Network Modeling Analysis in Health Informatics and Bioinformatics (2021) 10:46
https://doi.org/10.1007/s13721-021-00323-5

ORIGINAL ARTICLE

COVID-19 Community Temporal Visualizer: a new methodology for the network-based analysis and visualization of COVID-19 data

Marianna Milano[1,2] · **Chiara Zucco**[1,2] · **Mario Cannataro**[1,2]

Received: 3 March 2021 / Revised: 21 May 2021 / Accepted: 14 June 2021 / Published online: 2 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

Understanding the evolution of the spread of the COVID-19 pandemic requires the analysis of several data at the spatial and temporal levels. Here, we present a new network-based methodology to analyze COVID-19 data measures containing spatial and temporal features and its application on a real dataset. The goal of the methodology is to analyze sets of homogeneous datasets (i.e. COVID-19 data taken in different periods and in several regions) using a statistical test to find similar/dissimilar datasets, mapping such similarity information on a graph and then using a community detection algorithm to visualize and analyze the spatio-temporal evolution of data. We evaluated diverse Italian COVID-19 data made publicly available by the Italian Protezione Civile Department at https://github.com/pcm-dpc/COVID-19/. Furthermore, we considered the climate data related to two periods and we integrated them with COVID-19 data measures to detect new communities related to climate changes. In conclusion, the application of the proposed methodology provides a network-based representation of the COVID-19 measures by highlighting the different behaviour of regions with respect to pandemics data released by Protezione Civile and climate data. The methodology and its implementation as R function are publicly available at https://github.com/mmilano87/analyzeC19D.

Keywords COVID-19 · Network analysis · Community detection

1 Introduction

The coronavirus epidemic started in China, in November 2019 (Wu and McGoogan 2020) and it has quickly spread in hundreds of countries in the world. COVID-19 disease is a viral infection produced by Sars-CoV-2, a new coronavirus (Wu and McGoogan 2020). The high transmissibility, the high level of infectivity (Huang et al. 2020), and the initial absence of a COVID-19 vaccine caused a huge number of deaths. In March 2020, World Health Organization (WHO) declared COVID-19 as a serious epidemic. In Italy, the first transmission of COVID-19 disease was reported on February, 2020, in the Italian northern regions (Lai et al. 2020) and it spreaded rapidly in the rest of the regions. Italy was seriously influenced by the epidemic with 2697,296 infected and 93,045 deaths recorded at the end of February, 2020. At the time to writing the infected are 4172,525 and the deaths are 124,646.

The data about COVID-19 cases include spatial information (e.g.the geographical regions where data are recorded), and temporal information (e.g., the day of measurement). Thus, an effective analysis methodology must include these two features. This requirement is reinforced if COVID-19 evolution would be related to clinical data that have similar spatio-temporal features.

In this work, we present COVID-19 Community Temporal Visualizer (CCTV) a new methodology for the network-based analysis and visualization of COVID-19 data.

CCTV methodology enables to depict COVID-19 data as networks where each node represents an Italian region and each edge connects statistically similar regions. Then, CCTV methodology ensures to extract clusters of regions

✉ Marianna Milano
m.milano@unicz.it

Chiara Zucco
chiara.zucco@unicz.it

Mario Cannataro
cannataro@unicz.it

[1] Department of Medical and Surgical Sciences, University of Catanzaro, Catanzaro 88100, Italy

[2] Data Analytics Research Center, University of Catanzaro, Catanzaro, Catanzaro 88100, Italy

with similar behaviour along time by using network-based community detection algorithms.

In literature, there are different works (Reich et al. 2020; Herrmann and Schwartz 2020; Kuzdeuov et al. 2020; Kumar 2020; Wang et al. 2020b) that applied graph theory for the analysis of COVID-19 pandemic diffusion. However, these works exploited the network-based representation for the application of predictive models or to perform statistical and network analysis on heterogeneous network.

The strength of our methodology consists in proposing a new approach to analyze homogenous data i.e. COVID-19 data relating to different Italian regions and in different time intervals, that may be generalized to all data that vary along time and position. In fact, CCTV recurs to the graph formalism to map the homogeneous data into network; this enables to conduct network-based analysis, i.e. the community detection that otherwise would not be applicable. Also, in this way homogeneous subcategories of data that do not have a label can be easily categorized by placing them in communities, thus highlighting groups of data that behave in a similar way. Finally, CCTV ensures a direct interpretation of the data for the user, avoiding the explainability problem that affects Deep Learning approaches (Samek et al. 2019; Zucco et al. 2018). Thus, our methodology enables to evaluate the impact of clinical evolution of COVID-19 pandemic by integrating several clinical data on geographical and temporal data by evaluating the evolution of communities coherence in relation to different data. In detail, the CCTV methodology comprises four steps: (i) Application of statistical test to identify the regions that present similar/dissimilar behaviour with respect to COVID-19 measures; (ii) Building of similarity matrices; (iii) Mapping of each matrix of similarity into a network where each node is an Italian region and each edge depicts similarity connections; (iv) Identification of communities by applying community detection algorithms. In particular, we analyze the data released every day from the the Italian Protezione Civile Department at https://github.com/pcm-dpc/COVID-19/, by focusing on the data provided on the period February 24-April 26, 2020 (1st wave), on the period September 28-November 29, 2020 (2nd wave) and then, we compared these two periods. Furthermore, since several studies have found a relationship between climate and virus spread, in this study we wanted to use network alignment techniques to integrate climate information on COVID-19 aggregated data. In particular, we have aligned the network built for a type of data (e.g. a measures about COVID-19)with a previously built climate network, in such a way that a final network is obtained that incorporates the climate data. A community detection algorithm was applied to these new networks with the aim of evaluating how communities vary by taking climate data into consideration. The rest of the paper is organized as follows: Sect. 2 discusses the background on community detection on networks and the background on the correlation among climate data and COVID-19 epidemic, Sect. 3 presents the CCTV methodology and the application on Italian COVID-19 data, Sect. 4 presents and discusses the results. Finally, Sect. 5 concludes the paper.

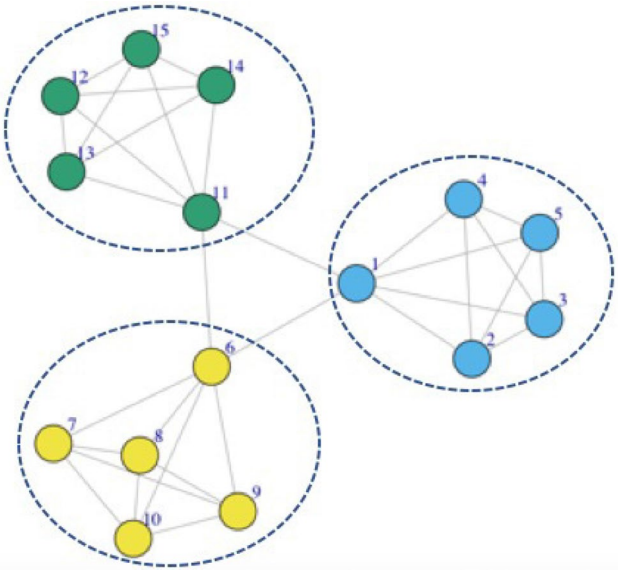


Fig. 1 A community structure example. The figure depicts community 1, community 2 and community 3 within the network

2 Background

In this section we present the background about community detection and the algorithms for community discovery presented in literature. Then, we present the background about the relationship between the spread of COVID-19 and the climate-related measures.

2.1 Algorithms for community discovery

In disparate domains, like diffusion phenomena, social media, and computational biology, data can be modelled as networks. In fact, network formalism is a powerful model able to give the understanding of complex data at different zoom, from the organization of entire networks to individual elements. One network feature is community structure, also called a cluster, that presents a high density of connections within each group of nodes with respect to the entire network (Fortunato and Hric 2016). Figure 1 reports a case of community structure.

Different real networks show natural community structure, i.e. nodes within the same group with common properties or similar functions (Wang et al. 2018). For example, in PPI (Protein-Protein Interaction) networks, communities

are formed by proteins involved in a similar function; or in brain networks, communities can be represented by Regions of Interest (ROIs) that are active during tasks, in social networks, communities can be groups of friends or colleagues. The community structure of networks is a significant characteristic for understanding the organization level of objects within the network. Thus, community detection is an interesting area for the analysis of networks. In the following, different community detection algorithms are summarized.

Fast Greedy (Clauset et al. 2004) and Louvain (Blondel et al. 2008) are two algorithms based on modularity optimization for the identification of communities, that differ in the way they optimise the modularity score, where modularity measures how well a network decomposes into modular communities.

Fast Greedy (Clauset et al. 2004) startsby setting each nodes a single community and it joins pairs of communities by applying a basic greedy approach. At each step, the communities are added according to the increasing of modularity (Yang et al. 2016). Fast Greedy joins pairs of communities until the merging of community pairs does not increase the modularity score.

Louvain (Blondel et al. 2008) includes a community aggregation step to improve communities detection process. The algorithm joins a node with each one of its neighbours communities according to the increasing of modularity, otherwise, the node remains in its initial community. Louvain terminates the procedure when none improvement in modularity is obtained. After that, a new network is built whose nodes represent the detected communities and inter and intra community edges are represented by weighted edges and self-loops respectively.

Multilevel algorithm (Blondel et al. 2008) uses a greedy approach for optimising the modularity. At first, a community is assigned to each node. Then, according to the increasing of modularity, Multilevel inserts a node into a community formed by its neighbours.

WalkTrap (Pons and Latapy 2005) is a hierarchical clustering algorithm that applies a random walk distance measure. Initially, WalkTrap computes the distances between all adjacent nodes in the network. Then, it starts with a node and randomly selects one of its neighbour; it merges them in a community and it updates distances between communities. The assumption is that short random walks remain in an equal community.

InfoMod (Rosvall and Bergstrom 2007) and InfoMap (Rosvall and Bergstrom 2008) represent the community as parts showing regularities in topology. A community is considered better when it has high compactness and low information loss.

The InfoMod and InfoMap algorithms represent the community structure in different ways. InfoMod represents the community structure as community matrices and membership vectors associating each node to a community, whereas, InfoMap depicts the communities by considering 2 levels: the first one to categorize a community within the network and the second one to categorize nodes within the communities.

The edge betweenness algorithm (Girvan and Newman 2002) is based on edges betweenness that is the generalised Freeman's betweenness centrality measure (Freeman 1978). The assumption is that the edges that connect communities show a high value of edge betweenness measure. Thus, by deleting the edges with high edge betweenness, the community topology is detected. Spinglass (Reichardt and Bornholdt 2006) algorithm is based on physical spin glass models (i.e. the model that describes a magnetic state characterized by randomness, besides cooperative behavior in freezing of spins at a temperature). The algorithm aims to discover the ground state of a spin glass model on the basis that the edges should link nodes with an equal spin state, i.e. equal community. Otherwise, the nodes with different spin states should be disconnected.

Label propagation algorithm (Raghavan et al. 2007) detects the communities considering how the information is transmitted in the network. The assumption is that the nodes of the same community are characterized by high efficiency in the exchange of information. The algorithm initialises a diverse label, i.e. community, for each node. Afterwards, it randomly lists the nodes according to a consecutive order. Then, by considering the list, the node is labeled as most of its neighbors. The process ends when all neighbours nodes have the same label. An extended version of Label propagation algorithm is Community Overlap Propagation Algorithm (COPRA) (Gregory 2010) that enables the detection of communities in weighted and bipartite networks, besides unweighted ones.

MarkovCluster algorithm (Van Dongen 2008) works by simulating a stochastic (Markov) flow in a weighted network, where the nodes are data points, while the adjacency matrix stores the edge weights. When the algorithm converges, it produces the new edge weights that define the new connected components of the graph (i.e. the clusters).

The leading eigenvector algorithm (Newman 2006) computes the eigenvectors of the modularity matrix for the optimization of modularity score. The algorithm computes the leading eigenvector and then it splits the network to maximize the modularity based on the leading eigenvector.

Until now, community detection algorithms have not been used to analyze COVID-19 data. In fact, all of these algorithms require that input data are provided as networks, on the other hand, in many complex phenomena the data is not natively available in the network format, so a prerequisite to use such algorithms is a deep preprocessing of data. Our contribution goes in this direction by providing a comprehensive pipeline for the network representation of data, their

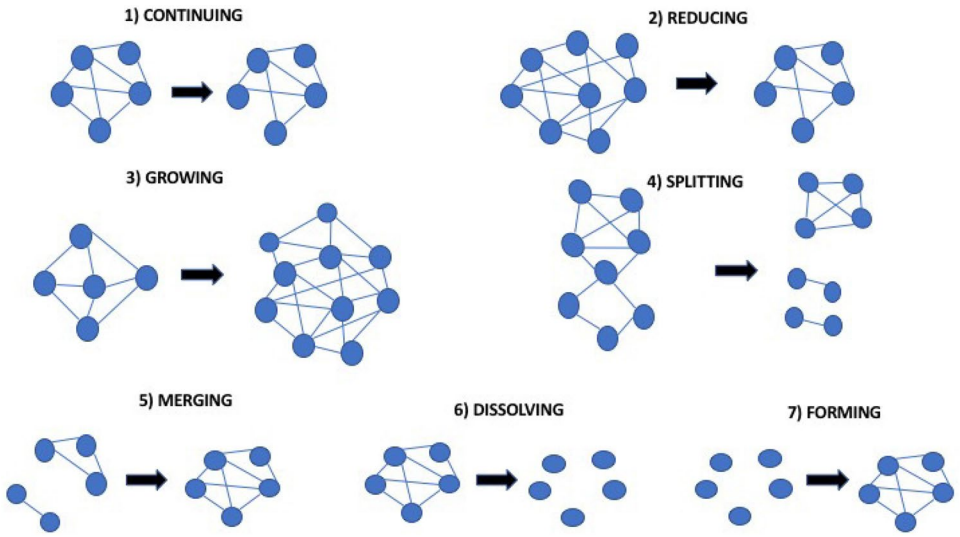


Fig. 2 Common changes in community structure evolution

visualization and their detection of communities through one of these algorithms.

Temporal evolution of communities

The algorithms discussed so far are usually applied to static networks for community detection. Furthermore, in real-world systems, the networks evolve across time by presenting modifications of community structure. In general, the evolution of community structure along the time can be summarized in seven most common changes (see Fig. 2) (Palla et al. 2007, Asur et al. (2009), Bródka et al. (2012)):

1. the community maintains its structure during the temporal evolution;
2. the community reduces when several members leave a group;
3. the community growths when new members join into a group;
4. the community splits into two or more groups;
5. the community is formed from merging of different groups;
6. the community ends its existence and all the nodes are disconnected;
7. the community is formed from groups of nodes initially disconnected.

Figure 2 depicts different scenarios of community structure evolution.

In literature, there exist different methods for community detection with respect to time. These approaches can be categorized into:

- Traditional static community detection (Lancichinetti et al. 2011; Palla et al. 2007);
- Evolutionary clustering (Chakrabarti et al. 2006);
- Incremental clustering (Li et al. 2015).

In the first method, the network evolution is divided into different timeframes, and the extraction of communities is obtained by traditional static community detection methods.

The evolutionary clustering method aims to find the best community topology that depicts the network at a specific time and to evaluate the similarity of a current community with the structure of a previous time by adding a cost related to temporal smoothness.

Finally, the incremental clustering method applies the community topology of the first temporal interval to conform to the community property of incremental nodes for the rest of the temporal intervals.

2.2 Relationship between COVID-19 and climate data

The relationship between COVID-19 spread and the environmental humidity or other climate-related measures has raised growing interest over the past decade (Casanova et al. 2010; Caspi et al. 2020; Gardner et al. 2019; Geller et al. 2012). Concerning SARS-COV-2, recent studies have been proposed with the goal of investigating the effects of different meteorological conditions on the spread of COVID-19. However, the scientific community is still divided between studies that may suggest an existing relationship between air temperature and COVID-19 incidence rates (Zhu and Xie

2020; Ma et al. 2020; Wang et al. 2020a), and studies that did not detect any evidence supporting that minimum and maximum temperature, rainfall, humidity and other climate-related variables are correlated with the COVID-19 spread (Tosepu et al. 2020; Bashir et al. 2020; Cheval et al. 2020).

More in details, attempts to model the spread of COVID-19 in Italy and to correlate it with several variables have been made. For instance, the work of Gatto et al. (2020) tried to estimate the effects of lockdown restrictions and the consequent changes in human mobility with the epidemiological data.

In Zoran et al. (2020), the authors carried out a Pearson correlation analysis to discuss the existence of relations between gaseous air pollutants and climate data with the spread of COVID-19 in Milan. In particular, the paired correlation between the time series related to the daily number of total cases, the daily number of new cases, and the daily count of total deaths with gaseous air pollutants O_3 and NO_2 and several climate factors as air temperature, relative humidity, wind speed intensity, precipitation rate, were investigated. The results showed an inverse correlation of COVID-19 spread indicators with daily average relative air humidity and precipitation rate, a positive correlation with air temperature, a negative correlation of NO_2 with confirmed total COVID-19 infections, daily new positive, and total deaths cases.

Under a Granger-causality perspective, Delnevo et al. (2020) discussed the possible existence of a directed relationship between the series of the measures of the daily values of PM2.5, PM10, and and NO_2 with the series of the new daily COVID-19 cases in February-April 2020, in the Italian region of Emilia-Romagna. The statistical tests were performed in two temporal windows referring to the period before and after the application of lockdown restrictions in Italy, that officially took effect on 8–10th March 2020, depending on the considered province.

Then, guided by the assumption that a predictive relationship exists between the presence in the air of particulate pollutants, such as PM2.5, PM10, and NO_2, and the spread of the virus, the same authors proposed a Machine Learning based model to forecast the spread of the disease in Emilia-Romagna (Mirri et al. 2020).

The work of Fazzini et al. (2020) exploits Mannk-Kendall rank correlation, multi-regression, and stepwise forward analysis to examine the possible correlation between climate variables and the ratio between the first positive swabs and the total number of swabs collected per day, in different areas of the Lombardy region. The results of the Mank-Kendall test did show an acceptable correlation only with the relative humidity, while a multiple linear regression analysis showed that relative humidity, solar radiation, and maximum temperature cover approximately the 60% of the positive swabs explained variability.

However, although the works citate so far, we can conclude that the research in this field is still limited and requires further developments.

3 CCTV methodology

The proposed CCTV (COVID-19 Community Temporal Visualizer) methodology implements the network-based analysis and visualization of generic input datasets organized as collection of homogeneous measures (e.g. COVID-19 data in similar regions).

Our methodology presents a general design, for which it can be applied to analyze different types of data. We implemented our methodology using the R software (Milano 2019).[1] CCTV function requires the upload of the igraph libraries (Csardi et al. 2006).

CCTV methodology comprises four steps:

1. Building of the similarity matrix;
2. Mapping similarity matrices to networks (i.e. Network building;
3. Network analysis over time;
4. Community detection.

Figure 3 shows the pipeline of the CCTV methodology and Algorithm 1 describes those steps.

Algorithm 1 CCTV Methodology Pseudocode

Data: D // Dataset
Data: w // Similarity Measure
Result: C (Network Communities)

$M \leftarrow$ SimilarityMatrix (D,w);
$N \leftarrow$ SimilarityNetwork (M);
$C \leftarrow$ CommunityDetection (N)
Return (C);

[1] The CCTV methodology as R function is available at https://github.com/mmilano87/analyzeC19D.

