



**HAL**  
open science

# Epidemic Threshold: A Laplacian Spectral and Structural Approach of Prediction

Claude Kanyou, Etienne Kouokam, Norbert Tsope

► **To cite this version:**

Claude Kanyou, Etienne Kouokam, Norbert Tsope. Epidemic Threshold: A Laplacian Spectral and Structural Approach of Prediction. Journal of Complex Networks, In press, 10.1093/comnet/xxx000 . hal-04430909

**HAL Id: hal-04430909**

**<https://hal.science/hal-04430909v1>**

Submitted on 1 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## Epidemic Threshold : A Laplacian Spectral and Structural Approach of Prediction

CLAUDE KANYOU<sup>†</sup>

*Department of Computer Science, Faculty of Science, Université de Yaoundé I, Yaoundé, Cameroon*

<sup>†</sup>Corresponding author: kanyouclaude@gmail.com

ETIENNE KOUOKAM

*Department of Computer Science, Faculty of Science, Université de Yaoundé I, Yaoundé, Cameroon*

AND

NORBERT TSOPZE

*Department of Computer Science, Faculty of Science, Université de Yaoundé I, Yaoundé, Cameroon*

[Received on 30 November 2023]

In epidemiology, control processes usually rely on some key parameters, such as  $R_0$  and the epidemic threshold over a contact network. Driven by network-based models, knowledge of network structures can improve the prediction of the epidemic threshold, which is a challenge when using contact network properties. There are several structural approaches to predicting the epidemic threshold. The common QMF (Quenched Mean-Field) approach uses the spectral radius as a single parameter. However, prediction can be improved by using compound diffusion parameters such as node number and Laplacian energy of graph. In this paper, a new structural and spectral prediction approach of the epidemic threshold called KSEL (K Spectral Energy of Laplacian) is designed and experimented with at different levels. Theoretical and formal levels establish mathematical foundations, while the comparative level computes a descriptive statistics summary. Qualitative and quantitative analysis are achieved at the experimental level using some data analytic and visualization techniques. Simulations are performed using a large and heterogeneous dataset with different network types and topologies. Results show that the new approach further captures the full network structure, connectivity, and network diffusion properties. KSEL is similar, shares a common rolling trend, and performs really good performance compared to the previous prediction approaches, including the most used QMF. There is a strong positive correlation and similar value distribution between KSEL and the previous prediction approaches that accept the null hypothesis by ANOVA analysis. Therefore, the new approach offers a Laplacian structural and spectral area to analyse spreading processes over a network. The results can have a practical interest to improve the control and prediction of spreading processes in networks. Particularly, they can be significant to advise an effective epidemiological control policy.

*Keywords:* Network structures; Dynamics on networks; Eigenvalues ; Laplacian energy of graph ; Epidemic threshold ; Epidemiology

## 1. Introduction

Networks are everywhere. Several real phenomena, like disease spreading, and rumour propagation, are described and understood as a spreading process in the complex system [6, 14, 17, 19]. These processes are widely modelled using networks or graphs. Therefore, networks are of great interest and become a fertile and flexible tools for scientific modelling and analysis of complex systems [17, 21] such as an infectious disease spread. Disease transmission can occur when there is direct or indirect contact between infected individuals or groups [15, 18] over a contact network.

In the study of infectious disease spread, the basic reproduction number  $R_0$  is the average of the expected secondary infection number caused by a primary infectious individual introduced into a fully susceptible host population. Critically,  $R_0$  depends not only on the disease but also on the host population structure [12]. Therefore, network-based models of epidemiological contact have emerged as an important tool in understanding and predicting the spread of an infectious disease [3, 15, 17]. Understanding the network structure improves the control of the micro and macro propagation [1, 12, 19], and even improves the predictions. As it happens, we need sophisticated tools such as the spectral theory of graph for analysis and visualization of network structure [2, 3]. So, predicting if a disease will disappear or become an epidemic refers to the *epidemic threshold* noted  $\tau$ .

$\tau$  is the incidence level of a disease at which it can be considered as an epidemic. In this sense,  $\tau$  describes a fundamental question in epidemiology and related areas [22]. Hence, in the study of infectious diseases on networks, researchers compute epidemic thresholds to help forecast whether or not a disease will eventually infect a large fraction of a population [4] to be considered as an epidemic. Formally,  $\tau$  is the critical  $\beta/\gamma$  ratio value beyond which an infection becomes an epidemic [26]. However,  $\tau$  depends not only on the transmission and recovery rates of a disease but also fundamentally on the network structure [26]. Therefore, using contact network properties to predict the epidemic threshold is a challenge. By using structural and spectral network properties, we can improve the prediction of the epidemic threshold. So, this paper aims to design and experiment a new structural and spectral prediction approach of the epidemic threshold called *KSEL (K Spectral Energy of Laplacian)*.

Indeed, the epidemic threshold is closed in the diffusion context. Diffusion is a spreading process by which information, epidemics, viruses, and any other spreading process occur over networks [18]. In the literature, to study diffusion or spreading processes in a network like an infectious disease spread, the Laplacian matrix of a graph and its spectrum have useful features. They can be used to determine the average path length, degree distribution, connectivity of the graph, number of connected components, and other determinants of diffusion over a network [2, 5, 7, 18]; these determinants are useful for understanding and predicting a diffusion process like an infectious disease spread. Related to the previous works, to improve the prediction of  $\tau$  as a spreading process, we can use the Laplacian matrix and its spectrum around the Laplacian energy of graph as a compound global parameter.

The layout of this paper is organised as follows: Section 2 presents the concept definitions and previous structural approaches of prediction. Section 3 describes the proposed Laplacian spectral and structural prediction approach of the epidemic threshold with its root concepts and mathematical foundations. Section 4 is dedicated to the experimentations, results, and discussions. We conclude in Section 5.

## 2. Concepts definition and Previous structural approaches of prediction

### 2.1 Concepts definition

DEFINITION 2.1 (Overview of general concepts) In Table 1 bellow, we define some basic concepts used as notations in this work.

Symbol	Short description
$G$	A connected graph $G = (V, E)$ with $n$ nodes in $V$ and $m$ edges in $E$ .
$A$	Adjacency matrix of the graph.
$\langle k \rangle, \langle k^2 \rangle$	Respectively the first moments (average connectivity), and the second moments (connectivity divergence) of the degree distribution.
$\lambda_{max}$	Spectral radius (largest eigenvalue) of the matrix $A$ .
$\beta$	Infection rate. It's the rate of infection or transmission from an infected individual to a susceptible individual per effective contact.
$\gamma$	Recovery rate. It's the rate that an infected individual will recover per unit time (in continuous-time models) or per time step (in discrete-time models).
$\tau$	Epidemic threshold, scaled by $\gamma^{-1}$ so that $\tau = \beta/\gamma$ .
$\tau_c$	Critical epidemic threshold.

