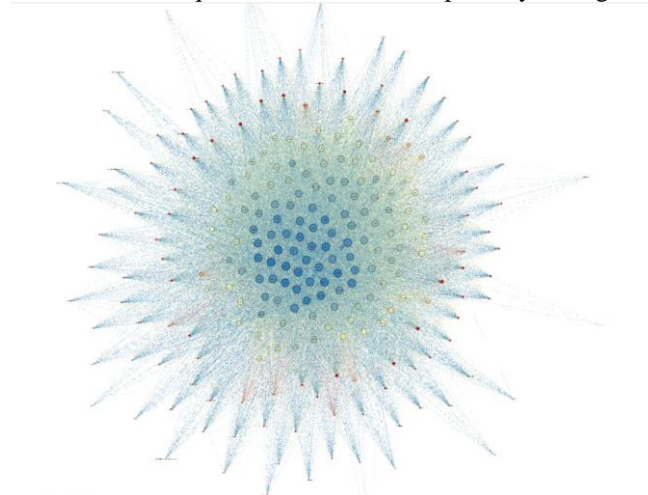


Fig. 2.Global Trade Network

by the United Nations through an open-source API <https://comtrade.un.org/>. Data was obtained month-wise from 2010 through 2012 for all countries and then heuristically combined. Using this Dataset “Trade Between Countries”, Weighted Global Trade network is formed.

Using the Global Trade Network, as shown in Fig.2, the ego network is formed for the countries where disasters have taken place in the year 2010-2012. Upon isolating the graph and data for those particular countries, a weighted directed graph is obtained. Centrality and Degree measures are calculated for each node.

B. Community Detection

Communities aim to identify highly connected groups amongst a complex social network. To identify countries that act as an Import-Export hub for other countries, the detection of relevant communities amongst the Global Trade Network is essential. Louvain Algorithm[5] for community detection is used since it was built to work extremely well with complex and dense networks such as a Global Trade Network. It follows a hierarchical clustering approach which recursively merges communities into a singular node and implements the modularity clustering on the condensed graphs. It is evaluated how much more densely connected the nodes within a community are, compared to how connected they would be in a random network. Algorithm for Louvain Community Detection is mentioned as follows:

Algorithm 1: Community Detection

```

Input: graph
1 Initialization: Each node is its own community
2 for each node  $\epsilon$  graph do
3   Remove node from current community
4   Place node in the community of its neighbour
5   Compute Modularity Change,  $Q$ 

$$Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

6   if  $Q > 0$  then
7     Node stays in Current Community
8     Ties are resolved randomly

```

C. Social Network Analysis

From the collected data, the obtained social network is analyzed on several parameters using networkX[7]. “networkX” is a python library for analyzing Complex Social Networks. The parameters are explained below[8]:

▪ **Betweenness Centrality**

This measure allows us to evaluate how many times a node has been used as a bridge for information flow inside the network. It helps us determine which country acts as a hub between import-export transfers.

▪ **Eigenvector Centrality**

Eigenvector Centrality evaluates the influence of a node over the information flow in the network. It helps us determine which country has a significant hold over the import-export inside the ego network. This parameter, in harmony with others, effectively helps us check where the trade relations have taken a hit.

▪ **Closeness Centrality**

Inside the Ego network, closeness centrality plays an important role. It helps us evaluate how close a particular node is to other nodes inside the network. This relationship in our social network is based on the weight of the edge between nodes. The weight represents the import-export volume between countries.

▪ **Degree Centrality**

Degree centrality evaluates the number of neighbouring nodes for a particular node. Countries with higher degree centrality have a high number of trade relations in the network.

The dependency of nodes on each other can be analyzed using these metrics, here, with respect to which countries are trading with each other.

IV. MACHINE LEARNING MODEL

A. Decision Tree

In the machine learning approach, the relevant data is processed by taking 16 countries, where disasters took place, into consideration and isolating them from the world: Afghanistan, Bangladesh, Bosnia Herzegovina, Brazil, Chile, China, Croatia, Fiji, Haiti, India, Indonesia, Iran, Ireland, Japan, Mexico, Myanmar. Taking monthly data between year 2010-2012. With these 16 countries, a directed graph is built. Using ego networks of each country various parameters are found, namely degree and centrality measures, that will be helpful for our prediction such as link density, number of links, degree of node, average nearest neighbor degree of

node, weighted average nearest neighbor degree of node, random walk betweenness centrality of node, weighted degree/strength of node and manually labelled the dataset with “crisis” or “no crisis”. Decision Tree approach is used for classification of binary data, i.e “crisis” or “no crisis” and for building decision tree algorithm called ID3 which employs a top-down, greedy search through the space of possible. ID3 uses Entropy (1) and Information Gain (2) to construct a decision tree.