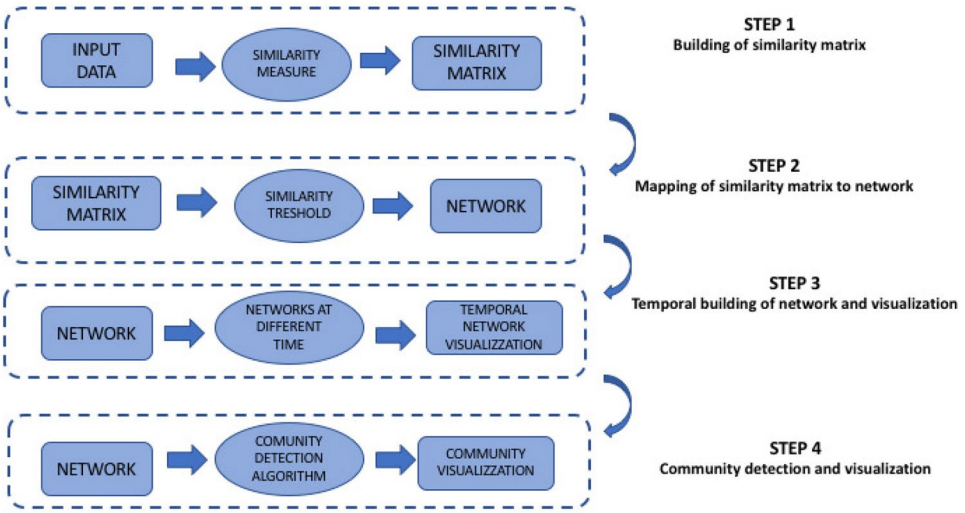


Fig. 3 CCTV Methodology pipeline

In the following, we report the steps of the **CCTV.R** function:

- **Building of the similarity matrix.** At first, the function takes as input the datasets consisting of the collected data for all regions. The user selects the kind of aggregate data on which he/she want to perform the analysis and the Wilcoxon test is applied to compute the pair-wise similarity among the regions. After that, a similarity matrix is built. The matrix contains the p value resulting of the Wilcoxon test for each couple (i, j) of regions;
- **Network building and Visualization** Starting from the similarity matrix, the CCTV function builds a network. At the end of this step, CCTV.R plots the built network.
- **Community detection.** Finally, the community detection algorithm is applied to mine the communities on the network built in the previous step. At the end of this step, CCTV.R plots the detected communities.

In the following paragraphs, we explain in detail each CCTV step by presenting the application of our methodology to analyze the Italian COVID-19 dataset.

Dataset

The COVID-19 datasets that we used of the analysis, are released daily by the Italian Protezione Civile at https://github.com/pcm-dpc/COVID-19. The overall database contains a dataset for each day, starting from February 24, 2020 and for each Italian region. The regions are: Abruzzo, Basilicata, Calabria, Campania, Emilia, Friuli, Lazio, Liguria, Lombardy, Marche, Molise, Piedmont, Puglia, Sardinia, Sicily, Tuscany, Umbria, Valle d'Aosta, Veneto, plus the autonomous provinces of Bolzano and Trento, for a total of 21 regions. The dataset of each region in a time point (day) contains the following 10 data measurements:

- Hospitalised with Symptoms, relating to the count of COVID-19 patients in the hospital;
- Intensive Care, relating to the count of COVID-19 patients in Intensive Care Units;
- Total Hospitalised, relating to the sum of Hospitalised with Symptoms and Intensive Care measured;
- Home Isolation, relating to the count of subjects in quarantine at home;
- Total Currently Positive, relating to the count of COVID-19 positive subjects;
- New Currently Positive, relating to the daily count of COVID-19 positive subjects;
- Discharged/Healed, relating to the count of healed or discharged from hospital subjects;
- Deceased, relating to the count of deaths;
- Total Cases, relating to the count of subjects affected by COVID-19;
- Swabs, relating to the count of test swab carried on COVID-19 positive subjects and on suspected COVID-19 positivity subjects.

For each time point, the dataset D of a region is a vector containing 10 integer values, e.g. at time $t = 1$, $D^1_1 = D^1_{\text{abruzzo}} = [x, y, ..]$. The data occupies 1 Gbytes of memory.

For all analysis we focus on two period: February 24-April 26, 2020 (1st wave), and September 28-November 29, 2020 (2nd wave) and then we compared the two periods.

At first, we focus on COVID-19 data related to the February 24-April 26, 2020 period.

Then, we considered the data evolution by analyzing five time intervals:

- the first week that starts on February 24 and ends on March 1;
- three weeks that comprise the period February 24–March 15;

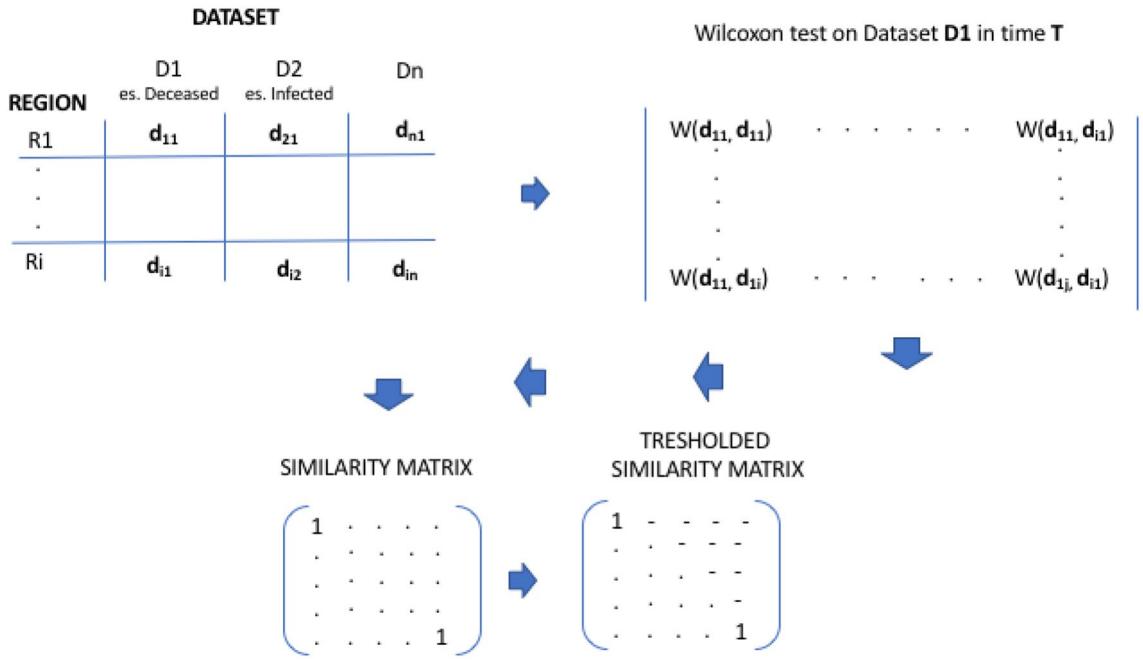


Fig. 4 Definition of the similarity matrix

– five weeks that comprise the period February 24–March 29;
– seven weeks that comprise the period February 24–April 12;
– nine weeks that comprise the period February 24–April 26;

Then, we analyze the COVID-19 data related to the September 29-November 24, 2020 period (that, for convenience, we called second wave or observation period).

Then, we considered the data evolution by analyzing five-time intervals:

– first week that starts on September 29 and ends on October 4;
– three weeks that comprise the period September 29–October 18;
– five weeks that comprise the period September 29–November 1;
– seven weeks that comprise the period September 29–November 15;
– nine weeks that comprise the period September 29–November 29.

3.1 Building of similarity matrices

The first step of CCTV methodology consists of the building of similarity matrix. Before this, the needs to identify similarity measures arises. So, with the goal to choose which similarity measure we can use to build a similarity matrix, we wanted to evaluate the data distribution. So, we used Pearson's chi-square test to evaluate the distribution of each type of data.

Since p value resulted less than 0.05, we chose to apply non-parametric test for subsequent similarity matrix building.

So, we recurred to the use of the Wilcoxon Sum Rank test (Gehan 1965) to analyze the same type of data on the observation period, consisting of 9 weeks and then, for single time intervals. The Wilcoxon test enables to assess the difference among conditions when the samples are correlated. Thus, we used the Wilcoxon test to compare the Italian regions with the aim to evidence statistically similar distributions among them.

Once a similarity measure (Wilcoxon test) is applied to dataset D, a similarity matrix M is built. Let's matrix M for dataset D over a time period T, each (i, j) element represents a value obtained by performing a similarity measure. A typical input dataset $D_1 ... D_n$ is a collection of measured data varying long time. Usually, a generic D_i refers to data collected in location i (i.e. region or geographical position).

So, the dataset i means the dataset collected at Region i.

In this work, we considered the p value as a measure of similarity. So, we applied a statistical test, the Wilcoxon test, to input datasets $D_i ... D_i$ to build a similarity matrix between each $D_i, D_i, i \neq j$. In detail, the (i,j) value of the similarity matrix, related to data D_i (for example deceased data), describes the p value of the Wilcoxon test achieved

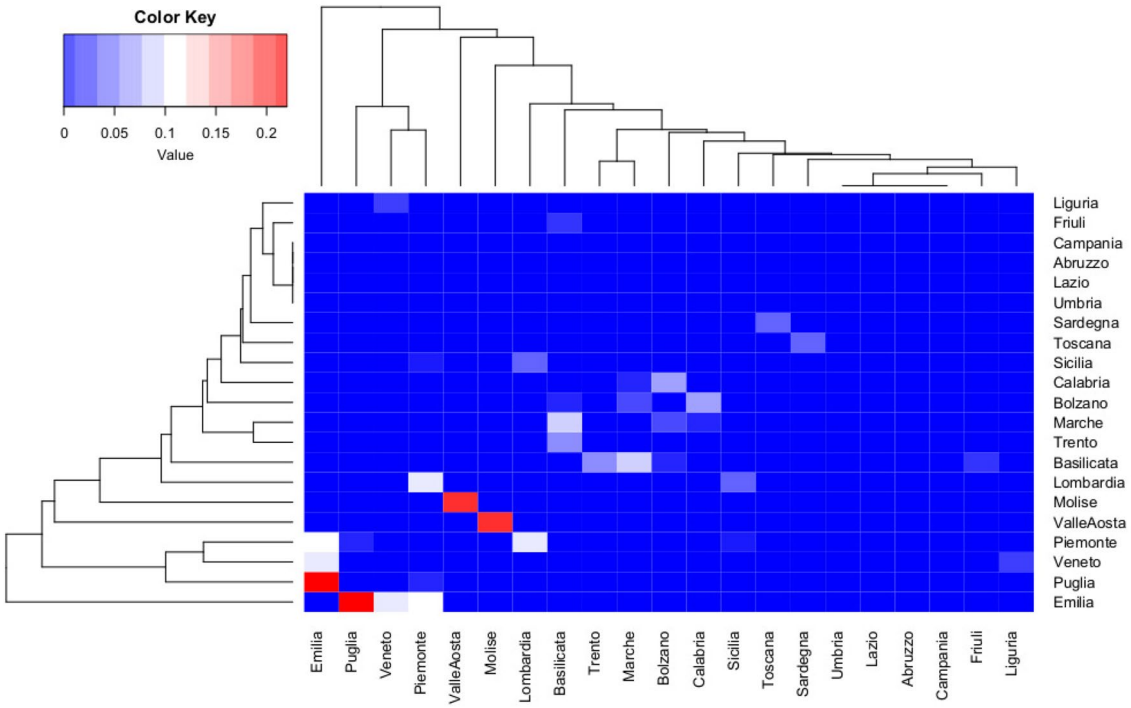


Fig. 5 An example of similarity matrix of Intensive Care network after the first week

by applying the test on a given measure (e.g. the number of deceased) of the region i versus the region j in a given time interval. A lower p value implies that two regions are different according to that measure. Otherwise, a higher *p-value* implies that regions show a similarity according to that measure. We considered the conventional significance threshold of 0.05, and for this reason, we built similarity matrices that contain only $p$values ≥ 0.05. We mapped the p values < 0.05 equal to zero. An example of the similarity matrix definition is reported in Fig. 4.

Thus, we constructed ten similarity matrices for all COVID-19 measures and for each time interval, that report the statistical comparison among each couple of regions.

Figure 5 reports the heatmap describing the similarity values among regions related to Intensive Care data after the first week (February 24–March 29). All similarity tables related to Italian COVID-19 data are computed for the first and second observation period and for a single week, and they are reported as Supplementary material, for the lack of space.

3.2 Converting similarity matrix to network

The second step consists of the converts similarity matrix into networks, i.e the building of similarity networks. According to the similarity matrix, a similarity threshold is fixed. Then, each similarity matrix $M(i, j)$ is mapped to a network N, whose nodes are the Italian regions and the edges connect them only when the similarity value among two regions (i,j), resulting in the matrix, exceeds the similarity threshold. Edges are weighted according to the similarity values. The weights of the edges result inversely proportional to similarity values. Thus, we mapped each similarity matrix $M(i, j)$ into a network N (Agapito et al. 2017). The nodes are the Italian regions, and the edges link two regions (i,j) when the p value is greater than the threshold, otherwise (p value < 0.05) no edge is added. Each edge is weighted according to p value resulting from the Wilcoxon test. In this way, the edge length in the network corresponds to weight and it results inversely proportional to similarity.

3.3 Network analysis over time

The third step consists of the building of the network at different time intervals. Starting from the consideration that the COVID-19 data present a temporal evolution, for each one, the corresponding networks at diverse time intervals and for an observation period are built.

We performed a temporal analysis by building ten networks related to the COVID-19 measures (Hospitalised with Symptoms, Intensive Care data, Total Hospitalised, Home Isolation, Total Currently Positive, New Currently

Positive, Discharged/ Healed, Deceased, Total Cases, Swab) by considering the period February 24–April 26 (the observation period) and then, by building the networks of the same data at different time intervals. After that, we built the networks related to the second observation period and single weeks. Furthermore, this step enables the visualization of the networks changes at different times.

The networks built for the first observation period and for the different time intervals and the networks built for the second observation period and for the different time intervals are reported in Supplementary for lack of space.

3.4 Community detection and temporal evolution

In the last step the extraction of the communities from the built networks is performed. The choice of an appropriate community detection algorithm is important to achieve the best results. In this step, an algorithm for community detection is applied to identify communities, i.e. groups of regions sharing similarity, on the networks built by different time intervals and for an observation period. For this aim, we used the WalkTrap community finding algorithm (Pons and Latapy 2005). WalkTrap is able to identify subgraphs with high density, i.e. communities, in a network through random walks. We selected the WalkTrap community detection algorithm because it outperforms other methods as discussed in (de Sousa and Zhao 2014). Thus, for each network related to different time intervals, we identify regions that form a community according to their similarity. Finally, this step enables the visualization of the community at different time intervals.

Figures 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15 show the identified communities related to Italian COVID-19 data in the first observation period. Figures 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25 show the identified communities related to Italian COVID-19 data in the second observation period and in a single time interval. Then, for each community of different data, we computed the related centroid, i.e. the mean value, for the first and second observation period. We reported the table related to the centroids of the extracted communities for the first and second observation period in the Supplementary material, for the lack of space.

3.5 Integrating climate data by using network alignment method

Different studies (Briz-Redón and Serrano-Aroca 2020; Bashir et al. 2020; Malki et al. 2020; Xie and Zhu 2020; De Natale et al. 2020) evidenced a correlation between climatic factors and the spread of the virus. In fact, the laboratory studies highlighted that the winter climate that characterized the city of Wuhan in 2019 was very similar to the climate observed between February and March 2020 in the Italian provinces that hit hardest by COVID-19. Furthermore, it was observed that, between January and March 2020, there was a higher growth of cases in the colder countries and a more contained spread in the warmer ones. Starting from these considerations, we wanted to relate the Italian COVID-19 measures with climate data. From a general point of view, we need to extend the pre-processing step of our methodology to incorporate other data, in particular climate data. For this aim, we recurred to network alignment technique. Network alignment consists of finding a mapping between nodes of two or more input graphs guided by a scoring function that represents the quality of the alignment (Milano et al. 2020). In literature, different network alignment methods have been proposed, that can be distinguished as local or global network alignment. In this work we selected MAGNA++ (Vijayan et al. 2015), a global network aligner, to combine the climate data with a specific kind of Italian COVID-19 data: Intensive Care, Total Hospitalised and Total Currently Positive data. For analysis, we referred to automatically collected weather-related data for every Italian province through standard web scraping techniques, focusing on the period February 24-April 26 and September 28–November 29.

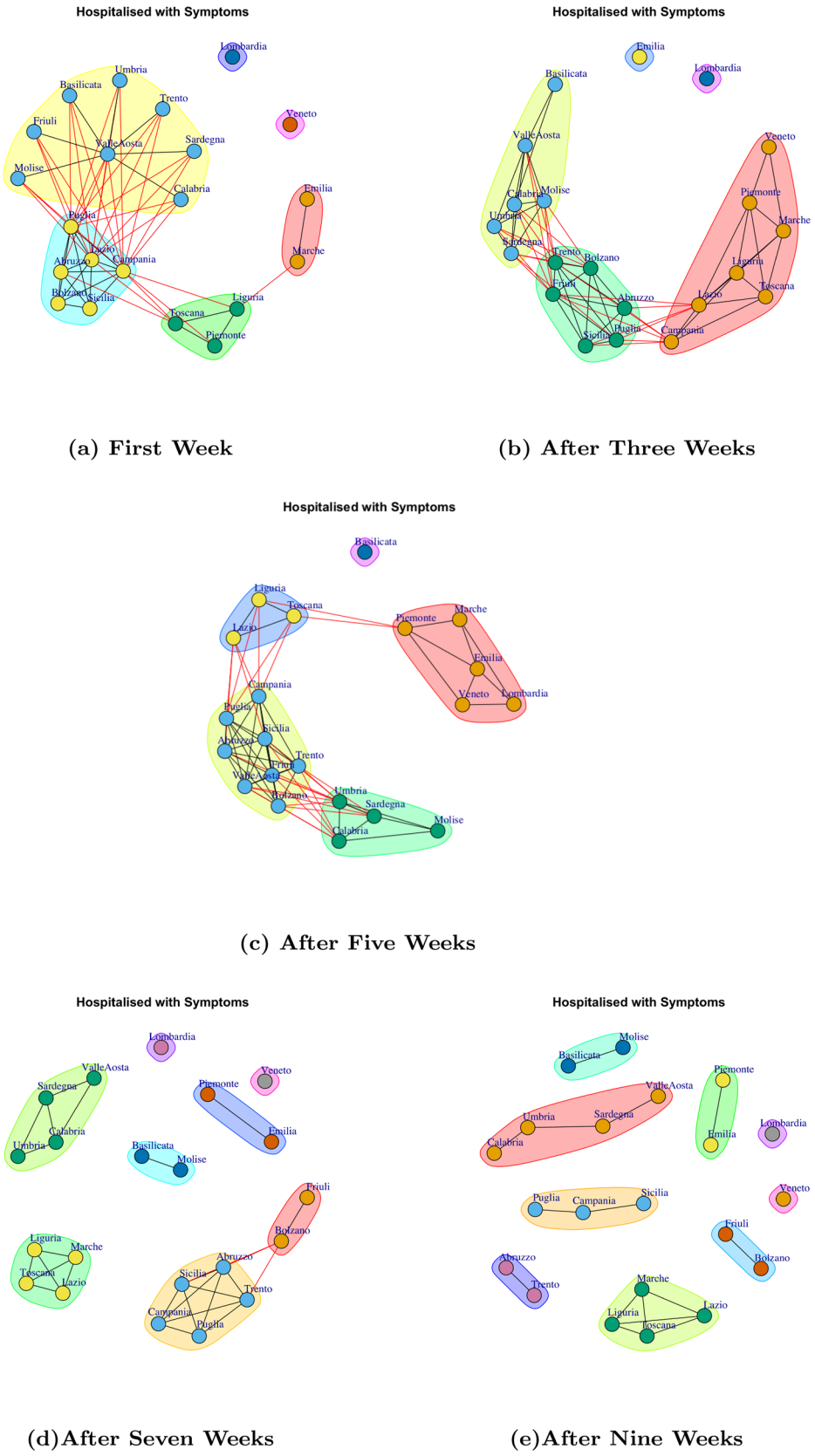
In particular weather data is collected from the historical weather database provided by the Italian website "IlMeteo.it"[2] by following the methodology presented in Agapito et al. (2020).

