Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Matemática

Danillo Barros de Souza

Topological and Geometric approaches in Epidemiology

Recife
2021

Danillo Barros de Souza

Topological and Geometric approaches in Epidemiology

Tese de Doutorado apresentada ao Programa de Pós-graduação em Matemática do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Doutor em Matemática.

Área de Concentração: Geometria

Orientador: Prof. Fernando Antônio Nóbrega Santos

Recife
2021

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S729t Souza, Danillo Barros de
Topological and geometric approaches in epidemiology / Danillo Barros de Souza. – 2021.
61 f.: il., fig.

Orientador: Fernando Antônio Nóbrega Santos.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Matemática, Recife, 2021.
Inclui referências.

1. Geometria. 2. Epidemiologia. I. Santos, Fernando Antônio Nóbrega (orientador). II. Título.

516 CDD (23. ed.) UFPE- CCEN 2021 - 45

DANILLO BARROS DE SOUZA

TOPOLOGICAL AND GEOMETRIC APPROACHES IN EPIDEMIOLOGY

Tese apresentada ao Programa de Pós-graduação do Departamento de Matemática da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutorado em Matemática.

Aprovado em: 26/02/2021

BANCA EXAMINADORA

_____
Prof. Dr**.** Fernando Antônio Nóbrega Santos (Orientador)
Universidade Federal de Pernambuco

_____
Prof. Dr. César Augusto Rodrigues Castilho (Examinador Interno)
Universidade Federal de Pernambuco

_____
Prof. Dr. Manoel José Machado Soares Lemos (Examinador Interno)
Universidade Federal de Pernambuco

_____
Prof. Dr. Jones Oliveira de Albuquerque (Examinador Externo))
Universidade Federal Rural de Pernambuco

_____
Prof. Dr. Fernando Jorge Sampaio Moraes (Examinador Externo)
Universidade Federal Rural de Pernambuco

Dedico este trabalho à minha família, amigos e colegas de trabalho que fizeram parte da minha tragetória acadêmica.

ACKNOWLEDGMENTS

I would like to thank my family and my advisor Fernando Nóbrega for the patience and emotional support thought my academic career. I also would like to thank Cesar Castilho for inspiring my work through the epidemic area and for giving my directions during my research. I am really thankful for the criticisms and mind support provided by Dayane Torres during the writing and review of this work. I also would like to thank the Department of Mathematics for all the knowledge given so the dream of becoming an academic researcher could come true. In special, I would like to thank André, Antônio da Silva, Élida Ferreira, Everlon dos Santos and Jonatas Teodomiro for the technical support and brain storming for the results in this thesis. Last, but not least, I would like to thank Jones Albuquerque for making part of the structural construction of the content of my thesis.

”A day without smile is a day wasted” (CHAPLIN, 1899–1976, p. 4).

ABSTRACT

Science has tried to use advanced modelling to understand nature behaviours. Dynamical systems often rise as the first option to forecast epidemic courses from parameter estimation and epidemic curve fitting. Even though such techniques do not always lead to successful results, the answer to this fact might be coupled with the variety of data sets and presence of noise and delays in the time collection, which might interview directly in the model precision. As a well-known parallel data-driven approach, Topological Data Analysis (TDA) emerges as a trend since the last decade as a promising and powerful tool for data science. Currently, infectious diseases have threatening humanity and concerning health system worldwide. Recent Dengue outbreaks have been reported over the past years, concerned with world health care and became our motivation to use and build geometric and topological techniques for a more in-depth, data-driven understanding of this topic. Yet, the damage of the novel Coronavirus disease (COVID-19) is reaching unprecedented scales. Numerous classical epidemiology models are trying to quantify epidemiology metrics. In the work of this thesis, we propose a data-driven, parameter-free, topological and geometric approaches to access the emergence of a pandemic states by studying the Euler characteristics and Ricci curvature discretizations. We first compute this curvatures in toy-models of epidemic time-series, which allows us to create epidemic networks. Those curvatures allow us to detect early warning signs of the emergence of the pandemic. The advantage of our method lies in providing an early geometrical data marker for the pandemic state, regardless of parameter estimation and stochastic modelling. This work opens the possibility of using discrete geometry to study epidemic networks.

Keywords: Topological Data Analysis. Dengue. COVID-19. Epidemiology. Ricci Curvatures. Euler Characteristic.

RESUMO

A ciência tentou usar modelos avançados para entender os comportamentos da natureza. Os sistemas dinâmicos surgem com bastante frequência como a primeira opção para prever cursos epidêmicos a partir da estimativa de parâmetros e do ajuste da curva epidêmica. Mesmo assim, tais técnicas nem sempre levam a resultados bem-sucedidos, a resposta a esse fato pode estar atrelada à variedade de conjuntos de dados e presença de ruídos e atrasos na coleta, que podem interferir diretamente na precisão do modelo. Como uma abordagem bem conhecida e orientada a dados, Análise Topológica de Dados (em inglês, Topological Data Analysis - TDA) surge como uma tendência desde a última década, além de ser uma ferramenta poderosa e promissora para a ciência de dados. Atualmente, as doenças infecciosas vêm ameaçando a humanidade e preocupando o sistema de saúde em todo o mundo. Surtos recentes de dengue têm sido relatados nos últimos anos, se tornou uma preocupação para a saúde mundial e tornou-se nossa motivação para usar e construir técnicas geométricas e topológicas na tentativa de se obter um entendimento mais aprofundado a partir dos dados neste tópico. No entanto, os danos advindos do novo coronavírus (COVID-19) estão atingindo escalas sem precedentes. Numerosos modelos clássicos de epidemiologia estão tentando quantificar as métricas da epidemiologia. No trabalho desta tese, propomos uma abordagem baseada em dados, livre de parâmetros, topológica e geométrica para acessar a emergência de estados de pandemia estudando as características de Euler e as discretizações da curvatura de Ricci. Calculamos primeiro essas curvaturas em modelos-teste para gerar séries temporais, o que nos permite criar redes epidêmicas. Essas curvaturas nos permitem detectar os primeiros sinais de alerta do surgimento da pandemia. A vantagem do nosso método reside em fornecer um marcador de dados geométricos iniciais para o estado de pandemia, independentemente da estimativa de parâmetro e modelagem estocástica. Este trabalho abre a possibilidade de usar geometria discreta para estudar redes epidêmicas.

Palavras-chaves: Análise Topológica de Dados. Dengue. COVID-19. Epidemiologia. Curvaturas de Ricci. Caractarística de Euler.

LIST OF FIGURES

SUMÁRIO

1 INTRODUCTION

Epidemic outbreaks represent a significant concern for global health. Recent Dengue outbreaks have been reported over the past years, concerned with world health care and became our motivation to use and build geometric and topological techniques for a more in-depth, data-driven understanding of this topic. The COVID-19 outbreak has caught the attention of researchers worldwide due to its rapid spread, high fluctuation in the incubation time and uncertain health and economic outcomes.

Ubiquitous and various epidemic outbreaks have motivated the development of research in epidemic behavior across the years, as well its different approaches ((DIAS et al., 2008; JOSSERAN et al., 2006; SCALLAN et al., 2011; NAFFAKH; WERF, 2009)).

Time varying models together with statistic and stochastic techniques were presented in attempt to understand the inherent complexity of outbreak events, as the well known techniques. One of those pioneers approaches were performed in the 50's and suggested by Kermack and McKendrick (BAILEY et al., 1975; BRAUER, 2005). Over the past years, the emerge of complex network theory made possible to deal with epidemic behavior at theoretical, statistical and big data frameworks. This synergy brings another point of view for the classical techniques, e.g, the network data visualization of presenting population dynamics behavior and parameter estimation control for modelling, (PASTOR-SATORRAS et al., 2015; KEELING; EAMES, 2005; NEWMAN, 2002; PASTOR-SATORRAS; VESPIGNANI, 2001; MOORE; NEWMAN, 2000).

One of the urgent challenges of this outbreak concerns the development of a coordinated and continuous data-driven feedback system that could quantify the spread and the risk of the epidemic, without strongly depending on parameter estimation (such as the contagion rate or the basic reproduction number) and even when data is heterogeneous or subject to noise. Such a data-driven system would allow to develop adequate responses at different scales (global, national or local) and to allocate limited resources in the most effective ways.

In this work, we aim to use topological data analysis in order to investigate epidemic behavior in real data trough an exploratory perspective. Topological Data Analysis (TDA) (ZOMORODIAN, 2012; WASSERMAN, 2018; PASCUCCI et al., 2010; TAYLOR et al., 2015; EDELSBRUNNER; HARER, 2010) represents a recent perspective regarding real data treatment and has bringing unreachable results over the past years across many fields (OTTER et al., 2017; BUBENIK, 2015; SAGGAR et al., 2018; MATAMALAS; GÓMEZ; ARENAS, 2019). The numerous applicability of those topogical and geometrical ideas demonstrates the powerful tool of this techniques on several scientific fields. For example, the Ricci curvature discretization has been useful in many approaches, such as stock market (SANDHU; GEORGIOU; TANNENBAUM, 2016) and cancer diagnostics (SANDHU et al., 2015a), while

Euler characteristics has been well applied to brain activity (SANTOS et al., 2019).

Dengue disease represents a huge health concern for the world scenario, where epidemic control still a challenge for Science and Society. Furthermore, its unpredictable spreading dynamics has been threatening humanity, particularly surrounding several tropical countries worldwide (ORGANIZATION et al., 2000; ARENAS; ZANOTTO, 2013).

In parallel, the COVID-19 pandemic scenario is also challenging science. Several papers are using network approaches to investigate spreading control, isolation policies and social distancing strategies as an attempt to provide a better understanding of the COVID-19 pandemic (WEITZ et al., 2020; BLOCK et al., 2020; BISWAS; KHALEQUE; SEN, 2020; PRASSE; MIEGHEM, 2020; PRASSE et al., 2020).

Recent developments in topological and geometric data analysis (EDELSBRUNNER; HARER, 2010; PASCUCCI et al., 2010; TAYLOR et al., 2015; WASSERMAN, 2018; ZOMORODIAN, 2012) offer useful perspectives regarding real data treatment, having yielded outstanding results over the past years across many fields (OTTER et al., 2017; BUBENIK, 2015; SAGGAR et al., 2018; MATAMALAS; GÓMEZ; ARENAS, 2019). As an emerging and promising approach in network science and complex systems more generally (PETRI et al., 2013), topological and geometric data analysis describes the shape of the data by associating data with high dimensional objects (EDELSBRUNNER; HARER, 2010; SAGGAR et al., 2018; TOTH; O'ROURKE; GOODMAN, 2017).

Among the numerous successful interdisciplinary applications of applied geometry and topology, ranging from differentiating cancer networks (SANDHU et al., 2015a) to modeling phase transitions in brain networks (SANTOS et al., 2019), one idea in particular can be beneficial to infer the emergence of a network pandemic state from data by using network geometry. As this approach is completely data-driven, it could provide a geometric way to give insights into the pandemic without the need for parameter estimations. Our idea is inspired by earlier results obtained for financial networks (SANDHU; GEORGIOU; TANNENBAUM, 2016), where the authors showed that it was possible to relate financial network fragility with the Ollivier-Ricci curvature of a network. Most importantly, the Ollivier-Ricci curvature emerged as a data-driven *crash hallmark* for major changes in stock markets over the past 15 years. In their study of market fragility, they used these geometric tools to analyse and characterize the interaction between the economic agents (the nodes of a financial network) and their correlation levels (which defines the edges' weights). In addition, these tools also allowed them to track the curvature of the financial network as a function of time, i.e. how the shape of the financial network changed according to a dynamic economic scenario. This geometric study also motivated us to test the use of topological structures, as the Euler characteristics.

As a result, the Ollivier-Ricci curvature emerged as a strong quantitative indicator of the systemic risk in financial networks. Motivated by this result, we initially used this discrete version of Ricci curvatures for our geometric approach to epidemic networks. From a

pure topological point-of-view and motivated by other applications, we also included the Euler characteristics in our studies. From an implementation perspective, however, the Forman-Ricci curvatures had proved to be an alternative, simpler discretization for computing the curvature. It empirically correlates with the Ollivier-Ricci curvature even being purely combinatorial, with the added value that the Forman-Ricci curvature has a faster computation time for large-scale, real-world networks (SAMAL et al., 2018). Furthermore, the Euler characteristics showed to be a low noise tool to detect fragility on networks provided by epidemic data, despite the computational complexity.

Taking all this features in consideration, here, we are going to compare our results for both Ricci Curvatures and Euler characteristics applied to epidemic networks.

This thesis is written as follows: In chapter 2 is provided a quick background regarding network theory. In chapters 3 and 4 the reader can find the classical topological and geometric approaches used in our work. In chapter 5 is described our methods for processing and filtering data, followed by chapter 6, where we test the techniques to synthetic data. In chapter 7 the results for Dengue disease and COVID-19 data are shown. Finally, in chapter 8 we conclude all the results reached in this work.

2 GRAPH THEORY

In this chapter, we are going to introduce the basics for the understanding of our work on graph theory. We will also introduce the reader into quick and well-known results regarding random graph theory. This notions will be crucial for a better understanding of our results.

2.1 NETWORKS

The definition of graphs can be found in many classical works (BONDY; MURTY, 2008). In many approaches, a *simple graph*, or a *simple network*, is often used to encode information from a complex system. In the literature, there are many ways to define graphs. For simplicity and easy-way for applications, we will restrict our approach to *simple undirected graphs*, and for the rest of the thesis, we will refer them as graphs (or networks) only.

Formally, let V be a set of countable objects. We define E the set such that if $e \in E$, than $e = \{x, y\}$, for some $x, y \in V$, and $x \neq y$.

Well call V the *set of nodes* (or *vertexes*), while E is the *set of edges* (or *links*).

We say that $H = (V', E') \subseteq G$ is a *subgraph* of G if $V' \subseteq V$ and $E' \subset E$.

The pair $G = (V, E)$ is said a *simple undirected graph*. In our approach, we are restricting our approach for finite graphs, *i.e.,* the sets of nodes and edges are finite.

Let $x \in V$. We say that y is a *neighbor* of x when there is an edge that connects x to y, and we denote by $x \sim y$. We define the *set of neighbors* of x as

$$\text{neigh(x)} = \{y \in V,\, x \sim y\}. \tag{2.1}$$

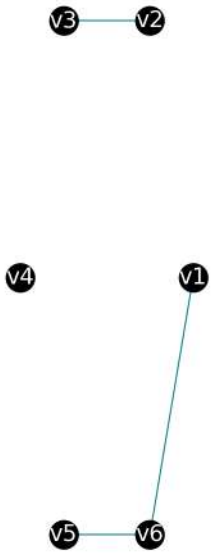
The *degree* of x is denoted by $deg(x)$ and is defined as follows:

$$\text{deg(x)} = \#\text{neigh(x)}. \tag{2.2}$$

Remark: This last definition makes sense in our approach due to the lack of *selfloops* (edges connecting a node to itself) in our definition of simple graphs. From now on, we are going to refer the natural number n for the number of nodes in a graph.

Figure 1 displays an example our representation of a simple graph. Once properly defined the graph structure, we are now able to bring some classic results from the random graph theory, which will be presented in the next section.

Figure 1 – Example of a simple graph containing 6 nodes (black dots) and 3 edges (blue lines). The degree of each node is also shown in a table.



Node	Degree
v1	1
v2	1
v3	1
v4	0
v5	1
v6	2

Source: self-provided.

2.2 ERDOS-RENYI THEORY

In attempt to study discrete curvature behavior, this section is dedicated to Erdos-Renyi random graph theory ((BOLLOBÁS; BÉLA, 2001),(ERDŐS; RÉNYI, 1960)). This classical model paved the way of the studies of randomness in network theory, despite its simplicity. In one of its versions, in the Erdos-Renyi model, we start with n isolated nodes while each edge is sequentially included according a probability p of having a connection. This edges are generated by sampling a random uniform number r in $[0, 1]$, for each combination of node pairs. Thus, for a fixed value of p, the presence of each pair of nodes $(x, y) \in V, x \neq y$ is going to depend of how often the number r is smaller than p. For $p = 0$, we have a totally unconnected graph, while $p = 1$ makes a complete graph. A *Erdos-Renyi graph* which has n nodes and probability p of connection is denoted by $G_{ER}(n, p)$.

The expected number of edges in $G_{ER}(n, p)$ is $\langle e \rangle = \binom{n}{2} p$, the expected node degree is $\langle \deg(x) \rangle = (n-1)p$, for every node x in $G_{ER}(n, p)$. The degree distribution of any node x is given by a binomial distribution as follows:

$$P(\deg(x) = k) = \binom{n-1}{k} p^k (1-p)^{(n-k-1)}. \tag{2.3}$$

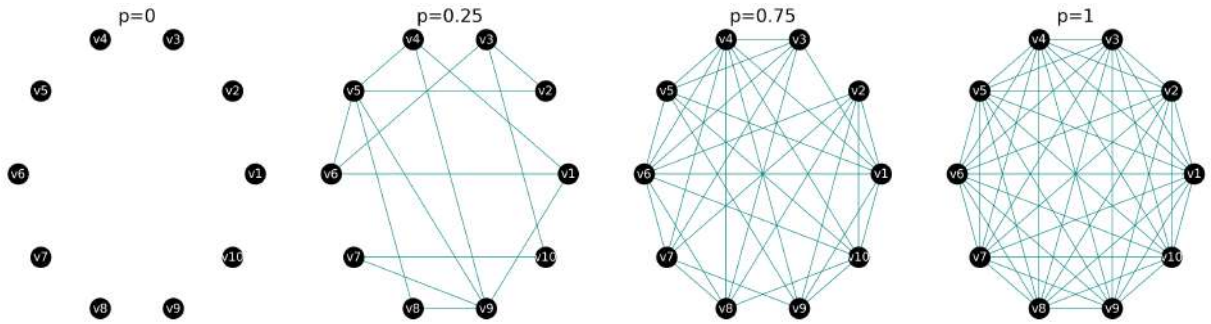
As well-known results, we also have:

- If $\langle \deg(x) \rangle < 1$, all components tend to be very small and the largest component

has size $O(\ln n)$;

- If $\langle \deg(x) \rangle = 1$, the largest component has size $O(n^{\frac{2}{3}})$;
- If $\langle \deg(x) \rangle > 1$, the rest of components are very small, with the second largest having size around $\ln n$.

Figure 2 – Examples of Erdos-Renyi graphs, for different values of probability p.



Source: self-provided.

Figure 2 provides the illustration of random graphs for different values of p. Now, we are able to make a deeper analysis over topological and geometrical changes in the graph structures based on the probability of links from Erdos-Renyi theory. This will give us estimations of how the topological and geometric curvatures might behave as the probability of links changes. Although real data, in general, does not behave as random graphs, it could give as insights, as we will see during the thesis.

