The Fourth International Conference on Government Education Management and Tourism
(ICoGEMT-4)
Bandung, Indonesia, January 25, 2025

DYNAMIC GRAPH REPRESENTATION LEARNING FOR FLIGHT DATA AS A FAST ECONOMIC INDICATOR

Wismu Sunarmodo

National Research and Innovation Agency (BRIN), Indonesia

Author's email:
wism002@brin.go.id

Corresponding author: wism002@brin.go.id

Abstract. *Future economic disruptions, such as natural disasters or pandemics, can impact a country's economy rapidly. However, current GDP-based economic indicators are slow to compile, taking months to release. This delay makes GDP growth unsuitable for timely policy decisions. A faster economic indicator is needed, one that correlates with official GDP, uses open data, and maintains historical consistency for predictive purposes. Big data, including shipping data, population mobility, and sentiment analysis, is being used to predict economic growth in near real-time. Air travel, a significant mode of transport in trade and tourism, has shown strong correlations with economic growth. Research indicates that the presence of airports and increased flight frequencies contribute to local economic activity. Flight data, which forms a global network between airports, can be modeled as a graph, where nodes represent airports and edges represent flight connections. Using graph analysis methods, these connections can be explored, and representation learning can automatically generate vector embeddings for nodes, simplifying the analysis of large and complex networks. These embeddings can be further used in machine learning models for economic predictions.*

Keywords: *Economic Indicator, Flight Data, Graph, Representative Learning.*

1. INTRODUCTION

In 1997, Indonesia and several Southeast Asian countries faced a severe financial crisis, which quickly escalated into an economic crisis. In 2020, the COVID-19 pandemic affected nearly all countries worldwide, leading to travel restrictions that slowed economic growth and caused recessions in many nations. The economic crisis resulted in business closures, job losses, food shortages, and high inflation (Belitski, 2022).

In the future, events such as natural disasters, disease outbreaks, global warming, and wars could rapidly impact a country's economy. Current economic indicators, such as Gross Domestic Product (GDP), take a long time to compile and release, making them less effective for quick decision-making. Therefore, a faster economic indicator with more frequent releases is needed.

Haldane A. (2021) proposed criteria for a new indicator, including high correlation with GDP, use of open data that can be verified, and historical consistency for predictions. Some researchers have used big data to predict economic growth in real time, such as shipping data, population mobility, and news sentiment analysis (UNSTATS, 2023). Flight data also correlates with economic growth because air travel supports global trade and tourism (Wang, 2020; Zhang, 2024). According to the International Air Transport Association (IATA), flights account for 35% of global trade and 54% of business relationships (IATA, 2022).

Research shows that airports can boost local economic activity, and economic growth increases flight frequency (Fangni & Daniel, 2020; Aunurrofik, 2018; Yongbin, 2020). Flight data, which includes departure and arrival schedules, forms a network between airports that can be analyzed using graphs. A graph consists of nodes (airports) and edges (flight routes), which can provide insights into airport connections. This study aims to develop a fast economic indicator derived from flight data using a graph representation approach. However, large and complex graphs are difficult to analyze manually. One method for automatic graph analysis is representation learning, which generates vector representations for each node, edge, or graph. Nodes with higher similarity will have similar vector representations, which can be used for further analysis or as input for other machine learning models.

2. LITERATURE REVIEW

2.1 Economic Network

In social life, economic life also involves other parties to function properly. When more parties become involved in the economy, they create an economic network. This network is formed through various channels that allow people to meet or communicate with each other. The types of networks that support the formation of an economic network include transportation, communication, and energy networks, among others. For example, transportation networks allow goods to be shipped from the seller to the buyer, making transportation closely related to the economy. Similarly, communication networks help economic players communicate quickly without needing face-to-face meetings. Many studies have been conducted to link economic networks with supporting networks. Research has been done on the connection between economic networks and transportation (Tsiotas, 2021), and how disruptions in transportation networks affect the economy.

A network is a natural form of a graph, so it's possible to model economic networks or their supporting networks as graphs. The relationship and representation of an economic network as a graph were discussed in a study by Chen et al. (2005). With the advancement of analysis methods and machine learning on graphs, deeper analysis can now be done using various methods. Setayesh et al. (2022) analyzed global trade networks using Exponential Random Graph Models (ERGMs). Understanding the global trade network allows for analysis of a country's economic situation in the context of the global economy.

