
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 177 (2020) 204–211

Procedia Computer Science

www.elsevier.com/locate/procedia

The 10th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2020)
November 2-5, 2020, Madeira, Portugal

# Towards Using Graph Analytics for Tracking Covid-19

Zakariyaa Ait El Mouden[1,*], Rachida Moulay Taj[1], Abdeslam Jakimi[1], Moha Hajar[1]

[a]*Software Engineering & Information Systems Engineering, FST Errachidia, Moulay Ismail University, Meknes, Morocco*
[b]*Operational Research & Computer Science, FST Errachidia, Moulay Ismail University, Meknes, Morocco*

**Abstract**

Graph analytics are now considered the state-of-the-art in many applications of communities detection. The combination between the graph's definition in mathematics and the graphs in computer science as an abstract data structure is the key behind the success of graph-based approaches in machine learning. Based on graphs, several approaches have been developed such as shortest path first (SPF) algorithms, subgraphs extraction, social media analytics, transportation networks, bioinformatic algorithms, etc. While SPF algorithms are widely used in optimization problems, Spectral clustering (SC) algorithms have overcome the limits of the most state-of-art approaches in communities detection. The purpose of this paper is to introduce a graph-based approach of communities detection in the novel coronavirus Covid-19 countries' datasets. The motivation behind this work is to overcome the limitations of multiclass classification, as SC is an unsupervised clustering algorithm, there is no need to predefine the output clusters as a preprocessing step. Our proposed approach is based on a previous contribution on an automatic estimation of the $k$ number of the output clusters. Based on dynamic statistical data for more than 200 countries, each cluster is supposed to group countries having similar behaviors of Covid-19 propagation.



*Keywords:* Covid-19; Coronavirus; Machine learning; Graph analytics; Spectral clustering; Communities detection

## 1. Introduction

In late December 2019, an increasing number of pneumonia cases was noticed in Wuhan city, China [1, 2]. Initially, those cases were classified as caused by unknown sources, but after one week, the novel coronavirus was identified and temporarily named 2019-nCoV [3].

Coronavirus or CoV is a large family of viruses discovered in 1930s in mammals and birds, later in 1960s, coronaviruses were discovered in humans. A novel coronavirus or nCoV is a new race that has not been previously iden-

* Corresponding author. Tel.: +212-642-818637.
*E-mail address:* mouden.zakariyaa@outlook.com

1877-0509 
10.1016/j.procs.2020.10.029

tified in humans. Thereafter, the novel coronavirus disease was named COVID-19 in February 11, 2020 [2, 4] which is caused by Severe Acute Respiratory Syndrome Coronavirus 2 or SARS-CoV-2. After two months, the novel virus was characterized as a pandemic as the statistics exceed 100,000 cases and 4,000 deaths in 114 different countries according to the World Health Organization (WHO).

The symptoms of COVID-19 disease can be divided into two parts; *i) Systematic disorders* such as fever, cough, fatigue, headache, hemoptysis, acute cardiac injury, hypoxemia, dyspnea, diarrhea and lymphopenia. *ii) Respiratory disorders* such as rhinorrhea, sneezing, sore throat, pneumonia, ground-glass opacities, RNAaemia and acute respiratory distress syndrome [5, 6].

From infection to the first symptoms, the incubation period is from 2 days to 14 days in most cases, but a maximal value of incubation period was observed with 27 days in $22^{th}$ February 2020, other values of 19 and 24 were noticed after, which proves that the incubation period can vary widely among patients and makes the subject of serious of coronavirus and its ability to spread. While searching for vaccines, the majority of countries implemented quarantine and travel restrictions to influence to spread of this novel coronavirus [7, 8].

While writing this paper, the number of COVID-19 cases in the world has reached 6149738 including 370497 deaths (6%) and 2729904 recovered cases (44.4%) according the last update on Worldometers in May 31, 2020. After more than 5 months, China has moved from the top of the table of cases to the 17th place, leaving the first place to USA with more than 1816122 cases (29.53%) of total cases in the world) and 105584 deaths (25% of total deaths in the world).