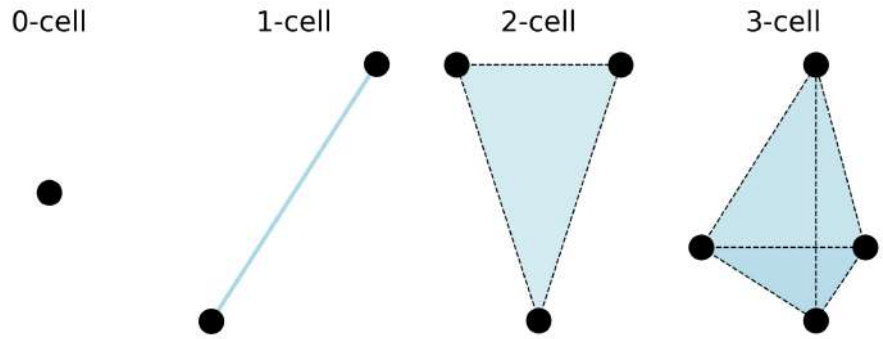
2.3 CELLS OF A GRAPH

The graph complexity not always can be summarized only by the presence of edges in a graph. It will be important to understand the meaning and the analysis of higher order structures in a graph (BATTISTON et al., 2020). Among many ways to define this structures, the common ways to refer to this structures are by *cells* or also, *cliques*, depending on the approach (topological or geometric). All the following information can be found at (HATCHER, 2005). Formally, let $G = (V, E)$ a graph and $d \in \mathbb{N} \bigcup \{0\}$. The structures equivalent to the d-dimensional open disks in the continuous approach are called *d-cells* ,or *d-faces*, or also *d-simplexes*. They are equivalent to a $(d+1)$-vertex clique. Depending on the approach, we will call them by the best notation as required.

A *clique of size* $d+1$, or a $(d+1)-$*vertex clique*, is a subset of V with cardinality $d+1$ in which all pairs of nodes are connected by an edge. They are also called *complete subgraphs* of G. An example of this d-cells is shown at the Figure 3, for $d = 0, 1, 2, 3$.

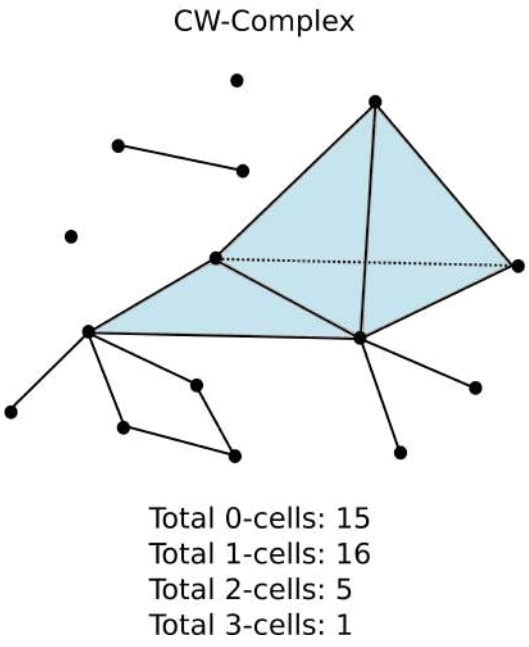
We will denote a d-cell by e^d_λ, for some $\lambda \in \Delta$, where Δ is a collection of indexes. We also will denote *d-skeleton* by X^d and define it recursively by $X^d = X^{d-1} \bigcup\limits_{\lambda \in \Delta} e^d_\lambda$, and X^0 is the set of all 0-cells.

Figure 3 – Examples of *cells* (or cliques) of different dimensions.



Source: self-provided.

Figure 4 – Example of a CW-Complex.



Source: self-provided.

The space constructed from the graph G is called a *Cell Complex* or a *CW-Complex* (FORMAN, 2003; HATCHER, 2005), and we will denote by X. The set of all d-cells of X will be denoted by $\mathcal{F}^d(X)$, and each d-cell will be expressed in terms of the set of nodes, *i.e.*, $e^d = [v_{\sigma(1)}, \cdots, v_{\sigma(d+1)}]$, where $\sigma : \{1, \cdots, d+1\} \to \{1, \cdots, n\}$ is an injective function. For example, in Figure 2, we have $\mathcal{F}^2(X) = \{[v_1, v_4, v_9], [v_4, v_5, v_9], [v_5, v_8, v_9]\}$.

In Figure 4 we have an example of a CW-Complex constructed from d-skeletons, for $d \in \{0, 1, 2, 3\}$. Clearly, X^0 is the set of nodes, X^1 is the union of X^0 with the set of edges, X^2 is the union of X^1 with the set of triangles and X^3 is the result of the union between X^2 and the tetrahedron. Once we have defined this structures, we are able to bring the notion of Betti numbers, which can be seen in the next chapter.

3 TOPOLOGICAL APPROACHES

In this chapter, we are going to introduce the definitions of the topological functions that we used in our approach. Those definitions can also be found at (EDELSBRUNNER; HARER, 2010; GREENBERG, 2018).

3.1 BETTI NUMBERS

As many other topological tools, the Betti number is one of those that is used to distinguish topological spaces based on the connectivity of n-dimensional simplicial complexes (EDELSBRUNNER; HARER, 2010; HATCHER, 2005; GREENBERG, 2018). Its applications is notable in neuroscience (SANTOS et al., 2019). Considering the definition given of a d-simplex in Chapter 2, we will construct the definition of Betti numbers. To start, we will induce an orientation to a d-simplex. An *orientation* of a d-simplex is given by an ordering of the vertices and denoted by $[v_1, \cdots, v_{d+1}]$. Two orderings induce the same orientation if and only if they differ by an even permutation of the vertices. In other words, for a permutation π on $[d]$,

$$[v_1, \cdots, v_{d+1}] = (-1)^{\text{sgn}(\pi)}[v_{\pi(1)}, \cdots, v_{\pi(d+1)}], \tag{3.1}$$

where $\text{sgn}(\pi) = (-1)^{N(\pi)}$ and $N(\pi)$ is the number of inversions in π.We assume that each simplex in our complex is assigned a specific orientation (*i.e.*, ordering).

Let $\mathbb{F}$ be a field. A simplicial d-chain is a formal sum of oriented d-simplices $\sum_i c_i e_i^d, c_i \in \mathbb{F}$. The free abelian group generated by the d-chain is denoted by the $C_d(X)$, the d-th chain group, *i.e.,*

$$C_d(X) = \left\{ \sum_i c_i e_i^d : c_i \in \mathbb{F}, e_i^d \in \mathcal{F}^d(X) \right\}, \tag{3.2}$$

for $d > 0$ and $C_{-1} = \mathbb{F}$, $C_d = \emptyset$, for $d = -2, -3, \cdots$. It is easy to verify that $C_d(X) \subseteq \mathbb{F}^d$ is a vector space. We also define the *boundary operator*, $\partial_d : C_d \to C_{d-1}$, as follows:

$$\partial_d([v_1, \cdots, v_{d+1}]) = \sum_{i=1}^{d+1} (-1)^{i-1}[v_1, \cdots, \hat{v_i}, \cdots, v_{d+1}], \tag{3.3}$$

where $\hat{v_i}$ denotes the omitted variable. Is is easy to verify that $\partial_{d-1}\partial_d = 0$. In our special case, we have $\mathbb{F} = \mathbb{Z}_2$. As an example, let $X = \{[1,2,3],[3,4]\}$. We have that $C_1(X) = \{[1,2],[1,3],[2,3],[3,4]\}$, $C_2(X) = \{[1,2,3]\}$ and $\partial_2([1,2,3]) = [2,3] - [1,3] + [1,2]$, $\partial_1\partial_2([1,2,3]) = \partial_1([2,3] - [1,3] + [1,2]) = [3] - [2] - ([3] - [1]) + [2] - [1] = 0$. We also have that the structural simplicity of this space allow as to express the boundary operator in terms of the vector basis $B_2 = [[1,2,3]]$, $B_1 = [[1,2],[1,3],[2,3],[3,4]]$ and

$B_0 = [[1], [2], [3], [4]]$ as follows:

$$[\partial_2]_{B_1}^{B_2} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \tag{3.4}$$

$$[\partial_1]_{B_0}^{B_1} = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.5}$$

In general, for every finite simplicial complex X, with $C_d(X)$ built over $\mathbb{Z}_2$ we can induce the same representation.

The d-th *boundary space* of X is denoted by $B_d = B_d(X)$ and is defined as the $\text{im}\partial_{d+1}$, and the d-th *cycle space*, $Z_d = Z_d(X)$ is $\ker \partial_d$. This way, we can construct the *d-dimentional homology group* by the quocient space given by

$$H_d = \frac{Z_d}{B_d}. \tag{3.6}$$

Finally, the *d-th Betti number* of the simplicial complex X is defined as

$$\beta_d(X) = rank(H_d(X)), \tag{3.7}$$

or simply, can be rewritten as

$$\beta_d(X) = \dim \ker \partial_d - \dim \text{im}\partial_{d+1}. \tag{3.8}$$

Using the same example, $X = \{[1, 2, 3], [3, 4]\}$ and (3.4),(3.5), we have that $\beta_0(X) = 4 - 3 = 1$ and $\beta_1(X) = 1 - 1 = 0$ and $\beta_i = 0$ for $i \geq 2$.

3.2 EULER CHARACTERISTICS

The Euler characteristics is one of the topological invariants often used to explore big data (CARLSSON, 2009; EDELSBRUNNER; HARER, 2010) and has been used to analyse networks in a topological, high-order perspective (PASCUCCI et al., 2010; TAYLOR et al., 2015; WASSERMAN, 2018; ZOMORODIAN, 2012). For us, the Euler Characteristics can induce an alternative curvature (KNILL, 2011a; WU et al., 2015) in attempt to contrast the results with the discrete Ricci curvatures.

The applications across fields has been bringing astounding results for science. For instance, the use of topological transictions as bio-markers in protein-protein networks (FILHO, 2019) and for distinguish disease by brain activities (SANTOS et al., 2019).

Furthermore, a strong result in (WEBER; SAUCAN; JOST, 2018) revels a connection between the Forman-Ricci curvatures and Euler Characteristics in an alternative Gauss-Bonnet Theorem for 2-dimensional complexes.

There are many ways to compute the Euler characteristics across fields, one being towards the Betti numbers (HATCHER, 2005; GREENBERG, 2018), by

$$\chi(X) = \sum_{i=1}^{\infty} (-1)^i \beta_i(X). \tag{3.9}$$

Remark: Originally, the sum in 3.9 is infinite, however, in our approach, as we are dealing with finite graphs, the sum is also finite. Another one is by using the *Euler curvature* (KNILL, 2011b) in each node of the graph. This method were more suitable for us due to the reduction of computational complexity. The Euler characteristics of CW complex is also given as follows:

Given $G = (V, E)$ an undirected finite graph, we define the *Euler curvature* of a node $v \in V$ as ((KNILL, 2012b)),((KNILL, 2011a)),((KNILL, 2012a))

$$K(v) = \sum_{d=0}^{d_{\max}} (-1)^d \frac{S_{d-1}(v)}{d+1}, \tag{3.10}$$

where S_d is the number of $d-1$ cells containing v, $S_{-1}(v) = 1$, and $d_{\max}$ is the highest cell dimension of the CW complex. The d-cells are the same defined at the previous section. In our case, for finite graph without self-loops and multiple connections, the Gauss-Bonnet Theorem is satisfied ((KNILL, 2011a)),((KNILL, 2012b)),((KNILL, 2014)),((KNILL, 2013)) i.e.,

$$\chi(G) = \sum_{v \in V} K(v), \tag{3.11}$$

where $K(v)$ is the curvature defined in (3.10) and can also be rewritten as

$$\chi(G) = \sum_{d=0}^{d_{\max}} (-1)^{(d+1)} k_d, \tag{3.12}$$

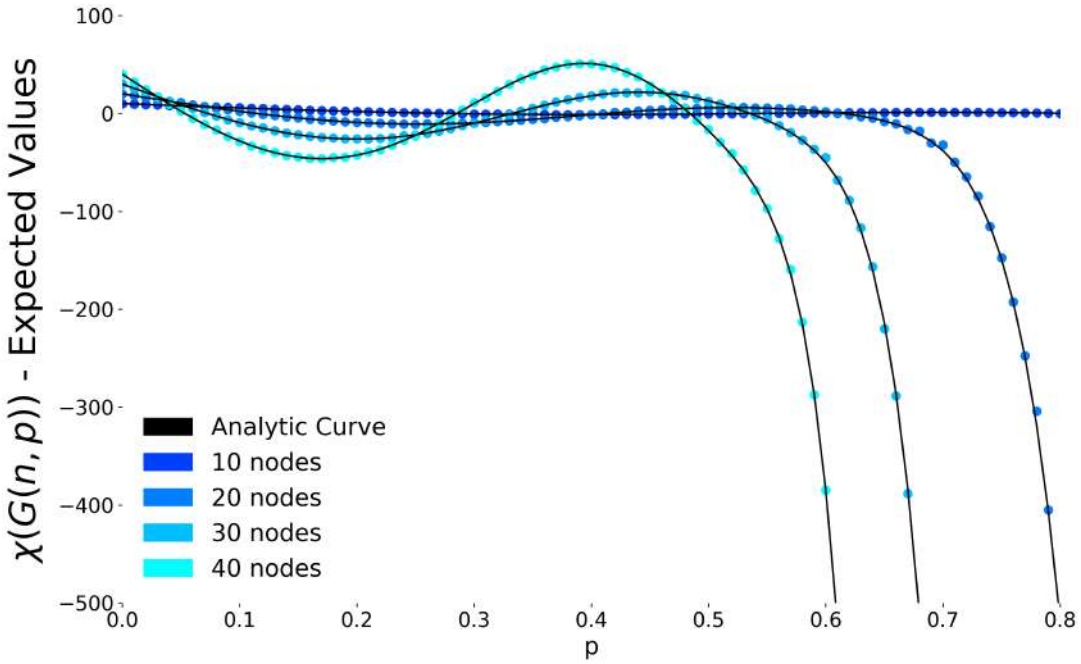
where k_d is the number of d-cells in G. For example, if $X = \{[1,2,3],[3,4]\}$, as before, from the calculus of Betti numbers, we have $\chi(X) = 1 - 0 = 1$.

The analytic expected Euler characteristic of an Erdos-Renyi graph is given by (KNILL, 2011b)

$$\mathrm{E}[\chi(G(n,p))] = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k+1} p^{\binom{k}{2}} \tag{3.13}$$

Due to computational limitations, we restricted our calculus to cells of dimension at most $d_{\max} = 6$. Figure 5 shows the results of our calculus in comparison to the expected Euler characteristics, provided by the equation (3.13). The original result can be found at (KNILL, 2011b). At this point, it was possible to analyse how the topological structure behaves as the probability of connection increases. Once compared to Forman-Ricci curvature, it was clear to realise the persistence of the high absolute values in curvatures

Figure 5 – Euler characteristic: average of 500 Erdos-Renyi graph realizations for diferent amount of nodes, restricted to dimention $d_{\max} = 6$.



Source: self-provided.

for values of p greater than 0.5. The main difference remains on the maximum of this absolute values, which are extremely higher than the Forman curvature, once it is limited to the number of nodes in the graph.

4 GEOMETRIC APPROACHES

Geometrizations are present everywhere in network science e.g. in quantum (KIBBLE, 1979; CASETTI; PETTINI; COHEN, 2000; SAVILLE; WOOD, 2012; PAWLOWSKY-GLAHN; EGOZCUE, 2001) and also offer useful perspectives regarding real data treatment and applications across fields (SANDHU; GEORGIOU; TANNENBAUM, 2016; FAROOQ et al., 2019). In this chapter, we are going to discuss some of the geometric approaches currently used in our data analysis.

4.1 OLLIVIER-RICCI CURVATURE

The Ollivier-Ricci curvature for networks is defined as follows (LIN; LU; YAU, 2011; SAMAL et al., 2018): Let $G = (V, E)$ be a graph. The path length function $d : V \times V \to \mathbb{R}^+$ defined as the length of the shortest path between two nodes induces a metric for the set of nodes of G. The neighborhood of a node $x \in V$ is the subset of nodes connected to x by an edge. Let $\alpha \in [0, 1]$ and $x \in V$. We define m_x^α, a probability measure over the set of nodes as

$$m_x^\alpha(y) = \begin{cases} \alpha & \text{if } y = x \\ (1-\alpha)/\deg(x) & \text{if } y \in \text{neigh}(x) \\ 0 & \text{otherwise} \end{cases} . \tag{4.1}$$

Finally, the *Olliver-Ricci curvature* of an edge $e = (x, y) \in E$ is defined as

$$\kappa(x, y) = 1 - \frac{W(m_x^\alpha, m_y^\alpha)}{d(x, y)}, \tag{4.2}$$

where W is the discrete Wasserstein distance (WASSERMAN, 2018) given by

$$W(m_x^\alpha, m_y^\alpha) \quad = \quad \inf_{\mu \in \Pi(m_x^\alpha, m_y^\alpha)} \sum_{x', y' \in V} d(x', y')\mu(x', y'). \tag{4.3}$$

Here, $\Pi(m_x^\alpha, m_y^\alpha)$ denotes the set of all probability measures $\mu : V \times V \to \mathbb{R}^+$ that satisfy

$$\sum_{y' \in V} \mu(x', y') = m_x^\alpha(x'), \quad \sum_{x' \in V} \mu(x', y') = m_y^\alpha(y'). \tag{4.4}$$

In this way, the computation of the Ollivier-Ricci curvature, in equation (4.2), requires solving an optimization and linear programming problem provided by equations (4.3) and (4.4), respectively. In our work, we set $\alpha = 0.5$. The choice of the parameter $\alpha \in [0, 1]$ directly influences the contributions of the nodes and edges to the computation of the curvature, see equation (4.1). Although Ollivier-Ricci discretization has numerous applications across fields (SANDHU; GEORGIOU; TANNENBAUM, 2016; FAROOQ et al., 2019; SANDHU et al., 2015b), the implementation can be time consuming when compared to

other discrete version of network curvatures. For this reason, we used and compared, alternatively, the combinatorial version of the Forman-Ricci curvature in our work, as discussed below.

4.2 WEIGHTED FORMAN-RICCI CURVATURE

We defined Forman-Ricci curvature for weighted and unweigted graphs. Both has its peculiarities in terms of numeric and combinatorial results, as well as the time processing. Here, we are going to define each one and show the main difference between them. Considering $G = (V, E)$ an undirected weighted graph, $x \in V$ and π_x the neighborhood of x (i.e., the adjacent nodes to x), whe define the weighted version of Forman-Ricci curvature as ((SREEJITH et al., 2016),(POURYAHYA; MATHEWS; TANNENBAUM, 2017)):

$$\begin{aligned} \mathrm{F}(x,y) = w_x + w_y - w_{xy}\Big(& \sum_{z\in\pi_x,\, z\neq y} \frac{w_x}{\sqrt{w_{xy}w_{xz}}} \\ & + \sum_{s\in\pi_y,\, s\neq x} \frac{w_y}{\sqrt{w_{xy}w_{sy}}}\Big), \end{aligned} \tag{4.5}$$

where $(x, y) \in E$, w_x, w_y and w_{xy} are the weights of vertices x, y and edge (x, y), respectively. The sums in 4.5 vary over all neighbouring nodes of x and y, except y and x, respectively.

In a particular case, when considered all nodes and edges with unitary weights (*i.e,* $w_{xy} = w_x = w_y = 1, \forall (x, y) \in E$) we have that

$$\begin{aligned} \mathrm{F}(x,y) &= 2 - \Big(\sum_{z\in\pi_x,\, z\neq y} 1 + \sum_{s\in\pi_y,\, s\neq x} 1 \Big) \\ &= 2 - [(\deg(x) - 1) + (\deg(y) - 1)], \end{aligned} \tag{4.6}$$

i.e.,

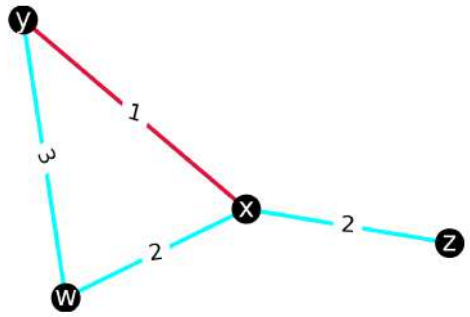
$$\mathrm{F}(x,y) = 4 - \deg(x) - \deg(y). \tag{4.7}$$

The formula in 4.7 allow us to compute curvatures from the basic graph structures, without the need of weights carried in the edges. The Formula in 4.5 expand this results by using the metric provided by edges.