DEFINITION 2.2 (Laplacian energy of graph) The original version of graph energy is based on the eigenvalues of  $A$  [9]. The graph energy concept has many variants including the Laplacian energy of graph. Giving a Laplacian matrix  $L$ , the Laplacian energy of graph is given by equation (2.1):

$$LE(G) = \sum_{i=1}^n |\mu_i - \frac{2m}{n}| \quad (2.1)$$

where  $\mu_i$  is the  $i^{th}$  eigenvalue of  $L$ . Nevertheless, the energy of graph concept found unexpected large applications in areas of engineering and science [10] such in [16], [11] with the epidemiological applications.

DEFINITION 2.3 (Epidemic threshold) An epidemic threshold  $\tau$  is the incidence level of a disease at which it can be considered as an epidemic. In other words,  $\tau$  is the minimum level of virulence to prevent a viral contagion from dying out quickly, and determining. Formally,  $\tau$  is the critical  $\beta/\gamma$  ratio value beyond which an infection becomes an epidemic [26].

### 2.2 Previous structural approaches of prediction

In the literature, there are many successful structural prediction approaches of the epidemic threshold. We denote various benchmarks widely used to structurally predict spreading dynamics in real networks. This includes the Mean-Field (MF), Degree-Based Mean-Field (DBMF) or Heterogeneous Mean-Field (HMF), Quenched Mean-Field (QMF) also called Individual-Based Mean-Field (IBMF), and K Spectral Energy (KSE).

2.2.1 *Mean-Field* The Mean-Field (MF) approach is based on the works of *Kephart* and *White* who adopted a modified homogeneous approach where directed graphs model the communication among

nodes or persons [13]. Formally, in a homogeneous network, the epidemic threshold is denoted by equation (2.2):

$$\tau_c^{MF} = \frac{1}{\langle k \rangle} \quad (2.2)$$

where  $\langle k \rangle$  is the first moment of degree distribution. The MF assumes that all nodes in the network are statistically equivalent; the interaction probabilities between any two nodes are the same. Therefore, the contact network structure is not considered. This approach can be inaccurate when the network degree distribution is asymmetric and heterogeneous.

**2.2.2 Heterogeneous Mean-Field** To better capture network structure, [20] improved the homogeneous MF approach to obtain the Heterogeneous Mean-Field (HMF) by making the assumption of the inability of a node (or person) to infect the node that infected it. Here, the epidemic threshold is given by equation (2.3):

$$\tau_c^{HMF} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (2.3)$$

where  $\langle k^2 \rangle$  is the second moment of degree distribution. The HMF is more used for uncorrelated networks [8]. This approach is useful under the mean-field assumption of independence between node's infectious states. Nevertheless, the HMF neglects dynamic correlations among neighbour states. Due to its parameters and assumptions, HMF can be inaccurate for quenched connections among the nodes.

**2.2.3 Quenched Mean-Field** Because neither the MF nor HMF approaches can capture enough of the contact network structure, the Quench Mean-Field (QMF) approach is developed using an adjacency matrix  $A$ . In [26], authors proposed a discrete-time formulation to predict the epidemic threshold problem with any assumption of homogeneous connectivity. Thus, the epidemic threshold is given by equation (2.4):

$$\tau_c^{QMF} = \frac{1}{\lambda_{max}} \quad (2.4)$$

where  $\lambda_{max}$  is the largest eigenvalue of  $A$ . The QMF approach depends only on network structures. It's commonly used to study spreading dynamics in a network [25]. QMF is an advanced approach that is more accurate than MF and HMF [25]. It has many variants, such as the N-intertwined approach [23], the Dynamical Message-Passing (DMP) using the non-backtracking matrix, and the Simplified DMP (SDMP). Nevertheless, in some specific situations, some research doubts the accuracy of the epidemic threshold value predicted using the QMF approach [8].

**2.2.4 K Spectral Energy** Because neither the MF, HMF, nor QMF approaches capture enough of the full contact network structure, the K Spectral Energy (KSE) approach is developed using  $A$ , node number, spectral radius, and energy of graph. The KSE approach is commonly used to study the spreading process over a network [11]. Here, the authors propose a discrete-time formulation to predict the epidemic threshold problem without any assumption of homogeneous connectivity. Hence, the epidemic threshold is given by equation (2.5):

$$\tau_c^{KSE} = \frac{kn}{E(G)} e^{-1/\lambda_{max}} \quad (2.5)$$

where  $E(G)$  is the graph energy, and  $k$  is a real parameter. The KSE prediction approach depends only on network structures.

As it happens, in the literature, there are many structural prediction approaches of the epidemic threshold. However, we are interested in the development of a new general structural and spectral prediction approach of the epidemic threshold. This should better capture the full network structure using network structural and spectral properties such as a node number, Laplacian matrix, spectral radius, and the Laplacian energy of graph. The new approach should be substantially similar to the previous structural approaches of prediction. Moreover, it should also be accurate. The new approach should offer a new general and spectral approach to analyse spreading processes in a network.

The adjacency matrix used in QMF [26], also in KSE [11] captures the network structure. However, the Laplacian matrix is closely related to the adjacency matrix but differs in some additional significant features that can also tell us much about network structure [18]. In the literature, a diffusion context takes into account some significant features, like the Laplacian matrix  $L$ . Hence, we can investigate the eigenvalues of  $L$  that include the Laplacian energy of graph.

### 3. Proposed approach for predicting the epidemic threshold

#### 3.1 Epidemic threshold and the spectral theory of graph

The spectral theory of graph and network science are used to understand how network structure predicts dynamic processes [10] such as the epidemic threshold. This theory analyses the relationships between the graph structure and its eigenvalues. It playing a key role in the fundamental understanding of networks [3, 5, 7]. However, a large literature on the spectral graph theory and its applications is available in several surveys, books, and monographs, including [5, 7].

3.1.1 *Largest eigenvalue of graph* The eigenvalues analysis allows to get useful information about a graph that might otherwise be difficult to obtain [7]. Indeed, eigenvalues have a strong relationship with the graph structures [5]. The largest eigenvalue of graph  $\lambda_1$  or  $\lambda_{max}$  is called the spectral radius of  $A$ .

