Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 3 (2025)

Leveraging Social Network Analysis For Optimized Community Detection: Insights From Global COVID-19 Datasets

Sandeep Ranjan[1*], Bhuvan Unhelkar[2], Deepak Prashar[3]

[1]CTIEMT, Jalandhar, India ersandeepranjan@yahoo.com
[2]University of South Florida, USA bhuvan.unhelkar@gmail.com
[3]Lovely Professional University, Phagwara, India deepak.prashar@lpu.co.in

Abstract:-
This research paper explores social network analysis for detecting communities within COVID-19 dataset to enhance pandemic outbreak response and prediction. Advancing community detection techniques, especially in the early stages of the pandemic, enables an improved understanding of its emergence and trajectory, resulting in optimized response effectiveness. The study employs network modeling, data visualization, and correlation analysis to identify communities in global COVID-19 data. Networks are created with case counts, deaths, and recoveries, and a community detection algorithm is applied to uncover clusters and pandemic trajectories. The clean and detailed COVID-19 data enables a comparative analysis of populations with varying mortality and recovery rates, enriching the community detection process. The paper also introduces a graph generator using Gaussian distributions to create realistic sparse networks, providing synthetic data for further analysis. The key contribution of this study is enhanced preparedness in identification, clustering, and tracking a potential pandemic. Future research will focus on dynamic networks to track the evolution of communities over time and integrate statistical models to enhance the realism of network structures.

Keywords: Community Detection, COVID-19 Data, Dynamic Networks, Outbreak Prediction, Social Network Analysis

Introduction:
Response to the COVID-19 pandemic presents an opportunity to learn how to optimize community detection and apply it to develop effective mitigation strategies for future events. The insights can help governments develop policies and frame decisions ((Pang & Lee, 2022) (Jairam-Owthar et al., 2022). This study utilizes social network analysis (SNA) to enhance community detection and forecast pandemic patterns. A comprehensive dataset in CSV format, encompassing variables such as WHO regions, confirmed cases, deaths, and recoveries from various countries is utilized. Initially, data is cleaned using Pandas to address missing values and duplicates. Next, advanced visualization tools like Matplotlib and Seaborn are used to identify key metrics such as confirmed cases and mortality rates. These visualizations aid in identifying trends and patterns among the most affected regions, enabling comparative analysis of the pandemic's progression. Additionally, we utilize network analysis techniques to model relationships between different international locations. Specifically, we use the Louvain modularity optimization approach to detect communities within the network, allowing us to identify clusters of closely linked regions based on COVID-19 features and gain valuable insights into the virus's global spread. Integrating community detection with predictive analysis enables this study to deepen understanding of pandemic clusters and pathways, resulting in effective and optimized response strategies.

Literature Review and Gap Analysis:

Network-based analyses have heavily influenced the examination of COVID-19's initial global dissemination. Wickramasinghe (2021) conducted a groundbreaking study that utilized actual travel data to map countries as nodes and travelers as connections (Wickramasinghe, 2021). Network theory and exponential random graph models, as applied in this study, have helped identify key locations and patterns in disease transmission. Four network discovery algorithms – Louvain, Infomap, Label Propagation, and Spin Glass – have been employed in this study to uncover community structures using COVID-19 data. The algorithms' performance is evaluated using similarity metrics such as the adjusted Rand index and normalized mutual information.

Wickramasinghe's study had a notable limitation due to the absence of genuine local data, prompting simulation techniques to generate synthetic networks. Two types of random graph generators have been used to study the spread of COVID-19: Gaussian random partition and a specialized Gaussian mixture technique. These simulations aimed to create sparse networks that resemble disease spread patterns. The results revealed that sparse networks with low edge probabilities in high-density areas were effective for local detection (Lee et al., 2022). The accuracy of community detection was affected by network parameters, including the average length and diversity of nodes. Heat maps illustrated the effects of various community variables on algorithm performance (Nallusamy & Easwarakumar, 2023). Among the algorithms tested, Louvain achieved the highest accuracy, Infomap performed well for sparse networks, while Label Propagation underperformed.

Shabdiz et al. (2023) analyzed the effect of COVID-19 on the iris of diabetic patients by integrating network science with community detection technologies to enhance COVID-19 pandemic transmission models (Shabdiz et al., 2024). A significant innovation in that study is the development of a novel random graph generator that leverages multiple Gaussian distributions to create more realistic, sparse network structures. This method addresses the limitations of earlier Gaussian random splitting techniques and generates synthetic data that more accurately reflects real-world disease patterns. A significant innovation in that study was the development of a new random graph generator that utilizes multiple Gaussian distributions to create more realistic sparse network structures. This approach addressed the deficiencies of previous Gaussian random splitting methods and produced synthetic data that better approximates actual disease data.

Despite these advancements, several gaps remain. The reliance on simulated data rather than authentic global data limits the applicability of findings to real-world scenarios. Bolaji (2021) emphasized the need for rigorous method comparison and performance evaluation in the context of network variables to bridge this gap (Bolaji, 2021). Westarb et al. (2023) emphasized the importance of understanding local detection methods and recommended a nuanced approach to algorithm selection based on network characteristics (Westarb et al., 2023). Social network data analysis focusing on community structures revealed interesting insights into donor behavior in non-profit organizations which has applications in predicting behavior patterns (Alsolbi et al., 2023; Gulati et al., 2024). Extensively harnessed social media data can be used to enable pattern identification and optimized decision-making (Unhelkar & Gonsalves, 2021).