2.2 Graph

A graph is a structure used to explain and analyze objects that have relationships or interactions. Formally, a graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of nodes (points) and $\mathcal{E}$ represents the set of edges (connections between the nodes). An edge connecting two nodes, $u \in \mathcal{V}$ and $v \in \mathcal{V}$, is denoted as $(u, v) \in \mathcal{E}$. According to Kazemi et al (2020), Based on whether there are changes in the graph, graphs are divided into two types: static graphs and dynamic graphs. If there are no changes in the graph, it is called a static graph. On the other hand, if there are changes in the graph, such as the addition or removal of nodes or edges, it is called a dynamic graph. Dynamic graphs are further divided into two categories based on how they are represented: continuous-time dynamic graphs (CTDG) and discrete-time dynamic graphs (DTDG). A CTDG is defined as $(\mathcal{G}, O)$, where $\mathcal{G}$ represents the graph at time t0, and O is the set of observations that describe changes from the initial graph $\mathcal{G}$. Each observation typically includes information such as (type of event, event, time). The types of events can include the addition or removal of nodes or edges.

Static node embedding techniques are methods used to represent nodes in a graph as vectors, which can be used for various tasks. There are two main approaches: random walk-based methods and graph neural networks (GNNs).

Random walk-based methods, like DeepWalk (Perozzi, 2014), node2vec (Grover, 2016), and Hub-aware random walk (Tomcic, 2024), focus only on the structure of the graph and don't consider the features or attributes of the nodes themselves. On the other hand, GNNs, such as Graph Convolutional Network (GCN) (Kipf, 2016), GraphSAGE (Hamilton, 2017), Graph Attention Network (GAT) (Vaswani, 2017), and Graph Isomorphism Network (GIN) (Xu, 2019), use both the graph structure and node attributes for better node representations.

GCN works by passing information between neighboring nodes using convolution, while GraphSAGE first samples nearby nodes before combining their features. GAT uses an attention mechanism to decide which neighboring nodes are more important, giving it more flexibility. GIN aims to improve the ability to distinguish different node types, making it especially useful for tasks like classification.

3. RESEARCH METHODS

The following hypotheses were used to develop the method: First, the economic growth of a region or country is correlated with the connectivity between its airports, meaning the more flight connections there are, and the higher their frequency, the higher the economic growth. Second, economic growth over time can be measured by the coefficient of connectivity over time. Third, the Euclidean distance between nodes in node embedding reflects their relative closeness in the graph, which represents the relative connectivity between airports. However, this closeness only shows connectivity at a single point in time, not across different time periods. Fourth, to assess connectivity over time, each period's flight graph is treated as an isolated subgraph within a spatiotemporal graph, as shown in Figure 2. The more connections a subgraph has, the closer its nodes will be in the embedding compared to subgraphs with fewer connections, as shown in Figure 1. Fifth, the density or closeness of nodes within each subgraph embedding (for each time period) indicates the level of flight connectivity during that period, relative to other periods. Lastly, in a skip-gram model, the output probability from the neural network reflects the centrality or influence of a node within the graph.

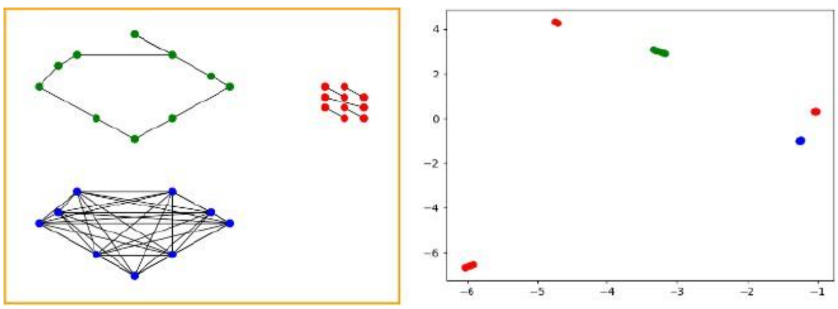
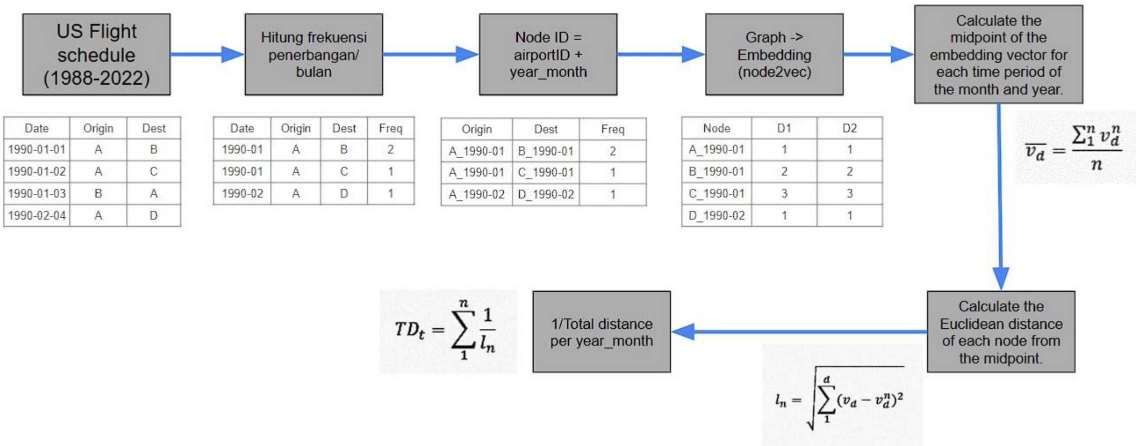