As an example, considering the graph in Figure 6, it is easily seen that $F(x, y) = 2 - [\sqrt{2} + \sqrt{3}/3]$ by the formula in 4.5 and $F(x, y) = 4 - 3 - 2 = -1$ by 4.7, thus they might provide completely different results.

In order to distinguish both curvatures, we are going to call 4.5 as *Weighted Forman-Ricci curvature* and 4.7 as simply *Forman-Ricci Curvature.*

Figure 6 – Example of a simple weighted graph.



Source: self-provided.

4.3 AUGMENTED FORMAN-RICCI CURVATURE

The study of the topology of a continuous manifold and its geometric counterparts is not a new concept, (LOTT; VILLANI, 2009; BOCHNER, 1946; COLDING, 1997; TIAN; YAU, 1990; MATSUMOTO, 2002; NAJMAN; ROMON, 2017). However, discrete versions of several results in Differential Geometry is quite a contemporaneous approach (BOBENKO, 2008) and It's an active area in pure mathematics. Over the past years, the concept of Ricci curvature, which will be used in our work, was formalized for discrete structures by Forman (BOCHNER, 1946; FORMAN, 2003; SREEJITH et al., 2016; POURYAHYA; MATHEWS; TANNENBAUM, 2017). Later, it was also formalized by Olliver Ricci (OLLIVIER, 2007; POURYAHYA; MATHEWS; TANNENBAUM, 2017). The ideas developed by Forman, also proven to be useful for complex systems and Networks. The concept is explained as follows: For an unweighted graph, we can define Forman Curvature from a topological approach provided at (LEVY, 1999). For more details regarding discretization and definitions, see (HATCHER, 2005). To clarify this concept, an example is given in Figure 3 and Figure 4.

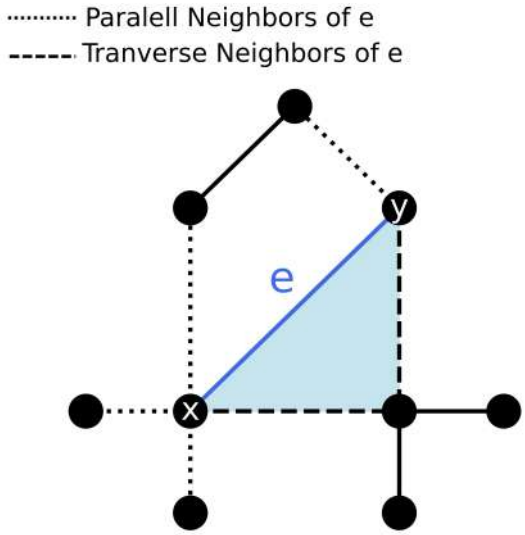
The curvature is calculated from the d-cells present in the graph. Considering a CW-complex (FORMAN, 2003; HATCHER, 2005) obtained from the graph and given two d-cells α_1 and α_2, we denote that α_1 is contained in the boundary of α_2 by $\alpha_1 < \alpha_2$. We say that α_1 and α_2 are *neighbors*, if at least one of the following conditions are fulfilled:

1. There is a $(d+1)$-cell β such that $\alpha_1 < \beta$ and $\alpha_2 < \beta$;
2. There is a $(d-1)$-cell γ such that $\gamma < \alpha_1$ and $\gamma < \alpha_2$.

We say that that α_1 and α_2 are *parallel neighbors* if conditions (1) and (2) is true but not simultaneously. In case of both (1) and (2) are true, α_1 and α_2 are said *transverse neighbors*.

In Figure 7, considering $d = 1$, the dashed edges are parallel neighbors of edge $e = (x, y)$ (blue edge), while the dotted edges are transverse neighbors. The black edges are not neighbors of e, once that neither conditions are satisfied. The d-th Forman

Figure 7 – Example of parallel and transverse neighbors for an undirected graph. The nodes x and y have 5 and 3 neighbors, respectively. Following the definiton, the edge e_{xy} has 4 paralell neighbors (dotted edges) and 2 tranverse neighbors (dashed edges).



Source: self-provided

Curvature for an unweighted d-cell α is given by

$$\begin{aligned} \mathrm{F}_d(\alpha) = \quad & \#\{(d+1)\text{-cells } \beta > \alpha\} \\ & +\#\{(d-1)\text{-cells } \gamma < \alpha\} \\ & -\#\{\text{parallel neighbors of } \alpha\}. \end{aligned} \tag{4.8}$$

In our special case, for $d = 1$, the formula reduces to

$$\begin{aligned} \mathrm{F}(e) = \quad & \#\{2\text{-cells } \beta > e\} + 2 \\ & -\#\{\text{parallel neighbors of } e\}. \end{aligned} \tag{4.9}$$

Applying (4.9) for e in Figure 7, we have: one triangle (2-cell) that contains e, as well as 4 parallel neighbors. Therefore, $\mathrm{F}(e) = 1 + 2 - 4 = -1$. Once defined the Forman-Ricci curvature for edges, we can extend the concept for the nodes as follows:

$$\mathrm{F}(x) = \frac{1}{deg(x)} \sum_{z \in \pi_x} F(e_{xz}), \tag{4.10}$$

where e_{xz} is the edge that connects x to z and $\pi_x = \mathrm{neigh}(x)$. We can have a set theory analogy for the definition above. This analogy will prove to be useful during the numerical

implementation of the Forman-Ricci curvature. The following results provide a generalized formulation for the Forman-Ricci curvature for higher-order CW-complexes, and is a original result from our work.

4.4 APPLICATION TO RANDOM GRAPHS

In our thesis, we developed analytic expressions for the mean Augmented Forman-Ricci curvature for random graphs. Given a d-cell, α, we can write it as $\alpha = \{v_1, v_2, \cdots, v_{d+1}\}$, for $v_i \in V$, $i = 1, 2, \cdots, d+1$. Now, we can give an alternative formulation from equation 4.8. The explanation can be given from a set theory perspective: The number of $d+1$-cells containing α can be found by counting the number of all vertices in V contained simultaneously in all neighbourhood of each $v_i \in \alpha$. The number of cells contained in the boundary of α are always $d+1$. For each $v_i \in \alpha$, the number of parallel cells can be found by taking all $v \in V$ such that v belongs to the neighborhood of all $v_j \in \alpha$, except the neighborhood of v_i. Summarizing all this information, we can rewrite 4.8 as

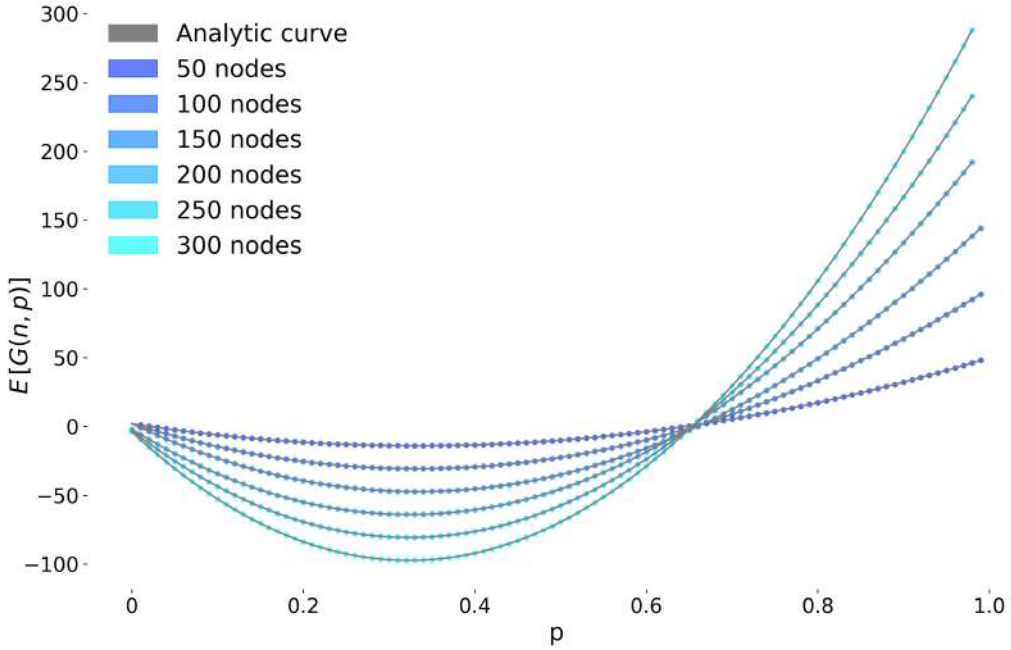
$$\begin{aligned} \mathrm{F}_d(\alpha) = & \quad \#\{v \in V \,|\, v \in \bigcap_{'i=1}^{d+1} \pi_{v_i}\} \\ & +(d+1) \\ & -\#\bigcup_{i=1}^{d+1}\{v \in V | v \in (\bigcap_{j=1,\, j\neq i}^{d} \pi_{v_j}) \setminus \pi_{v_i}\}, \end{aligned} \tag{4.11}$$

where $\pi_i = \mathrm{neigh}(v_i)$. Simplifying (4.11) with $d = 1$, we finally have the alternative Forman-Ricci formula for $e = \{x, y\}$, given by

$$\begin{aligned} \mathrm{F}_1(e) = & \quad \#\{v \in V \,|\, v \in \pi_x \cap \pi_y\}\} \\ & +2 \\ & -\#(\{v \in V \,|\, v \in \pi_x \setminus \pi_y\} \bigcup \{v \in V \,|\, v \in \pi_y \setminus \pi_x\}) \end{aligned} \tag{4.12}$$

This new version motivated us to find an analytic expression for the mean of the Augmented Forman-Curvature for random graphs. Let $e = \{x, y\}$ be and edge in a random graph. From the alternative equation in 4.12, we have that the number of triangles containing an edge can be counted by simply counting the number of nodes that belongs to the intersection of the neighbors of x and y. Hence, the expected number of triangles containing e is $(n-2)p^2$, once that for each edge different from e has probability p of appearing in the graph and we also have $(n-2)$ nodes in V for the choice. The conditions 1. and 2. of neighbouring cells ensure as that the parallel edges to e are the neighbors that does not satisfy the first condition. Thus, the number of edges parallel to e can also be explained as the number of neighbors of x (y), except y (x) that are not included in triangles containing e. Applying this result with the Erdos-Renyi theory presented in chapter 2 we have that the expected number of edges parallel to e is $2(n-2)p(1-p) = 2[(n-2)p - (n-2)p^2]$.

Figure 8 – Forman-Ricci Curvature average of 600 Erdos-Renyi graph realizations for diferent amount of nodes.



Source: self-provided.

Thus, we finally have from (4.12) the expected Forman-Ricci curvature of an Erdos-Renyi graph $G(n,p)$, namely

$$\mathrm{E}[F(G(n,p))] = 3(n-2)p^2 + 2 - 2(n-2)p, \tag{4.13}$$

for $p > 0$. The resultant quadratic equation in (4.13) has roots

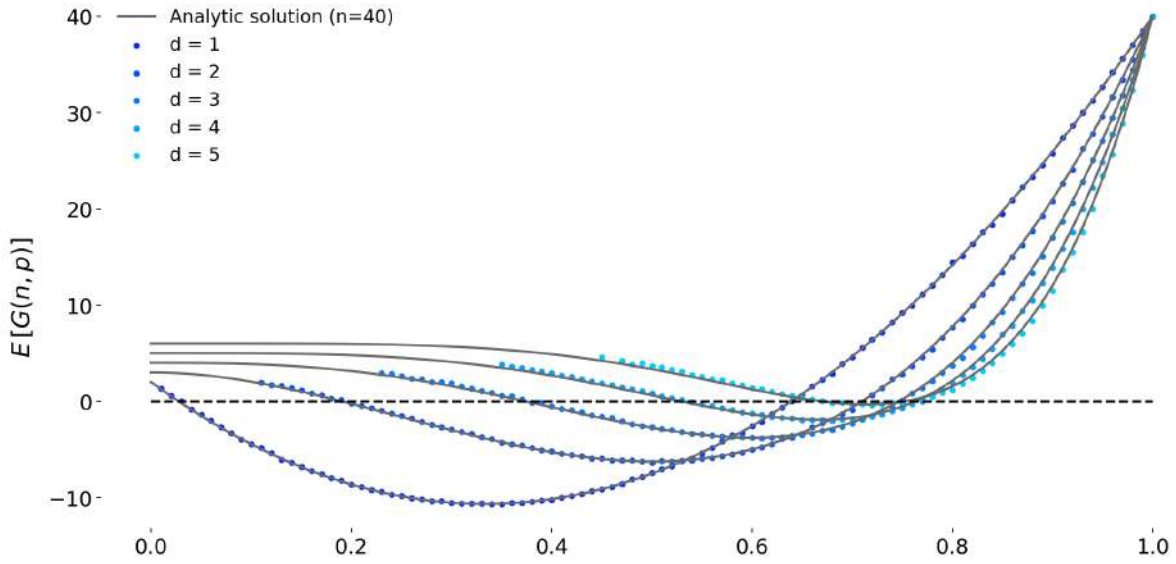
$$p^+, p^- = \frac{1}{3}\left[1 \pm \sqrt{\frac{(n-8)}{(n-2)}}\right], \tag{4.14}$$

for $n > 8$, and $p = 1/3$, for $n = 8$. For a large graph (when $n \longrightarrow \infty$) we have that the roots in 4.14 are close to 0 and 2/3.

Figure 8 shows the mean behavior of 600 realizations for the Forman-Ricci curvature of Erdos-Renyi graph. Furthermore, the roots of (4.13) signalize the geometric change from the curvature sign, which occurs, generally, for high values of p. The formulation in 4.13 can be generalized by following the previous argument for 1-cells as follows:

Let $\alpha = \{v_1, \ldots, v_{d+1}\}$ a d-cell. From the first parcel of equation 4.11, we have that the probability of a given vertex $v \in V$ be in $\bigcap_{i=1}^{d+1} \pi_{v_i}$ is p^{d+1}. We also have $(n-(d+1))$ possibilities for selecting v. Thus, the expected number of $(d+1)$-cells containing α is $(n-(d+1))p^{d+1}$. From the third parcel of equation 4.11, we have that, for each $i \in \{1, 2, \ldots, d+1\}$, the probability of a given $v \in V$ be in $(\bigcap_{j=1}^{d} \pi_{v_j}) \setminus \pi_{v_i}$ is $p^d(1-p)$, for every

Figure 9 – Comparison between the analytic expected Augmented Forman-Ricci curvatures for d-cells ($d = 1, 2, 3, 4, 5$) and the mean of 100 realizations in a Erdos-Renyi graph with $n = 40$ nodes.



Source: self-provided.

$j \in 1, ..., d+1, i \neq j$, which gives us $(n-(d+1))$ different choices, for each fixed node v_i in α. Hence, the expected number of parallel cells to α is $(d+1)(n-(d+1))p^d(1-p)$. Gathering all this information and applying to 4.11, we finally have

$$\mathrm{E}[F_d(G(n,p))] = (d+2)(n-(d+1))p^{(d+1)} + (d+1) - (d+1)(n-(d+1))p^d, \quad (4.15)$$

for $d = 1, 2, 3, \ldots$.

By setting $d = 1$, in 4.15 we bring back the formula in 4.13. Normalizing 4.15 by n, the number of nodes, we have that, as $n \longrightarrow \infty$, the resultant polynomial is $p^d[(d+2)p-(d+1)]$, which clearly has roots 0 and $(d+1)/(d+2)$. In Figure 9 we can see the comparison between the analytic formula obtained in 4.15 and the mean of 100 stochastic realizations, for different higher-order cells.

In the next chapter, we are going to introduce a topological approach for this data analysis, which can be compared to the geometric one, already made in this chapter.

5 TOPOLOGICAL AND GEOMETRIC APPLICATIONS TO EPIDEMIC DATA

In this chapter we are going to develop the main concepts used in this work. First, we are going to provide our definition of graphs generated from epidemic time-series, which we called *epidemic networks*, and then, all the steps needed for its construction and data visualisation.

5.1 EPIDEMIC NETWORKS

In fields where network science is applicable (BARABÁSI et al., 2016), and particularly in network neuroscience (FORNITO; ZALESKY; BULLMORE, 2016), weighted networks are build based on correlations between nodes of a network, or any other similarity measure. Similarly, in studies of financial networks (ONNELA; KASKI; KERTSZ, 2004; SANDHU; GEORGIOU; TANNENBAUM, 2016), the resultant weighted network is built from Pearson correlations between financial time-series (ONNELA; KASKI; KERTSZ, 2004). In contrast, in most of the network approaches to epidemiology (MILLER; KISS, 2014; TAO; ZHONGQIAN; BINGHONG, 2006), the structure of the contagion network is considered, i.e., the focus is studying an epidemic in a network (TAO; ZHONGQIAN; BINGHONG, 2006).

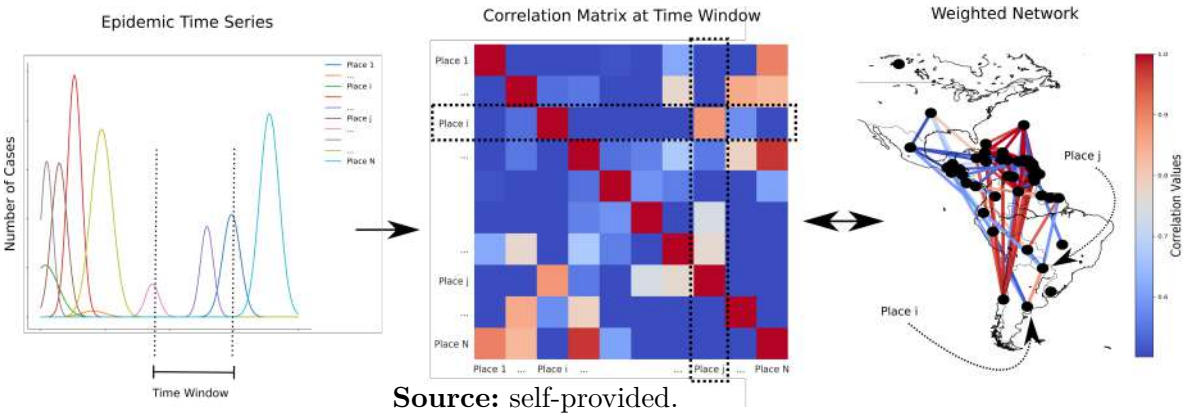
Here, following the network science approaches in neuroscience and economy, we introduce an *epidemic network* as a weighted network, where nodes are locations and the weights are defined by the Pearson correlation coefficient between two epidemic time-series. As far as we know, this way of defining an epidemic network from correlations was not yet introduced in the literature on network epidemiology. Therefore, not much is known regarding the relationship between our epidemic network approach and classic network epidemiology. This question deserves further investigation and can draw inspiration from other fields confronted with similar issues, in particular neuroscience (WANG et al., 2015; TEWARIE et al., 2020),

In this work, we create an epidemic network consisting of nodes and edges, based on the reported epidemic time-series. We define each spatial domain of the epidemic as the node of a network. The links between two locations are based on the Pearson correlation coefficient between their epidemic time-series.

