Correlation between Covid-19 and Mobility

Giovanni Fiordeponti, Antonio Ionta, Silvia Marchiori

***Abstract*—The paper deals with presenting a data visualization regarding the effects od the CoVid-19 pandemic in Italy, during the year 2020. This work focuses the attention towards an hypothetical governmental decision maker, who needs to evaluate the effectiveness of the submitted anti-contagion measures with the final objective of planning out the future actions. For accomplishing this task, through a visual environment a comparison in certain temporal interval between the variation of people's movement and some indicators of the pandemic situation is performed. Addressing the correlation between the pandemic aggressiveness trend and the variation of mobility of the population is the key point. The situation is framed with regional granularity, according with the recent Italian government decision of diversifying restrictive rules region by region. The work should try to give an answer to simple questions, that, at the same time, hides great complexity, like: has the lockdown be effective to reduce the contagion? In the case of an escalation in the severity of the pandemic, is there a direct connection with an increment in people's movement?**

I. Introduction

Public health officials often need to analyze simulations of epidemic models to improve preparedness, planning responses, and mitigate the impact of pandemics like COVID-19. During the planning and preparatory stages, they need to explore different scenarios and decision measures, and study the impact of these decision measures on controlling the pandemic's spread. In their analysis tasks, they often need to explore outputs in a spatiotemporal environment where they can analyze the spread patterns and the evolution of spread across space and time, interactively explore and filter, modify model parameters, and explore different scenarios. [AGSEHH] From these evidences the idea of building a visual environment able to fulfill a small part of the above mentioned requirements. The guidelines in the development of the project were the following:

- pandemic severity and mobility variation data are considered, which of restrictive rules application. The visual environment has to present in a coordinated way this source of information;
- the use of analytics, in particular of dimensionality reduction, is the chosen instrument for reveling possible correlation between the presented data, thus highlighting a sort of cause—effect relationship, if it actually exist;
- a form of visualization with space and temporal flexibility is needed;
- *the intended recipient*, i.e. a hypothetical decision-maker government figure, constitutes an audience with intermediate level of knowledge in the field of study: he is informed about the basic mechanisms which rule the pandemic, as his role requires, but, at the same time, he is not a proper expert of medical aspects. Therefore, the visualization can offer something not elementary, but should preserve the sufficient simplicity for not affecting the immediacy of the results interpretation
- the representation must be a dynamic one. The user should be guided in the system's browsing, having the perception that each interaction step is a part of a longer path toward the discovery of a correlation

II. Data

The data represented in the visual environment are collected from the following two public datasets:

- COVID-19 Italy data. To inform citizens and make the collected data available, useful for communication and information purposes only, the Italian Department of Civil makes available, under license CC-BY-4.0, information about the situation of COVID-19 pandemic in Italy;
- Google COVID-19 Community Mobility Reports These Community Mobility Reports aim to provide insights into what has changed in response to policies aimed at combating COVID-19. The reports chart movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.

A subset of the data enclosed in these sources were selected and, through an activity of pre-processing, were aggregated into a single CSV file feeding the visualization. Disregarding the elements referring to spatiotemporal collocation, the attributes reflecting the pandemic's severity, day by day, taken into account are: positives, death, healed, hospitalized, isolated, intensive care. The Community Mobility Reports show movement trends by region, across different categories of places. For each category in a region, reports show the changes comparing mobility for the report date to the baseline day, reported as a positive or negative percentage. A baseline day represents a normal value for that day of the week. The baseline day is the median value from the 5-week period Jan 3 – Feb 6, 2020. Places with similar characteristics for purposes of social distancing guidance are grouped: “Groceries-Pharmacy”, “Parks”, “Residential”, “Retail-Recreation”, “Transit”, “Workplaces”. All the data are referred to Italy as geographical index, and to the period Feb 24 – Dec 31, 2020 as time window.

III. The Visual Environment

The main result of this project is the production of the visual environment that is going to be described here, with a focus on each single main part. The ecosystem is composed by four interactive visualizations:

- a geographical map of Italy showing COVID-19 data about pandemic situation;
- a time series showing COVID-19 data about pandemic situation;
- a boxplot showing mobility data
- a scatter plot showing the output of the dimensionality reduction performed on the dataset.

The first two visualizations are, among the other things, of service for filtering the data from the geographical (regional granularity level) and temporal (daily granularity level) point of view.

A. Geographical Map

a) Visualization: The visualization acts as choropleth map for illustrating the spatial spread of the selected COVID-19 parameter (e.g. “New Cases”).

b) Interaction: The user can click on each region on the map for filtering the data focusing on specifical geographical area. The feedback the user receives for the selection is the region's borders becoming colored (with the specific color chosen for identifying that region data in the rest of the visual environment).

