Using Distributed Risk Maps by Consensus as a Complement to Contact Tracing Apps

Miguel Rebollo\textsuperscript{1,2(✉)}, Rosa M. Benito\textsuperscript{2}, Juan C. Losada\textsuperscript{2}, and Javier Galeano\textsuperscript{2}

\textsuperscript{1} VRAIn - Valencian Research Institute for Artificial Intelligence, Universitat Politècnica de València, Camino de Vera S/N 46022, Valencia, Spain
\text{mrebollo@vrain.upv.es}

\textsuperscript{2} Complex Systems Group, Universidad Politécnica de Madrid, c/ Ramiro de Maeztu, 7, 28040 Madrid, Spain
\{rosa.benito, juancarlos.losada, javier.galeano\}@upm.es

Abstract. The rapid spread of COVID-19 has demonstrated the need for accurate information to contain its diffusion. Technological solutions are a complement that can help citizens to be informed about the risk in their environment. Although measures such as contact traceability have been successful in some countries, their use raises society’s resistance. This paper proposes a variation of the consensus processes in directed networks to create a risk map of a determined area. The process shares information with trusted contacts: people we would notify in the case of being infected. When the process converges, each participant would have obtained the risk map for the selected zone.

A consensus simulation has been introduced in an SEIR model to evaluate how having available a risk map could affect the virus’s propagation. The scenario chosen is La Gomera Island: a region where the Spanish government has tested its contact tracing app (RadarCOVID). The paper also compares both strategies joint and separately: contact tracing to detect potential infections, and risk maps to avoid movements into conflictive areas. Contact tracing apps could work with 40\% of participants instead of 60\%. On the other hand, the elaboration of risk maps could work with just a 20\% of active installations. Nevertheless, the effect is to delay the propagation instead of reducing the contagion. With both strategies actives, we significantly reduce infected peoples with a relatively low participant number.

Keywords: Consensus · Complex network · COVID · Risk map · Collaboration · Contact tracing

1 Introduction

One of the current challenges to stem the spread of COVID-19 is to track people infected with coronaviruses that can spread the disease. Although technological
solutions such as contact traceability have been successful in some countries, they raise resistance in society due to privacy concerns [15]. The European Data Protection Board has published a guideline for the governments to use this kind of technology, guaranteeing privacy, and proper access to the data [13]. The solutions currently into consideration fall into two main groups: (1) personalized tracking of users from its geolocation, and (2) private tracking of contacts. Most governments have recommended using the second type of application, advising against proposals based on individuals’ geolocation.

Simko et al. [12] made a series of surveys over 100 participants to analyze their opinion about contact-tracking applications and privacy. It is a relevant study since the first part finished when some European countries were under different forms of lock-downs, and contact-tracing apps were not available yet. Between the first and the second study, several proposals appeared, such as the ones made by Apple and Google [1, 2], the Massachusetts Institute of Technology (MIT) [10], the University of Washington (UW) [4], PEPP-PT [9], Inria [3], WeTrace [16], and DP3T [14]. The study throws that people are more comfortable using an existing mapping application that adds tracking for COVID-19 instead of using new apps, with reservations even if they provide ‘perfect’ privacy. One of the main concerns is sharing data, preferring that Google or organizations such as the UN develop the application. In general, participants manifest a lack of trust in how their governments would use the citizens’ location data. They thought that it was unlikely that their government would erase the data after the crisis and also that they would use it for other purposes. For both studies, something in common was mixed feelings about using proximity tracking for the contacts and negative towards using any other data source.

The MIT Technology Review [8] has been collecting the different proposals that states have created. Currently, there are 43 registered apps. There are initiatives in the five continents, but most of the countries belong to Asia and Europe since they were the firsts places where COVID-19 appeared. The population that uses the applications varies from 9,000 inhabitants in Cyprus to 100,000,000 in India. The median value is 1,613,500. The Bluetooth technology is the solution that most countries have chosen, with 72% of the apps. Moreover, almost half of them use the API provided by Google and Apple. Despite the recommendation to avoid location services, 36% of the apps still use it.