5.2 NETWORK FROM EPIDEMIC TIME-SERIES DATA

Network-based approaches to epidemiology are usually performed by associating vertices to individuals and the edges are the infection probability (NEWMAN, 2002; BALL et al., 2019; BRITTON; JUHER; SALDAÑA, 2016; PASTOR-SATORRAS et al., 2015). The starting point of our approach for creating an epidemic network differs for the standard one since it is based in Pearson correlation coefficient from the time-series data. This approach is

Figure 10 – An explanation for the time varying process on epidemic data.



Source: self-provided.

inspired by network analysis in other fields, such as neuroscience (FORNITO; ZALESKY; BULLMORE,) or in financial networks (KIM; KIM; HA, 2007).

In this first step, we explain the process of conversion of time-series into evolving graphs as follows:

1. A moving window on data is pre-selected regarding the way of time measuring (days, weeks or months);

2. The resulting data will provide the time-series on each place to be analysed;

3. The place names are provided to the nodes, while the connection between epidemic time-series is given by edges between nodes via Pearson correlation coefficient, given by (KENNEY; KEEPING, 1957; SNEDECOR; COCHRAN, 1937; EDWARDS, 1984)

$$\rho = \frac{\text{cov}(X, Y)}{\text{var}(X)\text{var}(Y)}, \tag{5.1}$$

where $\text{cov}(X, Y)$ is the *co-variance* between two random variables, X and Y, while $\text{var}(X)$ and $\text{var}(Y)$ are the *variance* of the time-series X and Y, respectively.

4. The process is repeated for the next time period, until the whole analysed data be fully covered.

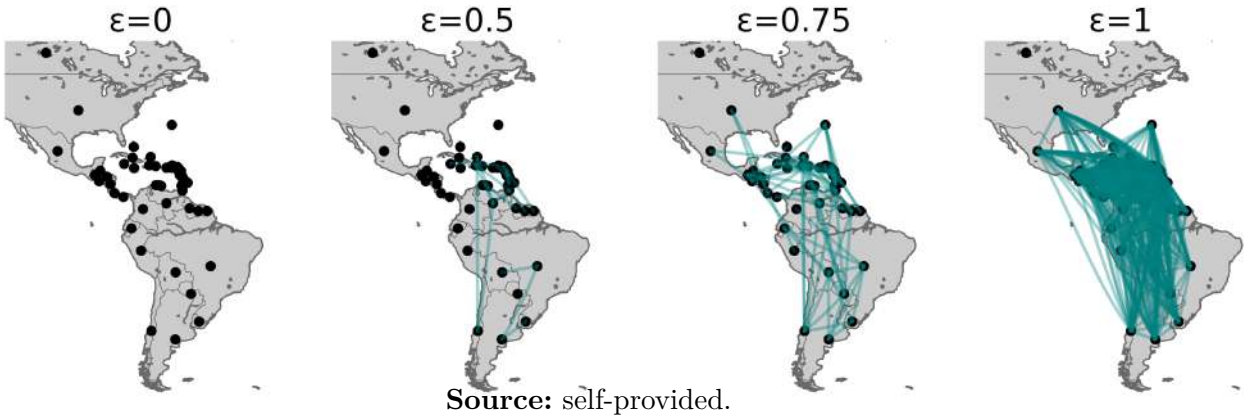
This way, each moving window provides a graph. For better understanding, see Figure 10.

Once we set the edges weights as the values found by applying 5.1, we are ready to perform all the topological and geometrical analysis in the epidemic data, as we will explain further.

5.3 FILTRATION

The network created using the procedure above often results into a dense graph, which may have several spurious links.

Figure 11 – Illustration of an epidemic network, for different filtration values.



Source: self-provided.

The reader may remember that we chose the links of the network according to the Pearson's correlation coefficient between two locations in descending order, which means that we include the strongest links in the network first, until the network reaches the same number of connected components of the original graph. The strategy for thresholding is based on the concept of filtration, which, for instance, (ZOMORODIAN, 2012) was used in neuroscience for classifying disease and control groups (SANTOS et al., 2019; GRACIA-TABUENCA et al., 2020).

More precisely, we define a graph $G = (V, E)$ as a structure constructed from the set of nodes (or vertexes) V and edges (links between nodes) E.

Formally, once V is the set of nodes of G, we define a *ϵ-neighbourhood* (ZOMORODIAN, 2005) on V as $G_\epsilon = (V, E_\epsilon)$, where

$$E_\epsilon = \{e = (x, y) \in E, x \neq y \,;\, d(x, y) \leq \epsilon\}, \tag{5.2}$$

where $d(x, y)$ is the distance between nodes x and y. The added value of this definition is the possibility to visualize an evolving graph as a function of a parameter ϵ, i.e., we can track the evolving from an empty graph, when $\epsilon = 0$, until the a dense network, when $\epsilon = 1$. The idea of filtration can be better understood via the example in Figure11.

The main idea is to start with an empty graph, and gradually add the *strongest* links to the evolving graph, i.e., we first add the edges with higher correlation coefficient. Therefore, we denoted the filtration E_ϵ by

$$E_\epsilon = \{e \in E;\ w_e \geq 1 - \epsilon\}. \tag{5.3}$$

Clearly, ϵ runs over the interval $[0, 2]$, and than $G_\epsilon \subset G$. In order to ignore redundant information for the time-varying analysis, we are going to find the critical percolation value of ϵ_c such that the graph still keeping the connections which is relevant in the skeleton structure. This threshold value is defined as follows:

$$\epsilon_c = \inf\{\epsilon \in [0, 2]\ ;\ |G_\epsilon| = |G|\}, \tag{5.4}$$

where $|G|$ denotes the number of connected components of $G = G_1$ (ERDOS; P., 1959; ERDŐS; RÉNYI, 1960). The idea is to keep the graph structure as the same of the crossing the *Betti numbers* of G ((LEVY, 1999)). An algorithm to obtain the approximate ϵ_c from small steps $\Delta\epsilon$ is given in bellow. In our studies, we used $\Delta\epsilon = 10^{-5}$.

Algorithm 1 Graph Threshold Algorithm

Result: ϵ_c
Input: G_1, $\Delta\epsilon$;
$\epsilon_c = 0$
while $|G_\epsilon| \neq |G_1|$ **do**
  $\epsilon_c := \epsilon_c + \Delta\epsilon$
end

This approach allow us to avoid the inclusion of spurious links between epidemic sites in a natural manner. Once the graph is created, a filtration is selected at the vicinity of the giant component transition. As is discussed in theoretical models in epidemic networks, an epidemic outbreak happens at the critical probability for the emergence of a giant component transition (MOORE; NEWMAN, 2000; NEWMAN, 2002), so that the detachment of connected components is imminent, i.e., we are seeking for the smallest value such it leads to the maximum amount of connected components graph. In order to start our tests in a data-driven approach, we started the topological and geometric applications to synthetic data, whose models are described in the next chapter.

6 ANALYTIC EPIDEMIC MODELS

In this chapter, we dedicate the sections to construct synthetic data from analytic models. As one of the first works that emerged on modelling COVID-19 data in Wuhan (ZIFF; ZIFF, 2020), we decided to chose the fractal model to generate our time-series for COVID data. For Dengue disease, we decided to use a recent work of modelling Dengue cases by using a modification of SIR model with interaction between sites used in Rio de Janeiro, Brazil (STOLERMAN; COOMBS; BOATTO, 2015).

6.1 FRACTAL MODEL

As a proof of concept, we decided to first investigate whether the curvatures are able to indicate the risk of a pandemic in a toy-model. For this, we simulated an epidemic network obtained from synthetic epidemic time-series. We will first build this simple toy-model heuristically and in a second step move towards the analysis of real COVID-19 data.

A simple way to access the number of cases in an epidemic network is to use the fractal growth hypothesis, as observed in (ZIFF; ZIFF, 2020), where the daily number of cases $n(t)$ in an epidemic follows a power-law distribution with an exponential cutoff:

$$n(t) = Kt^x \exp(-t/t_0), \tag{6.1}$$

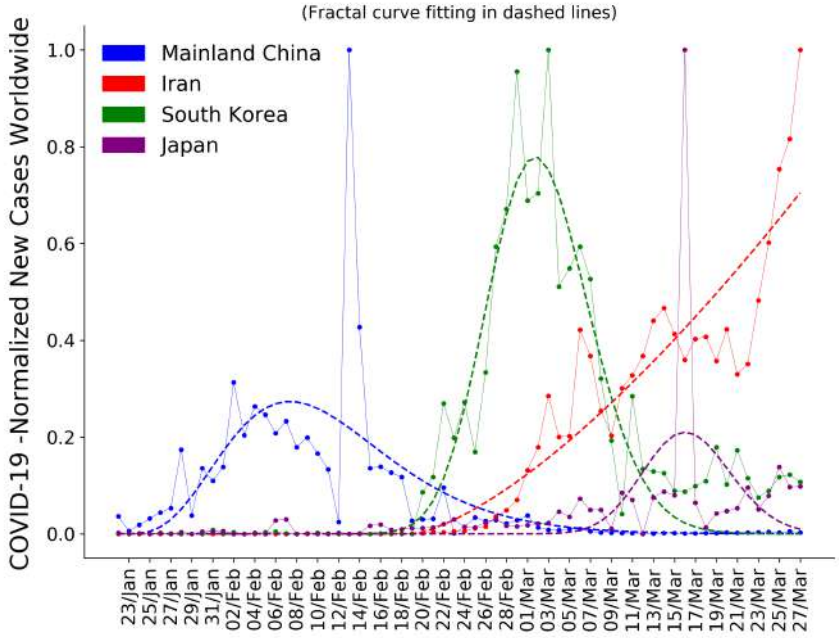
where, K, x and t_0 are fitting parameters. In Figure 12, we show examples of the fit between (6.1) and the number of reported COVID-19 new cases for four countries, namely, China, Iran, South Korea and Japan. This fit suggests that (6.1) paves a simple way for building a toy-model for epidemic time series. We stress that our aim here is not to find whether the best fit for the pandemic is exponential or power law, which was already addressed in (ZIFF; ZIFF, 2020; WODARZ; KOMAROVA, 2020), but to build a simple toy-model that allows us to test our hypothesis relating Forman-Ricci curvatures to epidemic networks.

Inspired by this equation, we can suggest a phenomenological toy-model for generating epidemic time series with noise that can capture the growth of an epidemic network. We assume that in each node i of the epidemic network, the daily number of cases follows a fractal epidemic growth with Gaussian noise $w_i(t)$ and a time delay d_i in relation to the epicenter:

$$n_i(t) = \begin{cases} w_i(t) & \text{if } t \leq d_i \\ K_i(t-d_i)^{x_i} \exp\left(-\frac{(t-d_i)}{t_0^i}\right) + w_i(t) & \text{if } t > d_i \end{cases}. \tag{6.2}$$

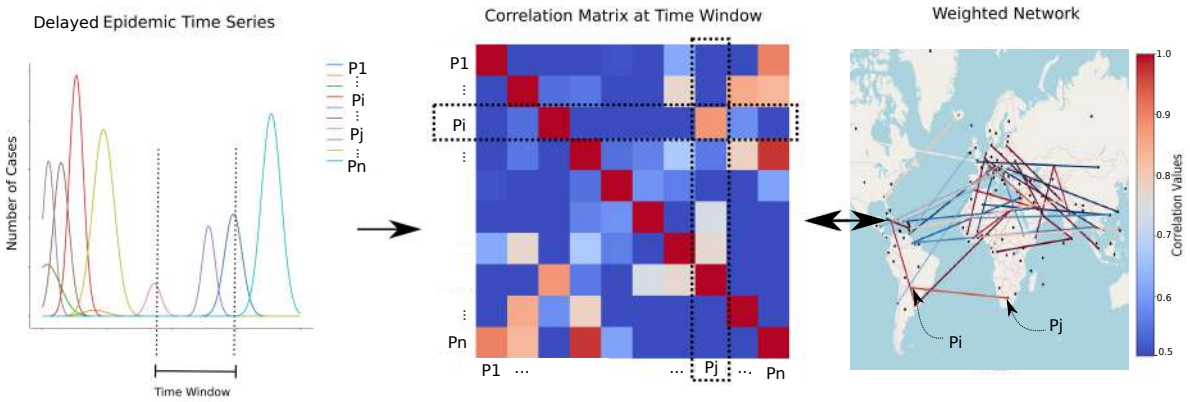
We now show that the Forman-Ricci curvature suffices to detect fragility and risk for the simulated epidemic network. The starting point for creating a fractal epidemic

Figure 12 – COVID-19 per country. Illustration of the number of cases and fitting through fractal growth (dashed lines), Eq. (6.1), for a representative number of countries.



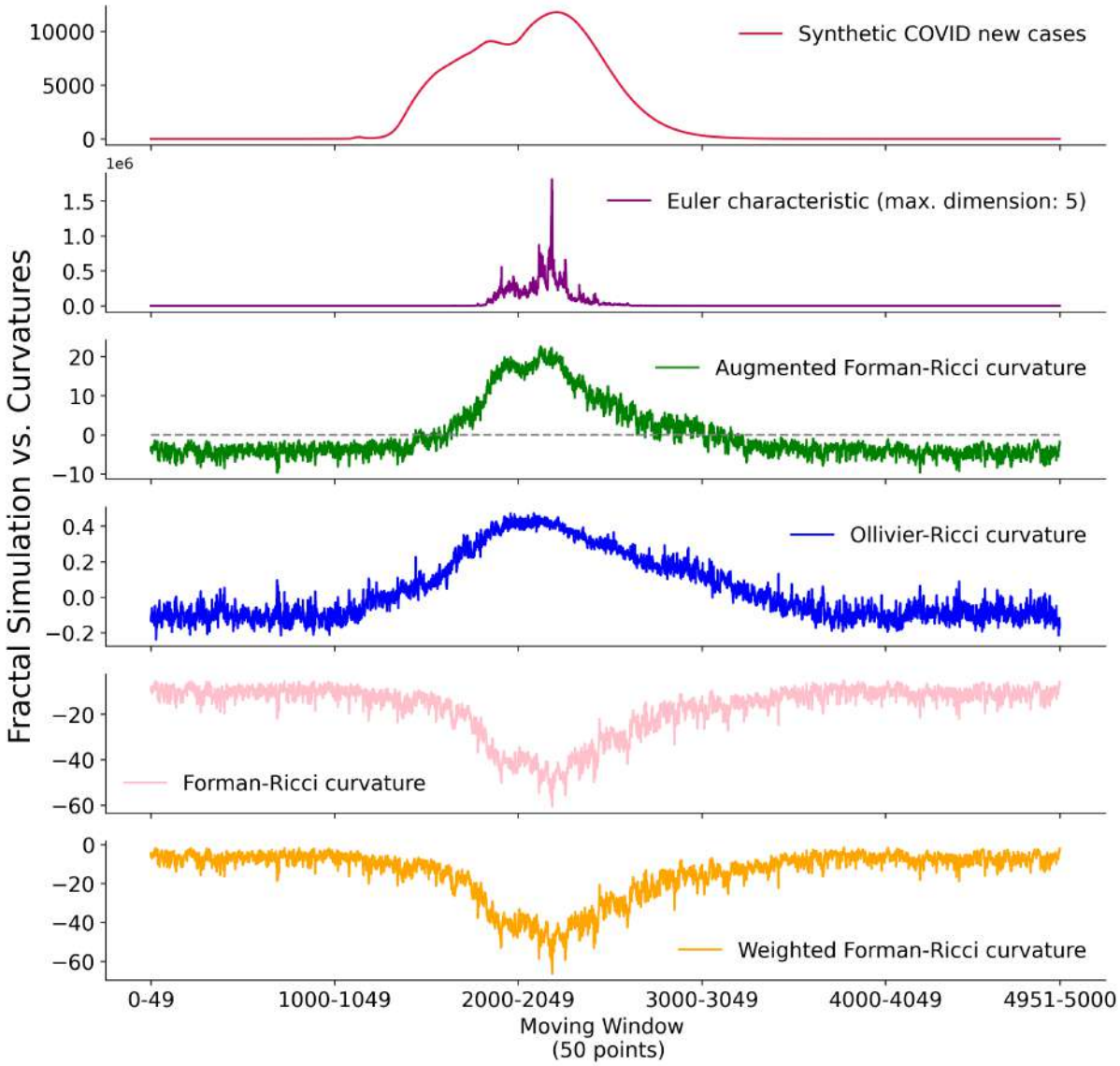
Source: self-provided.

Figure 13 – Illustration of the creation of epidemic networks based on the correlations between epidemic time series across spatial domains for a given time window. This approach allow us to infer network signatures for epidemic outbreaks without relying on parameter estimation of classic stochastic epidemic approaches.



Source: self-provided.

Figure 14 – Illustration of the toy-model epidemic curves for new cases (moving average), according to (6.2) and its respective curvatures, with white noise parameter $\sigma = 0.01$.



Source: self-provided.

network is based on simulating epidemic time series with delays from (6.2). In a second step, we define the weights of the epidemic network through the Pearson correlation coefficient between time series $n_i(t)$ and $n_j(t)$. The temporal epidemic network is computed for a given time window, and the process is repeated for the next time window, thus obtaining an evolving network. This approach is inspired by network analysis in other fields, such as neuroscience (FORNITO; ZALESKY; BULLMORE,) or finance (KIM; KIM; HA, 2007). We illustrate the delayed epidemic time series, its Pearson correlation matrix and its corresponding network for a given time point in Figure 13, resulting in a time evolving network.

The third step is to infer the fragility of the time evolving epidemic network by tracking geometric changes in this network as a function of time. More specifically, we observe the mean changes in the discrete version of the Forman-Ricci curvature (SREEJITH et al., 2016) for a selected moving window for each location affected by the epidemic and use the network curvature as a indicator for its fragility and risk. Thus, we assume that the application to epidemic time series follows an analogous behaviour to the one observed for stock markets in (SANDHU; GEORGIOU; TANNENBAUM, 2016).

As a proof of concept, we then investigate a simulated time series with delays in (6.2). We generated 50 time series with parameters K_i, x_i, d_i, and t_0^i randomly chosen in the interval $K_i \in [0, 20]$, $x_i \in [0, 5]$, $d_i \in [10, 21]$, and $t_0^i \in [0, 1]$. We also included a small white noise with zero mean and variance of $\sigma = 0.01$.

Figure 14 shows that the epidemic curve generated from our toy-model in Eq. (6.2) is compatible with an epidemic outbreak and contrasts the simulated epidemic curve with its curvatures. We observe that the curvature is constant before the starting of the simulated epidemic, grows during its progression and reaches its maximum during the peak of the simulated outbreak. After the end of the simulated epidemic, the curvature comes back to its initial level. We emphasize that the inclusion of white noise $w_i(t)$ in our model was very important to destroy spurious deterministic correlations that appear at the end of the outbreak.

6.2 SIR MODEL WITH BIRTH AND DEATH

Repeating the steps of the previous section, we will apply our techniques to a simple and classic infection disease model, named SIR. In addiction, we are considering the birth and death parameters in our modelling.

In this part of the research, we based our approach on a SIR social dynamic over networks applied to Dengue disease with variations regarding population dynamics. Basically, for simplicity, we decided to introduce birth and death on the model in order to include a natural noise and periodic outbreaks, which makes the epidemic time series

closer to the data collection. More specifically, we are providing the interaction between human population as follows (STOLERMAN; COOMBS; BOATTO, 2015):

The total initial population (N) is divided by sites/places ($P_1, P_2, \cdots, P_m$). For each $j \in \{1, \cdots, m\}$ we have susceptible (S_j), infected (I_j) and recovered/removed (R_j) as local population.