Our contribution focuses on the use of machine learning algorithms to manipulate Covid-19 data; the existing approaches classifies countries according to predefined classes and using statistical data in function of time, which is a very basic classification that requires to define the classes before processing the algorithm. The high success of multiclass classification for Covid-19's medical images [9] doesn't make this approach applicable for other situation using other formats of data, the use of each approach will be discussed in the section of related works. Our work is based on Spectral Clustering (SC) which is a clustering approach and not a classification one; the difference is that in advance, we do not have any idea about the number or the structure of the output groups, it is the combinison between the features that makes a set of countries have similar behaviors and then form a strong cluster with minimal links with countries from other clusters. SC is not a foreign tool in medicine, many previous works linked SC to protein-to protein interactions [10, 11], medical imaging [12, 13] and we wish that our study will open doors on using SC in tracking epidemics.

The paper is structures as follows; After this introduction, Section 2 details the literature review, Section 3 presents briefly the graph analytics thematic and locates our approach in this thematic. Including a graphical abstract, Section 4 presents the different processes of our proposed approach. Section 5 concludes the paper and presents some perspectives for future works.

## 2. Related Work

Since the first appearance of Covid-19 disease, many disciplines have joined the scientific community of Covid-19. Artificial Intelligence (AI) in turn, contributed with various works to support Covid-19 challenges such as modeling, simulation, predictions, social networks analytics, Geographic Information Systems (GIS) for spatial segmentation and tracking, etc.

In [14], the authors highlight the importance of GIS and big data challenges against Covid-19 challenges represented for GIS by spatial tracking of confirmed cases, predictions of the epidemic transmission from region to region, spatial data visualization and segmentation, ... etc. For big data, the problem is the format of data collected from different sources which produces heterogeneous dataset that can't be processed with traditional techniques, to remedy this issue, the authors proposed that it is indisputable for the different interveners to discuss the formulation of those data especially the government and the academics. The study was presented for three scales; individual, group and regional. Another contribution in the same field is [15], where the authors studied the spatial variability of Covid-19 in the level of the continental United States, the authors implemented a Multiscale Geographically Weighted Regression for both global models and local models.

Medical images analysis can be considered as the AI field with the higher number of contributions since the first appearance of the novel coronavirus. In [16] the authors presented a comparative study between seven recent deep

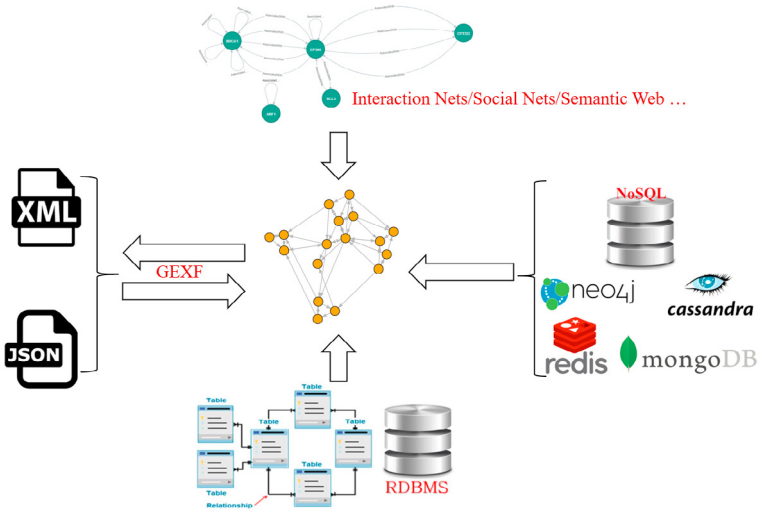


Fig. 1. Graph construction from different input data types (Documents, Relational databases, NoSQL db, interaction networks, . . . etc.)

learning methods for Covid-19's detection and classification using chest X-ray images. Using the same chest X-ray images, the authors of [17] proposed their deep learning model CovidX-Net based on the Convolutional Neural Networks (CNN) to classify the patients to either positive or negative cases, as results interpretation, the authors recommended the use of the Dense Convolutional Network models (DenseNet, $F_1 = 0.91$) which gives better results in comparison with other models such as InceptionV3 ($F_1 = 0.67$).

