On the Properties of Next Generation Wireless Backhaul

Gabriele Gemmi, *Student Member, IEEE,* Renato Lo Cigno, *Senior Member, IEEE*, Leonardo Maccari, *Member, IEEE*

Abstract—With the advent of 5G, cellular networks require a high number of base stations, possibly interconnected with wireless links, an evolution introduced in the last revision of 5G as the Integrated Access and Backhaul (IAB). Researchers are now working to optimize the complex topologies of the backhaul network, using synthetic models for the underlying visibility graph, i.e., the graph of possible connections between the base stations. The goal of this paper is to provide a novel methodology to generate visibility graphs starting from real data (and the data sets themselves together with the source code for their manipulation), in order to base the IAB design and optimization on assumptions that are as close as possible to reality. We introduce a GPU-based method to create visibility graphs from open data, we analyze the properties of the realistic visibility graphs, and we show that different geographic areas produce very different graphs. We run state-of-the-art algorithms to create wireless backhaul networks on top of visibility graphs, and we show that the results that exploit synthetic models are far from those that use our realistic graphs. Our conclusion is that the data-based approach we propose is essential to design mobile networks that work in a variety of real-world situations.

Index Terms—5G Backhauling, IAB, Mesh Topologies, Distributed Networks, Visibility Graphs

I. Introduction

The performance requirements of the fifth generation of cellular networks (5G) foresee an extreme densification of Base Stations, named next Generation NodeB (gNB). While the target of 4G was to reach 8-10 Base Stations per km², 5G will need tens and maybe hundreds of gNBs per km² [1], and networks beyond 5G probably even more. The interconnection of gNBs providing gigabit performance to mobile terminals requires novel solutions in the access network [2] [3]. Release 16 of the 3GPP standard for 5G introduced the concept of Integrated Access and Backhaul (IAB) to support densification without skyrocketing the cost [4]. The core idea of IAB is that gNBs functions can be split in two parts: the majority of the gNBs are *IAB-nodes* that collect the traffic from user terminals and have no wired connection, some gNBs are instead *IAB-donors* that are fiber connected to the network core. A wireless backhaul between gNBs needs to be created to route the user traffic from IAB-nodes to the closest (or best) IAB-donor.

The same concept is at the core of 6G, that plans to go beyond mmWave frequencies (as in 5G) aiming at THz communications [5]. Higher frequencies provide higher capacity,

R. Lo Cigno is with the Department of Information Engineering, University of Brescia, Italy (e-mail: renato.locigno@unibs.it)

G. Gemmi and L. Maccari are with the Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy. (e-mail: gabriele.gemmi@unive.it; leonardo.maccari@unive.it)

but have limited communication range and almost no capacity to penetrate obstacles, thus requiring Line of Sight (LoS) between transmission end-points, which reflects in an even higher densification of IAB-nodes. Release 16 foresees only a directed acyclic graph (DAG), possibly multi-hop topology, but the extension to meshed topology to enhance performance, reliability, and dependability is already under discussion.

Given this trend, wireless mesh networks will play a key role in future cellular networks, which we collectively refer to as `nextG`, whose backhauls will use mmWave and THz communications with massive MIMO antennas, enabling steerable directional beams that can be activated dynamically; finally they will integrate computing inside the network with Multi-Access Edge Computing (MEC) to support smart "verticals" like cooperative driving [6], [7]. We call this kind of network Next Generation Wireless Backhaul (NGWB).

Given a city made of hundreds of thousands of possible locations for IAB-nodes (buildings, traffic lights, light posts, etc.), and given some performance parameters (delay, throughput, reliability, energy efficiency, . . .), the design of NGWBs requires first the selection of the node placement, next the decision on links activation [8], [9] and where to place network functions [10], and finally how to properly route traffic.

It is clear that NGWBs open extremely interesting perspectives for the application of network science to communication networks. Yet, one of the limitations we face is the lack of data to design and test algorithms. So, as it happens in many contexts, synthetic topologies based on heuristic considerations are used in design and simulation, making scientific results depending on some abstract representation of reality. The accuracy and credibility of analysis and simulations in a context that cannot be validated with real data is an old problem [11], and it is still a problem today, as in very popular research areas only a tiny fraction of papers provide source code and data to reproduce results [12]. This paper moves a first step in the direction of providing a robust methodology to study the multi-faceted challenges that NGWB networks provide, with a data-based approach suitable to exploit network science, and Open Data and Software to build upon.

The main focus of this paper is the analysis of several network topologies obtained with a 3D ray-tracing approach boosted by the use of Graphical Processing Units (GPUs) to speed up computations. Starting from open data describing buildings elevation in 9 different regions populated by about 15 M people, we obtain the visibility graph, i.e., the graph $G_v(\mathcal{N}_v, \mathcal{E}_v)$ where $\mathcal{N}_v$ is the set of all locations on buildings in the area that we consider potential gNB locations and $\mathcal{E}_v$ is the

set of all the potential links that could be realized considering the Line-of-Sight (LoS) constraint. We do not consider instead street lamps and similar locations as they can, at most, be used as DAG extensions to a meshed NGWB. Since LoS is a prerequisite to make wireless links at high frequencies, the visibility graphs enable the study of NGWB with a novel perspective. We try to answer some key questions like: *i) is a wireless mesh backhaul network feasible in areas with different population densities?* and, *ii) is it reasonable to assume that the visibility graphs of realistic networks in different settings have common features?* We anticipate that the answer to the first question is positive, while the second answer is negative, which confirms the base intuition that motivated our work: We need real data to obtain realistic results. We corroborate this intuition by comparing some key metrics computed on realistic topologies with graphs obtained from SoA synthetic models, showing extreme differences. Finally, we go beyond the observation of graph properties, and we apply one state of the art algorithm for the creation of a wireless mesh backbone using both the realistic graphs and the original model used by the authors. The analysis confirms once more that there is a striking difference between the results on synthetic models and on realistic graphs.

A wireless mesh network like NGWB adds one more level below the physical connection graph: The visibility graphs $G_v(\mathcal{N}_v, \mathcal{E}_v)$ that defines the *potential* communication links between nodes, as contrasted to the *actual* communications links that define the graph of the physical connections $G_\varphi(\mathcal{N}_\varphi, \mathcal{E}_\varphi)$. The former depends on the properties of the 3D space where the network is built, and is made of all the potential nodes that are in LoS and all the edges connecting them. The latter depends on the algorithm that is used to *embed* a communication network in a visibility graph, i.e., to choose which nodes and which edges are activated. Needless to say, the properties of G_v strongly influence the construction of G_φ, which in turn strongly influences the performance of applications running on G_φ.

Summing up, the contribution of this work moves along three intertwined lines:

- The introduction of a data-driven methodology for the computation of the visibility G_v, which is the fundamental constraint to build the communication graph G_φ;
- The description and analysis of G_v to understand its properties and features;
- The study of how the features of a realistic G_v impact the design of G_φ with algorithms at the state of the art.

II. Background and State of the Art

We divide the SoA analysis in three different parts detailed in the subsections that follows.

1) SoA on Visibility Graphs G_v: The study of visibility graphs is essential also outside the networking area, for instance, in landscaping [13]. Early studies on the impact of G_v in communications are limited to non-urban sites due to the low accuracy of Digital Elevation Models (DEMs) and to small areas due to the inherent complexity of determining the visibility [14]. The availability of accurate digital maps and

technologies such as Laser Imaging Detection and Ranging (LiDAR), which increase the vertical accuracy of building height estimation down to centimeters and sub-meter horizontal mapping, and the use of GPUs allowing extra-fast computation of some tasks, empowers visibility analysis with classical ray-tracing algorithms [15] in large urban areas [16]. In this paper we use [15] as a building block to calculate the viewshed from a single point, and we implemented a new methodology that enables the computation of the visibility graph on large areas with tens of thousands of buildings. The open source code for Nvidia GPUs is available online.[1]

2) SoA on Wireless Backhaul Creation: In the recent past, several works proposed algorithms to design a wireless backhaul G_φ for 5G networks trying to maximize different metrics: reliability [17], energy efficiency [18], and cost [19]. These studies rely on Millimeter Wave (mmWave) links or free-space-optical links [20], both requiring LoS between the endpoints. The introduction of IAB in the latest revision of 5G provided a concrete application and reinforced the interest in the creation of efficient backhaul topologies [4], [8], [9], [21]. The design of G_φ is not only interesting per se, but has direct implications on MEC and Virtual Network Function placement, that inevitably depend on the network topology [22], [23].