Despite the efforts to develop technological solutions to track the propagation of COVID-19, the usage of the apps is not extended enough. That is why we propose a third method: a process of dissemination in local environments. The method exchanges information with known and trusted contacts only. The consensus for COVID (C4C) method is a variation of the consensus processes proposed by Olfati–Saber and Murray [7]. It is a dissemination process that allows a distributed calculation of the value of a function in a network, exchanging information only with direct neighbors without having global knowledge of the structure, size, values, or other characteristics of the graph. This process converges to a final single value for the calculated function. With this approach, privacy is maintained, the administration obtains aggregated information, and
citizens and the administration have the same data, promoting transparency. One relevant limitation is that some critical mass is still necessary.

The rest of the paper is structured as follows. Section 2 explains how citizens can collaboratively create risk maps using a consensus process with their close contacts. Section 3 shows the results using La Gomera as an example: one of the Canary Islands, with 21,550 inhabitants. The Spanish government carried out a pilot project with their contact tracing application (RadarCOVID) on that island, so we decided to use the same scenario. Finally, Sect. 4 summarizes the main findings of this work.

2 Collaborative Risk Map Generation

The consensus for COVID (C4C) proposal works over a contact network with non-reciprocal relationships. It is needed to avoid the presence of hubs with an elevate number of contacts. Therefore, the underlying structure is a directed graph. The original consensus algorithm works over non-directed graphs, so we have to extend the model to consider this case.

2.1 Extension of the Consensus Process

Let \( G = (V, E) \) be a non-directed network formed by a set of vertices \( V \) and a set of links \( E \subseteq V \times V \) where \( (i, j) \in E \) if there is a link between the nodes \( i \) and \( j \). We denote by \( N_i \) the set formed by the neighbors of \( i \). A vector \( x = (x_1, \ldots, x_n)^T \) contains the initial values of the variables associated with each node. Olfati–Saber and Murray [7] propose an algorithm whose iterative application converges to the mean value of \( x \).

\[
x_i(t + 1) = x_i(t) + \varepsilon \sum_{j \in N_i} [x_j(t) - x_i(t)]
\]  

(1)

The authors demonstrated that this consensus process converges to the average of the initial values when \( \varepsilon < \frac{1}{\max d_i} \), being \( d_i \) the degree of node \( i \). There is an equivalent matricial formulation.

\[
x(t + 1) = (I - \varepsilon L) x(t), \quad \text{with} \quad L = D_{AG} - A_G
\]  

(2)

where \( I \) denotes the identity matrix and \( L \) is the laplacian of the adjacency matrix of the graph \( G \). This expression \( P \) is called the Perron–Frobenius matrix and governs the consensus process’s collective dynamics.

If each component contains a vector \( x_i = (x_i^1, \ldots, x_i^m) \in \mathbb{R}^m \), the process carries out a consensus over \( m \) independent variables. By expanding the vector with one additional element \( y_i \in \mathbb{R} \), we can determine the size of the network at the same times as follows: \( x_i \oplus y_i = (x_i^1, \ldots, x_i^m, y_i) \). Initially, \( y_i = 0 \ \forall i \). Without losing generality, we can introduce an additional node in the network
Distributed Risk Maps by Consensus

whose initial values are \( x_0 \oplus y_0 = (0, \ldots, 0, 1) \). As the process converges to the average value, \( y_i = 1/|V| \) and, therefore, \(|V| = 1/y_i \) is the size of the network.

We need the Perron \( P \) matrix to be doubly stochastic for the consensus to work: a matrix whose rows and columns add up to one. However, in directed networks, we obtain a row stochastic one. Inspired in the Dominguez-Garcia and Hadjicostis matrix scaling algorithm [5], we define an iterative process to convert the Perron matrix into a double stochastic one. The process begins with a row stochastic matrix \( \overline{P} \), and, in each iteration, the matrix scales following the expression.

\[
P(t) = \overline{P} \Delta(t) + [I - \Delta(t)]
\]

where \( \overline{P} \) is a local Perron matrix and \( \Delta(t) \) is updated as Algorithm 1 describes. The only consideration is that the Perron matrix \( \overline{P} \) is defined using the degree of each node instead of a common \( \varepsilon \) value for all the nodes (see line 4). Furthermore, as \( P(t) \) is based on the Perron matrix, we can combine the matrix’s scaling with the consensus value calculation in the same step (line 12).

