A method based on Graph Theory and Three Way Decisions to evaluate critical regions in epidemic diffusion:

An analysis of COVID-19 in Italy

Angelo Gaeta[1] · Vincenzo Loia[1] · Francesco Orciuoli[1]

Abstract

The paper reports the results of an analysis of COVID-19 diffusion in Italy. The analysis was carried out with a new method based on the combined use of a 3 Way Decisions model and graph theory. Specifically, the data about infected people in the Italian regions is assessed by means of an evaluation function which allows the tri-partitioning of Italy and the identification of high, medium or low critical regions. The tri-partition is performed, along the temporal evolution of the COVID-19 diffusion, by calculating two threshold values which take into account the containment actions that, from time to time, the decision makers have implemented. The effects of a containment action are related to a reduction in the centrality value of a region. To estimate the effect of containment actions, we evaluated two approaches. The first is based on a uniform reduction in the centrality values of the regions, the second estimates the effects of containment actions starting from the mobility changes data provided by the Google Community Mobility reports. The results of our evaluation based on real data of the COVID-19 diffusion in Italy are encouraging and represent a good starting point for future extensions of the method.

1 Introduction

In Rome, on January 31, 2020 two tourists from China had symptoms of COVID-19. About a month later, on February 21, 2020, a first outbreak with about 16 cases was found in Lombardia, in Codogno in the province of Lodi. Once the first outbreak was discovered, 11 municipalities in northern Italy (in Lombardia and Veneto) have been quarantined. In the following periods, other actions have been taken by decision-makers such as the total closure of the municipalities with active outbreaks, the suspension of demonstrations and events in the municipalities themselves. These restriction measures become progressively more articulated and extended gradually to the whole national territory. However, also in light of these restrictive measures, COVID-19 has had a significant diffusion and a dramatic impact in Italy.

1.1 Contribution of the paper and its distinctive aspects.

This paper presents and evaluates a new analysis method that allows to improve the comprehension of the COVID-19 diffusion in Italy by taking into account the relative importance of each region in the country and the actions taken to contain the epidemic diffusion.

The analysis method supports decision-makers in evaluating the spread of COVID-19 taking into account that Italy is divided into regions with very different characteristics, such as different levels of productivity and mobility, which can contribute to spread the virus in a different way. Therefore, our goal is to try to analyse the diffusion at national level considering these peculiarities as well as the effect of containment actions taken by decision-makers.

✉ Angelo Gaeta
agaeta@unisa.it

Vincenzo Loia
loia@unisa.it

Francesco Orciuoli
forciuoli@unisa.it

1   Dipartimento di Scienze Aziendali - Management and Innovation Systems (DISA-MIS), Università degli Studi di Salerno, Via Giovanni Paolo II, 132, 84084, Fisciano, Italy

The common approaches used to describe the virus spread are based on the calculation of the number of infected, deceased and cured people (e.g., see the map of the Johns Hopkins Institute[1] or the similar dashboard from the Italian Civil Protection department[2]) and on indicators that take into account the daily variations of data compared to the number of swabs. National data can be obtained through the aggregation of data at regional level and the latter can be obtained starting from a lower level of granularity, such as cities.

The analysis that is carried out is therefore punctual and takes into account the actions taken by decision-makers indirectly.

To better understand this fact let us give an example. If a region has decided for a containment action related to the closure of some production activities on day x, this action is reflected in a lower movement of people which leads to a reduction of infected people after a few days (usually, $x + 14$).

Now, to determine whether the containment action within the specific region has been effective, we can intuitively compare the number of infected people with the number of total cases. If the difference between these two values increases, then in the period of containment actions the spread of the virus has been reduced.

Replicating this method at the nation level is not so simple as the specific regions have a different importance within the country system. It is clear that a lock-down of a low productivity region affects the country system quite differently from the lock-down of a high productivity region. In comparing the number of infected people with that of the total cases, it will be necessary to take into account how much the total cases of the individual regions weight to carry out reasoning at the country system level.

This is where the method proposed in this paper differs from other approaches.

Fundamental to our analysis is the concept of *criticality* or *critical level* of a region. The criticality of a region, in our model, is a measure of how much that region can contribute to the spread of the virus both within it and among other regions of the country. Therefore, this measure takes into account not only the data relating to the spread of the epidemic (such as, for example, the number of infected people) but also how central a region is in the system in light of the current containment actions and restrictive measures applied.

An example can help to understand. Let us consider a region A with a high number of infected people and which has applied significant restrictive measures, such as voluntary quarantine. Region B, on the other hand, has a smaller number of infected but has not applied restrictive measures. Establishing whether region A is more or less

critical than B is not trivial. Intuitively one might think that region A, due to the application of restrictive measures, may be less critical than B where, in the absence of such measures, the virus could continue to circulate and spread throughout the country system. This consideration, however, does not take into account the centrality value of the regions within the country system: if A, in fact, is a region with a high productivity rate, it could be more critical than B due to the interactions (albeit limited by the restrictive measures) with the other regions. So, a low or high number of infected during a specific period may characterize a region as more or less critical depending on other factors.

The solution we propose to this problem consists in establishing threshold values that allow us to evaluate the criticality of the regions, and combines measures of centrality of the graphs, effects of the actions to contain the spread of the virus and, finally, data relating to the evolution of epidemic in individual regions.

1.2 Motivations and limitations

The motivations behind our work are related to the need of defining and evaluating a method of analysis to better understand the situation concerning to the spread of COVID-19 at national level. The model we present in the paper is able to:

1. include in the analysis phase the effects of containment actions that the government and specific regions have undertaken;
2. consider the importance of a specific region within the country system.

The proposed method is based on modeling the country system as a graph and on the use of approximate reasoning mechanisms, such as evaluation-based Three-Way Decisions. The perspective of our study is typical of descriptive analytics and the method develops a form of Three-Way decision-making that leverages graph theory centrality measures to establish reference thresholds allowing a decision-maker to assess whether a region is more or less critical for the spread of the virus in the country system.

As reported in Section 2, there are other works combining Three-Way Decisions and Graph Theory even if for different purposes. For our application, the motivation behind this combination is to be able to consider the *heterogeneity* of the different Italian regions in evaluating the contribution they can provide to the spread of the virus in the country system. Based on this assumption of heterogeneity, which is realistic in Italy where very productive regions and economically depressed regions coexist, the method helps decision-makers to identify the regions that represent a greater risk for the spread of the

virus at the national level, also considering the impact that containment measures can have on reducing the centrality of these regions[3].

Hence, the method does not intend to propose a model for predicting or estimating the spread of the virus. The aim is to provide decision-makers with a better awareness of the epidemic situation at the national level considering the estimated effects of the containment actions, the current state of the epidemic spread and the centrality of specific Italian regions. As mentioned, we start from the assumption that Italy includes heterogeneous regions and therefore there exist conditions for which a region has a higher capacity to worsen the epidemic situation than another region (and also with respect to the overall country system). These conditions are substantiated in the fact that a region A dominates over at least another region B both in terms of epidemic values and in relation to its centrality in the country (represented as a graph). As a conservative assumption, we consider that this occurs when the minimum of the values describing the epidemic trend and the centrality of a region is higher than the maximum of the analogous values of another region.

This implies, as a limitation of the proposed method, its inadequacy to assess the criticality of the regions when these have homogeneous epidemiological and centrality values or do not show significant variations between region and region.

1.3 Organization of the paper

The paper is organized as follows. Section 2 presents an overview of related works. Section 3 gives an overview of the analysis method. Section 4 presents the graph-based modeling used to determine the importance of Italian region. The importance is given by the *centrality* of the region. Section 5 describes the Three-Way Decisions (3WD) evaluation approach. Section 6 describes the overall method with the help of pseudo-code. Section 7 reports the experimentation results and the evaluation on real data related to the diffusion of COVID-19 in Italy and a brief discussion. Section 8 presents conclusions and draws future works.

2 Related works

This section presents an overview of: *i)* 3WD and some research results combining 3WD and Graph Theory and *ii)* research works focused on COVID-19 to clarify the differences with our method.

---

[3]In this regard, we mention that the Italian government has made use of a tri-partition of the regions, based on 21 indicators, to manage the so-called second wave in Italy. This allows to take into account the differences between the different regions and to implement restrictions and action plans with regional granularity.