#### 3.2 New structural and spectral prediction approach

Some research doubts the accuracy of prediction values by the QMF approach [8]. Also, neither the MF, HMF, QMF, nor KSE approaches capture enough of the full contact network structure; the *K Spectral Energy of Laplacian* approach is developed using the laplacian matrix  $L$ , node number, spectral radius, and the Laplacian energy of graph. For any undirected connected network, we focus to develop a new structural and spectral prediction approach of the epidemic threshold. However, in the epidemic threshold study, one of the challenges is to capture the essence of the full network structure with accuracy, connectivity, and flexibility. The new prediction approach does not assume homogeneous connectivity or any particular topology at a discrete time. We assume that during each time interval, an infected node  $i$  tries to infect its neighbours with probability  $\beta$ . At the same time,  $i$  may be cured with probability  $\gamma$ . So, the new epidemic threshold approach  $\tau_c^{KSEL}$  is given by equation (3.1):

$$\tau_c^{KSEL} = \frac{kn}{LE(G)} e^{-1/\lambda_{max}} \quad (3.1)$$

Here,  $LE(G)$  is the Laplacian energy of graph,  $n$  is the node number,  $\lambda_{max}$  is the spectral radius, and  $k$  is a specific real scale parameter.  $\tau_c^{KSEL}$  means *K Spectral Energy of Laplacian* prediction approach of

the epidemic threshold. Regarding the mathematical foundations of the new  $\tau_c^{KSEL}$  prediction approach of the epidemic threshold, we can show that  $\tau_c^{KSEL}$  is related to the context of diffusion processes like an epidemic on networks. In fact, diffusion processes are sometimes used as a simple model of spread over a network, such as the spread of an idea or an infectious disease [18]. This process is described as a process by which information, epidemics, viruses, and other processes spread over a network. It's known that a diffusion context takes into account some significant features, like the Laplacian matrix  $L$  [18] used in  $\tau_c^{KSEL}$ . In a simple undirected connected network, think about a quantity of substance  $\phi_i$  (heat) at each node  $i$  at time  $t$ , the diffusion of heat over a network is given by the equation (3.2) [18]:

$$\frac{d\phi_i}{dt} = c \sum_j A_{ij}(\phi_j - \phi_i) \quad (3.2)$$

In the matrix notation,

$$\frac{d\phi}{dt} + cL\phi = 0, \quad \phi(0) = \phi_0 \quad (3.3)$$

Whose solution is,

$$\phi_i(t) = \phi_i(0)e^{-c\lambda_i t} \quad (3.4)$$

An initial condition for the system can be specified by the quantity  $\phi(0)$ . We can solve the equation (3.4) for the state at any later time. Thus, the new prediction approach of the epidemic threshold  $\tau_c^{KSEL}$  is based on the solution of the diffusion equation (3.4). The epidemic threshold is closed to the diffusion context.

In the other point of view, related to the mathematical foundations of the new  $\tau_c^{KSEL}$  prediction approach, it is proven that the more highly connected a network is, the larger is  $\lambda_{max}$  [24], and the smaller is  $1/\lambda_{max}$  as an epidemic threshold. This can exhibit a basic exponential decay model  $\phi(t)$ , where:

$$\phi(t) = e^{\frac{-1}{\lambda_{max}} t} \quad (3.5)$$

with the single parameter  $\lambda_{max}$ .  $\phi(0) = 1$ . To consider each eigenvalue, we are interested to the Laplacian energy of graph according to its definition and diffusion features. So, about the fraction of the Laplacian energy of graph on each node, we define:

$$\Delta = \frac{LE(G)}{n} \quad (3.6)$$

According to the salient features like critical or threshold values of the reciprocal model, we consider the simple reciprocal model  $y = k(\frac{1}{x})$ , where  $x$  is a variable and  $k$  is a constant or scale parameter. Hence, the reciprocal of  $\Delta$  is:

$$k\left(\frac{1}{\frac{LE(G)}{n}}\right) = \frac{kn}{LE(G)} \quad (3.7)$$

Related to this reciprocal, the intuition to observe the rate of  $\phi(t)$  at  $t = 1$  leads to:

$$\frac{kn}{LE(G)} \times e^{-1/\lambda_{max}} = \tau_c^{KSEL} \quad (3.8)$$

Thus, the new prediction approach of the epidemic threshold  $\tau_c^{KSEL}$  is an application that associates each Laplacian matrix to a specific decay relative composition eigenvalues  $\mu_i$  relating to  $\Delta$ .

Furthermore, we can observe that the new  $\tau_c^{KSEL}$  has some specific structural and spectral features:

- (i) *Spectral*: by using parameter  $\lambda_{max}, LE(G)$  as a compound of different eigenvalues. This should further capture the full network structure.
- (ii) *Connectivity*: by using the Laplacian matrix  $L$  related to the clustering coefficient, average path length, degree distribution, connected components, and diffusion over a network.
- (iii) *Diffusion over network*: similar to the solution of heat diffusion process over the network that is  $a_i(t) = a_i(0)e^{-C\lambda_i t}$ , where  $C$  is a constant called the diffusion constant [18].
- (iv) *Flexibility*: by using of  $\{\lambda_{max}, LE(G), n, k\}$  structural and spectral parameters in strong relationship.
- (v) *General*: to a contact network of an undirected connected graph  $G$ .

#### 4. Experimentation, results and discussions

The simulations have been carried out to answer the question of how the new approach  $\tau_c^{KSEL}$  is substantially similar and performs in real a good performance than the previous prediction approaches, including the most commonly used QMF. This is performed through various data analytic, and data visualisation techniques on the experimental heterogeneous dataset in Fig. 1. Then, different sets of predicted values of MF, HMF, QMF, KSE, and the new KSEL prediction approach of the epidemic threshold are computed, analysed, visualised, and discussed. To evaluate the new approach, we achieve some qualitative, quantitative, and comparative analysis.

The dataset described in Fig. 1 contains various real networks of infectious disease spread, small-world, random, and regular networks in spreading processes overall. There are 31 different types and topologies of networks: 17 real social networks, 9 generated social networks, 3 random networks, and 2 regular random networks. Related to the network,  $Id$  refers to the identifier,  $kmax$  refers to the maximum node degree,  $k$  denotes the first moment of degree,  $k2$  the second moment of degree,  $den$  refers to the density, and  $cc$  the clustering coefficient.

Fig. 2 emphasises the shape distribution of each prediction approach of the MF, HMF, QMF, and KSE. Also, their common quantitative values that are distributed across the  $[0.005, 0.977]$  range. Those previous approaches share a similar peak around 0.106; this helps track their approximative central values and also displays where overall different values are concentrated over the interval  $[0.04, 0.16]$ . However, the shape of KSE shows a specific sharp peak compared to other approaches. This reflects some relative outlier values concentrated in  $[0.04, 0.16]$ . Hence, there must be some weaknesses and worrying prediction cases with the KSE approach; therefore, the new KSEL approach will overcome these weaknesses.

Figure 3 highlights how, at a given network  $Id$ , the prediction values of the epidemic threshold are similar to each other in relative amplitude. There is a strong relationship and few differences in predicted values. Overall, this emphasises similar or near-predicted values through different approaches. This shows no significant glance changes among different predicted values of the epidemic threshold.

