Distributed Consensus based COVID-19 Hotspot Density Estimation

Monalisa Achalla
Department of ECE, CITeR,
Clarkson University, NY

Gowtham Muniraju
School of ECEE, SenSIP Center,
Arizona State University

Mahesh K. Banavar
Department of ECE, CITeR,
Clarkson University, NY

Cihan Tepedelenlioglu
School of ECEE, SenSIP Center,
Arizona State University

Andreas Spanias
School of ECEE, SenSIP Center,
Arizona State University

Stephanie Schuckers
Department of ECE, CITeR
Clarkson University, NY

***Abstract*—The primary focus of this work is an application of consensus and distributed algorithms to detect COVID-19 transmission hotspots and to assess the risks for infection. More specifically, we design consensus-based distributed strategies to estimate the size and density of COVID-19 hotspots. We assume every person has a mobile device and rely on data collected from the user devices, such as Bluetooth and WiFi, to detect transmission hotspots. To estimate the number of people in a specific outdoor geographic location and their proximity to each other, we first perform consensus-based distributed clustering to group people into sub-clusters and then estimate the number of users in a cluster. Our algorithm has been configured to work for indoor settings where we consider the signal attenuation due to walls and other obstructions, which are detected by using the Canny edge detection and Hough transforms on the floor maps of the indoor space. Our results on indoor and outdoor hotspot simulations consistently show an accurate estimate of the number of persons in a region.**

***Index Terms*—Consensus, distributed estimation, density, transmission hotspots, applications of ad-hoc networks.**

I. Introduction

The alarming rise of COVID-19 pandemic starting early 2020 brought the whole world to a halt. The pandemic affected world economies and tremendously impacted the livelihood of people. The development and distribution of vaccinations helped in combating the virus to an extent, but there is always a risk from new variants and future pandemics. Hence, there is a need to assess whether certain locations and events pose additional risks for disease spread. In this work, we are developing a suite of solutions that help minimize the spread of COVID-19 including mobile and network applications for real-time assessments and predictions of risks [1] [2]. We focus on the estimation of potential transmission hotspots, using network size estimation and counting of mobile network nodes. This approach together with location mapping techniques can provide real-time risk information to mobile phone users about the areas that they plan to visit.

In this paper, we propose to use consensus-based strategies for estimating network size [3]–[6], node locations [7] and node counts [8] in a network based on minimal transmit-receive data. We assume every person has a mobile device and rely on data collected from user devices, such as Bluetooth [9] and WiFi, to detect transmission hotspots. In order to estimate the number of people in a specific outdoor geographic location and their proximity to each other, we first perform consensus-based distributed clustering to group people into sub clusters and then estimate the number of users in a cluster. Our algorithm has also been configured to work for indoor settings. We consider the signal attenuation due to walls and other obstructions and detect them using Canny edge detection and Hough transforms. Since consensus methods allow us to use a decentralized architecture and distributed processing, our approach will show improvements in terms of accuracy, compactness and power consumption.

We will also present our graphical user interface (GUI) that performs network size estimation and node counting over a specific area in real-time, only using local communications. The GUI has controls to perform clustering using the density based spatial clustering of applications with noise (DBSCAN) algorithm and then selects a cluster to run distributed algorithms to obtain network size and node count. Our results on indoor and outdoor hotspot simulations consistently show an accurate estimate of the number of people in a region and their proximity.

II. System model and Problem statement

We consider a region of interest with users to be a network of devices/users (nodes), which can be grouped into sub-networks based on the density of nodes. Given a network of N nodes, we model the communication among the nodes as a connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \cdots, N\}$ is the set of nodes and $\mathcal{E}$ is the set of edges connecting the nodes. In case of indoor setting, we consider the signal attenuation due to walls and other obstructions and then decide the existence of an edge between the nodes. The set of neighbors of node i is denoted by $\mathbb{N}_i = \{j | \{i, j\} \in \mathcal{E}\}$. The degree of the i^{th} node, denoted by $d_i = |\mathbb{N}_i|$, is the number of neighbors of the i^{th} node. The degree matrix $\mathbf{D}$ is a diagonal matrix that contains the degrees of the nodes.

The authors from Arizona State University and Clarkson University are funded by NSF Awards 2032114 and 2032106, respectively.

