2021

Understanding the Behavior of the COVID-19 Pandemic Using Data Analytics

Mohammad Reza Davahli
University of Central Florida

UNDERSTANDING THE BEHAVIOR OF THE COVID-19 PANDEMIC USING DATA ANALYTICS

by

MOHAMMAD REZA DAVAHLI
B.S. Isfahan University of Technology, 2007
M.Sc. University of Central Florida, 2019

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering & Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2021

Major Professor: Waldemar Karwowski

ABSTRACT

In December 2019, China announced the breakout of a new virus identified as coronavirus SARS-CoV-2 (COVID-19), which soon grew exponentially and became a global pandemic. Despite strict actions to mitigate the spread of the virus in various countries, the COVID-19 pandemic resulted in a significant loss of human life in 2020 and 2021. To better understand the pandemic, this doctoral research incorporated data analytics to evaluate the behavior and impacts of the virus. The doctoral research contributed to the scientific body of the knowledge in different ways including (1) presenting a systematic literature review of current research and topics about impacts of the COVID-19 pandemic; (2) predicting the dynamics of the COVID-19 pandemic using deterministic and stochastic Recurrent Neural Networks; (3) predicting the dynamics of the COVID-19 pandemic using graph neural networks; and (4) analyzing the dynamics of the COVID-19 pandemic using graph theoretical method. This dissertation is sorted out as a manuscript-style including four published journal articles. The results of this doctoral research provide a comprehensive view of the behavior and impacts of the COVID-19 pandemic.

To my mother…

ACKNOWLEDGMENT

First things first, praise is due to ALLAH who kindly help me to complete my PhD journey. I want to thank my late father, who taught me to work hard. Deepest thanks to my mother who constantly serves as a source of inspiration, encouragement, and love.

I would like to especially thank my PhD advisor, Dr. Waldemar Karwowski, for his endless support and guidance over years. You taught me how to conduct high-quality research. You were very patient and kind to me. You have set an example of excellence as a mentor, researcher, instructor, and role model.

I wish also to offer my special thanks to Dr. Thomas O'Neal and Dr. Vernet Lasrado, for trusting me and hiring me as a graduate researcher in 2018. I still remember the day I came to Dr. O'Neal office and he kindly agreed to hire me and give me funding. I will never forget your kindness.

I would also present my gratitude to members of my dissertation committee, Dr. Ahmad Elshennawy, and Dr. Piotr Mikusinski; it has been an honor to have you all on this committee.

I would like to especially thank Dr. Krzysztof Fiok for helping me to develop proper methodologies. You have had a huge influence on my career path, and I am very grateful for that.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AI	Artificial Intelligence
AMR	Antimicrobial Resistance
ARIMA	Autoregressive integrated moving average
DL	Deep Learning
EMD	Empirical Mode Decomposition
ESOM	Evolutionary Self-Organizing Map
GNN	Graph Neural Networks
GPU	Graphic Processing Unit
GTNN	Graph Theory based Neural Networks
IMFs	Intrinsic Mode Functions
LEs	Lyapunov Exponents
LSTM	Long Short Term Memory
ML	Machine Learning
MTN	Mixture Density Networks
NGNN	Neighborhood-based Graph Neural Networks
NLP	Natural Language Processing

RNN	Recurrent Neural Networks
SEIR	Susceptible-Exposed-Infectious-Recovered
SIMDR	Susceptible-Infected Model with Multi-Drug Resistance
SIR	Susceptible-Infectious-Recovered
SLR	Systematic Literature Review
sMAPE	Symmetric Mean Absolute Percentage Error
SOM	Self-Organizing Map
UN/DESA	The Department of Economic and Social Affairs of the United Nations Secretariat

CHAPTER 1: INTRODUCTION OF THE COVID-19 PANDEMIC AND SEQUENCE-LEARNING PREDICITVE MODELS

In this chapter of the thesis, overall introduction about understanding the COVID-19 pandemic by using data analytics is presented. The structure of the dissertation is explained in the following introduction.

1.1 Overview

On December 8, 2019, the government of Wuhan, China, announced that health authorities were treating dozens of cases of a new virus, identified as coronavirus disease 2019 (COVID-19) [1]. Since then, COVID-19, a new strain of SARS (SARS-CoV-2), has grown into a global pandemic and spreading across many countries. A highly transmissible respiratory disease, COVID-19 spreads through contact with other infected individuals, with symptoms such as fever, cough, and difficulty breathing [2]. Transmission can also occur from asymptomatic individuals, with up to 40% of infected persons remaining asymptomatic [3]. Other factors that facilitate infection include (1) speed and efficiency of COVID-19 transmission; (2) airborne transmission [4]; (3) close contact between infected and non-infected individuals; (4) vulnerability of immunocompromised individuals with specific underlying health conditions (e.g., hypertension, diabetes, cardiovascular disease, respiratory problems); (5) susceptibility of persons over 65; and (7) contact with persons who have traveled to locations with a high number of cases [5].

Critical global responses to control the spreading of the COVID-19 pandemic have included travel restrictions, shelter-in-place, social distancing orders, and developing vaccines. Most countries around the world have imposed partial or complete border closures (at the time of writing), with travel bans affecting the world's population [6]. With millions suddenly unemployed, uncertainty over economic recovery, and global fears of continuing COVID-19 spread and its future waves and variants, the world economy was under threat [7].

Regarding the COVID-19 pandemic, some of the main questions are including when the pandemic is going to end, how accurately we can predict the pandemic, how serious the pandemic is, and what are the main impacts of the pandemic. Majority of researchers developed simulations and mathematical models to better understand the dynamics of the pandemic. However, developing analytics based on data is very powerful methodology to determine the behavior and impacts of the pandemic.

1.2 Research Objectives

This research particularly investigates the behavior and impacts of the COVID-19 pandemic through developing different data analytics. The main objectives of this study are to 1) study the impacts of the COVID-19 pandemic by using a systematic literature review technique [8]; 2) develop a sequence learning models without considering message passing [9]; and 3) developing a sequence learning models with message passing [10].

1.3 Organization of Dissertation

The organization of this dissertation is represented in Figure 1. Chapter 1 provides an introduction and general view for the whole dissertation. Chapter 2 discussed the impacts of the COVID-19 through conducting systematic literature review [8]. Fifty peer reviewed journal articles were included and analyzed to draw comprehensive conclusion about the impacts of the pandemic. Chapter 3 is about developing sequence-learning models by using stochastic and deterministic recurrent neural networks [9]. Chapter 4 is about developing sequence-learning models by using graph neural networks [10]. Chapter 5 focuses on graph theory and its application on the pandemic and finally chapter 6 drew a conclusion with overall results, discussed future work.

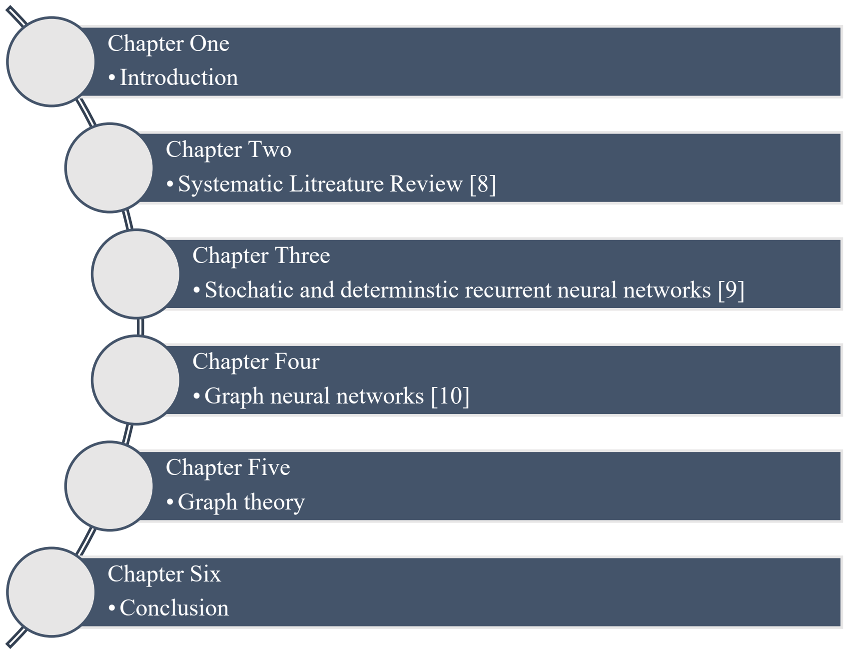

Figure 1: Organization of the dissertation

1.4 References

[1] N. A. Bakar and S. Rosbi, “Effect of Coronavirus disease (COVID-19) to tourism industry,” *International Journal of Advanced Engineering Research and Science*, vol. 7, no. 4, 2020, [Online]. Available: Retrieved from http://journal-repository.com/index.php/ijaers/article/view/1858

[2] Y. Yang, H. Zhang, and X. Chen, “Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak,” *Annals of Tourism Research*, p. 102913, 2020, doi: Advance online publication. https://doi.org/10.1016/j.annals.2020.102913.

[3] D. P. Oron and E. J. Topol, “Getting a handle on asymptomatic SARS-CoV-2 infection,” Apr. 2020, [Online]. Available: Retrieved from https://www.scripps.edu/science-and-medicine/translational-institute/about/news/sarc-cov-2-infection/index.html, April 20, 2020

[4] L. Morawska and J. Cao, “Airborne transmission of SARS-CoV-2: The world should face the reality,” *Environment International*, vol. 139, p. 105730, 2020, doi: https://doi.org/10.1016/j.envint.2020.105730.

[5] N. C. Peeri *et al.*, “The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned?,” *International journal of epidemiology*, 2020, doi: https://doi.org/10.1093/ije/dyaa033.

[6] P. Connor, “More than nine-in-ten people worldwide live in countries with travel restrictions amid COVID-19,” *Pew Research Center*, 2020. Retrieved from https://www.pewresearch.org/fact-tank/2020/04/01/more-than-nine-in-ten-people-

worldwide-live-in-countries-with-travel-restrictions-amid-covid-19/, April 1, 2020 (accessed May 29, 2020).

[7] A. Tappe and T. Luhby, "22 million Americans have filed for unemployment benefits in the last four weeks," *CNN*, 2020. Retrieved from https://www.cnn.com/2020/04/16/economy/unemployment-benefits-coronavirus/index.html, April 16, 2020 (accessed May 22, 2020).

[8] M. R. Davahli, W. Karwowski, S. Sonmez, and Y. Apostolopoulos, "The Hospitality Industry in the Face of the COVID-19 Pandemic: Current Topics and Research Methods," *International Journal of Environmental Research and Public Health*, vol. 17, no. 20, p. 7366, 2020.

[9] M. R. Davahli, W. Karwowski, and K. Fiok, "Optimizing COVID-19 vaccine distribution across the United States using deterministic and stochastic recurrent neural networks," *PLOS ONE*, vol. 16, no. 7, p. e0253925, Jul. 2021, doi: 10.1371/journal.pone.0253925.

[10] M. R. Davahli, K. Fiok, W. Karwowski, A. M. Aljuaid, and R. Taiar, "Predicting the Dynamics of the COVID-19 Pandemic in the United States Using Graph Theory-Based Neural Networks," *International Journal of Environmental Research and Public Health*, vol. 18, no. 7, Art. no. 7, Jan. 2021, doi: 10.3390/ijerph18073834.

[11] N. Sapkota *et al.*, "The Chaotic Behavior of the Spread of Infection During the COVID-19 Pandemic in the United States and Globally," *IEEE Access*, vol. 9, pp. 80692–80702, 2021, doi: 10.1109/ACCESS.2021.3085240.

[12] K. Fiok *et al.*, "A Study of the Effects of the COVID-19 Pandemic on the Experience of Back Pain Reported on Twitter® in the United States: A Natural Language Processing Approach,"

International Journal of Environmental Research and Public Health, vol. 18, no. 9, Art. no. 9, Jan. 2021, doi: 10.3390/ijerph18094543.

CHAPTER 2: SYSTIMATIC LITREATURE REVIEW

This chapter contained material previously published in: M. R. Davahli, W. Karwowski, S. Sonmez, and Y. Apostolopoulos, "The Hospitality Industry in the Face of the COVID-19 Pandemic: Current Topics and Research Methods," International Journal of Environmental Research and Public Health, vol. 17, no. 20, p. 7366, 2020.

The present study focuses on understanding the state of current research on the topic of the hospitality industry in the face of the COVID-19 pandemic. A systematic literature review (SLR) of the current literature is considered to identify and classify research that focuses on COVID-19 concerning the hospitality industry. The primary purpose of a systematic review is to identify, summarize, and analyze the findings of all relevant individual studies that are addressing predefined research questions [1]. The preferred reporting items for systematic reviews and meta-analyses (PRISMA) is a structured guideline for ensuring reliable and meaningful results of the systematic literature review studies. The PRISMA protocol consists of 27 items that help researchers prepare and report scientific evidence accurately and reliably, which improves the quality of research [1]. This review is structured as follows: the methodology section discusses inclusion and exclusion criteria and the risk of bias; the results and discussion section provides outputs of the literature search and describes the status of the hospitality industry at the time of COVID-19.

2.1 SLR on the COVID-19 Pandemic

The literature review follows Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [1], [2] and contains two main features: developing research questions and determining search strategy. The following research questions have guided this review:

RQ1. How research on the hospitality industry in the face of COVID-19 is conducted?

RQ2. What does current research reveal about the status of the hospitality industry at the time of COVID-19?

To answer these questions, a search strategy was developed to list and review all relevant scientific papers, by (a) defining keywords and identifying all relevant materials, (b) filtering the identified records, and (c) addressing the risk of any bias [1]. One of the main concerns in a systematic review is developing specific keywords. Herein, our objective was to target all critical segments of the hospitality industry (e.g., hotels, restaurants) as well as the broader tourism industry. Therefore, the keywords were defined, as shown in Table 1.

Table 1. Keywords used in the literature search.

Row	**Keywords**
Search 1	*COVID-19 AND hospitality industry*
Search 2	*COVID-19 AND event industry*
Search 3	*COVID-19 AND hotel industry*
Search 4	*COVID-19 AND restaurant industry*
Search 5	*COVID-19 AND tourism industry*

Web of Science, Science Direct, and Google Scholar were used as database search tools. Keywords were used to discover relevant articles and identify 175 articles with relevant content. Because this topic is rapidly evolving, it is important to mention that article discovery was finished at the end of July 2020. After the development of the main database and the identification of all relevant papers, a formal screening process based on specific exclusion and inclusion criteria was followed. The inclusion criteria were articles related to the hospitality industry and COVID-19, articles related to the research questions, and articles written in English. The exclusion criteria were papers written in other languages, book chapters, articles from secondary sources that were not free or open access, letters, newspaper articles, viewpoints, presentations, anecdotes, duplicated studies, and posters.

The screening of the titles, abstracts, conclusions, and keywords in the included papers after removing duplication (n = 168) resulted in the exclusion of articles (n = 115) because of not enough covering the hospitality industry. The remaining articles (n = 53) were read in full against the eligibility principle and three articles were excluded for poor quality and not representing the methodology.

Selection bias in a systematic review can occur by the erroneous application of inclusion/exclusion criteria and/or the specification of dimensions of included papers. To address the first type of bias, two researchers (MD and WK) independently reviewed the title, abstract, and conclusions of the identified records in order to select articles for the full-text review. Subsequently, the two researchers compared their selected articles to reach consensus. After reading the full text of the selected papers, the authors decided whether or not to include the article—which was considered and included upon reaching an agreement. Disagreements were resolved by the input of the other two authors (SS and YA). To address the second type of bias, two researchers (MD and WK) independently specified the classification of the included papers, and subsequently compared the results, resolving disagreements by consultation with the other authors (SS and YA). The selection strategy, as per PRISMA guidelines, is illustrated in Figure 2.

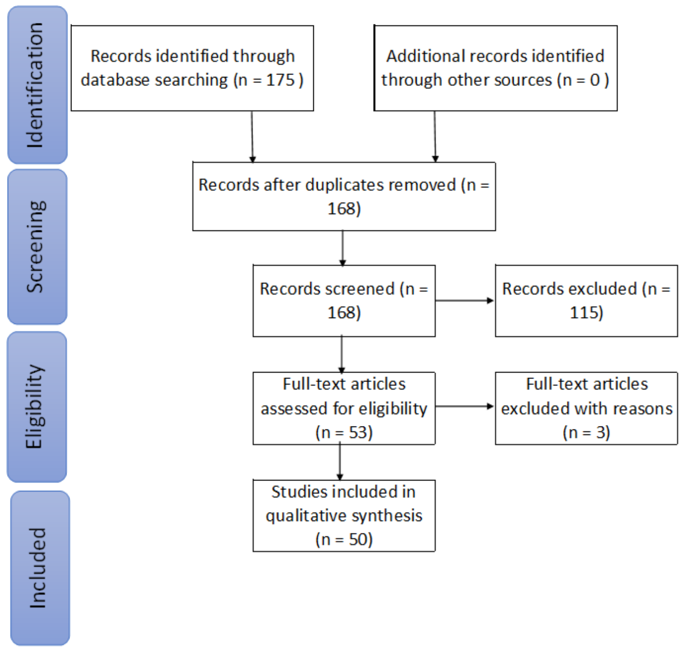

Figure 2. Chart of the selection strategy following PRISMA guidelines [1].

2.2 SLR Summary

All identified articles were categorized and stored in the main database according to year, source of publication, the segment of the industry, geographic location, and investigation approach. The list of included papers with their categorization is represented in Table 2.

Table 2. List of included papers

Reference	**Segment of industry**	**Geographic Location**	**Approach**
[3]	*Tourism industry*	*Global*	*Comparing COVID-19 with previous public health crises*
[4]	*Tourism industry*	*Global*	*Comparing COVID-19 with previous public health crises*
[5]	*Restaurant industry*	*United States*	*Conducting survey*
[6]	*Tourism industry*	*China*	*Conducting survey*

Reference	Segment of industry	Geographic Location	Approach
[7]	*Tourism industry*	*China*	*Conducting survey*
[8]	*Restaurant industry*	*United States*	*Conducting survey*
[9]	*Hospitality industry*	*Global*	*Conducting survey*
[10]	*Hospitality industry*	*United States*	*Measuring the impact of COVID-19*
[11]	*Hospitality industry*	*Philippine*	*Measuring the impact of COVID-19*
[12]	*Tourism industry*	*Turkey*	*Measuring the impact of COVID-19*
[13]	*Tourism industry*	*India*	*Measuring the impact of COVID-19*
[14]	*Hotel industry*	*Global*	*Measuring the impact of COVID-19*
[15]	*Tourism industry*	*The Diamond Princess cruise ship*	*Developing simulation & scenario modeling*
[16]	*Tourism industry*	*The Diamond Princess cruise ship*	*Developing simulation & scenario modeling*
[17]	*Tourism industry*	*The Diamond Princess cruise ship*	*Developing simulation & scenario modeling*
[18]	*Tourism industry*	*The Diamond Princess cruise ship*	*Developing simulation & scenario modeling*
[19]	*Tourism industry*	*The Diamond Princess cruise ship*	*Developing simulation & scenario modeling*
[20]	*Hospitality industry*	*Global*	*Developing simulation & scenario modeling*
[21]	*Tourism industry*	*Italy*	*Developing simulation & scenario modeling*

Reference	Segment of industry	Geographic Location	Approach
[22]	*Hospitality industry*	*Global*	*Developing simulation & scenario modeling*
[23]	*Tourism industry*	*Austria*	*Reporting the impacts of the Covid-19 pandemic*
[24]	*Restaurant industry*	*China*	*Reporting the impacts of the Covid-19 pandemic*
[25]	*Hospitality industry*	*China*	*Reporting the impacts of the COVID-19 pandemic*
[26]	*Tourism industry*	*China*	*Reporting the impacts of the COVID-19 pandemic*
[27]	*Hospitality industry*	*China*	*Reporting the impacts of the COVID-19 pandemic*
[28]	*Hotel industry*	*China*	*Reporting the impacts of the COVID-19 pandemic*
[29]	*Tourism industry*	*India*	*Reporting the impacts of the COVID-19 pandemic*
[30]	*Hospitality industry*	*Malaysia*	*Reporting the impacts of the COVID-19 pandemic*
[31]	*Tourism Industry*	*India*	*Reporting the impacts of the COVID-19 pandemic*
[32]	*Tourism industry*	*India*	*Reporting the impacts of the COVID-19 pandemic*
[33]	*Hospitality industry*	*Global*	*Reporting the impacts of the COVID-19 pandemic*
[34]	*Hotel industry*	*Global*	*Reporting the impacts of the COVID-19 pandemic*
[35]	*Hospitality industry*	*Global*	*Reporting the impacts of the COVID-19 pandemic*
[36]	*Tourism industry*	*Ghana*	*Reporting the impacts of the COVID-19 pandemic*
[37]	*Tourism industry*	*Nepal*	*Reporting the impacts of the COVID-19 pandemic*

Reference	Segment of industry	Geographic Location	Approach
[38]	*Tourism industry*	*China*	*Reporting the impacts of the COVID-19 pandemic*
[39]	*Hospitality industry*	*Europe*	*Reporting the impacts of the COVID-19 pandemic*
[40]	*Tourism industry*	*Indonesia*	*Review and recommendation*
[41]	*Hospitality industry*	*Global*	*Review and recommendation*
[42]	*Restaurant industry*	*India*	*Review and recommendation*
[43]	*Hotel industry*	*India*	*Review and recommendation*
[44]	*Hospitality industry*	*Canada*	*Review and recommendation*
[45]	*Hospitality industry*	*Global*	*Review and recommendation*
[46]	*Hospitality industry*	*Global*	*Review and recommendation*
[47]	*Tourism industry*	*Russia*	*Review and recommendation*
[48]	*Hotel industry*	*Global*	*Review and recommendation*
[49]	*Hotel industry*	*China*	*Review and recommendation*
[50]	*Hospitality industry*	*United States*	*Review and recommendation*
[51]	*Tourism industry*	*Global*	*Review and recommendation*
[52]	*Hotel industry*	*Global*	*Review and recommendation*

The Source of publication among included papers is represented in Figure 3. The most popular publication sources are including Tourism Geographies, International Journal of Infection Diseases, and Journal of Tourism and Hospitality Education.