For another visualization layout and features, Fig. 4 shows the similar trend and shape of different predicted values of the epidemic threshold using area visualization. The different areas overlap each



Id	Network	Type	Nodes	Links	kmax	k	k2	den	cc
0	Sociopatterns-infectious	Real social network	410	2765	50	13.488	252.434	0.032978	0.456
1	Airline	Real social network	36	57	20	3.167	33.389	0.090476	0.000
2	Internet	Real social network	40	61	10	3.050	13.000	0.078205	0.154
3	Karate club	Real social network	34	78	17	4.588	35.647	0.139037	0.571
4	Davis Southern Women	Real social network	32	89	14	5.562	39.062	0.179435	0.000
5	Florentine families	Real social network	15	20	6	2.667	8.933	0.190476	0.160
6	Les miserables	Real social network	77	254	36	6.597	79.532	0.086808	0.573
7	Watts Strogatz 1	Generated social network	1000	2000	10	4.000	17.898	0.004004	0.007
8	Watts Strogatz 2	Generated social network	3000	12000	18	8.000	67.741	0.002668	0.008
9	Connected Watts Strogatz 1	Generated social network	1000	2000	11	4.000	17.778	0.004004	0.008
10	Connected Watts Strogatz 2	Generated social network	3000	12000	16	8.000	68.055	0.002668	0.007
11	Newman Watts Strogatz 1	Generated social network	1000	3613	12	7.226	54.172	0.007233	0.157
12	Newman Watts Strogatz 2	Generated social network	5000	36028	23	14.411	211.592	0.002883	0.200
13	Newman Watts Strogatz 3	Generated social network	24	84	9	7.000	49.667	0.304348	0.498
14	Barabasi Albert	Generated social network	1000	4975	150	9.950	211.636	0.009960	0.044
15	Barbell	Generated social network	1005	1010	5	2.010	4.066	0.002002	0.005
16	Random 1	Generated network	1000	3500	17	7.000	56.174	0.007007	0.009
17	Random 2	Generated network	140	6811	111	97.300	9495.843	0.700000	0.700
18	Dense gnm Random	Generated network	1000	3500	16	7.000	56.016	0.007007	0.009
19	Random regular 1	Generated network	1000	1500	3	3.000	9.000	0.003003	0.003
20	Random regular 2	Generated network	1000	4500	9	9.000	81.000	0.009009	0.007
21	Facebook 1	Real social network	52	146	18	5.615	48.692	0.110106	0.462
22	Facebook 2	Real social network	61	270	29	8.852	109.705	0.147541	0.733
23	Facebook 3	Real social network	168	1656	77	19.714	645.321	0.118050	0.534
24	Facebook 4	Real social network	150	1693	57	22.573	680.240	0.151499	0.670
25	Facebook 5	Real social network	333	2519	77	15.129	469.526	0.045570	0.508
26	Facebook 6	Real social network	224	3192	99	28.500	1312.554	0.127803	0.544
27	Facebook 7	Real social network	534	4813	107	18.026	539.884	0.033820	0.544
28	Facebook 8	Real social network	786	14024	136	35.684	2086.852	0.045458	0.476
29	Facebook 9	Real social network	1034	26749	253	51.739	4886.236	0.050086	0.526
30	Facebook 10	Real social network	747	30025	293	80.388	10593.861	0.107759	0.635

Figure 1. The summary of structural information about networks in the dataset

other. There is a strong relationship between the different predicted values of the epidemic thresholds MF, HMF, QMF, KSE, and the new KSEL. There is a particular similar trend, a common shape, and a strong correlation between the commonly used QMF, KSE, and the new KSEL prediction approach. This highlights the similarity of the different prediction approaches, including the new KSEL. Therefore,

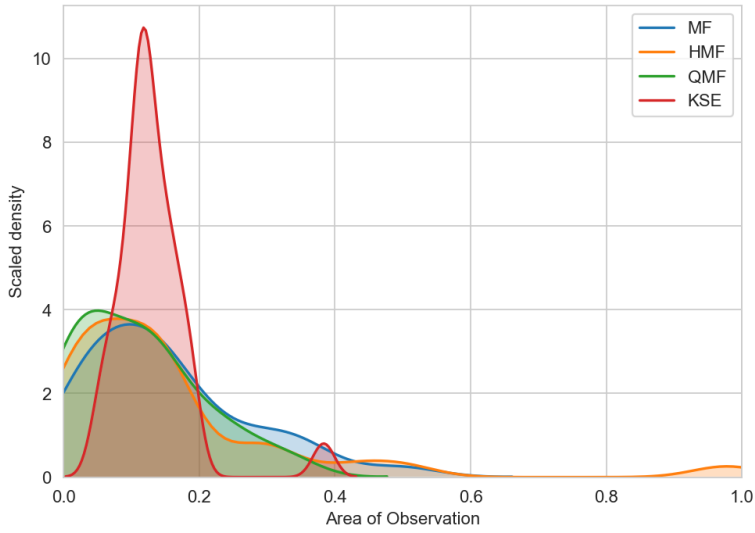


Figure 2. The shape and value distribution of MF, HMF, QMF and KSE prediction approach of the epidemic threshold, using scaled density visualization

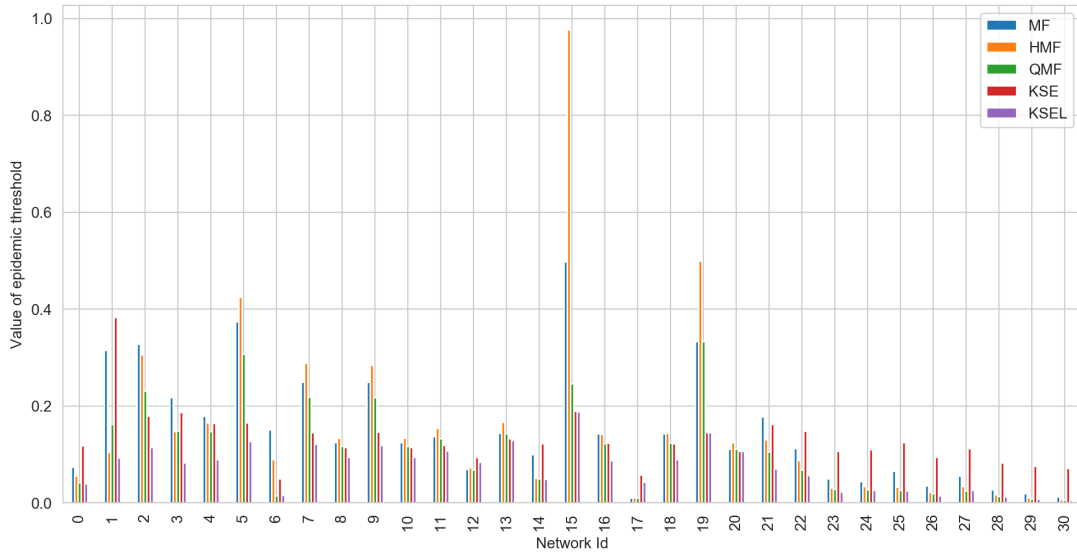


Figure 3. The relative comparison and relationship between MF, HMF, QMF, KSE and the new KSEL prediction approach of the epidemic threshold, using bar visualization

the sets of these prediction approaches are related to each other.