Neural network models were not limited to chest X-ray images analysis, other works used those models to study the efficiency of quarantine systems in different countries, such as [18], where the authors analyze the results of quarantine and isolation in China, Italy, South Korea and USA. The authors concluded that any relaxing of quarantine measures before estimated dates will lead to higher spread of Covid-19 disease in USA. The study was elaborated for data available from the start of the epidemic to March, 2020.

In the other hand, Spectral clustering has also proven its efficiency against big data challenges, with numerous applications in computer science such as communities' detection [19, 20], bioinformatics [21], image processing [22, 23], . . . etc. Recent works combined between spectral methods and deep learning models, such as the case of [24] where the authors presented their deep clustering approach to cluster data using both neural networks and graph analytics. The image segmentation presented by the authors of [22] is a combinison between SC and regularization which gives good results for high-dimensional image segmentation in comparison the state-of-art algorithms.

## 3. Graph Analytics

Graph analytics is a mathematical field of study that regroups all the algorithms and approaches based on graphs, we cite for example discovering the meaningful patterns using mathematical properties of graphs. One of the many powerful sides of graphs, is that those data structures can be built from any type of structured, semi-structured or even heterogeneous data (See Fig. 1), also graphs can be imported and exported as objects using different formats, one of the main formats to describe graphs is GEXF (Graph Exchange XML Format), more details about building GEXF graph schemas from relational data can be found in our previous contribution [25].

### *3.1. Connectivity analytics*

How easy is to break the graph by removing a few nodes or edges?
How can we compare two or more graphs?

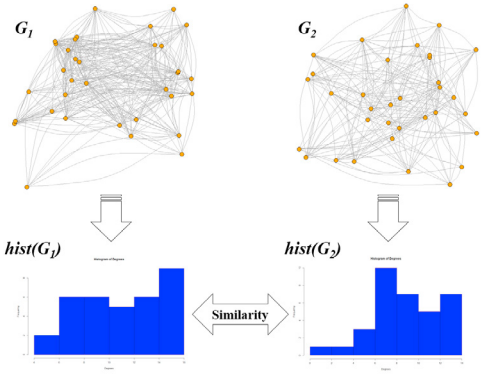


Fig. 2. Degrees' histogram

Those two questions can summarize the connectivity analytics problems. A graph is connected if it contains a path from $u$ to $v$ or from $v$ to $u$ for each pair of nodes $(u, v)$. In Connectivity analytics we study the robustness of a graph, the disconnection of graphs based on their nodes or edges and the similarity analysis (Fig.2).

As comparison between two graphs, degrees' histogram are widely used to analyze the connectivity between the nodes. In Fig. 2, the two graphs are $\varepsilon$-neighborhood graphs with different values of thresholding $\varepsilon = 0.3$ (left) and $\varepsilon = 0.5$ ( right), the two graphs describe the same data and only the connectivity is different. A graph with smaller $\varepsilon$ contains higher number of edges in comparison with a higher value of $\varepsilon$ and higher degrees, which explains the difference between the two dergrees' histograms.

### 3.2. Centrality analytics

A centrality is the measure of importance of a node or an edge based in its position in the graph. A centralization is measured for the entire graph or network as the variation in the centrality scores among the nodes or edges.

We distinguish between four types of centrality; *i) Degree centrality*, which is a measure that tells how much the graph is maximally-connected or close to a clique. *ii) Group degree centrality* is close the first type of centrality but we consider a set of nodes as a single node and we measure how much is that subset of nodes close to a clique. *iii) Closeness centrality* is the sum of shortest paths from all other nodes to a single node, a low closeness centrality means that our node has short distances from other nodes which makes it a key player in the graph by receiving information sooner and influencing other nodes directly or indirectly. *iv) Betweenness centrality* is the ratio of pairwise shortest paths that flows through a node $i$ and count all the shortest paths in the graph, a low betweenness centrality means that the node $i$ can reach other nodes faster than other nodes in the graph.