B. Time Series

a) Visualization: It shows the trend in time of the pandemic, plotting the data of a particular COVID-19 related parameter, chosen by a select drop-down list. On the x axis the time, scanned by the months indication; on the y axis the units of the corresponding parameter, as raw number of people.

b) Interaction: Through brushing the user can temporally filter the data, with a modification which is coordinated in the entire visual environment. The minimal granularity of the filter is one day.

C. Boxplot

a) Visualization: It represents the mobility data belonging to a certain category with indications of their statistical distribution. The category is chosen by a select drop-down list. The vertical axis encodes the percentage reflecting the variation of the movement with respect to a “normal” day.

b) Interaction: Through brushing the user can filter the data focusing on a particular range of variation in mobility. The modification is coordinated with the entire visual environment.

D. Scatter Plot

a) Visualization: It plots the result of dimensionality reduction and clustering executed on demand on the subset of data currently selected.

b) Interaction: The user can trigger the computation of multidimensional projection specifying also the number of clusters that the K-Means algorithm is going to use as main parameter. The scatter plot supports brushing for zooming on a specific portion of the graphic, and double click for resetting to default zoom.

IV. Analytics

The reasoning part of the project is based on the execution of dimensionality reduction paired with a consecutive step of clustering on the currently selected portion of the dataset, using, respectively, the algorithms t-SNE and K-Means. This process is performed on demand, exploiting a back-end service deployed on an Application server and continuously listening for requests. The back-end endorses the RESTful paradigm and it is developed in Python, with the support of Flask framework, and the use of sklearn library for executing the dimensionality reduction and clustering algorithms. t-distributed Stochastic Neighbor Embedding was chosen as algorithm for performing dimensionality reduction as it results, after some experimental attempts with other multidimensional projection techniques, the best approach for separating data into clusters, which is the ultimate objective of the analytics. In fact, the empirical evidence, is confirmed by the theory behind t-SNE, which is funded on plotting data according to the optimization of a cost function, focusing on similarities associated to distance, an approach which ends in amplifying separation between samples. After the dimensionality reduction step, the clustering one is actuated, with the simple aim of assigning a label to each single data point. K-means is adopted as algorithm, relying on the fact that is probably the most well-known and fast clustering algorithm, easy to comprehend and to implement. The principal disadvantage of K-means is that it requires to select in advance how many groups/classes there are, which is not a trivial task. This criticality is mitigated in the system thanks to the possibility of choosing on demand the number of groups on which to execute the clustering, thus having the possibility of tuning the process with an immediate visual feedback. Ideally, the user can choose the number of clusters which reflects visually the best surfaces of separation among data case by case.

V. Insights

Once the project has been described in its main components, this section gives some intuitions about the capacity of the system to answer to the questions it is intended for. Is it possible to perceive a correlation between pandemic's aggressiveness and mobility given common spatiotemporal coordinates? A semantical trip for reaching the objective is proposed as example. Fixed the temporal interval, the entry point is the geographical map: taking advantage of choropleth, you can identify an area characterized by a not strong COVID-19 manifestation, as confirmed by the time series report. You, then, restrict the dataset and can verify, with the help of the dedicated boxplot, if the samples in consideration are effectively related to a great decrease of mobility. This instinctive analysis can be corroborated or, on the contrary, weakened by the outcome of the dimensionality reduction plus clustering process: if selected data points are mainly grouped in the same cluster, the correlation is confirmed. The intended user, a decision maker with intermediate knowledge of the field of study, is able to associate to this kind of assessment the governmental restrictive procedures, with the plan of establishing if they were effective for slowing down the pandemic in a certain zone in a certain period. He may have different targets with respect to the elementary one faced in the above example; in any case, he is guided in the tour by the visualizations offered by the system, which support in picking spatiotemporal windows characterized by specific markers (of mobility or of pandemic) and then certifying the correlation using the analytics tuned through successive computations with more refined parameters, triggered by the user himself.