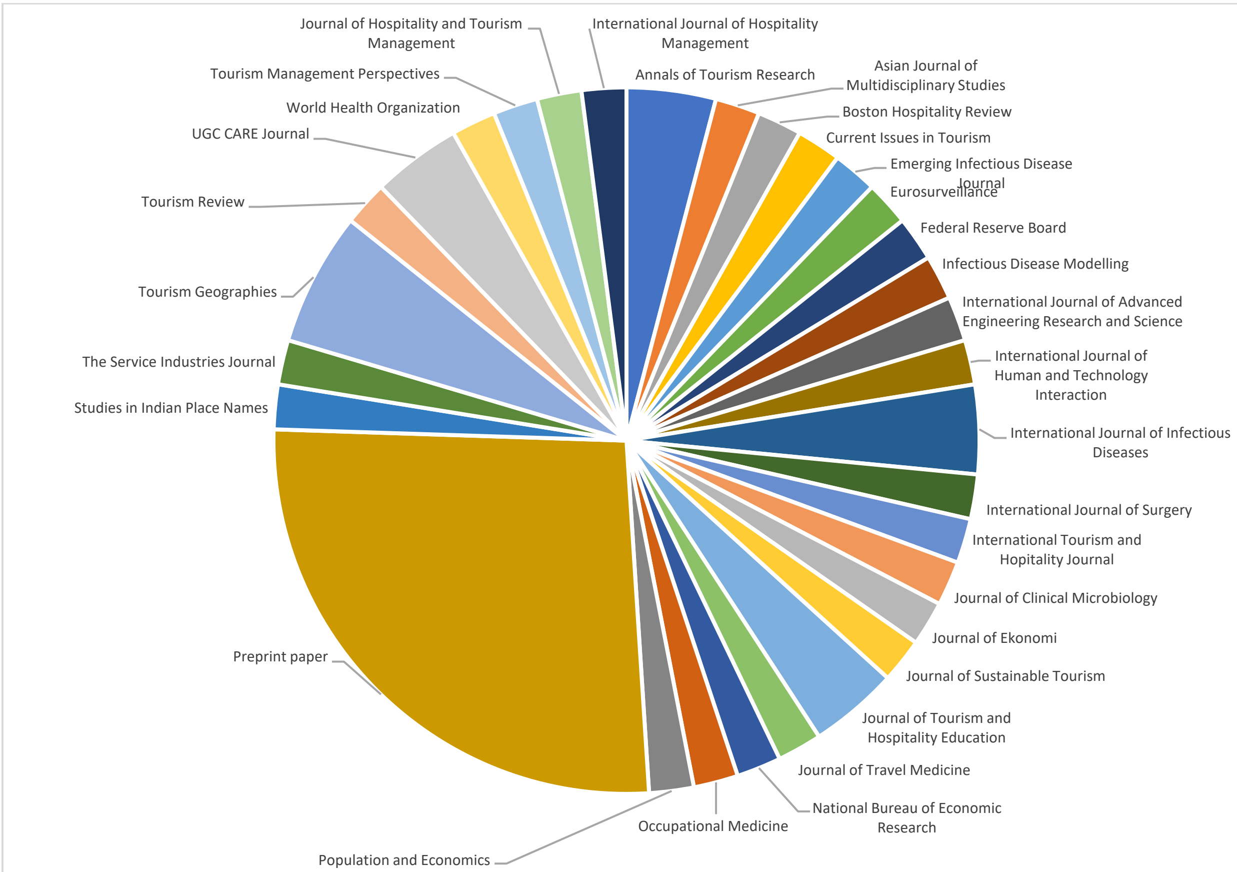

Figure 3. Publication source among included papers

The map of the co-occurrence of terms in the title and abstract of recorded papers is showed in Figure 4.

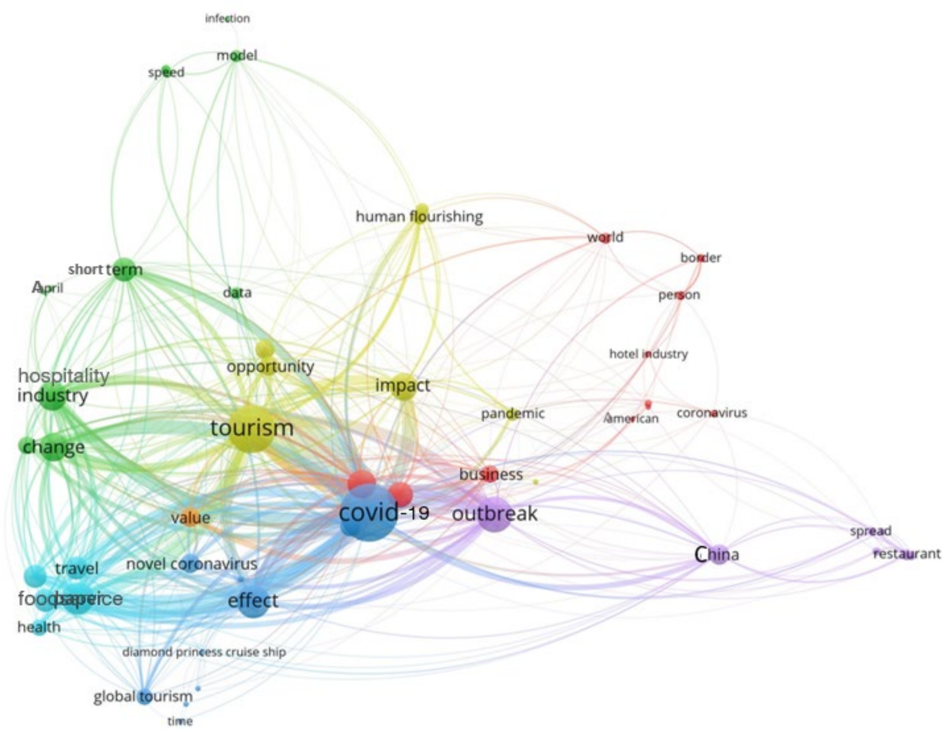

Figure 4. The map of co-occurrence of the terms in the title and abstract of recorded papers

In terms of the segment of the industry, the included papers mainly focused on the tourism industry, followed by the hospitality industry as it is represented in Figure 5. None of the included papers investigated the event industry.

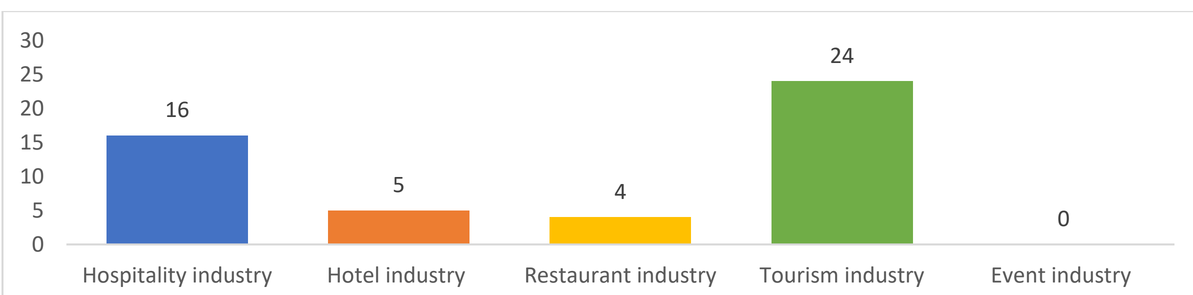

Figure 5. The segment of industry among included papers

Many included papers investigated the hospitality industry in the face of COVID-19 on the global scale as it is represented in Figure 6. Other papers focused on a specific country or location such as China, India, or the United States.

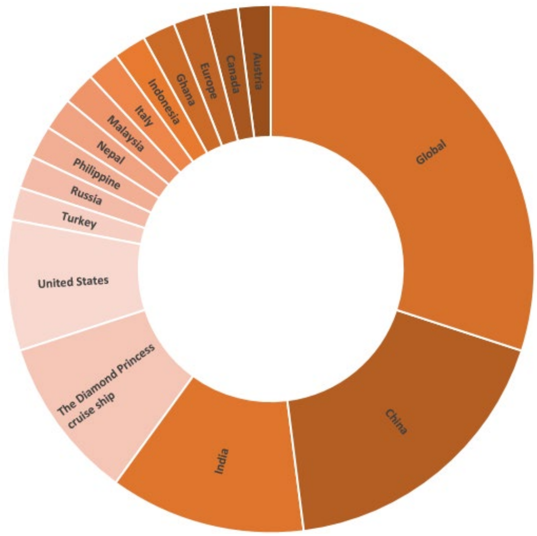

Figure 6. Geographic location among recorded papers

Included papers used different approaches to investigate the hospitality industry in the face of COVID-19. The most popular approach was using secondary data analysis to report the impacts of COVID-19 on the hospitality industry. Another popular approach was recommending different actions based on reviewing different documents as it is represented in Figure 7.

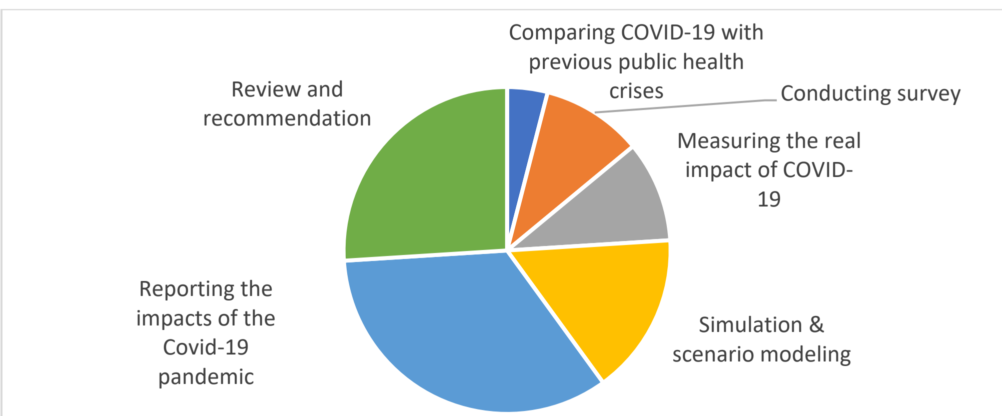

Figure 7. Investigation approach among included papers

2.3 The Impacts of the COVID-19 Pandemic

The reviewed papers focus on the variety of subjects related to the impact of COVID-19 on the hospitality industry. All papers have been classified into six groups as follows: developing simulation and scenario modeling, reporting impacts of the COVID-19 pandemic on the basis of secondary data analysis, comparing the COVID-19 pandemic with previous public health crises, measuring impacts of the COVID-19 pandemic in terms of percentages and dollars, recommending different actions on the basis of reviewing different documents, and conducting a survey. Since some of the reviewed papers belong to more than one group, these have been assigned to the dominant group.

2.3.1 Developing Simulation & Scenario Modeling

Eight included Papers in this review applied simulation & scenario modeling to estimate elements of tourism demand and the COVID-19 spreading pattern. The studies used different techniques including a dynamic stochastic general equilibrium (DSGE) model, supply and demand curve, agent-based model, epidemiological model, susceptible exposed infected and recovered (SEIR) model, and epidemic trajectory model.

Yang et al. [22] developed the DSGE model to investigate the impact of COVID-19. DSGE modeling is a technique in macroeconomics that depicts economic phenomena based on the general equilibrium framework [22]. To estimate the impact of COVID-19, Yang et al. [22] incorporated two indicators (health status, and health disaster) and three categories of decision-makers (the government, households, and producers) into the DSGE model concerning the tourism sector. Yang et al. [22] investigated the impacts of increasing health disaster risk and its persistence on the model parameters such as tourism demand. The findings are not surprising and point out that the longer pandemic will have a more devastating effect on the hospitality industry.

Bakar & Rosbi [20] utilized a supply and demand curve to analyze the economic impact of COVID-19 on the hospitality industry. To develop the supply and demand curve, the demand function was created by using factors of *price setting of selected goods, tastes and preferences of* *customers, customers' expectations, the average income of certain countries, and the number of buyers*. Meantime, the supply function is developed by using elements of *production techniques, resource price, price expectations, price of related goods, supply stocks, and numbers of sellers*. Then, the supply and demand curve was developed in the market equilibrium condition where the demand in the market is equal to the supply in the market. Finally, changes in market equilibrium as the result of the COVID-19 outbreak were investigated. The results indicate that the pandemic created some level of "panic" among people and consequently decreased overall demand in the tourism and hospitality industry [20]. The study urged governments to discover a vaccine as quickly as possible and identify policies to prevent the further decrease in demand for tourism and hospitality services during the post-pandemic period [20].

D'Orazio et al. [21] used an agent-based model to determine the virus spreading in tourist-oriented cities and consequently discover sustainable and resilient strategies [21]. The model represented the movement of simulated individuals and the contagion virus spreading approach (the epidemic rules based on previous studies) in a touristic urban area. The model calculated the probability that an infector i can infect a susceptible individual j based on a linear combination of the current incubation time of i, the exposure time, and the mask filter adopted by both i and j. The model evaluated the number of infectors within the touristic urban area over time and the number of visitors who return home being infected over time. After analyzing different scenarios, such as "social distancing-based measures" and "facial mask implementation," the results reveal that "social distancing-based measures" were related to significant economic losses [21]. This phenomenon appears to be an effective policy in locations with the highest infection rates [21]. However, "social distancing-based measures" lose their advantage in areas of low infection rates and a high degree of "facial mask implementation" [21].

Five studies investigated COVID-19 cases and spreading patterns on the Diamond Princess cruise ship. On February 1, 2020, a disembarked passenger from the ship tested positive for COVID-19 [53], after which the 3,711 passengers were quarantined [53]. By the end of the quarantine, more than 700 passengers were infected with COVID-19 [53]. Fang et al. [15] developed the flow of passengers on the Diamond Princess cruise ship, and then created the virus transmission rule between individuals to simulate the spread of the COVID-19 caused by the close contact during passengers' activities. Mizumoto et al. [16], [17] developed an epidemiological model based on discrete-time integral equations and daily incidence series. After estimating the model parameters, Mizumoto & Chowell [16] used a Monte Carlo Markov Chain technique to predict the number of the new COVID-19 cases. Rocklöv et al. [18] collected data on confirmed cases on the Diamond Princess cruise ship and used the SEIR model (compartmental technique estimating the number of susceptible (S), exposed (E), infected (I), and recovered (R) individuals) to calculate the basic reproduction number. The basic reproduction number is *the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection* [54]. Zhang et al. [19] collected data of daily incidence for COVID-19 on the Diamond Princess cruise ship, data of a serial interval distribution (*the time between successive cases in a chain of transmission* [55]), and applied "projections" package in R to calculate the basic reproduction number. The studies concluded that an immediate response by the cruise company in following recommended safety guidelines and early evacuation of all passengers have potential to prevent mass transmission of COVID-19 [15]–[19].

2.3.2 Reporting the Impacts of the COVID-19 Pandemic

Seventeen included Papers in this review applied secondary data analysis to report the impacts of COVID-19 pandemic on the hospitality industry. Because of the ongoing pandemic and publication time of included papers, secondary data sources have been invaluable for most studies in this review. The studies reported impacts of the pandemic on different elements of the hospitality industry including job loss, revenue losses, access to loans, market demand, emerging new markets, hostile behaviors towards foreigners, undocumented workers, and hotel cleaners.

Nicola et al. [33] summarized the impact of the pandemic on the global economy through the review of news distributed by mass-media, government reports, and published papers. To better understand impacts of the pandemic, the study divided the world economy into three sectors of primary (including agriculture, and petroleum & oil), secondary (including manufacturing industry), and tertiary (including education, finance industry, healthcare, hospitality tourism and aviation, real estate, sports industry, information technology, and food sector). In hospitality, tourism, and aviation, Nicola, et al. [33] reported job loss, revenue losses, and decreasing market demand. Ozili and Arun [35] provided a list of COVID-19 statistics, including confirmed cases, confirmed deaths, and recovered cases in different countries and continents, and discussed the global impact of COVID-19 on the travel and restaurant industries. The study reviewed different policy measures implemented by different countries around the world to deal with COVID-19. These policies were categorized by Ozili and Arun [35] into four groups of (1) human control measures; (2) public health measures; (3) fiscal measures; and (4) monetary measures. In the human control policies measures, different actions including foreign travel restrictions, internal travel restrictions, state of emergency declarations, limiting mass gathering, closing down of schools, and restricting shops and restaurants, have also been identified [35].

Several studies reported the effect of COVID-19 on specific critical domains of the hospitality industry, such as undocumented workers and hotel cleaners. Williams and Kayaoglu (2020) argued that the most vulnerable workers in the industry need governmental financial support, but are unable to receive assistance, most likely because they are undocumented immigrants [39]. Furthermore, Rosemberg [34] highlighted the issues of job insecurity, risk of exposure to COVID-19, lack of health insurance, added pressure due to increased workload, and extra time required for ensuring complete disinfection during the pandemic [34].

Other studies focused on the impacts of the pandemic on specific countries, including China, Malaysia, Nepal, and India. Several articles were reviewed on the consequences of COVID-19 on tourism in China and its hospitality industry, indicated that the impacts will last for an extended period [27]. Wen et al. [38] reviewed literature and news on Chinese tourist behavior, tourism marketing, and tourism management; they concluded the growing popularity of luxury trips, free and independent travel, and medical and wellness tourism in the post-COVID-19 period [38]. They indicated that new forms of tourism would be more popular in post-COVID-19 including (1) slow tourism which emphasizes on local destinations and longer lengths of stay; and (2) SMART tourism which uses data analytics to improve tourists' experiences [38]. Another study used automated content analysis to investigate newspaper articles and identified nine key themes among 499 newspaper articles, including *"COVID-19's impact on tourism, public sentiment, the role of the hospitality industry, control of tourism activities and cultural venues, tourism disputes and solutions, national command and local response, government assistance, corporate self-improvement strategies, and post-crisis tourism product"* [26].

2.3.3 Comparing COVID-19 with Previous Public Health Crises

Two included Papers in this review compared the COVID-19 pandemic with previous public health crises. In the first study, lessons learned from previous crises, and pandemics are discussed, including malaria, yellow fever, Ebola, Zika virus, Middle East respiratory syndrome (MERS-CoV), avian influenza (H5N1), Creutzfeldt-Jakob disease (Mad Cow disease), swine flu (H1N1), and severe acute respiratory syndrome (SARS) [4]. This paper concluded that the impacts of COVID-19 on the global economy and China's tourism and hospitality industry, in particular, are likely to differ from previous pandemics, from which the tourism and hospitality industry recovered relatively quickly [4].

Gössling et al. [3] reviewed the impact of previous crises on global tourism including the Middle East Respiratory Syndrome (MERS) outbreak (2015), the global economic crisis (2008-2009), the SARS outbreak (2003), and September 11 terrorist attacks (2001) [3]. The authors indicated that previous crises did not have long-term impacts on global tourism. The authors also warned about increasing pandemic threats for several reasons, including the fast-growing world population, rapidly developing global public transportation systems, and increasing consumption of processed/low-nutrition foods [3]. Gössling et al. [3] also discussed the impact of COVID-19 on different sectors of the hospitality industry. The authors distinguished the impact of COVID-19 in view of two different aspects of (1) observed impacts (e.g., declines in hotel occupancy rates, liquidity problems in the restaurant industry); and (2) projected impacts (e.g., revenue forecasts in the accommodations sector, estimation of revenues) [3].

The still-evolving understanding of the behavior of the coronavirus makes it difficult to predict the recovery of the industry in the near future. However, suggestions have already been made for post-COVID-19 management of the tourism and hospitality industry. These include: (1)

focusing primarily on domestic tourism; (2) ending mass tourism and pilgrimage tourism; (3) focusing more on conference tourism, virtual reality tourism, and medical tourism; and (4) building a more sustainable tourism and hospitality industry rather than a return to "business as usual" [3], [4].

2.3.4 Measuring the Impact of COVID-19 in Terms of Percentages and Dollars

Five included Papers in this review used different methods to measure the impacts of the pandemic on the hospitality industry in terms of percentages and dollars. The studies used different methods including seasonal autoregressive integrated moving average model, scenario analysis, and trend analysis.

The economic impact of COVID-19 on the tourism and hospitality industry has been examined in terms of lost earnings or jobs. Centeno and Marquez [11] developed seasonal autoregressive integrated moving average models for the Philippines' tourism and hospitality industry, forecasting the total earnings loss of around 170.5 billion PHP (Philippine Peso)—equivalent to $3.37 billion—from COVID-19 just until the end of July 2020. To ease the pandemic's effects on the hospitality industry, the authors propose dividing the country into two regions according to the level of infection risk (high-risk and low-risk of COVID-19) to allow domestic travel into the low-risk regions or areas [11].

Günay et al. [12] applied a scenario analysis technique to calculate the impact of COVID-19 on Turkey's tourism and hospitality industry. Their model predicts the total loss of revenues in the best and the worst scenarios as $1.5 billion and $15.2 billion, respectively for 2020 [12]. The worst-case scenario involves the closing of borders for four months without any economic recovery [12]. The authors indicated that under the worst-case scenario, this would be one of the worst tourism crises in Turkey, exceeding the losses from public health crises due to Swine flu, Avian Flu, and SARS [12].

Mehta [13] estimated the effect of COVID-19 on India's economy at an earnings loss of about $28 billion in 2020, along with 70% job losses for tourism and hospitality workers, and mass bankruptcies [13]. Trend analysis was also used to examine the impact of COVID-19 on the global tourism and hospitality industry and global GDP [14]. According to Priyadarshini [14], the real global GDP growth will drop from 2.9% in 2019 to 2.4% by the end of 2020, while global revenues for the tourism and hospitality industry will drop by 17% compared to 2019. The study also predicts that North America, Europe, and Asia will experience the most massive losses in terms of global revenues. The tourism and hospitality revenues will fall in the U.S., Germany, Italy, and China by 10%, 10%, 24%, and 40%, respectively [14].

Cajner et al. analyzed the impact of the COVID-19 pandemic on the U.S. labor market. The study calculated that about 13 million paid jobs were lost just between March 14 to 28, 2020. To better understand this number's significance, the authors pointed out that only nine million private payroll employment jobs were lost during the Great Recession of the 1930s (less than 70% of the pandemic job loss) [10]. The study also highlighted that in the current crisis, the leisure and hospitality industry was the hardest hit and most affected industrial sector [10].

2.3.5 Review and Recommendation

Thirteen included Papers reviewed different documents and recommended various actions for the resumption of activities during and after the pandemic. The consequences of COVID-19 on the hospitality industry, such as empty hotels and loss of jobs, are discussed in one paper that offers a positive outlook that the industry will receive a significant flow of guests upon the easing of travel bans and restrictions [47]. The author stressed the importance of support for the hospitality industry during the pandemic, and the need for proper guidance to assure the successful reopening during the post-pandemic period. Taking a different stance, another study suggests that the hospitality industry may not do well after the lifting of travel bans and mobility restrictions [42]. The study refers to a small survey that found more than half of the participants would not order food even after the pandemic ends. The author also recommends a series of different actions for restaurants to attract customers in the post-COVID-19 period, such as including island-sitting arrangements to assure maximum physical distances between people, live cooking counters to allow customers to watch their food being prepared to instill confidence in its safety and having appropriate hygiene and cleaning procedures throughout [42].