Moreover, Fig. 5 emphasises that the quantitative predicted values are distributed across a common

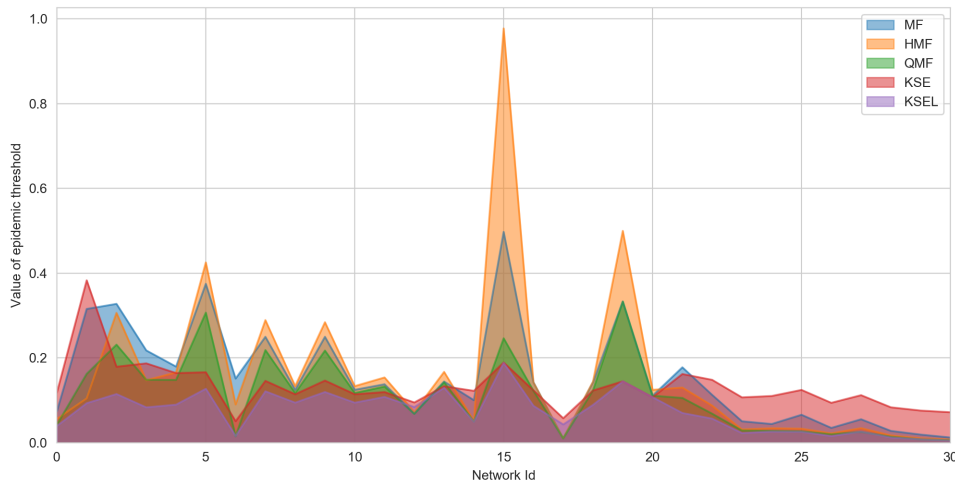


Figure 4. The trend and correlation on MF, HMF, QMF, KSE and the new KSEL prediction approach of the epidemic threshold, using area visualization

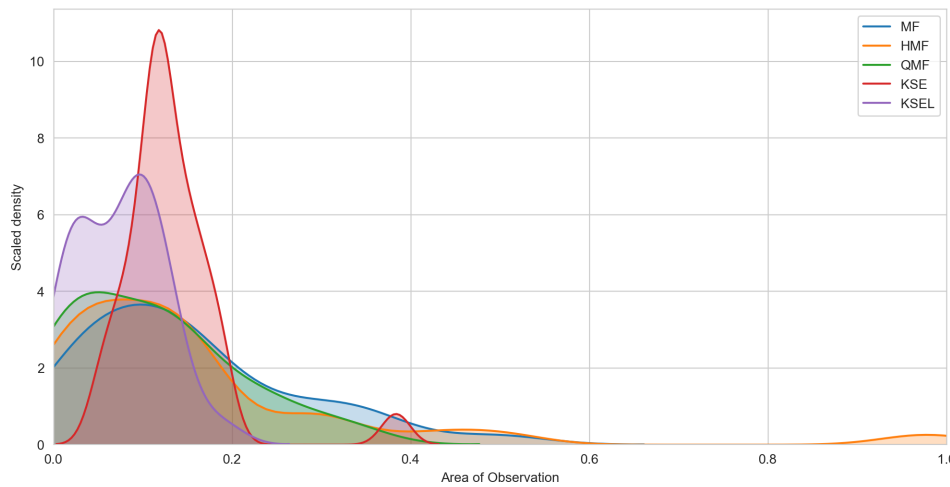


Figure 5. The shape and value distribution of MF, HMF, QMF, KSE and the new KSEL prediction approach of the epidemic threshold, using scaled density visualization

quantitative [0.005, 0.977] range. Every predicted value of the epidemic threshold shares a similar peak relatively around 0.106. This helps to track their approximative central values and to display where overall different predicted values are concentrated over the interval [0.03, 0.16]. Overall, there is a common shape and value distribution that overlap each other too much between [0.03, 0.16] that characterises the relatively similar peak, central values, and value distribution of the different prediction approaches.

Using descriptive analysis techniques, the Fig. 6 provides a concise summary of descriptive statistics that highlights some main features of a dataset. There are particular similar patterns and strong

relationships between predicted values using QMF, KSE, and KSEL.

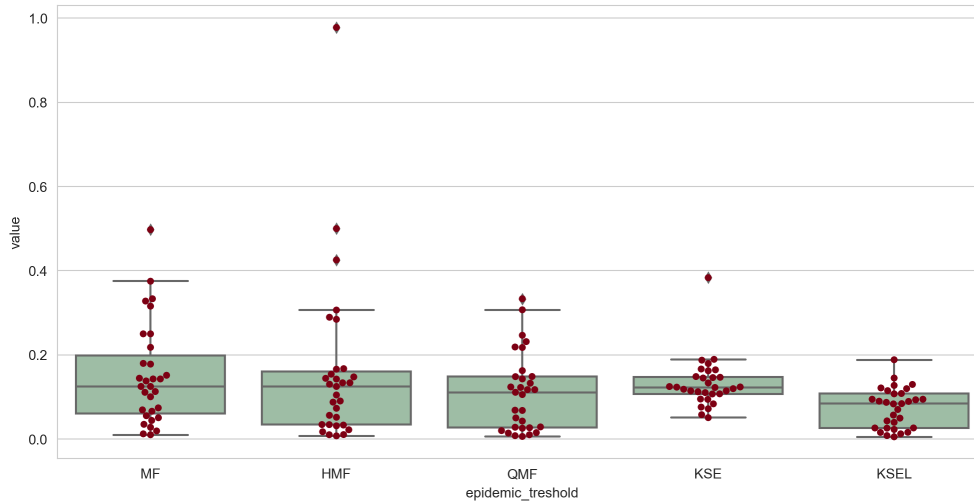


Figure 6. The box plots visualization of MF, HMF, QMF, KSE and the new KSEL prediction approach of the epidemic threshold

Concerning QMF as a common-use approach in the literature, Fig. 7 emphasises a relative comparison and strong relationship; Fig. 8 shows the overall similar rolling trend and relative pattern; and Fig. 9 shows the common shape and value distribution of QMF, KSE, and the new KSEL prediction approach. The predicted values of the KSEL approach are more similar to those of QMF compared to the predicted values of the KSE approach related to the QMF approach. Overall, of these 3 structural and spectral approaches, there is an emphasis on an overall rolling particular similar trend, a common shape, a relative curve, and a common large range of value distribution over predicted values by different network Id. This highlights no significant glance changes among these different prediction approaches. So, the results above mean that the new KSEL approach is substantially similar and shares major features related to the previous structural prediction approaches, specifically with the widely used, accurate QMF. Theoretically, those results come from the eigenvalues concept at the root of the QMF, KSE, and KSEL approaches. At the conceptual level, the new KSEL approach offers an insight into a relationship between three important and useful concepts in graph spectral theory: the spectral radius, Laplacian energy of graph, node number, and a specific scale parameter.

Furthermore, we can analyse an overview of the gaps or differences between previous prediction approaches of the epidemic threshold related to the new *KSEL*. The summary of the descriptive statistics of these gaps is shown in Table 2. Here, for any  $p, q$  epidemic threshold approach,  $e_{p,q}$  means the Euclidian gap or difference of  $p$  to  $q$ :  $p - q$ .

In Table 2, the standard deviation of the gap between the most used QMF and KSEL is  $0.0530$ ; the gap between the most used QMF and KSE is  $0.0544$  as all the gaps are relatively low. Here, the term relatively low is related to the previous prediction approaches, particularly lowest with the most commonly used QMF and KSE. Once again, this underlines the fact that the new KSEL prediction approach is similar to the previous. It shares major common features with the previous approach, specifically with the most accurate QMF, KSE.