Further research by Čížková (2022) analyzed disease transmission modeling through global COVID-19 mobility data and simulations (Čížková, 2022). This study used exponential random graph models to analyze cross-border link factors and applied community detection algorithms to identify clusters. However, the absence of established data was a significant limitation, indicating the need for improved simulation methods.

Zhang et al. (2021) and Abdulla & Khasawneh (2022) assert that networks with low inter-region boundaries and high intra-region edge probabilities enhance community recognition accuracy (Abdulla & Khasawneh, 2022; Zhang et al., 2021). Louvain method was identified as the most accurate, while Infomap showed improvement in sparse networks. Their study also indicates a need for enhanced graph generators and simulation techniques to advance disease transmission models.

Studies highlight significant advancements in utilizing network analytics for disease modeling yet critical gaps persist, especially concerning the availability and authenticity of data and the ensuing analytics. Addressing the current gaps is essential to enhance the accuracy of community detection and deepen our understanding of disease dynamics within complex networks. Integrating real-world data with advanced simulation methods and flexible statistical distributions helps model irregular community sizes and structures. Addressing these gaps enhanced the accuracy of community detection and improved our understanding of disease dynamics in complex networks (Fahad Alkhamees et al., 2021; Oza et al., 2023). The authors prioritized integrating real-world data with advanced simulation methods and flexible statistical distributions to model irregular community sizes and structures better.

Social media blogging platforms are a rich source of datasets to study the formation of online communities and complex networks. (Matherly & Greenwood, 2024; Safadi et al., 2024) Identifying community structures in these networks helps in predicting the spread of interactions and assess its effects on social network users. Positive and negative correlations can be helpful to the owners and promoters, or government entities assess the situation and adjust their future course of action. Normalization and standardization techniques facilitate comparative analysis of the general public's collective wisdom posted on social media platforms (Dadkhah et al., 2021). Customers use social media analytics in crucial decision-making circumstances, such as when selecting stocks for investments and making a buy/sell/hold decision (Havakhor et al., 2023; Ranjan et al., 2018).

Research Methodology:

The primary aim of this research is to apply Social Network Analysis (SNA) to identify communities in COVID-19 dataset to develop an efficient strategy for improved outbreak response and prediction. Cluster Analysis in this study uses Python to predict future pandemic patterns for effective response. The following are the objectives of this study:

1. Community Clustering: Use the Louvain Modularity algorithm to perform community clustering of the dataset, identifying and analyzing communities with similar pandemic trends. This clustering will enable the detection of regional patterns and group dynamics in pandemic progression.
2. Trend Identification: Analyze pandemic trends, including cases, survival rates, deaths, and recoveries, through data visualization. This will reveal how the pandemic has progressed differently across countries and help identify common trends and patterns.
3. Network Community Detection: Utilize network analysis techniques to detect communities with similar patterns across different countries. This will enhance the understanding of global outbreak spread based on the clusters produced by the Louvain Modularity algorithm.

Problem Statement:

The COVID-19 pandemic resulted in millions of deaths worldwide. Physicians and other healthcare professionals utilized social media platforms to disseminate knowledge and awareness of the pandemic (Y. Liu et al., 2023). Despite the wealth of worldwide statistics there remains a gap in understanding the complex mechanisms of pandemic trajectories. Social network community analysis can enhance the scientific understanding of the social, economic, and health factors predicting pandemic severity and spread. This research aims to address this gap by analyzing a dataset tracking COVID-19 indicators across 187 nations, employing a combination of clustering, correlation analysis, graph-based community detection, and data visualization techniques. The primary challenge is the development of a robust computational framework that highlights significant factors, connections, and trends in global pandemic dynamics.

Data Collection

The dataset used for this research is sourced from Kaggle and includes comprehensive information on COVID-19 cases (Devakumar, n.d.). This dataset was selected for its extensive quantitative metrics, which are crucial for modeling and analysis.
The dataset encompasses data on 187 countries, providing details on COVID-19 indicators such as admissions, deaths, recoveries, active cases, and weekly changes, as shown in Figure 1. It includes 15 columns with metrics such as deaths per 100, recoveries per 100, and cumulative statistics. The data allows for monitoring both weekly and cumulative epidemic trends and supports standardized cross-country comparisons through coefficient calculations. Additionally, the dataset categorizes countries by WHO regions, facilitating geographical analysis (Matta et al., 2023).

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region
0	Afghanistan	36263	1269	25198	9796	106	10	18	3.50	69.49	5.04	35526	737	2.07	Eastern Mediterranean
1	Albania	4880	144	2745	1991	117	6	63	2.95	56.25	5.25	4171	709	17.00	Europe
2	Algeria	27973	1163	18837	7973	616	8	749	4.16	67.34	6.17	23691	4282	18.07	Africa
3	Andorra	907	52	803	52	10	0	0	5.73	88.53	6.48	884	23	2.60	Europe
4	Angola	950	41	242	667	18	1	0	4.32	25.47	16.94	749	201	26.84	Africa

Figure 1: Displaying the first five rows of the dataset

The first five rows of the dataset are displayed to provide an overview of the dataset columns for the rest of the analysis. There is a significant disparity in the recovery rates. Afghanistan's rate hovers around 69%, while India's exceeds 93% (Rostami et al., 2023). The weekly percentage increase indicates fluctuations in this outbreak. Zambia's figure of 36.86% suggests the spread of an epidemic. This extensive dataset can be utilized for international and country-specific assessments of the impacts and responses to the COVID-19 pandemic globally.
The *df.info* function provides comprehensive dataset information that ensures the data aligns with expectations before proceeding with analysis and visualizations, enabling the early detection of abnormalities. Null values and columns are checked to determine the extent of cleaning required and the types of visualizations needed. Figure 2 provides a snapshot of the results generated by the *df.info* function.