Bagnera et al. [52] investigated the impact of COVID-19 on hotel operations and recommended a series of actions for hotel owners and managers, including using fewer rooms (reducing hotel capacity); emphasizing take-out or delivery options to reduce public dining, implementing intensified cleaning/sanitizing protocols; committing to the use of personal protective equipment (PPE) for workers and increasing attention to personal hygiene; communicating new COVID-19 policies to guests and employees; implementing physical distancing practices in public areas; and implementing protocols for guests exposed to or infected by COVID-19 [52]. It should be noted that the World Health Organization (WHO) produced a guide titled "Operational Considerations for COVID-19 Management in the Accommodations Sector" to provide practical assistance to the hospitality sector in particular [48]. The report is divided into sections for the management team, reception and concierge, technical and maintenance services, restaurants and dining rooms and bars, recreational areas for children, and cleaning and housekeeping with a list of responsibilities to help manage the threat of COVID-19 [48]. Furthermore, Jain [43] discussed different hotel industry strategies to bring back customers, including the use of disposable utensils in rooms, emphasizing staff health and hygiene, and using UV light to disinfect [43].

Specific steps for an exit strategy and the reopening of activities in different business sectors are presented by Peterson et al. [46]. Primary steps include implementing widespread COVID-19 testing, having enough supply of PPE, lifting social distancing and mobility restrictions, using electronic surveillance, and implementing strategies to decrease workplace transmission [46]. Emphasis was placed on the daily screening of hospitality sector staff for COVID-19 by using real-time reverse transcription polymerase chain reaction or serology tests [46]. In this aspect, another study used primary and secondary data and applied the descriptive analysis method to explore revitalization strategies for small and medium-sized businesses, especially in the tourism industry, after COVID-19 in Yogyakarta [40]. The study recommended several policies such as implementing credit policies by banks with simpler processes and lower interest [40].

Several papers discussed the theme of redesigning and transforming the tourism and hospitality industry. The proposed ideas include increasing resilience and security of the tourism and hospitality workforce in post-COVID-19 by cross-training and teaching different skills to workers [45]; exploiting the unique opportunity presented by COVID-19 to transform and refocus the tourism and hospitality industry towards local attractions rather than global destinations, and redesigning spaces to assure a 6-foot distance between tourists [41], [44], [51].

Hao et al. [49] developed a COVID-19 management framework as a result of reviewing the overall impacts of the COVID-19 pandemic on China's hotel industry. The framework contains three main elements of an anti-pandemic process, principles, and anti-pandemic strategies. The anti-pandemic process adopted the six phases of disaster management including the pre-event phase (taking prerequisite actions), the prodromal phase (observing the warning signs), the emergency phase (taking urgent actions), the intermediate phase (bringing back key community services), the recovery phase (taking self-healing measures), the resolution phase (restoring the normal routine). Hao et al. [49] recommended four principles for the different phases of disaster management including disaster assessment, ensuring the safety of employees, customer & property, self-saving, and activating & revitalizing business. Finally, the study discussed the main anti-pandemic strategies in the categories of leadership & communication, human resource, service provision, corporate social responsibility, finance, and standard operating procedure.

Sönmez et al. [50] reviewed the impacts of the COVID-19 pandemic on immigrant hospitality workers' health and safety. The study indicated that while a significant rise in occupational stress has been observed in immigrant hospitality workers over the past 15–20 years, the COVID-19 pandemic can add more pressure on workers and potentially deteriorate their mental and physical health condition. Sönmez et al. [50] recommended different actions in aspects of public and corporate policy, workplace policy, and future research areas.

2.3.6 Conducting Survey

Five included papers conducted a survey to investigate different elements of the hospitality industry including social costs, the theory of resilience, preference of customers, expected chance of survival, and travel behavior.

Qiu et al. [7] developed the contingent valuation method to estimate costs borne by residents of tourist destinations (social costs) as a result of the COVID-19 pandemic. Contingent valuation is *a survey-based economic technique for the valuation of non-market resources* [56]. The survey asks questions about how much money residents would be willing to pay to keep a specific resource. The study attempted to investigate how residents perceive the risk of tourism during the COVID-19 pandemic. By considering three Chinese urban destinations, Qiu et al. [7] quantified the social costs of tourism during the pandemic. The results indicate that most residents were willing to pay for risk reduction but amount this payment differ based on age and income of respondents.

Alonso et al. [9] focused on the theory of resilience and conducted a survey from a sample of 45 small hospitality businesses to answer questions of what are the main concerns of participants regarding the COVID-19 pandemic? How are small hospitality businesses handling this disruption? And what are the impacts of the pandemic on day-to-day activities? Alonso et al. [9] analyzed the qualitative responses to these questions through content analysis. The study highlighted nine theoretical dimensions about owners-managers' actions and alternatives when they confronted with the COVID-19 pandemic.

Kim & Lee [8] studied the impacts of the perceived threat of the COVID-19 pandemic on the preference of customers for private dining facilities. The study conducted a survey and concluded that the salience of the COVID-19 increases customers' preference for private dining facilities.

Bartik et al. [5] discussed the impact of COVID-19 on the U.S. small businesses, especially restaurants and tourism attractions, and highlighted their fragile nature in the face of a prolonged crisis. Such companies typically have low cash flow, and in the face of this pandemic, they will either have to declare bankruptcy, take out loans, or significantly cut expenses [5]. Their survey of restaurant owners found that the expected chance of survival during a crisis lasting one month is 72%, for a crisis that lasts four months is 30%, and for a crisis that lasts six months is 15%. The result also indicated that more than 70% of U.S. small businesses want to take up the CARES Act Paycheck Protection Program (PPP) loans, even though the majority of them believe it would be challenging to establish eligibility for receiving such loans [5].

A survey study by Nazneen et al. [6] investigated the pandemic's impact on travel behavior and reported that it had significant impacts on tourists' decisions to travel for the next 12 months. The authors also concluded that respondents are concerned about the safety and hygiene of hotels, recreational sites, and public transports [6]. It has also been postulated that hygiene and safety perception will play a significant role in travel decisions in post-COVID-19 times [6].

2.4 Conclusions

This paper provides a systematic review of the published research topics relevant to the understanding of the hospitality industry in the time of COVID-19. A total of 50 published and preprint papers that met the predefined inclusion criteria were included in the review. Two research questions have guided and answered as follows and as are represented in Figure 8:

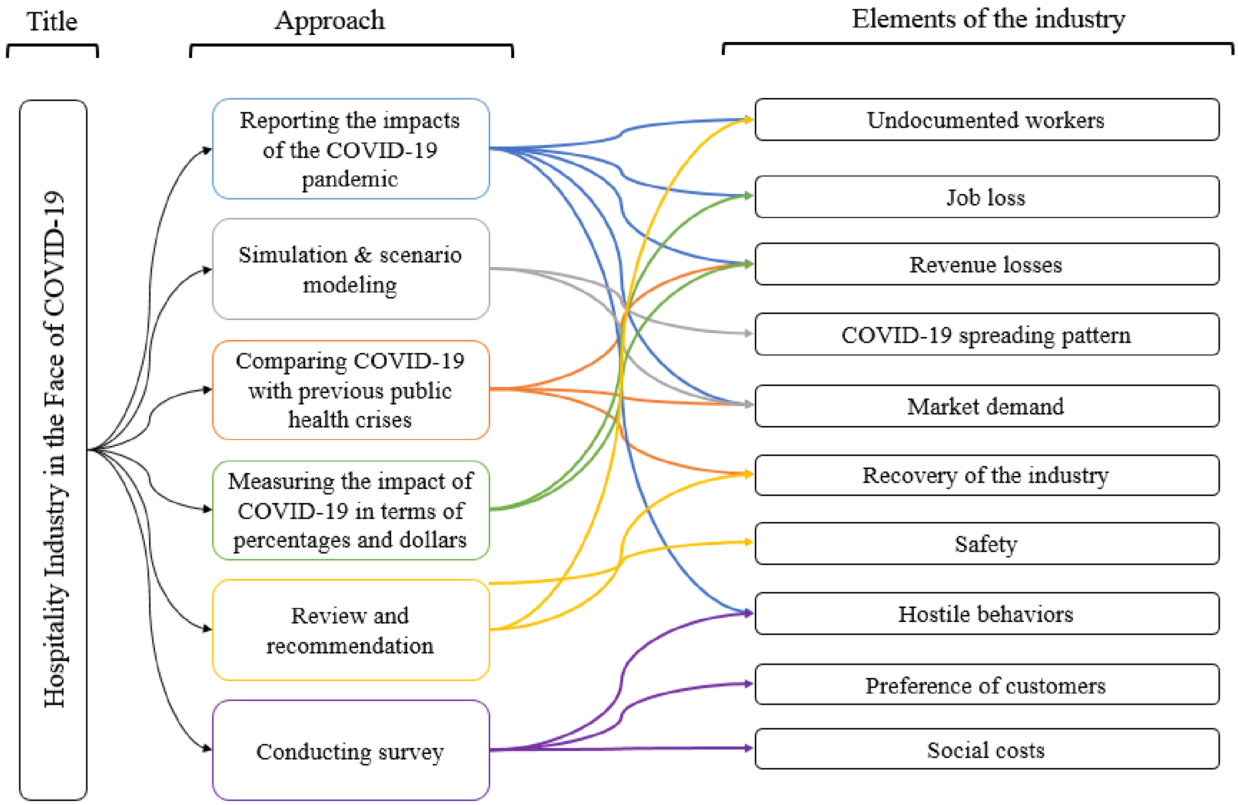

Figure 8. Investigation approach among included papers (answering research questions)

RQ1. How research on the hospitality industry in the face of COVID-19 is conducted? After identifying included papers, their methodologies have been investigated. The included papers used different approaches to study the hospitality industry in the face of COVID-19 including developing simulation and scenario modeling, reporting impacts of the COVID-19 pandemic, comparing the COVID-19 pandemic with the previous public health crises, measuring impacts of the COVID-19 pandemic, recommending different actions, and conducting a survey. For these approaches, included papers used different methodologies including secondary data analysis, dynamic stochastic general equilibrium (DSGE) model, supply and demand curve, agent-based model, epidemiological model, susceptible exposed infected and recovered (SEIR) model, epidemic trajectory model, seasonal autoregressive integrated moving average model, scenario analysis, trend analysis, and the contingent valuation method.

RQ2. What does current research reveal about the status of the hospitality industry at the time of COVID-19? After developing research questions and completing the search strategy, the status of the hospitality industry among included papers have been investigated. Even though included papers studied different elements of the hospitality industry, they mainly investigated the status of the hospitality industry in terms of undocumented workers, job loss, revenue losses, COVID-19 spreading pattern in the industry, market demand, recovery of the industry, safety, hostile behavior, and preferences of customers.

The papers identified in this systematic review provide useful knowledge about the scope of the recent research that focuses on the effects of the global COVD-19 pandemic on the tourism and hospitality industry. It should be noted that there are numerous other fertile research areas and methodologies that will need to be investigated and most likely implemented by multidisciplinary research teams. Due to the complex and dynamic nature of COVID-19 pandemic, the use of a wide array of complex systems science frameworks (e.g., syndemics) and methodologies (e.g., simulation modeling), can make an important contribution by examining how the synergistic effects of work and living conditions, as well as COVID-19 government and corporate responses, can influence the long-term health and safety of tourism and hospitality workers. Along these lines, the development and application of new technologies and equipment in the hospitality industry should protect guests and workers alike. Finally, other potential areas of research include the use of machine learning and artificial intelligence in the hospitality industry, best practices in building a more sustainable tourism and hospitality industry, and how impacts of travel and tourism activity on hosts, communities, and the environment can be minimized.

2.5 References

[1] A. Liberati *et al.*, “The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration,” *PLOS Medicine*, vol. 6, no. 7, p. e1000100, Jul. 2009, doi: 10.1371/journal.pmed.1000100.

[2] M. R. Davahli, W. Karwowski, and R. Taiar, “A System Dynamics Simulation Applied to Healthcare: A Systematic Review,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 16, Art. no. 16, Jan. 2020, doi: 10.3390/ijerph17165741.

[3] S. Gössling, D. Scott, and C. M. Hall, “Pandemics, tourism and global change: a rapid assessment of COVID-19,” *Journal of Sustainable Tourism*, pp. 1–20, 2020, doi: https://doi.org/10.1080/09669582.2020.1758708.

[4] A. N. Menegaki, “Hedging Feasibility Perspectives against the COVID-19 for the International Tourism Sector,” 2020, [Online]. Available: Retrieved from https://www.preprints.org/manuscript/202004.0536/v1, April 30, 2020

[5] A. W. Bartik, M. Bertrand, Z. B. Cullen, E. L. Glaeser, M. Luca, and C. T. Stanton, “How are small businesses adjusting to COVID-19? Early evidence from a survey,” National Bureau of Economic Research, 2020. [Online]. Available: Retrieved from https://www.nber.org/papers/w26989, April, 2020

[6] S. Nazneen, X. Hong, and N. Ud Din, “COVID-19 Crises and Tourist Travel Risk Perceptions,” *Available at SSRN 3592321*, 2020, [Online]. Available: Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3592321, May 4, 2020

[7] R. T. Qiu, J. Park, S. Li, and H. Song, “Social costs of tourism during the COVID-19 pandemic,” *Annals of Tourism Research*, p. 102994, 2020.

[8] J. Kim and J. C. Lee, "Effect of COVID-19 on Preference for Private Dining Facilities in Restaurants," *Journal of Hospitality and Tourism Management*, 2020.

[9] A. D. Alonso *et al.*, "COVID-19, aftermath, impacts, and hospitality firms: An international perspective," *International Journal of Hospitality Management*, p. 102654, 2020.

[10] T. Cajner, L. D. Crane, R. A. Decker, A. Hamins-Puertolas, and C. Kurz, "Tracking Labor Market Developments during the COVID-19 Pandemic: A Preliminary Assessment," Federal Reserve Board, 2020. [Online]. Available: Retrieved from https://www.federalreserve.gov/econres/feds/tracking-labor-market-developments-during-the-covid-19-pandemic.htm, April, 2020

[11] R. S. Centeno and J. P. Marquez, "How much did the Tourism Industry Lost? Estimating Earning Loss of Tourism in the Philippines," *arXiv preprint arXiv:2004.09952*, 2020, [Online]. Available: Retrieved from Cornell University's https://arXiv.org site, April 21, 2020

[12] F. GÜNAY, E. BAYRAKTAROĞLU, and K. ÖZKUL, "Assessing the Short-term Impacts of COVID-19 Pandemic on Foreign Visitor's Demand for Turkey: A Scenario Analysis," *Journal of Ekonomi*, vol. 2, no. 2, pp. 80–85, 2020.

[13] C. K. Mehta, "COVID-19: A nightmare for the Indian Economy," *Purakala with ISSN 0971-2143 is an UGC CARE Journal*, vol. 31, no. 20, pp. 333–347, 2020.

[14] I. Priyadarshini, "A Survey on some of the Global Effects of the COVID-19 Pandemic," 2020, [Online]. Available: Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C10&q=A+Survey+on+some+of+the+Global+Effects+of+the+COVID-19+Pandemic+Ishaani+Priyadarshini&btnG=, 2020

[15] Z. Fang *et al.*, “How many infections of covid-19 there will be in the" diamond princess"-predicted by a virus transmission model based on the simulation of crowd flow,” *arXiv preprint arXiv:2002.10616*, 2020, [Online]. Available: Retrieved from Cornell University’s https://arXiv.org site, April 25, 2020

[16] K. Mizumoto and G. Chowell, “Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020,” *Infectious Disease Modelling*, vol. 5, pp. 264–270, Jan. 2020, doi: 10.1016/j.idm.2020.02.003.

[17] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, “Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020,” *Eurosurveillance*, vol. 25, no. 10, p. 2000180, 2020, doi: https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180.

[18] J. Rocklöv, H. Sjödin, and A. Wilder-Smith, “COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures,” *Journal of travel medicine*, vol. 27, no. 3, 2020, doi: https://doi.org/10.1093/jtm/taaa030.

[19] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, and D. Chen, “Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis,” *International Journal of Infectious Diseases*, vol. 93, pp. 201–204, Apr. 2020, doi: 10.1016/j.ijid.2020.02.033.

[20] N. A. Bakar and S. Rosbi, “Effect of Coronavirus disease (COVID-19) to tourism industry,” *International Journal of Advanced Engineering Research and Science*, vol. 7, no. 4, 2020, [Online]. Available: Retrieved from http://journal-repository.com/index.php/ijaers/article/view/1858

[21] M. D'Orazio, G. Bernardini, and E. Quagliarini, "Sustainable and resilient strategies for touristic cities against COVID-19: an agent-based approach," *arXiv preprint arXiv:2005.12547*, 2020, [Online]. Available: Retrieved from Cornell University's https://arXiv.org site, May 26, 2020

[22] Y. Yang, H. Zhang, and X. Chen, "Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak," *Annals of Tourism Research*, p. 102913, 2020, doi: Advance online publication. https://doi.org/10.1016/j.annals.2020.102913.

[23] C. L. Correa-Martínez *et al.*, "A pandemic in times of global tourism: superspreading and exportation of COVID-19 cases from a ski area in Austria.," *Journal of clinical microbiology*, vol. 58, no. 6, pp. e00588-20, 2020, doi: https://doi.org/10.1128/JCM.00588-20.

[24] Jianyun Lu *et al.*, "COVID-19 Outbreak Associated with Air Conditioning in Restaurant, Guangzhou, China, 2020," *Emerging Infectious Disease journal*, vol. 26, no. 7, 2020, doi: 10.3201/eid2607.200764.

[25] Z. Ariawan and N. N. Ahmat, "Expected Tourist Attractions After Pandemic Covid-19," *International Journal of Human and Technology Interaction (IJHaTI)*, vol. 4, no. 1, pp. 107–112, 2020.

[26] H. Chen, X. Huang, and Z. Li, "A content analysis of Chinese news coverage on COVID-19 and tourism," *Current Issues in Tourism*, pp. 1–8, 2020, doi: https://doi.org/10.1080/13683500.2020.1763269.

[27] A. Hoque, F. A. Shikha, M. W. Hasanat, I. Arif, and A. B. A. Hamid, "The Effect of Coronavirus (COVID-19) in the Tourism Industry in China," *Asian Journal of Multidisciplinary Studies*, vol. 3, no. 1, pp. 52–58, 2020.

[28] M. R. Hu and A. D. Lee, "Airbnb, COVID-19 Risk and Lockdowns: Global Evidence," 2020, [Online]. Available: Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3589141, April 30, 2020

[29] M. D. Kalyankar and P. Patil, "Impact of Covid-19 Pandemic on the Tourism Sector," *Purakala with ISSN 0971-2143 is an UGC CARE Journal*, vol. 31, no. 8, pp. 611–617, 2020.

[30] W. Karim, A. Haque, Z. Anis, and M. A. Ulfy, "The Movement Control Order (MCO) for COVID-19 Crisis and its Impact on Tourism and Hospitality Sector in Malaysia," *International Tourism and Hopitality Yournal*, vol. 3, no. 2, pp. 1–07, 2020, doi: https://rpajournals.com/ithj-2020-02-09/.

[31] K. S. Kasare, "Effects of Coronavirus Disease (COVID-19) on Tourism Industry of India," *Studies in Indian Place Names*, vol. 40, no. 35, pp. 362–365, 2020.

[32] V. Kumar, "Indian Tourism Industry and COVID-19: Present Scenario," *Journal of Tourism and Hospitality Education*, vol. 10, pp. 179–185, 2020.

[33] M. Nicola *et al.*, "The Socio-Economic Implications of the Coronavirus and COVID-19 Pandemic: A Review," *International Journal of Surgery*, vol. 78, pp. 185–193, 2020, doi: https://doi.org/10.1016/j.ijsu.2020.04.018.

[34] M.-A. S. Rosemberg, "Health and safety considerations for hotel cleaners during Covid-19," *Occupational Medicine*, vol. 70, no. 3, pp. 214–215, 2020, doi: https://doi.org/10.1093/occmed/kqaa053.

[35] P. K. Ozili and T. Arun, "Spillover of COVID-19: impact on the Global Economy," *Available at SSRN 3562570*, 2020, [Online]. Available: Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3562570, March 30, 2020

[36] A. Thams, N. Zech, D. Rempel, and A. Ayia-Koi, "An initial assessment of economic impacts and operational challenges for the tourism & hospitality industry due to COVID-19," IUBH Discussion Papers-Tourismus & Hospitality, 2020. [Online]. Available: Retrieved from https://www.econstor.eu/handle/10419/216762

[37] N. Ulak, "A Preliminary Study of Novel Coronavirus Disease (COVID-19) Outbreak: A Pandemic Leading Crisis in Tourism Industry of Nepal," *Journal of Tourism and Hospitality Education*, vol. 10, pp. 108–131, 2020.

[38] J. Wen, M. Kozak, S. Yang, and F. Liu, "COVID-19: potential effects on Chinese citizens' lifestyle and travel," *Tourism Review*, 2020, Accessed: Jun. 05, 2020. [Online]. Available: Ahead-of-print, Retrieved from https://doi.org/10.1108/TR-03-2020-0110, May 8, 2020

[39] C. C. Williams and A. Kayaoglu, "COVID-19 and undeclared work: impacts and policy responses in Europe," *The Service Industries Journal*, pp. 1–18, 2020, doi: https://doi.org/10.1080/02642069.2020.1757073.

[40] S. Hadi, "Revitalization Strategy for Small and Medium Enterprises after Corona Virus Disease Pandemic (Covid-19) in Yogyakarta," *Journal of Xi'an University of Architecture & Technology*, 2020, [Online]. Available: Retrieved from https://www.researchgate.net/profile/Syamsul_Hadi11/publication/340870130_Revitalization_Strategy_for_Small_and_Medium_Enterprises_after_Corona_Virus_Disease_Pandemic_Covid-19_in_Yogyakarta/links/5ea1925492851c87d1ad7864/Revitalization-Strategy-

for-Small-and-Medium-Enterprises-after-Corona-Virus-Disease-Pandemic-Covid-19-in-Yogyakarta.pdf, April, 2020

[41] F. Higgins-Desbiolles, "Socialising tourism for social and ecological justice after COVID-19," *Tourism Geographies*, pp. 1–14, 2020, doi: https://doi.org/10.1080/14616688.2020.1757748.

[42] S. Jain, "Effect of COVID-19 on Restaurant Industry–How to Cope With Changing Demand," 2020, [Online]. Available: Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3577764, April 21, 2020

[43] S. Jain, "Would Hotel Industry Have to Redo the Rooms/Housekeeping Standards Post COVID? Instilling Greater Confidence to Bring Back the Customers," 2020, [Online]. Available: Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3587897, April 30, 2020

[44] D. Lapointe, "Reconnecting tourism after COVID-19: the paradox of alterity in tourism areas," *Tourism Geographies*, pp. 1–6, 2020, doi: https://doi.org/10.1080/14616688.2020.1762115.

[45] A. Martins, T. Riordan, and S. Dolnicar, "A post-COVID-19 model of tourism and hospitality workforce resilience," 2020, [Online]. Available: Retrieved from https://osf.io/preprints/socarxiv/4quga/, May 8, 2020

[46] E. Petersen *et al.*, "COVID-19–We urgently need to start developing an exit strategy," *International Journal of Infectious Diseases*, vol. 96, pp. 233–239, 2020, doi: https://doi.org/10.1016/j.ijid.2020.04.035.