More in details, starting from the COVID-19 dataset, also available at a city level, for each city, region, and date for which the epidemiological COVID-19 data were available, we reconstructed the URL of each page that contains the relative information in the meteorological website, that was in the format "https://www.ilmeteo.it/portale/archivio-meteo/City/Year/Month/Day."

We used the Python Beautiful Soap library to connect to the pages automatically and, by parsing the HTML document, we collected minimum, maximum, and average daily temperature, humidity, wind speed, average pressure, precipitation, and visibility. Raw data has been normalized (e.g., by fixing inconsistency, providing uniformity of unit measurement for each date and province) and cleaned (e.g., duplicate instances and redundant information removal). Finally, the weather data were stored in a CSV file format. The pipeline for data collection is reported in Fig. 26.

At first, we built a climate network for the first observation period and a climate network for the second observation period, where the nodes represent the Italian regions, and the edges connect statistically similar regions. Then, we aligned the climate network for the first observation

[2] https://www.ilmeteo.it/portale/archivio-meteo/.



(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

Fig. 6 Evolution of Hospitalised with Symptoms Network Communities in the first observation period

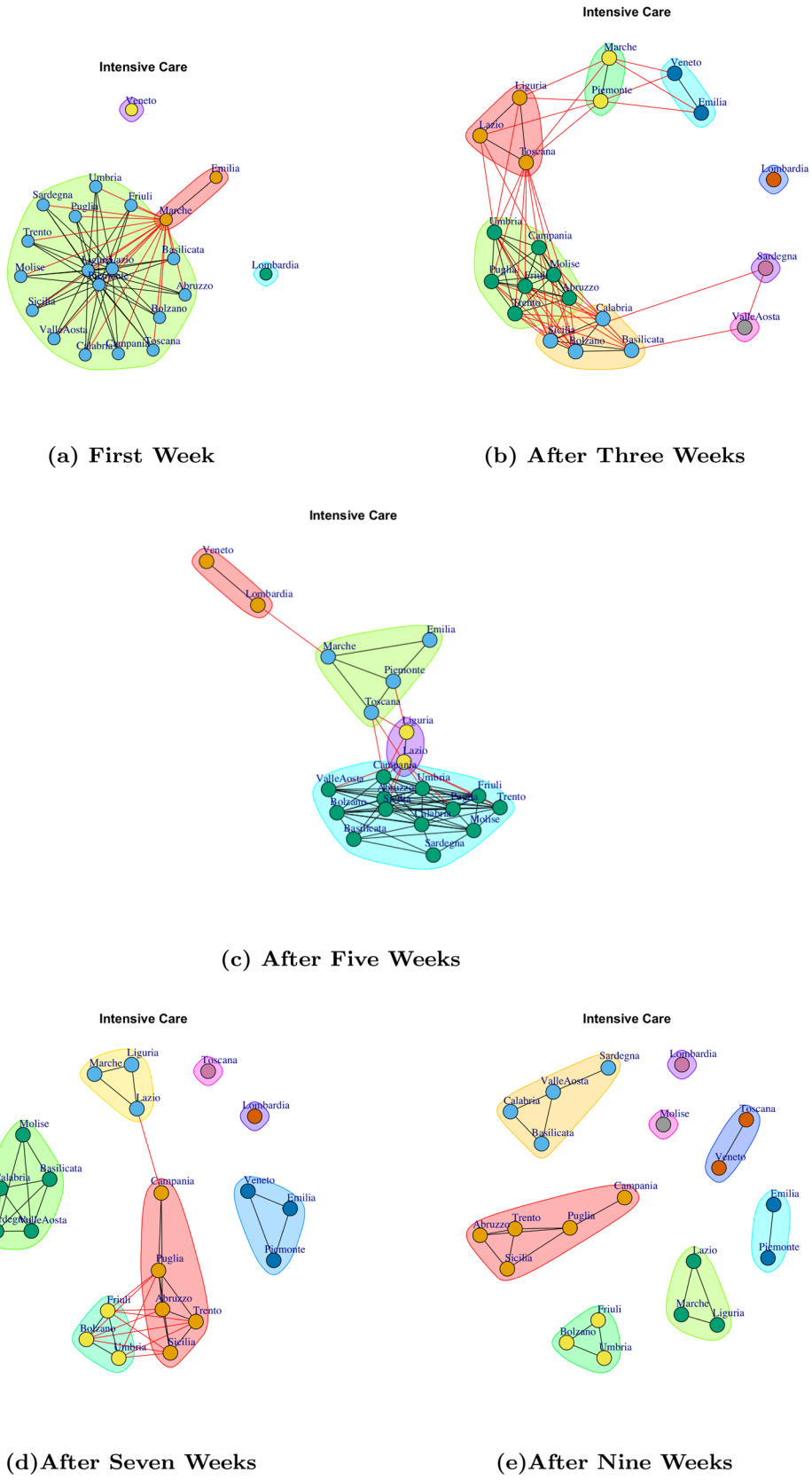


Fig. 7 Evolution of Intensive Care Network Communities in the first observation period

(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

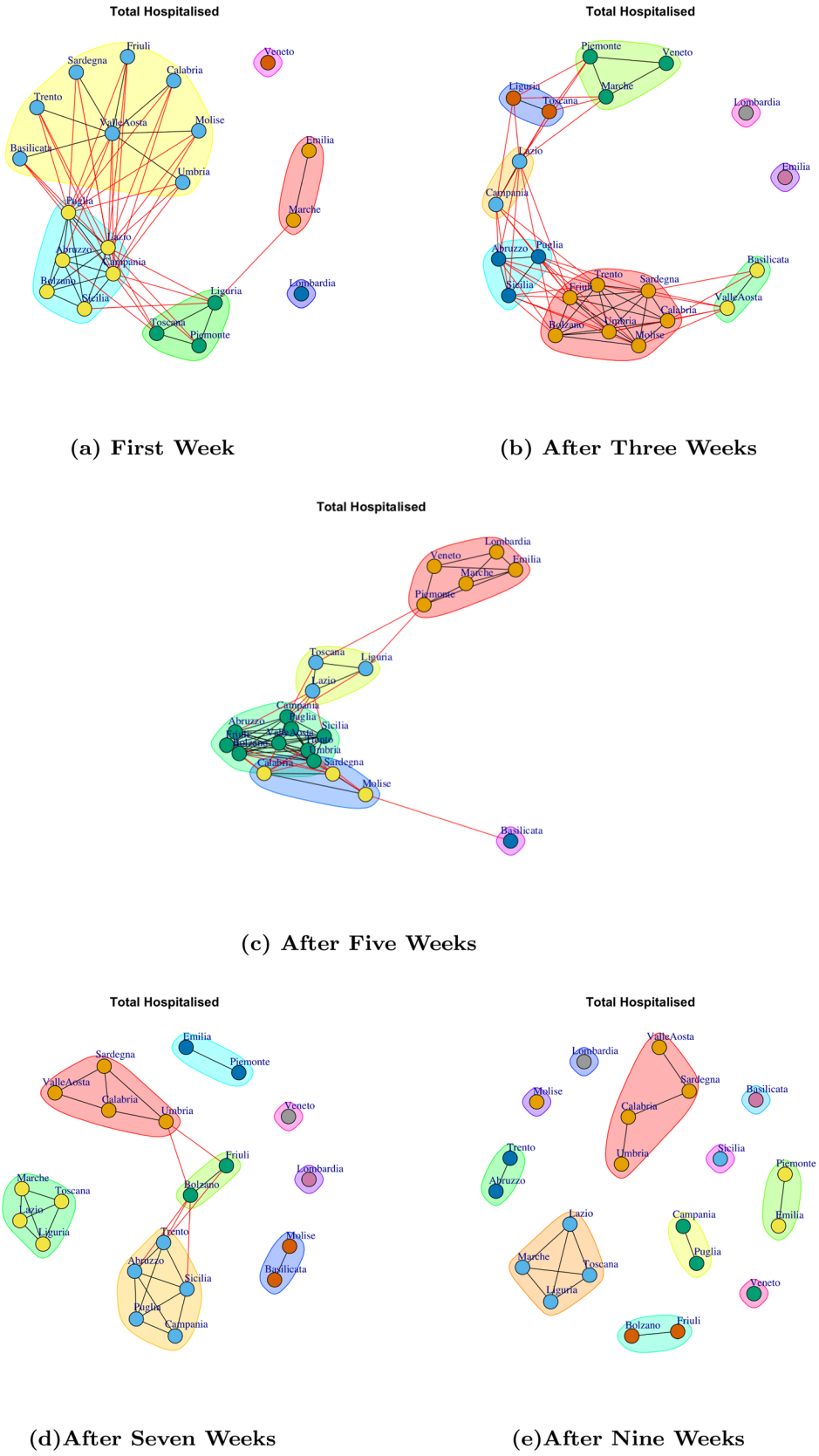


Fig. 8 Evolution of Total Hospitalised Network Communities in the first observation period

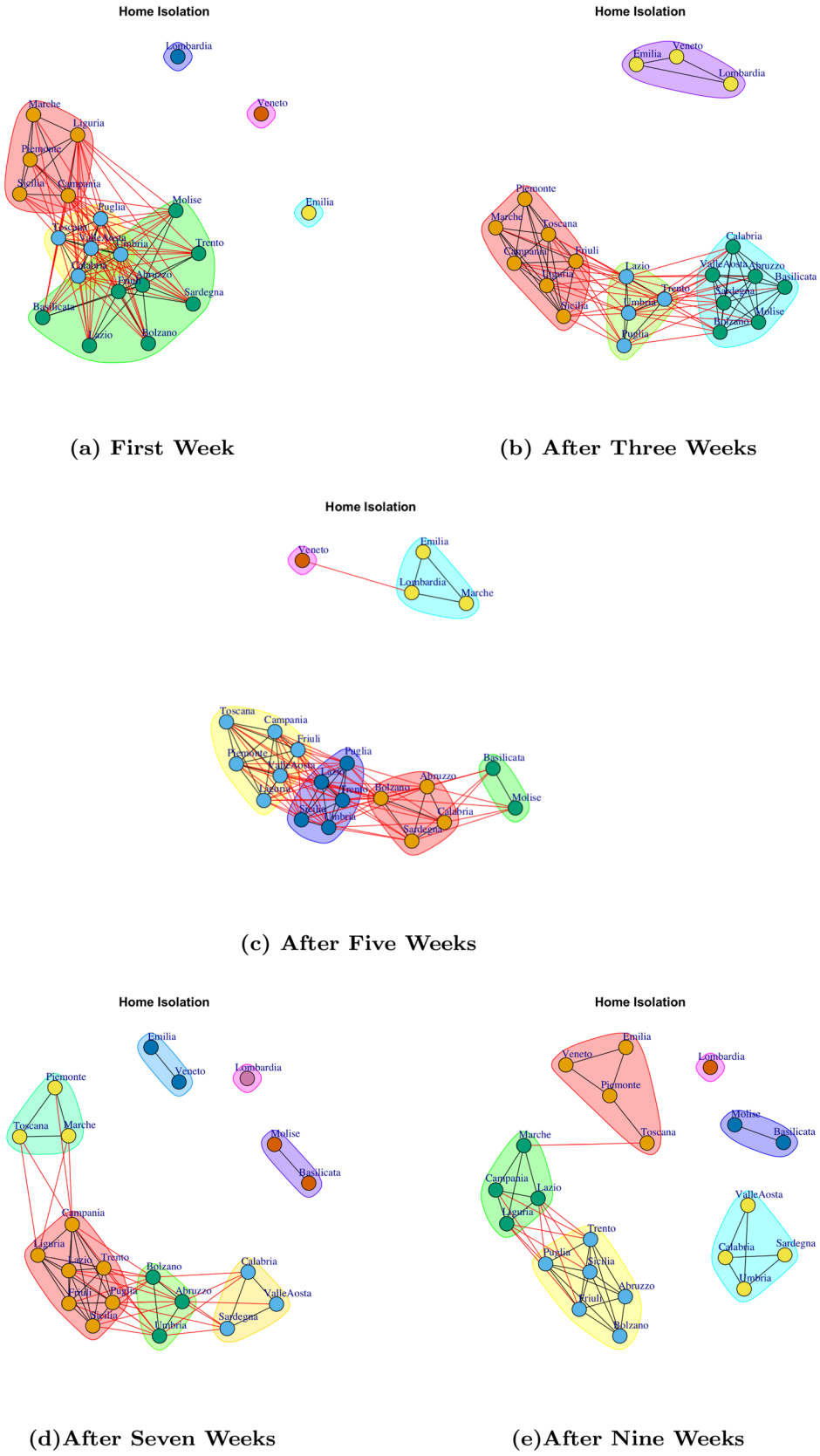


Fig. 9 Evolution of Home Isolation Network Communities in the first observation period

(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

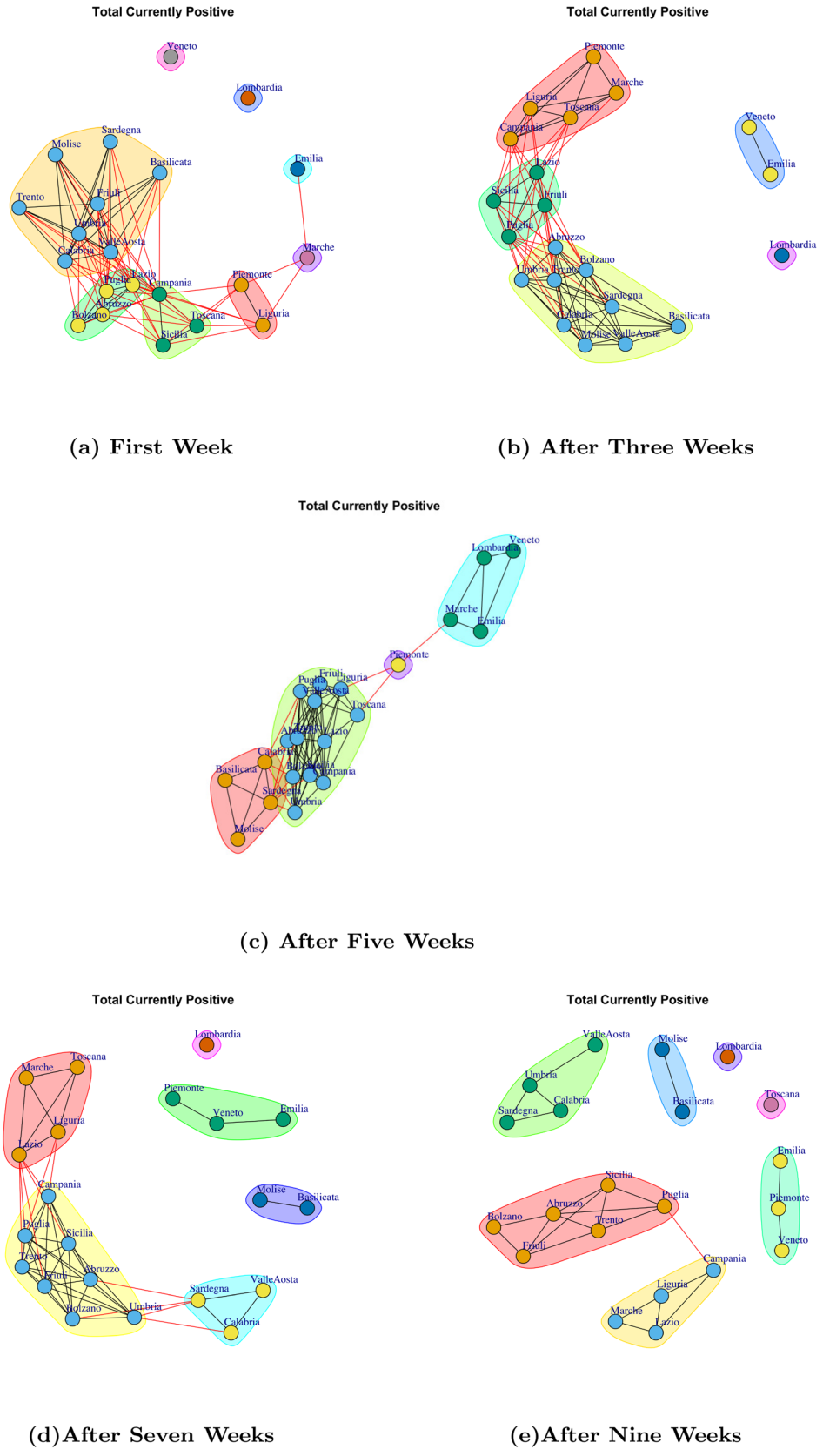


Fig. 10 Evolution of Total Currently Positive Network Communities in the first observation period

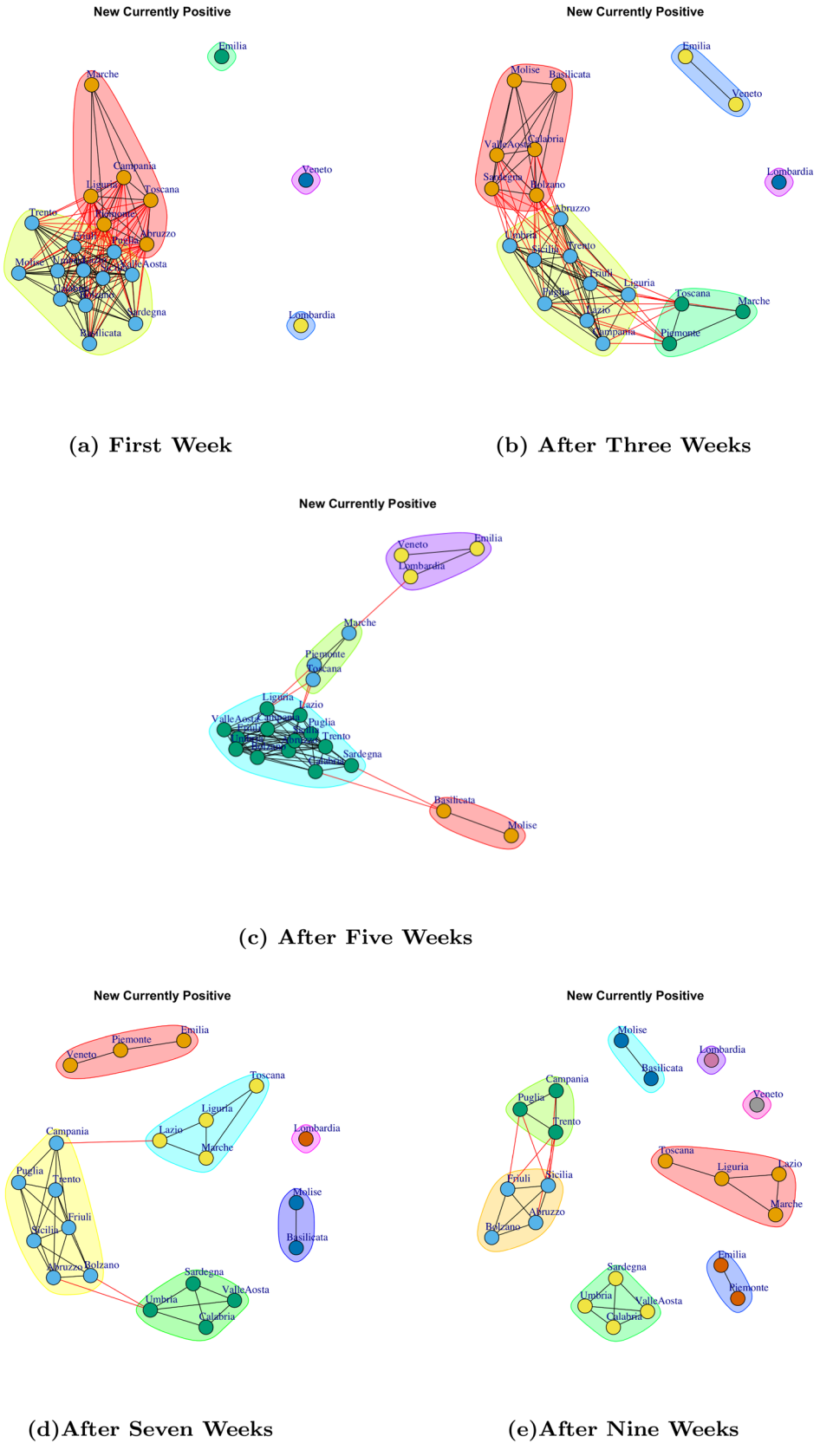
(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks



(a) First Week

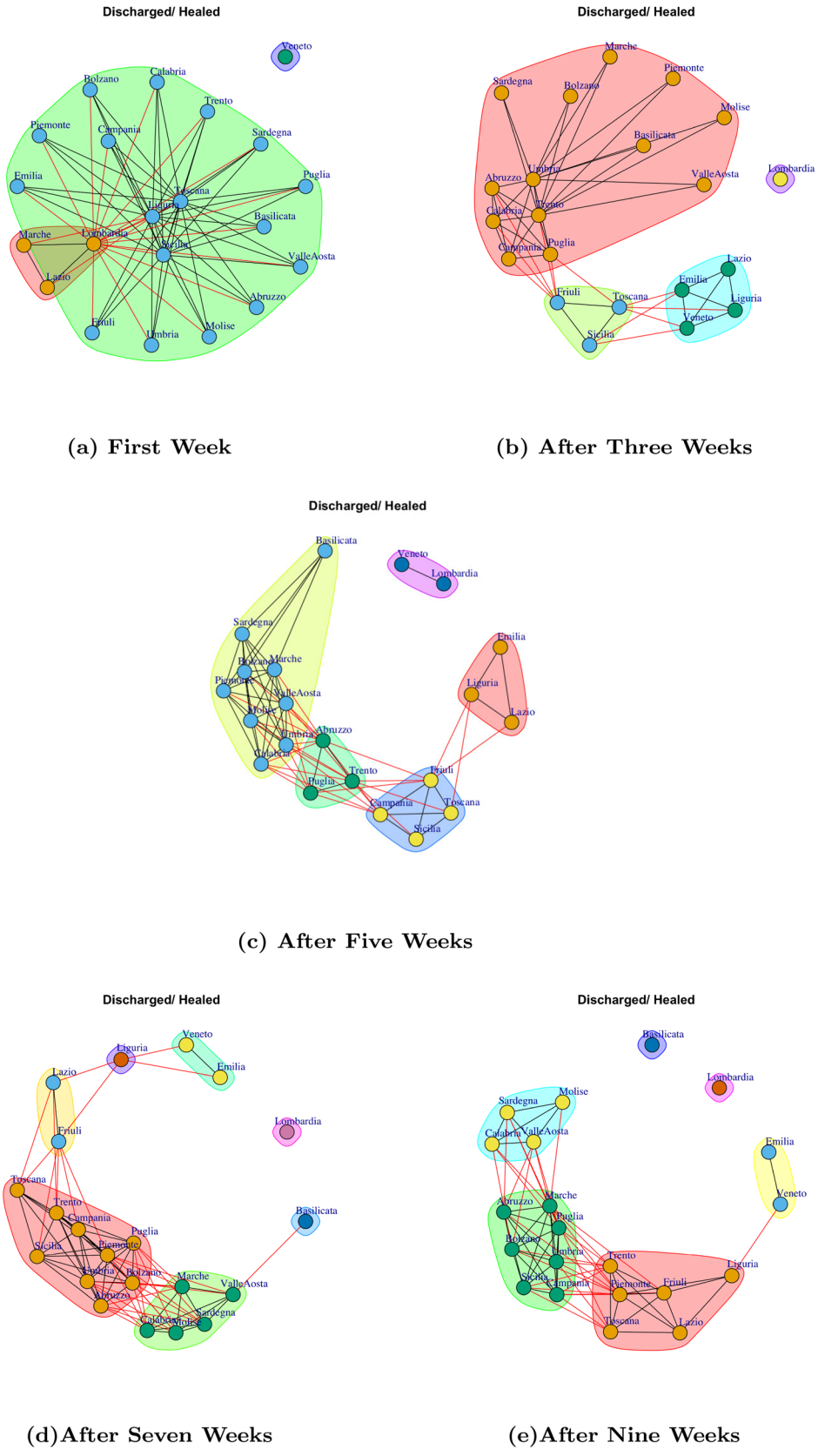
(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

Fig. 11 Evolution of New Currently Positive Network Communities in the first observation period



(a) First Week

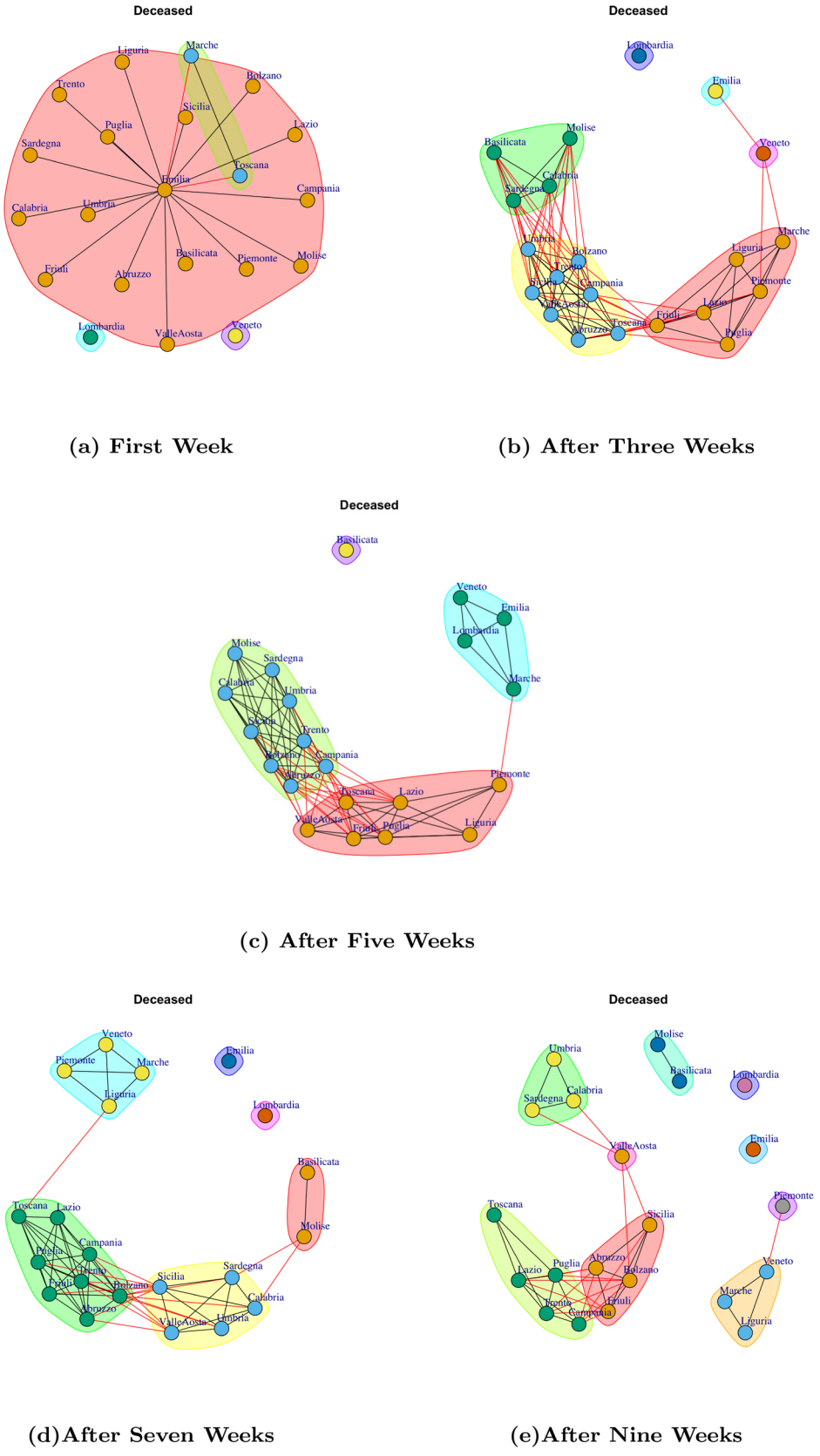
(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

Fig. 12 Evolution of Discharged/ Healed Network Communities in the first observation period



(a) First Week

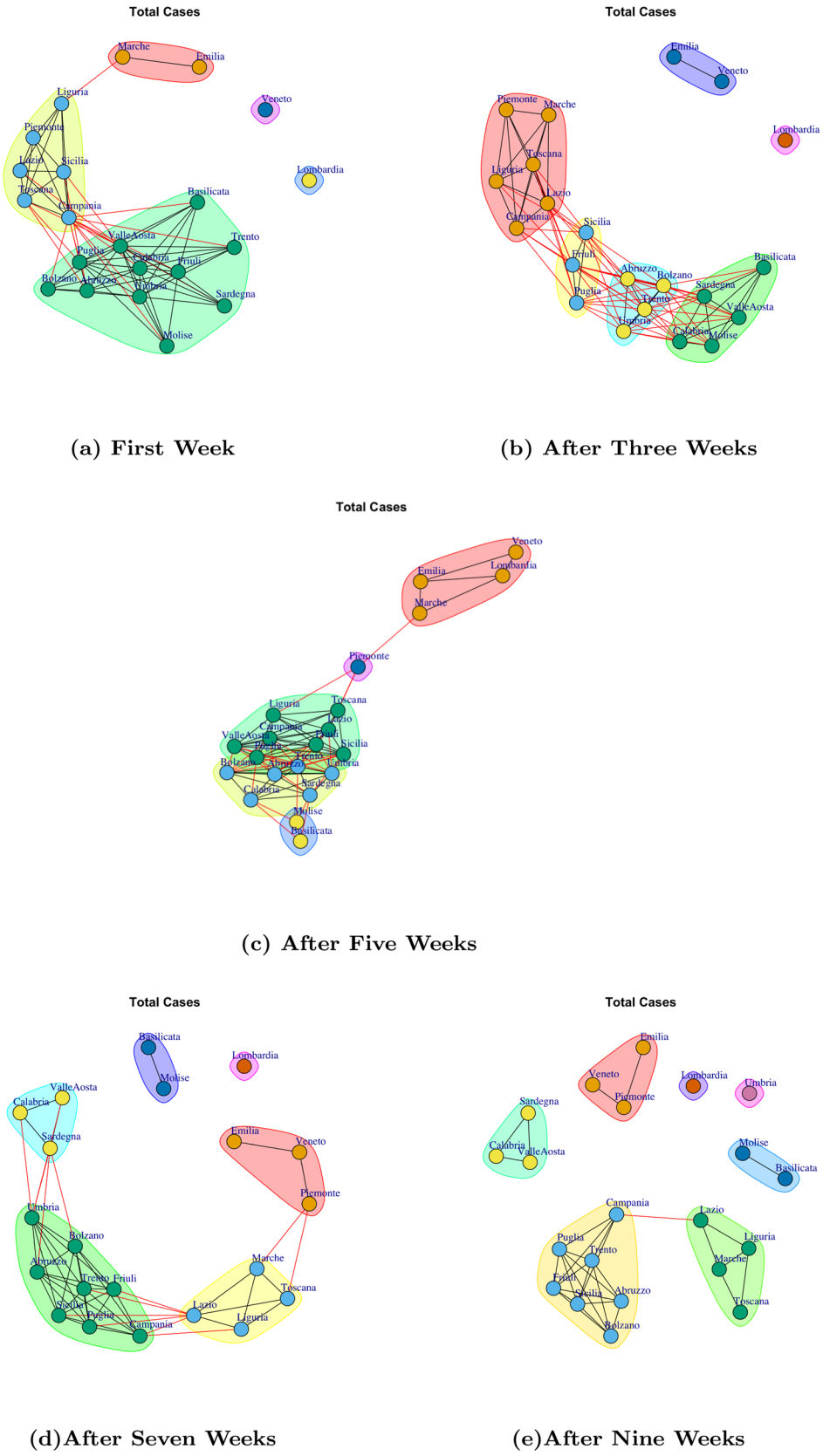
(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

Fig. 13 Evolution of Deceased Network Communities in the first observation period



(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

Fig. 14 Evolution of Total Cases Network Communities in the first observation period

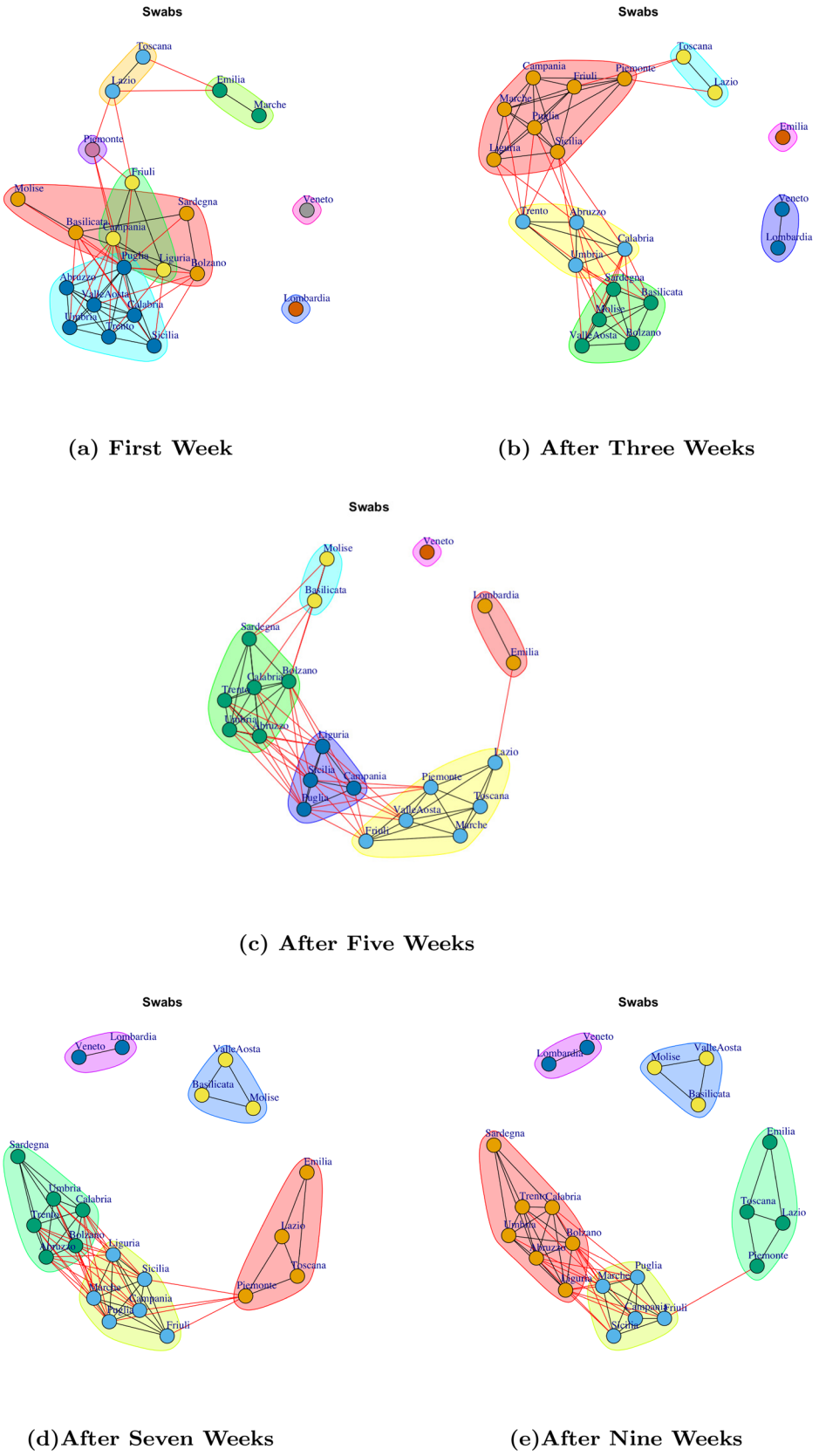


Fig. 15 Evolution of Swabs Network Communities in the first observation period

(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

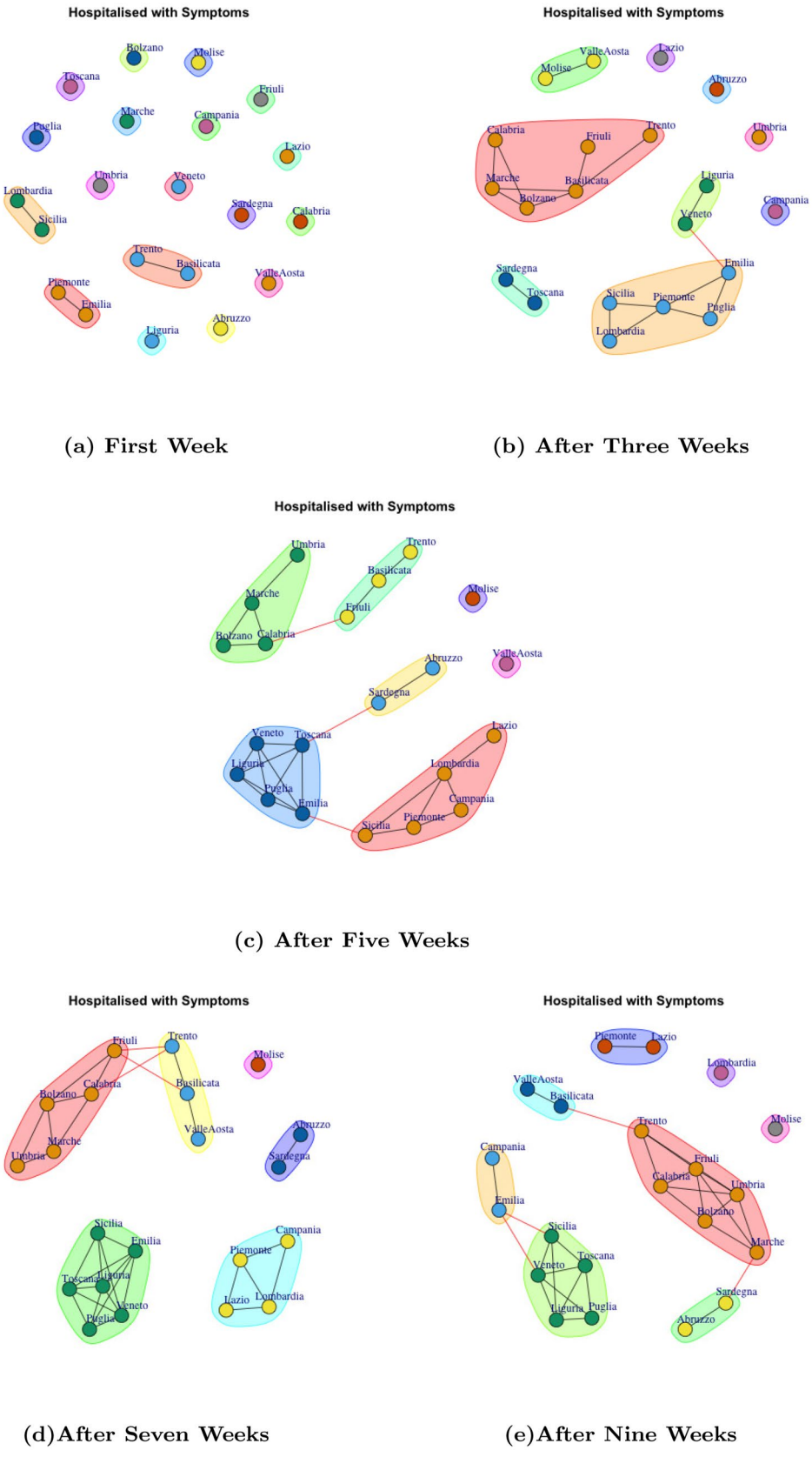


Fig. 16 Evolution of Hospitalised with Symptoms Network Communities in the second observation period