$$Entropy(S) = \sum -p(I) \log_2 p(I) \quad (1)$$

$$Gain(S,A) = Entropy(S) - \sum_{t \in T} p(t) Entropy(t) = Entropy(S) - Entropy(S|A) \quad (2)$$

For further increasing the usability of this model, the generated dataset was analyzed and it was found that it was imbalanced i.e. labels of ‘no crisis’ are much more than ‘crisis’. For solving this problem and improving our model, Synthetic Minority Over-sampling Technique (SMOTE)[9] was used. This technique is followed to avoid overfitting which occurs when there is an imbalanced class problem. Exact replicas of minority instances are added to the main dataset. A subset of data is taken from the minority class as an example and then new synthetic similar instances are generated. These synthetic instances are then added to the original dataset. The new dataset is used as a sample to train the classification model.

B. Performance Measures

Our model’s prediction is evaluated based on the F1 Score. F1 score is a performance metric which analyses the overall accuracy of the model. F1 Score is calculated as the harmonic mean of precision and recall. Precision calculates how often the predicted result is correct. In a dataset where imbalance is present, precision checks if the false positives are high.

$$Precision = \frac{truepositives}{truepositives + falsepositives}$$

Recall, as opposed to precision, calculates if the false negatives are high in number.

$$Recall = \frac{truepositives}{truepositives + falsenegatives}$$

Their harmonic mean forms the F1 score which ideally predicts the overall accuracy of our model. F1 score lies between 0 and 1, where 1 is considered a perfect score.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

C. Decision Tree without SMOTE

After repeating 10 iterations, twice, over our approach, one without applying SMOTE technique and one with, the following results were achieved.



Table I: Accuracy, Average Precision Score, Recall and F1 value without SMOTE

Sr. No.	Accuracy	Average Precision Score	Recall	F1
1	0.9306	0.0497	0.1667	0.143
2	0.9306	0.0497	0.1667	0.143
3	0.9306	0.0497	0.1667	0.143
4	0.9306	0.0497	0.1667	0.143
5	0.9306	0.0497	0.1667	0.143
6	0.919	0.0455	0.1667	0.125
7	0.9248	0.0474	0.1667	0.133
8	0.9248	0.0474	0.1667	0.133
9	0.9248	0.0474	0.1667	0.133
10	0.9306	0.0497	0.1667	0.143
AVERAGE	0.9277	0.04859	0.1667	0.138

An F1 score of 0.13 was achieved which is fairly low because of the class imbalance. The number of ‘non-crisis’ data points heavily outnumber the ‘crisis’ data points.

D. Decision Tree with SMOTE

Table II: Accuracy, Average Precision Score, Recall and F1 value with SMOTE

Sr. No.	Accuracy	Average Precision Score	Recall	F1
1	0.9717	0.9394	0.968	0.9629
2	0.925	0.8504	0.923	0.9056
3	0.9302	0.8565	0.8947	0.9042
4	0.9109	0.8049	0.8965	0.8764
5	0.9391	0.8224	0.9062	0.8923
6	0.9457	0.8846	0.9489	0.93
7	0.9271	0.8313	0.8915	0.8915
8	0.9421	0.8557	0.9367	0.9135
9	0.8954	0.6802	0.8181	0.7964
10	0.936	0.8664	0.9514	0.9201
Avg.	0.9323	0.8392	0.9135	0.8993

Once synthetic sampling was done, a superior average score of 0.89, as shown in Fig. 3, was achieved which indicated the overall accuracy of our approach as 89%.