To the best of our knowledge the works that focus on the creation of G_φ take a very simplistic approach to model the characteristics of G_v, due to the total absence of literature that characterizes the realistic properties of G_v. Since the publication of realistic G_v topologies is one of our contributions, we stress its importance showing that the performance of some state-of-the-art proposals for the creation of G_φ change significantly when applied to a real G_v or to one generated with simplistic assumptions. The work that is somehow closer to this paper is our previous contribution [24], which uses a similar ray-tracing approach, but with the objective of comparing growth strategies for Community Networks. The methodology is also different: This work uses a GPU-based approach that enables the computation of the whole G_v for entire cities to extend the analysis to an unprecedented scale. What we retain from our previous work is the name of the software tool, called *TrueNets*.

3) Path loss models for urban areas: In the context of cellular networks and flying networks, there is a strong interest in trying to estimate the probability of LoS and the path loss over a link connecting two points in space. We mention and later use for comparison two approaches taken from the literature, even if they were not derived in the context of rooftop backhauls. To the best of our knowledge, no synthetic model for rooftop backhauls (or in general mesh networks) LoS probability exists, thus we chose for comparison two models derived for scenarios that are close to our analysis. The results we present in Sect. V-D show that they are, as predictable, inadequate, thus calling for either data-driven network design or further efforts in LoS probability modeling.

The first one uses state-of-the-art stochastic models for LoS such as the ETSI 3GPP [25] or the WINNER II [26]

---

[1]https://github.com/UniVe-NeDS-Lab/TrueNets

models. These approaches model the probability of LoS with an exponential decay for different scenarios (Urban-Micro, Urban, Rural, Suburban). LoS probability for the Urban-Micro scenario, the one closer to our scenario, is defined as:

$$P_{\text{LoS}}^{\text{ETSI}}(d) = min\left(\frac{18}{d}, 1\right)\left(1 - exp\left(-\frac{d}{36}\right)\right) + exp\left(-\frac{d}{36}\right) \quad (1)$$

where d is the distance between the two points. The formulas for the other models differ only in the parameters' values and can be found in the WINNER II documentation. Albeit this model refers to the LoS probability between a base station and a terminal on the ground, it was used in the works that address the design of IABs [4], [21] that we use as a benchmark.

The second model we consider is provided by Al-Hourani [27] and is intended to estimate the LoS probability between two flying objects at height h_1 and h_2 respectively, at a certain distance d. LoS probability is defined as:

$$P_{\text{LoS}}^{\text{AL-H}}(d, h_1, h_2) = \exp\left(-2r_0\lambda_0 \int_0^{d-\frac{\pi}{2}r_0} G(h)dx\right) \quad (2)$$

where r_0 is the average radius of the buildings (approximated as cylinders), λ_0 is the average building density and $G(h)$ is the complementary cumulative distribution function (CCDF) of the building heights. Contrary to the ETSI model, the Al-Hourani model requires the knowledge of the properties of the area where it is applied. Whenever we use it, its parameters are fitted to the data of the area we consider.

III. VISIBILITY ANALYSIS: GENERATING G_v

We first define some fundamental concepts of visibility analysis, which we then apply to generate G_v.

A Digital Elevation Model (DEM) is a matrix of geolocated altitudes with a certain spatial density, which is the resolution of the digital map considered (in our case, 1 point per squared meter). DEMs can be obtained using different technologies, such as Radar, Photogrammetry, LiDAR, etc. However the only datasets which are both widely available and have sufficient precision are the ones obtained through LiDAR technologies by Public Administrations and released as Open Data. In our research we take advantage of the datasets published by the Italian Ministry of Environment[2]. Given the origin of the x/y plane set in $(1, 1)$, the DEM is a matrix of real numbers $\mathbf{E} \in \mathbb{R}^{m_x \times m_y}$, where m_x and m_y are the number of samples in the x and y dimensions respectively, and $\mathbf{E}_{x,y}$, is the elevation measured in $(x, y) \in \mathbb{N}^2$, which are the integer indexes of the points in the map. Consider an ordering $o(x, y)$ on the matrix indexes and let $i = o(x, y)$ and $j = o(x', y')$. Any ordering is valid, but we can consider $o(x, y) = xm_x + y$ if $m_x \geq m_y$, or $o(x, y) = ym_y + x$ otherwise, to fix ideas. We call p_i and p_j the points in space identified by the triplets (x, y, z) where $i = o(x, y)$, $z = \mathbf{E}_{x,y}$, and (x', y', z') where $j = o(x', y')$, $z' = \mathbf{E}_{x',y'}$ respectively. The *visibility analysis* is the process of determining, for every point p_i (called the observer point) all the other points p_j that have visibility with p_i.
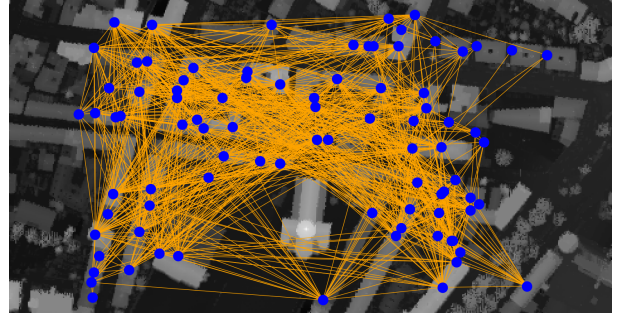

Fig. 1: Visibility graph of the sample area of Fig. 2 (for clarity, only a subset of the points are reported).

	Name	Extension (km^2)	Pop. dens. $(ab./km^2)$	Buildings	Build. dens. $(buil./km^2)$
U1	Trento	187.88	749.24	16490	104.45
U2	Firenze	102.32	3702.54	44598	435.87
U3	Napoli	119.02	8058.82	50879	423.90
S1	Mezzo-lombardo	13.88	517.10	1285	92.01
S2	Barberino di Mugello	133.29	82.05	2338	17.54
S3	Sorrento	9.95	1639.25	3856	387.32
R1	Predaia	80.05	83.39	2333	29.14
R2	Pontremoli	182.48	39.36	4079	22.35
R3	Visciano	10.90	401.60	1266	116.11

TABLE I: Main urbanization parameters for the 9 areas

Let $\Upsilon(\mathbf{E}, p_i, p_j)$ be a function that returns 1 if there is direct LoS between p_i and p_j and 0 otherwise. The visibility analysis is defined formally as the process of obtaining the *viewshed* from p_i, i.e., a matrix $\mathbf{V}^i$ of size $(m_x \times m_y)$ defined as:

$$\forall p_j \quad \mathbf{V}^i_{x',y'} = \Upsilon(\mathbf{E}, p_i, p_j) \quad (3)$$

Evaluating $\Upsilon(\mathbf{E}, p_i, p_j)$ is a computationally intensive task, but recent advances in GPU optimized algorithms [16] allow performing this process in areas that contain tens of thousands of buildings, with billions of potential links to be tested. The design of this process, together with the software provided to the community, is one of the contributions of this paper.

We obtained the DEM of 9 real-world areas, corresponding to 9 administrative municipalities in Italy. The areas belong to urban, suburban and rural areas[3], whose properties are reported in Tab. I. To process the data we use the Numba libraries that exploit the CUDA architecture for NVIDIA GPUs[4].

Our goal is to study the visibility graph G_v where each vertex of the graph corresponds to one point on the roof of an existing building, so we need the shapes of the buildings in the specified areas. This information is obtained using two different sources: Openstreetmap[5] and the open data made

---

[2]http://www.pcn.minambiente.it/mattm/en/

[3]The urban/suburban/rural label was obtained using the Eurostat definitions, see https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Territorial_typologies_manual_-_degree_of_urbanisation

[4]See the Numba documentation at https://numba.pydata.org/numba-doc/latest/cuda/index.html.

[5]More information on the project at https://www.openstreetmap.org/.

available by the single municipalities. The latter source is more accurate but not necessarily updated, while the first source is more frequently updated especially in urban areas, but may miss some buildings. For each area, we use the source that provides the highest density of buildings per km^2 without attempting to merge the data, a process that would be too error-prone. Therefore, given a closed polygon s_k that represents building b_k we implement a function that provides a binary matrix $\mathbf{S}^k$ of the same dimension of $\mathbf{E}$ so that:

$$\mathbf{S}^k_{x,y} = 1 \iff (x,y) \in A(s_k) \tag{4}$$

where $A()$ returns the area limited by s_k. In the following we detail the steps needed to create the annotated visibility graph G_v once we have $\mathbf{S}^k$ for all the buildings in a certain area. The code and the data used in the research are available to the community for further research and results' validation/falsification[6].

Algorithm 1 outlines the general methodology to build G_v once the nodes' locations $\mathcal{N}_v$ have been determined. The algorithm requires $\mathbf{E}$, which is the DEM and $\mathcal{N}_v$ which is the set of nodes' locations, and computes the line of sight for every couple of points in $\mathcal{N}_v$. The detailed algorithm that defines the function Υ is available in the additional material.