2.1 3WD and graphs

3WD [1] is an effective method to make a rapid decision. A model based on 3WD is able to mimic a particular way of human decision-making process that is based on a trisecting-and-acting model [2]. Usually, a 3WD model is based on two tasks: a division of the universal set into three pairwise disjoint regions and the definition of actions or strategies to act upon the objects of the three regions. The three regions are referred to as positive (POS), negative (NEG) and boundary (BND). In evaluation-based models, an evaluation function is defined for the objects in the universe and the result of the evaluation is compared with two thresholds, α and β. If the evaluation is greater than or equal to α the object is classified in the POS region, if the evaluation is lesser than or equal to β the object is classified in the NEG region, otherwise it is classified in the BND region. In literature, it is possible to recognize evaluation functions based on probabilistic rough sets [3], fuzzy neighbourhood covering functions [4], dominance relations and their extensions such as variable precision based [5].

Recently, the model has been generalized into a trisecting-acting-outcome (TAO) model [6], which takes into consideration the outcome. In brief, in the TAO model, a third aspect is introduced that is related to the evaluation of the effectiveness of both trisection result and strategy.

3WD has been investigated also in combination with Graph Theory. For instance, in [7], 3WD has been studied in the context of social networks and, specifically, location-based social networks to define a model of Proximal Three-Way decision-making. In [8, 9] and [10] 3WD models have been applied to the problem of community detection. In [8], the authors leverage on weighted graphs representations to take advantages of the global structure of the network. In [9] and [10], instead, authors take advantage of the local features of the network.

To some extent, the method we propose in this paper approaches [7] as we also try to take advantage of social network measures to make decisions. In our work, however, we use a centrality measure to estimate the relative influence of nodes in a social network.

2.2 COVID-19

Since data relating to the COVID-19 epidemic has been made available, many research works have been published.

Some refer to the study and modeling of the epidemiological and clinical characteristics of patients affected by the virus, (e.g., [11]), to gain comprehension of clinical course and risk factors (e.g., [12]), estimation and prediction of mortality (e.g., [13]).

Some others refer to the adoption of time series and autoregressive integrated moving average (ARIMA) models

to forecast epidemic evolution (e.g., [14, 15] and [16] for a case study on Italy). Also the adoption of Computational and Artificial intelligence to detect cases and predict the COVID-19 diffusion is gaining attention (e.g., [17–19]).

The method presented in this paper has a different purpose. Our objective is to support the decision-makers in gaining an improved awareness of COVID-19 diffusion in Italy by evaluating the criticality of regions taking into account also the effects of containment actions.

With respect to this last aspect, there are some recent results relating to the COVID-19 diffusion in Italy with the implementation of Governmental containment measures. In [20], authors examine and analyse the effect of containment measures for the diffusion of COVID-19 in Italy. Their work is based on epidemic models that include e.g. regional individual mobilities, the progression of social distancing, and local capacity of medical infrastructure. A similar work is proposed in [21] where a mathematical model (SIDARTHE) predicts the course of the epidemic to help plan possible scenarios of implementation of countermeasures. Of course, similar works are available also for other national case studies (e.g., for the U.K. [22]).

The method we propose in this paper differs from that one previously mentioned. We do not leverage on epidemic or mathematical models but, instead, on a graph modeling of Italy and on the adoption of approximate reasoning to evaluate the criticality of the Italian regions. Containment Actions, in our case, affect the reasoning method and are not part of a model. Obviously, in our case, the lack of mathematical modeling of the epidemic implies an approximate assessment of the criticality of the regions which, however, can be useful as a first indication on the regions that (based on current data and an estimate of the effects of containment) deserve more or less attention in relation to their criticality.

3 The analysis method

The method combines evaluation-based 3WD with graph theory and it is shown in Fig. 1

As done by other approaches to describe the diffusion of COVID-19, we gather data at the regional level. However, at the national level, we not only consider an aggregation procedure but use an approximate reasoning mechanism, i.e. 3WD evaluation, to understand what are the most critical regions of Italy with respect to the diffusion of the virus.

We model Italy as a Graph and employ a *centrality measure* to estimate the relative importance of regions considered as nodes, u, of the graph. The importance of a region is a function of the level of productive activities and mobility within the region and with other

Fig. 1 Overview of the analytic
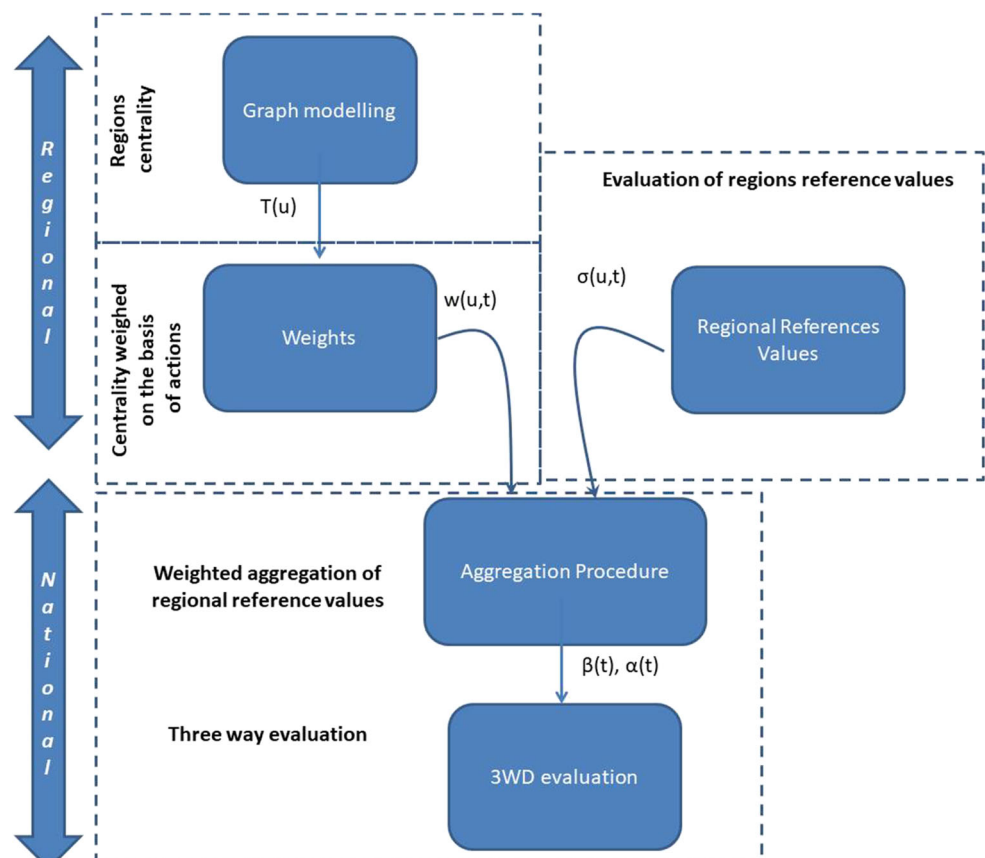
regions. This importance value does not remain constant during the evolution of the epidemic phenomenon. As previously mentioned, the value is modified by the effects of containment actions (such as the reduction of production activities) taken to contain the spread of the virus. So, on a given day t, our method considers a measure of centrality weighted for the specific containment action, $w(u, t)$. For each region we derive also some reference values, $\sigma(u, t)$. These values serve to understand if, on a given day t and for a given region u, there has been an increase in people positive for the virus.

The weights, $w(u, t)$, and reference values, $\sigma(u, t)$, are used to derive two thresholds, α and β, for the 3WD. The idea behind evaluation-based 3WD is to use an evaluation function, $E(u, t)$, to evaluate the infected people for each individual region u on a given day t. We compare the results of E with the thresholds α and β s.t. $1 \geq \alpha > \beta \geq 0$. If $E(u, t) \geq \alpha$ then region u is critical on day t, if $E(u, t) \leq \beta$ then u is not critical on day t, otherwise we should defer the decision.

In the following Sections 4 and 5 we detail the phases of graph modeling and 3WD evaluation. The overall procedure is formalized in the algorithm of Section 6.