**Algorithm 1.** Matrix scaling and consensus (collective)

1: init \( x(0) \)
2: \( L = D^\text{out}_A - A \)
3: \( \Delta(0) = D^\text{out}L^{-1} \)
4: \( \overline{P} = (I - \Delta(0) \ast L) \)
5: \( P(0) = \overline{P} \Delta(0) + [I - \Delta(0)] \)
6: \( \pi(0) = 0, \quad \eta(0) = 1 \)
7: repeat
8: \( \pi(t + 1) = P(0) \pi(t) \)
9: \( \eta(t + 1) = \max(\pi(t), \max \eta(t) A) \)
10: \( \Delta(t + 1) = \Delta(0) \frac{\pi(t+1)}{\eta(t+1)} \)
11: \( P(t + 1) = \overline{P} \Delta(t + 1) + [I - \Delta(t + 1)] \)
12: \( x(t + 1) = P(t + 1) x(t) \)
13: until \( x(t) \) converges

The adaptions of the scaling algorithm are the calculation of \( \Delta(t) \) (lines 3 and 10), the definition of the matrix \( \overline{P} \) as a local variation of the Perron matrix (line 4), and how \( \pi(t) \) updates (line 8).

### 2.2 Map Generation

Once we have defined the algorithm for consensus processes over directed networks, the goal is to create a citizen network in an area (town, province, state, or any other administrative division) that uses it to collaboratively create a risk map. We have chosen the census districts from the National Institute of Statistics (INE) of Spain. The sizes of the districts are relatively homogeneous, having between 900 and 3,000 inhabitants each one. It is easily scalable, aggregating
the information in bigger administrative units. Moreover, they never provide statistics with less than 100 persons to avoid reidentification.

Inhabitants share a risk index (RI) that measures their probability of being infected by COVID-19. The risk in a census district depends on the RI of all the people that live in it. The RI could integrate data from different sources: medical symptoms, symptoms from the close contacts, age, family situation, or habitability conditions. In this work, we use the same measure as the emergency service 112. The risk value depends on medical symptoms: shortness of breath (60 points), fever (15), coughing (15), or close contact (29). Over 30 points, it is considered that the person has been infected. We use the following notation:

- $ri_i$: risk index of node $i$, $i = 1, \ldots, n$
- $x_i \oplus y_i = (x_i^1, \ldots, x_i^n, y_i)$: vector with the risk map values in node $i$.
- $R_i = (ri_1, \ldots, ri_n)$: complete risk map

Let us assume an extra node representing an administrative unit, such as the town hall, acts as the $x_0$ node. Algorithm 2 describes the complete process.

Some important remarks related to the process are:

1. the position of $ri_i$ in $x_i(0)$ corresponds to its census district
2. each node executes a local version of Algorithm 1
3. the first exchange is the only moment in which vectors contain individual values: the risk and the census district of $i$. We assume that there are no privacy concerns since the node would share this information with its $N_i$
4. in the following exchanges, the vectors received $x_j(t)$ contains aggregated information. As the neighbors of $j$ remain unknown for $i$, it is not easy to track back the data.

It is a successive refinement mechanism: there is a map available at any time, and the longer the algorithm executes, the fittest the risk values are (see Fig. 4). The final risk map $R$ is the same one that a centralized process would obtain with all the risk indexes available.

3 Results

The purpose of this section is to validate the algorithm proposed to create risk maps in a scenario similar to the conditions of the real world. The ideal situation would be to launch a pilot project in a controlled environment. However,
we consider that a previous simulation is essential. Therefore, the population model, the mobility patterns, and the SEIR models are defined to provide an environment with the characteristics relevant for the algorithm.

### 3.1 Population and Infection Model

As an application example, we have chosen La Gomera: one of the Canary Islands with 21,550 residents. The National Institute of Statistics divides the island into 14 census districts. The population that lives in each area is publicly available. The network has as many nodes as inhabitants. For each node, we generate the coordinates for their home address (Fig. 1, left). They are random coordinates following the density distribution of the census districts.