The connectivity structure of the graph is characterized by the adjacency matrix $\mathbf{A}$ defined by, $a_{ij} = 1$ if $\{i, j\} \in \mathcal{E}$ and $a_{ij} = 0$, otherwise. We consider a potential hotspot as wireless sensor network of devices (nodes), wherein, a node i can communicate with other nodes that are within the communication radius, which depends on the transmission power of the node. Each node maintains a real valued state and can generate a random vector, which is used to propagate cluster labels. At each iteration, nodes broadcast their state values to their neighbors in a synchronized fashion. Every node is capable of locally estimating their own location [7], [10] and degree by exchanging information with their neighboring nodes. We develop algorithms to form sub-networks and reach consensus on the total number of active node and edges of a network, using only local communications.

III. Background

As detailed in section II, in a given geographical location, we create networks only based on local communications between the devices in a distributed fashion. We rely on the signal strength between the nodes for proximity calculation. However, in indoor environments, the proximity calculation often becomes tricky as the signal between nodes is attenuated due to the presence of walls and obstructions between the nodes. Hence, wall detection becomes a crucial step in network creation and estimation for indoor spaces and various methods have been discussed in literature for wall detection. Estimation of presence line of sight between nodes [11] using RSSI signal quality and using floor maps are among the commonly used wall detection methods. In this paper, we take floor maps of each environment and use image processing techniques like Canny edge detection and Hough Transform for wall detection as described in section III-A.

Once environment-appropriate networks have been created, we create clusters of users in close proximity. It has been observed that clustering users and using these clusters to estimate local hotspot density, where the spread of disease would be higher than in the network as a whole, is advantageous. Among the existing distributed clustering algorithms such as k-means, spectral clustering [12], and expectation maximization algorithms, DBSCAN (density-based spatial clustering of applications with noise) is optimal for distributed implementation as it is robust to communication noise and quantization, and requires lower computational complexity to implement.

DBSCAN is then followed by density estimation in each cluster. This is performed using distributed note counting. DBSCAN and the other distributed algorithms we use do not store any information about individual nodes and run only using local communications between nodes. This ensures that the privacy of all users/nodes is preserved.

A. Wall detection algorithm

To detect walls or obstacles in an indoor setting, we take floor maps of indoor spaces and transform the floor map into Hough space [13]. De-Houghing the transformed images provides us with information about the locations of walls. The Hough transform is implemented by quantizing the Hough parameter space into finite intervals or accumulator cells [14]. As the algorithm runs, each edge pixel (p_i, q_i) is transformed into a discretized (r, θ) curve and the accumulator cells which lie along this curve are incremented. Resulting peaks in the accumulator array represent strong evidence that a corresponding straight line exists in the image. Mapping back from Hough transform space into Cartesian space yields a set of line descriptions (walls) of the image subject. It has been reported in literature [15] that, the presence of a solid obstacle between nodes increases the attenuation of the signal and reduces the proximity by about twice the actual proximity. Once the walls are detected and networks are created accordingly, we cluster the networks to create sub-networks using Distributed DBSCAN algorithm [16]. It is a well known clustering algorithm to group the network of nodes into sub-networks in a distributed way. Once sub-networks are formed, the number of nodes are estimated using methods detailed in section III-B.

B. Distributed Average Consensus for node counting

In an average consensus problem, the nodes reach consensus on the global average of all sensed data based on only local communications [17]. It is assumed that nodes can communicate only with their neighbors and there is no communication noise between nodes. Consider $\mathbf{x}(0)$ to be the initial state values of the nodes in the network. Distributed node counting [3], [8] algorithm is formulated by relating the network size N of a graph with average consensus as,

$$N \approx \frac{\frac{1}{N}\sum_{i=1}^{N} x_i^2(0)}{\mathbb{E}\big[(\frac{1}{N}\sum_{i=1}^{N} n_i x_i(0))^2\big]}, \tag{1}$$

where, $n_i \sim \mathcal{N}(0, 1)$ is an i.i.d Gaussian random variable. The node counting algorithm involves 3 steps: First, estimate the value of the numerator in equation 1, using the average consensus on the node state values. Assuming consensus is reached in t^* iterations,

$$z_i(t^*) \approx \lim_{t\to\infty} z_i(t) = \frac{1}{N}\sum_{i=1}^{N} x_i^2(0) = \frac{1}{N}||\mathbf{x}(0)||_2^2. \tag{2}$$