4 Graph-based modeling

The system Italy is modeled as a weighted directed graphs $G = (U, E)$ where nodes in the set U represent regions and edges E represent connections among regions. Every edge $(i, j) \in E$ and $i, j \in U$, has a weight $d(i, j)$. U contains all Italian regions which are studied in the analysis proposed by such work. Moreover, also foreign regions (e.g., China) belong to U and although they are excluded from the analysis, they are used to compute some values useful for the evaluation of critical levels in connected Italian regions. Fig. 2 provides an example, of such modeling, focused on the *Lombardia* region. Firstly, the reader can observe that the example, in the aforementioned figure, uses two types of edges, the first one (black) connecting regions with their geographical neighbors and, the second one (grey) connecting regions having relevant logistical links for commerce, tourism, etc.

The goal of such modeling is to calculate an importance value $T(u)$, for each region u, representing how much the region can contribute to the diffusion of COVID19 according only to mobility due to the productivity level of the region. The idea is to obtain such value through a *centrality* measure. In particular, the selected measure is the *Katz Centrality* [23], typically used for estimating the relative influence of actors in a social network:

$$T(u) = \gamma \sum_{j} d(u, j) T(j) + \phi_u \tag{1}$$

In the (1), the *importance* of region u is proportional to the sum of the *importance* values of its neighbors (geographical and logistical) j. Additional parameters γ and ϕ_u represent respectively a balance parameter and a non-network factor characterizing region u with respect to its intrinsic importance depending on the inhabitants. γ and ϕ_u (for all u) have to be chosen opportunely with the aim at considering more important the first term of the (1) than the second one or vice versa. The edge weights are calculated by considering two factors related to *gross domestic product*
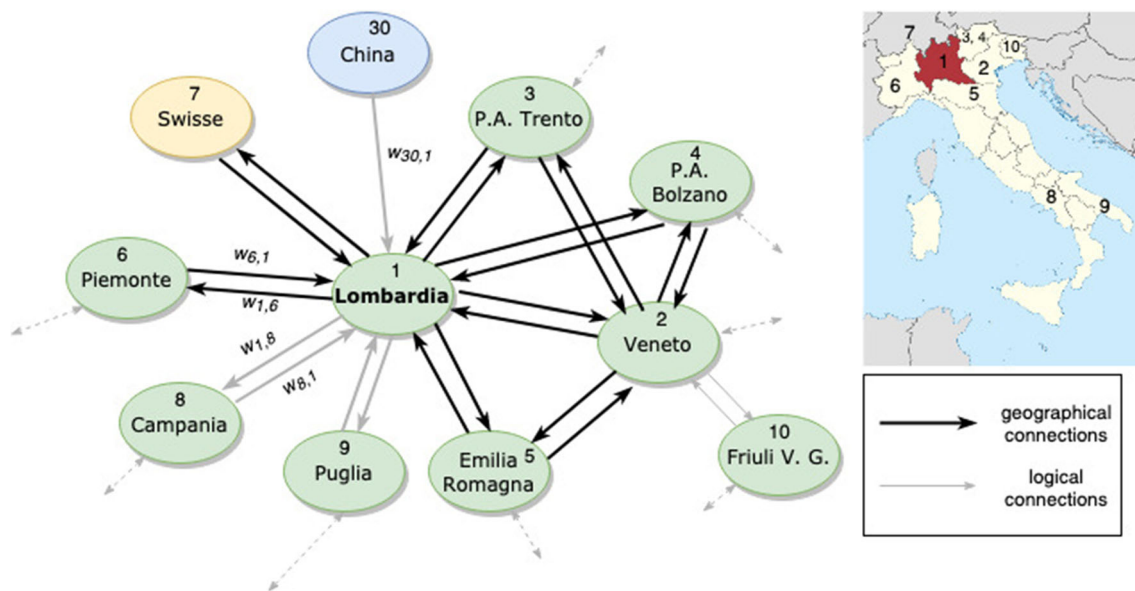

Fig. 2 Graph modeling: sample

(GDP). A possible way to calculate the weight of the edge between node u (source) and node j (destination) is this:

$$d(u, j) = \frac{g_j}{g_u}, \qquad (2)$$

where g_u is the estimate of the GPD in region u and g_j is the estimate of the GDP in region j. Such weight approximates the measure of the flow of people traveling from u to j on business, by assuming that business travelers go from the place with lower GDP to the place with higher GDP.

5 3WD evaluation

First of all, it is needed to calculate the thresholds α and β to tri-partitioning the Italian regions, with respect to their criticality, by using the 3WD. The difficulty consists in the fact that these two thresholds must be representative of heterogeneous regions which have different importance in the country system and face different epidemic situations. Thus, the idea is to use appropriate aggregation operators allowing to consider the aforementioned aspects.

In order to take into account the *different epidemic situations* of the regions, we need to derive some reference values. Let us consider a region u on a day t and define: $\sigma(u, t) = \frac{TC(u,t)}{Swabs(u)}$ where $TC(u, t)$ is the number of total cases (including: positive people, healed and deceased) of the region u on day t and $Swabs(u)$ is the number of total swabs executed in the region u. High values of $\sigma(u, t)$ could indicate a severe epidemic situation in a region which, however, must be compared with the number of currently positive people in the region. In fact, the total number of cases may include people who tested positive for COVID-19 some time ago and not yet clinically cured. To this purpose, let us define $E(u, t) = \frac{AP(u,t)}{Swabs(u)}$ where $AP(u, t)$ is the number of currently positive people of the region u on day t. It is observed that $0 \le E(u, t) \le 1$ and $E(u, t) \le \sigma(u, t)$. If on a given day t, $E(u, t)$ tends to $\sigma(u, t)$, this indicates that the number of current infected on that day contributes significantly to the total number of cases. Hence, in this case there is a significant increase in positive cases in a region u that is due to the diffusion of the virus. Therefore, for our purpose, the values of $\sigma(u, t)$ can be considered as regional references to establish a level of criticality when compared with the actual positive people.

To account for the *different importance* of a region in the country, we have to consider the effect of containment actions. Having modeled Italy as a graph in which regions are nodes, the idea is to start from the Katz centrality values, $T(u)$, of the Italian regions and derive some weights, $w(u, t)$, that include the effects of containment actions which are taken in specific regions (e.g., a full or partial closure of some commercial activities, prohibition of inter-regional travels). These actions tend to reduce the centrality values of the regions. In fact, these actions make regions less important with respect to the diffusion at the national level since, as result of a containment action, productivity and mobility within the region and between regions decreases. So $w(u, t) = CA(u, t) \times T(u)$ where $CA(u, t)$ indicates a percentage of reduction of the centrality of region u due to the specific containment action.

Once we have derived the values of $\sigma(u, t)$ and $w(u, t)$ for all regions, we can aggregate them to compute α and β. We use two aggregation operators named Weighted Max and Weighted Min to evaluate α and β respectively[4]:

$$\alpha = WMAX(\sigma(u_i, t), w(u_i, t)) = \vee_{i=1}^{|U|}(w(u_i, t) \wedge \sigma(u_i, t)) \quad (3)$$

$$\beta = WMIN(\sigma(u_i, t), w(u_i, t)) = \wedge_{i=1}^{|U|}(w(u_i, t) \vee \sigma(u_i, t)) \quad (4)$$

where $|U|$ is the cardinality of the set of the regions.

Equation (3) tends to aggregate towards the highest of the values of $\sigma(u_i, t)$ which weigh more. In fact, when a weight $w(u_i, t)$ is "small" (i.e., lower than the corresponding value of $\sigma(u_i, t)$), this element will be included in the subsequent search for the max and discarding, therefore, the corresponding value of $\sigma(u_i, t)$. Conversely, when the weight $w(u_i, t)$ is higher than the corresponding value of $\sigma(u_i, t)$, the weight will be excluded from the search for the max in favor of the corresponding value of $\sigma(u_i, t)$.

With the same reasoning, we understand that (4) tends to aggregate towards the smallest values of $\sigma(u_i, t)$ that weigh less.