The movements of the people along the day are simulated using recurrent Lévy flights [6]. Each person has assigned a path with 96 points (every 5 min for 8 h) that begins and ends at his or her home location (Fig. 1, right). We have validated the model comparing the movements with the data available in the study on mobility based on mobile phone carried out by the Spanish National Statistics Institute (INE) in 2019.² In this study, La Gomera was divided into two areas. The flows represent travels from commuters between both areas. No external sources, such as ferry or plane trips from other islands or the peninsula, would undoubtedly be relevant. This data, if available, could be added to the model. The daily mobility between them was 450 persons leaving San Sebastián de la Gomera area and 550 enter (the inverse from the northern area viewpoint). The simulation with Lévy flights throws an output flow of 464 persons and an input flow of 668. The movements are in the same magnitude order, so we assume that they are coherent.

² https://www.ine.es/en/experimental/movilidad/experimental_em_en.htm.
The population moves alternatively between the home and the working location. Carriers can infect other people in both places. The cycle consists of a sequence
\[ \text{movement} \rightarrow \text{infection} \rightarrow \text{movement} \rightarrow \text{infection}. \]

To simulate the close contacts, we use the same criteria as the contact tracing app: a close contact is defined when two persons are at 5 meters (with 2 meters only obtaining a 78% accuracy) at most and during 15 min. The result is a daily risky contact matrix of dimension 21,550 × 21,550.

Finally, to simulate the spread of the COVID-19, we use an SEIR model. Its parameters follow the findings from the literature that has analyzed the COVID-19 propagation [17]. Particularly, the incubation time is 7 days, so $\beta = \frac{1}{7}$, the probability of infection $\sigma = 0.1$ and the recovery time is 15 days, so $\gamma = \frac{1}{15}$.

The purpose of the model is not to predict precisely the behavior of the disease. The model provides the consensus process with different scenarios to check the accurateness of the risk maps.

People start at their home location. They move along the day, interacting with the other persons. Nodes update their state according to the epidemic model and the contact matrix, and they go back to their home locations. A new infection stage is performed at home since, in COVID-19, some researches demonstrate the family to be a strong transmission source. Once completed the update, a new cycle begins (Fig. 2).

### 3.2 Risk Map Creation

The consensus process described in Algorithm 2 obtains the actual risk map if all the inhabitants participate in the process. However, as we have seen with contact tracing apps, this is a utopic scenario. Therefore, we assume that just 3,000 persons participate in creating the risk map (same volume as in the RadarCOVID contact tracing app.

To create the contact network, we have analyzed the network formed by the followers of the Twitter account of the town halls of cities of similar sizes. The degree of those networks follows a power-law distribution of parameter $\alpha \approx -1.7$. To model the contact network of the entire population, we have generated a preferential attachment network following the same distribution (Fig. 3, left). The resultant network had 58,000 contacts and a maximum degree of 876. As a potential application would bound the number of closer contacts, we choose a
Fig. 3. Cumulated degree distribution of the networks (Left) Complete contact network with a power-law with $\alpha = -1.7$ (Right) Random selection of contacts. It follows a Weibull distribution with parameters $\alpha = 13.4$, $\gamma = 0.48$.

subset of the potential links. A reasonable limit is 15 contacts, five of each type (family, colleagues, and friends). If each person choose randomly $9 \pm 2$ contacts, the resulting network has 26,500 contacts and the number of connections vary from 4 to 16 (Fig. 3, right).

Therefore, we have obtained a network with 3,000 nodes and mean degree 10, varying from 4 to 16, generated from social network profiles. Over this scenario, the inhabitants can determine their town’s risk map at the end of the day. We assume that no additional measures, such as social distancing or limitations of movements, are taken.

As an example, let us consider the situation after 30 days. People have moved during this period as described in Sect. 3.1, and the contagion has evolved following the SEIR model. After 30 days, the situation of the COVID-19 in La Gomera appears in Fig. 4, with a clear breakout in San Sebastian de La Gomera (in red). The 3,000 persons determine their risk index (some of them are already infected) and share the value with their direct contacts, following the consensus process from C4C.