### 3.3. Community analytics

A community or a cluster is dense subgraph where the nodes are more connected to each other than nodes outside the graph. From this definition, we can model a communities' detection in a graph in a graph as a multi-objective optimization problem; Let $G$ be a graph and $c$ a connected subgraph of $G$,

For each cluster $c$, we compute its intra-cluster density $\sigma_{int}$ and its inter-cluster density $\sigma_{ext}$, where:

$$\sigma_{int} = \frac{\#\ of\ internal\ edges\ in\ c}{n_c(n_c - 1)/2} \tag{1}$$

$$\sigma_{ext} = \frac{\#\ of\ intercluster\ edges\ in\ c}{n_c(n - n_c)} \tag{2}$$

With $n$ the number of nodes in $G$ and $n_c$ the number of nodes in the cluster $c$. A good communities' detection (also called graph clustering) is the optimization function with minimal value of $\sigma_{ext}$ and maximal value of $\sigma_{int}$. Louvain community detection [26] and normalized spectral clustering [27, 28] are the most used communities' detection algorithms for graphs and complex networks.

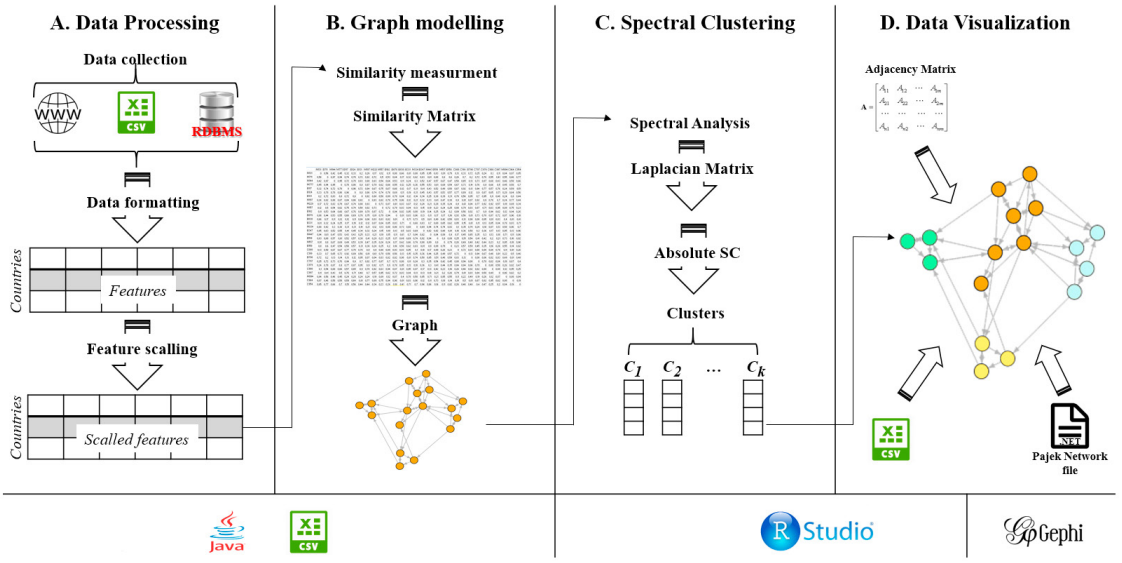


Fig. 3. Approach overview

## 4. Proposed approach

Our proposed approach consists of a SC based communities detection where the objective is to have an unsupervised grouping of countries having similar behaviors of Covid-19 spreading. The approach is meant to be applied dynamically to track the coronavirus behaviors in the world. The approach is divided into four steps (see Fig. 3).

### *4.1. Data processing*

*"It's not who has the best algorithm that wins. It's who has the most data"*. Andrew Ng.

Data processing is a key process in every machine learning model, even with an efficient model, a bad data collection, formatting or scaling can lead to very bad results. As we know, data sources are no longer consistent, data are available and easily accessible from different devices, but each source provides different format of data (text, csv, multimedia, html, ... etc.) with produces heterogeneous collected data.