Algorithm 1 Algorithm for the computation of G_v.

Require: $\mathbf{E}$ (DEM of the area), $\mathcal{N}_v$ (Set of points),
1: **procedure** GENERATE G_v($\mathbf{E}$, $\mathcal{N}_v$)
2:     $G_v = \mathbf{0}$
3:     **for** $p_i \in \mathcal{N}_v$ **do**
4:         **for** $p_j \in \mathcal{N}_v \setminus \{p_i\}$ **do**
5:             $G_v[i,j] = \Upsilon(\mathbf{E}, p_i, p_j)$
6:     **return** G_v

A. Roof Node Placement

The first step for the computation of G_v is to determine $\mathcal{N}_v$, which requires to attribute one point p_i to each building b_k, so $|\mathcal{N}_v|$ equals the number of buildings in the area. The precise position is important because it influences the chances of having LoS with the other chosen points. In the real world this is done inspecting the roof and visually searching for the place with the best visibility towards other buildings. In an automated procedure, we need an algorithm that tries to maximize the probability of LoS towards other buildings.

For every s_k we are looking for a point p_i with coordinates $(x,y) \in A(s_k)$ so that the resulting graph G_v has the highest number of edges. An exact algorithm needs to explore all the possible combinations of all the $(x,y) \in s_k$ for every $k \in [1, \ldots |\mathcal{N}_v|]$, whose number grows exponentially with $|\mathcal{N}_v|$. By modeling the problem as a colored graph, where each point k on the roof is colored with the same color, we can reduce it as the search for the connected rainbow subgraph, which has been proven to be NP-Hard [28]. For this reason we need a heuristic.

A reasonable assumption is that points that have a good visibility toward the centroids of buildings will also have good visibility toward the good visibility points on the buildings, so that, starting from centroids, we can select these points with an algorithm that scales quadratically with the number of buildings. However, centroids may actually fall outside the building roof (e.g., internal courts), so we use the point c_k, which is the point of coordinates (x_k, y_k, z_k) where (x_k, y_k) are the coordinates of the a suitable point that ensures $(x_k, y_k) \in A(s_k)$ as defined by the C++ function `GEOSPointOnSurface` from the `GEOS` library. For the sake of readability we'll call these points pseudocentroids.

We define the pseudocentroid $c_k = (x_k, y_k, z_k{=}\mathbf{E}_{x,y}{+}2)$ adding 2 m to z_k since we assume gNB antennas are elevated on the roof with a pole[7] and we compute all the viewsheds $\mathbf{V}^k$ from any point c_k to any other point also elevated by 2 m. Since Υ is a symmetric function, given a point $(x, y, \mathbf{E}_{x,y}{+}2)$, each element $\mathbf{V}^k_{x,y}$ represents the availability of LoS from $(x, y, \mathbf{E}_{x,y}{+}2)$ to c_k. Summing all $\mathbf{V}^k$ we obtain $\mathbf{V} = \sum_{k=1}^{|\mathcal{N}_v|} \mathbf{V}^k$, which is a cumulative visibility matrix whose values range from 0 to $|\mathcal{N}_v|$ indicating how many centroids c_k are in LoS from any elevated point on every building.

The second step is searching, for every building b_k, the point $\bar{p}_i$ that has the highest visibility of centroids. This is obtained masking $\mathbf{V}$ with $\mathbf{S}^k$: $\mathbf{V} * \mathbf{S}^k$ (where $*$ is the element-wise multiplication) selecting all the values of $\mathbf{V}$ belonging to points inside b_k, and returning the coordinates of the maximum:

$$(\bar{x}, \bar{y}) = arg\,max_{(x,y)}(\mathbf{V} * \mathbf{S}^k) \tag{5}$$

then $\bar{p}_i = (\bar{x}, \bar{y}, \bar{z}_k = \mathbf{E}_{\bar{x},\bar{y}} + 2)$ is going to be the position of the gNB antennas for building b_k. We repeat this process for every building and we obtain $\mathcal{N}_v$ as the collection of all the $\bar{p}_i$. From now on, when we mention a generic point p_i, we always refer to points chosen with this procedure.[8]

Fig. 2 shows a sample of the visibility matrix $\mathbf{V}$ as a greyscale image in which every pixel is a point in $\mathbf{V}$, with lighter color meaning higher visibility index; the chosen $\bar{p}_i \in \mathcal{N}_v$ are reported in yellow.

B. Building the visibility graph

Once the set $\mathcal{N}_v$ has been determined we need to build the set $\mathcal{E}_v$ computing $\Upsilon(p_i, p_j)$ for each couple of points $p_i, p_j \in \mathcal{N}_v$, tracing a ray between p_i and p_j and checking whether any obstacle intersects the ray.

As in the previous task, implementing this algorithm in a GPU allows parallelizing the task among the large number of cores available. Each core is assigned to one point $p_i \in \mathcal{N}_v$ and calculates the LoS towards any other point $p_j \neq p_i \in \mathcal{N}_v$ using well known algorithms from ray-tracing literature [15]. The pseudocode of our specific implementation is available in the additional material. Fig. 1 shows the visibility graph

---

[6]The code archive is at https://github.com/UniVe-NeDS-Lab/TrueNets and the data are at http://doi.org/10.5281/zenodo.4905536

[7]This is an arbitrary choice that can be modified according to the design constraints and requirements.

[8]Note that this introduces a little abuse of notation. So far we used i as the index of point p_i in the ordering $o(x, y)$ for the coordinates in the DEM, and k as an index for buildings, so the two indexes are uncorrelated. From now on we will simply refer to p_i as the point chosen for building i.
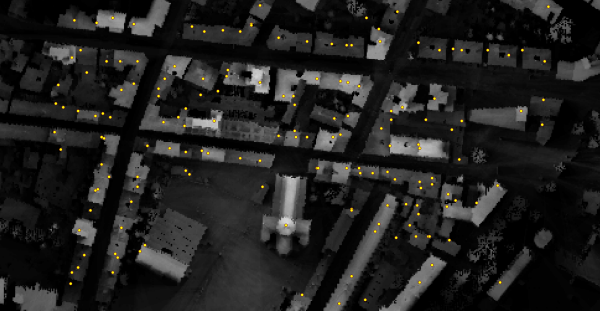
Fig. 2: Outcome of the selection of the best visibility point –in yellow– for each building in a small portion of an urban area, the shade of grey indicates the visibility index with light grey indicating higher visibility.

of the sample area. The properties of the resulting graphs are reported in Tab. II and commented in Sect. V.

C. Complexity analysis

The worst-case time complexity for building G_v is $O(|\mathcal{N}_v|^2) \cdot C(\Upsilon)$, where the first part represents the iteration on every pair of nodes and the second part represents the complexity of the LoS computation between two given points. $C(\Upsilon)$ depends on the longest LoS link for which the algorithm needs to evaluate every space element intersected by the ray. Assuming that $\mathbf{E}$ is a square matrix, the longest link would be its diagonal which is $\sqrt{2|E|}$, which leads to $C(\Upsilon) = O(\sqrt{|E|})$, thus a general complexity equals to $O(|\mathcal{N}_v|^2 \cdot \sqrt{|E|})$. In order to express the complexity solely on the size of the area we can take advantage of the fact that the number of buildings grows linearly with the area. This leads to an overall complexity of $O(|E| \cdot \sqrt{|E|})$

On the other hand, the SoA algorithm for viewshed analysis [15] has a complexity of $O(|E|^2)$ to calculate a single viewshed. In order to calculate the visibility graph the complexity would be $O(|\mathcal{N}_v| \cdot |E|^2)$, which again expressed only in terms of the area is $O(|E|^2 \cdot \sqrt{|E|})$.

Computing such an algorithm on datasets like U3 (more than 50 k nodes and 100 km^2) would not be feasible using a normal CPU, however modern GPUs, with their high number of cores and large RAM allow to speed up the process and compute the whole visibility graph in a reasonable time. We used an NVIDIA Tesla P100 GPU which has 3584 cores and 16 GB of memory. This allows the computation of the whole process for the largest city in the data-set in roughly one hour at a speed of 40 M-links per second.

IV. Generation of G_φ

Given G_v and the position of a set $\mathcal{N}_\varphi \subset \mathcal{N}_v$ of gNBs, G_φ is embedded into G_v choosing a set $\mathcal{E}_\varphi \subset \mathcal{E}_v$ that will create the wireless backhaul. Our goal is to show that the realistic data-set of G_v topologies that we publish is key to obtain realistic results in the design of a next generation wireless backhaul G_φ. Contrarily to classical large-scale, low density mesh networks that span across multiple municipalities [29] a NGWB is expected to extend a wired backhaul in localized

regions, so in this section we describe three processes: i) how we create realistic localized G_v and $\mathcal{N}_\varphi$; ii) how we create simplistic localized G_v and $\mathcal{N}_\varphi$ for comparison; iii) two strategies from the state of the art to choose $\mathcal{E}_\varphi$, whose performance will strongly vary when applied to realistic or simplistic data.