Research Design

The research design outlines a structured approach for systematically collecting and analyzing important data (Zarezadeh et al., 2022). The methodology to identify the community structures in the COVID-19 dataset has been designed for examining the structure and addressing the subject of the study. In this research, a multivariate regression analysis is employed alongside a quantitative correlational design (Rustamaji et al., 2023).

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 3 (2025)

```
[ ] <bound method DataFrame.info of          Country/Region  Confirmed  Deaths  Recovered  Active  New cases  \
    0           Afghanistan      36263    1269      25198    9796        106
    1               Albania       4880     144       2745    1991        117
    2               Algeria      27973    1163      18837    7973        616
    3               Andorra        907      52        803      52         10
    4                Angola        950      41        242     667         18
    ..                  ...        ...     ...        ...     ...        ...
    182  West Bank and Gaza      10621      78       3752    6791        152
    183      Western Sahara         10       1          8       1          0
    184               Yemen       1691     483        833     375         10
    185              Zambia       4552     140       2815    1597         71
    186            Zimbabwe       2704      36        542    2126        192

         New deaths  New recovered  Deaths / 100 Cases  Recovered / 100 Cases  \
    0            10             18                3.50                  69.49
    1             6             63                2.95                  56.25
    2             8            749                4.16                  67.34
    3             0              0                5.73                  88.53
    4             1              0                4.32                  25.47
    ..          ...            ...                 ...                    ...
    182           2              0                0.73                  35.33
    183           0              0               10.00                  80.00
    184           4             36               28.56                  49.26
    185           1            465                3.08                  61.84
    186           2             24                1.33                  20.04

         Deaths / 100 Recovered  Confirmed last week  1 week change  \
    0                      5.04                35526            737
    1                      5.25                 4171            709
    2                      6.17                23691           4282
    3                      6.48                  884             23
    4                     16.94                  749            201
    ..                      ...                  ...            ...
    182                    2.08                 8916           1705
    183                   12.50                   10              0
    184                   57.98                 1619             72
    185                    4.97                 3326           1226
    186                    6.64                 1713            991

         1 week % increase             WHO Region
    0                 2.07  Eastern Mediterranean
    1                17.00                 Europe
    2                18.07                 Africa
    3                 2.60                 Europe
    4                26.84                 Africa
    ..                 ...                    ...
```

Figure 2: Snapshot of the df.info function results

We use quantitative research design, as recommended by Bi et al. (2020) for time variant community structures (Bi et al., 2020). We have used multivariate regression analysis to examine multiple predictor variables and their relationships with life satisfaction simultaneously. This method allows for a more comprehensive examination of the interactions among independent variables compared to bivariate correlation analysis, which only examines relationships between pairs of variables. By using multivariate regression, we aim to gain deeper insights into how age, income, and education level collectively influence life satisfaction, thus improving the predictive accuracy and relevance of our findings.

Research Strategy

This study employs a comprehensive algorithmic approach to analyze the global impact of COVID-19 using a dataset that tracks pandemic indicators such as confirmed cases, deaths, and recoveries across 187 nations. The research is structured around three key areas:

1. **Community-Based Network Analysis**: The first phase involves applying community detection algorithms to identify clusters of nations that exhibit similar pandemic trajectories. By grouping countries with comparable patterns in case progression, fatalities, and recoveries, the algorithm will facilitate a comparative analysis aimed at identifying the differences and

similarities in community-level responses. This approach will help isolate common patterns and responses to the pandemic (Yassine et al., 2021).

2.

3. **Correlation Analysis**: In the second phase, statistical modeling will be used to assess the relationships between diverse COVID-19 metrics (e.g., confirmed cases, fatalities, recoveries) and derivative indicators such as recovery rates and death rates over time. The goal is to uncover significant correlations and predictive relationships that could provide insight into the dynamics of pandemic containment or spread. This process will highlight key factors that may influence the course of the pandemic in different nations (Biju et al., 2023).

4. **Data Visualization**: The final phase focuses on creating visualizations of critical COVID-19 trends, offering clear representations of the relationships between various pandemic indicators. By mapping these metrics, the study aims to identify specific country-level trends and patterns. These visualizations will not only display pandemic progress over time but also allow for a more intuitive understanding of how various factors interrelate (Scepanovic et al., 2020).

Integrating community detection, correlation analysis, and visualization techniques, the proposed study models for an improved outbreak response and prediction of the global pandemic. The model aims to provide a clearer interpretation of the statistics of global dynamics of the pandemic. This model allows for further comparative analysis aimed at identifying the differences and similarities among community-level experiences depicted in the code snippet in figure 3.

```
# Adding nodes and edges
for idx, row in df.iterrows():
    G.add_node(idx, **row.to_dict())

# Connecting nodes based on some criteria (you may need to customize this)
for i in range(len(df)):
    for j in range(i + 1, len(df)):
        G.add_edge(i, j)

# Applying Louvain Modularity
communities = list(greedy_modularity_communities(G))

# Mapping nodes to community
community_mapping = {node: i for i, comm in enumerate(communities) for node in comm}
nx.set_node_attributes(G, community_mapping, 'community')

# Drawing the graph with nodes colored by community
pos = nx.spring_layout(G)  # You can use different layout algorithms
node_colors = [community_mapping[node] for node in G.nodes]
nx.draw(G, pos, node_color=node_colors, with_labels=True, cmap=plt.cm.get_cmap("viridis"))
```

Figure 3: Code For Community Detection using Louvain Modularity

Sampling Technique

The Kaggle dataset lacks information on the origin and sampling methodology. The dataset covers a wide range of countries with different geographic locations, economic statuses, and pandemic outcomes. A stratified sampling approach ensures representation across various global parameters such as case counts, income levels, and regional locations. The selection pattern includes a diverse range of nations. As a result, the dataset captures a representative cross-section of international COVID-19 data, allowing for comparative analysis across different country groups based on specific characteristics.