With the high spread of the novel coronavirus, data related to this disease keep growing, which causes two major problems, the selection of trusted sources and data formatting. For the application, data were collected from *EU Open Data Portal* which is a trusted source where data can be collected and reused free of charge and without any copyright restrictions. This portal offers a daily updated data in XLSX format, which can be easily converted to CSV (Comma-Separated Values) for further processing.

The portal presents Covid-19's data for statistical studies, where the most important feature is time, the presentation of data according to time produces a high redundancy for a lot of static data, such as the country name, its geographical code, the continent, the population value, ... etc. To remedy the redundancy problem, we proposed a preprocessing step of data formatting where we applied the object-oriented programming concepts; each country was converted to an object with a set of attributes and methods.

For each country object, we are interested in collecting a set of features (Table 1), then those features will be used to calculate another set of features (Table 2) in have more data to manipulate (See Table 1 and Table 2).

The vectors $vTests$, $vCases$, $vDeaths$ and $vRecovers$ store the daily value for Covid-19's tests, cases, deaths and recovers respectively for each country, from the first appearance of the coronavirus in the country to the day of executing the model which produces vectors with different sizes for different countries, but for each single county, the vectors will have the same size. The size of the vectors is stored as an extra feature called Contamination days. The sums of the values of each vector are stored in the features $sumTest$, $sumCases$, $SumDeaths$, $SumRecovers$. For

Table 1. Collected Features

| Attribute | Type | Range of Values |
|---|---|---|
| Name | Text | - |
| Continent | Text | IN [Africa, Asia, Europe, Oceania, America] |
| Population | Decimal | $\in \mathbf{R}^*$ |
| vTests<br>vCases<br>vDeaths<br>vRecovers | Vector of Integers | $\in \mathbf{N}^*$ |

Table 2. Calculated Features

| Attribute | Type | Range of Values |
|---|---|---|
| vIFR | Vector of Decimals | $\in [0, 1]$ |
| sumTest<br>sumCases<br>sumDeaths<br>sumRecovers<br>ContaminationDays | Integer $\in \mathbf{N}^*$ | |

each day we store the value of the Infection Fatality Rate ($IFR$) which produces a vector with same size of the four previous vectors, we named this vector as $vIFR$.

In Machine learning models, the use of features with different scales can never lead to a meaningful result, the features with higher range of values (Population feature for example) will always have more coefficient than features will smaller scales ($IFR$ for example). To remedy this problem, a feature scaling process is necessary, especially when we plan to use a similarity kernel in the next step of the approach. The use of a mean normalization will produce a second CSV dataset which will be used for the rest of the process.

### *4.2. Graphical modelling*

After the feature scaling, we apply the Gaussian similarity kernel to measure the similarities between each pair of countries $c_i$ and $c_j$ basing on the vectorization of the predefined features in the first step of the process. Gaussian similarity is defined as follows:

$$s(c_i, c_j) = exp(-\frac{\|c_i - c_j\|^2}{2\sigma^2}) \tag{3}$$

With $\|c_i - c_j\|$ the Euclidean distance between the corresponding vectors of the pair of countries and $\sigma$ a manually-chosen positive scaling parameter to control the size of the neighborhood of our produced graph.

The result of the similarities measurement is a similarity square matrix $S \in \mathbf{R}^{n \times n}$ where $n$ is the number of the countries in the dataset and each element of the matrix $S_{ij} = s(c_i, c_j)$. This matrix will be used to build our graph, where the set of vertices is the countries of our dataset, and the edges are weighed with the similarities of $S$.

### *4.3. Spectral clustering*

For spectral clustering process, we prefer the Absolute SC algorithm with an unsupervised choice of $k$ number of clusters. This version of SC is based on the use of the Absolute Laplacian matrix defined as follows:

$$L_{abs} = D^{-1/2} W D^{-1/2} \tag{4}$$

Where $D$ is the diagonal degrees matrix and $W$ the weights matrix, both $D$ and $W$ are square matrices of size $n \times n$. The Absolute SC can be defined by the following steps:

- Compute $Labs$;
- Extract $k$ eigenvectors associated to the $k$ largest eigenvalues of $Labs$ in absolute values;
- Store the extracted eigenvectors as columns of a matrix $U \in \mathbf{R}^{n \times k}$;
- Using $k$-means clustering, cluster the lines of $U$ into $k$ clusters.