Figure 1. The left image shows a graph made up of several separate subgraphs, while the right image shows the embedding results, where the more relationships there are in a subgraph, the closer the embedding distance. In the blue subgraph, which has the most relationships in the graph, the nodes are closer together in the embedding. Conversely, in the red subgraph, which has the fewest relationships, the nodes are farther apart in the embedding.

Figure 2. The principle of a diachronic encoder is applied to the flight relationship graph, where the node ID is composed of the airport ID along with a time parameter.

The method and data illustration used to identify trends in flight connectivity are shown in Figure 3. The data source is domestic flight departure and arrival data in the United States from 1988 to 2022, obtained from the Bureau of Transportation Statistics (BTS), Department of Transportation (DOT) U.S. This flight schedule data is then transformed into monthly flight frequency relationships. For economic growth data, monthly GDP data of the United States from S&P was used, with the same time range as the flight data. In this study, an undirected graph is used, meaning the flight relationship from airport A to airport B is treated the same as the flight from airport B to airport A.



Date	Origin	Dest
1990-01-01	A	B
1990-01-02	A	C
1990-01-03	B	A
1990-02-04	A	D

Date	Origin	Dest	Freq
1990-01	A	B	2
1990-01	A	C	1
1990-02	A	D	1

Origin	Dest	Freq
A_1990-01	B_1990-01	2
A_1990-01	C_1990-01	1
A_1990-02	D_1990-02	1

Node	D1	D2
A_1990-01	1	1
B_1990-01	2	2
C_1990-01	3	3
D_1990-02	1	1

Figure 3. Flowchart of Proposed Method

Using node2vec for node representation learning, each airport per time period is mapped into an embedding vector. The model combines random walk and skip-gram techniques. The random walk converts the graph into a list of nodes based on relationships and flight frequencies, with parameters p and q influencing global and local node proximity, respectively. The walk length determines the number of node samples. These samples are split by window size, and the target-context pairs are used in the skip-gram model to generate embedding vectors. The model optimizes the probability of context given a target node, and the embedding vector is based on the hidden layer's input weights.

The embeddings are grouped by time period, and the closeness between nodes is calculated using the Euclidean distance to the center of each period's embedding. The total distance (TDt) reflects the relative density of each subgraph compared to others.

4. RESULTS AND DISCUSSION

4.1 Connectivity Trend

In the trial of the proposed method to determine flight connectivity trends, three combinations of parameter values for p, q, and walk length in node2vec were tested, as shown in Figure 4. The Pearson correlation coefficient for each parameter combination was compared with monthly GDP (mGDP), as shown in Table 1. The results showed that all three parameter variations yielded good Pearson correlation coefficients. However, changes in the q parameter had little impact, even though p and q both had higher accuracy. The walk length parameter had a greater influence on the results, with longer walk lengths resulting in better correlation.

Visually, the trend between the connectivity index graph and the mGDP pattern tends to be similar. However, the graph on the connectivity index tends to be volatile, which may be due to seasonal air passengers traveling by plane during holidays. This can be seen from the tendency of increased connectivity values during holiday months. In addition, another factor that may influence the results is the use of only domestic flight data from the United States, excluding international flights. Meanwhile, economic growth is also influenced by the interactions between one country and others. From the

Pearson correlation index between the connectivity index value and mGDP, it appears that the correlation value is high for all scenarios, with the highest correlation index value being more influenced by the walk length parameter. This may be because the longer or more walk lengths better represent the simulation of air traffic. However, increasing the walk length also significantly increases the processing time.