A. A Realistic Localized G_v and $\mathcal{N}_\varphi$

In each of the 9 areas we select 5 sub-areas of approximately 1×1 km, and in each sub-area we call Γ the border of the convex hull that includes all the buildings fully contained in the sub-area, with $A(\Gamma) \leq 1$ km^2 its area. We choose $\mathcal{N}_v$ assigning a point p_i to each building in the area with the procedure described in Sect. III, and finally we produce the localized realistic G_v applying Algorithm 1 to these points. Given a desired density of gNB per km^2 ($\rho \in \{30, 60\}$), we set the size $|\mathcal{N}_\varphi| = \rho A(\Gamma)$ and we pick a random set $\mathcal{N}_\varphi \subset \mathcal{N}_v$. In the figures we simply refer to data generated with this process with the TrueNets label.

B. Simplistic Localized G_v and $\mathcal{N}_\varphi$

Given the same 9 areas and $|\mathcal{N}_\varphi|$ as defined above, we generate two versions of a simplistic G_v. As a baseline we pick $\mathcal{N}_v = \mathcal{N}_\varphi$ using a Homogeneous Poisson Point Process (which we refer to as HPPP) in which locations are chosen with a random uniform choice in $A(\Gamma)$, without any relation to the building maps. In order to create $\mathcal{E}_v$, for each couple of nodes in $\mathcal{N}_v$, we add an edge with a probability given by the ETSI model using Eq. (1). This simple strategy is the one used in most of the papers that propose approaches for the creation of wireless backhauls, such as the ones from Polese et al. [4], [21] mentioned later on.

In a second, slightly more realistic approach, we pick $\mathcal{N}_\varphi$ as in the realistic case of Sect. IV-A, but we use the ETSI model to generate the edges $\mathcal{E}_v$. This is an intermediate model in which the 2D distribution of the points is not completely uncorrelated from the city map, but is similar to the 2D distribution of the buildings. Yet, without 3D information, the edges are chosen with a synthetic model and not with the Υ function. As we use the OpenStreetMap data, we refer to this process as OSM.

C. Choosing $\mathcal{E}_\varphi$

Both the described processes produce a visibility graph G_v and a set $\mathcal{N}_\varphi$ of gNBs. In the IAB terminology, some of the nodes are *donors*, i.e., they are connected to the wired backbone, and the other nodes need to build a multi-hop path to some donor. We randomly choose $\lceil 0.1 \rho A(\Gamma) \rceil$ donors (at least 10% of the gNBs), as in [21]. For each donor a Directed Acyclic Graph (DAG) is created that interconnects the reachable nodes, and the union of all the DAGs provides G_φ. We mentioned in Sect. II-2 that there are several proposals to choose the DAGs, and thus create the backhaul graphs, among which we pick two. The first is one of the heuristic proposed by Polese et al. [21] that assumes no centralized coordination. We report the results for the algorithm named

	Name	Giant Component	Links	Avg. Degree
U1	Trento	16348 (99.1%)	21531378	1317
U2	Firenze	44221 (99.2%)	36981097	836
U3	Napoli	50552 (99.3%)	64141794	1268
S1	Mezzo-lombardo	1276 (99.2%)	327579	256
S2	Barberino di Mugello	2288 (97.9%)	474618	207
S3	Sorrento	3828 (99.2%)	2041968	533
R1	Predaia	1954 (83.7%)	834965	427
R2	Pontremoli	3869 (94.8%)	621069	160
R3	Visciano	1161 (91.7%)	169083	145

TABLE II: G_φ Network metrics for the 9 areas.
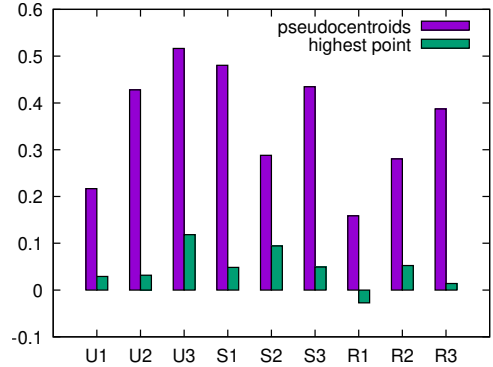

Fig. 3: Relative gain on the number of edges between our heuristic and other point selection strategies.

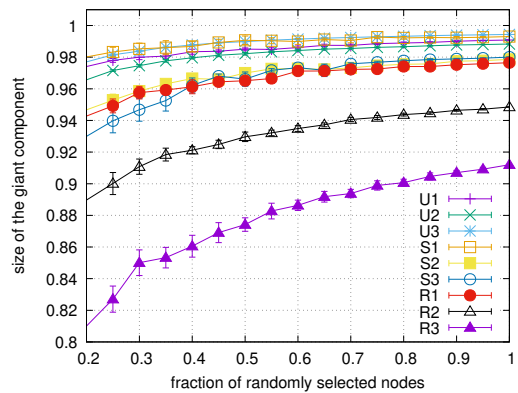Fig. 4: Size of the giant component with a randomly sampled subset of the nodes.

DPS_WF, in which each node tries to connect with a multi-hop path to the physically closest donor. The algorithm is distributed and greedy, thus not optimal. The second one is an optimal centralized strategy for the creation of the backhaul with the smallest distance (in hops) from each node to its donor. The strategy chooses $\mathcal{E}_\varphi$ as the union of all the edges that are in the shortest path from any node to the closest donor, computed with classical Dijkstra's algorithm. It is optimal in the sense that it minimizes the distance between each node and its donor.

Both algorithms are taken from literature and we use them to test the impact of G_v on the properties of G_φ.

V. ANALYSIS OF G_v

This section presents the features of the visibility graph G_v generated considering the whole areas and the 5 sub-areas of $1\,km^2$. Due to space constraints in the rest of the paper we include and comment only a small set of the figures we generated, that are enough to robustly support our conclusions. The rest of the figures can be found in the supplementary material, together with the links to the data sets and the source code.

A. Size of the giant component

Before we provide the results on the analysis of G_v, we want to highlight the importance of the algorithm chosen to select $\bar{p}_i$. We compute G_v with three different strategies: *i)* using the heuristic described in Sect. III-A; *ii)* using the pseudocentroid c_k; and *iii)* using the highest point on the roof. Fig. 3 shows the relative difference in terms of number of edges between our heuristic and the other two. For instance, in case of comparison between the heuristic (h) and the pseudocentroid (c) placement the metric is $\frac{|\mathcal{E}_v^h| - |\mathcal{E}_v^c|}{|\mathcal{E}_v^h|}$.

It is clear from the figure that the improvement with respect to the pseudocentroids is substantial. In fact in areas such as U3 the number of edges doubles. The comparison with the highest point of the roof still shows a relevant gain, up to 10% in U3, with one exception: R1, which is a mountain rural area composed of isolated hamlets at very different heights, where overall visibility is more influenced by the position of the hamlets than the characteristics of the buildings. Note

that, in absolute terms in U3 we gain 7,696,210 edges while in R1 we lose 13,187 edges, so overall the advantage is considerable. Moreover, the highest point in the roof could be hardly accessible (e.g., a chimney).

This confirms the importance of a solid and repeatable methodology to produce the visibility graphs like the one we provide. It also suggests that in the real world small differences in node positions reflect in large differences in the graph properties, i.e., the network density in the real world is extremely sensitive to small differences in nodes placement. This makes it an interesting challenge to define generic synthetic models able to capture that variability. From now on we only consider the points p_i selected with the heuristic.

Tab. II reports the number of nodes that can not be connected to the giant network component. It is always less than 2.1% for the urban and suburban areas and below 10% in the rural areas, with the exception of R1 where, due to the morphology of the area, roughly 17% of the nodes are not in the giant component. However, in the process of network construction that we followed we did not devote any effort to have full connectivity while in a real setting several ad-hoc solutions can be introduced, e.g., higher trellises on roofs, repeater nodes in strategic locations, even in the absence of a

building – recall that R1 is in a mountain area. So the first key finding is that a NGWB mesh network covering almost entire cities or vast suburban/rural areas is possible in 8 out of 9 settings, without any specific attempt to maximize coverage. This is itself an extremely interesting result, as it confirms that the concept of IAB is feasible in practice and encourages several applications that rely on LoS links, such as backup networks, community networks, the extension of wired access where it can not be provided, or even dedicated networks made of Free Space Optical links for quantum key distribution [30]. An additional result that provides further insights is shown in Fig. 4, where the size of the giant component is evaluated in the case when only a subset of the buildings, randomly selected, are used. The figure shows that even with a small percentage of randomly selected buildings it is still possible to build a backhaul connecting most of the nodes. In fact with just 30% of the buildings it is possible to connect more than 95% of the nodes in 7 out of 9 areas, and in all cases more than 80% of the selected nodes are connected. Note that this positive result should be considered as a lower bound, as operators can easily improve it choosing high buildings or those in strategic positions, instead of choosing at random.