Analysis, Results and Discussion

Data Cleaning and Preprocessing

To manage the COVID-19 indicator data for the five sample countries, an initial dictionary, titled `Data`, is created to store relevant metrics. Each key in the dictionary represents a specific data column, such as recovery percentage or confirmed cases, and each row corresponds to a country's data. This dictionary format organizes the data, preparing it for conversion into a DataFrame (Weir, 2020).

Using the Pandas library, the dictionary is transformed into a DataFrame object named `*df* through the DataFrame constructor. This conversion organizes the data into a tabular format with rows and columns, accommodating fifteen distinct metrics and indicators. Pandas facilitates data management by offering a high-level, optimized interface for statistical data operations, including effectively handling null values. The `.isnull()` method is employed to identify missing values in each column, ensuring no zeros and confirming the presence of missing data.

Subsequent data cleaning steps include using `.drop_duplicates()` to remove any duplicate rows, thus maintaining data integrity and ensuring that the results of the study are not skewed by repeated entries. The `.astype()` method is then used to typecast the WHO Region field as a categorical variable, enhancing readability by representing area names with alphanumeric characters rather than numerical codes.

The first five entries of the DataFrame are printed to verify the preprocessing and cleaning. This provides a snapshot of country names and confirms that the data has been correctly structured and cleaned. With these steps completed, the data is prepared for statistical analysis, machine learning model training, or further simulations (Song et al., 2019). Data purity is crucial for maintaining quality and avoiding spurious results in subsequent analyses.

Data Visualization

This code utilizes the Seaborn and Matplotlib packages to create visualizations of the data. The Seaborn attribute is set to a whitegrid style, presenting data with a white background and grey gridlines. The first visualization is a bar chart displaying the top ten countries based on the number of confirmed cases. The DataFrame is sorted to achieve this, and the top ten rows are selected. The bar plot uses the "viridis" color palette, with countries on the y-axis and cases on the x-axis. Figure 4 shows the first 5 countries (alphabetically) from the dataset and their confirmed cases of COVID-19 infection.

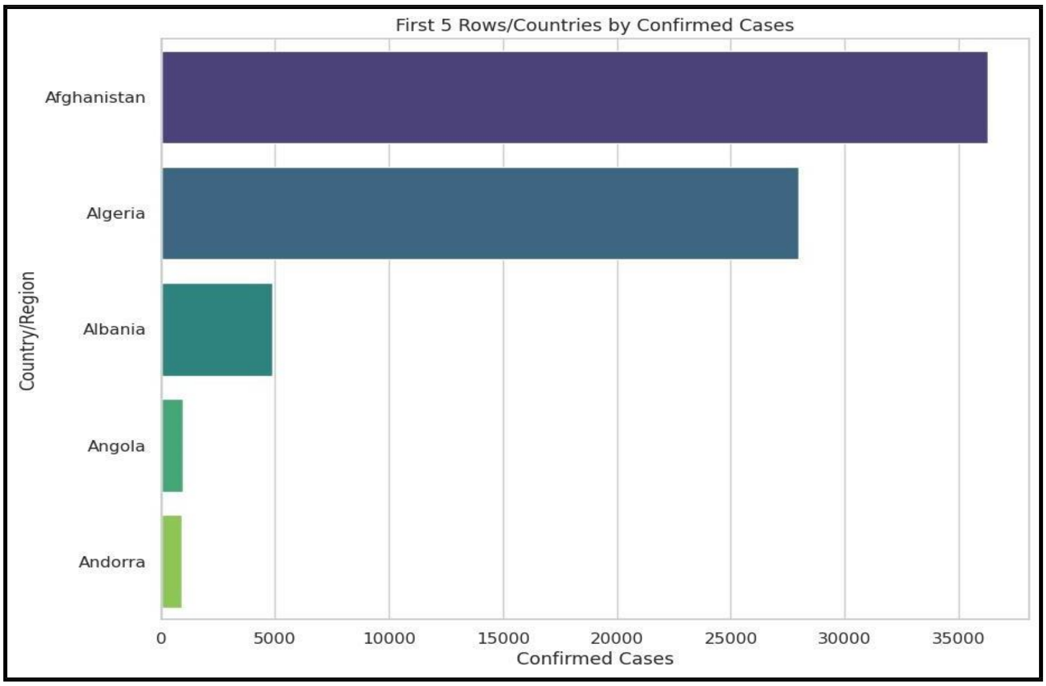

Figure 4: Bar plot for the first five rows/countries with confirmed COVID-19 cases

The process of visualizing data involves several important steps for effectively presenting and analyzing COVID-19 metrics for specific countries. These steps are as follows:

1. Bar Plot: A bar plot is created to compare the number of confirmed COVID-19 cases across five countries: Afghanistan, Algeria, Albania, Angola, and Andorra. The plot shows that Afghanistan has the highest number of confirmed cases, followed by Algeria. This visual representation makes it easy to compare the magnitude of confirmed cases among these countries (Abdulla & Khasawneh, 2022).
2. Scatter Plot: The second visualization is a scatter plot that compares fatalities per 100 cases with recoveries per 100 cases shown in figure 5. Each dot on the plot represents a country and is color-coded according to the WHO region, with the size of the dot proportional to the number of cases. The scatter plot uses the "Set1" color palette for differentiation. The plot is carefully formatted with a legend indicating the WHO regions, clear labels on the axes, and an informative title to enhance readability (Zheng et al., 2021).
3. Correlation Matrix Heat Map: A correlation matrix heat map visualizes the relationships between variables. This heat map, generated using Seaborn’s heatmap tool, displays the correlation coefficients ranging from -1 to 1, indicating the strength and direction of relationships between variables shown in figure 6. The heat map includes detailed dimensions, formatted fonts, titles, and a color bar legend to aid in interpreting the correlations (Cassule et al., 2023; Ma et al., 2020).