Figure 4. The comparison of the flight connectivity trend graph with mGDP is as follows: the yellow line represents the mGDP graph, the blue line represents the connectivity trend with parameters P=1, Q=1, walk length=20, the orange line represents the connectivity trend with parameters P=1, Q=0.5, walk length=20, and the gray line represents the connectivity trend with parameters P=1, Q=1, walk length=200.

Table 1. Pearson Correlation of Connectivity Index

No.	P	Q	Walk length	Pearson Correlation
1	1	1	20	0.959
2	1	0.5	20	0.956
3	1	1	200	0.981

CONCLUSION

In the preliminary research trial that was conducted, the results showed a fairly good correlation between economic growth and the proximity of points in the embedding. This indicates that flight data can be used as an indicator of economic growth. Meanwhile, the airport influence ranking using the node influence method in the embedding domain produced better accuracy compared to the point ranking method in the graph domain. Further research is still needed because the method used in this study only identifies the overall trend of connectivity without examining the connectivity trend for each individual airport or region. The method also still combines spatial and temporal graph components into one, which may result in the loss of temporal trends at each airport, as proximity values are measured relatively to all points at all times. In fact, each region may have different seasonal trend rates.

REFERENCES

Aunurrofik, A. (2018). "The Effect of Air Transportation on Regional Economic Development: Evidence From Indonesian Regencies." Signifikan, vol. 7, no. 1, 2018, pp. 45-58, doi:10.15408/sjie.v7i1.6178.

Belitski M, Guenther C, Kritikos AS, Thurik R. (2022). Economic effects of the COVID-19 pandemic on entrepreneurship and small businesses. Small Bus Econ. 2022;58(2):593–609. doi: 10.1007/s11187-021-00544-y. Epub 2021 Sep 12. PMCID: PMC8435174.

Chen, S.H., and Chen C.L. (2005). Graphs and Network Economics. 2005 International Conference on Neural Networks and Brain, Beijing, 2005, pp. 1639-1645

Fangni Zhang & Daniel J. Graham (2020): Air transport and economic growth: a review of the impact mechanism and causal relationships, Transport Reviews, DOI: 10.1080/01441647.2020.1738587

Grover, A. and Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 855–864. https://doi.org/10.1145/2939672.2939754

Haldane, A., Chowla, S. (2021). Fast economic indicators. Nat Rev Phys 3, 68–69. https://doi.org/10.1038/s42254-020-0236-y

Hamilton, W. L., Ying R., and Leskovec J.. (2017). Inductive representation learning on large graphs. In NeurIPS, 2017b.

IATA. (2022). 2022 Air Connectivity: Measuring The Connections That Drive Economic Growth. International Air Transport Association, https://www.iata.org/en/iata-repository/publications/economic-reports/air-connectivity-measu ring-the-connections-that-drive-economic-growth/

Kazemi, S.M., Goel R., Jain K, et. al. (2020). Representation learning for dynamic graphs: a survey. J. Mach. Learn. Res. 21, 1, Article 70 (January 2020), 73 pages.

Kipf, T.N., & Welling. M. (2016). Semi-supervised classification with graph convolutional networks. In ICLR, 2016a.

Perozzi, R., Al-Rfou, R., and Skiena, S. (2014). DeepWalk: online learning of social representations. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14). Association for Computing Machinery, New York, NY, USA, 701–710. https://doi.org/10.1145/2623330.2623732

Setayesh, A., Sourati Hassan Zadeh, Z. & Bahrak, B. (2022). Analisis jaringan perdagangan global menggunakan model grafik acak eksponensial. Aplikasi Netw Sci 7 , 38

Tsiotas, D. (2021). Drawing indicators of economic performance from network topology: The case of the interregional road transportation in Greece. Research in Transportation Economics, Elsevier, vol. 90(C).

UNSTATS. (2023). Faster economic indicators from big data. United Nations Statistics Division, diakses 20 Desember 2023, https://unstats.un.org/unsd/bigdata/conferences/2019/presentations/conference/day1/econo mics/tom-smith-uk%20faster-economic-indicators.pdf

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. arXiv:1706. 03762 [cs.CL], 2017.

Wang Y.(2020). The correlation between GDP and different transport modes turnover based on grey correlation analysis. J. Phys.: Conf. Ser. 1486 072008

Zhang F. & Graham D., J. (2020): Air transport and economic growth: a review of the impact mechanism and causal relationships, Transport Reviews, DOI: 10.1080/01441647.2020.1738587