We can now apply 3WD. Let U be our universal set consisting of all the Italian regions, $u \in U$ is an Italian region and t is a given day, $E(u, t) = \frac{AP(u,t)}{Swabs(u)}$ an evaluation function, and α and β two thresholds, we define the following three subsets:

$$POS(U) = \{u \in U | E(u, t) \ge \alpha\}$$
$$BND(U) = \{u \in U | \beta < E(u, t) < \alpha\} \quad (5)$$
$$NEG(U) = \{u \in U | E(u, t) \le \beta\}$$

It is possible to establish the criticality of each region as follows: if $u \in POS(u)$ than u is critical on day t, if $u \in NEG(u)$ than u is not critical on day t, otherwise u presents a medium critical level on day t.

α and β are good threshold indicators at the national level for our objectives. In fact, α will tend (upwards or downwards, depending on the value of the weights) towards the maximum reference value $\sigma(u_i, t)$ of the most important regions. For this reason, if we exceed the value of α, the classification in the POS area is correct. Similarly, β will approach the minimum reference values of the less

---

[4]We have followed the Weighted Maximum and Minimum implementation of the agop R package: https://cran.r-project.org/web/packages/agop/index.html

important regions and, therefore, we are sure that if the evaluation function is lesser than α, the region can be classified in the NEG area.

6 Formalization of the method

Algorithm 1 describes the method using pseudo-code:

Algorithm 1 3WD with graphs.

Input: A graph: $G(U, E)$, The parameters: $\gamma, \phi_u,$
$\quad\quad CA(u, t)$
Result: A tri-partition: $POS(U), BND(U), NEG(U)$
for $u \leftarrow 1$ **to** $|U|$ **do**
$\quad$ **for** $j \leftarrow 1$ **to** $|U|$ **do**
$\quad\quad$ Evaluate weights on edges: $d(u, j) = \frac{g_j}{g_u}$;
$\quad$ **end**
end
Apply Katz Centrality algorithm to evaluate network centrality: $T(u) = \gamma \sum_j d(u, j) T(j) + \phi_u$;
for $u \leftarrow 1$ **to** $|U|$ **do**
$\quad$ Include containment actions:
$\quad$ $w(u, t) = CA(u, t) \times T(u)$;
$\quad$ Calculate reference values: $\sigma(u, t) = \frac{TC(u,t)}{Swabs(u)}$;
end
Evaluate α as Weighted Max:
$\alpha = WMAX(w(u, t), \sigma(u, t))$;
Evaluate β as Weighted Min:
$\alpha = WMIN(w(u, t), \sigma(u, t))$;
for $u \leftarrow 1$ **to** $|U|$ **do**
$\quad$ Evaluate E(u,t);
$\quad$ **if** $E(u, t) \geq \alpha$ **then**
$\quad\quad$ $u \leftarrow POS(U)$
$\quad$ **end**
$\quad$ **else if** $E(u, t) \leq$ β **then**
$\quad\quad$ $u \leftarrow NEG(U)$
$\quad$ **end**
$\quad$ **else**
$\quad\quad$ $u \leftarrow BND(U)$
$\quad$ **end**
end

---

7 Experimentation and evaluation

The method has been experimented on real data on the COVID-19 diffusion in Italy[5] from the Civil Protection Department. Before presenting the evaluation results, we are going to briefly describe the scenario, the data, the graph centrality values of the regions and the approach used to estimate $CA(u, t)$.

7.1 Scenario

The evaluation was carried out over three time windows corresponding to the application of different restrictive decrees by the Italian government:

– February 25, 2020 - March 8, 2020. This time window corresponds to the implementation of the so-called "red areas' Decree which implemented restrictions (e.g., reduced mobility) in some municipalities where epidemic outbreaks were present, mainly in the regions of Lombardia and Veneto. Other regions, however, have also taken steps to reduce mobility.
– March 9, 2020 - March 25, 2020. This time window corresponds to the implementation of the so-called "I stay at home" Decree which implemented restrictions in all Italian regions.
– March 26, 2020 - April 9, 2020. This time window corresponds to the implementation of the so-called "Close Italy" Decree which implemented further restrictions and reduced activities and services in all Italian regions to only essential ones.

7.2 Data

As mentioned before, the evaluation has been done on real data related to the COVID-19 diffusion in Italy. The data are provided by the Civil Protection Department of the Italian government and inform on the daily trend of COVID-19 at different levels of granularity: national, regional and provincial.

The data used in our evaluation informs about the regional trend, and is described with 20 attributes. The attributes related to the evolution of the epidemic concern, among the others, hospitalised patients with symptoms, in intensive care and home confined; total and actual positive; total tests (i.e., swabs made).

The dataset is available on github[6].

7.3 Graph and Centrality values

Figure 3 shows the graph model and the centrality values, $T(u)$, of the regions. As mentioned in Section 4 we included in the modeling phase also regions outside Italy, mainly those with the highest trade.

The regions with the highest centrality value are shown at the edges of the figure. It is not surprising that the highest value is relative to Lombardia, the most productive region of
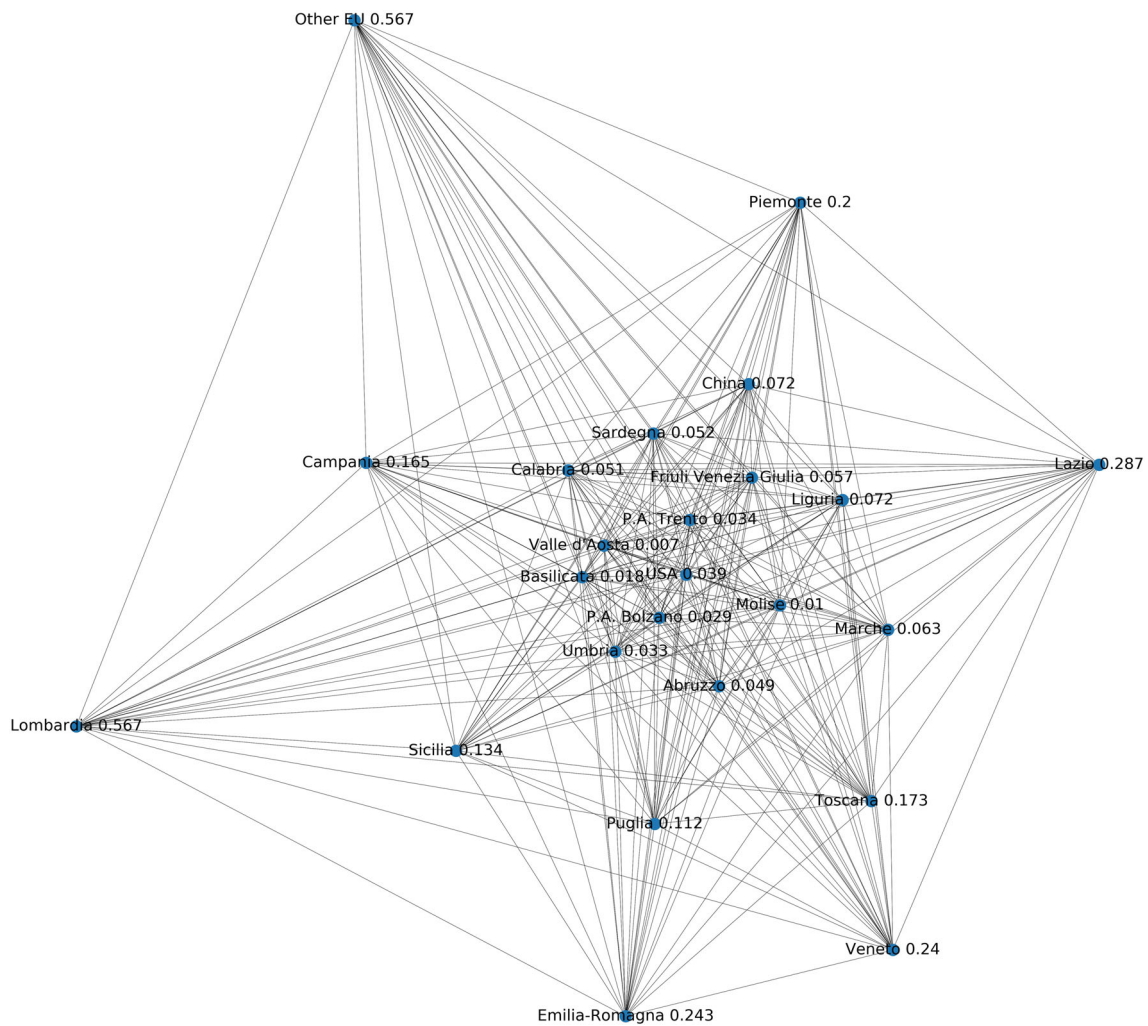

Fig. 3 Graph and Centrality values

Italy, followed by other regions of the northern area of the country (such as Veneto, Emilia-Romagna, Piemonte). As far as the Center-South is concerned, Lazio and Campania are highlighted in the figure.