Next, in order to compute the denominator in equation 1, each node i generates a $K \times 1$ initial state vector, $\mathbf{y}_i(0) = [y_i^{(1)}(0), \cdots, y_i^{(K)}(0)]$, where the k^{th} element $y_i^{(k)}(0) = n_i^{(k)} x_i(0)$ and $n_i^{(k)}$ is a random variable drawn from any continuous distribution with mean 0 and variance 1. The average consensus is performed in parallel on each k^{th} element of $\mathbf{y}_i(0)$ using distributed average consensus. Assuming average consensus is reached at after t^* iterations, $\mathbf{y}_i(t^*) \approx \lim_{t\to\infty} \mathbf{y}_i(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i(0)$. Next, a post processing function $g(\cdot)$ is applied at each node by squaring each element in the $K \times 1$ state vector $\mathbf{y}_i(t^*)$ and computing the average. For node i, the post processing is performed as, $g(\mathbf{y}_i(t^*)) = \frac{1}{K}\sum_{k=1}^{K} \big(y_i^{(k)}(t^*)\big)^2$. The post processed result $g(\mathbf{y}_i(t^*))$ can be related to the l_2 norm of the initial measurements $\mathbf{x}(0)$ as,

$$g(\mathbf{y}_i(t^*)) \approx \frac{1}{N^2}||\mathbf{x}(0)||_2^2. \tag{3}$$

The estimate of network size $\hat{N}$ is computed by taking the ratio of node state values computed in previous steps, i.e, equations (2) and (3), as follows,

$$\hat{N}_i(t^*) = \frac{z_i(t^*)}{g(\mathbf{y}_i(t^*))} \tag{4}$$

Note that, computation of $z_i(t^*)$ and $g(\mathbf{y}_i(t^*))$ can be performed simultaneously, reducing the computation time.

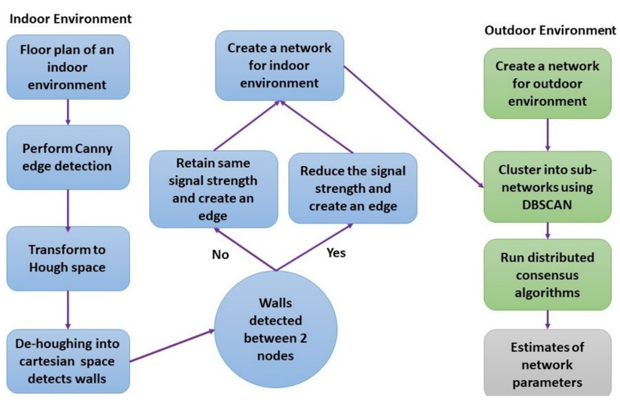

Fig. 1. Flowchart of the proposed approach to estimate network size of a potential hotspot, as an application of consensus and distributed algorithms.

IV. Distributed network size estimation

In this section, we discuss the approach to estimate the number of nodes and edges in the COVID-19 hotspots. The flowchart of our proposed is shown in Figure 1. The global characteristics of a network can be inferred using the consensus based distributed algorithms on the larger network. However, in order to find the local characteristics, clustering algorithms are run on larger network to form the sub-networks, over which distributed algorithms are run to obtain local characteristics. We use a wall detection algorithm, as discussed in Section III-A, to detect the walls in the indoor setting, and then establish an edge between two nodes which contains a wall between only if the signal strength between them is above a threshold τ. Alternatively, in case of outdoor setting, an edge is formed between two nodes if they are within the communication radius ϵ within each other. DBSCAN algorithm forms clusters by sharing the label to the neighboring nodes (label propagation), if there are at least η number of nodes withing the communication radius of ϵ.

Finally, after the sub-networks are formed, we run the distributed node counting (Section III-B) and distributed edge counting algorithms, to reach consensus on the number of nodes and edges in each sub-network. Our solutions thus allows a user to accurately estimate the number of devices and the size of the network within a certain radius.

V. Simulation Results

In this section, we discuss the Graphical user interface (GUIs) that have been developed for network size estimation in both indoor and outdoor settings. The GUIs are designed to give the estimates of number of devices (nodes) and number of edges (connections between the nodes) in any given network. Our algorithm can be actively tuned to provide the risk assessment based on the most recent public health guidelines.

To demonstarte our results for an indoor setting we have used the floor map of Science Center, Clarkson University, shown in figure 2(a), and detect the presence of walls. The first step in wall detection is performing Canny edge detection to detect the edges in an image. Once the edges are detected through a canny edge detector,the next step is to apply hough transform for exactly detecting the wall locations. The image in Hough space is De-houghed and the walls, obstacles in the floor map are reconstructed and detected as shown in figure 2(c). Once walls are detected, we create network in the given