[47] M. Y. Sheresheva, "Coronavirus and tourism," *Population and Economics*, vol. 4, no. 2, pp. 72–76, 2020, doi: https://doi. org/10.3897/popecon.4.e53574.

[48] World Health Organization, “Operational considerations for COVID-19 management in the accommodation sector: interim guidance, 31 March 2020,” World Health Organization, 2020. [Online]. Available: Retrieved from https://apps.who.int/iris/bitstream/handle/10665/331638/WHO-2019-nCoV-Hotels-2020.1-eng.pdf, March 31, 2020

[49] F. Hao, Q. Xiao, and K. Chon, “COVID-19 and China’s Hotel Industry: Impacts, a Disaster Management Framework, and Post-Pandemic Agenda,” *International Journal of Hospitality Management*, p. 102636, 2020.

[50] S. Sönmez, Y. Apostolopoulos, M. K. Lemke, and Y.-C. J. Hsieh, “Understanding the effects of COVID-19 on the health and safety of immigrant hospitality workers in the United States,” *Tourism Management Perspectives*, vol. 35, p. 100717, 2020.

[51] P. Brouder, “Reset redux: possible evolutionary pathways towards the transformation of tourism in a COVID-19 world,” *Tourism Geographies*, pp. 1–7, 2020, doi: https://doi.org/10.1080/14616688.2020.1760928.

[52] S. M. Bagnera, E. Stewart, and S. Edition, “Navigating hotel operations in times of COVID-19,” *Boston Hospitality Review*, 2020, [Online]. Available: Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C10&q=Navigating+Hotel+Operations+in+Times+of+Covid-19+By+Dr.+Suzanne+Markham+Bagnera+and+Emily+Stewar&btnG=

[53] S. Mallapaty, “What the cruise-ship outbreaks reveal about COVID-19,” *Nature*, vol. 580, no. 7801, Art. no. 7801, Mar. 2020, doi: 10.1038/d41586-020-00885-w.

[54] Basic reproduction number in Wikipedia, “Basic reproduction number,” *Wikipedia*. Aug. 24, 2020. Accessed: Aug. 25, 2020. [Online]. Available:

https://en.wikipedia.org/w/index.php?title=Basic_reproduction_number&oldid=974641299

[55] M. L. John, *A dictionary of epidemiology*. Oxford university press, 2001.

[56] Contingent valuation in Wikipedia, “Contingent valuation,” *Wikipedia*. 2020. Accessed: Aug. 26, 2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Contingent_valuation&oldid=916108680

CHAPTER 3: STOCHASTIC AND DETERMINISTIC NEURAL NETWROKS

This chapter contained material previously published in: M. R. Davahli, W. Karwowski, and K. Fiok, "Optimizing COVID-19 vaccine distribution across the United States using deterministic and stochastic recurrent neural networks," PLOS ONE, vol. 16, no. 7, p. e0253925, Jul. 2021, doi: 10.1371/journal.pone.0253925.

In this article, we develop real-time approaches for predicting the behavior of COVID-19 in all US states. We use data from the Centers for Disease and Prevention website and create two time-series datasets of the number of confirmed cases, and the effective reproduction numbers for all US states. The effective reproduction number, R_t, is defined as "the average number of secondary cases of disease caused by a single infected individual over her or his infectious period" [1].

To avoid training the models for all states, we use a self-organizing map (SOM) [2] to categorize all states into four groups according to their similarity in the reported effective reproduction numbers. In each group, we select the leading state (the state with earliest outbreaks). A deterministic Long Short Term Memory (LSTM) model [3], recurrent neural network (RNN) model, and stochastic Mixture Density Network (MDN) model [4] are then trained on data from each of the leading states.

In the deterministic LSTM model, the network output is the number of confirmed cases and the value of effective reproduction number in the next time-step. We use an LSTM RNN because (1) more confirmed cases can lead to more potential infection among populations in the future, and therefore, retaining all relevant historical information is important, and (2) this intelligent sequence analysis model has been reported by several studies to have high efficiency in time series forecasting problems [5].

In the stochastic MDN model, the network output is parameters of mixture distributions rather than a direct prediction value. The proposed MDN model is a combination of LSTM layers and a mixture of distributions. In this model, LSTM layers supply parameters for one or several distributions, which are then combined with weighting [4]. Finally, a sample of data can be extracted from the developed mixture distributions as an actual prediction [6].

We then compare the performance of developed models with a baseline linear regression model [7]. We aim to study whether using deterministic and stochastic sequence-learning models might have better predictive performance than linear regression. We also use an Augmented Dickey Fuller test [8] to assess the stationary and non-stationary status of the input dataset. We then remove seasonality and trend from the non-stationary datasets to investigate their effects on predictive performance.

This article is structured as follows. Section two discusses a published article on using artificial intelligence and machine learning to predict the behavior of the COVID-19 pandemic. Section three presents a brief mathematical explanation of R_t, seasonal-trend decomposition, SOMs, RNNs, and mixture density networks (MDNs). Section four discusses the development of sequence learning predictive models. Finally, section five explains the experimental setup, performance metrics, and results.

3.1 Reviewing Published Literature on RNN

On December 8, 2019, the government of China reported treatment of several new virus cases of a disease later named coronavirus disease 2019 (COVID-19) [9]. Since then, COVID-19 has spread across many countries and become a pandemic. COVID-19 is a highly transmissible respiratory disease with symptoms such as cough, fever, and breathing problems; it spreads through contact with infected individuals [10]. In January 2020, the US reported its first confirmed case of COVID-19; in mid-February 2020, the COVID-19 pandemic began to cause unprecedented social and economic consequences [9]. On December 14, 2020, the CDC reported 16,113,148 confirmed COVID-19 cases and 298,266 deaths in the US [11]. In this dire situation, the successful prior application of artificial intelligence and machine learning in critical problems inspired researchers to use these techniques against the COVID-19 pandemic. Artificial intelligence and machine learning have been used in various areas of predicting, contact tracing, screening, forecasting, and drug development for the COVID-19 pandemic [12].

Ribeiro et al. [13] have used cumulative confirmed Brazilian COVID-19 cases to train a support vector regression algorithm to forecast case numbers 6 days in advance. Chakraborty and Ghosh [14] have developed a hybrid method based on a Wavelet-based forecasting model and autoregressive integrated moving average model to forecast case numbers 10 days in advance for France, India, Canada, South Korea, and the UK. Chakraborty and Ghosh [14] have indicated that these forecast numbers of COVID-19 cases can act as an early-warning for policymakers and can be useful for the efficient allocation of health care resources. Kapoor et al. [15] have used mobility data and Graph Neural Networks to predict COVID-19 cases and have reported a 6% lower root mean squared logarithmic error than the best-performing baseline models.

Hartono [16] has indicated that developing an efficient predictive model is difficult because of the unknown characteristics of the virus causing COVID-19, as well as the political and geographical influences. Hartono [16] has used a topological autoencoder (TA), a topological neural network, to map the transmission dynamics of COVID-19 spread in several countries. TA produces a two-dimensional map in which countries with similar transmission dynamics are located close to each other. After selection of a target location for forecasting, TA has been used to identify a reference location with similar transmission dynamics that experienced earlier spread of the virus causing COVID-19. Finally, LSTM has been trained on data from the reference location to forecast the COVID-19 distribution in the target location.

Tomar and Gupta [17] have used LSTM and curve fitting to predict the number of COVID-19 positive cases and the number of recovered cases in India 30 days in advance. In that study, the data were collected from January 30, 2020 to April 4, 2020; 80% of the data were used for training, and 20% were used for testing. Li et al. [18, p. 19] have developed an integrated spatiotemporal model based on RNNs and epidemic differential equations to predict the number of COVID-19 cases in Italy 7 days in advance.

Arora et al. [5] have used RNN based LSTM variants including Deep LSTM, Bidirectional LSTM, and Convolutional LSTM to predict the number of COVID-19 cases in India 1 day and 1 week in advance. In that study, the states of India are categorized into different areas according to the daily growth rate and the number of confirmed COVID-19 cases. The dataset contains time-series data of confirmed COVID-19 cases from March 14, 2020 to May 14, 2020 for each state in India [5]. Arora et al. [5] have conducted an experiment on open source libraries and have used the Adam optimizer to optimize the mean squared error loss. The authors used the mean absolute percentage error (MAPE) to compare the performance of several predictive methods and found an average MAPE of 3.22% for bi-directional LSTM, 4.81% for Stacked LSTM, and 5.05% for conv-LSTM.

Shahid et al. [19] have used support vector regression, autoregressive integrated moving average, LSTM, and Bidirectional LSTM for predicting confirmed COVID-19 cases, deaths, and recoveries in Israel, Russia, Brazil, Spain, the UK, Germany, Italy, China, India, and the US. The study used the mean absolute error, root mean square error, and r^2_score indices to measure the performance of the models. The methods were found to rank as follows from best performance to worst performance: Bidirectional LSTM, LSTM, support vector regression, and autoregressive integrated moving average.

Chimmula and Zhang [8] have collected data on the numbers of confirmed COVID-19 cases, of fatalities, and recovered patients in a time series format from the Canadian Health Authority and Johns Hopkins University. The Augmented Dickey Fuller test was used to identify the effects of trends on the dataset and to report the stationary and non-stationary nature of the data [8]. The study has also developed an LSTM model to forecast the pandemic outbreak in Canada.

3.2 Mathematical Models

In this section, the mathematical formulae of effective reproduction numbers, SOMs, RNNs, and MDNs are explained.

3.2.1 Effective Reproduction Number

The effective reproduction number, R_t, is defined as “the expected number of new infections caused by an infectious individual in a population where some individuals may no longer be susceptible” [20]. One of the main reasons for calculating R_t is to determine how interventions and control efforts in population immunity, policy, and other elements affect transmission in specific time-steps [21]. Furthermore, R_t can be used to study real-time changes in COVID-19 transmission [20]. To bring the pandemic under control, R_t must be decreased to less than 1 and as close to 0 as possible [1]. Therefore, predicting R_t, which is situation- and time-specific, can aid in understanding the pathogen transmissibility during the COVID-19 pandemic in the future. Several methods have been developed to estimate R_t but we use the method of Cori et al. [1], in which the effective reproduction number is as follows:

$$R_t = \frac{I_t}{\sum_{s=1}^{t} I_{t-s} w_s} \tag{1}$$

where I_t is the number of incidents of infections on day t, and w_s is the generation interval, which is defined as "the time between the infection time of an infected person and the infection time of his or her infector" [22]. In this equation, the generation interval is the only parametric assumption adopted from Nishiura et al. [23]. That study obtained 28 infector-infectee pairs and used the log-normal distribution and the discretized gamma distributions to generate the results. Nishiura et al. [23] have reported the standard deviation and mean of the serial interval at 2.9 days (95% credible interval (CrI): 1.9, 4.9) and 4.7 days (95% CrI: 3.7, 6.0). For estimating R_t, the Excel file of EpiEstim package was borrowed from Cori et al. [1](Please refer to https://github.com/RezaDavahli for input data; 10 February 2021) [24].

3.2.2 Seasonal-Trend Decomposition

Normally, time series data can be decomposed into the trend, seasonality, and residual, as represented in the following equation:

$$q = \tau_t + s_t + r_t \tag{2}$$

Where t = 1, 2, · · ·, N; x_t is an original signal at time t; τ_t is the trend; s_t is the seasonality, which is the patterns that repeat with a period of time; and r_t is the residual. Several decomposition algorithms have been proposed for periodic and non-periodic datasets [25]. In this article, we use Seasonal-Trend Decomposition in six steps, which have been fully discussed by Qin et al. [26].

Before removing the seasonality and trend, we apply the Dickey Fuller test to determine whether the datasets are stationary or non-stationary. For the stationary dataset, seasonality and trend are not removed.

3.2.3 Self-Organizing Map

Teuvo Kohonen developed the SOM as a new form of neural network architecture and learning algorithm in the 1980s [2]. SOM uses an unsupervised learning process to analyze and represent the basic structures of a dataset as a map [27]. Therefore, SOM is commonly used to convert high-dimensional datasets into one- or two-dimensional maps [28]. Suppose that the input variables are $X = (x_1,x_2,\cdots x_p)'$; the weight vector assigned to the node l is $u_l = (u_{l1},u_{l2},\cdots u_{lp})'$; u_{lj} is the weight associated with node l of input variable x_j; and p is the number of input variables [29].

The learning concept of SOM involves detecting and moving the winning node closer to each training case. For this purpose, the Euclidean distance d_i between the weight vector and the input variables is calculated for each item i in the training case. Subsequently, the weights of the winning node with the smallest d_i are updated by a learning rule. In each step, the index q of the winning node is:

$$q = argmin\|u_l^s - x_i\| \tag{3}$$

Where u_l^s is the weight for the l_{th} node on the sth step, α^s is the learning rate for the sth step, and x_i is the input variable for the i_{th} training case. For the winner node, the update rule is:

$$u_q^{s+1} = u_q^s(1-\alpha^s) + x_i\alpha^s = u_q^s + \alpha^s\left(x_i - u_q^s\right) \tag{4}$$

Where u_l^{s+1} is set to u_l^s for all non-winning nodes.

3.2.4 Recurrent Neural Networks

Deep learning methods are effective for prediction because they automatically extract appropriate features from datasets [30]. RNN, a deep learning method, can store extensive historical information and use it to accurately predict the next steps in time-series problems [31]. However, its main disadvantage is long training time, because of vanishing gradient problems [17]. To overcome this problem, the LSTM structure, comprising a cell, an input gate, an output gate, and a forget gate, was developed to consider a long-term dependency [3]. In this structure, the cell stores values over arbitrary time intervals, and the gates adjust the flow of information in the recurrent hidden layer [17] [32].

The states of an input gate, an output gate, and a forget gate can be demonstrated mathematically by five equations:

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_t\right) \tag{5}$$

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\tilde{C}_t = tanh(W_c . [h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{8}$$

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t * tanh(C_t) \tag{10}$$

In these equations, σ is the logistic sigmoid activation function; C_t is the cell state; W indicates the weight matrices; and i, o, and f indicate the input gate, output gate, and forget gate, respectively [32]. In this structure, the input gate specifies the flow of information and protects the cell from irrelevant information, the forget gate deletes irrelevant information, and the output gate regulates the flow of information passing through the rest of the network [5].

3.2.5 Mixture Density Networks

MDNs are a combination of a neural network and a mixture of distributions. In MDNs, neural networks are used to model a mixture of components [33]. The main aspects of MDNs include the type of neural network, the number and size of the hidden layers, the dimension of the output, the number of input parameters, the type of distribution, and the number of distributions [33]. Unlike the LSTM deterministic model with fully determined outputs, MDNs estimate probability distributions of potential outcomes[34].

In the following equation, the mixture of the probability density function (PDF) p(x) is represented as a combination of the m PDFs with weights $\Omega = \{\omega_0, \ldots, \omega_{m-1}\}$, where the sum of weights is equal to 1:

$$p(x) = \sum_{j=0}^{m-1} \omega_j p_j(x) \tag{11}$$

Each p_j is a normal distribution defined by a variance σ_j and a mean μ_j, according to the following equation:

$$p(x) = \sum_{j=0}^{m-1} \frac{\omega_j}{\sqrt{2\pi\sigma_j^2}} exp(\frac{-1}{2\sigma_j^2}(x-\mu_j)^2) \tag{12}$$

The model can be fit to the following objective loss function:

$$f(x) = -\sum_{i=0}^{n-1} log \sum_{j=0}^{m-1} \omega_j p_j(x) \tag{13}$$

In this study, RNNs are used to output the parameters of a mixture model including the mixing coefficient of each Gaussian kernel (the probability of each kernel), and the mean and variance of each Gaussian kernel.

3.3 The COVID-19 Predictive Models

In this section, the deterministic and stochastic sequence-learning models are explained. These models are used to predict the number of confirmed COVID-19 cases and the effective reproduction numbers in all states in the US. We use data from the Centers for Disease and Prevention website, and have developed a dataset of the number of confirmed COVID-19 cases in all states of the US from January 22, 2020, to November 26, 2020, as indicated in Table 3.

Table 3. The confirmed case dataset at one time-step

Date	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delaware	Florida	…
3/29/2020	110	12	146	34	480	246	469	18	891	…
…										

Next, we use the EpiEstim package to compute effective reproduction numbers for all time-steps and all states, as represented in Table 4.

Table 4. The Rt dataset at one time-step

Date	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delaware	Florida	…
3/29/2020	2.06	1.89	2.11	1.28	1.77	1.92	2.39	1.91	2.26	…
…										

Both datasets contain 310 rows (time-step-days) and 50 columns (US states). To decrease the dimensionality of datasets, we use SOM to categorize all states into four categories. We apply the Minisom package [35] to a dataset containing the effective reproduction numbers from August 26, 2020 to November 26, 2020 for all US states. In the dataset, time-steps are considered features, and states are nodes. We have categorized all states into four groups according to the behavior of the effective reproduction numbers over time, as represented in Figure 8.

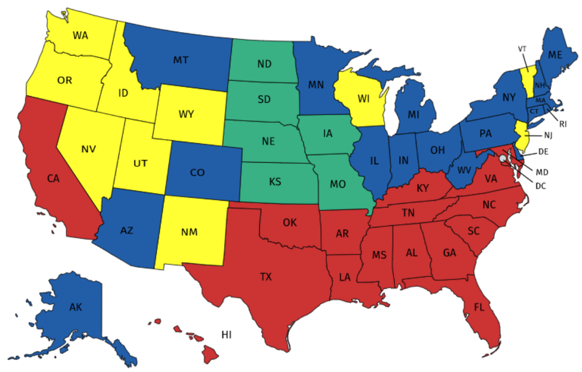

Figure 9. Categorization of all states according to the effective reproduction numbers over time (red: group one, blue: group two, green: group three, yellow: group four).

As shown in Fig 3, most neighboring states are interestingly clustered into the same group, thus indicating that the COVID-19 behavior is similar in close states. This conclusion appears logical, because there is more commuting and traveling between neighboring states.

We also use the R package Chorddig [36, p. 2] to visualize all relationships among states according to their similarities in effective reproduction number (Figure 9).

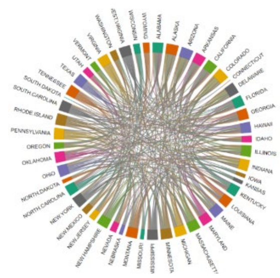

Figure 10. The relationships among states in terms of the similarity of effective reproduction numbers

After categorizing the states into four groups, we select the state with the earliest outbreaks as the leading state in each group. These leading states are used for training the models. Two sequence-learning models are considered: a deterministic LSTM model and a stochastic LSTM/MDN model. Figure 10 represents the structure of the stochastic LSTM/MDN model.

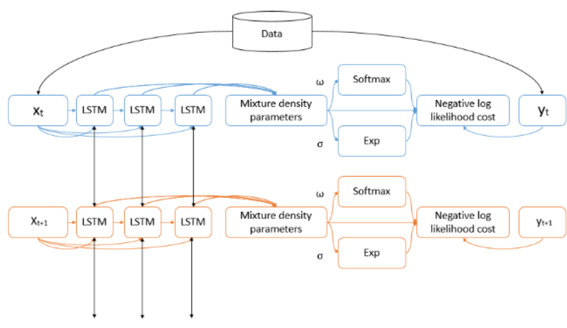

Figure 11. The LSTM-MDN learning model through time-steps.

In the stochastic LSTM/MDN model, the neurons corresponding to the means $\mu_k(x)$ are passed to the negative log likelihood cost, but neurons corresponding to the variances $\sigma_k(x)$ are passed through an exponential function before moving to the negative log likelihood cost. To satisfy the constraint of a sum of weights equal to 1 ($\Omega = \{\omega_0, . . ., \omega_{m-1}\}$), the neuron corresponding to weights passes through the softmax function. Softmax creates probabilities between 0 and 1 from real values that add up to 1:

$$Softmax(z)_j = \frac{e_j^z}{\sum_{k=1}^{n} e^{zk}} \tag{14}$$

As described earlier, the probability density of y_t can be calculated according to the following equation:

$$p(y_t|x) = \sum_{k=1}^{M} \omega_k(x) g_k(y_t|x) \tag{15}$$

Where $g_k(y_t|x)$ is represented in the following equation as the k_{th} multivariate Gaussian kernel.

$$g_k(y_t|x) = \frac{1}{(2\pi)^{N/2}} exp\left\{\frac{||y_t - \mu_k^2(x)||^2}{2\sigma_k(x)^2}\right\} \tag{16}$$

Where the vector $\mu_k(x)$ is the center of k_{th} kernel. Finally, the error function is represented as follows:

$$E_t = -ln\{\sum_{k=1}^{M} \omega_k(x) g_k(y_t|x)\} \tag{17}$$

Both deterministic and stochastic models were trained to provide predictions for time-step t + 1 after input of values up to time-step t. However, the output of the LSTM model is a value, whereas the output of the LSTM/MDN model is a mixture density parameters of a Gaussian mixture distribution. Therefore, for the stochastic model, a sample selected from this Gaussian mixture distribution is considered a prediction of the next time-step.

3.4 Experimental Study

In this section, the developed stochastic and deterministic models are evaluated on two datasets of confirmed COVID-19 cases and effective reproduction numbers (Please refer to https://github.com/RezaDavahli for models and input data; 10 February 2021). Then they are compared with a linear regression model to better understand their predictive ability. In the next experiment, after performing an Augmented Dickey Fuller test, we remove the seasonality and trend of the non-stationary dataset. We then investigate the performance of the developed models trained on the residuals dataset.

3.4.1 Experimental Setup

The performance of the developed deterministic and stochastic models is evaluated with the datasets of confirmed COVID-19 cases and effective reproduction numbers. The datasets contain values from January 22, 2020 through November 26, 2020 (Please refer to https://github.com/RezaDavahli for models and input data; 10 February 2021). In each dataset, 95% of the data are used for training (including 76% for training and 19% for validation), and 5% are used for testing. The testing set is considered from November 11, 2020 to November 26, 2020. The number of days for the testing set was borrowed from Arora et al. [5] and Hartono [16] aiming to provide comparability of our results. For developing the training dataset, 14 previous days are used in one batch to train the model and predict the value for the next day (1 day in advance). The Tensorflow [37] and Keras [38] libraries are used for developing the networks. The list of parameters in the two models is shown in Table 3.

Table 5. List of parameters in the two models.

Elements	**LSTM**	**LSTM/MDN**
Time step length	*Day*	*Day*
Normalization	*Yes*	*Yes*
Number of sequences	*14*	*14*
Number of hidden layers	*3*	*2*
Number of nodes in each hidden layer	*50*	*10*
Number of mixture Gaussian kernels	-	*1*

3.4.2 Performance Metrics

We use Mean Absolute Percentage Error (MAPE), which is the percentile error of the models, to test the performance of the developed predictive models [39]. As represented in the following equation, $y_{i,t}$ is the real value in state i at time-step t, whereas $\hat{y}_{i,t}$ is the predicted value.

$$MAPE_i = \frac{1}{T}\sum_{t=1}^{T}\frac{|y_{i,t} - \hat{y}_{i,t}|}{y_{i,t}} \tag{18}$$

We compare the developed stochastic and deterministic predictions with that of linear regression to better understand the performance of the models.