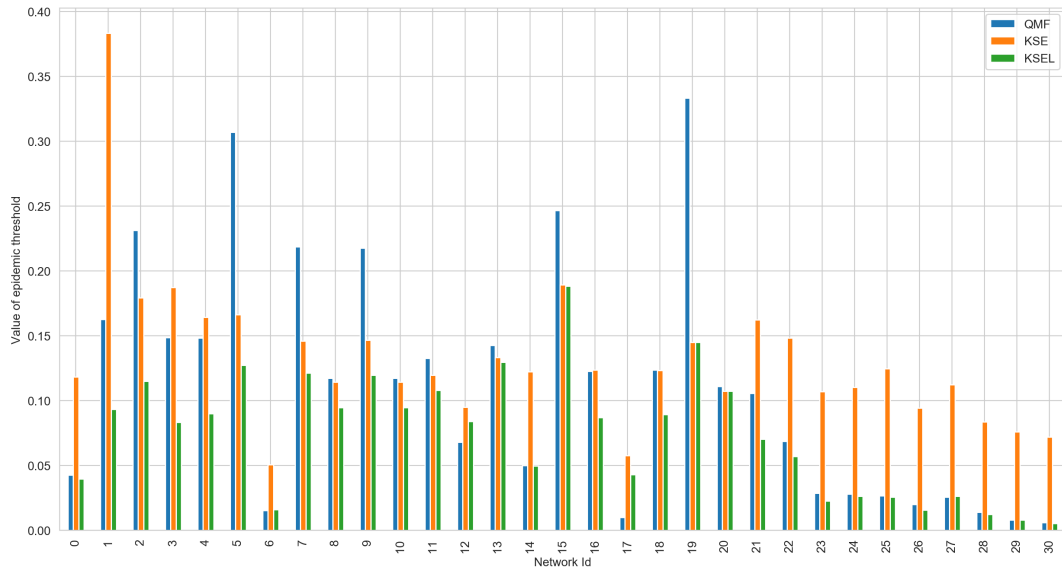


Figure 7. An emphasis on the relative comparison and relationship between QMF, KSE and the new KSEL prediction approach of the epidemic threshold, using bar visualization

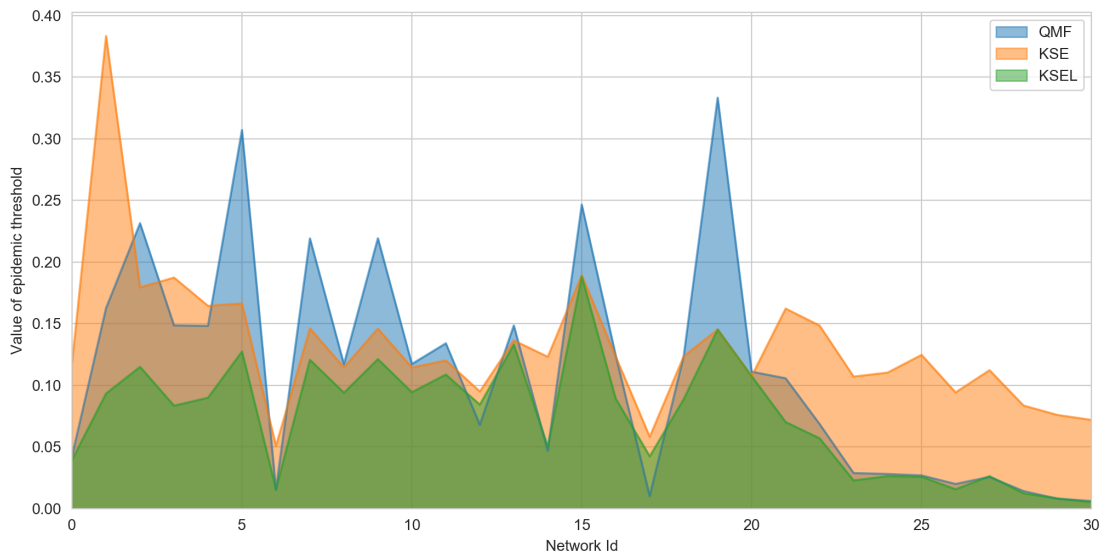


Figure 8. An emphasis on the trend and correlation on QMF, KSE and the new KSEL prediction approach of the epidemic threshold, using area visualization

Besides those precedent results, we analyse the Pearson’s correlation coefficient between MF, HMF,

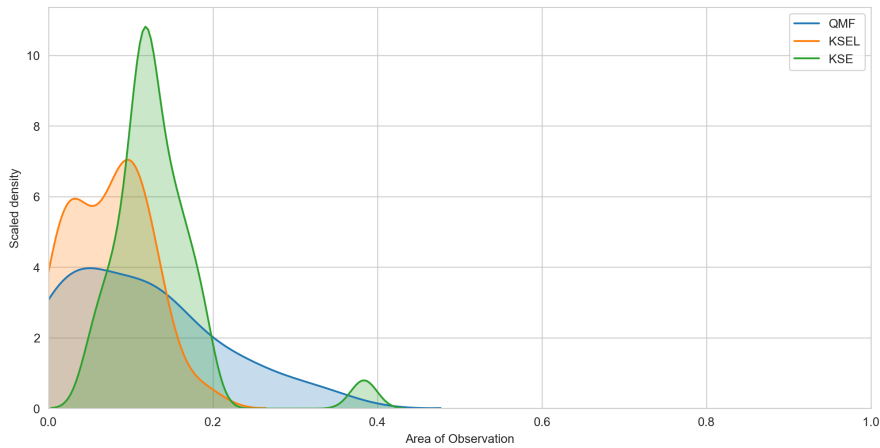


Figure 9. An emphasis on the shape and value distribution of QMF, KSE and the new KSEL prediction approach of the epidemic threshold, using scaled density visualization

Table 2. Summary of the descriptive statistic values of the gaps or differences between MF, HMF, QMF, and KSE approach related to the new KSEL prediction approach of the epidemic threshold

	e_MF_KSEL	e_HMF_KSEL	e_QMF_KSEL	e_KSE_KSEL
count	31.0000	31.0000	31.0000	31.0000
mean	0.0770	0.0834	0.0355	0.0579
std	0.0845	0.1579	0.0530	0.0544
min	-0.0314	-0.0313	-0.0314	-0.0000
25%	0.0186	0.0076	0.0014	0.0201
50%	0.0401	0.0307	0.0116	0.0644
75%	0.1292	0.0692	0.0582	0.0813
max	0.3091	0.7893	0.1881	0.2899
IQ	0.1106	0.0616	0.0569	0.0612
range	0.3404	0.8206	0.2195	0.2899

QMF, KSE, and KSEL.