(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

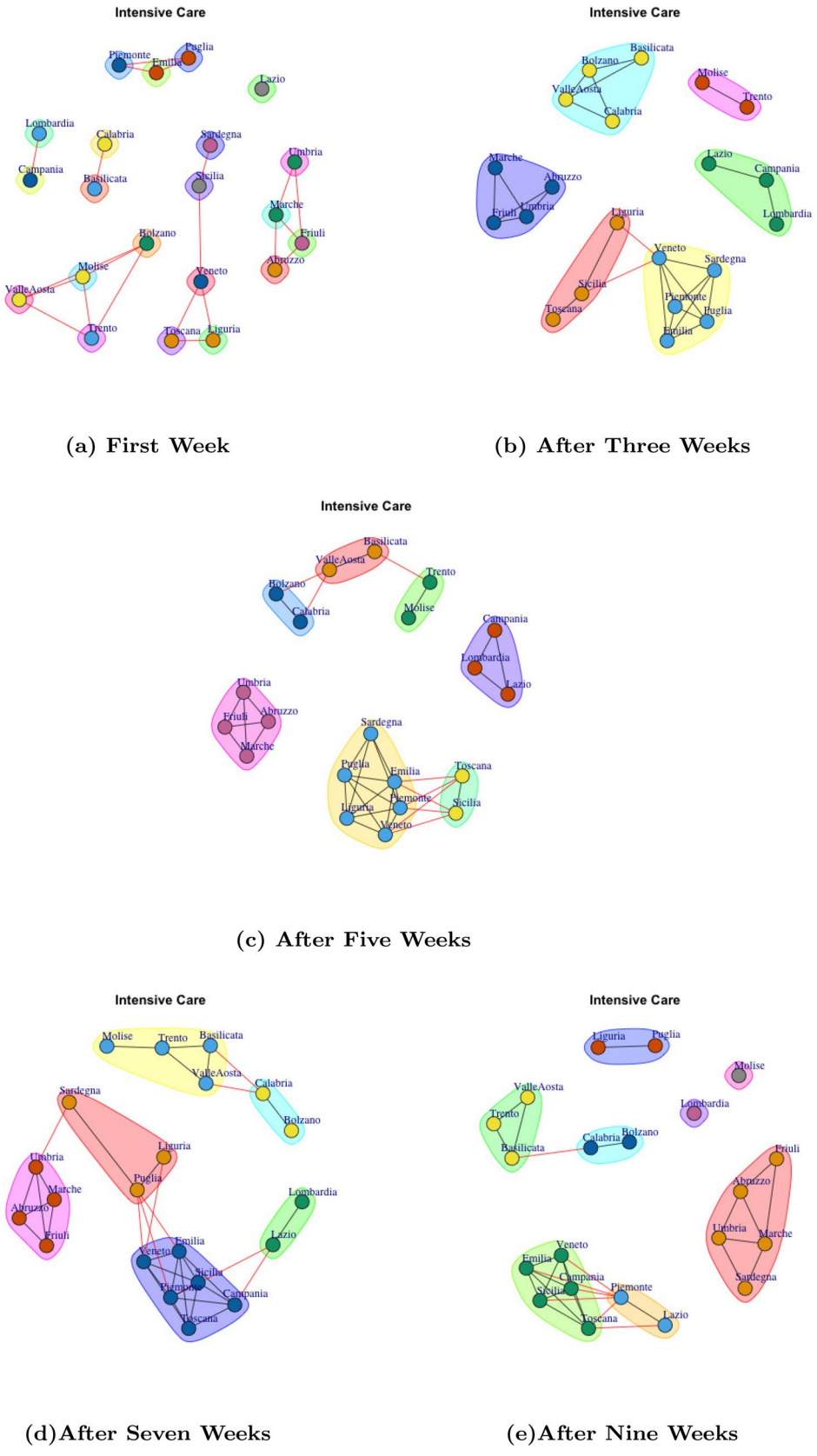


Fig. 17 Evolution of Intensive Care Network Communities in the second observation period

(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

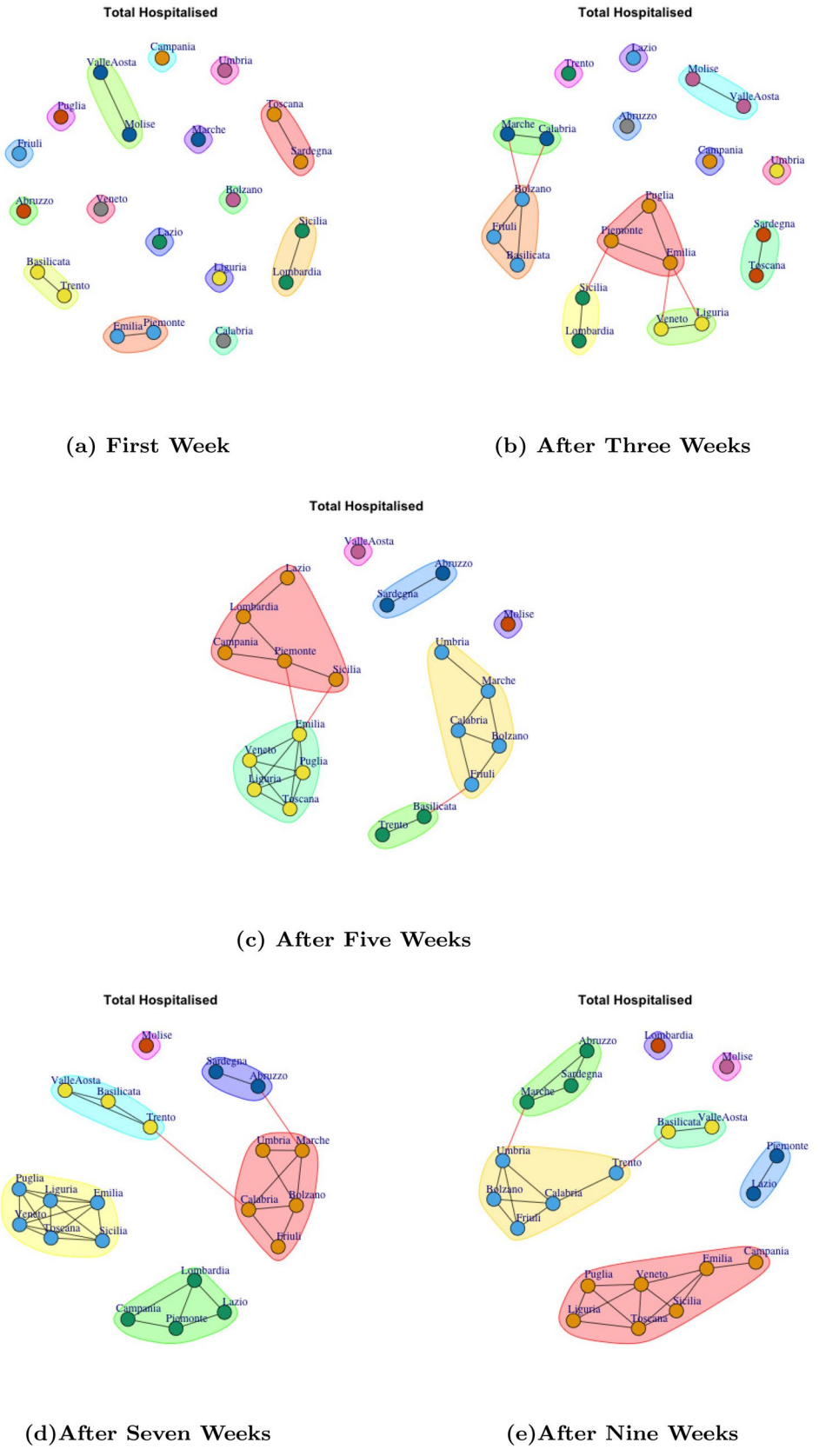
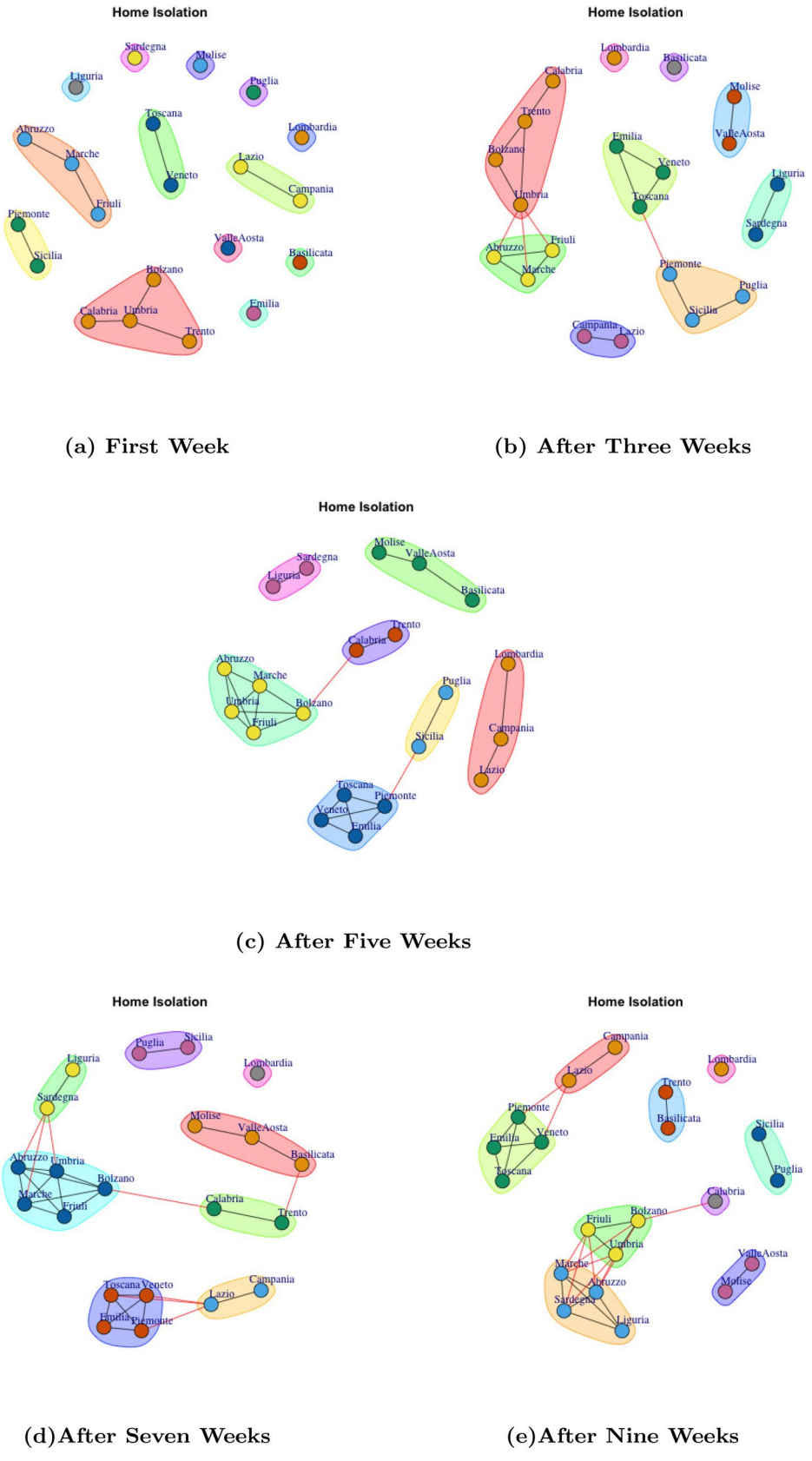


Fig. 18 Evolution of Total Hospitalised Network Communities in the second observation period



(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

Fig. 19 Evolution of Home Isolation Network Communities in the second observation period

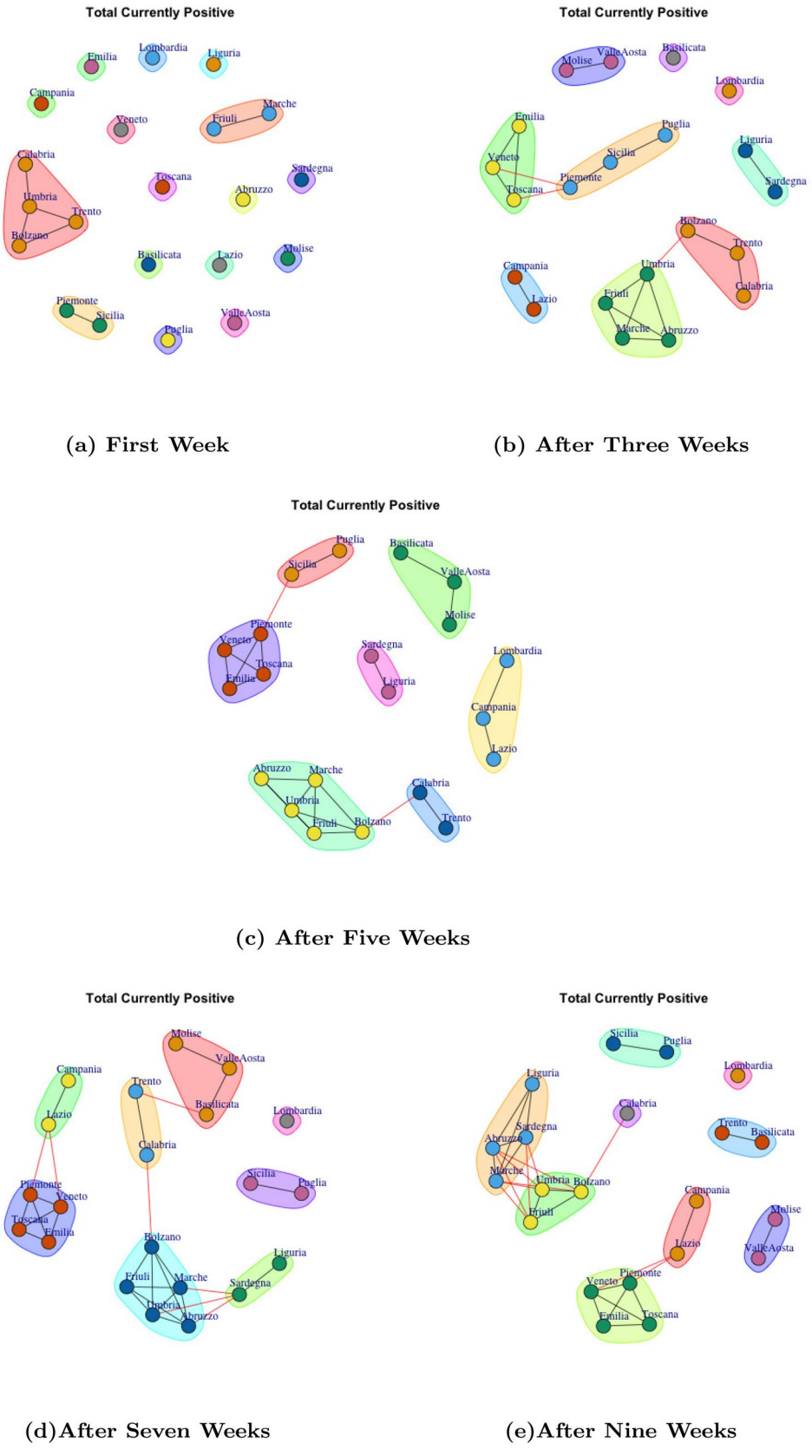


Fig. 20 Evolution of Total Currently Positive Network Communities in the second observation period

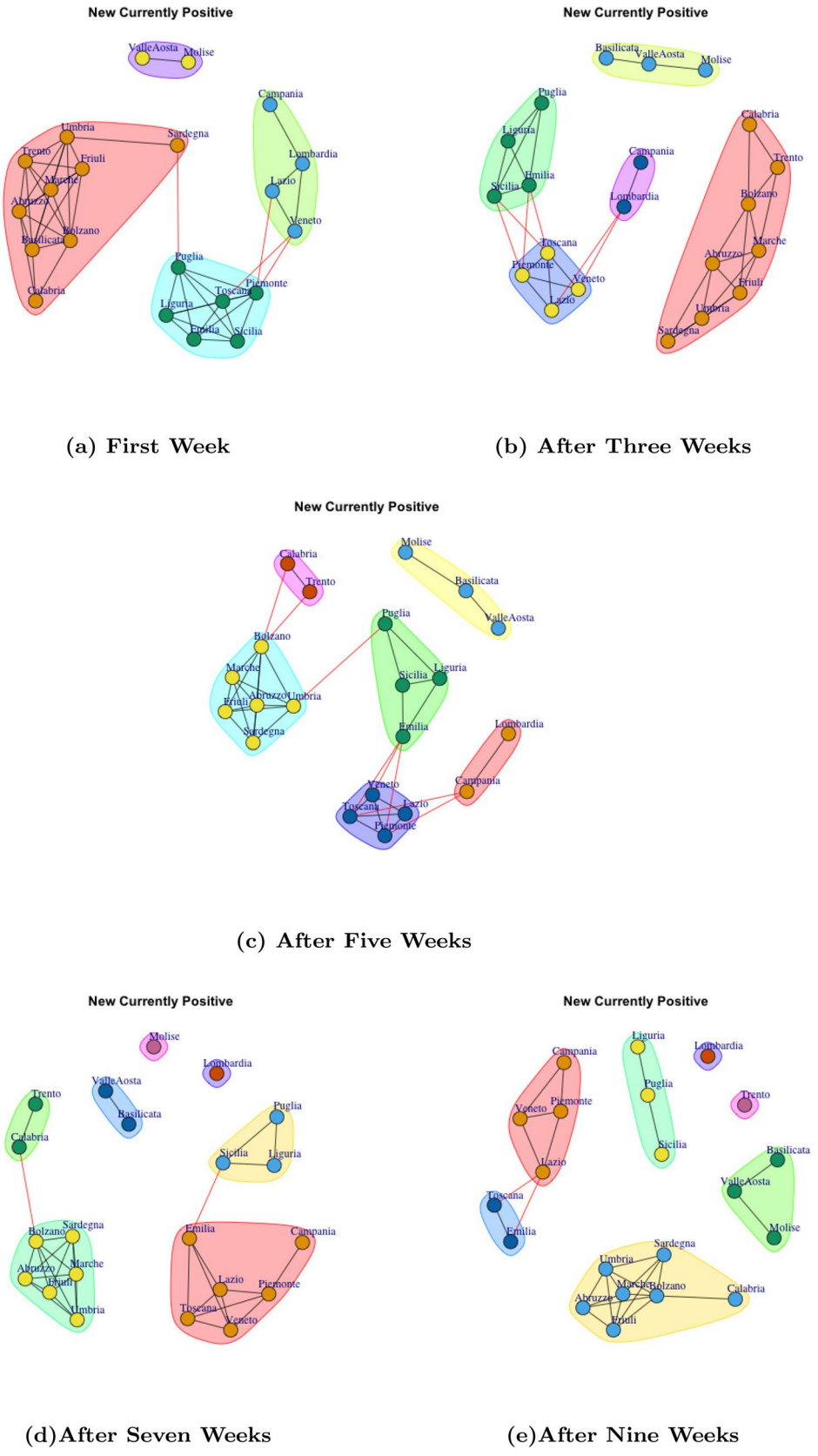


Fig. 21 Evolution of New Currently Positive Network Communities in the second observation period