B. Coverage, Degree Distribution, Link Length, LoS Probability

The topological properties of G_v do not show clear regularities among different areas or even inside the same area. Fig. 5 reports the degree distribution of all rural and urban areas, and shows that while the distributions for U2 and U3 seem to suggest a power law trend, the same trend is less distinguishable for U1 (also due to a more compact distribution). Rural areas have a noisy behavior indicating irregular degrees, also due to the smaller number of nodes in the area, and R3 has a second mode close to the maximum degree. The average degree is high, as it ranges from 145 to 1317. In the network with the lowest (highest) edge density, the average degree corresponds to 1.8% (20%) of the number of nodes.

The physical length of the links of G_v displays large differences even among areas of the same kind. Fig. 6 reports the cumulative density function of the link length for all the areas, and it shows that the curves for the three areas of the same kind are always distinct and they start to diverge very early in the graph.

These observations have two key direct implications for the performance of NGWBs, the first is that a high average degree implies a huge number of G_φ networks that can be embedded on G_v. This is very important because it enables to divide the physical network in a large number of virtual backbones to support different applications. This can be an enabler of the *network slicing* features of 5G: A high density of links provides many possible physical IABs to map slices on, each one with different performance in terms of delay, robustness, etc. The second implication is that the diversity in the link length strongly impacts the performance of the network, and the choice of the technology to build links on the selected edges, as the propagation of signals changes significantly with

the technology selected. This calls for techniques to build G_φ that are tailored for the specific target area, discouraging a one-size-fits-all approach, and justifies the need for the real-world data sets we publish.

C. Antenna elevation

The differences in G_v shown in Sect. V-B are due to two concurrent factors, a different distribution of building elevation, mostly due to terrain factors, and a different distribution of buildings in the 2D map of the area. Here we focus on the first one.

Fig. 7 reports the ECDF of the z value for all the points p_i (accounting for the ground elevation, building height and 2 m pole) in all areas and 2 sets of sub-areas. The z values are referred to the lowest point of the area.

Fig. 7 shows clear differences between areas, without a recognizable pattern even among areas of the same kind. R3 shows a bimodal behavior that is due to the earth altitude, rather than the buildings' height, while R2 shows a smooth trend that is quite different from the other curves. If we zoom on a single area, we find again different situations. U1 is surrounded by mountains, so that a different choice of the sub-area yield even more variability (Fig. 7b). The same can not be said for S2, in which the five sub-areas show a very similar behavior (Fig. 7c). Again, we observe that the variability of the data do not allow a single model for all the areas, and not even for sub-areas inside the same one. This has a strong impact on the accuracy of simulations that use a single model to describe every scenario, as we discuss in Sect. VI.

D. Comparison with SoA LoS models

We now compare the probability of LoS estimated on the graphs generated with TrueNets with the two LoS models introduced in Sect. II-3: The ETSI (or WINNER) in Eq. (1), and the Al-Hourani model in Eq. (2). As explained, these two models were not derived for rooftop backhauls, thus the goal of this analysis is understanding if these model are somehow adequate for the design of rooftop backhauls or not.

Consider a certain area on which we compute G_v, let us call $\mathcal{E}_p$ the set of all the *potential edges*, i.e. all the couples (p_i, p_j) of the points in $\mathcal{N}_v$, with $i \neq j$. An edge $e = (p_i, p_j)$ is present in $\mathcal{E}_p$ even if $\Upsilon(p_i, p_j) = 0$. We call $d(e)$ the length of e. For computational reasons, we extract a random fraction r ($r = 1\%$ in urban areas and 10% in sub-urban and rural areas) of edges from $\mathcal{E}_p$ called $\overline{\mathcal{E}}_p$, and we bin the edges in $\overline{\mathcal{E}}_p$ based on the edge length, with interval $\Delta = 200\,\mathrm{m}$:

$$B_l = \{e \in \overline{\mathcal{E}}_p \mid l\Delta < d(e) < (l+1)\Delta\} \tag{6}$$

Then for each $e \in B_l$ we compute the LoS probability $P_{\mathrm{LoS}}^{\mathrm{ETSI}}(d(e))$ using the ETSI model as in Eq. (1), and we calculate the average on the whole bin:

$$M_{\mathrm{LoS}}^{\mathrm{ETSI}}(l) = \frac{1}{|B_l|} \sum_{e \in B_l} P_{\mathrm{LoS}}^{\mathrm{ETSI}}(d(e)) \tag{7}$$

We repeat the same procedure to obtain $M_{\mathrm{LoS}}^{\mathrm{WINNER}}(l)$ (using the WINNER model for rural areas) and $M_{\mathrm{LoS}}^{\mathrm{AL-H}}(l)$ (for
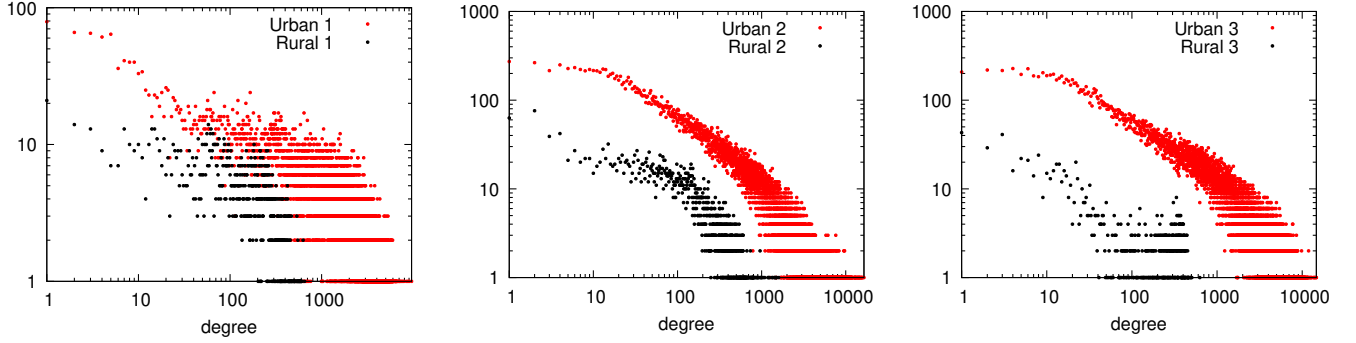
Fig. 5: Degree distribution for every location p_i in Urban and Rural areas.

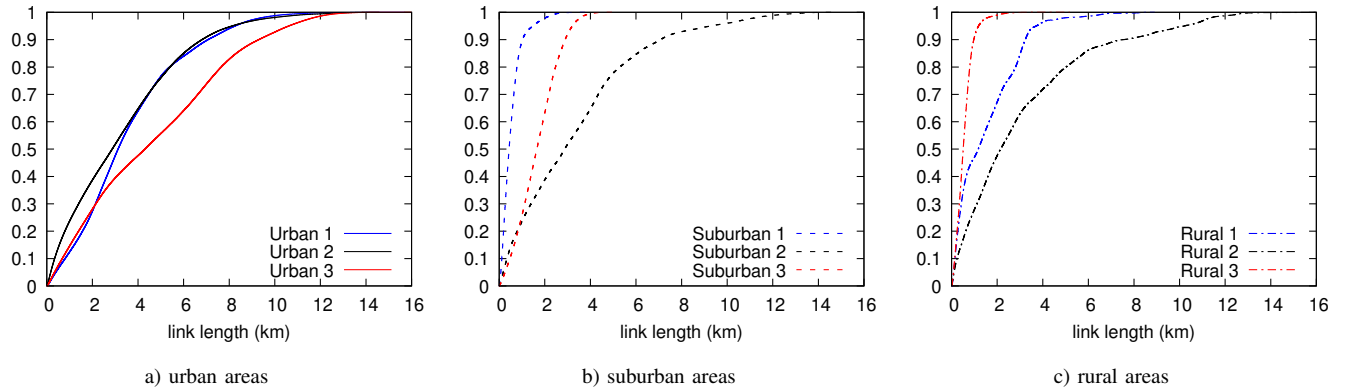a) urban areas

b) suburban areas

c) rural areas

Fig. 6: Empirical Cumulative Density Function (CDF) of links length for all the 9 areas

the Al-Hourani model). In the latter case we compute $P_{\text{LoS}}^{\text{AL-H}}(d(e), z, z')$ using Eq. (2), with $e = (p_i, p_j)$ and z, z' the elevation values of p_i and p_j. We replace $P_{\text{LoS}}^{\text{ETSI}}$ with $P_{\text{LoS}}^{\text{AL-H}}$ in Eq. (7) and we obtain $M_{\text{LoS}}^{\text{AL-H}}(l)$. Finally, we use TrueNets to compute $\Upsilon(p_i, p_j)$ and we have a fourth value:

$$M_{\text{LoS}}(l) = \frac{1}{|B_l|} \sum_{(p_i, p_j) \in B_l} \Upsilon(p_i, p_j) \qquad (8)$$