Table 3. The Pearson’s matrix correlation coefficient between MF, HMF, QMF, KSE and the new KSEL prediction approach of the epidemic threshold, with emphasis on the density and clustering coefficient

	MF	HMF	QMF	KSE	KSEL	den	cc
MF	1.000						
HMF	0.885	1.000					
QMF	0.902	0.796	1.000				
KSE	0.655	0.346	0.534	1.000			
KSEL	0.838	0.818	0.902	0.485	1.000		
den	-0.204	-0.200	-0.198	-0.126	-0.171	1.000	
cc	-0.583	-0.505	-0.674	-0.404	-0.728	0.502	1.000

In Table 3, the new KSEL is more strongly and positively correlated to the most used QMF, even to the previous prediction approaches, including the KSE in [11]. Particularly, the correlation between KSE and QMF is 0.5340, for 0.9020 between KSEL and QMF. All the prediction approaches, including the new KSEL, are strongly and positively correlated. Moreover, by definition, an epidemic threshold in a contact network should be negatively correlated with density and clustering coefficient. Table 3 highlights that all prediction approaches of the epidemic thresholds, including the new KSEL, are significantly correlated with a negative sign to the density (den). Also, Table 3 highlights that all prediction approaches, including the new KSEL, are significantly correlated with a negative sign to the clustering coefficient; this is with a greater negative value of  $-0.7280$  for the new KSEL approach.

Additionally, using ANalysis Of VAriance (ANOVA), we analyse the statistical difference among different experimental sets of MF, HMF, QMF, KSE, and KSEL predicted values of the epidemic threshold. We have used the univariate ANOVA test through the Ordinary Least Squares (OLS) model and the Bioinfokit Python package. Then, Table 4 summarizes the output of ANOVA  $F$  and  $p$ -value, where  $sum_{sq}$  denotes the sum of squares,  $df$  denotes the degree of freedom,  $F$  the F-statistic, and  $PR$  the P-value.

Table 4. The ANOVA  $F$  and  $p$ -value using the Ordinary Least Squares to MF, HMF, QMF, KSE related to KSEL prediction approach of the epidemic threshold

	df	sum_sq	mean_sq	F	$PR(> F)$
C( $\tau$ )	4.0000	0.1431	0.0358	2.7161	0.0320
Residual	150.0000	1.9755	0.0132	nan	nan

In Table 4,  $p$ -value  $0.0320 < 0.05$ . Thus, there are statistically significant differences overall in different sets of predicted values of the epidemic threshold. However, related to the previous results that showed similar features between MF, HMF, QMF, KSE, and KSEL, particularly from QMF, KSE, and KSEL, we are interested in studying why the overall  $p$ -value  $0.0320 < 0.05$ . This leads to an examination of multiple pairwise differences (post hoc comparison) analysis for all unplanned comparisons using Tukey's honestly significantly differenced (HSD) test summarizes in Table 5.

Table 5. The Tukey's honestly significantly differenced (HSD) using Bioinfokit Python package to MF, HMF, QMF, KSE, KSEL prediction approach of the epidemic threshold

ID	group1	group2	Diff	Lower	Upper	q-value	p-value
0	MF	HMF	0.0065	-0.0740	0.0870	0.3136	0.9000
1	MF	QMF	0.0415	-0.0390	0.1220	2.0124	0.5991
2	MF	KSE	0.0191	-0.0614	0.0996	0.9259	0.9000
3	MF	KSEL	0.0770	-0.0035	0.1575	3.7346	<b>0.0683</b>
4	HMF	QMF	0.0479	-0.0325	0.1284	2.3261	0.4723
5	HMF	KSE	0.0255	-0.0549	0.1060	1.2396	0.9000
6	HMF	KSEL	0.0834	0.0029	0.1639	4.0482	<b>0.0381</b>
7	QMF	KSE	0.0224	-0.0581	0.1029	1.0865	0.9000
8	QMF	KSEL	0.0355	-0.0450	0.1160	1.7221	0.7147
9	KSE	KSEL	0.0579	-0.0226	0.1384	2.8086	0.2779

In Table 5, except the pairwise (MF, KSEL) and (HMF, KSEL), all other pairwise predicted values of the epidemic threshold *accept* the null hypothesis since  $p$ -value  $\in [0.2779, 0.9000]$  is greater than  $0.10$ . This indicates globally that there is "not significant" statistical difference between different predicted values set of the epidemic threshold, except for pairwise (MF, KSEL) and (HMF, KSEL).

**The potential advantages and benefits of the KSEL new approach compared to the previous**

We established an analytical comparative study in Table 6. Here, the term *relatively* is related to the context and dataset of this study. This term refers to possible suggestive theoretical interpretations or missing formal proofs. Moreover, contextually in Table 6, the criteria *accuracy* refers to the quality to capture the full network structure; *transparency* is the quality to assess the rule and function of each parameter in the formula, even the assessment of the parameters in a relationship; *flexibility* refers to the ability to change or be real scale easily; and *parameter* refers to the quality of parameter(s), their number, and also their meaning in a relationship.

Table 6. The potential advantages and benefits of the new approach related to the previous: a qualitative comparison between MF, HMF, QMF, KSE and the new KSEL prediction approach of the epidemic threshold

Model	Accuracy	Transparency	Flexibility	Parameter
MF	Relatively poor fit: network structure isn't considered.	Relatively easy: single parameter $\langle k \rangle$ .	Relatively poor: due to its assumptions.	The use of a single parameter $\langle k \rangle$ .
HMF	Relatively poor fit: due to its parameters can be inaccurate.	Relatively medium: can assess the role of $\langle k \rangle, \langle k^2 \rangle$ .	Relatively medium: due to its assumptions.	The use of 2 parameters $\langle k \rangle, \langle k^2 \rangle$ .
QMF	Relatively medium fit: captures network structure using only $\lambda_{max}$	Relatively easy: due to it single parameter $\lambda_{max}$ .	Relatively good: due to its assumptions.	The use of a single parameter $\lambda_{max}$ .
KSE	Relatively high fit: captures the full network structure using $\{\lambda_{max}, E(G), n, k\}$ .	Relatively medium: parameter assessment in relationships can be complex.	Relatively improved: due to its assumptions, using $\{\lambda_{max}, E(G), n\}$ and a scale $k$ .	The use of $\{\lambda_{max}, E(G), n, k\}$ structural and spectral parameters in <i>relationship</i> .
KSEL	Relatively high fit: captures the full network structure using $\{\lambda_{max}, LE(G), n, k\}$ .	Relatively medium: parameter assessment in relationships can be complex.	Relatively better: due to its assumptions, using $\{\lambda_{max}, LE(G), n\}$ and a scale $k$ .	The use of $\{\lambda_{max}, LE(G), n, k\}$ structural and spectral parameters in strong relationship.

Nevertheless, no model or approach is perfect; the new *KSEL* can have interesting potentials and an attractive balance in terms of accuracy, transparency, and flexibility.