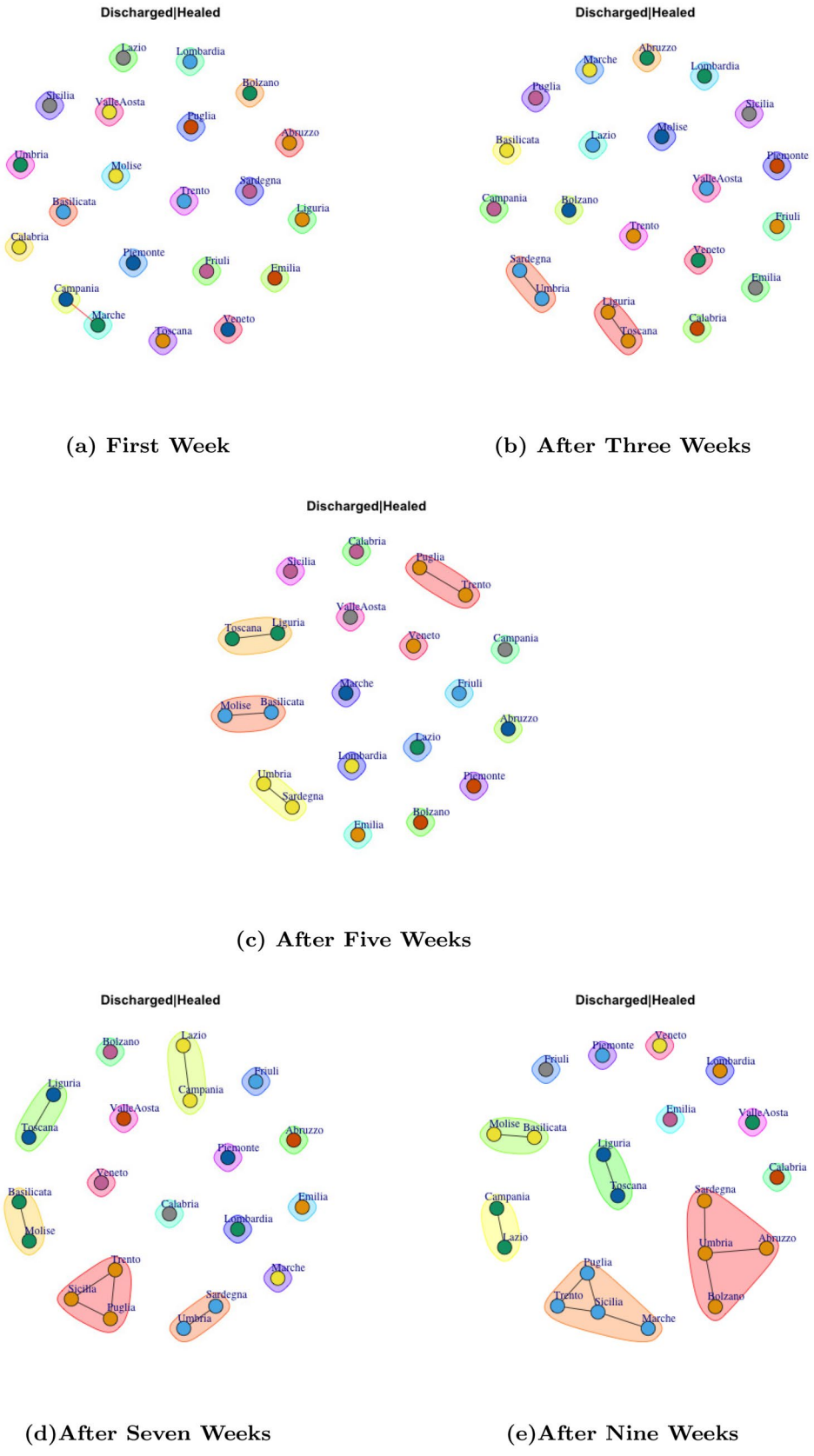


Fig. 22 Evolution of Discharged/ Healed Network Communities in the second observation period

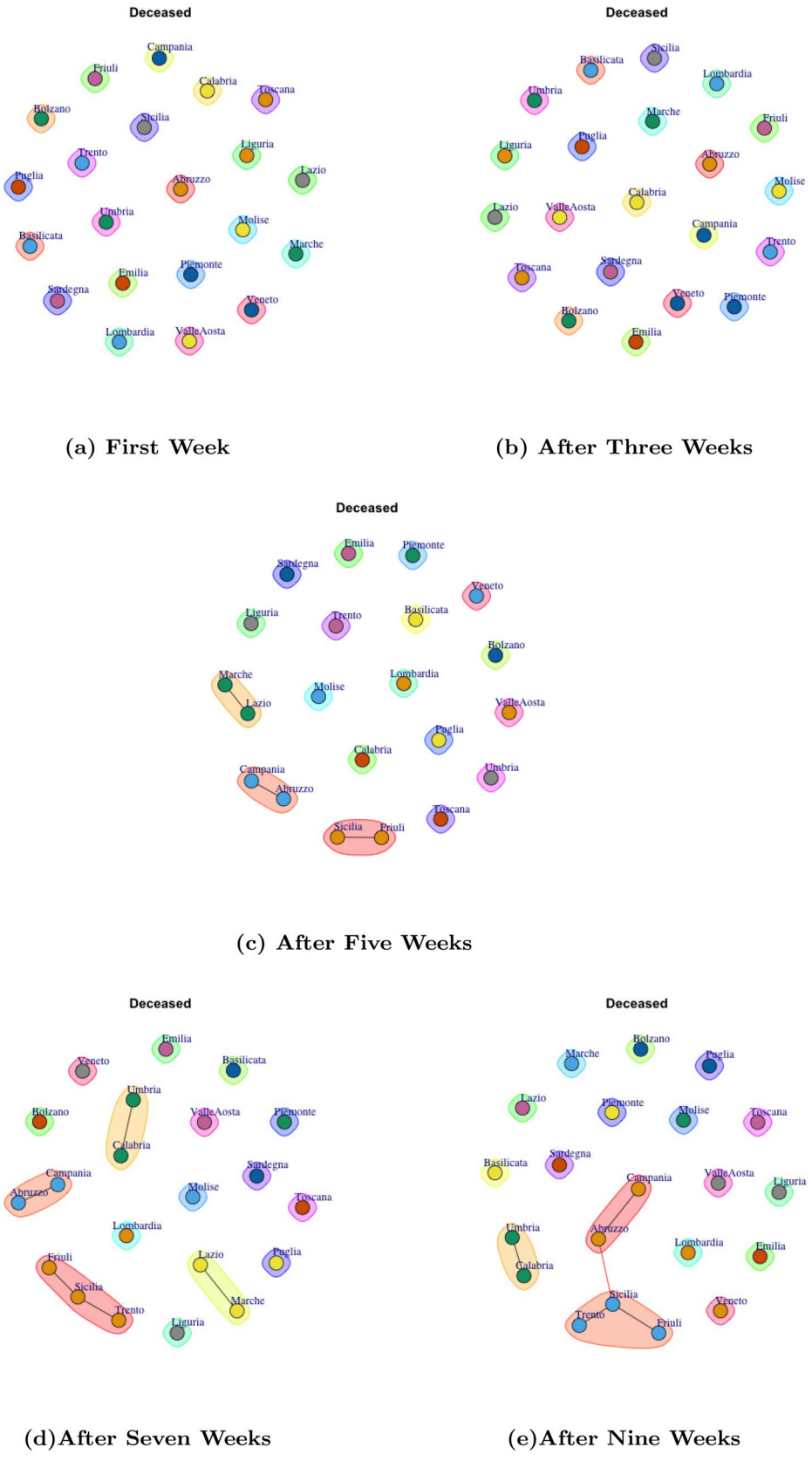


Fig. 23 Evolution of Deceased Network Communities in the second observation period

(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d) After Seven Weeks

(e) After Nine Weeks

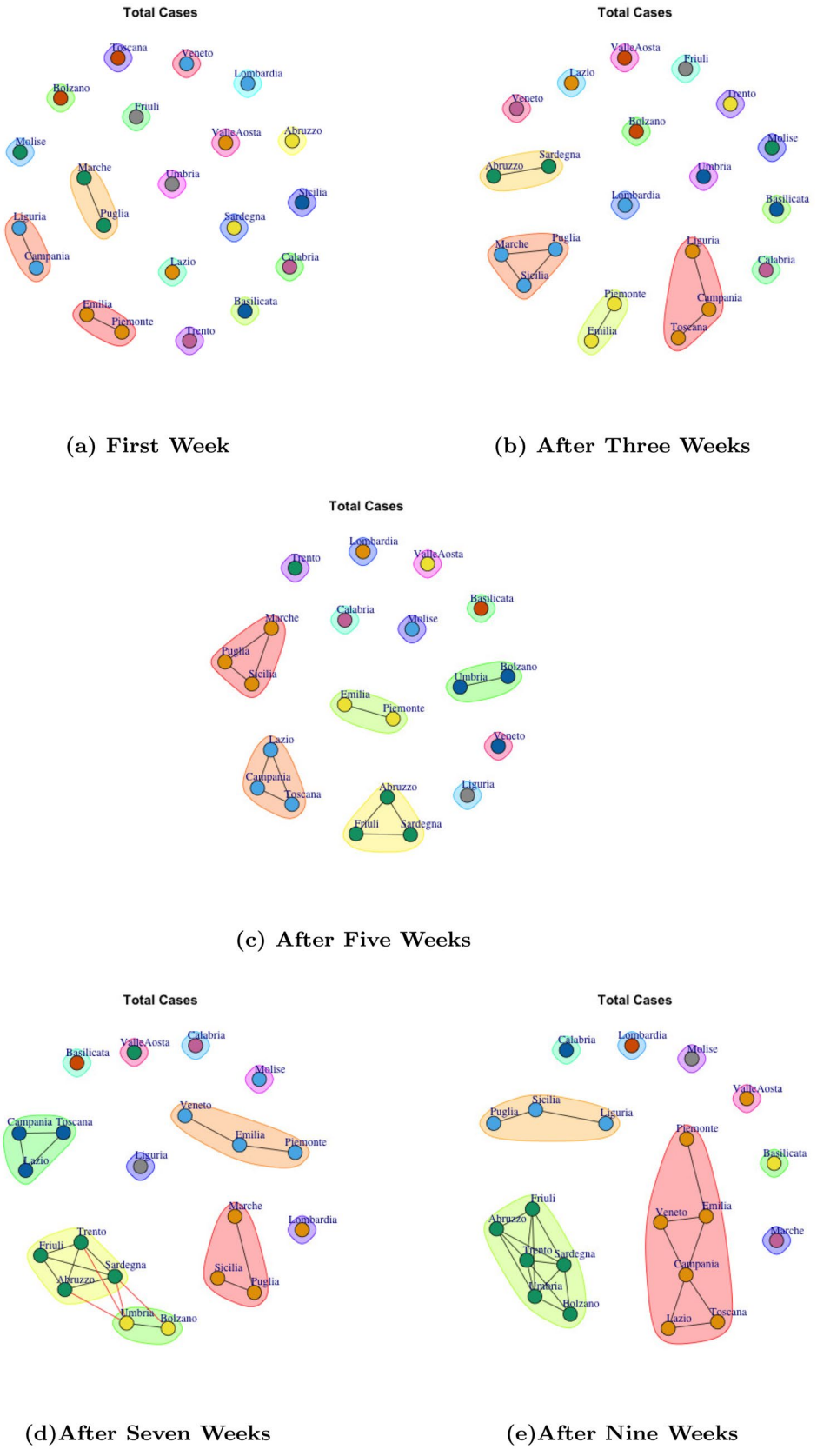
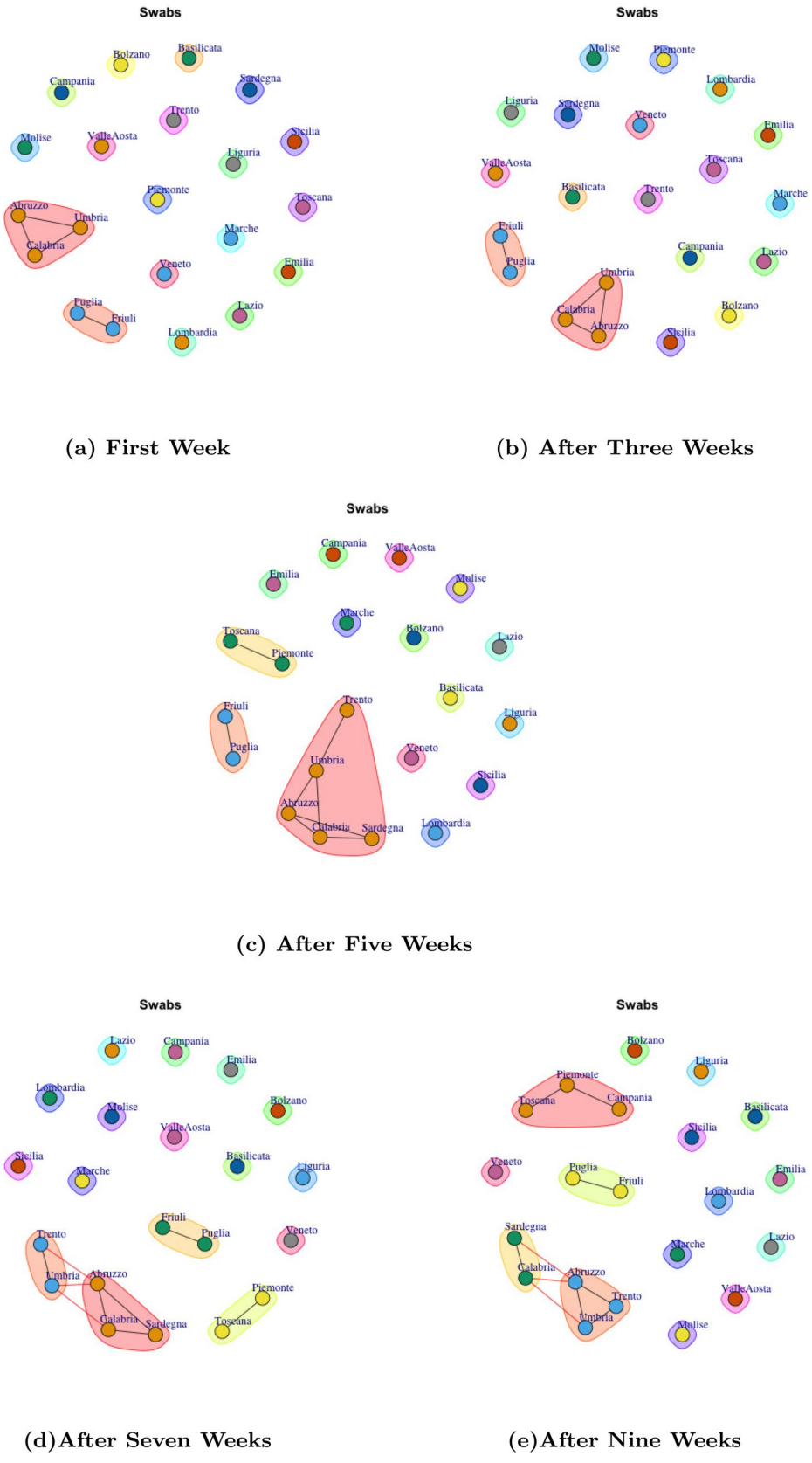


Fig. 24 Evolution of Total Cases Network Communities in the second observation period



(a) First Week

(b) After Three Weeks

(c) After Five Weeks

(d)After Seven Weeks

(e)After Nine Weeks

Fig. 25 Evolution of Swabs Network Communities in the second observation period

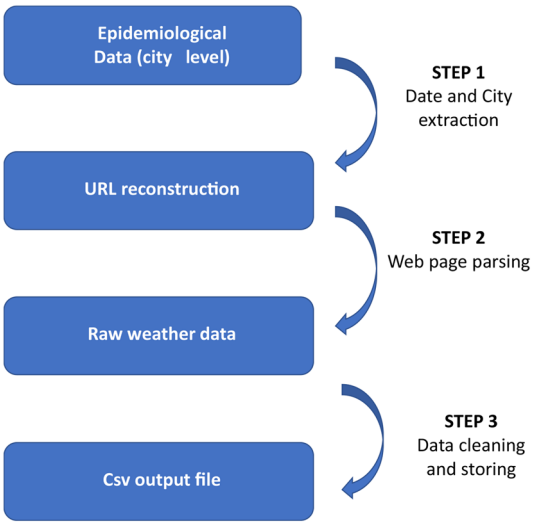


Fig. 26 Weather data collection pipeline. A list of dates and city names are extracted from the epidemiological data previously described. Then, the cities are combined with all dates to reconstruct the URL of each page that contains historical weather information on the meteorological website. Data is extracted through data scraping, then raw data is normalized and cleaned. Finally, data are stored in a CSV file format

period with the Intensive Care network, Total Hospitalised network and Total Currently Positive network by applying MAGNA++. After that, we aligned the climate network for the second observation period with the Intensive Care network, Total Hospitalised network, and Total Currently Positive network. We ran MAGNA++ as the global aligner with the following parameters: S3 as a measure of Edge Conservation, the Edge-Node Weight equals to 0, to consider only topologies, whereas the population size, number of generation, fraction of elite members were set to default values. At the end of this step, 3 global alignments were built that integrate respectively Intensive Care, Total Hospitalised and Total Currently Positive network data with climate data for the first observation period (Fig. 30a–c) and 3 global alignments for the second observation period (Fig. 31a–c). Finally, we applied the WalkTrap community finding algorithm to mine the communities for the global alignments built. Figure 30a1–c1 show the identified communities related to networks built with climate data interaction in the first observation period. Figure 31a1–c1 show the identified communities related to networks built with climate data interaction in the second observation period. The goal is to evaluate how the communities vary by integrating the climate data.

4 Results and discussion

In this section, we analyze the temporal evolution of the detected communities in the fist and second observation period and the new communities obtained by integrating climate data with Intensive Care, Total Hospitalised and Total Currently Positive data. The aim is to highlight that the communities may be diverse according to (i) different data analyzed and (ii) different time intervals when considering the same data.

For the first observation period the temporal evolution of the communities is computed at the following different time intervals:

- at the end of the first week (February 24–March 1);
- after 3 weeks (February 24–March 15);
- after 5 weeks (February 24–March 29);
- after 7 weeks (February 24–April 12);
- after 9 weeks that we called the observation period (February 24–April 26);

Then, we considered the temporal evolution of the communities for the second observation period at following different time intervals:

- at the end of the first week (September 29–October 4);
- after 3 weeks (September 29–October 18);
- after 5 weeks (September 29–November 1);
- after 7 weeks (September 29–November 15);
- after 9 weeks that we called the second observation period (September 29–November 29).

Furthermore, we analyzed the communities on the networks obtained by integrating climate data with three types of Italian COVID-19 data, Intensive Care, Total Hospitalised, and Total Currently Positive data.

Our goal is to assess: (i) if diverse COVID-19 measures show similar or dissimilar communities and (ii) if the communities are similar or dissimilar analyzing diverse time intervals on the same COVID-19 measure.