7.4 Estimation of the effect of containment actions

The estimation of $CA(u, t)$ is critical in our model. We recall that $CA(u, t) \in [0, 1]$ is a factor that reduces the Centrality value, $T(u)$, of a region. This reduction derives from the implementation of containment actions. In other words, $CA(u, t)$ has to quantify the effect on the region u of a containment action on date t.

As mentioned earlier, it is critical because it determines the weights, $w(u, t)$ that are used to establish the threshold values, α and β, which allow regions to be partitioned.

We evaluated two cases.

In the first case we have estimated the $CA(u, t)$ for the regions as follows:

– equal to 0.6 for the Lombardia and Veneto regions and 0.9 for the others during the "Red Areas" Decree.
– equal to 0.5 for all the regions during the "I Stay at home" Decree.
– equal to 0.2 for all the regions during the "Close Italy" (lock-down) Decree.

This estimate tends to replicate the increase in the level of restrictions that the three decrees have entailed.

The second case uses data from Google COVID-19 Community Mobility Reports[7]. Google understood that their data (e.g., used in products such as Google Map) could be helpful to make critical decisions to combat COVID-19, and produced and made available Community Mobility Reports to provide insights into what has changed

---

[7]Google LLC "Google COVID-19 Community Mobility Reports". https://www.google.com/covid19/mobility/ Accessed: 01/09/2020

Table 1 Estimation of $CA(u, t)$ based on Google Mobility data

| Region	25-feb	08-mar	25-mar	09-apr
Molise	1.016	0.878	0.278	0.366
Sardegna	1.122	0.96	0.32	0.378
Piemonte	0.898	0.872	0.276	0.346
Veneto	0.912	0.904	0.292	0.356
Marche	1.13	0.88	0.264	0.348
Friuli Venezia Giulia	0.912	1.044	0.3	0.366
Lombardia	0.858	0.812	0.242	0.304
Lazio	1.02	0.908	0.252	0.31
Sicilia	1.042	0.788	0.214	0.326
Campania	0.972	0.828	0.24	0.328
Calabria	1.022	0.772	0.23	0.332
Valle d'Aosta	1.27	0.87	0.268	0.316
P.A. Bolzano	1	0.96	0.238	0.29
P.A. Trento	1	0.96	0.238	0.29
Umbria	1.062	0.95	0.3	0.384
Basilicata	1.012	0.824	0.288	0.392
Abruzzo	1.086	0.858	0.254	0.346
Liguria	0.942	0.922	0.272	0.338
Toscana	1.074	0.916	0.27	0.34
Emilia-Romagna	0.952	0.9	0.282	0.356
Puglia	1.034	0.818	0.27	0.338
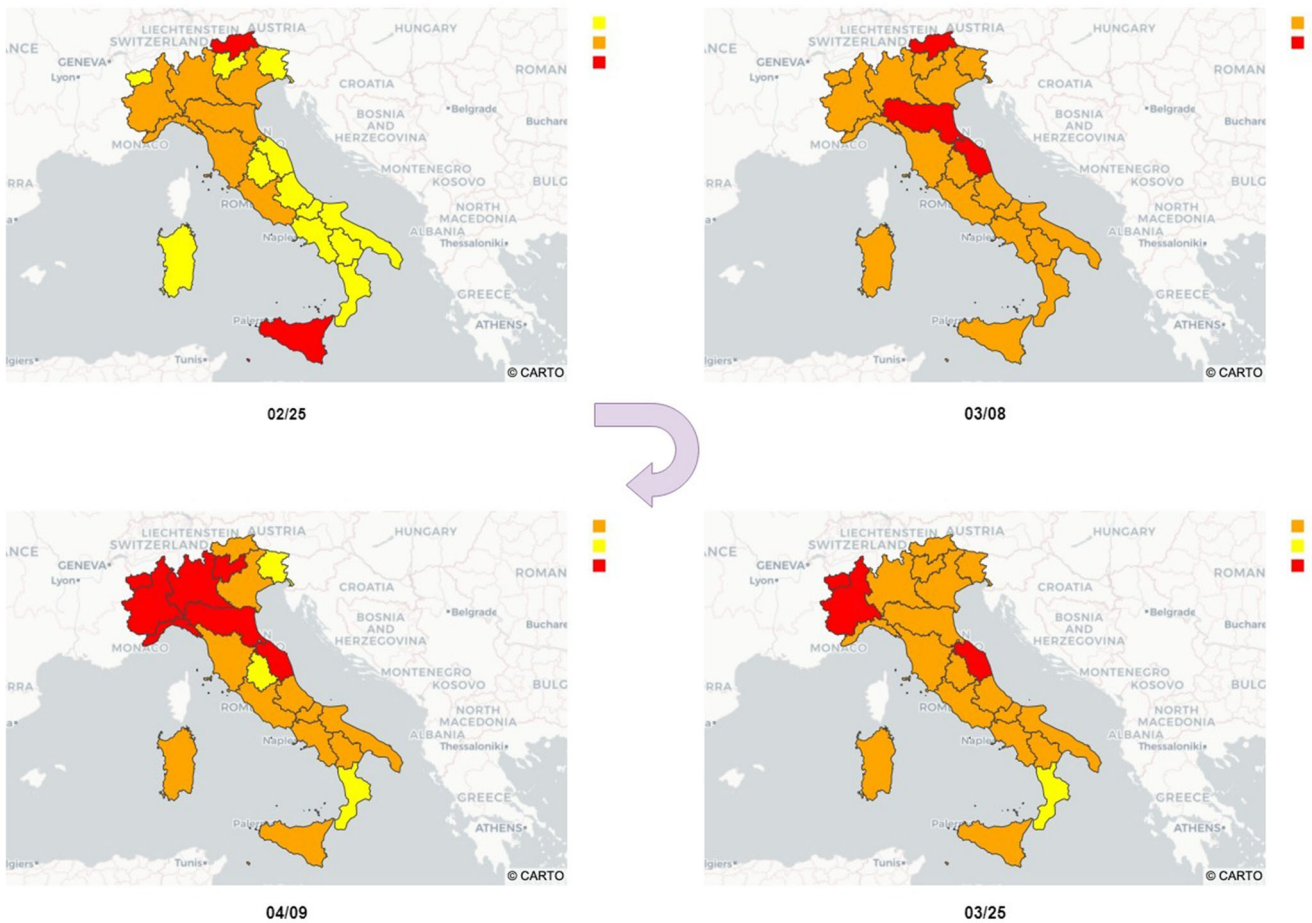
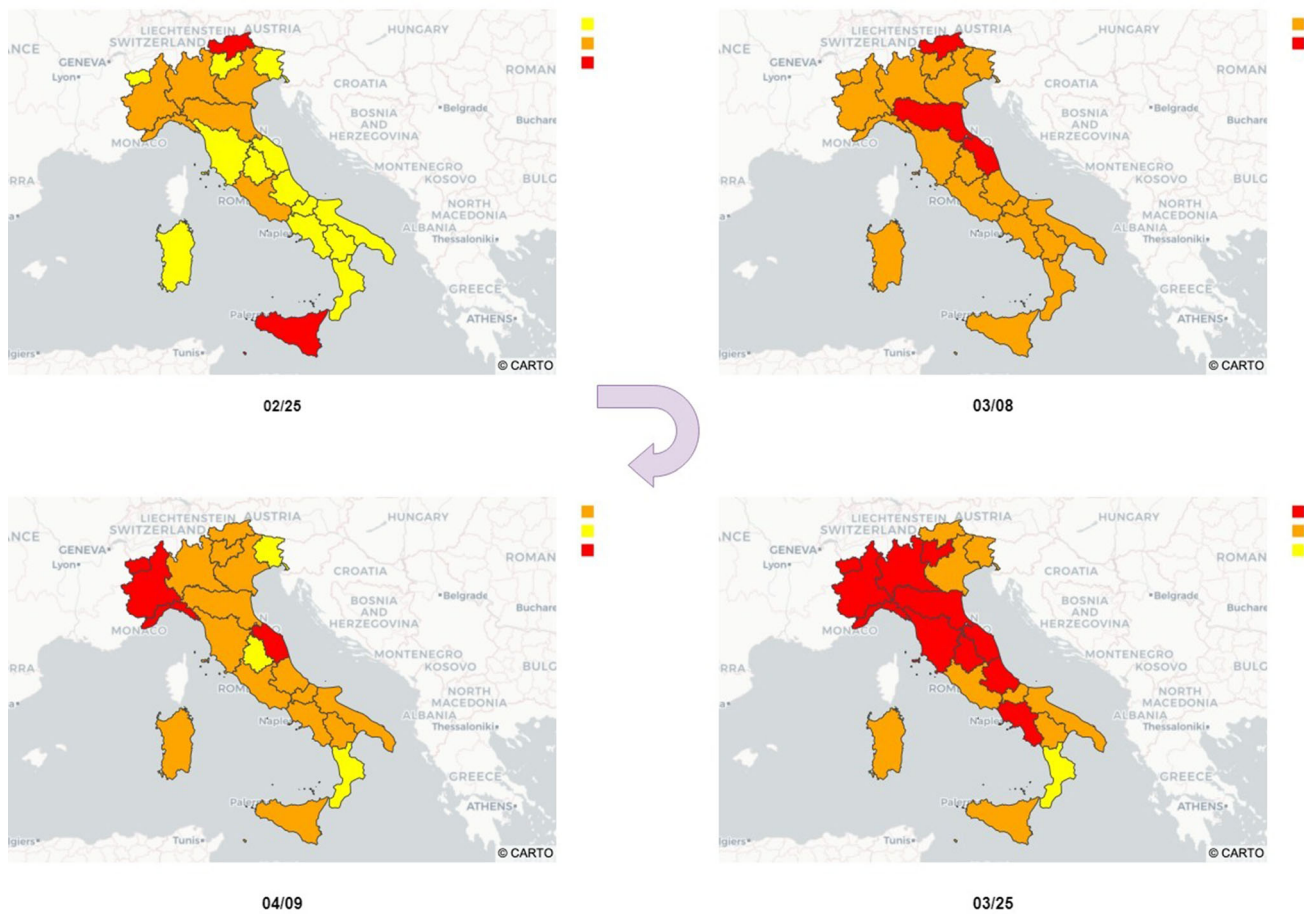**Fig. 4** Results of 3WD - Flat Case