VI. Related Works

How can be this work located with respect to existing literature about similar topics? To answer this question a succinct analysis of related scientific papers is going to be executed. For the unique purpose of disambiguating the references to the present work, intending the overall visual environment plus analytics process, in the following discussion we are going to indicate it with the initials PW. [ZSPBBP] proposes a huge panoramic of COVID-19 crisis visualizations, which offers numerous causes for reflection. First of all, it is important to highlight a fundamental point of discontinuity between the visualizations taken into account in [ZSPBBP] and PW. The firsts refer to something created to communicate information to the public, with the key concept that Information exposure in a public health crisis can impact people's attitudes towards and responses to the crisis and risks, and ultimately the trajectory of a pandemic. Therefore, the intended consumers are common people, with a non-deep knowledge of the analytics tools an no responsibility about political measures. This clearly differs from the intended user of PW, already discussed. According to [ZSPBBP] COVID-19 crisis visualizations communicate a wide range of information about the pandemic. These messages can be classified as reflecting six purposes: informing of the severity; forecasting trends and influences; explaining the nature of the crisis; guiding risk mitigation; communicating risk, vulnerability, and equity; gauging the multifaceted impacts of the crisis. Note that these message categories are not mutually exclusive. It is possible to insert PW in the categories (1) and (4). Let's examine the visualizations in these groups. Temporal visualizations focused on depicting the trajectory of the pandemic over time. The most basic solution was to visualize daily and cumulative numbers over a time period mostly using bar charts, line charts, and area charts. The wide adoption of linear scale charts may be because they are easier to understand for the general public. Geospatial visualizations displayed variables of interest over geographical maps to demonstrate which regions were impacted and compare how the impact differs by regions. Hence, for informing the severity of pandemic PW uses the most common solutions explored in the corpus of [ZSPBBP], which can be defined as the state of art for the matter, thus centering the objective of providing a non-ambiguous, almost familiar, visual environment concerning COVID-19 data. While choropleth maps were the most frequently used geospatial visualization in the corpus of [ZSPBBP], they do present a challenge in that color-coding raw data may mislead viewers. For example, if a highly populated zone has the same case number as a less-populated zone, the color will be the same on the map. A reader may interpret the map as conveying that the intensity of cases is similar in the two zones, while in actuality, the intensity in the less-populated zone is higher. Normalizing data may help remove this type of bias, and, indeed, this advice is followed in PW. Conceptual flattening-the-curve charts have become prevalent, particularly with the addition of a horizontal line marking "healthcare system capacity". They appear to emphasize personal responsibility in minimizing the unprecedented strain on the health system. However, these type of visuals have become controversial amidst the pandemic as they simplify complex pandemic situations (e.g., implying that it is good enough to keep the patient counts below the healthcare system's capacity) and they may also be misinterpreted by the public. Guiding risk mitigation is faced from another point of view PW, trying not to trivialize the situation which is, on the contrary, depicted with the rigor of an analytics process. [ZSPBBP] emphasizes the importance of reporting on the source (and even the source of source) and recency of data. Crises such as the COVID-19 pandemic introduce particular threats to the production of trustworthy visual information. Having a trustworthy data source is crucial for creating visualizations, especially in times of crisis where misinformation and disinformation are rampant. [ZSPBBP] highlights also the importance of reporting the recency of data, that is, the date of data collection and visualization production. And, the discussion of how COVID-19 visualizations have changed over time, highlights how visualizations produced at one point in time may not be sufficient for depicting the state of the crisis at a later time. Indications followed by PW. After an overview on a plethora of COVID-19 related visual environments, concreate examples of works similar to PW for objectives and datasets are now going to be examined.

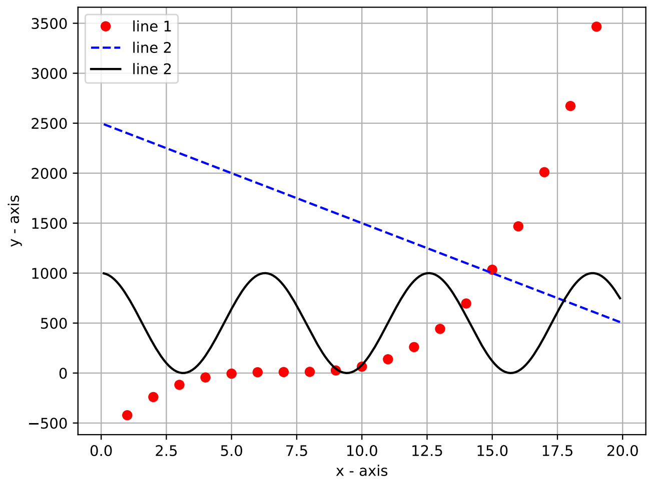

Fig. 1. Simulation results.

References

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [1]. Refer simply to the reference number, as in [1]—do not use "Ref. [1]" or "reference [1]" except at the beginning of a sentence: "Reference [1] was the first . . ."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al." [2]. Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [3]. Papers that have been accepted for publication should be cited as "in press" Capitalize only the first word in a paper title, except for proper nouns and element symbols.

Another examples are: book [4], [5], [5], periodical [6], report [7], conference [8], [9], [10], [11], [12], software document [13], patent [14], article [15], [16], [17], [18], data sheet [19], theses (M.S) [20], dissertation (Ph.D) [21], conference proceedings [22], [23], [2], [24], standards [25], and others [26], [27].