3.4.3 Performance Results

To fully understand the efficient model, we report the average MAPE for all leading states and for different combinations of models and datasets, as shown in Figures 11 and 12.

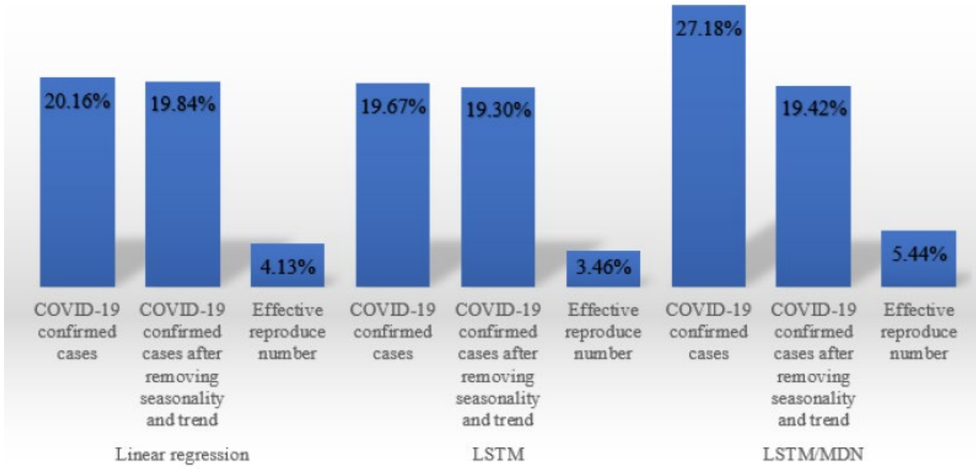

Figure 12. The performance of different combinations of models and datasets

Figure 13. The performance of different combinations of models and datasets

Several specific patterns are seen among the data. First, the predictive models trained on effective reproduction numbers show much better performance than models trained on confirmed cases. On average, there is a 16% difference between the predictions based on confirmed cases versus effective reproduction numbers. Second, unlike the confirmed cases dataset, the R_t dataset is stationary, and there is no need to remove the seasonality and trend. However, with the confirmed cases dataset, the greatest improvement in performance due to removal of seasonality and trend is seen in the stochastic LSTM/MDN model. Third, the deterministic LSTM model has the best performance for the two datasets. The LSTM model trained on the effective reproduction number has the best performance, with 3.46% MAPE among all fusions.

We also represent the performance of models from November 11, 2020, to November 26, 2020 in the leading state of California in group one in Figures 13 and 14.

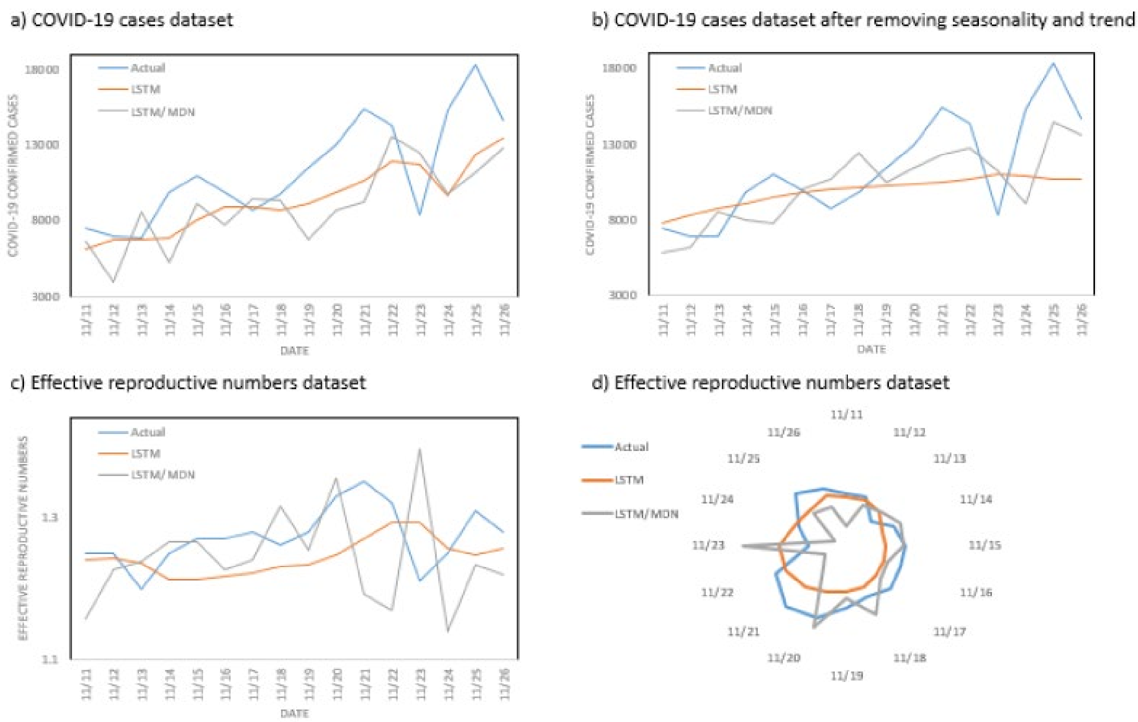

Figure 14. The performance of different combinations of models and datasets in the leading state of California in group one: (a) performance of deterministic and stochastic models trained on the COVID-19 cases dataset, (b) performance of deterministic and stochastic models trained on the dataset of COVID-19 cases after removal of seasonality and trend, (c) performance of deterministic and stochastic models trained on the effective reproduction numbers dataset, (d) performance of deterministic and stochastic models trained on the effective reproduction numbers dataset.

As shown in Figure 8, although deterministic LSTM has better performance, stochastic LSTM/MDN is more successful in following the trend of the actual data. However, stochastic LSTM/MDN is much more sensitive to large changes in the actual data.

We also show the performance of models on COVID-19 datasets when seasonality and trend are removed in comparison to the original datasets in the leading state of California (Figure 14).

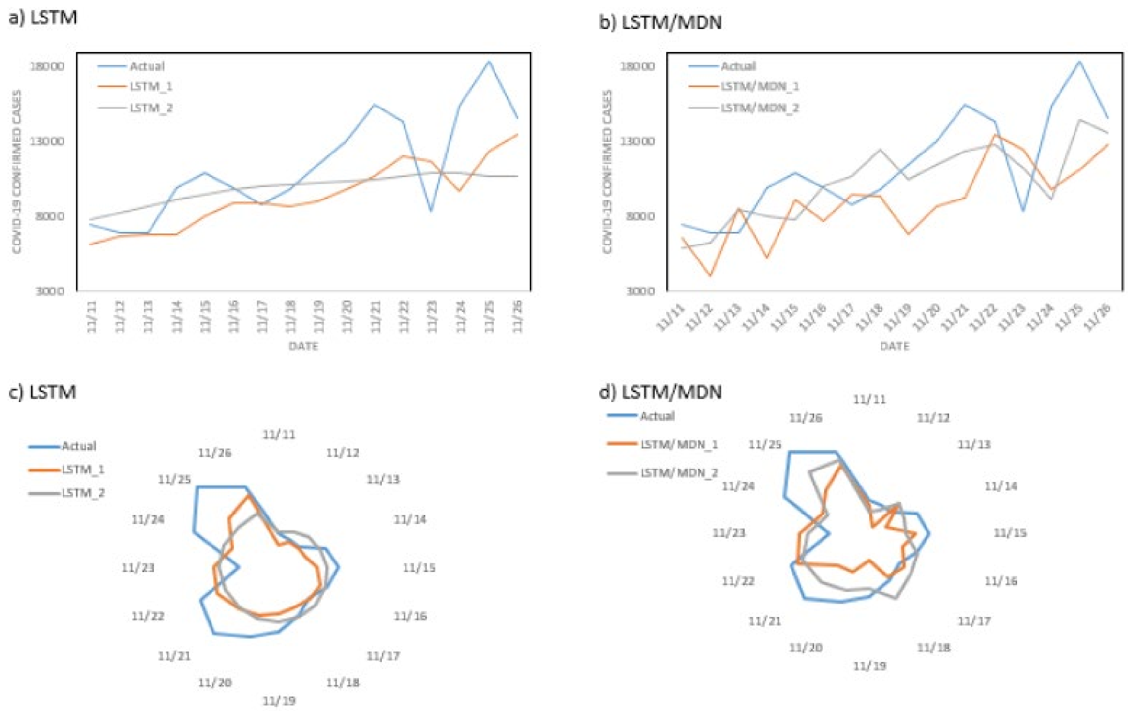

Figure 15. The performance of deterministic and stochastic models trained on the COVID-19 cases dataset with seasonality and trend removed, in comparison to the original dataset in the leading state of California in group one: (a, c) performance of a deterministic model trained on the COVID-19 cases dataset (LSTM_1: without removal of seasonality and trend; LSTM_2: with removal of seasonality and trend), (b, d) performance of the stochastic model trained on the COVID-19 cases dataset (LSTM/MDN_1: without removal of seasonality and trend; LSTM/MDN_2: with removal of seasonality and trend).

3.5 Limitations

In this study, we developed models to predict the behavior of COVID-19 within the leading states. Therefore, the main limitation is that we did not consider the effect of states on one another. Many states issued a stay-at-home order, asking residents to stay at home, which reduced mobility between states.

In our next study, we are going to investigate the impacts of mobility on the performance of the sequence learning models.

Although we indicated that the models trained on R_t have much better performance, there are some limitations associated with that. The main limitation is that R_t can be calculated from different methodologies, which do not give the same estimate. The final major limitation relates to using SOM for dividing US states into four groups. SOM uses an unsupervised learning process to analyze and represent the R_t dataset as a map. SOM decreased the dimensionality of the R_t dataset by clustering states based on similarities in their respective R_t numbers from August 26, 2020 to November 26, 2020. In the resulting map, most neighboring states were clustered together, but there were several exceptions. Because this is an unsupervised clustering technique, the reasoning behind the clusters and exceptions is not clear.

3.6 Conclusion

We developed stochastic and deterministic sequence learning models based on RNNs and MDNs to predict the behavior of COVID-19 in different US states. We trained the models on historical confirmed cases and R_t patterns. The developed models can predict geographic spreading of the active virus. The primary dataset contains 310 time-steps and 50 features (US states). To avoid training the models for all states, we used the unsupervised learning methods of SOM to categorize all states into four groups according to their similarity in COVID-19 behavior. After selecting one state from each group as the leading state (the state with the earliest outbreak), we trained the developed models. We found that the predictive models trained on R_t have much better performance than those trained on confirmed cases. In addition, the deterministic LSTM model has better performance than the stochastic LSTM/MDN and linear regression models. However, the stochastic model is more successful in predicting the trends in the actual dataset. Finally, LSTM trained on R_t has the best performance, with a MAPE value of 3.46%.

3.7 References

[1] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics," *Am J Epidemiol*, vol. 178, no. 9, pp. 1505–1512, Nov. 2013, doi: 10.1093/aje/kwt133.

[2] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] C. M. Bishop, "Mixture density networks," 1994.

[5] P. Arora, H. Kumar, and B. K. Panigrahi, "Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India," *Chaos, Solitons & Fractals*, vol. 139, p. 110017, 2020.

[6] J. Xu, R. Rahmatizadeh, L. Bölöni, and D. Turgut, "Real-time prediction of taxi demand using recurrent neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2572–2581, 2017.

[7] B. D. Craven and S. M. Islam, *Ordinary least-squares regression*. Sage Publications, 2011.

[8] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons & Fractals*, p. 109864, 2020.

[9] M. R. Davahli, W. Karwowski, S. Sonmez, and Y. Apostolopoulos, "The Hospitality Industry in the Face of the COVID-19 Pandemic: Current Topics and Research Methods," *International Journal of Environmental Research and Public Health*, vol. 17, no. 20, p. 7366, 2020.

[10] Y. Yang, H. Zhang, and X. Chen, "Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak," *Annals of Tourism Research*, p. 102913, 2020, doi: Advance online publication. https://doi.org/10.1016/j.annals.2020.102913.

[11] CDC, "COVID-19 Cases, Deaths, and Trends in the US | CDC COVID Data Tracker," *Centers for Disease Control and Prevention*, Mar. 28, 2020. https://covid.cdc.gov/covid-data-tracker (accessed Dec. 14, 2020).

[12] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review," *Chaos, Solitons & Fractals*, p. 110059, 2020.

[13] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos Santos Coelho, "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil," *Chaos, Solitons & Fractals*, vol. 135, p. 109853, 2020.

[14] T. Chakraborty and I. Ghosh, "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis," *Chaos, Solitons & Fractals*, vol. 135, p. 109850, 2020.

[15] A. Kapoor *et al.*, "Examining covid-19 forecasting using spatio-temporal graph neural networks," *arXiv preprint arXiv:2007.03113*, 2020, [Online]. Available: https://arxiv.org/abs/2007.03113

[16] P. Hartono, "Similarity maps and pairwise predictions for transmission dynamics of covid-19 with neural networks," *Informatics in medicine unlocked*, vol. 20, p. 100386, 2020.

[17] A. Tomar and N. Gupta, "Prediction for the spread of COVID-19 in India and effectiveness of preventive measures," *Science of The Total Environment*, vol. 728, p. 138762, 2020.

[18] Z. Li, Y. Zheng, J. Xin, and G. Zhou, "A recurrent neural network and differential equation based spatiotemporal infectious disease model with application to covid-19," *arXiv preprint arXiv:2007.10929*, 2020, [Online]. Available: https://arxiv.org/abs/2007.10929

[19] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons & Fractals*, vol. 140, p. 110212, 2020.

[20] K. M. Gostic *et al.*, "Practical considerations for measuring the effective reproductive number, Rt," *medRxiv*, 2020, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7325187.1/

[21] J. Sciré *et al.*, "Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft," *Swiss Medical Weekly*, vol. 150, no. 19–20, p. w20271, 2020.

[22] E. Kenah, M. Lipsitch, and J. M. Robins, "Generation interval contraction and epidemic data analysis," *Math Biosci*, vol. 213, no. 1, pp. 71–79, May 2008, doi: 10.1016/j.mbs.2008.02.007.

[23] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, "Serial interval of novel coronavirus (COVID-19) infections," *International journal of infectious diseases*, vol. 93, pp. 284–286, 2020.

[24] M. R. Davahli, "Input datasets, developed models, and live figures." 2020. [Online]. Available: https://github.com/RezaDavahli/COVID-19

[25] J. Gao, X. Song, Q. Wen, P. Wang, L. Sun, and H. Xu, "RobustTAD: Robust time series anomaly detection via decomposition and convolutional neural networks," *arXiv preprint arXiv:2002.09545*, 2020, [Online]. Available: https://arxiv.org/abs/2002.09545

[26] L. Qin, W. Li, and S. Li, “Effective passenger flow forecasting using STL and ESN based on two improvement strategies,” *Neurocomputing*, vol. 356, pp. 244–256, 2019.

[27] T. Kohonen, “Essentials of the self-organizing map,” *Neural networks*, vol. 37, pp. 52–65, 2013.

[28] G. Douzas and F. Bacao, “Self-Organizing Map Oversampling (SOMO) for imbalanced data set learning,” *Expert systems with Applications*, vol. 82, pp. 40–52, 2017.

[29] F. Huang, K. Yin, J. Huang, L. Gui, and P. Wang, “Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine,” *Engineering Geology*, vol. 223, pp. 11–22, 2017.

[30] K. Fiok, W. Karwowski, E. Gutierrez, and M. Reza-Davahli, “Comparing the Quality and Speed of Sentence Classification with Modern Language Models,” *Applied Sciences*, vol. 10, no. 10, Art. no. 10, Jan. 2020, doi: 10.3390/app10103386.

[31] J. T. Connor, R. D. Martin, and L. E. Atlas, “Recurrent neural networks and robust time series prediction,” *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 240–254, 1994.

[32] colah’s blog, “Understanding LSTM Networks,” 2015. http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed Dec. 05, 2020).

[33] C. N. Davis, T. D. Hollingsworth, Q. Caudron, and M. A. Irvine, “The use of mixture density networks in the emulation of complex epidemiological individual-based models,” *PLoS computational biology*, vol. 16, no. 3, p. e1006869, 2020.

[34] D. O. Borchers, “A Hitchhiker’s Guide to Mixture Density Networks,” *Medium*, Feb. 15, 2019. https://towardsdatascience.com/a-hitchhikers-guide-to-mixture-density-networks-76b435826cca (accessed Dec. 05, 2020).

[35] G. Vettigli, “MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map,” 2013.

[36] M. Flor, *chorddiag: Interactive Chord Diagrams. R package version 0.1. 2*. 2019.

[37] M. Abadi *et al.*, “Tensorflow: A system for large-scale machine learning,” in *12th symposium on operating systems design and implementation*, 2016, pp. 265–283.

[38] F. Chollet, “Keras: The python deep learning library,” *ascl*, p. ascl-1806, 2018.

[39] A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, “Mean absolute percentage error for regression models,” *Neurocomputing*, vol. 192, pp. 38–48, 2016.

CHAPTER 4: GRAPH NEURAL NETWORKS

This chapter contained material previously published in: M. R. Davahli, K. Fiok, W. Karwowski, A. M. Aljuaid, and R. Taiar, "Predicting the Dynamics of the COVID-19 Pandemic in the United States Using Graph Theory-Based Neural Networks," International Journal of Environmental Research and Public Health, vol. 18, no. 7, Art. no. 7, Jan. 2021, doi: 10.3390/ijerph18073834.

On December 8, 2020, Margaret Keenan received the first dose of vaccine for SARS-CoV-2 (COVID-19) outside a clinical trial [1]. After that, the COVID-19 vaccine is being distributed and used around the world. Among different countries, the US was one of the leading countries in administering the COVID-19 vaccine [2]. The US reported its first COVID-19 case on January 20, 2020; since then the pandemic has imposed unprecedented social and economic consequences [3]. Even though herd immunity is on the horizon because of using the vaccine, the basic question of when the pandemic might be over in the US must be answered. To find the accurate answer to this question, it is important to select (1) a reliable indicator of the pandemic's condition and status (2) an accurate prediction model.

One of the good indicators of pandemics is the effective reproduction number, R_t, which is defined as "the average number of secondary cases of disease caused by a single infected individual over her or his infectious period" [4]. This indicator, which is the situation- and time-specific, is used to study changes in pathogen transmissibility. These changes are as the results of implementing different policies, changes in population immunity, and/or other factors during the pandemic [5]. To bring the pandemic under control, it is important to decrease R_t to less than 1 and close to 0. Therefore, studying R_t over time can explicitly represent feedback on the use of vaccine intervention.

Calculating accurate R_t is challenging. However, two main methodologies are used to calculate R_t including (1) the case reproductive number and (2) the instantaneous reproductive number [4]. While the first method calculates transmission by a certain cohort of people, the second method calculates transmission at a certain point in time. It is reported that the instantaneous reproductive number is a more accurate method for estimating R_t and we used this method in this study [4].

Regarding the accurate prediction model, we select Graph neural networks (GNN) because (1) this model considers the impacts of neighbor states on the target state, (2) this method has high efficiency in time series forecasting [6]. GNN is a combination of (1) graphs that are a data structure with two main components of nodes and edges, (2) neural network architectures. We consider two types of GNN model including graph theory based model (GTNN) and neighborhood-based model (NGNN). Nodes represent states of the US in both graphs; but edges in GTNN indicate high correlation between time series COVID-19 confirmed cases data of states (functional connectivity); and edges in NGNN represents neighbor states. In both models, each node learns embedding information about its connected nodes. This embedding is used to solve different problems such as node features prediction [6].

In this article, an approach is developed to predict the end of the COVID-19 pandemic in all US states. For this objective, the number of COVID-19 confirmed cases are obtained from the Centers for Disease and Prevention website and R_t is calculated for all US states over time. Following that, GTNN and NGNN models are trained to forecast R_t for all states of the US simultaneously. Finally, efficiency of two models are compared with each other and with baseline deterministic recurrent neural networks (Long short-term memory) model.

This article is structured as follows. Section (2) briefly explains published articles on using GNN to predict time series pandemic-related data. Section (3) describes base models of graph theory, R_t and GNN. Section (4) represents different steps for developing the predictive models. Section (5) indicates the setup of the experiment, performance metrics, and outcomes. Finally, section (6) discusses limitations of this study.

4.1 Reviewing Published Articles

On December 8, 2019, the first COVID-19 cases were officially reported in China; In January 2020, the first COVID-19 confirmed case was identified in the US [3]. In mid-February and March 2020, COVID-19 became a pandemic affecting all states of the US and causing unprecedented consequences [3]. On December 8, 2020, exactly one year after officially reporting the virus in China, the first dose of vaccine was received outside a clinical trial [1]. Meanwhile in this day, On December 8, 2020, the CDC reported 285,351 COVID-19 related deaths and 15,208,638 confirmed cases in the US [7]. In this dire situation, accurate predicting of the end of the pandemic can have significant social and economic impacts on the US [8]. One of the most accurate prediction methods is GNN and different articles reported results of this method on COVID-19 time series data.

Zheng et al. [9, p. 19] proposed a hybrid Spatio-temporal model by combining susceptible-exposed-infectious-recovered (SEIR) and recurrent neural networks (RNN). The article represented features on the graph structure including (1) geographic neighbor effect (edge feature) and (2) local temporal infection trend (node feature). The study applied SEIR to node feature and

RNN to edge feature to achieve both efficiency and accuracy in training and predicting [9, p. 19]. The study used COVID-19 confirmed case data of the US states to predict case numbers 1-day and 7-day in advance. The article indicated that the hybrid Spatio-temporal model outperformed standard RNN, SEIR, and Autoregressive integrated moving average (ARIMA) models.

Panagopoulos et al. [10] used GNN to investigate the impact of human mobility on the geographical distribution of COVID-19 cases. In the developed GNN model, edges correspond to population movement between regions of a country, and nodes represented the country's regions. The study used this model to predict the number of COVID-19 confirmed cases 3-day, 7-day, and 14-days in advance. The study used data from four European countries and indicated that the developed model outperformed traditional LSTM, ARIMA, and PROPHET models.

Cao et al. [11] focused on multivariate time-series forecasting techniques that analyze a set of time-series as a unified entity. The study proposed Spectral Temporal Graph Neural Network (StemGNN) to forecast the number of COVID-19 confirmed cases. StemGNN modeled both temporal dependencies (by applying Discrete Fourier Transform) and inter-series correlations (by applying Graph Fourier Transform) together in the spectral domain [11]. The study used time-series of 25 countries from January 22, 2020 to May 10, 2020 to predict number of COVID-19 confirmed cases 7-day, 14-day, and 28-days in advance.