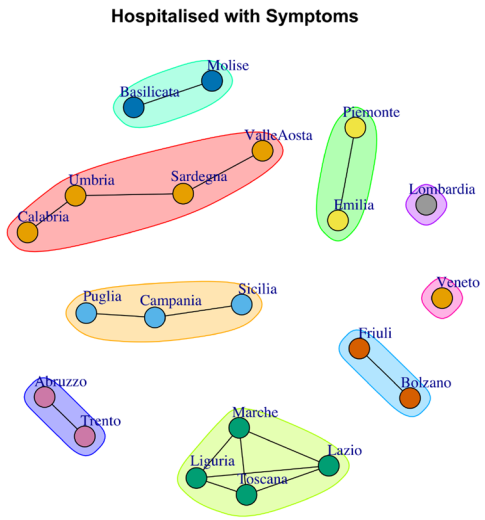
4.1 First observation period (February 24–April 26, 2020 (1st wave))

We analyze Fig. 6 that shows the development of Hospitalised with Symptoms Network Communities. In the first week, six communities are identified (Fig. 6a): (1) Lombardy, (2) Veneto, (3) Emilia and Marche, (4) Liguria, Tuscany and Piedmont, (5) Puglia, Lazio, Campania, Abruzzo, Bolzano and Sicily, (6) Umbria, Sardinia, Calabria, Molise, Valle d'Aosta, Friuli, Basilicata, Trento. After 3 weeks, the regions leave the previous communities, and they move to

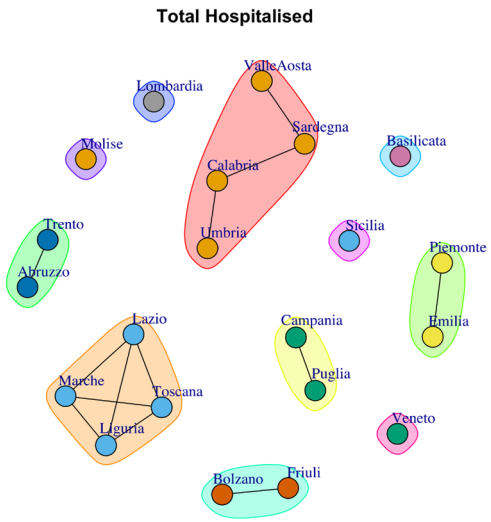
other ones. For example, Veneto, that was an alone community in the first week, migrates to another community after 3 weeks. Also, Emilia represents a single community, whereas it formed a community with Marche in the first week. In Fig. 6b all identified communities after 3 weeks are reported: (1) Lombardy, (2) Emilia, (3) Veneto, Marche, Piedmont, Liguria, Tuscany and Lazio, (4) Trento, Bolzano, Abruzzo, Friuli, Sicily, Puglia, (5) Campania, Umbria, Sardinia, Calabria, Molise, Valle d'Aosta, Basilicata. After 5 weeks, five communities are extracted as Fig. 6c depicts. The first community is formed by Basilicata that becomes a single one by leaving the previous community; the second community is formed by Piedmont, Marche, Emilia, Lombardy and Veneto; the third community is formed by Liguria, Lazio, Tuscany; the fourth one is composed by Campania, Puglia, Sicily, Abruzzo, Valle d'Aosta, Friuli, Trento and Bolzano; the last community consists of Umbria, Sardinia, Calabria, Molise. Figure 6d shows the development of the communities after 7 weeks. The initial network is split into seven groups, and eight communities are formed. (1) The first one consists Lombardy; (2) the second community is represented by Veneto; (3) the third one consists of Piedmont and Emilia; (4) the fourth one consists of Molise and Basilicata; (5) the fifth one consists of Friuli and Bolzano; (6) Valle d'Aosta, Sardinia, Umbria and Calabria form the sixth community; (7) the seventh one is formed by Umbria, Marche, Liguria, Tuscany and Lazio; (8)Abruzzo, Sicily, Campania, Puglia, and Trento compose the eighth community. After 9 weeks, the network splits into more groups by forming nine communities, reported in Fig. 6e: (1) Lombardy; (2) Veneto; (3) Piedmont and Emilia; (4) Molise and Basilicata; (5) Friuli and Bolzano; (6) Abruzzo and Trento; (7) Valle d'Aosta, Sardinia, Umbria, and Calabria; (8) Sicily, Campania and Puglia; (9) Marche, Liguria, Tuscany and Lazio.

It is possible to notice that the progression of the Hospitalised with Symptoms Network Communities is similar to the Total Hospitalised Network Communities, reported in Fig. 8.

In fact, in the first week, Fig. 8a, six communities are mined. The first one consists of Liguria, Tuscany and Piedmont, the second community comprises Sardinia, Umbria, Calabria, Basilicata, Valle d'Aosta, Friuli, Trento and Molise; the third community consists of Lazio, Campania, Sicily, Abruzzo, Bolzano and Puglia; the fourth one consists of Marche, Emilia; the fifth is represented by Lombardy, and the sixth one is formed by Veneto. Then, after 9 weeks, the network splits into eleven communities, seven of them are equal to the communities identified in Hospitalised with Symptoms Network: (1) Lombardy; (2) Veneto; (3) Piedmont and Emilia; (5) Friuli and Bolzano; (6) Marche, Liguria, Tuscany and Lazio, (7) Valle d'Aosta, Sardinia, Umbria and Calabria.



(a) Hospitalised with Symptoms Communities



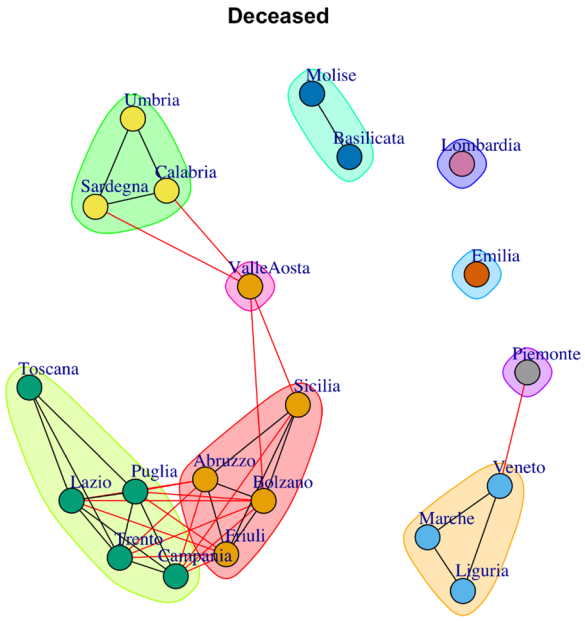
(b) Total Hospitalised Communities

Fig. 27 The comparison among Hospitalised with Symptoms Network Communities and Total Hospitalised Network Communities at the end of the 9 weeks of first observation period

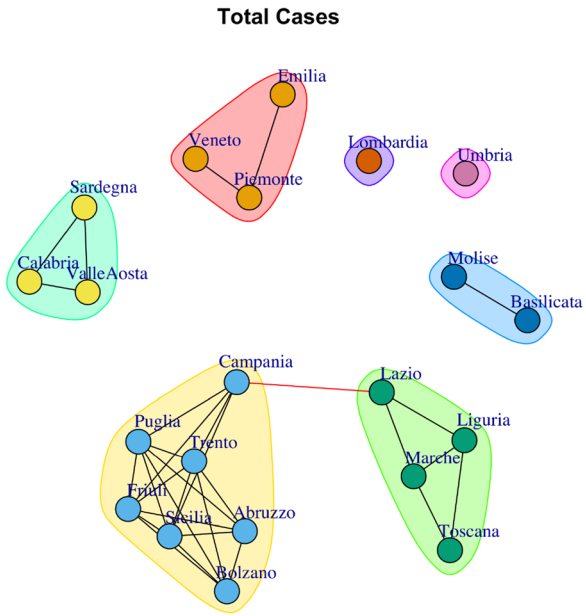
The comparison among Hospitalised with Symptoms Network Communities and Total Hospitalised Network Communities after 9 weeks is reported in Fig. 27.

Moreover, Deceased Network Communities, Fig. 13, and Total Cases Network Communities, Fig. 14, report similar evolution.

Figure 13a presents the mined communities of the Deceased Network at the end of the first week. The detected communities consist of: a considerable community formed



(a)Deceased Communities



(b)Total Cases Communities

Fig. 28 The comparison among Deceased Network Communities and Total Cases Network Communities at the end of the 9 weeks of first observation period

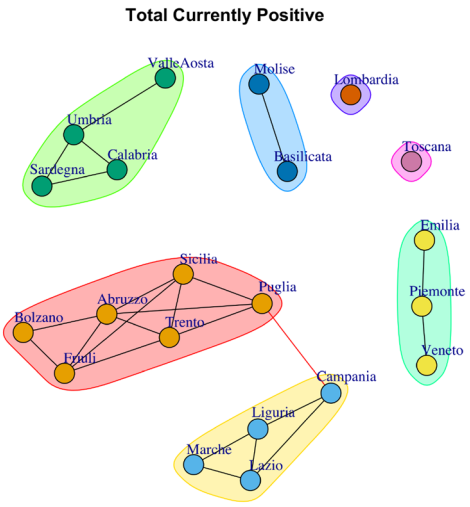
by 17 regions: Emilia, Piedmont, Liguria, Campania, Abruzzo, Puglia, Valle d'Aosta, Umbria, Calabria, Sicily, Campania, Trento, Lazio, Sardinia, Bolzano, Basilicata, Molise and Friuli; a single community formed by Lombardy: a single community represented by Veneto; and a single community consisting of Marche and Tuscany.

Figure 14a shows the five communities mined from Total Cases Network at the end of the first week. The first one comprises Liguria, Tuscany Lazio, Piedmont, Campania and Sicily; the second is composed by Sardinia, Abruzzo, Umbria, Calabria, Basilicata, Bolzano, Valle d'Aosta, Friuli, Trento, Molise and Puglia; the third cluster is represented by Marche and Emilia; the forth one comprises only Lombardy and the last one Veneto. Although the communities are different in the first week, at the end of the 9 weeks, the community evolution shows a similar trend. In fact, both networks split into more similar communities. The comparison among Deceased Network Communities and Total Cases Network Communities in the observation period is reported in Fig. 28.

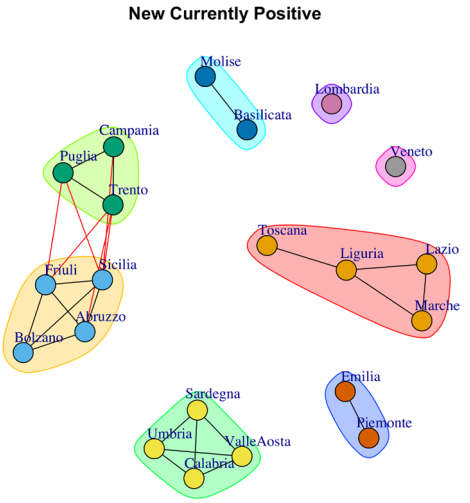
Also, Total Currently Positive Network Communities, Fig. 10, and New Currently Positive Network Communities, Fig. 11 report similar evolution.

Figure 10 shows the development of Total Currently Positive Communities. Figure 10a depicts the detected community at the end of the first week. The first one comprises Umbria, Sardinia, Basilicata, Molise, Friuli Tuscany, Calabria, Valle d'Aosta and Trento; the second community is represented by Bolzano, Lazio, Abruzzo and Puglia; the third one is represented by Campania, Sicily and Liguria; Piedmont and Lazio form the forth community; Marche forms the fifth community; the sixth one is formed by Emilia; the seventh is composed of Lombardy and the eight community consists of Veneto. At the end of 5 weeks, the detected communities further decline, as reported in Fig. 10c. The first one comprises Veneto, Lombardy, Emilia and Marche; the second one is represented by Piedmont; the third one groups Basilicata, Molise, Calabria and Sardinia; the fourth one comprises Tuscany, Lazio, Friuli, Valle d' Aosta, Sicily, Campania, Liguria Puglia, Abruzzo, Umbria, Trento and Bolzano.

Figure 11 depicts the development of New Currently Positive Communities. Figure 11a shows the detected communities at the end of the first week. Lombardy, Veneto and Emilia composes a single communities, and then there are two considerable communities: the first community consists of Marche, Piedmont, Liguria, Campania, Abruzzo and Tuscany; the second one is formed by Puglia, Valle d'Aosta, Umbria, Calabria, Sicily, Campania,



(a) Total Currently Positive Communities



(b) New Currently Positive Communities

Fig. 29 The comparison among Total Currently Positive Network Communities and New Currently Positive Network Communities at the end of the 9 weeks of first observation period

Trento, Lazio, Sardinia, Bolzano, Basilicata, Molise and Friuli. After 5 weeks, the number of communities further decline, as reported in Fig. 11c. In fact, four communities are detected: (1) Piedmont, Marche, Tuscany; (2) Lombardy, Veneto and Emilia, (3) Basilicata and Molise, (4) Puglia, Friuli, Valle d'Aosta, Lazio, Abruzzo, Umbria, Campania, Trento, Liguria Bolzano, Calabria, Sardinia and Sicily.

Figures 10e and 11e report the Total Currently Positive Network Communities and New Currently Positive Network Communities after 9 weeks. It is possible to notice that both networks split into more similar groups such as: (1) Lombardy community; (2) Basilicata and Molise community; (3) Valle d'Aosta, Umbria, Calabria and Sardinia community; (4) Emilia and Piedmont community.

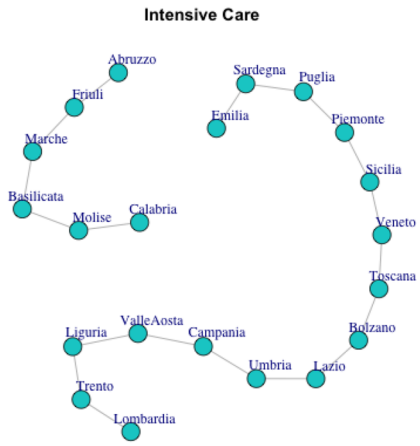
The comparison among Total Currently Positive Network Communities and New Currently Positive Network Communities at the end of the 9 weeks is reported in Fig. 29 in the first observation period.

In contrast, there are different networks such as, Intensive Care Network, Discharged/Healed Network, Swabs Network, whose communities evolve differently over time, resulting reduced at the end of 9 weeks due to several members that leave different groups, see Figs. 7, 12 and 15.

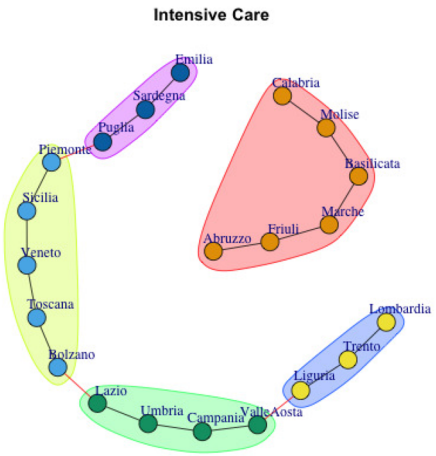
4.2 Second observation period (September 28–November 29, 2020 (2nd wave))

By considering the networks built in the second observation period it is possible to note that the topology evolves from a sparse to a dense structure and this reflects on the discovered communities. In fact, by analyzing the first week in Hospitalised with Symptoms Network (Fig. 15a), Total Hospitalised Network Communities (Fig. 17a), Total Currently Positive Network Communities (Fig. 19a), Discharged/ Healed Network Communities (Fig. 21a), Deceased Network Communities (Fig. 22a), Total Cases Network Communities (Fig. 23a), it is possible to notice that the Italian regions represent single communities that is reflected by the sparse network topology. Especially, in Deceased Network, each region forms a single community in the first week and after three weeks, whereas, after five weeks, three communities are composed by two regions (Fig. 22c): (i) Marche and Lazio, (ii) Campania and Abruzzo, (iii) Sicily and Friuli. After seven weeks (Fig. 22d), the Italian regions form single communities with the exception of three communities formed respectively by: Umbria and Calabria, Campania and Abruzzo, Lazio and Marche; and a community formed by three regions, Friuli, Sicily and Trento. Instead, after 7 weeks (Fig. 22e), the community composed by Lazio and Marche splits into two single communities whereas Umbria and Calabria, Campania and Abruzzo, and Friuli, Sicily, Trento continue to form communities. Furthermore, by comparing the communities extracted from the networks in the first observation period and those discovered in the second observation period, it is possible to notice a substantial diversity among them. This reflects the different impact that the spread of the virus has had on the Italian regions considering the different observation periods. In fact, the mined communities are different for each kind of COVID-19 measure and for each considered week. This implies that highlighted original clusters of regions with respect to COVID-19 data are discovered.

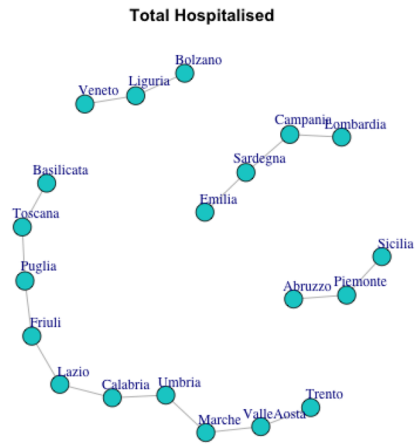
Thus, it should be noted that the network representation of similar/dissimilar regions allows showing in a graphical



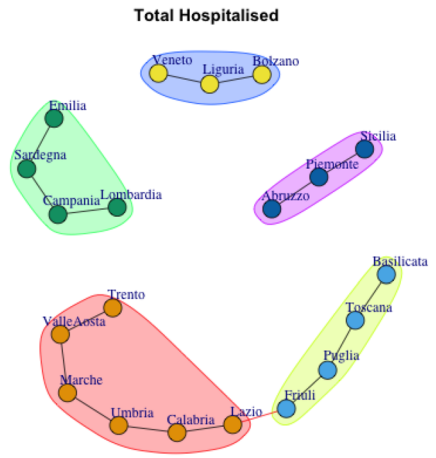
(a) Intensive Care Network



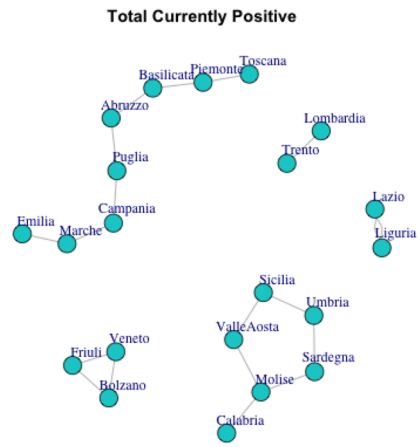
(a1)Intensive Care Communities



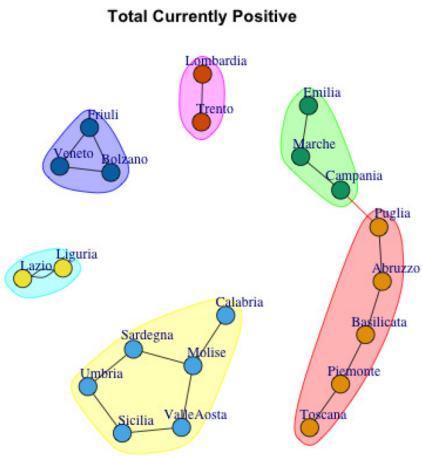
(b) Tot. Hospitalised Network



(b1)Tot. Hospitalised Communities

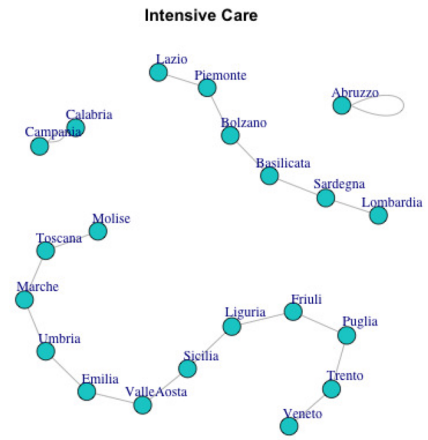


(c) Tot. Currently Positive Network

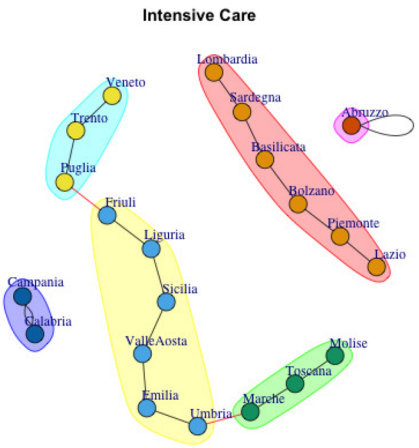


(c1) Tot. Currently Positive Communities

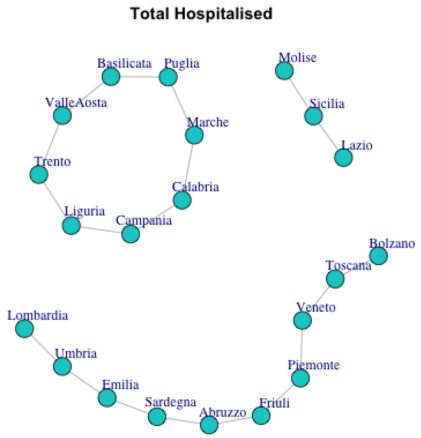
Fig. 30 Networks built integrating climate in the first observation period. The formed communities take climate data into account