The contagion dynamics between the sites is ruled by a probability matrix, $\Phi = (\phi_{ij})_{m\times m}$. Each class (susceptible, infected or recovered) has a *rate of death* (δ_{S_j},δ_{I_j} and δ_{R_j}), and only new susceptible arrives with *rate of birth* α_i, for all $j \in \{1, \cdots, m\}$. From this dynamics, we can express a system of ordinary differential equations from rate equation techniques provided at (BAEZ; BIAMONTE, 2012),

$$\begin{cases} \frac{dS_i}{dt} &= \alpha_i - \sum_{j,k=1}^{m} \beta_j \phi_{ij}\phi_{kj} S_i \frac{I_k}{N_j^p} - \delta_{S_i} S_i \\ \\ \frac{dI_i}{dt} &= \sum_{j,k=1}^{m} \beta_j \phi_{ij}\phi_{kj} S_i \frac{I_k}{N_j^p} - (\gamma_i + \delta_{I_i}) I_i \\ \\ \frac{dR_i}{dt} &= \gamma_i I_i - \delta_{R_i} R_i \end{cases} , \tag{6.3}$$

for all $j, k \in \{1, \cdots, m\}$, and satisfying the constant initial condition equation

$$S_j(0) + I_j(0) + R_j(0) = N_j \ , \tag{6.4}$$

for $N_j \in \mathrm{N}^+$, where $\sum_{j=1}^{m} N_j = N$, $N_j^p = \sum_{k=1}^{m} \phi_{kj} N_k$, and ϕ_{ij} is an element of the *Flux Matrix* satisfying

$$\sum_{k=1}^{m} \phi_{jk} = 1, \tag{6.5}$$

for all $j \in \{1, \cdots m\}$. In our approach, we are considering

$$\phi_{ij}(\rho) = \begin{cases} \rho, & i \neq j \\ 1 - (m-1)\rho, & i = j \end{cases} , \tag{6.6}$$

for some $\rho \in [0, 1/(m-1)]$. We generated the epidemic time series from a Monte-Carlo simulation as chemical reactions (ERBAN; CHAPMAN; MAINI, 2007) as follows:

1. Select the time: t;
2. Take $r \in (0, 1)$ a random variable uniformly distributed;
3. Update the propensities, i.e., for all $i \in 1, \ldots, m$:
   - $a_i = \sum_{j,k=1}^{m} \beta_j \phi_{ij}\phi_{kj} S_i(t) \frac{I_k(t)}{N_j^p}$;
   - $b_i = \gamma_i I_i(t)$;

- $c_i = S_i(t)$;
- $d_i = \delta_{S_j} S_j(t)$;
- $e_i = \delta_{I_j} I_j(t)$;
- $f_i = \delta_{R_j} R_j(t)$;
- $r_0 = \sum_{i=1}^{m} a_i + b_i + c_i + d_i + e_i + f_i$;

4. Compute the extreme points of intervals for the next reaction:

$$r_i = \begin{cases} \frac{1}{r_0}\sum_{j=1}^{i} a_j, & 1 \le i \le m \\ \frac{1}{r_0}(\sum_{j=1}^{m} a_j + \sum_{j=1}^{i-m} b_j), & m+1 \le i \le 2m \\ \frac{1}{r_0}(\sum_{j=1}^{m}(a_j + b_j) + \sum_{j=1}^{i-2m} c_j), & 2m+1 \le i \le 3m \\ \frac{1}{r_0}(\sum_{j=1}^{m}(a_j + b_j + c_j) + \sum_{j=1}^{i-3m} d_j), & 3m+1 \le i \le 4m \\ \frac{1}{r_0}(\sum_{j=1}^{m}(a_j + b_j + c_j + d_j) + \sum_{j=1}^{i-4m} e_j), & 4m+1 \le i \le 5m \\ \frac{1}{r_0}(\sum_{j=1}^{m(} a_j + b_j + c_j + d_j + e_j) + \sum_{j=1}^{i-5m} f_j), & 5m+1 \le i \le 6m \end{cases}$$

5. Update time, $t := t + 1$;

6. Update population:

   - If $0 \le r < r_1$, then $S_1(t+1) = S_1(t) - 1$ and $I_1(t+1) = I_1(t) + 1$;

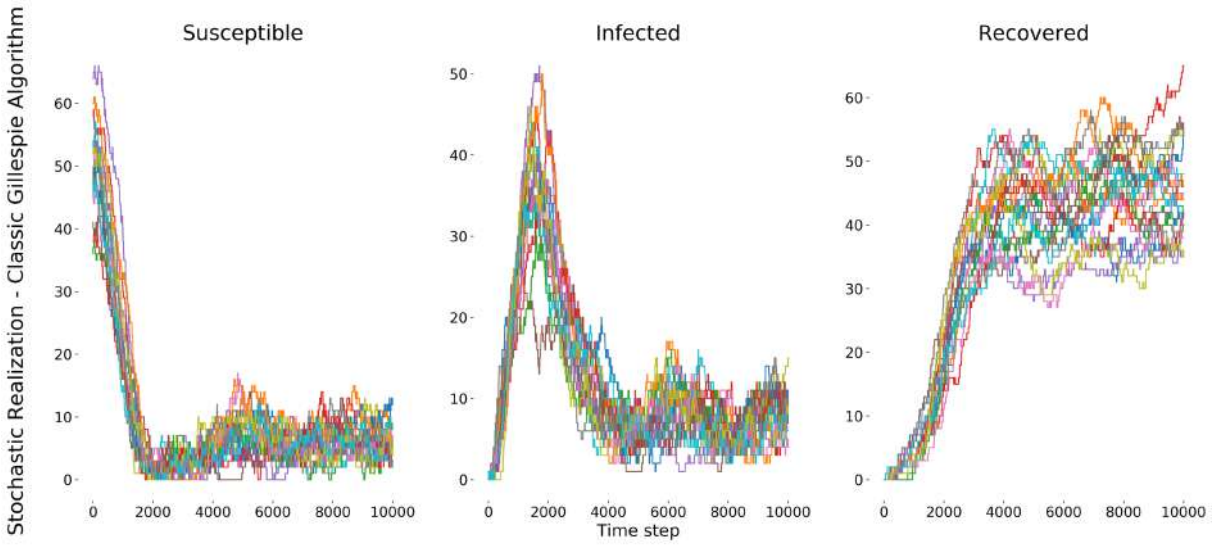
   Considering the other possible cases, when $r_i \le r < r_{i+1}$:

   - If $i \in \{1, \cdots, m-1\}$, then $S_{i+1}(t+1) = S_{i+1}(t) - 1$ and $I_{i+1}(t+1) = I_{i+1}(t) + 1$;
   - If $i \in \{m, \cdots, 2m-1\}$, then $I_{i+1-m}(t+1) = I_{i+1-m}(t) - 1$ and $R_{i+1-m}(t+1) = R_{i+1-m}(t) + 1$;
   - If $i \in \{2m, \cdots, 3m-1\}$, then $S_{i+1-2m}(t+1) = S_{i+1-2m}(t) + 1$;
   - If $i \in \{3m, \cdots, 4m-1\}$, then $S_{i+1-3m}(t+1) = S_{i+1-3m}(t) - 1$;
   - If $i \in \{4m, \cdots, 5m-1\}$, then $I_{i+1-4m}(t+1) = I_{i+1-4m}(t) - 1$;
   - If $i \in \{5m, \cdots, 6m-1\}$, then $R_{i+1-5m}(t+1) = R_{i+1-5m}(t) - 1$;

7. Return to 1. and repeat the algorithm for the next time step.

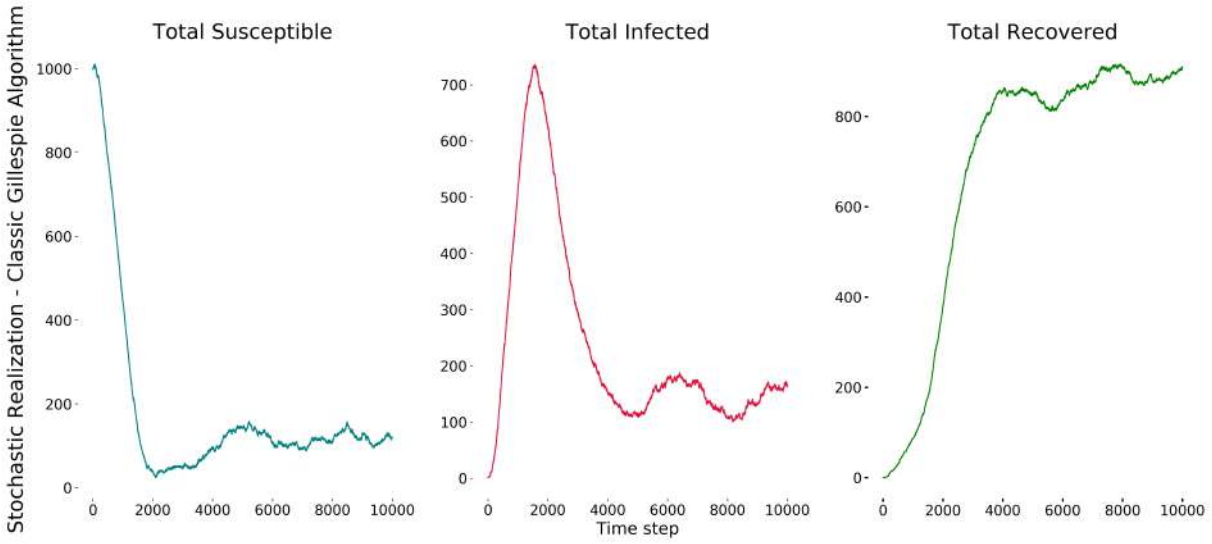
In our study, we computed the algorithm in order to obtain the infected time series for a population of size $N = 2000$, with one infected at time $t = 0$. The population were uniformly distributed in 20 sites ($m = 20$) and parameters $\beta_i = 1$, $\gamma_i = 0.1$, $\alpha_i = 1$, $\delta_{S_i} = \delta_{I_i} = \delta_{R_i} = 1/60$, for all $i \in \{1, \cdots, m\}$, and a total of 10000 steps. The next site infection parameter was $\rho = 1/m = 0.05$.The resulting simulation is provided at Figure 15 and Figure 16. The result of curvature approach over this simulated epidemic cases is provided at Figure 17. As observed, the Forman-Ricci curvature and Euler characteristics stay elevated at the vicinity of the epidemic periods, and the relevant manifestation occurs before or during the epidemic peaks, for different size of epidemics.

Figure 15 – Stochastic simulation of SIR epidemic model with birth and death. Each color represent a site (20 sites in total).



Source: self-provided.

Figure 16 – Total population from simulation in Figure 15.



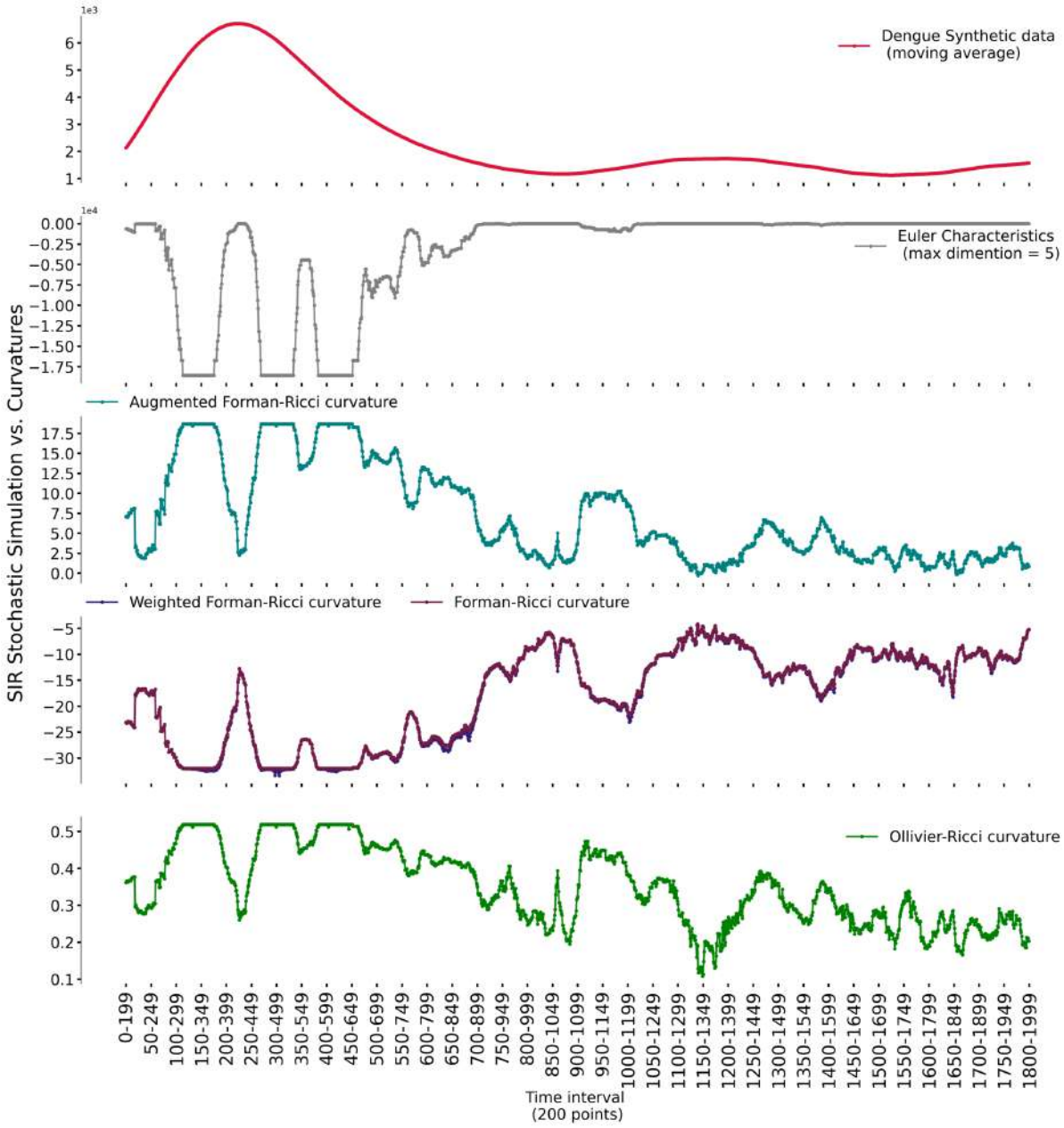
Source: self-provided.

Overall, it is evident the presence of higher amplitude values of the curvatures while approaching to epidemic periods, regardless the geometric or topological approach. This pattern is repeated in three scales of epidemic intensity.

All curvatures are high correlated and bring analogous information from the epidemic network. Augmented Forman-Ricci and Ollivier Ricci curvatures presented similar behaviours and are different only in the scale of values. Forman-Ricci curvatures, both weighted and unweighted versions, were almost identical and also provided good results, once compared to the other Ricci discretizations.

Visually, Euler characteristics provided a better marker for epidemic states, once that

Figure 17 – Comparison between Synthetic data of Dengue Disease and curvatures.



Source: self-provided.

the noise is minimal compared to the geometric approach.

7 RESULTS

Having illustrated that we can use the Ricci curvatures and Euler characteristics as estimators for the risk of reaching a pandemic state in a toy-model for epidemic networks, we are now ready to test whether the curvatures are reliable network fragility metrics for real data. In this chapter we are going to summarize the main information of our results for COVID-19 pandemic and Dengue disease.

7.1 TOPOLOGICAL AND GEOMETRIC APPROACH TO COVID-19

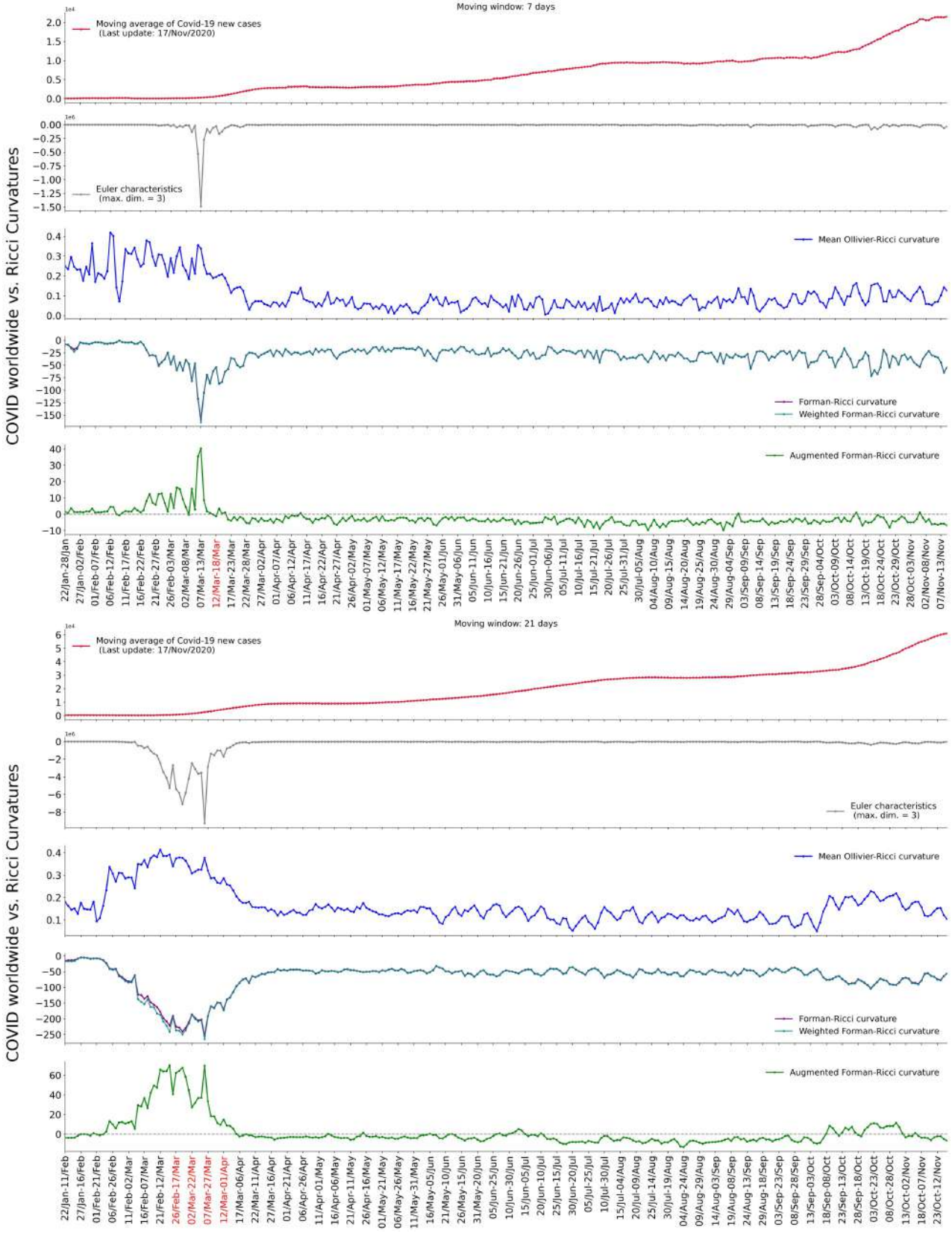
In Figure 18 we illustrate the epidemic curve and the curvatures for the COVID-19 database (ORGANIZATION et al., 2020). As in the simulated data, the curvature was stable at the beginning and grows over time, signaling increasing risk and fragility of the epidemic network. Remarkably, we observe that the curvature of the epidemic network might give early warning signs for the emergence of the pandemics, as the curvature starts to increase weeks before the exponential growth in number of cases is observed and the WHO declares COVID-19 as a pandemic (e.g., see Figure 18, in red). Figure 19 provides an additional geographical illustration of the distribution of the Augmented Forman-Ricci curvature and Ollivier-Ricci curvature across countries for three time windows.