La Gatta et al. [12] used users' mobility data and proposed a model to determine parameters of an epidemiological model such as recovery rates and contact rates. For developing the model, the article combined LSTMs and Graph Convolutional Neural Networks with temporal and spatial features. The study used the model to forecast the COVID-19 dynamics in Italy from February 24 to May 5, 2020.

Kapoor et al. [13] developed a forecasting model by using GNN and mobility data to predict COVID-19 cases. In the proposed large-scale Spatio-temporal graph, spatial edges indicated the population movement based on inter-region connectivity, nodes indicated the region-level population movement, and temporal edges indicate node features through time [13]. The article applied the model to the US county-level COVID-19 dataset and compared the results with traditional LSTM and ARIMA models. Finally, Shah et al. [14] emphasized the importance of early contact-tracing in the COVID-19 pandemic and used GNN to locate patient zero (the source of an epidemic).

4.2 Basic Models

In this section, the basics of graph theory and functional connectivity, R_t, and GNN are represented.

4.2.1 Graph Theory

The graph theory and network analysis have been used to address problems in variety of fields such as electrical power infrastructures, transportation systems, big data environments, social networks, biological neural networks, and complex brain networks [15]. In this theory, a graph consists of a number of nodes that are linked by edges. Graph edges can be unweighted direct, weighted indirect, weighted direct, and unweighted indirect. Direct and indirect focus on the flow of information; unweighted and weighted emphasis on the strength of connections. The following eight steps explain the pipeline for developing of a functional network with graph theory.

1. Defining the nodes of the network: nodes can be changed depend on the objective of networks.

2. Preprocessing of time series data: the time series data should be preprocessed to remove noise and artifact. Different preprocessing methods can be used such as Empirical Mode Decomposition (EMD). EMD is more suitable to process nonstationary and nonlinear data such as COVID-19 time series data. EMD decomposes time series data into finite number of multiple oscillatory modes called intrinsic mode functions (IMFs) [16]. For given a time-series dataset $x(t)$, EMD can be described as follows [17].

A. Identifying all extrema of $x(t)$,

B. Interpolanting between all local maxima (minima) to create upper (lower) envelopes of $e_{\max}(t)$ ($e_{\min}(t)$),

C. Computing the average $m(t) = [e_{\max}(t) + e_{\min}(t)]/2$,

D. Extracting the detail $h(t) = x(t) - m(t)$,

E. Iterating on the residual $r(t) = x(t) - c(t)$.

As it is summarize in Eq. (19), by using EMD, $x(t)$ can be decomposed into n IMFs and a residue.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{19}$$

3. Defining the edges: The edges indicate links and connections between nodes and display different patterns of functional or structural connectivity. The edges, in functional connectivity, indicate the time series correlation of nodes [18].

4. Computing the connectivity matrix: The connectivity matrix, adjacency matrix, consists of information concerning connectivity patterns of nodes. In this matrix, the connectivity is explained by an N × N symmetric matrix, in which the columns (j) and rows (i) represent nodes; and matrix entries (aij) represent edges [18].

5. Converting the connectivity matrix into a binary matrix: Matrix binarization is used to develop an unweighted undirected matrix from the adjacent matrix [15]. For this purpose, in the first step, a threshold value is considered. If the correlation between two nodes in the connectivity matrix exceed the threshold, the value of edge corresponding to those nodes considers one otherwise zero.

6. Selecting the threshold value: this value is used to simplify the complexity of network by removing insignificant and weak edges from the network.

7. Selecting and applying a functional connectivity measurement: different measurements can be used to calculate functional connectivity such as correlation, magnitude squared coherence, phase locking value, mutual information, and transfer entropy [18].

8. Construct the network: by following mentioned steps, the network can be constructed.

4.2.2 Effective Reproduction Number

R_t (effective reproduction number) is "the expected number of new infections caused by an infectious individual in a population where some individuals may no longer be susceptible" [4]. Calculating R_t can determine when the vaccine intervention affects COVID-19 transmission in certain time-steps [19]. To eliminate the pandemic, R_t should be reduced to 0 and to bring the pandemic under control R_t must be less than 1 [5]. Therefore, estimating situation- and time-specific R_t can help to understand the pathogen transmissibility. Among different methods that have been developed to estimate R_t, we selected the instantaneous reproductive number method specifically the method developed by Cori et al. [5] as follows.

$$R_t = \frac{I_t}{\sum_{s=1}^{t} I_{t-s} w_s} \tag{20}$$

- w_s is the generation interval, which is "the time between the infection time of an infected person and the infection time of her or his infector",
- I_t is defined as the number of incidents of infections on day t [20].

The generation interval is borrowed from Nishiura et al. [21] where the standard deviation and mean of the serial interval 2.9 days (95% CrI: 1.9, 4.9) and 4.7 days (95% CrI: 3.7, 6.0) were reported. Finally, the R package EpiEstim is used for determining R_t [22].

4.2.3 GNN

Graph neural networks have been used for different applications such as computer vision, natural language processing, and chemistry [12]. However, one of the main applications of GNN is in time-series forecasting. In GNN, a pair of (1) information from a node's connections and (2) the input node's signal, is used to efficiently inform the hidden state of the input layer [13]. GNN is a combination of graphs and structure of convolutional neural networks. In specific, the Graph Convolutional Networks is the modification of the standard Convolutional Neural Networks which use to detect low-level features from data based on nodes' characteristics and their neighborhoods' topology and aspect [12]. These low-level features can be used for different tasks such as node prediction, node labeling, and edge prediction. In this study, models were developed based on Graph Convolutional Networks. In a graph G = (V, E) where E, V are the edges and nodes sets, and A is its adjacency matrix; $H^{(l)}$ layer recursively is [23]:

$$\mathrm{H}^{(l+1)} = f\left(H^{(l)}, A\right) = \sigma(\check{\mathrm{D}}^{-\frac{1}{2}}.\hat{\mathrm{A}}.\check{\mathrm{D}}^{-\frac{1}{2}}.H^{(l)}.W^{(l)}) \tag{21}$$

Where, $\hat{A} = A+I$ (I is the identity matrix), σ is activation function, $W^{(l)}$ is the weight matrix for the l-th layer, $\check{D}$ is the nodes diagonal node degree matrix of $\hat{A}$.

4.3 The Pandemic Predictive Model

For developing the models, we considered two types of graph. For NGNN we linked the neighbor states together and construct the graph. For GTNN, we constructed a graph based on graph theory and functional connectivity. In this graph, nodes are states of the US. By considering the spreading the COVID-19 virus across the US as a very complex network, graph theory can help us to analyze the spreading the virus by representing a mathematical relationships between different states of the US.

For this purpose, time series data were collected for all states of the US from the Centers for Disease and Prevention website. Empirical Mode Decomposition (EMD) is used for preprocessing the collected time series COVID-19 confirmed cases data. We used PyEMD library for Python implementation of Empirical Mode Decomposition [24]. By applying EMD, data of all the US states were divided into seven or eight IMFs. For example, IMFs of Alabama state is shown in Figure 15.

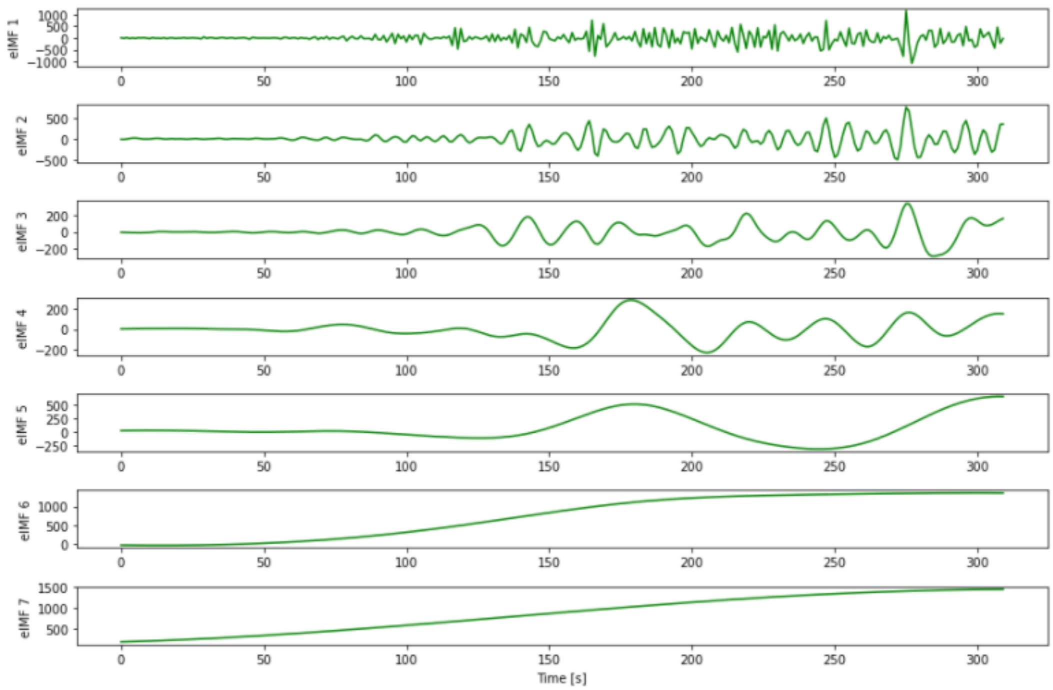

Figure 16. IMFs of Alabama State

By removing highly oscillated IMFs including IMF 1 (which was oscillated daily), IMF 2 (which was oscillated almost bi-weekly), we combine remaining IMFs to construct smoothed time series data for all states. Then, we calculated Pearson's correlation coefficients (r) between time series from all states of the US. Because r among all states were high, we computed Pearson's correlation coefficients between percent changes in time series to make sure the correlation coefficient represent actual connection between time series of the US states. The results delivered a symmetric correlation matrix Cij (size 51 × 51), in which an element in the i, j position indicated a correlation between percent changes of time series of states i and j. By considering 0.3 as a threshold value, we developed a binary matrix to maintain the strongest links between time series of different states and remove the weak connections. Finally the results were used to construct the COVID-19 correlation network as it is indicated in Figure 16. For example, COVID-19 time series of Arizona state was correlated with 21 states and Utah states was not correlated with any state.

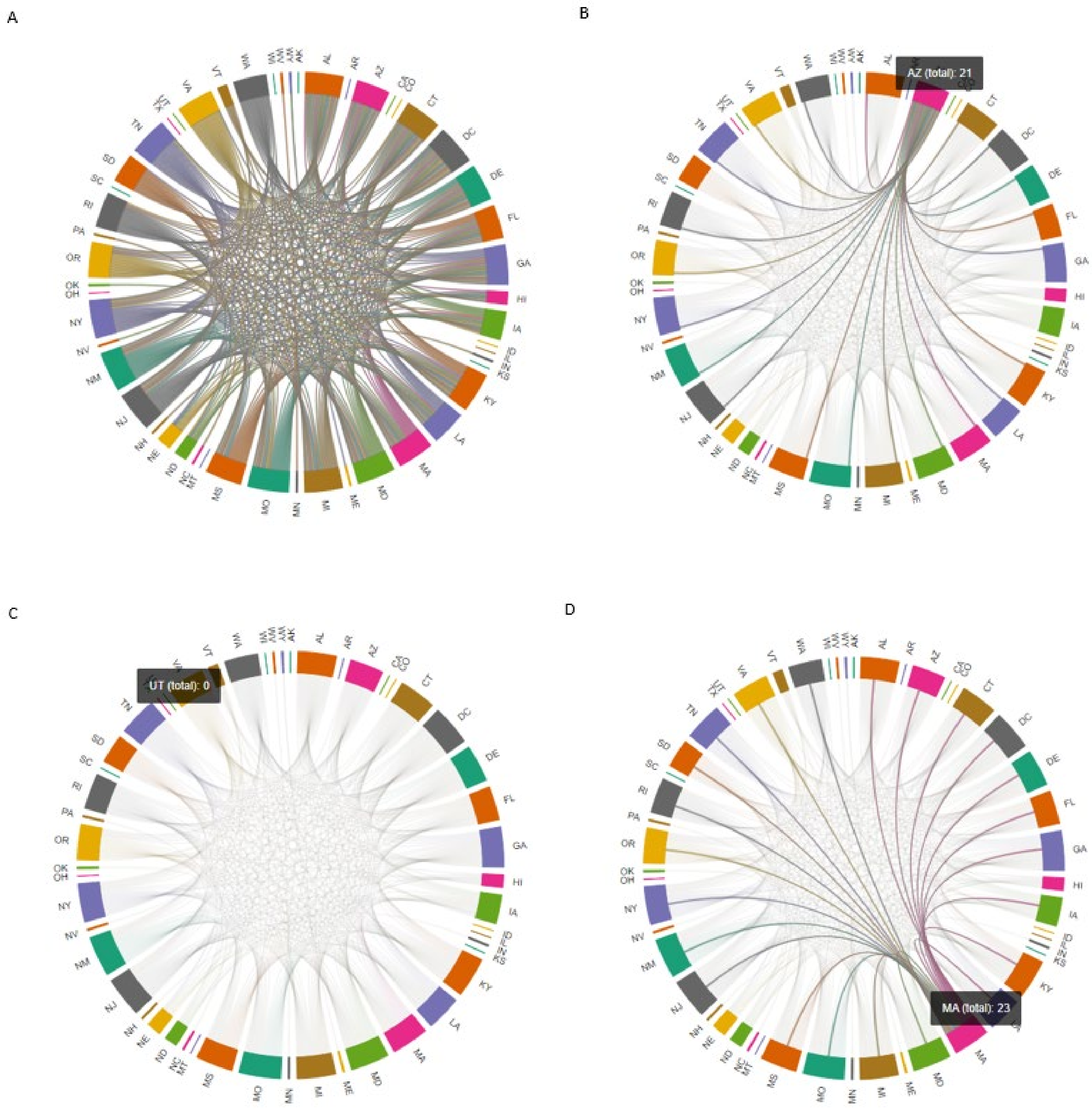

Figure 17. The COVID-19 correlation network for all states (A), Arizona (B), Utah (C), and Massachusetts (D) states

The GTNN and NGNN models are used to predict the effective reproduction numbers in all states in the US. The number of confirmed COVID-19 cases data from January 22, 2020, to November 26, 2020, is extracted from the Centers for Disease and Prevention website. The R package EpiEstim is used to calculate R_t for all states. For example, R_t for all states of the US on November 26, 2020, is represented in Figure 17.

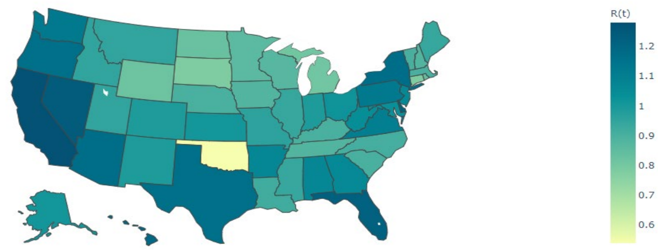

Figure 18. Rt for all states on November 26, 2020

After calculating R_t, the GTNN and NGNN models are trained. Information flow in the model contains several steps as follows:

- Passing and receiving information between connected nodes (message passing)
- Aggregating embedding of connected nodes
- Passing information to the activation function
- Apply regularizations such as dropout as it is represented in Figure 4 [6].

4.4 Experimental Study

In this section, the developed GTNN and NGNN models are evaluated on the dataset of the effective reproduction numbers. Then they are compared with each other and with baseline LSTM model to understand their predictive accuracy.

4.4.1 Experimental Setup

The performance of the models are evaluated with the datasets of effective reproduction numbers with values from January 22, 2020, through November 26, 2020. In the models, nodes are the states; edges are connections between neighborhood states for NGNN model and functional connectivity for GTNN model; node features are R_t of time steps; there are no edge features. For training and testing the dataset, R_t of each state for 4 previous days are used (as a node features) to train and predict the value R_t for each state (node feature) on the next day (1 day in advance). Dataset is divided into training dataset and testing dataset; we considered 98% of data for training (which contains randomly selected 75% training set and 23 % validation set) and 2% for testing.

Therefore, the main task of the models are predicting node feature of all states based on previous node features. The Pytorch [25] and PyTorch Geometric [26] are used for developing dataset objects and networks. We utilize an ADAM optimizer; we have three hidden layers of conv1, 2, 3.

4.4.2 Performance Metrics

The percentile error of the models, Symmetric Mean Absolute Percentage Error (sMAPE),, is used to evaluate the performance of the model as follows [27], [28].

$$sMAPE_i = \frac{1}{T}\sum_{t=1}^{T}\frac{|y_{i,t} - \hat{y}_{i,t}|}{y_{i,t} + \hat{y}_{i,t} + C} \tag{22}$$

- $y_{i,t}$ is the real value in state i at time-step t,
- $\hat{y}_{i,t}$ is the predicted value.

The constant C is added to prevent an error in calculation.

4.4.3 Performance Results

To better compare the performance of models, we computed sMAPE for seven days of testing as it is represented in Figure 23. The Average sMAPE of the GTNN model was 5.38% while sMAPE of NGNN was 5.71% and LSTM was 6.51%.

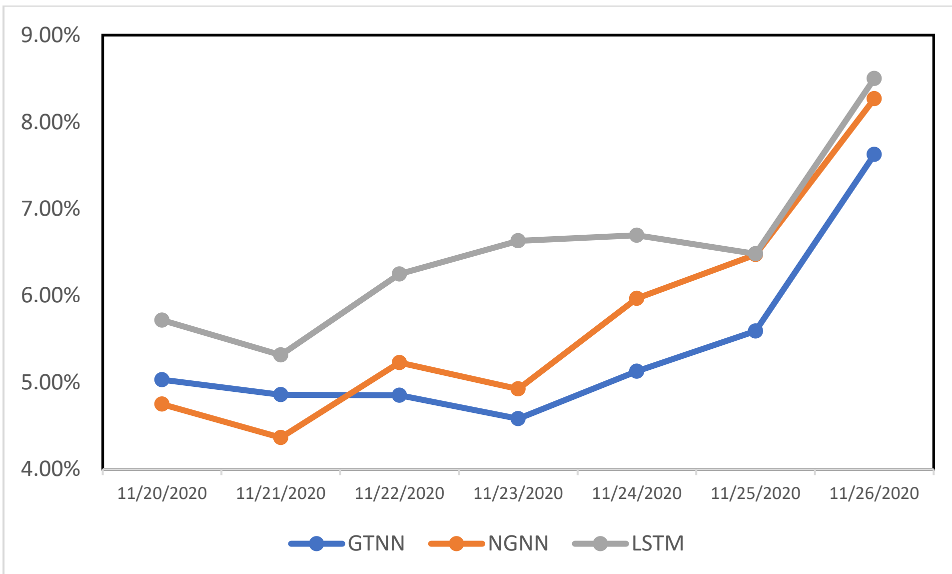

Figure 19. sMAPE for GTNN, NGNN, and baseline LSTM model for seven days of testing

Also, we represented the forecasting performance of GTNN, NGNN, and LSTM for all states of the US on November 23, 2020 in Figure 24. As it is represented in the Figure, the performance of both GTNN and NGNN outperformed baseline LSTM model.

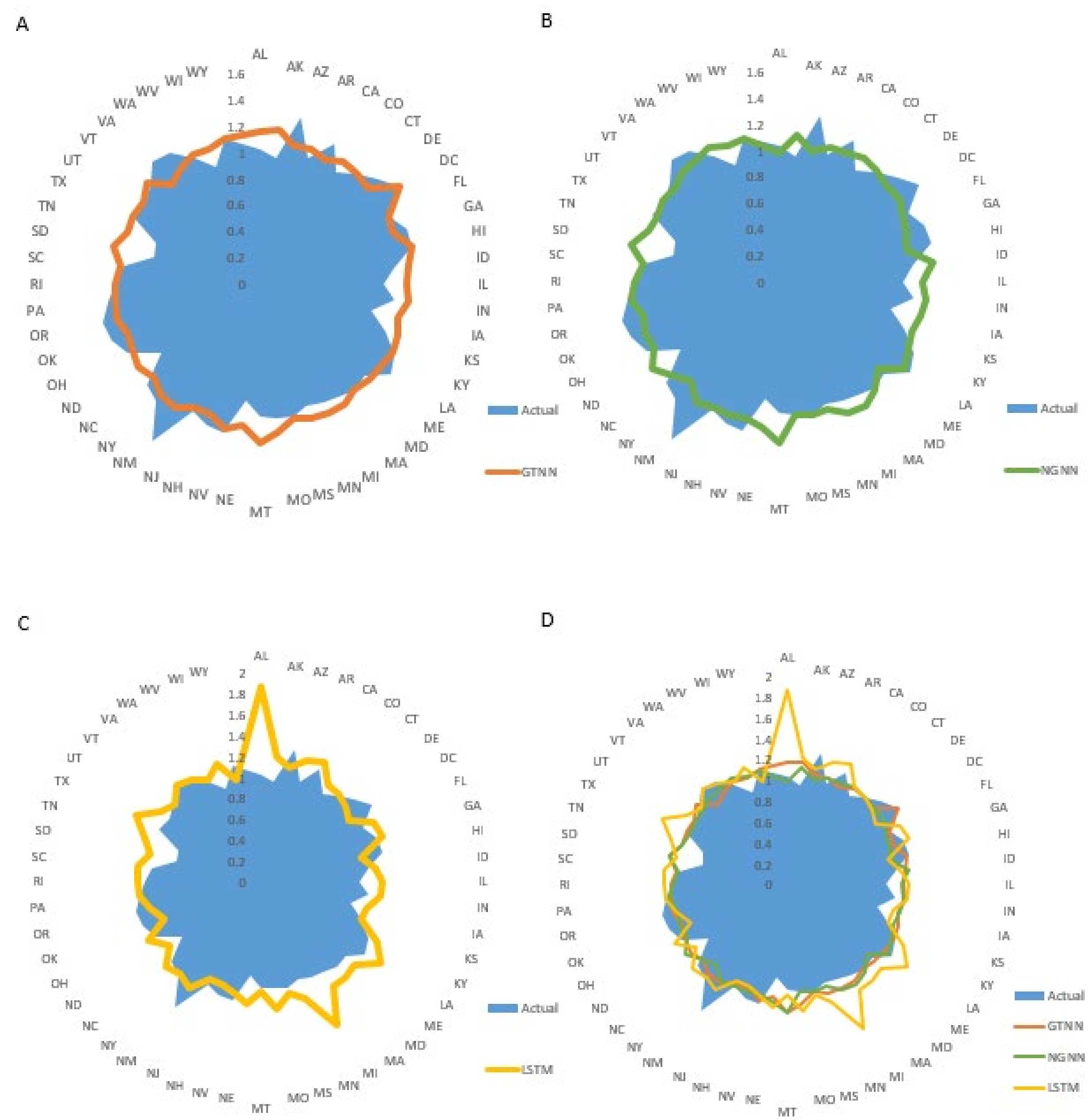

Figure 20. Actual and predicted Rt for all states on November 23, 2020 for GTNN (A) NGNN (B) LSTM (C) and all models (D)

We also calculated the average sMAPE of GTNN model for all states of the US as it is represented in Figure 25. As it is represented, states of New Jersey, Arkansas, Pennsylvania, and Texas had the minimum sMAPE and states of South Dakota, Oklahoma, Iowa, North Dakota had the maximum sMAPE over seven days of prediction (Please refer to https://github.com/RezaDavahli/Graph-neural-networks for model, input data).