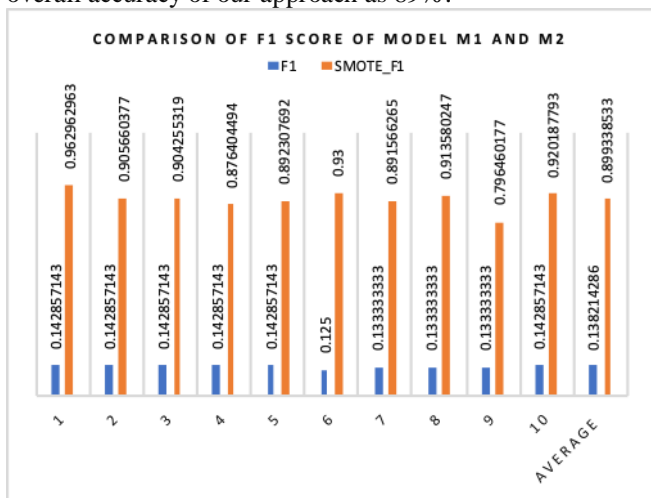


Fig. 3. Comparison of F1 Score for Model M1 and M2

V. RESULT AND DISCUSSION

Using ego networks in a complex social information

network isolated the countries where some disaster has taken place. This helped us focus on countries and their trade relations in a smaller network. After successful calculation of the Network analysis parameters, Decision Tree was used year wise to predict the occurrence of crisis. Using the Random walk betweenness centrality, we determine if a crisis has happened. If the value decreases for certain number of years, with the strength of neighboring node also decreasing means a crisis has occurred. Other features that affect our machine learning model include the number of links of node c, Average weight of neighbors of node c, and Average strength of the links of node c. List of crises is present in the Appendix. An average precision score of 0.48 and Average Recall of 0.16 was achieved in the imbalanced dataset. Once over sampling technique was applied, Precision and Recall score were drastically improved to 0.83 and 0.91 respectively. F1 Score acts as an overall measure for accuracy and a score of 89% was achieved which is a drastic improvement on the baseline approach. An updated dataset including recent trade volumes and flow with recent disasters will increase the accuracy of our model.

APPENDIX

Different natural disasters that are labelled are:

Earthquake

- Haiti, Jan 12, 2010 (Crisis Month: Jan, Feb 2010)
- Chile, Feb 27, Mar 11, 2010 (Crisis Month: Mar, Apr 2010)
- Mexico and Southern California, April 4, 2010 (Crisis Month: Apr, May 2010)
- Indonesia, Oct 25, 2010 (Crisis Month: Nov, Dec 2010)
- Bonin Island, Japan, Nov 30, 2010 (uninhabited, potentially not much impact)
- Tohoku Japan (+ Tsunami), Mar 11, 2011 (Crisis Month: Mar, Apr 2011)
- Iran Earthquake, Aug 11, 2012 (Crisis Month: Aug, Sep 2012)
- Afghanistan Earthquake, Jun 11, 2012 (Crisis Month: Jun, Jul 2012)

Floods

- Bangladesh, June 15, 2010 (Crisis Month: Jun, Jul 2010)
- China, May 10, 2010 (Crisis Month: May, Jun 2010)
- Beijing, China, July 2012 (Crisis Month: Jul, Aug 2012)
- Ireland, April 2012 - early 2013 (Crisis Month: Apr 2012 - Feb 2013)
- Fiji, Jan 21 - Feb 12, 2012 (Crisis Month: Feb, Mar 2012)

Hurricanes/Typhoons

- Hurricane Tomas, Category 2, Costa Rica / Haiti, Oct. 29 - Nov 7, 2010 (Crisis Month: Nov, Dec 2010)
- Hurricane Frank, Category 1, Mexico, Aug 21-28, 2010 (Crisis Month: Sep, Oct 2010)
- Cyclone Giri, Category 4, Myanmar, Oct. 20-23, 2010 (Crisis Month: Nov, Dec 2010)
- Hurricane Irene, Eastern US, Aug 21 - 30 2011 (Crisis Month: Sep, Oct 2011)



Isolated Countries:

- Afghanistan
- Bangladesh
- Bosnia Herzegovina
- Brazil
- Chile
- China
- Croatia
- Fiji
- Haiti
- India
- Indonesia
- Iran
- Ireland
- Japan
- Mexico
- Myanmar

REFERENCES

1. He, J., & Deem, M. W. (2010). Structure and response in the world trade network. *Physical review letters*, 105(19), 198701.
2. Kumar, P., & Zhang, K. (2007, October). Social network analysis of online marketplaces. In *IEEE International Conference on e-Business Engineering (ICEBE'07)* (pp. 363-367). IEEE.
3. Kim, S., & Shin, E. H. (2002). A longitudinal analysis of globalization and regionalization in international trade: A social network approach. *Social forces*, 81(2), 445-468.
4. Christina Kao, Lili Yang, Ye Yuan (2015, December). Predicting Crisis in Global Trade Network, Stanford University (Unpublished).
5. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.
6. Schönfeld, Mirco & Pfeffer, Juergen. (2019). Fruchterman/Reingold (1991): Graph Drawing by Force-Directed Placement. 10.1007/978-3-658-21742-6_49.
7. Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX*(No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States)
8. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.
9. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
10. Goyal, S., Chauhan, R. K., & Parveen, S. (2016, December). Spam detection using KNN and decision tree mechanism in social network. In *2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 522-526). IEEE.

AUTHORS PROFILE



Vashisht Marhwal is currently pursuing his Bachelor of Technology in Computer Science and Engineering at Vellore Institute of Technology, Chennai, India. His research interests include intelligent systems and their application in the industry. He also has keen interest towards the area of

Natural Language Processing.



Piyush Bamel is currently pursuing his Bachelor of Technology in Computer Science and Engineering at Vellore Institute of Technology, Chennai, India. His research interests include machine learning, artificial intelligence and their applications in the industry. He also has keen interest towards the area of computer

vision.



Tanay Agarwal is currently pursuing his Bachelor of Technology in Computer Science and Engineering at Vellore Institute of Technology, Chennai, India. His research interests include intelligent systems and