We also tested the sensitiveness of our approach to different threshold values. We find that the Augmented Forman-Ricci curvature is able to identify the risk of an epidemic in both synthetic and real data for threshold values $\epsilon \in \{0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75\}$, apart from $\epsilon = 0$, which leads to an empty graph, and apart from $\epsilon = 2$, which leads to a fully connected graph and thus constant Augmented Forman-Ricci curvature as can be seen in Figure 20.

We illustrate the sensitiveness of the Ollivier-Ricci curvature to COVID-19 data for $\alpha \in \{0, 0.25, 0.5, 0.75\}$, apart from $\alpha = 1$, which considers the major weight influence from the nodes. The result can be seen in figure Figure 21.

Applications of discrete curvature to complex systems are rather new (FAROOQ et al., 2019; SANDHU; GEORGIOU; TANNENBAUM, 2016; SANDHU et al., 2015a). Therefore, it is relevant to contrast our findings with classic network metrics on COVID-19 data, both for daily new reported cases and for daily deaths. We focus our comparison on the following nine metrics: the degree assortativity coefficient (NEWMAN, 2003; FOSTER et al., 2010), the percolation threshold, modularity (CLAUSET; NEWMAN; MOORE, 2004; YANG, 2013), the mean diameter, the average clustering coefficient (SARAMÄKI et al., 2007; KAISER, 2008), communicability, betweenness centrality, the first non-zero eigenvalue of the Laplacian and the Perron-Frobenius eigenvalue (YANG, 2013). We also performed the dynamic network filtration provided by equation (5.3) before computing each metric with time window of

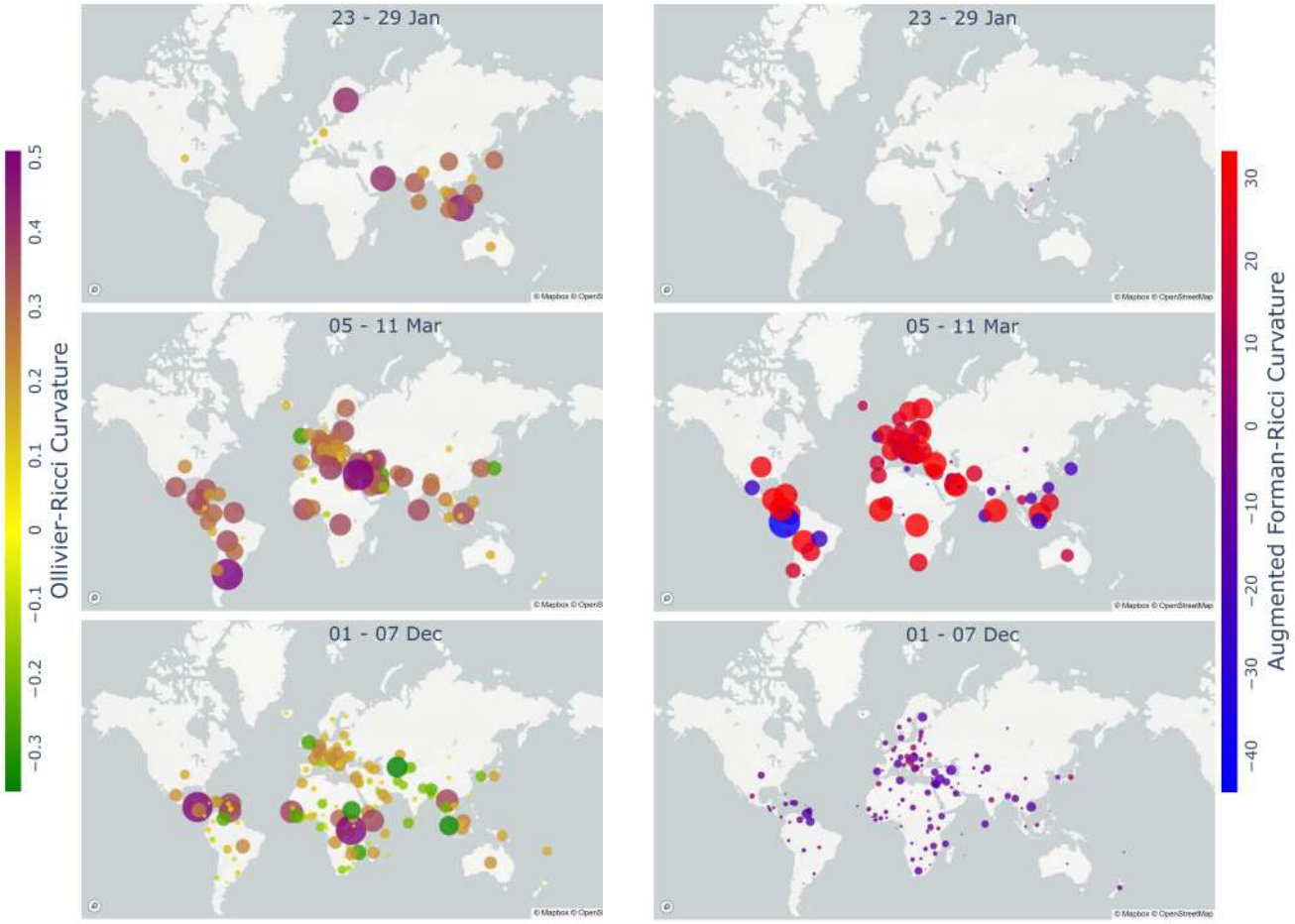
Figure 18 – Reported COVID epidemic cases per time window, for new cases, vs. its curvatures for a time window of 7 days (Top) and 21 days (bottom). In red, we indicate the moment when the WHO declared COVID-19 as a pandemic.



Source: self-provided.

Figure 19 – Illustration of the distribution of the Ollivier-Ricci curvature (a) and Forman-Ricci curvature (b) for three different time windows based on new cases corresponding to periods before, during, and after the beginning of the COVID-19 pandemic.

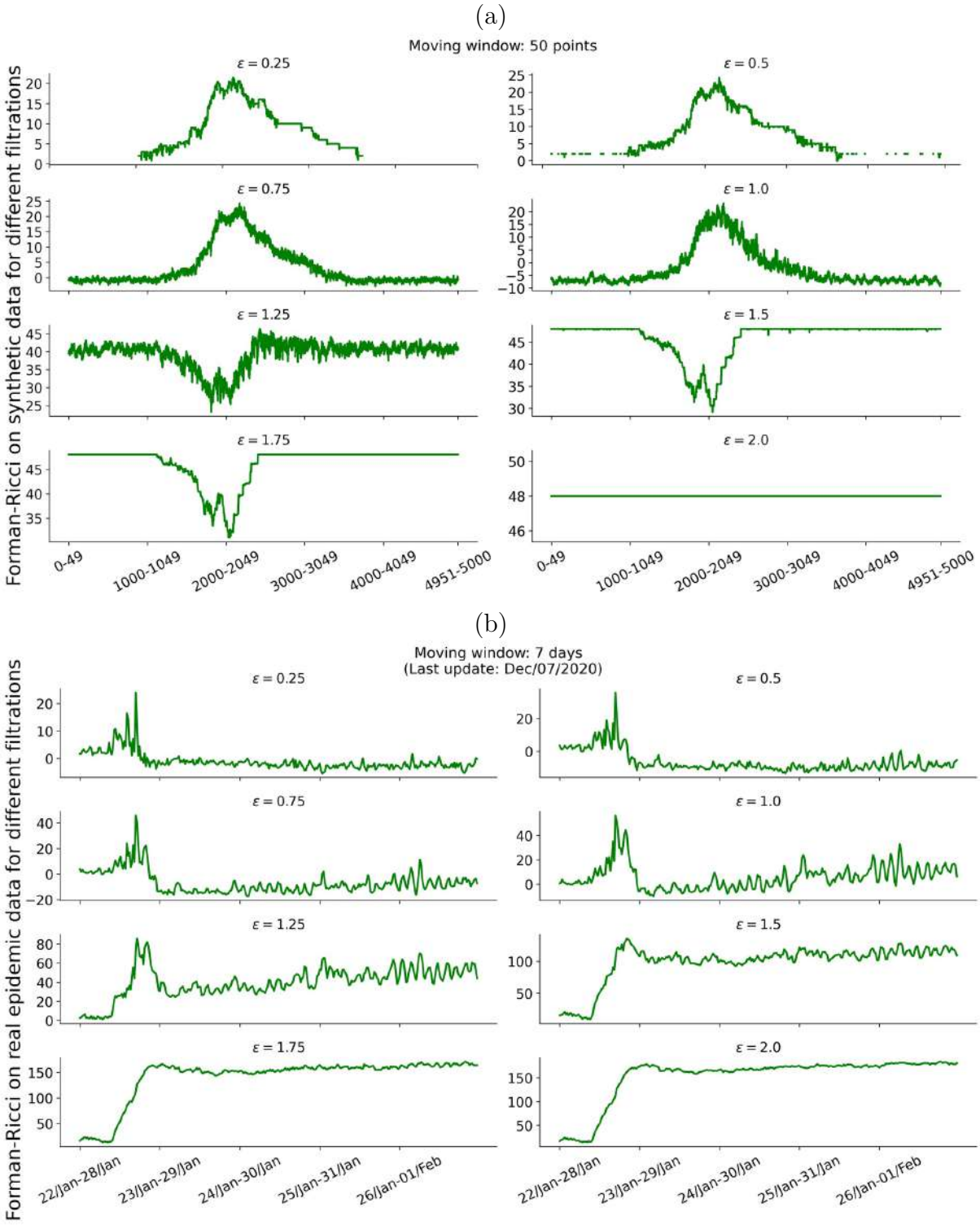
(a) (b)



Source: self-provided.

seven days. All those classic network metrics are easily computed using python NetworkX package (HAGBERG; SWART; CHULT, 2008). We summarize the results of our comparison in Figure 22 for new cases and Figure 23 for new deaths. For better comparability, we also computed the Pearson correlation coefficient of all these metrics and illustrate it in a cluster map in Figure 24. Some numerical exploration on the relationship between discrete curvature and classic network metrics in complex systems was made in (SREEJITH et al., 2016), but a rigorous connection between classic network metrics and discrete curvatures deserves further investigation. Our systematic comparison shows that the Ricci curvatures provide analogous results to some of the classic metrics for both new cases new deaths. In particular, the percolation threshold, betweenness centrality, and average clustering coefficients for new cases are the three classic metrics with the highest absolute correla-

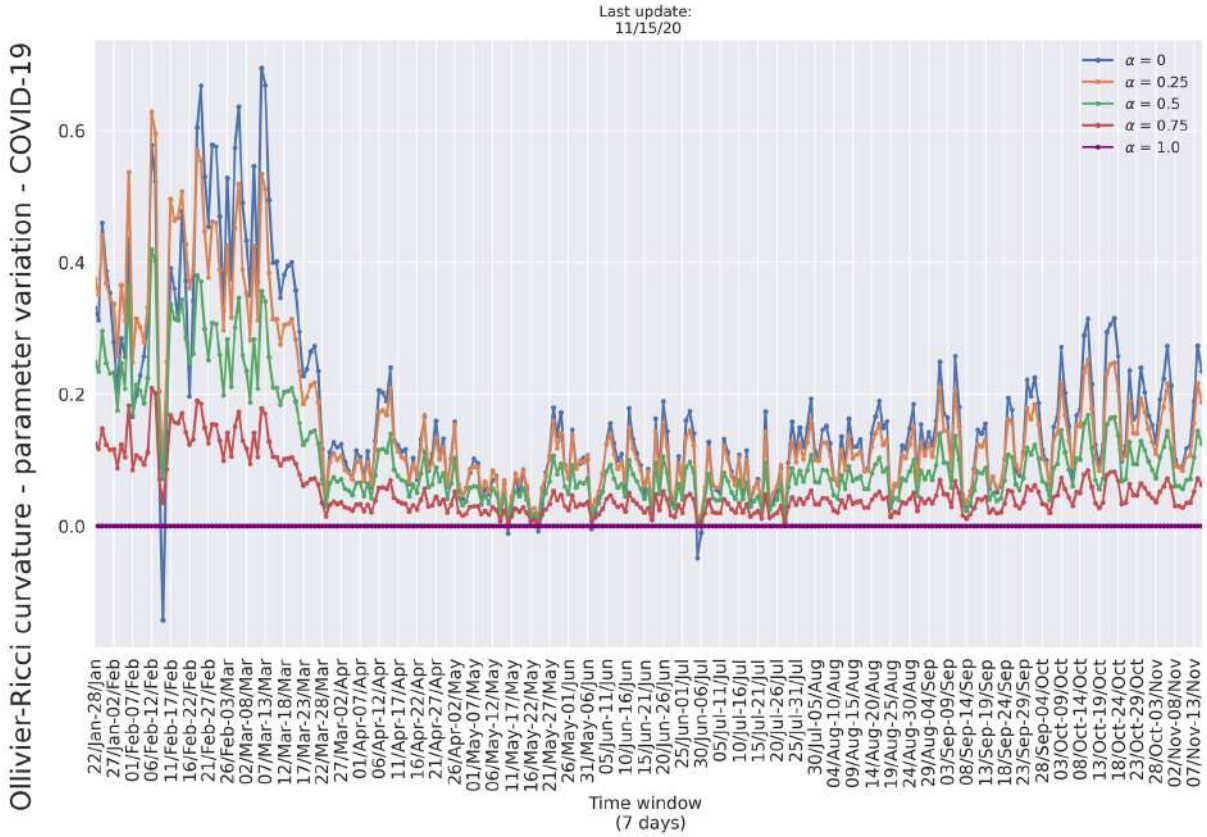
Figure 20 – Augmented Forman-Ricci curvature for different threshold values, for synthetic data (a) and real data (b).



Source: self-provided.

tion with Ollivier-Ricci curvatures, see Figure 24 (a). In addition, the average clustering coefficient, mean betweenness centrality and mean diameter, for new deaths are the three classic metrics with the highest absolute correlation with Ollivier-Ricci curvatures, see Figure 24 (b). However, the Forman-Ricci curvature has the advantage of fast time pro-

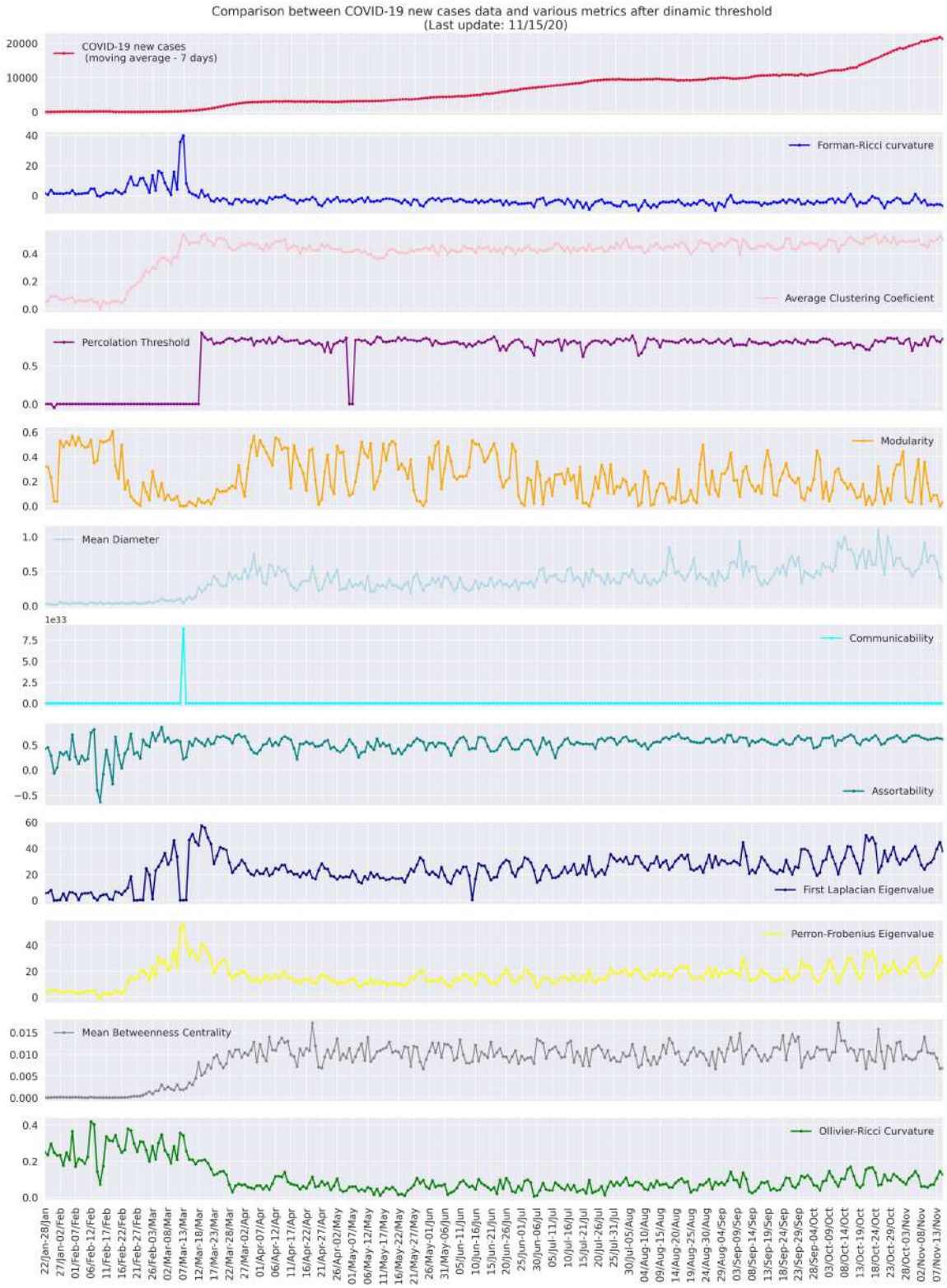
Figure 21 – Ollivier-Ricci curvature of COVID-19 (new cases), for different parameter values.



Source: self-provided.