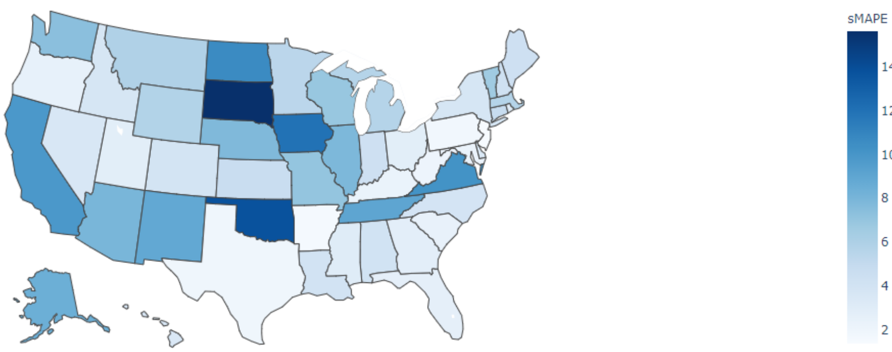

Figure 21. Average sMAPE for all states of the US for GTNN model

4.5 Limitation

One of the limitations of this study was that we did not consider the impacts of states on each other in baseline model. Some reviewed literature considered the interaction between states, especially in LSTM baseline model. Another limitation was not considering the R_t values after distributing and using the vaccine. Because we started working on this study before COVID-19 vaccination, we did not consider the impact of vaccination on the R_t number. In ths next step, we are going to calculate R_t values after initiating vaccination in the US, then calculate efficiency of developed models for predicting R_t values.

4.6 Conclusion

The COVID-19 vaccines were developed and started distributing and administering in different countries, especially in the US that hit hard by the pandemic. In this situation, one of the main questions is that when the pandemic is going to end, and normal life will start. Our objective was to develop a methodology to predict the end of the pandemic in the US. Two main elements of this methodology were predictive model and indicator of the pandemic's condition. We considered the effective reproduction number as an indicator of the pandemic. To bring the pandemic under control, R_t must be less than 1, and to eliminate the pandemic this number should be close to zero. Therefore, this number can be the perfect indicator of the end of the pandemic. For the predictive method, we select the GNN models to consider the impact of different states on each other. In addition, this method is very effective in time series forecasting. We trained the models on historical R_t patterns. We trained the GNN models and compared the results with baseline LSTM model.

4.7 References

[1] S. Charumilind, M. Craven, J. Lamb, A. Sabow, and M. Wilson, "When will the COVID-19 pandemic end? | McKinsey," 2021. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/when-will-the-covid-19-pandemic-end# (accessed Feb. 03, 2021).

[2] T. L. Microbe, "COVID-19 vaccines: the pandemic will not end overnight," *The Lancet Microbe*, vol. 2, no. 1, p. e1, Jan. 2021, doi: 10.1016/S2666-5247(20)30226-3.

[3] M. R. Davahli, W. Karwowski, S. Sonmez, and Y. Apostolopoulos, "The Hospitality Industry in the Face of the COVID-19 Pandemic: Current Topics and Research Methods," *International Journal of Environmental Research and Public Health*, vol. 17, no. 20, p. 7366, 2020.

[4] K. M. Gostic *et al.*, "Practical considerations for measuring the effective reproductive number, Rt," *medRxiv*, 2020, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7325187.1/

[5] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics," *Am J Epidemiol*, vol. 178, no. 9, pp. 1505–1512, Nov. 2013, doi: 10.1093/aje/kwt133.

[6] N. Sharma, "Introduction to Graph Neural Networks," *Medium*, Oct. 28, 2020. https://heartbeat.fritz.ai/introduction-to-graph-neural-networks-c5a9f4aa9e99 (accessed Feb. 04, 2021).

[7] CDC, "COVID-19 Cases, Deaths, and Trends in the US | CDC COVID Data Tracker," *Centers for Disease Control and Prevention*, Mar. 28, 2020. https://covid.cdc.gov/covid-data-tracker (accessed Dec. 14, 2020).

[8] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review," *Chaos, Solitons & Fractals*, p. 110059, 2020.

[9] Y. Zheng, Z. Li, J. Xin, and G. Zhou, "A Spatial-Temporal Graph Based Hybrid Infectious Disease Model with Application to COVID-19," *arXiv preprint arXiv:2010.09077*, 2020.

[10] G. Panagopoulos, G. Nikolentzos, and M. Vazirgiannis, "Transfer Graph Neural Networks for Pandemic Forecasting," 2020.

[11] D. Cao *et al.*, "Spectral Temporal Graph Neural Network for Multivariate Time-series Forecasting," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[12] V. La Gatta, V. Moscato, M. Postiglione, and G. Sperli, "An Epidemiological Neural network exploiting Dynamic Graph Structured Data applied to the COVID-19 outbreak," *IEEE Transactions on Big Data*, 2020.

[13] A. Kapoor *et al.*, “Examining covid-19 forecasting using spatio-temporal graph neural networks,” *arXiv preprint arXiv:2007.03113*, 2020, [Online]. Available: https://arxiv.org/abs/2007.03113

[14] C. Shah *et al.*, “Finding Patient Zero: Learning Contagion Source with Graph Neural Networks,” *arXiv preprint arXiv:2006.11913*, 2020.

[15] F. V. Farahani, W. Karwowski, and N. R. Lighthall, “Application of graph theory for identifying connectivity patterns in human brain networks: a systematic review,” *frontiers in Neuroscience*, vol. 13, p. 585, 2019.

[16] D. A. Cummings *et al.*, “Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand,” *Nature*, vol. 427, no. 6972, pp. 344–347, 2004.

[17] P. Flandrin, G. Rilling, and P. Goncalves, “Empirical mode decomposition as a filter bank,” *IEEE signal processing letters*, vol. 11, no. 2, pp. 112–114, 2004.

[18] L. E. Ismail and W. Karwowski, “A Graph Theory-Based Modeling of Functional Brain Connectivity Based on EEG: A Systematic Review in the Context of Neuroergonomics,” *IEEE Access*, vol. 8, pp. 155103–155135, 2020.

[19] J. Sciré *et al.*, “Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft,” *Swiss Medical Weekly*, vol. 150, no. 19–20, p. w20271, 2020.

[20] E. Kenah, M. Lipsitch, and J. M. Robins, “Generation interval contraction and epidemic data analysis,” *Math Biosci*, vol. 213, no. 1, pp. 71–79, May 2008, doi: 10.1016/j.mbs.2008.02.007.

[21] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, "Serial interval of novel coronavirus (COVID-19) infections," *International journal of infectious diseases*, vol. 93, pp. 284–286, 2020.

[22] M. R. Davahli, "Input datasets, GNN developed models, and figures." 2021. [Online]. Available: https://github.com/RezaDavahli/Graph-neural-networks

[23] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[24] D. Laszuk, *EMD-signal: Implementation of the Empirical Mode Decomposition (EMD) and its variations*. 2021. Accessed: Mar. 08, 2021. [Online]. Available: https://github.com/laszukdawid/PyEMD

[25] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703*, 2019.

[26] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," *arXiv preprint arXiv:1903.02428*, 2019.

[27] A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, 2016.

[28] J. Xu, R. Rahmatizadeh, L. Bölöni, and D. Turgut, "Real-time prediction of taxi demand using recurrent neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2572–2581, 2017.

CHAPTER 5: GRAPH THEORETICAL MODEL

Material of this chapter has not published in any journal or conference yet.

China officially reported the first case of a new virus identified as coronavirus (COVID-19) On December 8, 2019 [1]. Due to the failure of Chinese government in controlling the COVID-19 virus, it has been spread to many countries and turned into a global pandemic [2]. The US and Japan were among the countries affected by the COVID-19 virus.

The US reported its first confirmed case of COVID-19 on January 20, 2020 [3]. By the end of January, the number of confirmed cases increased to six cases, and consequently, the US government restricted travel from China and declared a public health emergency [4]. By the end of February, the number of COVID-19 confirmed cases grew to 60; on March 13, cases climbed to more than 2,100 and the US administration declared a national emergency over the COVID-19 outbreak [4].

From January 2020 to July 2021, the US faced three waves of COVID-19 infection. The first wave started from March 20, 2020 to June 10, 2020 with 2,104,956 confirmed cases and 118,464 death; the second wave from June 10, 2020 to September 16, 2020 with 4,889,694 confirmed cases and 84,521 death; the third wave started from September 16, 2020 to June 20, 2021 with 27,300,183 confirmed cases and 414,833 death [5].

To address the pandemic, the US federal and states governments focused on four key measures of (1) investing in research to accelerate production of vaccines, diagnostics, and treatments; (2) improving access to diagnostics and treatment; (3) improving health systems delivery to have a fast response to the COVID-19 outbreak; and (4) increasing the availability of data to improve surveillance [6]. However, the US was not successful in implementing all key measures and they have been partially addressed. The US was mainly successful in increasing funding for scientific research, developing vaccines, and changing regulations regarding telemedicine; however, the US is falling most behind in developing homogenous policy among all states, equitable access to treatment, and effective surveillance [6].

Japan reported the first confirmed case of the COVID-19 infection on January 16, 2020 [7]. By the end of February 2020, several confirmed cases were identified and consequently, the Japanese government closed all schools [8]. The number of COVID-19 cases increased considerably by mid-March, and the government declared a state of emergency on April 16, 2020 [7] [9].

From January 2020 to July 2021, Japan faced four waves of COVID-19 infection. The first wave started January 26 to May 31, 2020 with 16,582 confirmed cases and 898 death; the second wave June 1 to July 31with 19,120 confirmed cases and 114 death; the third wave started from October 10, 2020 to March 6, 2021 with 349,344 confirmed cases and 6,612 death; and the fourth wave started from March 6, 2021 to June 25, 2021 with 353,227 confirmed cases and 6,395 death[10].

At first, Japan seemed vulnerable to the COVID-19 pandemic for different reasons such as (1) the proximity of Japan to China and the high travel volumes between two countries, (2) heavy population density and high volumes of commuters in big cities, and (3) significant percentage of elderly people [11]. However, the Japanese government could reduce the number of COVID-19 cases and controlled the spread of the pandemic [12]. The government developed and implemented a comprehensive COVID-19 response including (1) decreasing the number of travelers and returnees from key affected areas, (2) increasing the testing capacity and strengthening medical capacity, (3) framing the Basic Countermeasure Policy based on suggestions from the expert committee, (4) providing a stronger legal basis for countermeasure policy, and (5) improving the recovery of the economy [11].

By looking at the COVID-19 confirmed cases and death, it is inferred that Japan is more successful than the US in controlling the pandemic. However, analyzing the time series of COVID-19 confirmed cases and death can improve our understanding regarding the behavior of the pandemic in two countries.

One way to better understand the behavior of the pandemic is developing COVID-19 graphs and using graph theory metrics to analyze them. In this paper, we adopted the functional connectivity approach from neuroscience to develop complex network of the COVID-19 pandemic in the US and Japan. Then, we applied graph theory analysis to investigate these networks. In our previous published paper, we indicated that using functional connectivity networks could improve accuracy of graph neural networks [1]. In this paper, we focused on developing the COVID-19 networks based on functional connectivity, analyzing the networks based on graph theory, and reported the results. This paper is structured as follows: the literature review sections reviews published articles concerning pandemic diffusion; the background section explains the the functional connectivity approach; the methodology section discusses final functional network; the results section describes the developed COVID-19 networks; the discussion section analyzes the developed networks and lists limitations.

5.1 Literature Review

The spread of an infectious disease normally follow one of these types of: (1) moving wave-like from its original center to other centers called contagious spreading, (2) moves progressively from large to small centers called hierarchical spreading, and (3) containing both contagious and hierarchical components [13]. The locations over which spread occur can be treated as a graph containing nodes (regions) and the links (diffusion process) between them [13]. The spreading pattern in different locations are frequently observed to fluctuate synchronously [14]. These synchronized fluctuations can be measured by different statistics and they often indicate connectivity between locations [15].

One study used the US influenza-related mortality data to investigate the between-state progression of the influenza pandemic [16]. The study used correlation analysis between different locations and indicated that there was higher pairwise synchrony between populous states. Another study focused on spatial structure of influenza transmission from June 1918 to April 1919 in England and Wales [14, pp. 1918–1919]. The study used statistical methods (average lags and correlation analysis) to better understand different spatial and temporal characteristics of the pandemic. One study investigated the spatial-temporal pattern of dengue hemorrhagic fever (DHF) incidence [17]. The study collected a time series dataset containing 850,000 DHF infections during the period 1983 to 1997occurred in 72 provinces of Thailand. The study used cross-correlation functions to provide metrics of the spatial dependency of temporal correlation among time series [17]. Another study tried to answer the question of how influenza spread in space within one cycle of an epidemic [18]. The study investigated the Spatio-temporal dynamics of influenza and concluded the importance of diffusion over long distances due to global transportation systems.

5.2 Background

Functional connectivity is about investigating statistical interdependence between signals [19], [20]. The most popular methodologies employed in functional connectivity are including cross-correlation and coherence analysis [19], [20]. In the cross-correlation method, a correlation can be calculated between signals that are functionally interconnected by using their recorded time series. Zero lag is commonly used to calculate correlation between time series. In the coherence analysis, the correlation concepts can be applied in the frequency domain [21].

Even though functional connectivity is developed for neuroscience study, this method can be applied for the COVID-19 pandemic as well. Because synchronized fluctuations in disease infection spreading are similar to functional connectivity, the COVID-19 data can be presented as a graph, containing nodes and edges [22]. The following steps are used to develop the functional network of COVID-19 data.

Step 1. Defining the nodes of the network. In the COVID-19 network, the nodes represented states and prefectures of the US and Japan in COVID-19 datasets.

Step 2. Preprocessing the COVID-19 data. It is important to identify noise and artifacts in the time series dataset and remove them. Different preprocessing methods such as Empirical Mode Decomposition (EMD) can be used [1].

Step 3. Defining the edges of the network. The edges represent statistical measures of association and connections between nodes. In the COVID-19 dataset, edges indicate a connection between two locations in terms of their COVID-19 behavior. The edges are classified into direct or indirect with or without weights [22].

Step 4. Selecting a methodology for the functional connectivity. As mentioned earlier, correlation and coherence analysis are the most common methodologies. We used the correlation method in this study.

Step 5. Calculating the connectivity matrix. Computing the connectivity of the nodes can create the connectivity matrix, which is known as the adjacency matrix. In this matrix, nodes are represented by rows (i) and columns (j) and edges by matrix entries (aij) [1].

Step 6. Selecting the threshold value. To simplify the network, a threshold value is selected to remove weak and insignificant edges from the matrix. We selected a threshold value of 0.7 for this study.

Step 7. Forming a binary matrix. The adjacency matrix can be used to create an unweighted unidirectional matrix called the binary matrix. For developing this matrix, a threshold value must be selected. The value of the edge between two nodes would change to one if the value of correlation between nodes in the connectivity matrix exceeds the threshold [1].

Step 8. Constructing the final network.

After developing the network, the topological properties of the network can be analyzed by using different graph theory metrics. These metrics can be used to extract features and quantify network's structure. These metrics can be divided into global (graph) and local (nodal) measures. Global measures are including the clustering coefficient (CC), characteristics path length (PL), small-worldness (σ), local efficiency (Elocal), network density, global efficiency (Eglobal), transitivity (T), modularity (Q), and assortativity (r); whereas, nodal measures are including degree centrality (K), nodal centrality, degree correlation, hubs, betweenness, degree distribution, eigenvalue centrality, closeness centrality, eccentricity centrality, nodal efficiency, and motifs [22] as it is represented in Table 1.

Table 6. Network measures [22]

Measurement	**Metrics**	**Description**
Global	*PL*	*The average of the shortest path lengths over all nodes*
	CC	*Existing edges / all possible connected edges*
	σ	*Normalized CC / normalized PL*
	Eglobal	*Inverse of PL*
	Elocal	*The efficiency of all pairs of nodes*
	T	*The number of triangles in the matrix*
	D	*Numbers of connections / maximum capacity*
	r	*The tendency of a node to connect to other nodes with similar number of edges*
	Q	*Combination of nodes that are more connected to each other than the rest of the network*
Local	*nodal centrality*	*Importance of a node for the network*
	K	*The number of edges connected to one node*
	hubs	*Node with the most edges*
	degree distribution	*Probability distribution of the degrees of all nodes*
	betweenness	*The tendency of a node to be more central than other nodes*
	eigenvalue centrality	*The accessibility of a node to other nodes*
	closeness centrality	*The closeness of a node to other nodes*
	nodal efficiency	*Propagate information from one node to others*
	motifs	*A small group of nodes connected in a specific way*

Some of the metrics are visually represented in Figure 1 that borrowed from Farahani et al. [23].

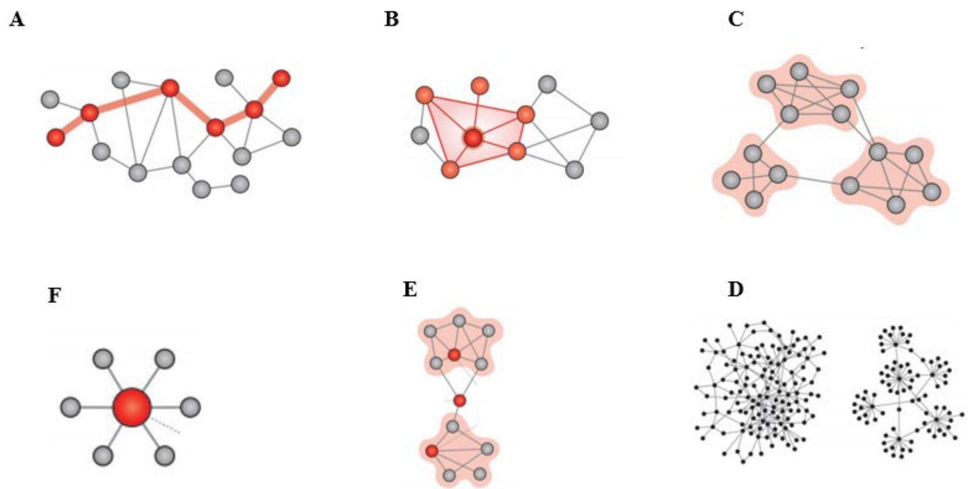

Figure 22.Schematic representation of graph theory metrics including (A) length path, (B) Clustering coefficient, (C) modularity, (D) assirtativity, (E) hubs, and (F) Degree centrality [23].

5.3 Methodology

To compare the behavior of the COVID-19 pandemic in the US with Japan, we used daily confirmed cases of COVID-19 from January 5, 2020, to July 31, 2021. For the US, we used daily records for all states plus New York city of the US from the Centers for Disease Control and Prevention [5]. For Japan, we used data for all prefectures from Japan Ministry of Health, Labour and Welfare [10].

By developing the COVID-19 network, graph theory can be used to understand the behavior of the pandemic by providing a mathematical representation of pairwise relations between different regions (states and prefectures). Figures 2 and 3 represent an overview of our pipeline. In this pipeline, after collecting COVID-19 data from CDC and MHLW, preprocessing step was implemented to make sure data is ready for next step. We developed two COVID-19 networks for the US and Japan. For the US dataset, the nodes represent US states plus New York City, District of Columbia, Puerto Rico, and Guam that were 54 nodes; for the Japan dataset, the node represented prefectures that were 47 nodes. Then, correlation coefficient was calculated between time series of all regions. The results were including two symmetric correlation matrixes Cij with size of (54 × 54) and (47 × 47). Then, by considering 0.7 as a threshold value, the binary adjacency matrixes were formed and weak connections were removed. Finally, we extracted the most important global and local graph metrics for both the US and Japan networks.

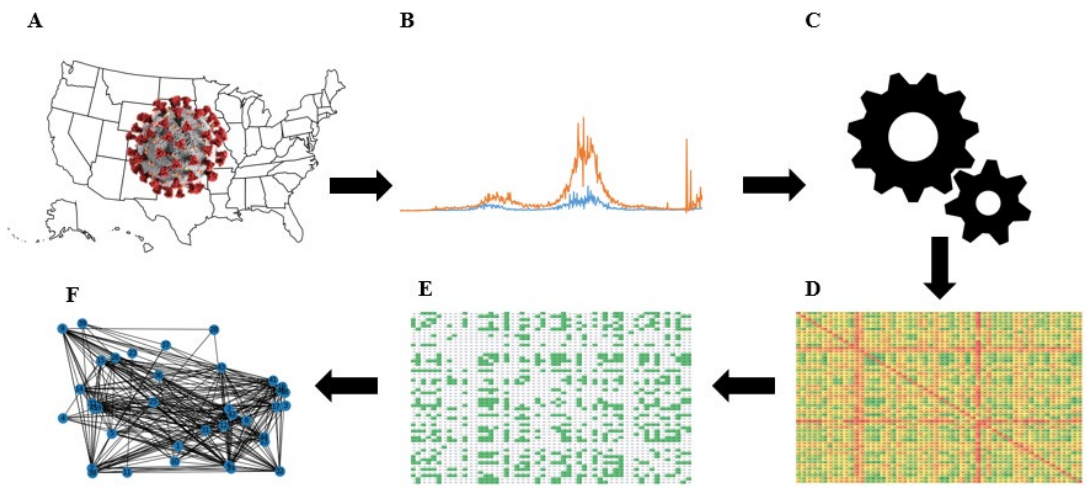

Figure 23.Schematic representation of the US COVID-19 network construction and graph theoretical analysis

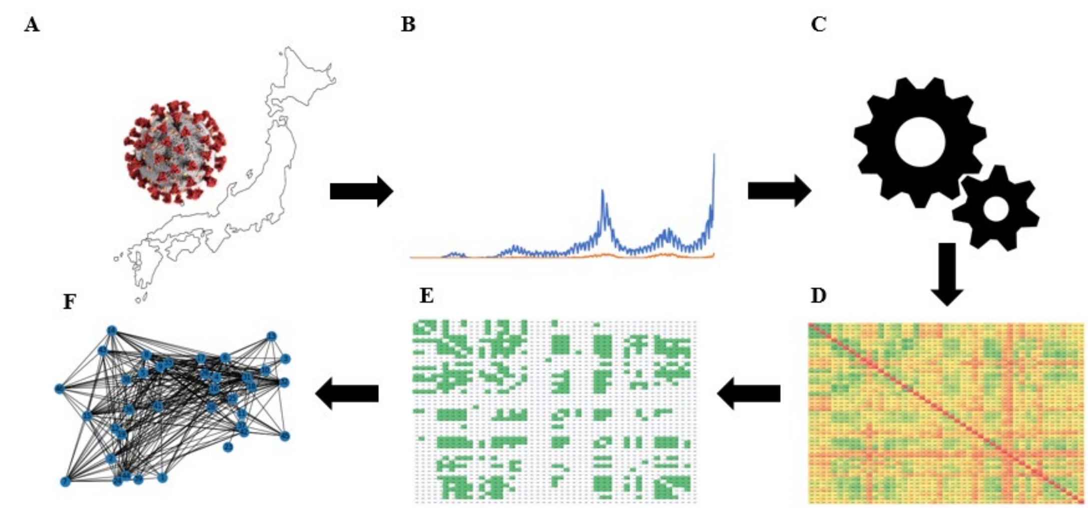

Figure 24.Schematic representation of Japan COVID-19 network construction and graph theoretical analysis

5.4 Results

By implementing the described methodology, the US and Japan graphs were created as it is shown in Figures 4 and 5.