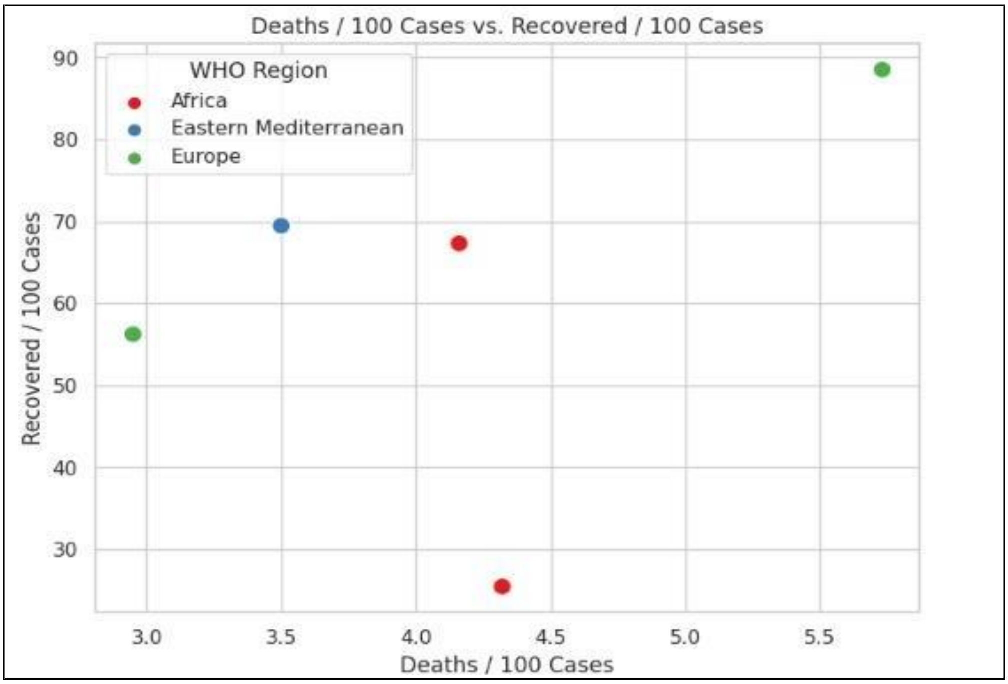

Figure 5: Displaying the scatter plot for deaths and recovery cases in WHO regions

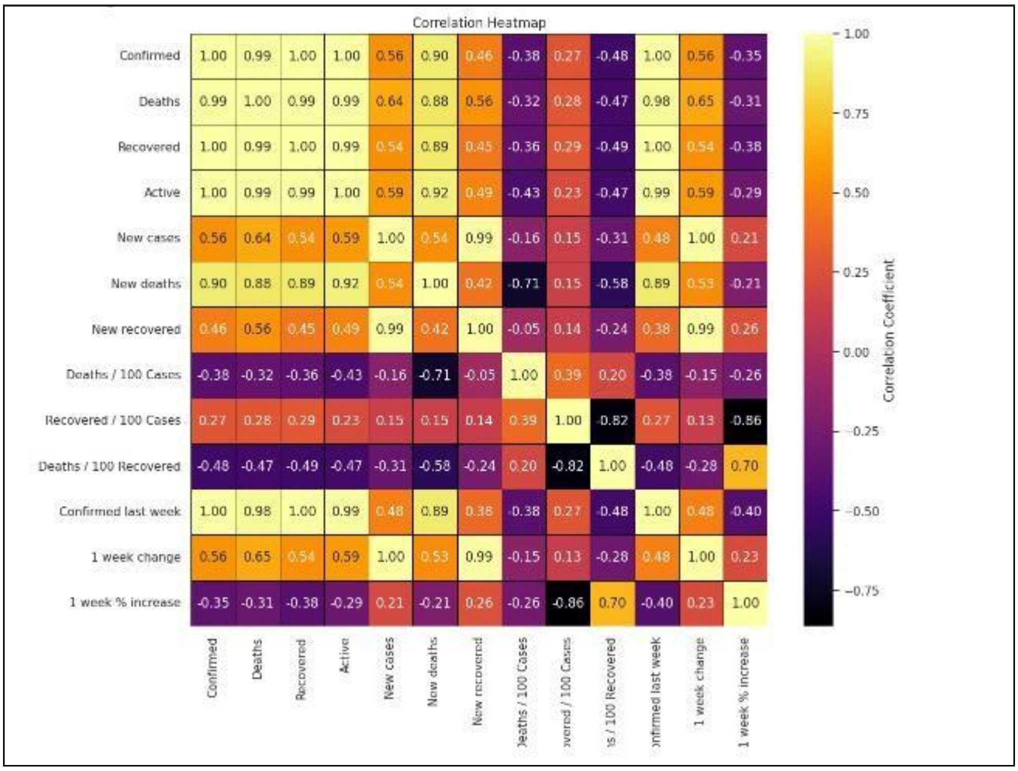

Figure 6: Correlation Heat map Visualization for infection and recovery statistics

The visualizations provide an in-depth exploration of the COVID dataset by analyzing correlations among variables, comparing results between countries, and examining mortality versus recovery (Lai et al., 2023). Using just a few lines of code, they leverage the integrated pandas and Matplotlib/seaborn functions to enable quick evaluation and visualization. As a result, these visualizations provide insights into the COVID-19 infection and recovery data for the decision-makers.