The output is a set of clusters where each cluster groups the countries having high similarities, which means that those countries have lived similar spreading of the coronavirus.
As contamination day feature gets higher values, our model produces better results in comparison to the first days of Covid-19 disease; This is due to the vectors getting more values day after day allowing the model to have more data for similarity measurement.

### *4.4. Data visualization*

*"A picture is worth a thousand words"*.
Data visualization is a supplementary task, as the main objective of the model is to compute the clusters. But with data visualization, a visual graph can describe the output of our model better than the mathematical results or any textual description.
A graph can be built from different sources and using different tools and programming languages. In R, a graph ca be built from an adjacency matrix of types $Matrix$ or $dgCMatrix$ of the package *matrixcalc*, graph functions are available in the package *igraph* and plotting functions in the package $gplots$, Pajek [29] network files are also an available option to create graphs in R. In Gephi [30], a graph can be constructed from a pair of two CSV files, the first file for the nodes and the second for the edges, or from a GEXF file which is an XML-based document to describe graphs.
Graph databases are also an important tool for data visualization. In addition to visualization, graph-based systems provide querying languages to interact with graphs. Neo4j [31] is a widely used NoSQL system which manipulates data modelled by graphs and offers Cypher language to query the stored graphs. Neo4j supports CSV files and manually creation using Cypher querying language which makes it the most powerful tool for graph creation, manipulation and visualization.

## 5. conclusions and future works

In this paper, we proposed a graph-based approach for clustering Covid-19 data using spectral clustering. Starting with data processing, we highlighted the importance of data collection and feature scaling in increasing the efficiency of each machine learning model. Then, we described the transitional phase between the heterogeneous data collected from different sources and the graph data structure, we were based on the use of Gaussian similarity kernel to build our graph. Next, spectral analysis was applied to the built graph, we selected the absolute spectral clustering, which is based on the absolute Laplacian matrix to cluster the vertices of the graph based on the similarities between their properties. The last process was the visualization of the output data, we proposed three different tools and programming languages, then we recommended the graph-based system Neo4j as it supports a querying language called Cypher to intact with the graph.
Ongoing work intends to link the different processes of the model, developed with two different programming languages (Java and R) to build a model able to cluster heterogeneous data based on graph analytics and spectral clustering for communities' detection. Furthermore, The data processing needs more improvement to collect more data and expand the set of features, as the model still gives better clustering for countries with more data whose are the countries that discovered coronavirus earlier, than the fresh contaminated countries whose gather less data and generally appears as common nodes between all the clusters due to the fact that majority of the countries had a very similar spreading of coronavirus in the first days of the contamination.

## References

[1] J. Yanga, Y. Zhenga, and X. Goua, "Prevalence of comorbidities and its effects in coronavirus disease 2019 patients: a systematic review and meta-analysis," Int. J. Infect. Dis., vol. 94, pp. 91-95, 2020.