Each node has a vector of 14 components, one for each census district $x_i \oplus y_i = (x^1_i, \ldots, x^{14}_i, 0)$. Let be $ri_i$ the risk index of $i$ and $cd_i = k$ the census district $i$ lives in. $x^k_i = ri_i$ and the rest $x^l_i = 0$, $l \neq k$. Each node executes the process detailed in Algorithm 2. The evolution appears in Fig. 4. The real risk values are and the values calculated by consensus in one of the 3,000 participants for each census district are

| Area | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Real risk | 9.9 | 10.1| 9.9 | 9.9 | 12.2| 11.3| 50.3| 55.6| 53.1| 15.4| 21.8| 9.9 | 10.9| 10.1|
| $R_i$ | 10.0| 8.1 | 12.3| 8.6 | 13.2| 9.5 | 49.2| 48.3| 40.1| 12.5| 22.0| 9.6 | 11.4| 8.8 |
Let us assume that people outside the risky areas do not move into them, and people who live in high-risk areas do not go out, depending on the risk map readings. The effect of having a risk map available and avoid areas with high risk does not reduce the total number of infections significantly. It reduces the peak but keeps the propagation active more days (see Fig. 5).

**Fig. 4.** Evolution of the consensus in the creation of a risk map. (Left) Convergence of the process. (Right) Evolution of the map calculated for one random node. (Bottom) Map created by consensus versus real risk map

**Fig. 5.** Evolution of infected with and without considering the risk map. (Left) Evolution (Right) Cumulated
3.3 Evolution with Contact Tracing App Active

A problem with tracking applications is that they need a large percentage of the population using the app. Some works suggest that tracking applications need at least 60% of penetration to be effective [11]. We have simulated the propagation of COVID-16 in three scenarios: (1) no measures taken, (2) all infected detected and isolated, and (3) people with contact tracing app isolated in 48 h from the first symptom. To evaluate the impact of the penetration of the app, we have considered a 20%, 40%, 60%, and 80% of the total population using the app (Fig. 6).

Fig. 6. Evolution of infected in three scenarios: no isolation (blue), total isolation (red) and isolation for traced users (yellow), from 20% to 80% of users. (top) total infected by day (bottom) cumulated infections

The effect of contact tracing apps in the propagation is almost irrelevant, with 20% of users. There are just little differences between 40% and 60%, so probably it is not necessary to arrive at this value. With 80%, the transmission is almost controlled.

3.4 Contact Tracing and Risk Maps Combined

Finally, we have tested the combination of risk maps and contact tracing. The behavior of the people would be

- if you live in an area with medium or high risk, you do not go out of it
- if you live in a low-risk area, do not go to risky ones
- tracing app notifies exposure in 48 h. If the person receives an alert, then is isolated

Five scenarios are analyzed: no isolation, total isolation, limited movement by risk map, isolation by contact tracing, and risk map plus contact tracing. Considering the case with 3,000 users (15% of the population of La Gomera), we see that contact tracing or risk maps have a low effect on their own in controlling the propagation (Fig. 7). However, with both strategies combined, we obtain a significant reduction in total infected, reducing a 50%.
4 Conclusions

Technology can be an essential ally to control the transmission of COVID-19. Nevertheless, concerns about privacy and the possible use of the data after the pandemic have made it difficult to implement technological solutions.

This work proposes an alternative for users to create risk maps collaboratively. This approach executes a consensus process that uses local information and data from the direct neighbors to calculate the value of a shared function. In our case, the values are the risk index of the different districts that form the town. Close contacts (family, colleagues, and friends) define the network relationships, whom we warned about being infected. The data exchanged is an aggregation, and it is not possible to reidentify the personal information. At the end of the process, all the participants obtain the same copy of the complete risk map. However, constraining the movements using the information on risk maps reduces the peak and smooth the evolution of the infection.

The use of contact tracing apps needs a considerable proportion of active users to work. Nevertheless, the combination of the information from risk maps to avoid areas with a high index of infections and alerts of exposure obtain good results even with relatively low penetration of the apps.

Acknowledgements. This research was supported by TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No 952215, and by the Spanish Ministry of Science, Innovation and Universities (MICIU) under Contract No. PGC2018-093854-B-I00b and RTI2018-095390-B-C32.