Data Modelling

To effectively visualize and analyze COVID-19 data, Matplotlib, NetworkX, and Pandas are employed. Initially, data is imported into Pandas and NetworkX for dataframe manipulation and network graph creation, respectively. A DataFrame, populated with COVID-19 metrics, serves as the foundation for the network model. A NetworkX `Graph` instance is instantiated to represent this network, where each row in the DataFrame corresponds to a node in the graph, with attributes derived from the DataFrame's rows (Pan et al., 2022). Edges are then established by iterating over all possible node pairs, with inclusion criteria tailored to the specific analysis requirements, resulting in a fully connected network that reflects node connectivity and similarity.

Community detection is performed using the Louvain algorithm implemented in NetworkX, which identifies clusters or modules of densely linked nodes, thereby revealing distinct communities within the network (M. Liu, 2022). For visualization, the graph is arranged using the Spring layout algorithm, which positions nodes based on their connectivity. Node colors are assigned using the continuous Viridis colormap, and node labels are derived from the DataFrame index, facilitating clear identification of local detection results and regional groupings (Dwi Putra Aditama & SN, 2020). The entire process, including the transformation of the DataFrame into a graph, clustering of nodes, and visualization, is executed efficiently in a succinct Python codebase, providing a robust framework for network analysis and visualization. Figure 7 shows the results of the Louvain modularity for the first five countries (alphabetically) from the dataset.

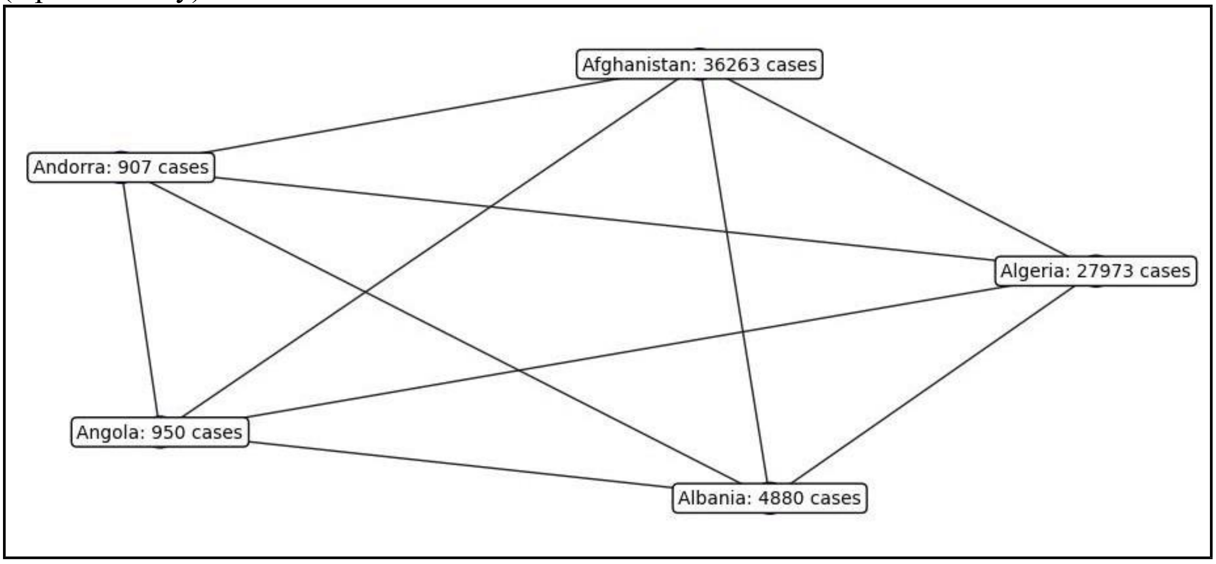

Figure 7: Plotting the result nodes generated using Louvain Modularity for alphabetically top 5 countries

Conclusion, Limitations, And Future Research Directions

This study utilized data visualization, correlation analysis, and network modeling techniques to conduct a comprehensive cross-national analysis of COVID-19 data. By developing a predictive model for identifying the early stages of major outbreaks, the research aimed to improve both the speed and precision of response strategies. The results highlight the immense potential of these models to enhance outbreak forecasting and enable timely interventions. Through thorough data cleaning and the creation of preliminary visualizations as comparative graphs and correlation matrix, the study uncovered key insights into the core relationships between the number of cases, deaths, and recoveries in each of the countries. Additionally, the application of community detection methods identified clusters of countries with similar transmission patterns, effectively modeling these nations as interconnected networks based on COVID-19 characteristics. These findings demonstrate the efficacy of integrating machine learning and network science approaches to support global pandemic response efforts and contribute to more informed decision-making during health crises.

This study has some limitations. First, the dataset relied solely on publicly available information, which may not fully capture the complexity of pandemic dynamics due to incomplete or inconsistent reporting across countries. Moreover, the static nature of the analysis does not account for the temporal evolution of the pandemic, potentially impacting the accuracy of the predictive models. Additionally, the simplified network models and assumptions in community detection may not adequately reflect the complexities of real-world epidemic spread. Finally, the generalizability of the model may be constrained by regional variations and the specific characteristics of the data used, limiting its applicability across different geographical contexts.