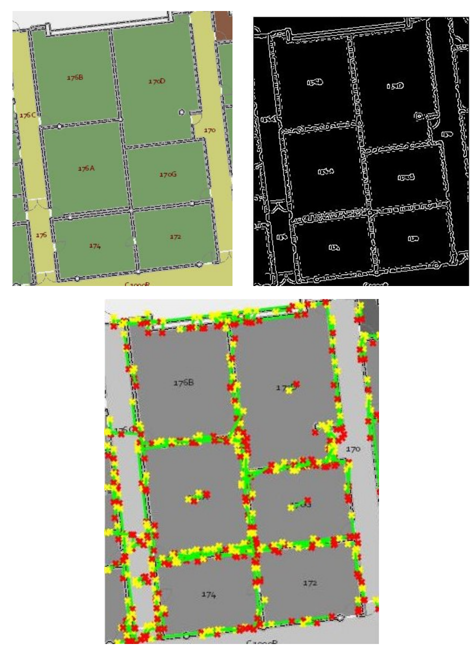

Fig. 2. a) Floor map of indoor environment for network estimation, b) Floor map after passing through the Canny Edge Detector. b) De-houghing the Hough Space to reconstruct walls of indoor environment.

indoor space using only local communication between nodes. We create a network with 160 nodes and run the distributed DBSCAN algorithm, with communication range ϵ as six feet. The DBSCAN clustered the nodes into 4 major clusters with nearly 6 nodes as outliers. We developed a GUI in which a user can create networks in an environment of their choice, fine-tune the parameters like network range, number of nodes. The GUI shows the clusters formed on the right side and also gives details about each cluster, estimated number of nodes as shown in figure 3. The distributed average consensus algorithms as discussed in section III-B are used on each cluster to estimate the number of nodes each cluster. The results of estimates in each cluster are shown in figure 4.

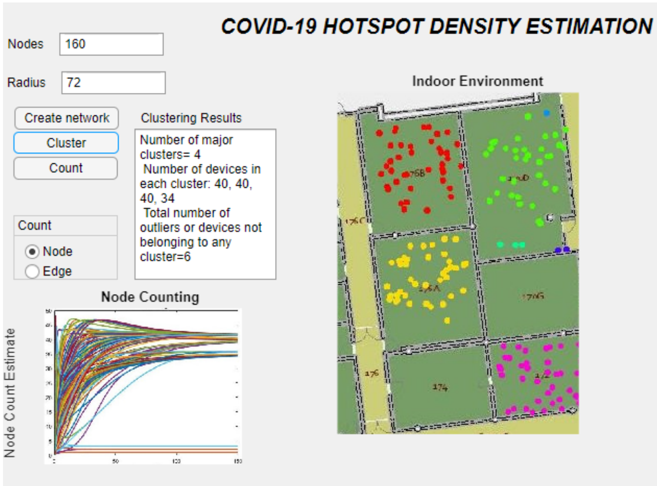

Fig. 3. GUI developed for hotspot detection application in an indoor setting. Wall detection algorithm is run on the floor map to detect the walls and the communication network was formed with 160 nodes and communication range of six feet. On clustering the network using DBSCAN algorithm, the algorithm primarily identified four major clusters. Six nodes were identified as outliers.

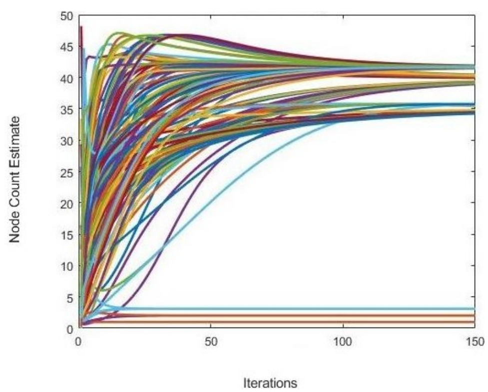

Fig. 4. The consensus algorithm is run for 150 iterations and it converges at a count of fourty nodes for three clusters, 34 for one cluster and the node count below five as shown in the image above corresponds to the outliers.

Figure 3 shows a user-freindly GUI developed for indoor setting, displaying the results of clustering and node counting algorithms (left corner). Additionally, the GUI has controls to tune η and ϵ parameters of DBSCAN algorithm and options to perform node and edge counting in a distributed manner.

VI. Conclusion

We presented a distributed algorithm to estimate the network size of each cluster of potential hotspots. We illustrated the working of our algorithm via simulation results displayed on the developed GUIs. Our algorithms will enable users to estimate the density of users in a region, leading to the estimation of the seriousness of risk. We envision to include the localization algorithm in our proposed approach, which will enable us to calculate the proximity among the users. Our approach together with location mapping techniques can provide real time risk information to mobile phone users about the areas that they plan to visit. Our research has applicability to other fields such as 5G+ communications, increasing the accuracy of location information, indoor user tracking, and infrastructure-free implementations applicable to robotics, location-aware patient care and other mobile health applications.