Fig. 5 Results of 3WD - Google Case

in response to policies aimed at combating COVID-19. The data behind these report is also available[8].

The data shows how visits to places, such as grocery stores and parks, are changing in each geographic region. The places included in the dataset are: grocery and pharmacy; parks; transit stations; retail and recreation; residential; workplace. For each day and region, the dataset provides the mobility variations related to these places. Changes are established according to a baseline.

For region u on date t, we estimated $CA(u, t)$ as the mean of mobility changes of all places except residential. Residential mobility, admitted for particular activities and limited periods even during the lock-down, within a single municipality or in the vicinity of the place of residence does not have significant effects in relation to the reduction of the centrality, $T(u)$ of a region in our model.

The values of $CA(u, t)$ estimated from Google mobility data are shown in Table 1.

In the following, to differentiate the two cases, we refer to the first as "Flat Case" and the second as "Google Case".

7.5 Results

The results of 3WD are shown in Figs. 4 and 5 that correspond to the two cases for estimation of $CA(u, t)$: Flat and Google. In Figs. 4 and 5, Red regions are POS classes, Orange regions are BND classes and Yellow regions are NEG classes. The names of the Italian regions are reported in Fig. 6. The maps have been produced with plotly[9] and CartoDB[10].

The Flat Case of Fig. 4 shows how many regions start to move from the NEG zone to the BND zone with an increase of their criticality during the Red Areas decree (i.e., from the 02/25 to 03/08). From the 03/09 to the 03/25, we can observe that the actions of the "I stay at home" decree on the one hand led to the stabilization of the Calabria region (which will henceforth always be present in NEG areas) but on the other hand they have not managed to contain the spread in many regions of Italy which remain at medium-high critical level. With the implementation of the "Close Italy" decree, from 03/26 to 04/09, it is observed that the

[8]https://www.google.com/covid19/mobility/data_documentation.html?hl=en

[9]https://plotly.com/python/

[10]https://github.com/CartoDB/carto-python

BND area tends to become empty and other regions, mainly in central and northern Italy, tend to become POS and NEG. This indicates a strong variation in this period and for these regions in their level of criticality due to the strong fluctuation of the virus in such regions.

Looking at Fig. 5 that reports the Google Case, we can observe a different trend for the periods related to "I stay at home" and "Close Italy" decrees. In fact, the POS area, first, tends to increase during the "I stay at home" (from 03/09 to 03/25) and, after, tends to reduce during the "Close Italy"

Table 2 Correct classification and model prediction for case 1

Region	g	Correct classification	Model prediction (03/08) - Flat case	Model prediction (03/08) - Google case
Abruzzo	0.108	BND	BND	BND
Basilicata	0.033	NEG	BND	BND
P.A. Bolzano	0.229	POS	POS	POS
Calabria	0.081	BND	BND	BND
Campania	0.103	BND	BND	BND
Emilia-Romagna	0.271	POS	POS	POS
Friuli Venezia Giulia	0.058	NEG	BND	BND
Lazio	0.044	NEG	BND	BND
Liguria	0.182	BND	BND	BND
Lombardia	0.212	POS	BND	BND
Marche	0.264	POS	POS	POS
Molise	0.121	BND	BND	BND
Piemonte	0.235	POS	BND	BND
Puglia	0.057	NEG	BND	BND
Sardegna	0.074	BND	BND	BND
Sicilia	0.061	NEG	BND	BND
Toscana	0.123	BND	BND	BND
P.A. Trento	0.102	BND	BND	BND
Umbria	0.162	BND	BND	BND
Valle D'Aosta	0.265	POS	BND	BND
Veneto	0.048	NEG	BND	BND

Table 3 Correct classification and model prediction for case 2

Region	g	Correct classification	Model prediction (03/25) - Flat case	Model prediction (03/25) - Google case
Abruzzo	0.149	BND	BND	POS
Basilicata	0.148	BND	BND	BND
P.A. Bolzano	0.112	BND	BND	BND
Calabria	0.066	NEG	NEG	NEG
Campania	0.159	BND	BND	POS
Emilia-Romagna	0.210	BND	BND	POS
Friuli Venezia Giulia	0.101	BND	BND	BND
Lazio	0.084	NEG	BND	BND
Liguria	0.289	POS	BND	POS
Lombardia	0.262	BND	BND	POS
Marche	0.350	POS	POS	POS
Molise	0.106	BND	BND	BND
Piemonte	0.349	POS	POS	POS
Puglia	0.130	BND	BND	BND
Sardegna	0.139	BND	BND	BND
Sicilia	0.118	BND	BND	BND
Toscana	0.162	BND	BND	POS
P.A. Trento	0.266	BND	BND	POS
Umbria	0.145	BND	BND	POS
Valle D'Aosta	0.318	POS	POS	POS
Veneto	0.092	BND	BND	BND