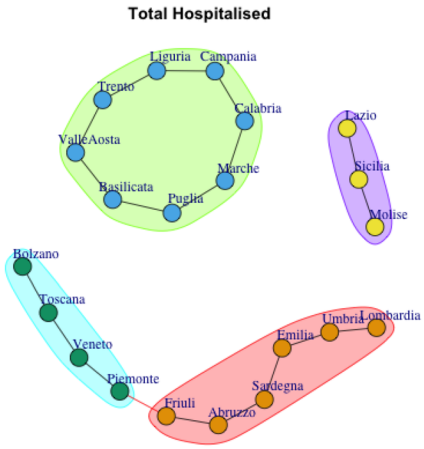
(a) Intensive Care Network



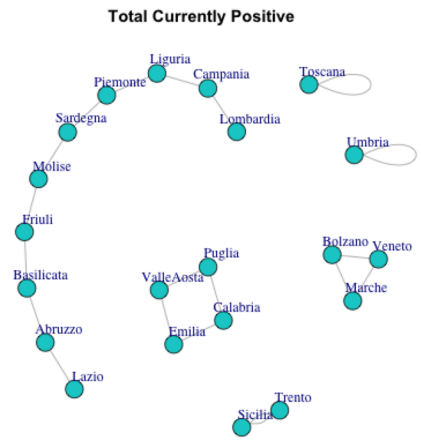
(a1)Intensive Care Communities



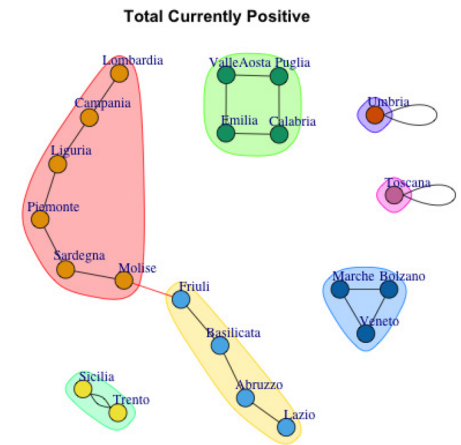
(b) Tot. Hospitalised Network



(b1)Tot. Hospitalised Communities



(c) Tot. Currently Positive Network



(c1) Tot. Currently Positive Communities

Fig. 31 Networks built integrating climate in the second observation period. The formed communities take climate data into account

way how this varies over time. In addition, community detection allows to better identify clusters of regions that can be shown both graphically and also considering objective measures such as community (cluster) centroid (e.g. mean measure). In fact, it is possible to notice that the community centroids are different for each individual community surveyed, considering both the different observation periods and the individual weeks. Therefore this is an indication of the variability of the communities.

The results evidence that community structure evolves. For example, the communities grow due to joining of regions or the communities reduces due to leaving of regions. This allows to evaluate the changes of community coherence in relation to different data and along time. We showed that the regions most affected by the epidemic (Lombardy, Veneto, Piedmont, and Emilia) behaved differently than the other regions. This aspect is reflected in the community detection analysis in which those regions formed a single community or a community among them. In addition, this study also allows to highlight that regions geographically distant may show similar behaviours and form community. For instance, far regions like Calabria, Sardinia, and Molise, form a community in Hospitalized with Symptoms Network, Total Hospitalized Network, Total Currently Positive Network, Discharged/Healed Network, Total Cases, Deceased Network, Intensive Care Network.

In conclusion the CCTV methodology is able to provide a snapshot of COVID-19 measures at a given time interval or depict their temporal evolution. The analysis of the results could lead to the identification of: (i) an event that did not cause a region to move from an initial community, (ii) an event that caused a region to move to a more supportive community, for example where fewer people were recorded cases or (iii) an event that led to a region to move into one. Once the event is identified, it might be possible to plan interventions to improve the behavior of critical regions such as increasing the number of intensive care units or swab tests. In this work, we conducted the analyzes at the regional level but the methodology can be applicable to different zooms, such as the analysis of the behavior of the cities in a region or the behavior of the states of in a country (e.g. Europe). Therefore, our methodology can be considered as a support tool for political and health decisions.

4.3 Integrating climate data with COVID-19 measures by using network alignment method

Finally, we wanted to integrate climate data to correlate them with a selection of three types of Italian COVID-19 measures, using network alignment techniques. At first, we built three new networks relating to the first observation period, three networks relating to the second observation period that integrate the chosen COVID-19 measures with climate data, and we extracted the communities. As it is possible to see in Figs. 30 and 31), the communities related to Intensive Care, Total Hospitalised and Total Currently Positive networks including climate data are different if compared with those extracted from the networks both for the first and for the second period on the same networks. This implies that the variation of the climate in the Italian territory affects the identification of the communities for Intensive Care, Total Hospitalised and Total Currently Positive data.

Let's consider the example of Lombardy in relation to Intensive Care Units data. In the first observation period, Lombardy formed a single community, except 5 weeks after the start of the pandemic in which it forms a community with Veneto. In the second observation period, Lombardy initially forms a community with Campania, then Lazio joins these two. After that, Campania leaves the community and finally, until 9 weeks, Lombardy forms a single community.

Instead, by considering the networks obtained by integrating the climate data, we can see that in the first observation period Lombardy forms a community with Trento, while in the second period it forms a community with Bolzano, Basilicata and Sardinia. Starting from these results, it is possible to build a forecast model based on climate data.

Hence the CCTV methodology can be used as a decision support tool. For example, given a target, such as a maximum number of infected per inhabitants, it could be analyze what variations in the climate data may modify those infected per inhabitants data. Therefore, a possible rule may be: if the climate remains in the positive (i.e. it positively affects the COVID-19 measures) side , there is no alarm on infected per inhabitants data, on the other hand, if the climate becomes negative (i.e. it negatively affects the COVID-19 measures), interventions such as avoiding travel or planning a lockdown can be planned. So, as affirmed before, our methodology can be considered as a tool to support different public health policy decisions.

5 Conclusion and future work

The COVID-19 pandemic has spread worldwide very quickly. In Italy, the outbreak of COVID-19 has begun in the north of the State, and it immediately included all the regions. In this work, we introduced a new methodology, named CCTV (COVID-19 Community Temporal Visualizer), to represent COVID-19 data as networks and to apply graph-based methods to evidence regions with similar behaviour along the time. The methodology allows to analyze the impact of clinical evolution of COVID-19 pandemic by integrating several clinical, pollution and climate data

along geographical and temporal dimensions, for two temporal periods. Since the methodology is able to provide a regional zoom of the trend of different factors, e.g. the count of COVID-19 patients in Intensive Care Units, the results could support the planning of interventions that can be conducted on regions that exhibit a similar response to the pandemic. Otherwise, it is possible to use the results as a baseline to conduct detailed analysis of why regions responded differently to the pandemic. The present methodology is general and can be applied for the network analysis of different data varying over time. As future work, we plan to implement the graphical user interface of CCTV, to make easier to use it, allowing dynamic data analysis configuration, as well as the selection of various community detection algorithms. Also, we plan to apply CCTV methodology to different PPI (Protein-Protein Interaction) human-COVID-19 to analyze the evolution over time.

Acknowledgements MM, CZ and MC have been partially supported by Data Analytics Research Center, Department of Medical and Surgical Sciences, University of Catanzaro. The authors wish to thank the Italian Protezione Civile Department for freely providing online COVID-19 data. Also, the authors would like to thank the Italian web site https://www.ilmeteo.it/portale/archivio-meteo/ as the source of weather information.

Availability of data and materials Data and supplementary materials are available at https://github.com/mmilano87/analyzeC19D

Declarations

Conflict of interest The authors declare that they have not conflict of interests.

Ethics approval and consent to participate Not applicable.

Consent to publish Not applicable.

References

Agapito G, Guzzi PH, Cannataro M (2017) Challenges and opportunities for visualization and analysis of graph-modeled medical data. Preprints 1:1. https://doi.org/10.20944/preprints201710.0018.v1

Agapito G, Zucco C, Cannataro M (2020) Covid-warehouse: a data warehouse of italian covid-19, pollution, and climate data. Int J Environ Res Public Health 17(15):5596

Asur S, Parthasarathy S, Ucar D (2009) An event-based framework for characterizing the evolutionary behavior of interaction graphs. ACM Trans Knowl Disc Data (TKDD) 3(4):1–36

Bashir MF, Ma B, Komal B, Bashir MA, Tan D, Bashir M et al (2020) Correlation between climate indicators and COVID-19 pandemic in New York, USA. Sci Total Environ 728:138835

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech 10:P10008

Briz-Redón Á, Serrano-Aroca Á (2020) The effect of climate on the spread of the covid-19 pandemic: a review of findings, and statistical and modelling techniques. Progress Phys Geogr 44(5):591–604

Bródka P, Kazienko P, Kołoszczyk B (2012) Predicting group evolution in the social network. In International Conference on Social Informatics, pages 54–67. Springer

Casanova LM, Jeon S, Rutala WA, Weber DJ, Sobsey MD (2010) Effects of air temperature and relative humidity on coronavirus survival on surfaces. Appl Environ Microbiol 76(9):2712–2717

Caspi G, Shalit U, Kristensen SL, Aronson D, Caspi L, Rossenberg O, Shina A, Caspi O (2020) Climate effect on COVID-19 spread rate: an online surveillance tool. MedRxiv. https://doi.org/10.1101/2020.03.26.20044727

Chakrabarti D, Kumar R, Tomkins A (2006) Evolutionary clustering. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 554–560

Cheval S, Mihai AC, Georgiadis T, Herrnegger M, Piticar A, Legates DR (2020) Observed and potential impacts of the COVID-19 pandemic on the environment. Int J Environ Res Public Health 17(11):4140

Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70(6):066111

Csardi G, Nepusz T et al (2006) The igraph software package for complex network research. Int J Complex Syst 1695(5):1–9

De Natale G, De Natale L, Troise C, Marchitelli V, Coviello A, Holmberg KG, Somma R (2020) The evolution of covid-19 in italy after the spring of 2020: An unpredicted summer respite followed by a second wave. Int J Environ Res Public Health, 17(23). ISSN 1660-4601. https://doi.org/10.3390/ijerph17238708. https://www.mdpi.com/1660-4601/17/23/8708

de Sousa FB, Zhao L (2014) Evaluating and comparing the igraph community detection algorithms. In: 2014 Brazilian Conference on Intelligent Systems, pages 408–413. IEEE

Delnevo G, Mirri S, Roccetti M (2020) Particulate matter and covid-19 disease diffusion in emilia-romagna (Italy). Already a cold case? Computation, 8(2). ISSN 2079-3197. https://doi.org/10.3390/computation8020059. https://www.mdpi.com/2079-3197/8/2/59

Fazzini M, Baresi C, Bisci C, Bna C, Cecili A, Giuliacci A, Illuminati S, Pregliasco F, Miccadei E (2020) Preliminary analysis of relationships between COVID19 and climate, morphology, and urbanization in the lombardy region (northern italy). Int J Environ Res Public Health, 17(19). ISSN 1660-4601. https://doi.org/10.3390/ijerph17196955. https://www.mdpi.com/1660-4601/17/19/6955

Fortunato S, Hric D (2016) Community detection in networks: a user guide. Phys Rep 659:1–44

Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Networks 1(3):215–239

Gardner EG, Kelton D, Poljak Z, Van Kerkhove M, von Dobschuetz S, Greer A (2019) A case-crossover analysis of the impact of weather on primary cases of middle east respiratory syndrome. BMC Infect Dis 19(1):113

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, 2020. ISSN 0027-8424. https://doi.org/10.1073/pnas.2004978117. https://www.pnas.org/content/117/19/10484

Gehan EA (1965) A generalized wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52(1–2):203–224

Geller C, Varbanov M, Duval RE (2012) Human coronaviruses: insights into environmental resistance and its influence on the development of new antiseptic strategies. Viruses 4(11):3044–3068

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826

Gregory S (2010) Finding overlapping communities in networks by label propagation. New J Phys 12(10):103018

Herrmann HA, Schwartz J-M (2020) Using network science to propose strategies for effectively dealing with pandemics: the COVID-19 example. medRxiv. https://doi.org/10.1101/2020.04.02.20050468

Huang C, Wang Y, Li X, Ren L, Zhao J, Yi H, Zhang L, Fan G, Jiuyang X, Xiaoying G et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 395(10223):497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Kumar A (2020) Modeling geographical spread of COVID-19 in India using network-based approach. medRxiv. https://doi.org/10.1101/2020.04.23.20076489

Kuzdeuov A, Baimukashev D, Karabay A, Ibragimov B, Mirzakhmetov A, Nurpeiissov M, Lewis M, Varol HA (2020) A network-based stochastic epidemic simulator: controlling COVID-19 with region-specific policies. medRxiv. https://doi.org/10.1101/2020.05.02.20089136

Lai A, Bergna A, Acciarri C, Galli M, Zehender G (2020) Early phylogenetic estimate of the effective reproduction number of sars-cov-2. J Med Virol. https://doi.org/10.1002/jmv.25723

Lancichinetti A, Radicchi F, Ramasco JJ, Fortunato S (2011) Finding statistically significant communities in networks. PloS one 6(4)

Li X, Wu B, Guo Q, Zeng X, Shi C (2015) Dynamic community detection algorithm based on incremental identification. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pages 900–907. IEEE

Malki Z, Atlam E-S, Hassanien AE, Dagnew G, Elhosseini MA, Gad I (2020) Association between weather data and COVID-19 pandemic predicting mortality rate: machine learning approaches. Chaos, Solitons Fractals 138:110137

Ma Y, Zhao Y, Liu J, He X, Wang B, Fu S, Yan J, Niu J, Luo B (2020) Effects of temperature variation and humidity on the mortality of COVID-19 in Wuhan. medRxiv. https://doi.org/10.1101/2020.03.15.20036426

Milano M (2019) Computing languages for bioinformatics. In: Gribskov MR, Nakai K, Schonbach C (eds) Encyclopedia of bioinformatics and computational biology, vol 1. Elsevier, Oxford, pp 889–895

Milano M, Milenković T, Cannataro M, Guzzi PH (2020) L-hetnetaligner: a novel algorithm for local alignment of heterogeneous biological networks. Sci Rep 10(1):1–20

Mirri S, Delnevo G, Roccetti M (2020) Is a COVID-19 second wave possible in emilia-romagna (italy)? forecasting a future outbreak with particulate pollution and machine learning. Computation, 8(3). ISSN 2079-3197. https://doi.org/10.3390/computation8030074. https://www.mdpi.com/2079-3197/8/3/74

Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E 74(3):036104

Palla G, Barabási A-L, Vicsek T (2007) Quantifying social group evolution. Nature 446(7136):664–667

Pons P, Latapy M (2005) Computing communities in large networks using random walks. In: International symposium on computer and information sciences, pages 284–293. Springer

Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E 76(3):036106

Reich O, Shalev G, Kalvari T (2020) Modeling COVID-19 on a network: super-spreaders, testing and containment. medRxiv. https://doi.org/10.1101/2020.04.30.20081828

Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. Phys Rev E 74(1):016110

Rosvall M, Bergstrom CT (2007) An information-theoretic framework for resolving community structure in complex networks. Proc Natl Acad Sci 104(18):7327–7331

Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci 105(4):1118–1123

Samek W, Montavon G, Vedaldi A, Hansen LK, Müller K-R (eds) Explainable AI: interpreting, explaining and visualizing deep learning, volume 11700 of Lecture Notes in Computer Science. Springer, 2019. ISBN 978-3-030-28953-9. https://doi.org/10.1007/978-3-030-28954-6. https://doi.org/10.1007/978-3-030-28954-6

Tosepu R, Gunawan J, Effendy DS, Lestari H, Bahar H, Asfian P et al (2020) Correlation between weather and COVID-19 pandemic in Jakarta, Indonesia. Sci Total Environ 725:138436

Van Dongen S (2008) Graph clustering via a discrete uncoupling process. SIAM J Matrix Anal Appl 30(1):121–141

Vijayan V, Saraph V, Milenković T (2015) Magna++: Maximizing accuracy in global network alignment via both node and edge conservation. Bioinformatics 31(14):2409–2411

Wang Z, Li Z, Yuan G, Sun Y, Rui X, Xiang X (2018) Tracking the evolution of overlapping communities in dynamic social networks. Knowl-Based Syst 157:81–97

Wang P, Jun-an L, Jin Y, Zhu M, Wang L, Chen S (2020b) Statistical and network analysis of 1212 COVID-19 patients in Henan, China. Int J Infect Dis. https://doi.org/10.1016/j.ijid.2020.04.051

Wang M, Jiang A, Gong L, Luo L, Guo W, Li C, Zheng J, Li C, Yang B, Zeng J, et al. (2020a) Temperature significant change COVID-19 transmission in 429 cities. medRxiv. https://doi.org/10.1101/2020.02.22.20025791

Wu Z, McGoogan JM (2020) Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese center for disease control and prevention. JAMA, 323(13):1239–1242, 04. ISSN 0098-7484. https://doi.org/10.1001/jama.2020.2648. https://doi.org/10.1001/jama.2020.2648

Xie J, Zhu Y (2020) Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci Total Environ 724:138201

Yang Z, Algesheimer R, Tessone CJ (2016) A comparative analysis of community detection algorithms on artificial networks. Sci Rep 6:30750

Zhu Y, Xie J (2020) Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci Total Environ page 138201

Zoran MA, Savastru RS, Savastru DM, Tautan MN (2020) Assessing the relationship between ground levels of ozone (O_3) and nitrogen dioxide (NO_2) with coronavirus (COVID-19) in Milan, Italy. Sci Total Environ, 740:140005. ISSN 0048-9697. https://doi.org/10.1016/j.scitotenv.2020.140005. https://www.sciencedirect.com/science/article/pii/S0048969720335257

Zucco C, Liang H, Fatta GD, Cannataro M (2018) Explainable sentiment analysis with applications in medicine. In: Zheng HJ, Callejas Z, Griol D, Wang H, Hu X, Schmidt HHHW, Baumbach J, Dickerson J, Zhang L (Eds) IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3–6, 2018, pages 1740–1747. IEEE Computer Society, https://doi.org/10.1109/BIBM.2018.8621359. http://doi.ieeecomputersociety.org/10.1109/BIBM.2018.8621359

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.