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

References

[1] M. Shell. (2007) IEEEtran homepage. [Online]. Available: http://www.michaelshell.org/tex/ieeetran/

[2] R. M. A. Dawson, Z. Shen, D. A. Furst, S. Connor, J. Hsu, M. G. Kane, R. G. Stewart, A. Ipri, C. N. King, P. J. Green, R. T. Flegal, S. Pearson, W. A. Barrow, E. Dickey, K. Ping, C. W. Tang, S. V. Slyke, F. Chen, J. Shi, J. C. Sturm, and M. H. Lu, "Design of an improved pixel for a polysilicon active-matrix organic LED display," in *SID Tech. Dig.*, 1998, vol. 29, pp. 11–14.

[3] T. J. Ott and N. Aggarwal, "TCP over ATM: ABR or UBR," unpublished.

[4] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[5] J. C. Candy and G. C. Temes, Eds., *Oversampling Delta-Sigma Data Converters Theory, Design and Simulation*. New York: IEEE Press., 1992.

[6] , *Special Issue on Wireless ATM*, vol. 3, Aug. 1996.

[7] R. Jain, K. K. Ramakrishnan, and D. M. Chiu, "Congestion avoidance in computer networks with a connectionless network layer," Digital Equipment Corporation, MA, Tech. Rep. DEC-TR-506, Aug. 1987.

[8] R. K. Gupta and S. D. Senturia, "Pull-in time dynamics as a measure of absolute pressure," in *Proc. IEEE International Workshop on Microelectromechanical Systems (MEMS'97)*, Nagoya, Japan, Jan. 1997, pp. 290–294.

[9] M. S. Yee and L. Hanzo, "Radial basis function decision feedback equaliser assisted burst-by-burst adaptive modulation," in *Proc. IEEE Globecom '99*, Rio de Janeiro, Brazil, Dec. 5–9, 1999, pp. 2183–2187.

[10] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.

[11] B. Mikkelsen, G. Raybon, R.-J. Essiambre, K. Dreyer, Y. Su., L. E. Nelson, J. E. Johnson, G. Shtengel, A. Bond, D. G. Moodie, and A. D. Ellis, "160 Gbit/s single-channel transmission over 300 km nonzero-dispersion fiber with semiconductor based transmitter and demultiplexer," in *Proc. ECOC'99*, 1999, postdeadline paper 2-3, pp. 28–29.

[12] M. Yajnik, S. B. Moon, J. Kurose, and D. Towsley, "Measurement and modeling of the temporal dependence in packet loss," in *Proc. IEEE INFOCOM'99*, vol. 1, New York, NY, Mar. 1999, pp. 345–352.

[13] *SaberDesigner Reference Manual*, Analogy, Inc., Beaverton, OR, 1998, release 4.3.

[14] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sep. 16, 1997.

[15] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," vol. 20, pp. 569–571, Nov. 1999.

[16] Y. Okada, K. Dejima, and T. Ohishi, "Analysis and comparison of PM synchronous motor and induction motor type magnetic bearings," vol. 31, pp. 1047–1053, Sep./Oct. 1995.

[17] F. Delorme *et al.*, "Butt-jointed DBR laser with 15 nm tunability grown in three MOVPE steps," *Electron. Lett.*, vol. 31, no. 15, pp. 1244–1245, 1995.

[18] A. Castaldini, A. Cavallini, B. Fraboni, P. Fernandez, and J. Piqueras, "Midgap traps related to compensation processes in CdTe alloys," *Phys. Rev. B.*, vol. 56, no. 23, pp. 14 897–14 900, 1997.

[19] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.

[20] N. C. Loh, "High-resolution micromachined interferometric accelerometer," Master's thesis, Massachusetts Institute of Technology, Cambridge, 1992.

[21] Q. Li, "Delay characterization and performance control of wide-area networks," Ph.D. dissertation, Univ. of Delaware, Newark, May 2000. [Online]. Available: http://www.ece.udel.edu/ qli

[22] P. Hedelin, P. Knagenhjelm, and M. Skoglund, "Theory for transmission of vector quantization data," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier Science, 1995, ch. 10, pp. 347–396.

[23] W. V. Sorin, "Optical reflectometry for component characterization," in *Fiber Optic Test and Measurement*, D. Derickson, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[24] J. B. Anderson and K. Tepe, "Properties of the tailbiting BCJR decoder," in *Codes, Systems and Graphical Models*, ser. IMA Volumes in Mathematics and Its Applications. New York: Springer-Verlag, 2000.

[25] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.

[26] L. Roberts, "Enhanced proportional rate control algorithm PRCA," ATM Forum Contribution 94-0735R1, Aug. 1994.

[27] K. K. Ramakrishnan and S. Floyd, "A proposal to add explicit congestion notification (ECN) to IP," RFC 2481, Jan. 1999.

Received: 30 May 2020;
Accepted: 14 September 2020;
Published: 22 September 2020;