Fig. 8 contains different curves for three different areas (the supplementary material contains all the figures for all the areas), the x axis of each graph is cut to avoid the noise introduced by bins with less than 0.1% of the sampled edges. The upper graphs report the numerosity of the bins, which shows that in the urban areas the density of buildings smooths the distribution, with some small fluctuations. We observed this behavior also in the other two urban areas. The sub-urban area maintains some regularity, while the rural area shows a completely different behavior. In this case, the area seems to be partitioned in small clusters that generate the multi-modal shape of Fig. 8c, due to the settlement structure done by small, dense hamlets scattered in a mountain environment: two hamlets facing each other across a valley at distance x give a very large contribution of LoS links around this value.

The curves in the bottom part of the figure report the four values of M_{LoS}. We observe that the Al-Hourani model seems to differ largely from TrueNets values, which is due to two effects. The first is that Al-Hourani was designed for short links (below 250 m) so its application to longer links extends beyond its initial purpose. In the supplementary material we report the curves for the range 0-500 m that show (for some cases) a trend that is closer to M_{LoS}. Second, Al-Hourani models the LoS of drones moving in a 3D space, so the position of the drones is in the empty area among buildings, while we put our nodes on top of the buildings. We already observed the large difference caused by a change in the position of p_i inside the same building area (Tab. II), so it is not surprising that the Al-Hourani model does not fit a real-world LoS probability between building roofs.

The ETSI model for Urban-Micro has a trend that is reasonably similar to M_{LoS} but on a totally different scale, which suggests that it could be adapted to fit the real data at least in the initial part of the curve. One of the most interesting observation is that the WINNER models for the area types we consider (suburban and rural) yield completely different results: Not only they are a pure exponential models, but they also follow a decay that is often completely different from the measured values.
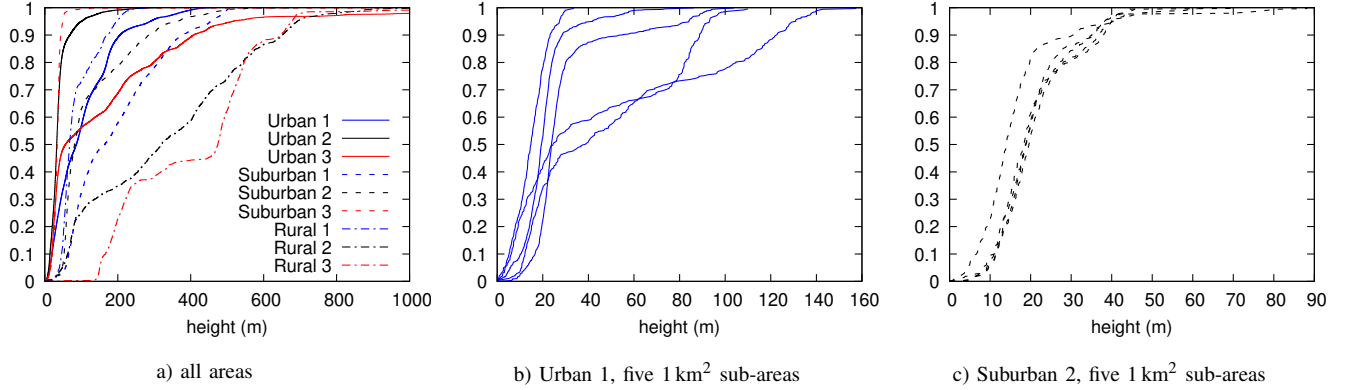
Fig. 7: ECDF of the z value (height) of all p_i in the 9 areas a) and for 5 subareas b) and c)

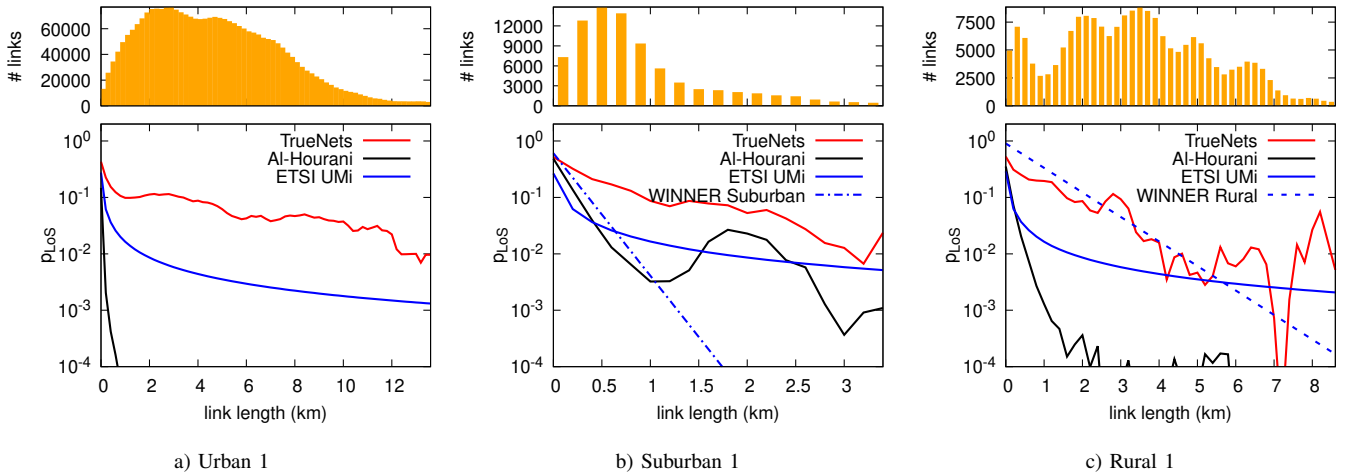a) Urban 1   b) Suburban 1   c) Rural 1

Fig. 8: Comparison between synthetic LoS probability models (ETSI and Al-Hourani) and measures with TrueNets.
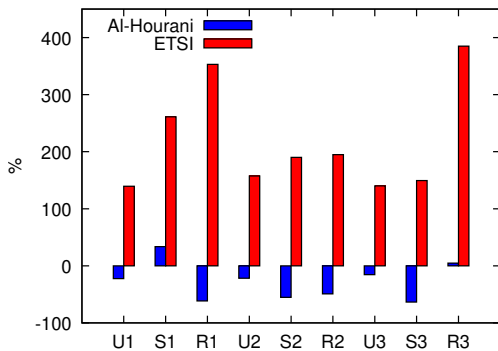
E. Robustness

Fig. 9: Effective Graph Resistance ξ of ETSI and Al-Hourani models as a percentage difference from ξ measured with TrueNets with 30 gNB/km^2

One intrinsic limit of both the ETSI and Al-Hourani model is that they use a radial symmetry, so given p_i they assume the probability of LoS with p_j is independent of the angle of the segment between p_i and p_j, which is of course not true in a real setting. To quantify the effect of the path loss model we need to evaluate the properties of visibility graphs generated with different LoS probability models. We already studied the properties of G_v on entire areas, so here we focus on G_v built among a subset of buildings in sub-areas that are most likely to be interesting for designing real communication graphs.

We use the methodology described in Sect. IV-A to select a random set of buildings in one sub-area, and then we build the visibility graph G_v among all the p_i on the selected buildings. We repeat the process ten times with different random seeds ($\rho = 30$ gNB/km^2) in all the 45 sub-areas obtaining a total of 45×10 random topologies.

Since availability and fault tolerance is a key requirement for nextG, we focus on a robustness metric: the *Effective Graph Resistance* ξ. ξ takes into account the presence of parallel (possibly disjoint) paths, and is computed as the average of the resistance between any two nodes s, d in the network, computed as if the graph were an electrical resistive circuit where the links have unitary resistances (the interested reader can refer to Ellens et al. for details [31]). ξ is defined

as:

$$\xi = |\mathcal{N}_v| \sum_{i=2}^{|\mathcal{N}_v|} \frac{1}{\lambda_i} \qquad (9)$$

where λ_i is the value of the i-th eigenvalue (ordered by their value) of the Laplacian matrix of graph G_v. ξ strictly decreases when an edge is added to the network, so the smaller ξ the more robust is in general the mesh, but also the larger is the overall capacity of a mesh built on top of it, as it can exploit more disjoint paths between nodes. For each generated graph we compute the relative percent difference as follows:

$$100 \frac{\xi_{\text{ETSI}} - \xi_{\text{TrueNets}}}{\xi_{\text{TrueNets}}} \qquad (10)$$

for the case of the ETSI model, and similarly for the Al-Hourani model.