References

[1] Yanfang Ye, Shifu Hou, Yujie Fan, Yiyue Qian, Yiming Zhang, Shiyu Sun, Qian Peng, and Kenneth Laparo, "α-satellite: An ai-driven system and benchmark datasets for hierarchical community-level risk assessment to help combat covid-19," *arXiv preprint arXiv:2003.12232*, 2020.

[2] Nuria Oliver, Emmanuel Letouzé, Harald Sterly, Sébastien Delataille, Marco De Nadai, Bruno Lepri, Renaud Lambiotte, Richard Benjamins, Ciro Cattuto, Vittoria Colizza, et al., "Mobile phone data and covid-19: Missing an opportunity?," *arXiv preprint arXiv:2003.12347*, 2020.

[3] Sai Zhang, Cihan Tepedelenlioglu, and Andreas Spanias, "Distributed network center area estimation," US Patent 10,440,553, Issued Oct. 2019.

[4] Ioannis D Schizas, Gonzalo Mateos, and Georgios B Giannakis, "Distributed lms for consensus-based in-network adaptive processing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2365–2382, 2009.

[5] Shanying Zhu, Cailian Chen, Xiaoli Ma, Bo Yang, and Xinping Guan, "Consensus based estimation over relay assisted sensor networks for situation monitoring," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 2, pp. 278–291, 2014.

[6] Demetri P Spanos, Reza Olfati-Saber, and Richard M Murray, "Dynamic consensus on mobile networks," in *IFAC world congress*, 2005, pp. 1–6.

[7] Sai Zhang, Cihan Tepedelenlioglu, Andreas Spanias, and Mahesh Banavar, "Distributed network structure estimation using consensus methods," *Synthesis Lectures on Communications, Morgan & Claypool*, vol. 10, no. 1, pp. 1–88, 2018.

[8] Sai Zhang, Cihan Tepedelenlioğlu, Mahesh K Banavar, and Andreas Spanias, "Distributed node counting in wireless sensor networks in the presence of communication noise," *IEEE Sensors Journal*, vol. 17, no. 4, pp. 1175–1186, 2016.

[9] Scott McLachlan, Peter Lucas, Kudakwashe Dube, Graham A Hitman, Magda Osman, Evangelia Kyrimi, Martin Neil, and Norman E Fenton, "Bluetooth smartphone apps: Are they the most private and effective solution for covid-19 contact tracing?," *arXiv preprint arXiv:2005.06621*, 2020.

[10] Xue Zhang, Cihan Tepedelenlioğlu, Mahesh K Banavar, Andreas Spanias, and Gowtham Muniraju, "Location estimation and detection in wireless sensor networks in the presence of fading," *Physical Communication*, vol. 32, pp. 62–74, 2019.

[11] Monalisa Achalla, Kevin Mack, Mahesh K Banavar, M Vanitha, and Harish Krishnamoorthi, "Statistical methods for fast los detection for ranging and localization," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. IEEE, 2020, pp. 1–5.

[12] Gowtham Muniraju, Sai Zhang, Cihan Tepedelenlioglu, Mahesh K Banavar, Andreas Spanias, Cesar Vargas-Rosales, and Rafaela Villalpando-Hernandez, "Location based distributed spectral clustering for wireless sensor networks," in *2017 Sensor Signal Processing for Defence Conference (SSPD)*. IEEE, pp. 1–5, London, Dec 2017.

[13] Yasutaka Furukawa and Yoshihisa Shinagawa, "Accurate and robust line segment extraction by analyzing distribution around peaks in hough space," *Computer Vision and Image Understanding*, vol. 92, no. 1, pp. 1–25, 2003.

[14] Varsha Kamat-Sadekar and Subramaniam Ganesan, "Complete description of multiple line segments using the hough transform," *Image and Vision Computing*, vol. 16, no. 9-10, pp. 597–613, 1998.

[15] Mahesh K Banavar, Shandeepa Wickramasinghe, Monalisa Achalla, and Jie Sun, "Ordinal unloc: Target localization with noisy and incomplete distance measures," *IEEE Internet of Things Journal*, 2021.

[16] Jinfei Liu, Joshua Zhexue Huang, Jun Luo, and Li Xiong, "Privacy preserving distributed dbscan clustering," in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, 2012, pp. 177–185.

[17] Angelia Nedic, Asuman Ozdaglar, and Pablo A Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.