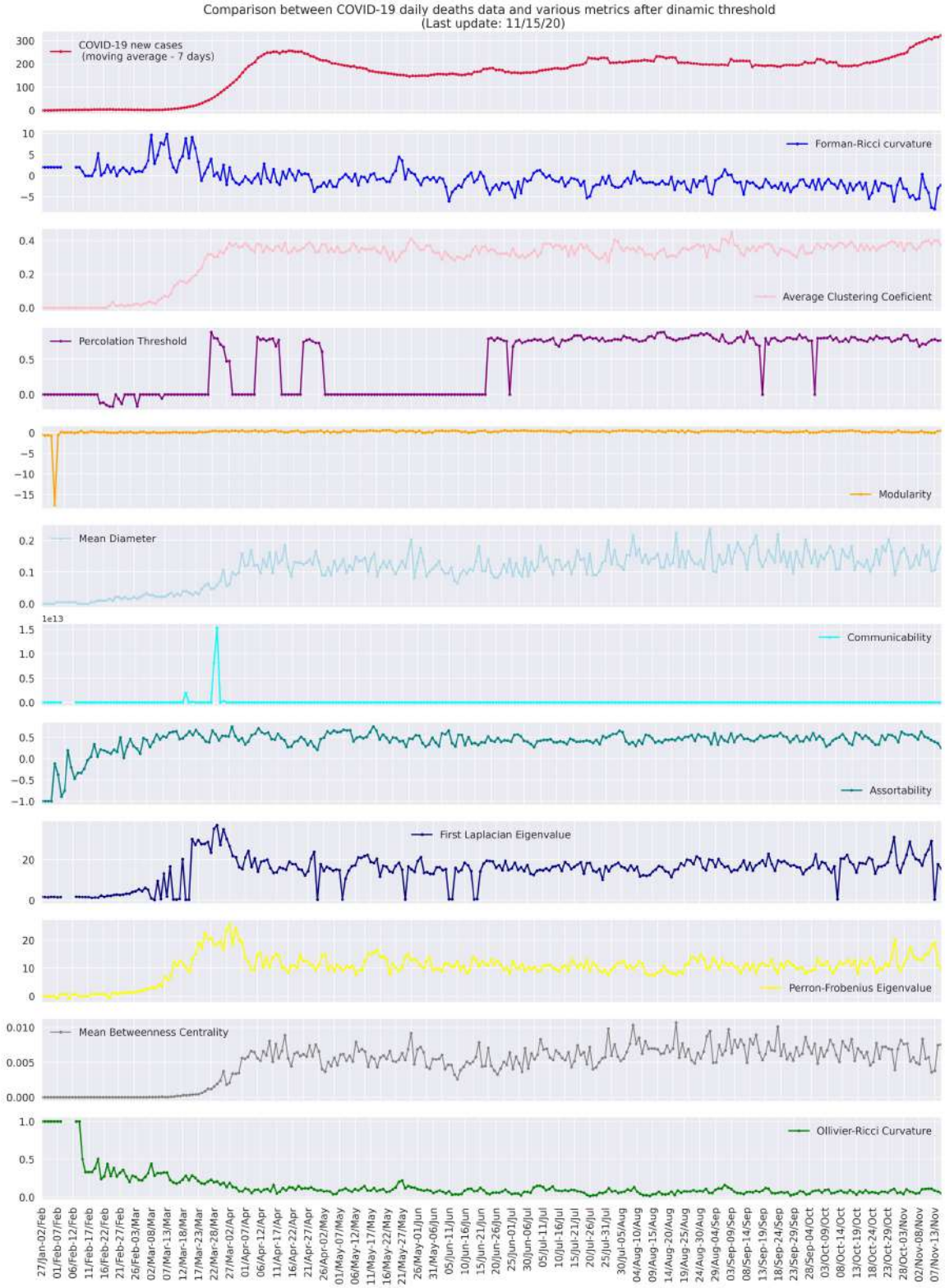
cessing and easy implementation. For new cases, in particular, the percolation threshold, mean betweenness centrality, and average communicability are the three classic metrics with the highest absolute correlation with Forman-Ricci curvatures, see Figure 24 (a). In addition, for new deaths, the average clustering coefficient, percolation threshold, and mean diameter are the three classic metrics with the highest absolute correlation with Forman-Ricci curvatures, see Figure 24 (b).

Figure 22 – Comparison between various network metrics and COVID-19 (new cases).



Source: self-prvided.

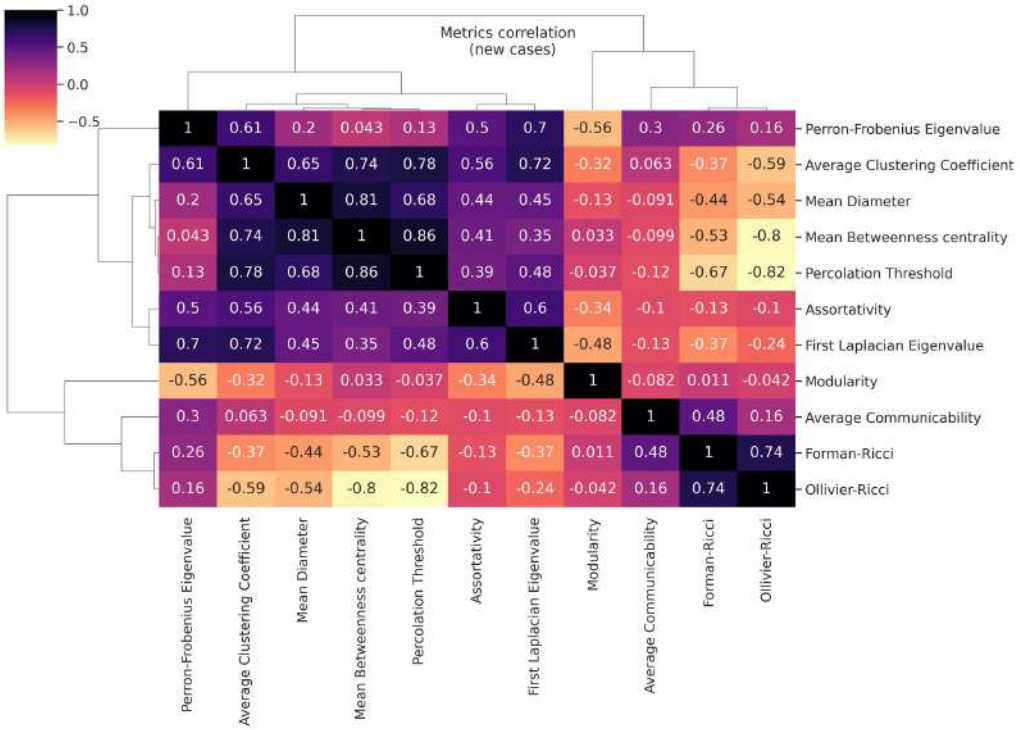
Figure 23 – Comparison between various network metrics and COVID-19 (new deaths).



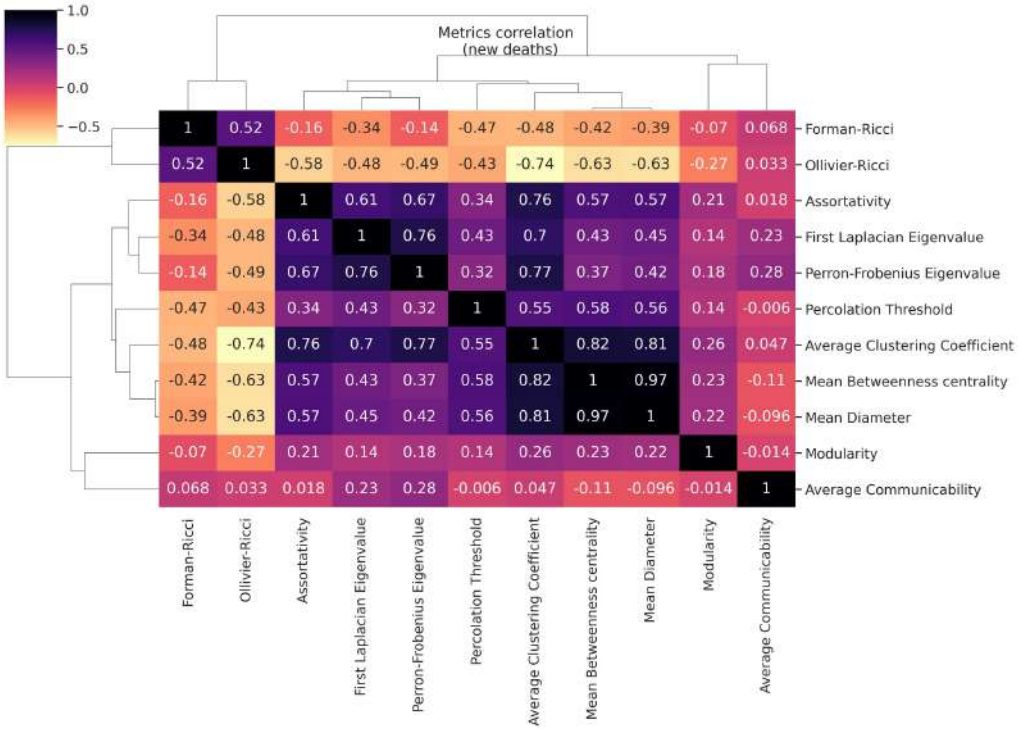
Source: self-provided.

Figure 24 – Cluster matrix comparing classic and geometric metrics in COVID-19 data, for new cases (a) and new deaths (b).

(a)



(b)

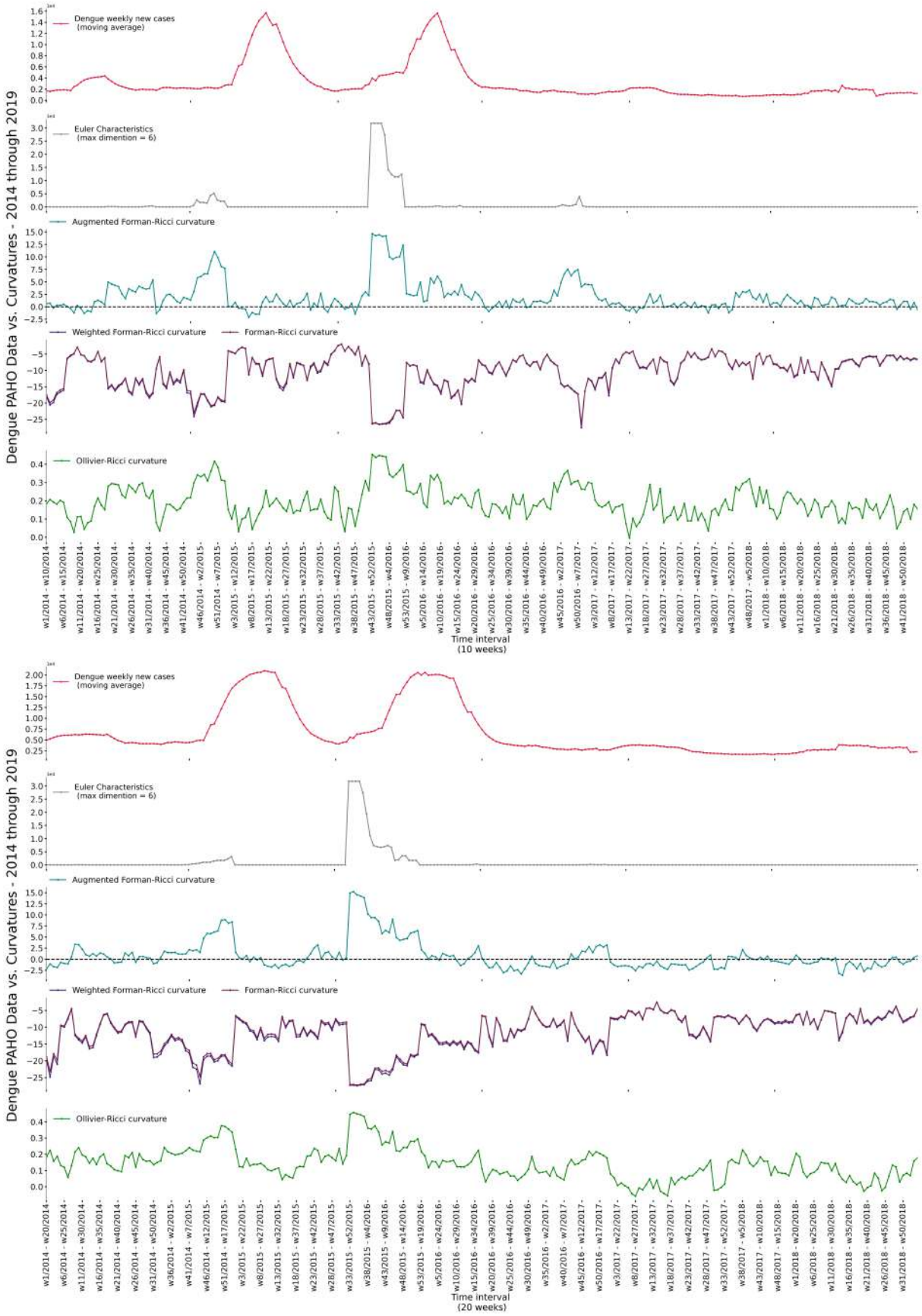


Source: self-provided.

7.2 TOPOLOGICAL AND GEOMETRIC APPROACH TO DENGUE

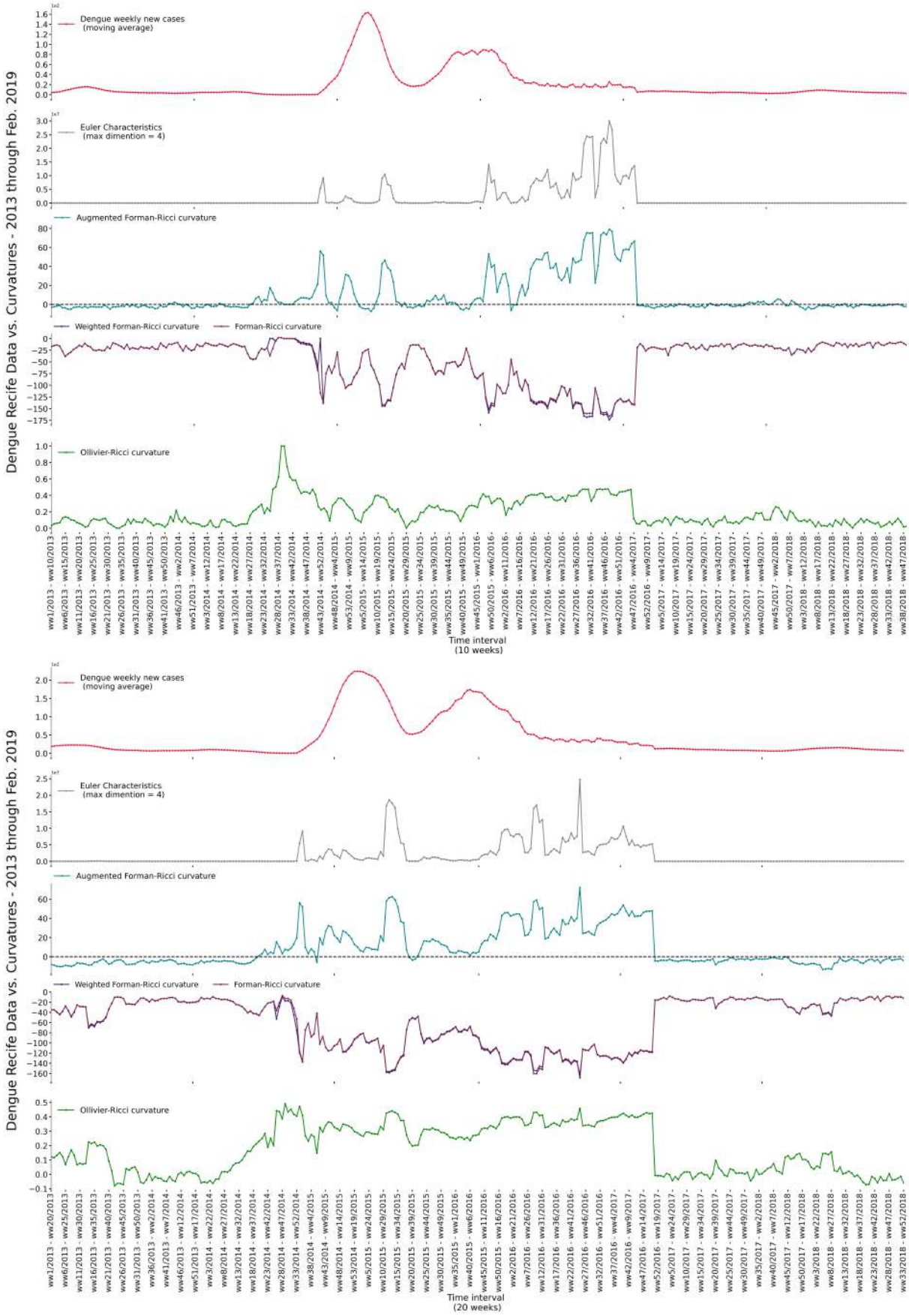
Once that we validated our technique on Dengue synthetic data, we are now able to apply the topological and geometric approaches to real Dengue data. In Figs. 25 and 26 we show the comparison between curvatures and the weekly collected Dengue database, provided by Pan American Health Organization (PAHO, 2018) and DadosRecife (DADOSRECIFE, 2019), respectively. Although the Dengue data were provided by weeks (which might reduce important information from daily cases, and than, increase noise in the data set), the results obtained from curvature studies were satisfactory. As in the application to synthetic data provided in the networks SIR model in section 6.2, we have an analogous behavior of the curvatures on detecting Dengue outbreaks, and this is clear when the high values of curvatures are present near epidemic peaks. All the curvatures presented signs for epidemic outbreaks, despite the different levels of noise. Euler characteristic showed to be the best option to network fragility due to the high levels of true pandemic signs, in contrast to the noise, which was barely visible. Even with a higher level of noise, the geometric approach also presented an easy implementation to detect epidemic outbreaks, once that the sensibility to smaller epidemic epochs are also detected. Here, we also used $\alpha = 0.5$ as the metric parameter in Ollivier-Ricci curvature calculus. The test for other parameter values, such for synthetic as real data, deserves further investigations.

Figure 25 – Reported PAHO Dengue epidemic cases per time window, for weekly new cases, vs. its curvatures for a time window of 10 weeks (Top) and 20 weeks (bottom).



Source: self-provided.

Figure 26 – Reported Recife Dengue epidemic cases per time window, for weekly new cases, vs. its curvatures for a time window of 10 weeks (Top) and 20 weeks (bottom).



Source: self-provided.

8 CONCLUSION

In this work, we showed that Ricci curvatures metrics, as well as the Euler characteristics, might be strong indicators for the identification of a pandemic state in epidemic networks, in special, Dengue and COVID-19. Consequently, it could have the potential to become a data-driven, geometric and topological approach in epidemiology more generally. All discrete curvatures presented similar results when it comes to the detection of geometric signatures on the epidemic network. While the Augmented Forman-Ricci discretization is purely combinatory and differs substantially from the Ollivier-Ricci computation, it presents substantial advantages, such as less noise and quicker computational convergence. Meanwhile, the Euler characteristics provides better visual results, once that mostly real information is not confused with the noise in data. However, as the computation algorithm for finding all cliques in a network is NP-complete, a huge computation power is required.

The main added value of a topological and geometric approaches, in contrast to classical stochastic and modelling simulations, is that the results emerge intrinsically and empirically independent of parameter estimations for the pandemic. We emphasise that this are the early stage of this field and deserves further investigations.

We also compared the Augmented Forman-Ricci and Ollivier-Ricci curvatures with classic network metrics, and found that our approach is consistent with results for network fragility detection. In particular, we found absolute correlation between the Ricci curvatures and some classical network metrics. Although some of the classical network metrics provided network fragility detection analogous to the Ricci curvatures, geometrization is ubiquitous across domains beyond network science, e.g. in quantum (KIBBLE, 1979) and statistical mechanics (CASETTI; PETTINI; COHEN, 2000), and statistics (SAVILLE; WOOD, 2012; PAWLOWSKY-GLAHN; EGOZCUE, 2001). We tested this approaches for synthetic data generated by the fractal model for COVID-19. We also used a modified version of SIR model for Dengue disease. However, the extension of this framework to other epidemic models deserves further investigation, once that the aim of our work was to provide geometric and topological approaches to epidemic data, instead of finding the best model for fitting.

Lastly, our work could pave the way for parameter-free, topological and geometric approaches to epidemic networks and open the possibility for studying epidemics from a topological and geometric perspective.