Future research will incorporate dynamicity in network analysis on integrated real-world pandemic datasets. This approach would improve the model's predictive capabilities, enabling it to identify early signs of potential future outbreaks more effectively. Integration of large-scale datasets with machine learning models will enhance the accuracy of predictions and adapt the model to evolving pandemic scenarios. Future studies will also incorporate temporal aspects of dynamic networks and apply unsupervised machine learning methods to generate timely predictions for better decision-making. Researchers may compare new mathematical models to improve the groundwork laid by this study. Additionally, integrating statistical distributions representing large, real-world irregular units could result in more realistic and localized architectural models, providing a global perspective on pandemic impacts extending beyond national boundaries.

References

1. Abdulla, M., & Khasawneh, M. T. (2022). Louvain clustering integration within density-based graph classification (Louvain dbGC) in Schizophrenia. *IISE Transactions on Healthcare Systems Engineering*, *12*(1). https://doi.org/10.1080/24725579.2021.1933268
2. Alsolbi, I., Agarwal, R., Unhelkar, B., Al-Jabri, T., Samarawickrama, M., Tafavogh, S., & Prasad, M. (2023). A Conceptual Design of an AI-Enabled Decision Support System for Analysing Donor Behaviour in Nonprofit Organisations. *Information (Switzerland)*, *14*(10), 578–602. https://doi.org/10.3390/info14100578
3. Bi, Q., Shen, L., Evans, R., Zhang, Z., Wang, S., Dai, W., & Liu, C. (2020). Determining the topic evolution and sentiment polarity for albinism in a Chinese online health community: Machine learning and social network analysis. *JMIR Medical Informatics*, *8*(5). https://doi.org/10.2196/17813
4. Biju, S. M., Mishra, A., & Kumar, M. (2023). Advanced interdisciplinary applications of machine learning python libraries for data science. In *IGI Global*. https://doi.org/10.4018/978-1-6684-8696-2
5. Bolaji, A. A. (Purdue U. G. S. (2021). *Community Detection of Anomalies in Large-Scale Network* (Issue May).
6. Cassule, M., Renato, L., & Silva, L. A. (2023). Discovery of Attention Deficit Hyperactivity Disorder Spectrum Using Community Detection. *Preprints*, *June*, 1–12.

https://doi.org/10.20944/preprints202306.1211.v1
7. Čížková, K. (2022). Comparing two main community detection algorithms and their applications on human brains. In *Charles University*. http://repositorio.unan.edu.ni/2986/1/5624.pdf%0Ahttp://fiskal.kemenkeu.go.id/ejournal%0Ahttp://dx.doi.org/10.1016/j.cirp.2016.06.001%0Ahttp://dx.doi.org/10.1016/j.powtec.2016.12.055%0Ahttps://doi.org/10.1016/j.ijfatigue.2019.02.006%0Ahttps://doi.org/10.1
8. Dadkhah, S., Shoeleh, F., Yadollahi, M. M., Zhang, X., & Ghorbani, A. A. (2021). A real-time hostile activities analyses and detection system. *Applied Soft Computing*, *104*. https://doi.org/10.1016/j.asoc.2021.107175
9. Devakumar, K. P. (n.d.). *COVID-19 Dataset*. Retrieved April 25, 2025, from https://www.kaggle.com/datasets/imdevskp/corona-virus-report
10. Dwi Putra Aditama, T. B., & SN, A. (2020). Determining Community Structure and Modularity in Social Network using Genetic Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *14*(3), 219. https://doi.org/10.22146/ijccs.57834
11. Fahad Alkhamees, B., Mosleh, M. A. A., Alsalman, H., & Azeem Akbar, M. (2021). An Effective Approach for Modular Community Detection in Bipartite Network Based on Integrating Rider with Harris Hawks Optimization Algorithms. In *Journal of Mathematics* (Vol. 2021). https://doi.org/10.1155/2021/9511425
12. Gulati, K., Unhelkar, B., Khatri, E., Abdul, S. M., Choubey, S., & Patni, I. (2024). The Impact of AI and IoT-Driven Systems on the Social and Psychological Aspects of Employee Management in the Banking Sector. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(8s), 357–366.
13. Havakhor, T., Golmohammadi, A., Sabherwal, R., & Gauri, D. K. (2023). Do Early Words From New Ventures Predict Fundraising? A Comparative View of Social Media Narratives. *MIS Quarterly: Management Information Systems*, *47*(2). https://doi.org/10.25300/MISQ/2022/16392
14. Jairam-Owthar, D., Naicker, V., & Unhelkar, B. (2022). The effect of performance management of strategy on an executive decision support framework for the financial service sector of a developing economy. *EUREKA: Social and Humanities*, *6*, 3–14. https://doi.org/10.21303/2504-5571.2022.002705
15. Lai, S., Li, J., & Lu, Y. (2023). A Comprehensive Review of Community Detection in Graphs. *ArXiv*.
16. Lee, J. H., Sato, N., Yano, K., & Miyake, Y. (2022). Universal association between depressive symptoms and social-network structures in the workplace. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-14366-9
17. Liu, M. (2022). Outbreak Detection From Virus Genetic Sequence Variation By Community Detection. In *The University of Western Ontario*. https://ir.lib.uwo.ca/etd/8781
18. Liu, Y., Xu, X., Jin, Y., & Deng, H. (2023). Understanding The Digital Resilience of Physicians During The Covid-19 Pandemic: An Empirical Study. *MIS Quarterly: Management Information Systems*, *47*(1), 391–422. https://doi.org/10.25300/MISQ/2022/17248
19. Ma, Y., Zhao, Y., Wang, J., Liu, M., Shen, W., & Ma, Y. (2020). Modularity-based incremental label propagation algorithm for community detection. *Applied Sciences (Switzerland)*, *10*(12). https://doi.org/10.3390/app10124060
20. Matherly, T., & Greenwood, B. (2024). No News is Bad News: Political Corruption, News Deserts, and the Decline of the Fourth Estate. *Management Information Systems Quarterly*, *48*(2), 699–714. https://doi.org/10.5465/ambpp.2021.10153abstract