References

1. Apple, Google: Exposure notifications: using technology to help public health authorities fight covid’19, 01 June 2020. https://www.google.com/covid19/exposurenotifications/
2. Apple, Google: Privacy-preserving contact tracing, 01 June 2020. https://www.apple.com/covid19/contacttracing
3. Castelluccia, C., Bielova, N., Boutet, A., Cunche, M., Lauradoux, C., Le Métayer, D., Roca, V.: ROBERT: ROBust and privacy-presERving proximity Tracing, May 2020. https://hal.inria.fr/hal-02611265/file/ROBERT-v1.1.pdf

4. Chan, J., Foster, D., Gollakota, S., Horvitz, E., Jaeger, J., Kakade, S., Kohno, T., Langford, J., Larson, J., Sharma, P., Singamanalluri, S., Sunshine, J., Tessaro, S.: Pact: privacy sensitive protocols and mechanisms for mobile contact tracing (2020). arXiv:2004.03544 [cs.CR]

5. Dominguez-Garcia, A.D., Hadjicostis, C.N.: Distributed matrix scaling and application to average consensus in directed graphs. IEEE Trans. Autom. Control 58(3), 667–681 (2013)

6. González, M., Hidalgo, C., Barabási, A.: Understanding individual human mobility patterns. Nature 453, 779–782 (2008)

7. Olfati-Saber, R., Fax, J.A., Murray, R.M.: Consensus and cooperation in networked multi-agent systems. Proc. IEEE 95(1), 215–233 (2007)

8. O’Neill, P.H., Ryan-Mosley, T., Johnson, B.: A flood of coronavirus apps are tracking us. Now it’s time to keep track of them, 07 May 2020. https://www.technologyreview.com/2020/05/07/1000961/launching-mitr-covid-tracing-tracker/

9. Pan-European privacy-preserving proximity tracing, 01 June 2020. https://www.pepp-pt.org/

10. Rivest, R.L., Weitzner, D.J., Ivers, L.C., Soibelman, I., Zissman, M.A.: Pact: private automated contact tracing, 01 June 2020. pact.mit.edu

11. Hincha, R., Probert, W., Nurtay, A., Kendall, M., Wyman, C., Hall, M., Lythgoe, K., Cruz, A.B., Zhao, L., Stewart, A., Ferretti, L., Parker, M., Meroehu, A., Mathias, B., Stevenson, S., Montero, D., Warren, J., Mather, N.K., Finkelstein, A., Abeler-Dörner, L., Bonsall, D., Fraser, C.: Effective Configurations of a Digital Contact Tracing App: A report to NHSX. University of Oxford, 16 April 2020

12. Simko, L., Calo, R., Roesner, F., Kohno, T.: Covid-19 contact tracing and privacy: studying opinion and preferences (2020). arXiv:2005.06056 [cs.CR]

13. The European Data Protection Board: Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, 01 June 2020. https://edpb.europa.eu/our-work-tools/our-documents/usmernenia/guidelines-042020-use-location-data-and-contact-tracing_en

14. Troncoso, C., Payer, M., Hubaux, J.P., Salath, M., Larus, J., Bugnion, E., Luks, W., Stadler, T., Pyrgelis, A., Antonioli, D., Barman, L., Chatel, S., Paterson, K., apkun, S., Basin, D., Beutel, J., Jackson, D., Preneel, B., Smart, N., Singelee, D., Abidin, A., Guerres, S., Veale, M., Cremers, C., Binns, R., Cattuto, C.: Decentralized privacy-preserving proximity tracing, April 2020. https://github.com/DP-3T/documents/blob/master/DP3T

15. Vinuesa, R., Theodorou, A., Battaglini, M., Dignum, V.: A socio-technical framework for digital contact tracing. Results Eng. 8, 100163 (2020)

16. WeTrace: Privacy-preserving bluetooth covid-19 contract tracing application, 01 June 2020. https://wetrace.ch/

17. Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., He, J.: Modified SEIR and AI prediction of the epidemics trend of covid-19 in china under public health interventions. J. Thorac. Dis. 12(3) (2020)