Table 4 Correct classification and model prediction for case 3

Region	g	Correct classification	Model prediction (04/09) - Flat case	Model prediction (04/09) - Google case
Abruzzo	0.062	BND	BND	BND
Basilicata	0.058	BND	BND	BND
P.A. Bolzano	0.043	BND	BND	BND
Calabria	0.040	BND	NEG	NEG
Campania	0.080	BND	BND	BND
Emilia-Romagna	0.112	POS	POS	BND
Friuli Venezia Giulia	0.028	NEG	NEG	NEG
Lazio	0.048	BND	BND	BND
Liguria	0.110	POS	POS	POS
Lombardia	0.077	BND	POS	BND
Marche	0.063	BND	POS	POS
Molise	0.077	BND	BND	BND
Piemonte	0.155	POS	POS	POS
Puglia	0.071	BND	BND	BND
Sardegna	0.076	BND	BND	BND
Sicilia	0.044	BND	BND	BND
Toscana	0.061	BND	BND	BND
P.A. Trento	0.091	BND	POS	BND
Umbria	0.002	NEG	NEG	NEG
Valle D'Aosta	0.119	POS	POS	POS
Veneto	0.047	BND	BND	BND

Table 5 Confusion matrix - case 1

			Truth		
			NEG	BND	POS
Pred		NEG	0	0	0
		BND	6	9	3
		POS	0	0	3

Table 6 Confusion matrix - case 2 flat

			Truth		
			NEG	BND	POS
Pred		NEG	1	0	0
		BND	1	15	1
		POS	0	0	3

Table 7 Confusion matrix - case 2 google

			Truth		
			NEG	BND	POS
		NEG	1	0	0
Pred		BND	1	8	0
		POS	0	7	4

Table 8 Confusion matrix - case 3 flat

			Truth		
			NEG	BND	POS
		NEG	2	1	0
Pred		BND	0	11	0
		POS	0	3	4

Table 9 Confusion matrix - case 3 google

			Truth		
			NEG	BND	POS
		NEG	2	1	0
Pred		BND	0	13	1
		POS	0	1	3

Table 10 Accuracy measures - Case 1

	Sens.	Spec.	PPV	NPV	Precision	Recall	F1	BA
Class: NEG	0.00	1.00		0.71		0.00		0.50
Class: BND	1.00	0.25	0.50	1.00	0.50	1.00	0.67	0.62
Class: POS	0.50	1.00	1.00	0.83	1.00	0.50	0.67	0.75

Table 11 Accuracy measures - Case 2 - Flat

	Sens.	Spec.	PPV	NPV	Precision	Recall	F1	BA
Class: NEG	0.50	1.00	1.00	0.95	1.00	0.50	0.67	0.75
Class: BND	1.00	0.67	0.88	1.00	0.88	1.00	0.94	0.83
Class: POS	0.75	1.00	1.00	0.94	1.00	0.75	0.86	0.88

Table 12 Accuracy measures - Case 2 - Google

	Sens.	Spec.	PPV	NPV	Precision	Recall	F1	BA
Class: NEG	0.50	1.00	1.00	0.95	1.00	0.50	0.67	0.75
Class: BND	0.53	0.83	0.89	0.42	0.89	0.53	0.67	0.68
Class: POS	1.00	0.59	0.36	1.00	0.36	1.00	0.53	0.79

decree (from 03/26 to 04/09). We recall that, in this case, $CA(u, t)$ are evaluated taking into account only mobility changes from the Google dataset.

7.6 Accuracy evaluation

To assess the accuracy of the results, we evaluated precision, recall, F1-measure, and balanced accuracy. We have derived a ground truth as the ratio between the increase in positive people and the increase in swabs between the end and the beginning of each time window. Let us define a time window as an interval $TW = [startDate, endDate]$, and let $g(u, TW)$ be defined as:

$$g(u, TW) = \frac{AP(u, endDate) - AP(u, startDate)}{Swabs(u, endDate) - Swabs(u, startDate)}. \quad (6)$$

The values g provide indications on the increase of the infected people compared to the number of swabs during a time window. They are, therefore, indicators of the effectiveness of the measures and, for our evaluation, of the correct assignment of the weights in deriving the thresholds, α and β, for processing POS, BND and NEG areas.

We consider u correctly classified in the POS subset if its g value exceeds the mean of the g values of the Italian regions increased by 50% (i.e., $mean(g) + 0.5 * mean(g)$), correctly classified in the negative zone if its g value is lower than the mean of the g values of the Italian regions reduced by 50% (i.e., $mean(g) - 0.5 * mean(g)$), otherwise u is correctly classified in the BND area.

7.6.1 Case 1: First Time Window - Decree "Red Areas"

The results of Case 1 are reported in Table 2, where it is possible to understand that according to values of g, rounded to the third decimal, the classes correctly classified as POS are those that exceed 0.202. Moreover, the classes correctly classified with NEG are those whose values of g are lesser than 0.068. Lastly, the other objects are correctly classified as BND.

7.6.2 Case 2: Second Time Window - Decree "I stay at home"

In Table 3, reporting the results of Case 2, it is shown that, according to the values of g, the classes correctly classified as POS are those that exceed 0.268. Moreover, the classes correctly classified with NEG are those whose values of g are lesser than 0.090. Lastly, the other objects are correctly classified as BND.

7.6.3 Case 3: Third Time Window - Decree "Close Italy"

From Table 4, that reports the results of Case 3, it is clear that, according to the values of g, the classes correctly classified as POS are those that exceed 0.105. Moreover, the classes correctly classified with NEG are those whose values of g are lesser than 0.035. Lastly, the other objects are correctly classified as BND.

7.6.4 Accuracy measures

We have evaluated accuracy measures using the Caret package of R[11]. Tables 5, 6, 7, 8 and 9 report the confusion matrices for the cases analysed. Table 5 is the same for the Flat and Google cases.

The following Tables 10, 11, 12, 13 and 14 report the values *sensitivity=recall*, *specificity*, *Positive Predicted Value (PPV) = precision*, *Negative Predicted Value (NPV)*, *F1 measure* and *Balanced Accuracy (BA)* evaluated on the basis of the confusion matrices. As before, Table 10 is the same for the Flat and Google cases.

The overall accuracy is 0.5714 for Case 1 (Flat and Google), 0.9048 for Case 2 - Flat, 0.619 for Case 2 - Google, 0.8095 for Case 3 - flat and 0.8571 for Case 3 Google).

7.6.5 Discussion

We discuss and compare results in the following.

– **Case 1 - Decree "Red Areas"**
  As we can observe from Table 10, our method fails to correctly classify regions with low critical levels (i.e.,

---

[11] https://cran.r-project.org/web/packages/caret/index.html

Table 13 Accuracy measures - Case 3 - Flat

	Sens.	Spec.	PPV	NPV	Precision	Recall	F1	BA
Class: NEG	1.00	0.95	0.67	1.00	0.67	1.00	0.80	0.97
Class: BND	0.73	1.00	1.00	0.60	1.00	0.73	0.85	0.87
Class: POS	1.00	0.82	0.57	1.00	0.57	1.00	0.73	0.91

NEG) for the case 1. In the other cases, the model offers better results according to the ground truth defined.

- **Case 2 - Decree "I stay at home"**

  The overall performance ($F1-measure$ and BA) in this case are better in both the situations when $CA(u, t)$ are flatly estimated or evaluated with Google data. In this last situation, however, the model performs worse mainly as regards the POS case. It can be seen from Table 12 that the precision value is very low. This is due to the high number of false positives predicted by the model in this case. The $CA(u, t)$ values derived from Google data, in this case, tend to reduce too much the weights $w(u, t)$ for some regions and thus to reduce the thresholds values, α and β, with the consequence that some true BND classes have been classified as POS. These classes increment the number of false negative for the BND classification and this is reflected also in the low value of the NPV of Table 12.

- **Case 3 - Decree "Close Italy"**

  The overall performance improves also in this case. The accuracy measures for the Flat and Google situations are quite similar.

As overall comparison, we observe that the adoption of flat values to estimate the effect of $CA(u, t)$ provides better performance for the Case 2 of "I stay at home" decree. We observe, furthermore, that the estimation of $CA(u, t)$ with Google data gives the better results the more restrictive are the measures applied in the three decrees under analysis. In fact, the Tables 10, 12 and 14 show a continuous increase in the BA values as we move form "Red Areas" to "Close Italy" decrees.

8 Conclusions and future works

The results of the experimentation are quite good in terms of accuracy. Obviously, the presented method should be considered the starting point of several further works aiming at improving the model and other aspects of the method.