Fig. 9 reports the values averaged on all graphs. It is evident that even using the ETSI model, that was somehow closer to the TrueNets data in Fig. 8, the robustness metric is completely different, and so the properties of the potential embedded G_φ will be. The Al-Hourani model maintains a higher similarity, but robustness still differs in a range between +33.6% and −63.4%. Again, this has direct implications on the design of reliable networks and in the support of integrated network slices.

VI. ANALYSIS OF G_φ

To underline the impact of realistic data in the evaluation of scientific contributions we selected two state-of-the-art algorithms for the generation of G_φ, and we compare its performance when applied to synthetic G_v models based on HPPPs (Homogeneous Poisson Point Processes) to place nodes and synthetic models of LoS probability, or to realistic G_v estimated with TrueNets. The first one is a greedy distributed algorithm from Polese et al. [21], which was proposed as an algorithm to deploy IAB networks [4]. The second one is the classical centralized shortest path algoritm from Dijkstra used to compute a multi-source spanning tree. As documented in Sect. IV we dissect the comparison in three parts, showing the impact of both modifications incrementally and finally together. The baseline, which we call HPPP+ETSI, is the synthetic model used by Polese [21]: $\mathcal{N}_\varphi$ is selected with an HPPP and $\mathcal{E}_\varphi$ is based on the ETSI UMi model. The intermediate one we call OSM+ETSI uses the realistic positions of $\mathcal{N}_\varphi$ from real-world data and retains the ETSI model for the selection of $\mathcal{E}_\varphi$. Finally, TrueNets uses both realistic node location and realistic edge LoS measured from our data set.

For each of the 45 sub-areas of $1 \, km^2$ and for each strategy we generate 10 different possible networks, each one with a different random choice of $\mathcal{N}_\varphi$ and donor nodes and a different resulting $\mathcal{E}_\varphi$ (the process is the same explained in Sect. IV-A) and we average the results. As usual, we report only the minimal set of results to support our conclusions, more results which corroborate the same conclusions are available in the supplementary material.
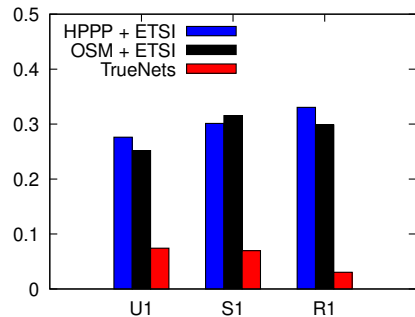
a) $\rho = 30 \, \text{gNB}/km^2$
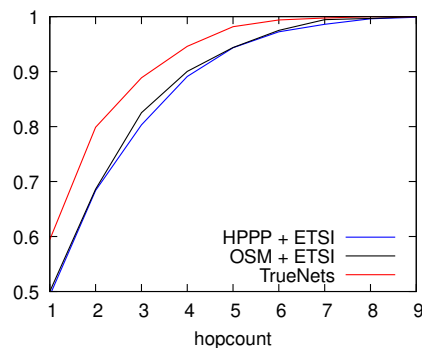
b) $\rho = 30 \, \text{gNB}/km^2$

Fig. 10: a) Ratio of unconnected nodes with DPS_WF strategy. b) ECDF of the distance of IAB nodes from the closest gNB donor in U3

A. Distributed Algorithm

The first result, reported in Fig. 10 a), shows the ratio of IAB-nodes that can not be connected with a multi-hop path to any IAB-donor. These nodes are isolated and can not be part of the backhaul network. This result indicates that the main difference is due to the LoS probability model. In fact, regardless of the nodes selection strategy for $\mathcal{N}_\varphi$, the fraction of disconnected nodes remains very similar. In all three areas the difference among the baseline and OSM+ETSI (blue and black bars) is below 30%. On the other hand, the usage of a realistic visibility graph dramatically impacts the ratio of unconnected nodes, with differences up to 700% due to the strong difference in the LoS probability as reported in Fig. 8. Increasing the density of gNBs, the fraction of disconnected nodes becomes marginal and also differences between strategies decrease as one may expect.

The second result, reported in plot b) of Fig. 10, is the metric used by the authors in the original research: The distance between each IAB node and its donor gNB in terms of hops, which is a key parameter to estimate the latency in an NGWB, but also to estimate the effective capacity given a technology to set up links on edges. The plots report the ECDF of the hopcount and show that not only the G_φ generated using synthetic data has between 20% and 30% fewer nodes (Fig. 10 a)), but also the length of the path to the donor is much longer.

B. *Centralized Algorithm*

The results found for the distributed algorithm generally holds true also for the centralized (and optimal) algorithm. The first result, reported in plot a) of Fig. 11 shows that the usage of a realistic visibility graph still provides some gains in the connectivity of nodes. In fact, albeit the values are more compressed, the usage of a realistic visibility graph diminish the number of unconnected nodes.

The second result, reported in plot b) of Fig. 11, also confirms the results found for the distributed algorithm. By using a realistic visibility graph, the paths are generally shorter due the fact that longer edges are less likely to be present in synthetic graphs than in reality. Additionally, we note that when employing a centralized algorithm the average number of hops diminishes significantly, suggesting that the algorithm proposed in [21] was not conceived to minimize the number of hops. Without further findings, a centralized algorithm may be preferable for delay-sensitive applications.

a) $\rho = 30\text{gNB}/km^2$
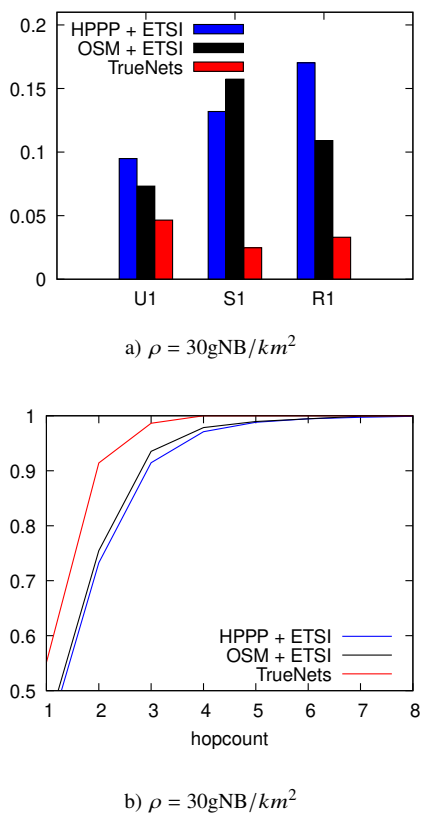

b) $\rho = 30\text{gNB}/km^2$

Fig. 11: a) Ratio of unconnected nodes with Dijkstra strategy. b) ECDF of the distance of IAB nodes from the closest gNB donor in U3

We can conclude that, in this case, the use of synthetic data produces results that are extremely pessimistic, primarily in terms of admission to the network, but also in terms of performance, so that an evaluation based on these models may lead to hamper the development of a technology that is instead very promising. These results do not change if we use synthetic models to estimate the LoS and realistic node positions, but they dramatically change when we use realistic data for both node positions and LoS availability, confirming the importance of data-based approaches and models for the generation of G_v.

VII. CONCLUSIONS

The key finding of this paper is the observation that a one-size-fits-all synthetic model for NGWB are not accurate enough to capture the complexity of the real world. Past experience tells us that when we ignore the variability of the real situations, systems simply fall short in delivering their expected social or economic value, or are not deployed due to pessimistic performance expectation.

For this reason, we propose a novel approach to the problem, that is to use open data together with ray-tracing to produce accurate data sets to study the performance of networks in their real environments, and we provide an initial data set together with open source code to extend it in the future. We believe the data and the methodology we provide help researchers to obtain robust results in several network planning tasks for NGWBs. Among them we mention:

- The study of the realistic coverage of `nextG` networks. We know that `nextG` requires critical densification of BSes, but only data can realistically quantify it;
- The study of the real applicability of flying networks to extend `nextG` coverage. Drones have been proposed for this task when needed, but again, their efficacy needs to be evaluated using real data;
- New, more accurate models that can capture the features of existing places for which we do not possess open data.