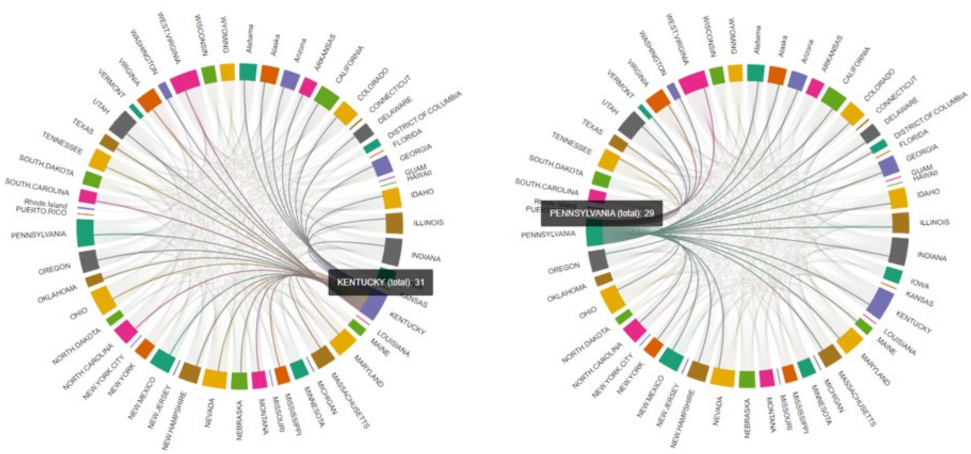

Figure 25.Schematic representation of the US graph for Kentucky and Pennsylvania

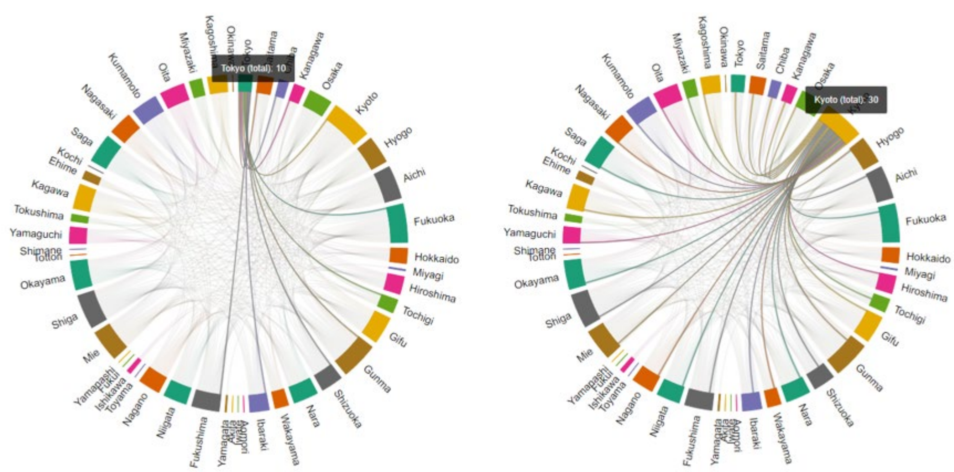

Figure 26.Schematic representation of the Japan graph for Tokyo and Kyoto

Furthermore, the results of graph theory analysis for both networks are represented in Table 2.

Table 7.The results of graph theory analysis

Metrics	The US	Japan
PL	*1.46*	*1.37*
CC	*0.72*	*0.74*
Eglobal	*0.68*	*0.73*
Elocal	*0.83*	*0.84*
Network density	*0.249*	*0.253*
r	*0.0055*	*0.019*
Q	*0.32*	*0.0077*
hubs	*Kentucky*	*Kyoto*

We explore each metrics as follows:

Path length (PL). Average shortest-path length is defined as *the average number of steps along the shortest paths for all possible pairs of network nodes* [24]. This metric indicates the efficiency of information transport on a developed network. The average shortest path length of a graph can be calculated using the following equation:

$$l_G = \frac{1}{n.(n-1)} \sum_{i \neq j} d(v_i, v_j) \qquad (23)$$

In this formula, d (v_i, v_j) indicates the length of the shortest path between two nodes. To calculate the average shortest path in a graph, sum of shortest paths between all nodes divided by number of all possible paths.

In the COVID-19 network, this metric measures the functional integration of different regions. The average shortest path for the US COVID-19 networks was 1.46 and for Japan was 1.37. A low PL represents a greater functional integration among regions, and it is an indication of the ease of COVID-19 spreading flow. Therefore, based on the US and Japan PL values, we can conclude that the COVID-19 pandemic spreads more easily in Japan rather than in the US.

Global efficiency (Eglobal). Eglobal is the inverse of PL and it is another metric to quantify the COVID-19 spread in a network. Eglobal for the US COVID-19 networks was 0.68 and for Japan was 0.73. A higher Eglobal indicates superior integration and faster transfer of the COVID-19 spreading in the scale of whole graph.

Clustering coefficient (CC). The clustering coefficient is used to better understand the function-structure of the network and CC is related to the number of triangles in a network [25]. CC of a graph can be calculate using following equation:

$$C_i = \frac{number\ of\ triangles\ connected\ to\ node\ i}{number\ of\ triples\ centered\ around\ node\ i} \tag{24}$$

In this formula, a set of two edges connected to node i is called a triple centered around node i. For the whole graph, CC is the average of the local values Ci. CC for the US COVID-19 networks was 0.72 and for Japan was 0.74. As it is indicated Japan has higher value of CC value than the US. A higher value of CC indicates more robust local interactions. Therefore, on local scale, the COVID-19 spreading pattern is faster in Japan than the US.

Network density. Network density is another metric for evaluating the effectiveness of the network. This metric is the actual number of connections in the network divided by its maximum capacity [22]. The network density for the US COVID-19 network was 0.249 and for Japan was 0.253. As it is indicated the network density of the US is very close to Japan.

Assortativity (r). Assortativity is about answering the questions of are large-degree nodes are primarily connected to low-degree nodes or is there a tendency of nodes with the same magnitude of degree to connect to each other [26]. For calculating Assortativity, we used methodology mentioned in a reference [27]. Assortativity for the US COVID-19 networks was 0.0055 and for Japan was 0.019. A higher Assortativity indicated the preference of a node in a network to attach to others that are similar. As it is indicated Assortativity for Japan network is much higher than the US. This means that in the US there is more likely that COVID-19 spreading transfer from COVID-19 epicenter to less affected areas than Japan.

Modularity (Q). Modularity metric measures the structure of a network based on the statistical arrangement of nodes [28]. Modularity can have values from -1 to 1 which value close to zero indicates that the community (modularity) division is not better than random and value close to 1 or -1 indicates strong community structure. The modularity of a graph can be calculated using the following equation:

$$Q = \sum_{i=1}^{k}(e_{ii} - a_i^2) \tag{25}$$

Where e_{ii} is the number of edges that have both ends in community i, k is the number of communities, and a_i is number of edges with one end in community i [28]. Modularity for the US COVID-19 networks was 0.32 and for Japan was 0.0077. The result indicated that the US COVID-19 spreading network is more structured and it is more module-based than Japan.

Local efficiency (Elocal). Efficiency in graph theory describes networks from the perspective of information flow [29]. The local efficiency of *a graph* is measured as follows:

$$E_{loc}(G) = \frac{1}{n}\sum_{i \in G} E_{glob}(G_i) \tag{26}$$

Where $E_{\text{glob}}(G_i)$ is the Eglobal of only a node i's immediate neighbors, but not the node i itself [29]. Elocal measures the ability of a network to spread COVID-19 at the local level. A higher Elocal indicates superior integration and faster transfer of COVID-19 spreading at the local scale. Elocal for the US COVID-19 networks was 0.83 and for Japan was 0.84. The results indicate that the COVID-19 spreading in Japan is more powerful than the US on the local scale.

Nodal centrality (K). Nodal centrality is about quantifying the importance of a node in a network and it can be measured by different metrics such as nodal efficiency, degree centrality, closeness centrality, and betweenness centrality [29]. Among these metrics, degree centrality is one of the most popular and it is the number of edges of a node. The higher the number of edges, the more central (hub) the node is. In the COVID-19 networks, for Japan, the regional hub is Kyoto and for the US, the regional hub is Kentucky.

5.5 Discussion

The covid-19 pandemic can be considered a nonlinear complex phenomenon because it broke out in one location in China and exponentially spread all over the world [30]. One of the main methods to better understand the underlying processes of the pandemic is through analyzing the COVID-19 related data. However, many traditional data analysis methods are developed only for linear and stationary data and their application in nonlinear and non-stationary data is problematic [31]. Therefore, to identify main patterns of COVID-19 behavior, it is important to take non-stationarity and nonlinearity of COVID-19 data into account. To discover the insights and implications hidden in COVID-19 related data, methodologies not only cannot impose irrelevant mathematical rules, but also they should be adaptive to underlying nature of the data [31]. To the best of our knowledge, this is the first study to apply the functional connectivity approach to analyze the behavior of the COVID-19 pandemic. Our investigation revealed two main findings by considering the graph theory metrics, respectively: (1) although Japan experienced much fewer COVID-19 cases and death, it is more vulnerable to strong waves of COVID-19 infection; (2) both Japan and the US have similarity in terms of network density and having global hubs.

To discover mentioned findings, we developed COVID-19 networks (graphs) by adopting the functional connectivity approach and analyzed the graphs' properties including path length, global and local efficiency, clustering coefficient, assortativity, modularity, network density, hubs, and degree centrality. The results of these analyses indicated that if a similar infection wave hits both countries, Japan is more vulnerable than the US. Based on graph theory analysis, in Japan, the COVID-19 pandemic spread more easily, faster, with superior integration especially on the local scale rather than the US. In Japan, Assortativity of COVID-19 networks was higher which means it is more likely that the pandemic moves progressively hierarchical from large to small centers. On the other hand, in the US, the pandemic diffusion is more likely to be contagious moving from one epicenter to another epicenters. Another main finding was that Kyoto and Kentucky were the hubs in the COVID-19 networks of Japan and the US.

Multiple challenges and future directions should be acknowledged regarding this study. First, we only focused on two countries of the US and Japan. Although the findings were significant, further investigations with more countries will be required to generalize the results. Second, we computed the correlation between time series with zero lag. However, further studies are required to investigate correlation between time series with lags. Third, even though we used correlation analysis to understand connectivity and to develop the COVID-19 network, other methodologies such as coherence analysis should be considered.

5.6 Conclusion

This study adopted the functional connectivity approach to developing the COVID-19 networks for the US and Japan. Then, graph theory analysis used to better understand the behavior of the pandemic in two countries. The main applied metrics of graph theory were including path length, global and local efficiency, clustering coefficient, assortativity, modularity, network density, hubs and degree centrality. Even though our results were significant, further studies are needed to investigate other graph theory metrics such as small-worldness, nodal centrality and network costs. Japan was much more successful in controlling the virus than the US; however, the results revealed the vulnerability of Japan to a strong infection wave. Therefore, it is necessary for the Japanese government to accelerate COVID-19 vaccination to reach herd immunity.

5.7 References

[1] M. R. Davahli, K. Fiok, W. Karwowski, A. M. Aljuaid, and R. Taiar, "Predicting the Dynamics of the COVID-19 Pandemic in the United States Using Graph Theory-Based Neural Networks," International Journal of Environmental Research and Public Health, vol. 18, no. 7, Art. no. 7, Jan. 2021, doi: 10.3390/ijerph18073834.

[2] M. R. Davahli, W. Karwowski, and K. Fiok, "Optimizing COVID-19 vaccine distribution across the United States using deterministic and stochastic recurrent neural networks," PLOS ONE, vol. 16, no. 7, p. e0253925, Jul. 2021, doi: 10.1371/journal.pone.0253925.

[3] M. R. Davahli, W. Karwowski, S. Sonmez, and Y. Apostolopoulos, "The Hospitality Industry in the Face of the COVID-19 Pandemic: Current Topics and Research Methods," *International Journal of Environmental Research and Public Health*, vol. 17, no. 20, p. 7366, 2020.

[4] S. Smith and C. McClean, "COVID-19 Overtakes Japan and the United States," *Comparative Connections*, May 15, 2020. http://cc.pacforum.org/2020/05/covid-19-overtakes-japan-and-the-united-states/ (accessed Aug. 02, 2021).

[5] CDC, "COVID-19 Cases, Deaths, and Trends in the US | CDC COVID Data Tracker," *Centers for Disease Control and Prevention*, 2021. https://covid.cdc.gov/covid-data-tracker (accessed Aug. 10, 2021).

[6] S. Bergquist, T. Otten, and N. Sarich, "COVID-19 pandemic in the United States," *Health Policy and Technology*, vol. 9, no. 4, pp. 623–638, Dec. 2020, doi: 10.1016/j.hlpt.2020.08.007.

[7] S. Bracke, A. Puls, and M. Inoue, "COVID-19 pandemic: Analyzing of spreading behavior, the impact of restrictions and prevention measures in Germany and Japan.," *medRxiv*, 2021.

[8] O. Amengual and T. Atsumi, "COVID-19 pandemic in Japan," *Rheumatology International*, vol. 41, no. 1, pp. 1–5, 2021.

[9] M.-K. Looi, "Covid-19: Japan declares state of emergency as Tokyo cases soar." British Medical Journal Publishing Group, 2020.

[10] MHLW, "Visualizing the data: information on COVID-19 infections," 2021. https://covid19.mhlw.go.jp/public/opendata/newly_confirmed_cases_daily.csv. (accessed Aug. 10, 2021).

[11] A. Tashiro and R. Shaw, "COVID-19 pandemic response in Japan: What is behind the initial flattening of the curve?," *Sustainability*, vol. 12, no. 13, p. 5250, 2020.

[12] G. Loewenthal *et al.*, “COVID-19 pandemic-related lockdown: response time is more important than its strictness,” *EMBO molecular medicine*, vol. 12, no. 11, p. e13171, 2020.

[13] M. Smallman-Raynor and A. D. Cliff, “Epidemic diffusion processes in a system of US military camps: Transfer diffusion and the spread of typhoid fever in the Spanish-American War, 1898,” *Annals of the Association of American Geographers*, vol. 91, no. 1, pp. 71–91, 2001.

[14] M. Smallman–Raynor, N. Johnson, and A. D. Cliff, “The spatial anatomy of an epidemic: influenza in London and the county boroughs of England and Wales, 1918–1919,” *Transactions of the Institute of British Geographers*, vol. 27, no. 4, pp. 452–470, 2002.

[15] J. P. Buonaccorsi, J. S. Elkinton, S. R. Evans, and A. M. Liebhold, “Measuring and testing for spatial synchrony,” *Ecology*, vol. 82, no. 6, pp. 1668–1679, 2001.

[16] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell, “Synchrony, waves, and spatial hierarchies in the spread of influenza,” *science*, vol. 312, no. 5772, pp. 447–451, 2006.

[17] D. A. Cummings *et al.*, “Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand,” *Nature*, vol. 427, no. 6972, pp. 344–347, 2004.

[18] E. Bonabeau, L. Toubiana, and A. Flahault, “The geographical spread of influenza,” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 265, no. 1413, pp. 2421–2425, 1998.

[19] F. V. Farahani and W. Karwowski, “Computational methods for analyzing functional and effective brain network connectivity using fMRI,” in *International Conference on Applied Human Factors and Ergonomics*, 2018, pp. 101–112.

[20] R. M. Hutchison *et al.*, “Dynamic functional connectivity: promise, issues, and interpretations,” *Neuroimage*, vol. 80, pp. 360–378, 2013.

[21] R. Baumgartner, L. Ryner, W. Richter, R. Summers, M. Jarmasz, and R. Somorjai, “Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis,” *Magnetic Resonance Imaging*, vol. 18, no. 1, pp. 89–94, 2000.

[22] L. E. Ismail and W. Karwowski, “A Graph Theory-Based Modeling of Functional Brain Connectivity Based on EEG: A Systematic Review in the Context of Neuroergonomics,” *IEEE Access*, vol. 8, pp. 155103–155135, 2020.

[23] F. V. Farahani *et al.*, “Effects of Chronic Sleep Restriction on the Brain Functional Network, as Revealed by Graph Theory,” *Front. Neurosci.*, vol. 0, 2019, doi: 10.3389/fnins.2019.01087.

[24] G. Mao and N. Zhang, “Analysis of Average Shortest-Path Length of Scale-Free Network,” *Journal of Applied Mathematics*, vol. 2013, p. e865643, Jul. 2013, doi: 10.1155/2013/865643.

[25] N. Masuda, M. Sakaki, T. Ezaki, and T. Watanabe, “Clustering Coefficients for Correlation Networks,” *Front. Neuroinform.*, vol. 0, 2018, doi: 10.3389/fninf.2018.00007.

[26] F. Barrenas, S. Chavali, P. Holme, R. Mobini, and M. Benson, “Network properties of complex human disease genes identified through genome-wide association studies,” *PloS one*, vol. 4, no. 11, p. e8090, 2009.

[27] R. Noldus and P. Van Mieghem, “Assortativity in complex networks,” *Journal of Complex Networks*, vol. 3, no. 4, pp. 507–542, 2015.

[28] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[29] J. Wang, X. Zuo, and Y. He, "Graph-Based Network Analysis of Resting-State Functional MRI," *Front Syst Neurosci*, vol. 4, p. 16, Jun. 2010, doi: 10.3389/fnsys.2010.00016.

[30] O. Postavaru, S. R. Anton, and A. Toma, "COVID-19 pandemic and chaos theory," *Mathematics and computers in simulation*, vol. 181, pp. 138–149, 2020.

[31] N. E. Huang and Z. Wu, "A review on Hilbert-Huang transform: Method and its applications to geophysical studies," *Reviews of geophysics*, vol. 46, no. 2, 2008.

CHAPTER 6: CONCLUSIONS, DISCUSSIONS AND FUTURE WORK

In this dissertation, the behavior and impacts of the COVID-19 pandemic are investigated. For this purpose, a systematic literature review was conducted to better understand the impacts of the pandemic. The included papers used different approaches to study the hospitality industry in the face of COVID-19 including developing simulation and scenario modeling, reporting impacts of the COVID-19 pandemic, comparing the COVID-19 pandemic with the previous public health crises, measuring impacts of the COVID-19 pandemic, recommending different actions, and conducting a survey. For these approaches, included papers used different methodologies including secondary data analysis, dynamic stochastic general equilibrium (DSGE) model, supply and demand curve, agent-based model, epidemiological model, susceptible exposed infected and recovered (SEIR) model, epidemic trajectory model, seasonal autoregressive integrated moving average model, scenario analysis, trend analysis, and the contingent valuation method. Even though included papers studied different elements of the hospitality industry, they mainly investigated the status of the hospitality industry in terms of undocumented workers, job loss, revenue losses, COVID-19 spreading pattern in the industry, market demand, recovery of the industry, safety, hostile behavior, and preferences of customers.

It should be noted that there are numerous other fertile research areas and methodologies that will need to be investigated and most likely implemented by multidisciplinary research teams. Due to the complex and dynamic nature of COVID-19 pandemic, the use of a wide array of complex systems science frameworks (e.g., syndemics) and methodologies (e.g., simulation modeling), can make an important contribution by examining how the synergistic effects of work and living conditions, as well as COVID-19 government and corporate responses, can influence the long-term health and safety of tourism and hospitality workers. Along these lines, the development and application of new technologies and equipment in the hospitality industry should protect guests and workers alike. Finally, other potential areas of research include the use of machine learning and artificial intelligence in the hospitality industry, best practices in building a more sustainable tourism and hospitality industry, and how impacts of travel and tourism activity on hosts, communities, and the environment can be minimized.

In the next step, we developed stochastic and deterministic sequence learning models based on RNNs and MDNs to predict the behavior of COVID-19 in different US states. We trained the models on historical confirmed cases and R_t patterns. The developed models can predict geographic spreading of the active virus. The primary dataset contains 310 time-steps and 50 features (US states). To avoid training the models for all states, we used the unsupervised learning methods of SOM to categorize all states into four groups according to their similarity in COVID-19 behavior. After selecting one state from each group as the leading state (the state with the earliest outbreak), we trained the developed models. We found that the predictive models trained on R_t have much better performance than those trained on confirmed cases. In addition, the deterministic LSTM model has better performance than the stochastic LSTM/MDN and linear regression models. However, the stochastic model is more successful in predicting the trends in the actual dataset. Finally, LSTM trained on R_t has the best performance, with a MAPE value of 3.46%.

In the following step, our objective was to develop a methodology to predict the end of the pandemic in the US. Two main elements of this methodology were predictive model and indicator of the pandemic's condition. We considered the effective reproduction number as an indicator of the pandemic. To bring the pandemic under control, R_t must be less than 1, and to eliminate the pandemic this number should be close to zero. Therefore, this number can be the perfect indicator of the end of the pandemic. For the predictive method, we select the GNN models to consider the impact of different states on each other. In addition, this method is very effective in time series forecasting. We trained the models on historical R_t patterns. We trained the GNN models and compared the results with baseline LSTM model.

Finally, we evaluated evidence of the pandemic behavior in the US and Japan. For this objective, we adopted the functional connectivity approach to develop the covid-19 graphs in the US and Japan. Then, we used graph theory metrics to compare the behavior of the pandemic in the US with Japan. These metrics enable the characterization of the behavior of COVID-19 that cannot be explained by simple linear methodologies. These metrics were including path length, global and local efficiency, clustering coefficient, assortativity, modularity, network density, hubs and degree centrality. Our investigation revealed two main findings. First, although Japan experienced much fewer COVID-19 cases and death, Japan is more vulnerable to a strong COVID-19 wave if both countries hit with a similar infection wave. Based on graph theory analysis, in Japan, the COVID-19 pandemic spread more easily, faster, with superior integration especially on the local scale compared to the US. In Japan, the pandemic diffusion is more hierarchical (the pandemic moves progressively from a large center to small centers) compared to the US, which the pandemic more likely to move from a large center to other large centers. Second, both Japan and the US have similar network density and Kyoto and Kentucky were the hubs in their COVID-19 networks.