A more in-depth study on how to derive $CA(u, t)$ and, thus, the weights for the determination of α and β is underway. We are investigating two directions: the adoption of more complex operators, such as Ordered Weighted Averaging (OWA) operators [24], to aggregate Google data and uses fuzzy cognitive maps [25] and granular functional networks [26] to consider the effects of retention measures in greater detail.

Furthermore, also other approaches to define the thresholds α and β are under investigation. We are evaluating the approach adopted in [27] that takes into account prior and a posterior probability of an event in the calculation of the thresholds.

Although the paper has been focused on the analysis of the diffusion of COVID-19 in Italy, the method is also applicable in other contexts, such as Critical Infrastructures [28], in which the objects of a universe can be considered nodes of an interconnected system having different importance in establishing the overall criticality of the system. To this purpose, an additional research line is devoted to the integration of the methods in a more general framework to support Situation Awareness, such as our previous work on Granular Situation Awareness [29–31], and ontologies and Situation Awareness [32].

Table 14 Accuracy measures - Case 3 - Google

	Sens.	Spec.	PPV	NPV	Precision	Recall	F1	BA
Class: NEG	1.00	0.95	0.67	1.00	0.67	1.00	0.80	0.97
Class: BND	0.87	0.83	0.93	0.71	0.93	0.87	0.90	0.85
Class: POS	0.75	0.94	0.75	0.94	0.75	0.75	0.75	0.85

References

1. Yao Y (2012) An outline of a theory of three-way decisions. In: International Conference on Rough Sets and Current Trends in Computing. Springer, pp 1–17

2. Yao Y (2016) Three-way decisions and cognitive computing. Cogn Comput 8(4):543–554

3. Yao Y (2011) The superiority of three-way decisions in probabilistic rough set models. Inf Sci 181(6):1080–1096

4. Yue XD, Chen YF, Miao DQ, Fujita H (2020) Fuzzy neighborhood covering for three-way classification. Inf Sci 507:795–808

5. Yang D, Deng T, Fujita H (2020) Partial-overall dominance three-way decision models in interval-valued decision systems. Int J Approx Reason 126:308–325

6. Yao Y (2018) Three-way decision and granular computing. Int J Approx Reason 103:107–123

7. Peters JF, Ramanna S (2016) Proximal three-way decisions: theory and applications in social networks. Knowl-Based Syst 91:4–15

8. Chen J, Li Y, Zhao S, Wang X, Zhang Y (2020) Three-way decisions community detection model based on weighted graph representation. In: International Joint Conference on Rough Sets. Springer, pp 153–165

9. Chen J, Fang L, Zhang Y, Zhao S, Zhang L, Qian F, Liu F, Wang X, Hu X (2017) Three-way dicision community detection algorithm based on local group information. In: International Joint Conference on Rough Sets. Springer, pp 171–182

10. Liu Y, Pan L, Jia X, Wang C, Xie J (2013) Three-way decision based overlapping community detection. In: International Conference on Rough Sets and Knowledge Technology. Springer, pp 279–290

11. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. Lancet 395(10223):497–506

12. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X et al (2020) Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. The Lancet

13. Wang L, Li J, Guo S, Xie N, Yao L, Cao Y, Day SW, Howard SC, Graff JC, Gu T et al (2020) Real-time estimation and prediction of mortality caused by covid-19 with patient information based algorithm. Sci Total Environ:138394

14. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M (2020) Application of the arima model on the covid-2019 epidemic dataset. Data Brief:105340

15. Hernandez-Matamoros A, Fujita H, Hayashi T, Perez-Meana H (2020) Forecasting of covid19 per regions using arima models and polynomial functions. Appl Soft Comput 96:106610. https://doi.org/10.1016/j.asoc.2020.106610, http://www.sciencedirect.com/science/article/pii/S1568494620305482

16. Chintalapudi N, Battineni G, Amenta F (2020) Covid-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in italy: A data driven model approach. J Microbiol Immunol Infect

17. Vaishya R, Javaid M, Khan IH, Haleem A (2020) Artificial intelligence (ai) applications for covid-19 pandemic. Diabetes & Metabolic Syndrome: Clinical Research & Reviews

18. Mei X, Lee H-C, Diao K, Huang M, Lin B, Liu C, Xie Z, Ma Y, Robson PM, Chung M et al (2020) Artificial intelligence–enabled rapid diagnosis of patients with covid-19. Nat Med:1–5

19. Adly AS, Adly AS, Adly MS (2020) Approaches based on artificial intelligence and the internet of intelligent things to prevent the spread of covid-19: Scoping review. J Med Internet Res 22(8):e19104

20. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A (2020) Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures. Proc Natl Acad Sci 117(19):10484–10491

21. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M (2020) Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. Nat Med:1–6

22. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, Edmunds WJ (2020) Quantifying the impact of physical distance measures on the transmission of covid-19 in the uk. BMC Med 18:1–10

23. Katz L (1953) A new status index derived from sociometric analysis. Psychometrika 18(1):39–43

24. Yager RR (1988) On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Trans Syst Man Cybern 18(1):183–190

25. D'Aniello G, Gaeta A, Gaeta M, Loia V, Reformat MZ (2016) Collective awareness in smart city with fuzzy cognitive maps and fuzzy sets. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, pp 1554–1561

26. Loia V, Parente D, Pedrycz W, Tomasiello S (2018) A granular functional network with delay: some dynamical properties and application to the sign prediction in social networks. Neurocomputing 321:61–71

27. Fujita H, Gaeta A, Loia V, Orciuoli F (2020) Hypotheses analysis and assessment in counterterrorism activities: A method based on owa and fuzzy probabilistic rough sets. IEEE Trans Fuzzy Syst 28(5):831–845

28. Fujita H, Gaeta A, Loia V, Orciuoli F (2018) Resilience analysis of critical infrastructures: a cognitive approach based on granular computing. IEEE Trans Cybern 49(5):1835–1848

29. Loia V, D'Aniello G, Gaeta A, Orciuoli F (2016) Enforcing situation awareness with granular computing: a systematic overview and new perspectives. Granular Comput 1(2):127–143

30. D'Aniello G, Gaeta A, Loia V, Orciuoli F (2017) A granular computing framework for approximate reasoning in situation awareness. Granular Comput 2(3):141–158

31. Fujita H, Gaeta A, Loia V, Orciuoli F (2019) Improving awareness in early stages of security analysis: A zone partition method based on grc. Appl Intell 49(3):1063–1077

32. D'Aniello G, Gaeta A, Loia V, Orciuoli F (2016) Integrating gso and saw ontologies to enable situation awareness in green fleet management. In: 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA). IEEE, pp 138–144
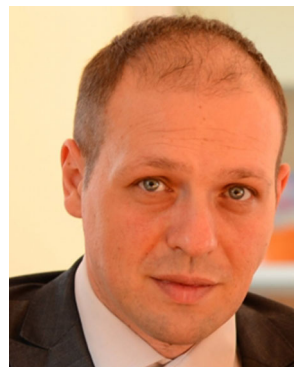
Angelo Gaeta received the master's degree cum laude in electronic engineering and Ph.D. degree in Management & Information Technology from the University of Salerno (Italy). He is currently research assistant (RTD-A) in computer science at the University of Salerno. His research interests relate to distributed computing, situation awareness, and computational intelligence.

Vincenzo Loia received the master's degree in computer science from the University of Salerno (Italy) in 1985, and the Ph.D. degree in computer science from the University of Paris 6, Paris, France, in 1989. He is Full Professor of computer science with the University of Salerno. He is Editor-in-Chief and Founder of the international journal "Ambient Intelligence and Humanized Computing" (Springer-Verlag), Editor-in-Chief of the international journal "Evolutionary Intelligence" (Springer-Verlag) and Associate Editor of "IEEE Transaction Systems, Man and Cybernetics: Systems, "IEEE Transactions on Fuzzy Systems and IEEE Transactions on Cognitive and Developmental Systems. He is author of more than 300 scientific publications indexed by SCOPUS.

Francesco Orciuoli received the master's degree cum laude in computer science from the University of Salerno (Italy) in 1999. He is Associate Professor of computer science with the University of Salerno. He is currently focusing his research activities on semantic technologies and computational intelligence to support human cognitive processes (e.g., decision-making, learning, reasoning). He is author of more than 120 scientific publications indexed by SCOPUS.