The last point is an intriguing one. With more datasets we can abstract new models that enable to estimate the performance of networks even in the absence of *all* the data. These models may depend on data that are easier to access than the full DEM, such as satellite images, ground elevation profiles (without building elevation), statistical distributions of building height in existing areas. Yet these models need new instruments coming from network science and data analysis to complement the classical measurement campaigns that generated the ones we use today.

ACKNOWLEDGMENT

REFERENCES

[1] C. Madapatha, B. Makki, C. Fang, O. Teyeb, E. Dahlman, M.-S. Alouini, and T. Svensson, "On integrated access and backhaul networks: Current status and potentials," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1374–1389, 2020.

[2] M. K. Shehzad, A. Ahmad, S. A. Hassan, and H. Jung, "Backhaul-Aware Intelligent Positioning of UAVs and Association of Terrestrial Base Stations for Fronthaul Connectivity," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 4, pp. 2742–2755, May 2021.

[3] J. Gui, X. Dai, and N. Xiong, "Design and analysis of network behaviors for optimizing network energy efficiency in 5G mmWave systems," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1842–1861, Apr 2021.

[4] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated Access and Backhaul in 5G mmWave Networks: Potential and Challenges," *IEEE Communications Magazine*, vol. 58, no. 3, 2020.

[5] M. Polese, J. M. Jornet, T. Melodia, and M. Zorzi, "Toward End-to-End, Full-Stack 6G Terahertz Networks," *IEEE Communications Magazine*, vol. 58, no. 11, 2020.

[6] D. Sabella, A. Alleman, E. Liao, M. Filippou, Z. Ding, L. G. Baltar, S. Srikanteswara, K. Bhuyan, O. Oyman, G. Schatzberg *et al.*, "Edge computing: from standard to actual infrastructure deployment and software development," *ETSI White paper*, 2019.

[7] F. Dressler, F. Klingler, M. Segata, and R. Lo Cigno, "Cooperative Driving and the Tactile Internet," *Proceedings of the IEEE*, vol. 107, pp. 436–446, Feb. 2019.

[8] B. Zhai, M. Yu, A. Tang, and X. Wang, "Mesh Architecture for Efficient Integrated Access and Backhaul Networking," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2020.

[9] M. Simsek, O. Orhan, M. Nassar, O. Elibol, and H. Nikopour, "IAB Topology Design: A Graph Embedding and Deep Reinforcement Learning Approach," *IEEE Communications Letters*, vol. 25, no. 2, pp. 489–493, Feb 2021.

[10] F. Chiti, R. Fantacci, F. Paganelli, and B. Picano, "Virtual functions placement with time constraints in fog computing: A matching theory perspective," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 980–989, 2019.

[11] T. R. Andel and A. Yasinsac, "On the credibility of manet simulations," *Computer*, vol. 39, no. 7, 2006.

[12] E. R. Cavalcanti, J. A. R. de Souza, M. A. Spohn, R. C. d. M. Gomes, and A. F. B. F. da Costa, "Vanets' research over the past decade: overview, credibility, and trends," *ACM SIGCOMM Computer Communication Review*, vol. 48, no. 2, pp. 31–39, 2018.

[13] K. Kvarfordt, "Planning for Closure of the Logan City/Cache County Landfill and Surrounding Landscape," *Master's Thesis*, May 2010.

[14] S. Mansfield, K. Veenstra, and K. Obraczka, "Terrainlos: An outdoor propagation model for realistic sensor network simulation," in *2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, London,UK, Sept. 2016.

[15] W. R. Franklin and C. K. Ray, "Higher isn't necessarily better: Visibility algorithms and experiments," in *Advances in GIS Research: Sixth International Symposium on Spatial Data Handling*. Taylor & Francis, 1994, pp. 751–770.

[16] A. Osterman and P. Ritosa, "Radio Propagation Calculation: A Technique Using 3D Fresnel Zones for Decimeter Radio Waves on Lidar Data," *IEEE Antennas and Propagation Magazine*, vol. 61, no. 6, pp. 31–43, Dec. 2019.

[17] F. Yaghoubi, M. Furdek, A. Rostami, P. Öhlén, and L. Wosinska, "Consistency-Aware Weather Disruption-Tolerant Routing in SDN-Based Wireless Mesh Networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 2, pp. 582–595, Jun. 2018.

[18] O. Taghizadeh, P. Sirvi, S. Narasimha, J. A. L. Calvo, and R. Mathar, "Environment-Aware Minimum-Cost Wireless Backhaul Network Planning With Full-Duplex Links," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2582–2593, Sep 2019.

[19] Y. Li, A. Cai, G. Qiao, L. Shi, S. K. Bose, and G. Shen, "Multi-Objective Topology Planning for Microwave-Based Wireless Backhaul Networks," *IEEE Access*, vol. 4, pp. 5742–5754, Jun. 2016.

[20] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Hybrid Radio/Free-Space Optical Design for Next Generation Backhaul Systems," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2563–2577, Jun. 2016.

[21] M. Polese, M. Giordani, A. Roy, D. Castor, and M. Zorzi, "Distributed path selection strategies for integrated access and backhaul at mmwaves," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018.

[22] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," *IEEE Communications Magazine (COMMAG)*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[23] C.-X. Wang, M. Di Renzo, S. Stańczak, S. Wang, and E. G. Larsson, "Artificial Intelligence Enabled Wireless Networking for 5G and Beyond: Recent Advances and Future Challenges," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 16–23, Feb. 2020.

[24] L. Maccari, G. Gemmi, R. Lo Cigno, M. Karaliopoulos, and L. Navarro, "Towards Scalable Community Networks Topologies," *Ad Hoc Networks*, vol. 94, pp. 1570–8705, Nov. 2019.

[25] ETSI, "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP 38.901 (V16.0.0), Tech. Rep., Oct 2019.

[26] P. Kyösti *et al.*, "IST-4-027756 WINNER II D1. 1.2 V1. 2 WINNER II Channel Models," *EBITG, TUI, UOULU, CU/CRC, NOKIA, Tech. Rep*, 2007.

[27] A. Al-Hourani, "On the Probability of Line-of-Sight in Urban Environments," *IEEE Wireless Communications Letters*, vol. 9, no. 8, Aug. 2020.

[28] J.-A. A. d'Auriac, N. Cohen, H. El Mafthoui, A. Harutyunyan, S. Legay, and Y. Manoussakis, "Connected tropical subgraphs in vertex-colored graphs," *Discrete Mathematics and Theoretical Computer Science*, vol. 17, no. 3, pp. 327–348, 2016.

[29] L. Maccari and R. Lo Cigno, "A week in the life of three large wireless community networks," *Ad Hoc Networks*, vol. 24, Part B, 2015.

[30] P. V. Trinh, A. T. Pham, A. Carrasco-Casado, and M. Toyoshima, "Quantum key distribution over fso: Current development and future perspectives," in *Progress in Electromagnetics Research Symposium*, Tokyo, Japan, Aug 2018.

[31] W. Ellens, F. Spieksma, V. P., A. Jamakovic, and R. Kooij, "Effective graph resistance," *Linear Algebra and its Applications*, vol. 435, no. 10, pp. 2491–2506, Nov 2011.

Gabriele Gemmi received a MSc in Computer Science from the University of Trento in 2020. He is now pursuing a PhD in Computer Science at the Ca' Foscari University of Venice in cotutelle with the Polytechnic University of Catalonia. His research interests are in wireless communication networks and distributed systems.

Renato Lo Cigno is Full Professor at the University of Brescia, Italy. He received a degree in Electronic Engineering with a specialization in Telecommunications from Politecnico di Torino; in 1998/9, he was Visiting Scholar at UCLA; from 2002 to 2019 he was with the University of Trento, Italy. He is currently Associate Editor for IEEE/ACM TNET and Editor for Elsevier Computer Communications. His current research interests are in performance evaluation of wired and wireless networks, modeling and simulation techniques, vehicular networks and networks beyond 5G, and networked systems in general. Renato Lo Cigno is senior member of IEEE and ACM and has co-authored around 200 papers in peer reviewed journals and conferences.

Leonardo Maccari Leonardo Maccari is an associate professor at the Ca' Foscari University of Venice, Italy. He received a Ph.D in the Faculty of Computer Science Engineering in the University of Florence in 2011 and he was a Marie Curie fellow and an Assistant Professor at the University of Trento (Italy) till 2019. He is an IEEE and ACM member, he co-authored more than 60 publications in refereed conferences, journals and book chapters. He is also among the authors of European and US patents. He was part of the TPC of relevant conferences such as IEEE INFOCOM, IEEE ICDCS, IFIP Networking, IEEE Globecom, IEEE ICC. Among his research interest there are network protocols for large-scale wireless mesh networks, Community Networks, interdisciplinary studies and network security.