21. Matta, J., Singh, V., Auten, T., & Sanjel, P. (2023). Inferred networks, machine learning, and health data. *PLoS ONE*, *18*(1 January). https://doi.org/10.1371/journal.pone.0280910
22. Nallusamy, K., & Easwarakumar, K. S. (2023). Classifying schizophrenic and controls from fMRI data using graph theoretic framework and community detection. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *12*(1). https://doi.org/10.1007/s13721-023-00415-4
23. Oza, P., Agrawal, S., Ravaliya, D., & Kakkar, R. (2023). EVALUATING THE IGRAPH COMMUNITY DETECTION ALGORITHMS ON DIFFERENT REAL NETWORKS. *Scalable Computing*, *24*(2). https://doi.org/10.12694/scpe.v24i2.2102
24. Pan, Y., Zhang, L., Unwin, J., & Skibniewski, M. J. (2022). Discovering spatial-temporal patterns via complex networks in investigating COVID-19 pandemic in the United States. *Sustainable Cities and Society*, *77*, 103508. https://doi.org/10.1016/j.scs.2021.103508
25. Pang, M. S., & Lee, G. (2022). The Impact of IT Decision-Making Authority on IT Project Performance in The U.S. Federal Government. *MIS Quarterly: Management Information Systems*, *46*(3), 1759–1776. https://doi.org/10.25300/MISQ/2022/16898
26. Ranjan, S., Singh, I., Dua, S., & Sood, S. (2018). Sentiment Analysis of Stock Blog Network Communities for Prediction of Stock Price Trends. *Indian Journal of Finance*, *12*(12), 7–21.
27. Rostami, M., Oussalah, M., Berahmand, K., & Farrahi, V. (2023). Community Detection Algorithms in Healthcare Applications: A Systematic Review. *IEEE Access*, *11*, 30247–30272. https://doi.org/10.1109/ACCESS.2023.3260652
28. Rustamaji, H. C., Kusuma, W. A., Nurdiati, S., & Batubara, I. (2023). Community-Consideration Centrality, a Case Study of Lung Cancer Proteins. *IEEE Access*, *11*, 134340–134354. https://doi.org/10.1109/ACCESS.2023.3337375
29. Safadi, H., Lalor, J., & Berente, N. (2024). The Effect of Bots on Human Interaction in Online Communities. *Management Information Systems Quarterly*, *48*(3), 1279–1296.
30. Scepanovic, S., Martin-Lopez, E., Quercia, D., & Baykaner, K. (2020). Extracting medical entities from social media. *ACM CHIL 2020 - Proceedings of the 2020 ACM Conference on Health, Inference, and Learning*, 170–181. https://doi.org/10.1145/3368555.3384467
31. Shabdiz, M., Azarbar, A., & Azgomi, H. (2024). Hyperbolic decision-making based on RBF for diabetic/COVID-19 iris. *Engineering Applications of Artificial Intelligence*, *130*. https://doi.org/10.1016/j.engappai.2023.107589
32. Song, Y., Dana, L. P., & Berger, R. (2019). The entrepreneurial process and online social networks: forecasting survival rate. *Small Business Economics*. https://doi.org/10.1007/s11187-019-00261-7
33. Unhelkar, B., & Gonsalves, T. (2021). Artificial Intelligence for Business Optimization. In *Artificial Intelligence for Business Optimization*. https://doi.org/10.1201/9781003120926
34. Weir, W. H. (2020). *Modularity-based approaches to community detection in multilayer networks with applications toward precision medicine*.
35. Westarb, G., Stefenon, S. F., Hoppe, A. F., Sartori, A., Klaar, A. C. R., & Leithardt, V. R. Q. (2023). Complex graph neural networks for medication interaction verification. *Journal of Intelligent and Fuzzy Systems*, *44*(6). https://doi.org/10.3233/JIFS-223656
36. Wickramasinghe, A. N. (2021). *Community detection in social networks with an application to COVID-19 data*.
37. Yassine, S., Kadry, S., & Sicilia, M. A. (2021). Application of community detection algorithms on learning networks. The case of Khan Academy repository. *Computer*

Applications in Engineering Education, *29*(2). https://doi.org/10.1002/cae.22212

38. Zarezadeh, M., Nourani, E., & Bouyer, A. (2022). DPNLP: distance based peripheral nodes label propagation algorithm for community detection in social networks. *World Wide Web*, *25*(1). https://doi.org/10.1007/s11280-021-00966-4
39. Zhang, L., Priestley, J., Demaio, J., Ni, S., & Tian, X. (2021). Measuring Customer Similarity and Identifying Cross-Selling Products by Community Detection. *Big Data*, *9*(2). https://doi.org/10.1089/big.2020.0044
40. Zheng, M., Domanskyi, S., Piermarocchi, C., & Mias, G. I. (2021). Visibility graph based temporal community detection with applications in biological time series. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-84838-x