REFERENCES

ARENAS, C. J. Villabona; ZANOTTO, P. M. de A. Worldwide spread of dengue virus type 1. *PloS one*, Public Library of Science, v. 8, n. 5, p. e62649, 2013.

BAEZ, J. C.; BIAMONTE, J. Quantum techniques for stochastic mechanics. *arXiv preprint arXiv:1209.3632*, 2012.

BAILEY, N. T. et al. *The mathematical theory of infectious diseases and its applications*. [S.l.]: Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

BALL, F.; BRITTON, T.; LEUNG, K. Y.; SIRL, D. A stochastic sir network epidemic model with preventive dropping of edges. *Journal of mathematical biology*, Springer, v. 78, n. 6, p. 1875–1951, 2019.

BARABÁSI, A.-L. et al. *Network science*. [S.l.]: Cambridge university press, 2016.

BATTISTON, F.; CENCETTI, G.; IACOPINI, I.; LATORA, V.; LUCAS, M.; PATANIA, A.; YOUNG, J.-G.; PETRI, G. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, Elsevier, 2020.

BISWAS, K.; KHALEQUE, A.; SEN, P. Covid-19 spread: Reproduction of data and prediction using a sir model on euclidean network. *arXiv preprint arXiv:2003.07063*, 2020.

BLOCK, P.; HOFFMAN, M.; RAABE, I. J.; DOWD, J. B.; RAHAL, C.; KASHYAP, R.; MILLS, M. C. Social network-based distancing strategies to flatten the covid-19 curve in a post-lockdown world. *Nature Human Behaviour*, Nature Publishing Group, p. 1–9, 2020.

BOBENKO, A. *Discrete differential geometry*. [S.l.]: Springer, 2008.

BOCHNER, S. Vector fields and ricci curvature. *Bulletin of the American Mathematical Society*, v. 52, n. 9, p. 776–797, 1946.

BOLLOBÁS, B.; BÉLA, B. *Random graphs*. [S.l.]: Cambridge university press, 2001.

BONDY, J. A. J. A.; MURTY, U. S. R. *Graph theory*. [S.l.]: Springer, 2008. 651 p. ISSN 0072-5285.

BRAUER, F. The kermack–mckendrick epidemic model revisited. *Mathematical Biosciences*, Elsevier, v. 198, n. 2, p. 119–131, 2005.

BRITTON, T.; JUHER, D.; SALDAÑA, J. A network epidemic model with preventive rewiring: comparative analysis of the initial phase. *Bulletin of mathematical biology*, Springer, v. 78, n. 12, p. 2427–2454, 2016.

BUBENIK, P. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, JMLR. org, v. 16, n. 1, p. 77–102, 2015.

CARLSSON, G. Topology and data. *Bulletin of the American Mathematical Society*, v. 46, n. 2, p. 255–308, 1 2009. ISSN 0273-0979. Disponível em: <http://www.ams.org/journal-getitem?pii=S0273-0979-09-01249-X>.

CASETTI, L.; PETTINI, M.; COHEN, E. Geometric approach to hamiltonian dynamics and statistical mechanics. *Physics Reports*, Elsevier, v. 337, n. 3, p. 237–341, 2000.

CHAPLIN, C. 1899–1976.

CLAUSET, A.; NEWMAN, M. E.; MOORE, C. Finding community structure in very large networks. *Physical review E*, APS, v. 70, n. 6, p. 066111, 2004.

COLDING, T. H. Ricci curvature and volume convergence. *Annals of mathematics*, JSTOR, p. 477–501, 1997.

DADOSRECIFE. *Dados Recife*. 2019. Disponível em: <http://dados.recife.pe.gov.br/dataset/casos-de-dengue-zika-e-chikungunya>.

DIAS, J. P.; BASTOS, C.; ARAÚJO, E.; MASCARENHAS, A. V.; NETTO, E. M.; GRASSI, F.; SILVA, M.; TATTO, E.; MENDONÇA, J.; ARAÚJO, R. F. et al. Acute chagas disease outbreak associated with oral transmission. *Revista da Sociedade Brasileira de Medicina Tropical*, SciELO Brasil, v. 41, n. 3, p. 296–300, 2008.

EDELSBRUNNER, H.; HARER, J. J. *Computational topology : an introduction*. [S.l.]: American Mathematical Society, 2010. 241 p. ISBN 9780821849255.

EDWARDS, A. L. *An introduction to linear regression and correlation*. [S.l.], 1984.

ERBAN, R.; CHAPMAN, J.; MAINI, P. A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv preprint arXiv:0704.1908*, 2007.

ERDOS; P. On random graphs. *Publicationes Mathematicae*, v. 6, p. 290–297, 1959. Disponível em: <https://ci.nii.ac.jp/naid/10018689248/>.

ERDŐS, P.; RÉNYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, v. 5, n. 1, p. 17–60, 1960.

FAROOQ, H.; CHEN, Y.; GEORGIOU, T. T.; TANNENBAUM, A.; LENGLET, C. Network curvature as a hallmark of brain structural connectivity. *Nature communications*, Nature Publishing Group, v. 10, n. 1, p. 1–11, 2019.

FILHO, E. C. d. A. Topological transitions on protein-protein interaction networks. Universidade Federal de Pernambuco, 2019.

FORMAN, R. Bochner's method for cell complexes and combinatorial ricci curvature. *Discrete and Computational Geometry*, Springer, v. 29, n. 3, p. 323–374, 2003.

FORNITO, A.; ZALESKY, A.; BULLMORE, E. *Fundamentals of brain network analysis*. [S.l.]: Academic Press, 2016.

FORNITO, A.; ZALESKY, A.; BULLMORE, E. T. *Fundamentals of brain network analysis*. [S.l.: s.n.]. 476 p. ISBN 9780124081185.

FOSTER, J. G.; FOSTER, D. V.; GRASSBERGER, P.; PACZUSKI, M. Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 107, n. 24, p. 10815–10820, 2010.

GRACIA-TABUENCA, Z.; DÍAZ-PATIÑO, J. C.; ARELIO, I.; ALCAUTER, S. Topological data analysis reveals robust alterations in the whole-brain and frontal lobe functional connectomes in attention-deficit/hyperactivity disorder. *Eneuro*, Society for Neuroscience, v. 7, n. 3, 2020.

GREENBERG, M. J. *Algebraic topology: a first course*. [S.l.]: CRC Press, 2018.

HAGBERG, A.; SWART, P.; CHULT, D. S. *Exploring network structure, dynamics, and function using NetworkX*. [S.l.], 2008.

HATCHER, A. *Algebraic topology*. [S.l.]: , 2005.

JOSSERAN, L.; PAQUET, C.; ZEHGNOUN, A.; CAILLERE, N.; LE, A. T.; SOLET, J.-L.; LEDRANS, M. Chikungunya disease outbreak, reunion island. *Emerging infectious diseases*, Centers for Disease Control and Prevention, v. 12, n. 12, p. 1994–1995, 2006.

KAISER, M. Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. *New Journal of Physics*, IOP Publishing, v. 10, n. 8, p. 083042, 2008.

KEELING, M. J.; EAMES, K. T. Networks and epidemic models. *Journal of the Royal Society Interface*, The Royal Society London, v. 2, n. 4, p. 295–307, 2005.

KENNEY, J. F.; KEEPING, E. S. *Mathematics of statistics*. [S.l.]: van Nostrand, 1957. v. 2.

KIBBLE, T. W. Geometrization of quantum mechanics. *Communications in Mathematical Physics*, Springer, v. 65, n. 2, p. 189–201, 1979.

KIM, K.; KIM, S. Y.; HA, D.-H. Characteristics of networks in financial markets. *Computer physics communications*, Elsevier, v. 177, n. 1-2, p. 184–185, 2007.

KNILL, O. A graph theoretical gauss-bonnet-chern theorem. *arXiv preprint arXiv:1111.5395*, 2011.

KNILL, O. On the dimension and euler characteristic of random graphs. 12 2011.

KNILL, O. An index formula for simple graphs. *arXiv preprint arXiv:1205.0306*, 2012.

KNILL, O. On index expectation and curvature for networks. *arXiv preprint arXiv:1202.4514*, 2012.

KNILL, O. The euler characteristic of an even-dimensional graph. *arXiv preprint arXiv:1307.3809*, 2013.

KNILL, O. Curvature from graph colorings. *arXiv preprint arXiv:1410.1217*, 2014.

LEVY, S. *New perspectives in algebraic combinatorics*. [S.l.]: Cambridge University Press, 1999. v. 38.

LIN, Y.; LU, L.; YAU, S.-T. Ricci curvature of graphs. *Tohoku Mathematical Journal*, v. 63, n. 4, p. 605–627, 2011. ISSN 0040-8735. Disponível em: <http://projecteuclid.org/euclid.tmj/1325886283>.

LOTT, J.; VILLANI, C. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, JSTOR, p. 903–991, 2009.

MATAMALAS, J. T.; GÓMEZ, S.; ARENAS, A. Abrupt phase transition of epidemic spreading in simplicial complexes. *arXiv preprint arXiv:1910.03069*, 2019.

MATSUMOTO, Y. *An introduction to Morse theory*. [S.l.]: American Mathematical Society, 2002. 219 p. ISBN 9780821810224.

MILLER, J. C.; KISS, I. Z. Epidemic spread in networks: Existing methods and current challenges. *Mathematical modelling of natural phenomena*, EDP Sciences, v. 9, n. 2, p. 4–42, 2014.

MOORE, C.; NEWMAN, M. E. Epidemics and percolation in small-world networks. *Physical Review E*, APS, v. 61, n. 5, p. 5678, 2000.

NAFFAKH, N.; WERF, S. V. D. April 2009: an outbreak of swine-origin influenza a (h1n1) virus with evidence for human-to-human transmission. *Microbes and infection*, Elsevier, v. 11, n. 8-9, p. 725–728, 2009.

NAJMAN, L.; ROMON, P. *Modern approaches to discrete curvature*. [S.l.]: Springer, 2017. v. 2184.

NEWMAN, M. E. Spread of epidemic disease on networks. *Physical review E*, APS, v. 66, n. 1, p. 016128, 2002.

NEWMAN, M. E. Mixing patterns in networks. *Physical review E*, APS, v. 67, n. 2, p. 026126, 2003.

OLLIVIER, Y. Ricci curvature of markov chains on metric spaces. *arXiv preprint math/0701886*, 2007.

ONNELA, J.-P.; KASKI, K.; KERTSZ, J. Clustering and information in correlation based financial networks. *The European Physical Journal B - Condensed Matter*, Springer Berlin Heidelberg, v. 38, n. 2, p. 353–362, 3 2004. ISSN 1434-6028. Disponível em: <http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1140/epjb/e2004-00128-7>.

ORGANIZATION, W. H. et al. Dengue/dengue haemorrhagic fever: situation in 2000. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire*, v. 75, n. 24, p. 193–196, 2000.

ORGANIZATION, W. H. et al. Coronavirus disease 2019 (covid-19): situation report, 82. World Health Organization, 2020.

OTTER, N.; PORTER, M. A.; TILLMANN, U.; GRINDROD, P.; HARRINGTON, H. A. A roadmap for the computation of persistent homology. *EPJ Data Science*, SpringerOpen, v. 6, n. 1, p. 17, 2017.

PAHO. *Pan American Health Organization*. 2018. Disponível em: <http://www.paho.org/data/index.php/en/>.

PASCUCCI, V.; TRICOCHE, X.; HAGEN, H.; TIERNY, J. *Topological Methods in Data Analysis and Visualization: Theory, Algorithms, and Applications*. [S.l.]: Springer Science & Business Media, 2010.

PASTOR-SATORRAS, R.; CASTELLANO, C.; MIEGHEM, P. V.; VESPIGNANI, A. Epidemic processes in complex networks. *Reviews of modern physics*, APS, v. 87, n. 3, p. 925, 2015.

PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, APS, v. 63, n. 6, p. 066117, 2001.

PAWLOWSKY-GLAHN, V.; EGOZCUE, J. J. Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 15, n. 5, p. 384–398, 2001.

PETRI, G.; SCOLAMIERO, M.; DONATO, I.; VACCARINO, F.; LAMBIOTTE, R. Topological Strata of Weighted Complex Networks. *PLoS ONE*, v. 8, n. 6, 2013. Disponível em: <http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0066506&type=printable>.

POURYAHYA, M.; MATHEWS, J.; TANNENBAUM, A. Comparing three notions of discrete ricci curvature on biological networks. *arXiv preprint arXiv:1712.02943*, 2017.

PRASSE, B.; ACHTERBERG, M. A.; MA, L.; MIEGHEM, P. V. Network-inference-based prediction of the covid-19 epidemic outbreak in the chinese province hubei. *Applied Network Science*, SpringerOpen, v. 5, n. 1, p. 1–11, 2020.

PRASSE, B.; MIEGHEM, P. V. Network reconstruction and prediction of epidemic outbreaks for general group-based compartmental epidemic models. *IEEE Transactions on Network Science and Engineering*, IEEE, 2020.

SAGGAR, M.; SPORNS, O.; GONZALEZ-CASTILLO, J.; BANDETTINI, P. A.; CARLSSON, G.; GLOVER, G.; REISS, A. L. Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature communications*, Nature Publishing Group, v. 9, n. 1, p. 1–14, 2018.

SAMAL, A.; SREEJITH, R.; GU, J.; LIU, S.; SAUCAN, E.; JOST, J. Comparative analysis of two discretizations of ricci curvature for complex networks. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 1–16, 2018.

SANDHU, R.; GEORGIOU, T.; REZNIK, E.; ZHU, L.; KOLESOV, I.; SENBABAOGLU, Y.; TANNENBAUM, A. Graph curvature for differentiating cancer networks. *Scientific reports*, Nature Publishing Group, v. 5, p. 12323, 2015.

SANDHU, R.; GEORGIOU, T.; REZNIK, E.; ZHU, L.; KOLESOV, I.; SENBABAOGLU, Y.; TANNENBAUM, A. Graph Curvature for Differentiating Cancer Networks. *Scientific Reports*, Nature Publishing Group, v. 5, n. 1, p. 12323, 12 2015. ISSN 2045-2322. Disponível em: <http://www.nature.com/articles/srep12323>.

SANDHU, R. S.; GEORGIOU, T. T.; TANNENBAUM, A. R. Ricci curvature: An economic indicator for market fragility and systemic risk. *Science advances*, American Association for the Advancement of Science, v. 2, n. 5, p. e1501495, 2016.

SANTOS, F. A. N.; RAPOSO, E. P.; COUTINHO-FILHO, M. D.; COPELLI, M.; STAM, C. J.; DOUW, L. Topological phase transitions in functional brain networks. *Phys. Rev. E*, American Physical Society, v. 100, p. 032414, Sep 2019. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevE.100.032414>.

SARAMÄKI, J.; KIVELÄ, M.; ONNELA, J.-P.; KASKI, K.; KERTESZ, J. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, APS, v. 75, n. 2, p. 027105, 2007.

SAVILLE, D. J.; WOOD, G. R. *Statistical methods: The geometric approach*. [S.l.]: Springer Science & Business Media, 2012.

SCALLAN, E.; HOEKSTRA, R. M.; ANGULO, F. J.; TAUXE, R. V.; WIDDOWSON, M.-A.; ROY, S. L.; JONES, J. L.; GRIFFIN, P. M. Foodborne illness acquired in the united states—major pathogens. *Emerging infectious diseases*, Centers for Disease Control and Prevention, v. 17, n. 1, p. 7, 2011.

SNEDECOR, G. W.; COCHRAN, W. G. *Statistical Methods-80-7**. [S.l.]: Iowa State Press, 1937.

SREEJITH, R.; MOHANRAJ, K.; JOST, J.; SAUCAN, E.; SAMAL, A. Forman curvature for complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, IOP Publishing, v. 2016, n. 6, p. 063206, 2016.

STOLERMAN, L. M.; COOMBS, D.; BOATTO, S. Sir-network model and its application to dengue fever. *SIAM Journal on Applied Mathematics*, SIAM, v. 75, n. 6, p. 2581–2609, 2015.

TAO, Z.; ZHONGQIAN, F.; BINGHONG, W. Epidemic dynamics on complex networks. *Progress in Natural Science*, Taylor & Francis, v. 16, n. 5, p. 452–457, 2006.

TAYLOR, D.; KLIMM, F.; HARRINGTON, H. A.; KRAMÁR, M.; MISCHAIKOW, K.; PORTER, M. A.; MUCHA, P. J. Topological data analysis of contagion maps for examining spreading processes on networks. *Nature communications*, Nature Publishing Group, v. 6, p. 7723, 2015.

TEWARIE, P.; PRASSE, B.; MEIER, J. M.; SANTOS, F. A.; DOUW, L.; SCHOONHEIM, M.; STAM, C. J.; MIEGHEM, P. V.; HILLEBRAND, A. Mapping functional brain networks from the structural connectome: Relating the series expansion and eigenmode approaches. *NeuroImage*, Elsevier, p. 116805, 2020.

TIAN, G.; YAU, S. T. Complete kahler manifolds with zero ricci curvature. i. *Journal of the American Mathematical Society*, JSTOR, v. 3, n. 3, p. 579–609, 1990.

WASSERMAN, L. Topological data analysis. *Annual Review of Statistics and Its Application*, Annual Reviews, v. 5, p. 501–532, 2018.

WEBER, M.; SAUCAN, E.; JOST, J. Coarse geometry of evolving networks. *Journal of Complex Networks*, Oxford University Press, v. 6, n. 5, p. 706–732, 2018.

WEITZ, J. S.; BECKETT, S. J.; COENEN, A. R.; DEMORY, D.; DOMINGUEZ-MIRAZO, M.; DUSHOFF, J.; LEUNG, C.-Y.; LI, G.; MĂGĂLIE, A.; PARK, S. W. et al. Modeling shield immunity to reduce covid-19 epidemic spread. *Nature medicine*, Nature Publishing Group, p. 1–6, 2020.

WODARZ, D.; KOMAROVA, N. L. Patterns of the covid19 epidemic spread around the world: exponential vs power laws. *medRxiv*, Cold Spring Harbor Laboratory Press, 2020.

ZIFF, A. L.; ZIFF, R. M. Fractal kinetics of covid-19 pandemic. *medRxiv*, Cold Spring Harbor Laboratory Press, 2020.

ZOMORODIAN, A. Topological data analysis. *Advances in applied and computational topology*, v. 70, p. 1–39, 2012.

ZOMORODIAN, A. J. *Topology for Computing*. Cambridge: Cambridge University Press, 2005. ISBN 9780511546945. Disponível em: <https://www.cambridge.org/core/product/identifier/9780511546945/type/book>.

TOTH, C. D.; O'ROURKE, J.; GOODMAN, J. E. *Handbook of discrete and computational geometry*. [S.l.]: Chapman and Hall/CRC, 2017.

WANG, Z.; DAI, Z.; GONG, G.; ZHOU, C.; HE, Y. Understanding structural-functional relationships in the human brain: a large-scale network perspective. *The Neuroscientist*, Sage Publications Sage CA: Los Angeles, CA, v. 21, n. 3, p. 290–305, 2015.

WU, Z.; MENICHETTI, G.; RAHMEDE, C.; BIANCONI, G. Emergent complex network geometry. *Scientific reports*, Nature Publishing Group, v. 5, n. 1, p. 1–12, 2015.

YANG, S. *M. E. J. Newman Networks: An Introduction, page 224. Oxford University Press, 2011*. [S.l.]: Taylor & Francis, 2013.