Overall experimentation, we use various data analysis and data visualisation techniques to analyse and evaluate the new structural and spectral prediction approach of the epidemic threshold *KSEL*. Therefore we deliver various views and filters for specific analysis, visualisation, and comparison of *KSEL* related to the previous structural prediction approaches. As it happens, the results show that



KSEL is similar and shares a common rolling trend, shape, curves, and range of distribution values related to the previous structural approaches, specifically more similar with the most accurate QMF [25]. KSEL constitutes a new general and spectral area to analyse the spreading processes in a network using the network's structural and spectral properties. Besides, concerning the results in [11, 25, 26], these results contribute to highlight the relationship between network-based models and mathematical modelling approaches, which could be useful for computing or predicting the epidemic threshold  $\tau$  and the reproduction number  $R_0$ . This highlights the usefulness of a network-based structural approach for the prediction of some key epidemiological parameters such as  $\tau$ ,  $R_0$ .

## 5. Conclusion

In the context of the spreading process, we address the problem of epidemic threshold prediction over a contact network. Network structures fundamentally influence spreading processes with boundary conditions like the epidemic threshold. To address this issue and improve prediction approaches, a new structural and spectral prediction approach called *KSEL* has been designed and experimented. The new approach is based on the node number, spectral radius, and Laplacian energy of graph. KSEL has been analysed and evaluated at theoretical, formal, and experimental levels using some qualitative and quantitative analysis and visualization techniques on a large heterogeneous dataset. The results show that KSEL is similar and shares a common rolling trend, shape, and range of distribution values related to the previous prediction approaches, specifically more similar to the most used QMF. KSEL performs really good performance compared to the previous approaches, including the most-used QMF. It further captures the full network structure, connectivity, network diffusion properties, and flexibility but is not limited over there. Furthermore, we observe a high positive and strong correlation between KSEL and the previous prediction approaches, as well as a negative correlation between different prediction approaches, including KSEL and the network density, or the clustering coefficient. The ANOVA analysis shows that all the different sets of epidemic threshold predicted values *accept* the null hypothesis. Therefore, KSEL is a new approach to predicting the epidemic threshold. It offers a new structural and spectral approach to analyse spreading processes in a network. Pragmatically, these results can have a practical interest to improve the control and prediction of spreading processes in networks. Particularly meaningful to decision-makers in public health who can use these results to improve the control of an infectious disease spread in order to advise policy for successful mitigation strategies. Future research could examine the full dynamic network constraints in the epidemic threshold prediction.

## REFERENCES

1. Alshahrani, M., Fuxi, Z., Sameh, A., Mekouar, S. & Huang, S. (2020) Efficient Algorithms based on Centrality Measures for Identification of Top-K Influential Users in Social Networks. *Information Sciences*, **527**.
2. Biggs, N. (1993) Algebraic Graph Theory. *Cambridge University Press*. (2nd ed.), Cambridge.
3. Bogachev, V. & Smolyanov, O. (2020) *Spectral Theory. Chapter of Real and Functional Analysis, Moscow Lectures*, pages 279–356.
4. Chen, Q. & Porter, M. A. (2021) Epidemic thresholds of infectious diseases on tie-decay networks. *Journal of Complex Networks*, **10**(1).
5. Chung, F. & on Recent Advances in Spectral Graph Theory (1994 : California State University (Fresno)), C. C. (1997) *Spectral graph theory*. CBMS-NSF regional conference series in mathematics, no. 92. Conference Board of the Mathematical Sciences.
6. Coscia, M. (2021) Pearson correlations on complex networks. *Journal of Complex Networks*, **9**(6).
7. Cvetkovic, D., Doob, M. & Sachs, H. (1980) Spectra of Graphs – Theory and Application. New York.
8. Ferreira, S., Castellano, C. & Pastor-Satorras, R. (2012) Epidemic thresholds of the Susceptible-Infected-Susceptible model on networks: A comparison of numerical and theoretical results. *Physical Review E, Statistical, nonlinear, biological, and soft matter physics*, **86**.

9. Gutman, I. (1978) The energy of a graph. *Ber. Math. Statist. Sect. Forschungsz. Graz*, **103**, 1–22.
10. Gutman, I. & Ramane, H. (2020) Research on graph energies in 2019. *MATCH Communications in Mathematical and in Computer Chemistry*, **84**, 277–292.
11. Kanyou, C., Kouokam, E. & Emvudu, Y. (2023) Epidemic threshold : A new spectral and structural approach of prediction. *Revue Africaine de Recherche en Informatique et Mathématiques Appliquées (ARIMA)*.
12. Keeling, M. & Rohani, P. (2011) *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
13. Kephart, J. & White, S. (1991) Directed-Graph Epidemiological Models of Computer Viruses.. *In Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 343–359.
14. Kresge, K. & Petruzelli, N. (2021) Analyzing Epidemic Thresholds on Dynamic Network Structures. *SIAM Undergraduate Research Online*, **14**.
15. Lang, J. C., De Sterck, H., Kaiser, J. L. & Miller, J. C. (2018) Analytic models for SIR disease spread on random spatial networks. *Journal of Complex Networks*, **6**(6), 948–970.
16. Mieghem, P. V. & Bovenkamp, R. V. d. (2015) Accuracy criterion for the mean-field approximation in susceptible-infected-susceptible epidemics on networks. *Physical Review E, Statistical, nonlinear, biological, and soft matter physics*, **91**.
17. Nevin, J., Groth, P. & Lees, M. (2023) An approach for analysing the impact of data integration on complex network diffusion models. *Journal of Complex Networks*, **11**(4).
18. Newman, M. E. J. (2010) *Networks: an introduction*. Oxford University Press, Oxford; New York.
19. Niranjana, U., Singh, A. & Kumar Agrawal, R. (2019) A mean-field-theoretic model for dual information propagation in networks. *Journal of Complex Networks*, **7**(4), 585–602.
20. Pastor-Satorras, R. & Vespignani, A. (2001) Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, **86**, 3200–3203.
21. Pellis, L., Ball, F., Bansal, S., Eames, K., House, T., Isham, V. & Trapman, P. (2015) Eight Challenges for network epidemic models. *Epidemics*, **10**, 58–62.
22. Prakash, B. A., Chakrabarti, D., Valler, N. C., Faloutsos, M. & Faloutsos, C. (2012) Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks. *Knowledge and Information Systems*, **33**, 549–575.
23. Prasse, B. & Mieghem, P. V. (2020) Time-dependent solution of the NIMFA equations around the epidemic threshold. *Journal of mathematical biology*, **81**.
24. Tinkler, K. J. (1972) The Physical Interpretation of Eigenfunctions of Dichotomous Matrices. *Transactions of the Institute of British Geographers*, **55**, 17–46.
25. Wang, W., Tang, M., Stanley, H. E. & Braunstein, L. (2016) Unification of theoretical approaches for epidemic spreading on complex networks. *Reports on Progress in Physics*, **80**.
26. Wang, Y., Chakrabarti, D., Wang, C. & Faloutsos, C. (2003) Epidemic spreading in real networks: An eigenvalue viewpoint. *22nd International Symposium on Reliable Distributed Systems (SRDS 2003). Proceedings*, pages 25–34.