[2] C. Sohrabi et al., "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," International Journal of Surgery, vol. 76, pp. 71-76, 2020.
[3] P. Liu and X.-z. Tan, "2019 novel coronavirus (2019-nCoV) pneumonia," Radiology, vol. 295, pp. 19-19, 2020.
[4] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," The Lancet, vol. 395, pp. 470-473, 2020.
[5] L. Pan, M. Mu, P. Yang, Y. Sun, R. Wang, J. Yan, et al., "Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: a descriptive, cross-sectional, multicenter study," The American journal of gastroenterology, vol. 115(5), pp. 766-773 2020.
[6] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," Journal of autoimmunity, vol. 109, p. 102433, 2020.
[7] S. Lai, I. I. Bogoch, A. Watts, K. Khan, Z. Li, and A. Tatem, "Preliminary risk analysis of 2019 novel coronavirus spread within and beyond China," ed, 2020.
[8] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, et al., "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," Science, vol. 368, pp. 395-400, 2020.
[9] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images," arXiv preprint arXiv:2003.09871, 2020.
[10] G. Qin and L. Gao, "Spectral clustering for detecting protein complexes in protein–protein interaction (PPI) networks," Mathematical and Computer Modelling, vol. 52, pp. 2066-2074, 2010.
[11] H. Mahmoud, F. Masulli, S. Rovetta, and G. Russo, "Community detection in protein-protein interaction networks using spectral and graph approaches," in International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics, LNCS, vol. 8452, 2013, pp. 62-75.
[12] C.-T. Kuo, P. B. Walker, O. Carmichael, and I. Davidson, "Spectral clustering for medical imaging," in 2014 IEEE International Conference on Data Mining, 2014, pp. 887-892.
[13] K. Xia, X. Gu, and Y. Zhang, "Oriented grouping-constrained spectral clustering for medical imaging segmentation," Multimedia Systems, vol. 26, pp. 27-36, 2020.
[14] C. Zhou, F. Su, T. Pei, A. Zhang, Y. Du, B. Luo, et al., "COVID-19: challenges to GIS with big data," Geography and Sustainability, vol. 1(1), pp. 77-87, 2020.
[15] A. Mollalo, B. Vahedi, and K. M. Rivera, "GIS-based spatial modeling of COVID-19 incidence rate in the continental United States," Science of The Total Environment, vol. 728, p. 138884, 2020.
[16] K. Elasnaoui and Y. Chawki, "Using X-ray Images and Deep Learning for Automated Detection of Coronavirus Disease," Journal of Biomolecular Structure and Dynamics, pp. 1-22, 2020.
[17] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," arXiv preprint arXiv:2003.11055, 2020.
[18] R. Dandekar and G. Barbastathis, "Neural Network aided quarantine control model estimation of global Covid-19 spread," arXiv preprint arXiv:2004.02752, 2020.
[19] Z. Ait El Mouden, R. M. Taj, A. Jakimi, and M. Hajar, "Towards for Using Spectral Clustering in Graph Mining," Communications in Computer and Information Science, vol. 872, pp. 144-159, 2018.
[20] Z. A. El Mouden, A. Jakimi, and M. Hajar, "An application of spectral clustering approach to detect communities in data modeled by graphs," in Proceedings of the 2nd International Conference on Networking, Information Systems & Security, ACM, 2019.
[21] N. Afiqah-Aleng, M. Altaf-Ul-Amin, S. Kanaya, and Z.-A. Mohamed-Hussein, "Polycystic ovarian syndrome novel proteins and significant pathways identified using graph clustering approach," Reproductive BioMedicine Online, 2019.
[22] M. Tang, D. Marin, I. B. Ayed, and Y. Boykov, "Kernel cuts: Kernel and spectral clustering meet regularization," International Journal of Computer Vision, vol. 127, pp. 477-511, 2019.
[23] W. Casaca, G. Taubin, and L. G. Nonato, "Graph laplacian for spectral clustering and seeded image segmentation," in Anais do XXVIII Concurso de Teses e Dissertações, 2020, pp. 31-36.
[24] X. Yang, C. Deng, F. Zheng, J. Yan, and W. Liu, "Deep spectral clustering using dual autoencoder network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4066-4075.
[25] Z. Ait El Mouden, A. Jakimi, and M. Hajar, "An Algorithm of Conversion Between Relational Data and Graph Schema," in International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning, 2019, Smart Innovation Systems and Technologies, vol. 111, pp. 594-602.
[26] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Generalized louvain method for community detection in large networks," in 2011 11th International Conference on Intelligent Systems Design and Applications, 2011, pp. 88-93.
[27] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in Advances in neural information processing systems, 2002, pp. 849-856.
[28] M. Afzalan and F. Jazizadeh, "An automated spectral clustering for multi-scale data," Neurocomputing, vol. 347, pp. 94-108, 2019.
[29] V. Batagelj and A. Mrvar, "Pajek: Program for analysis and visualization of large networks," Timeshift-The World in Twenty-Five Years: Ars Electronica, pp. 242-251, 2004.
[30] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks," in Third international AAAI conference on weblogs and social media, 2009.
[31] J. Webber, "A programmatic introduction to neo4j," in Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity, 2012, pp. 217-218.