NPART+:
Improving Wireless Network Topology Generators with Data from the Real World

Gabriele Gemmi
University of Trento
gabriele.gemmi@studenti.unitn.it

Leonardo Maccari
University of Venice
leonardo.maccari@unive.it

***Abstract*—Topology generators are a key asset for researchers in computer science and telecommunications that often need to test network protocols or distributed systems in simulated environments that resemble real scenarios. Despite that, in the research area of distributed wireless networks still many works use very simplistic models that do not have the characteristics of the currently existing large-scale wireless mesh networks. The only topology generator that tries to produce synthetic graphs that *look like* real networks is NPART [1].**

In this work we test the characteristics of NPART against another, completely different approach: `TrueNets` [2]. `TrueNets` uses accurate data representing land surface of a real world location to create topologies of networks *that could actually exist*. The downside of `TrueNets` is twofold: it can be used only when data-sets are available and generating topologies is computationally intensive. We show that using aggregate data from `TrueNets` we are able to improve NPART. We call the new generator NPART+ and we show that compared to topologies generated with `TrueNets`, NPART+ (or its variants) improves NPART in several metrics, but still it can not match the accuracy of `TrueNets`.

I. Introduction

The physical topology of a network strongly influences the performance of protocols and applications that run on it. In the field of telecommunications and networking it is essential to use the right topology to benchmark the performance of proposed protocols and applications but unfortunately, most of the times real topologies are not available. Even when they are, their number is limited and not sufficient to assess the performance of a new protocol in different but still realistic settings. For this reason in each specific area of research, topology generators with different properties have been created. In the area of wireless multi-hop networks, despite their importance, research on topology generators lags behind. Two models have been proposed, one by Cerdá-Alabert [3] and another (NPART, Node Placement Algorithm for Realistic Topologies) by Milic and Malek [1], both derived from the observation of existing networks. While the first one models networks with a mix of wireless and wired connections, the second one generates topologies with characteristics that have been measured on real-world wireless mesh networks made of hundreds of nodes.

NPART has two input parameters: a degree distribution to be matched and a target number of leaf nodes in the topology. These two parameters were derived by the analysis of two existing networks, and generated topologies try to reproduce them as closely as possible. This is one of the limitations of NPART, which is hard to generalize since we don't have degree distributions from many real networks, so the overall size of the network is limited to the ones we know. The second limitation is that NPART is essentially a random point placement process in a free space area that places links in order to match a degree distribution. While the input parameters influence the macroscopic features of the resulting graph, the local features may well be very different compared to real networks. Finally, NPART is based on the analysis of urban networks, that have characteristics that may strongly differ from networks in rural or suburban areas.

A recent work proposes a completely different approach [2]. `TrueNets` is a network emulator that takes advantage of morphological data provided by public administrations - such as elevation profiles obtained with LIDAR (Light Detection and Ranging) measurement campaigns - and building shapes obtained by OpenStreetMap. `TrueNets` is able to evaluate the presence of line-of-sight between two buildings in a city map and the networks generated with `TrueNets` are extremely realistic both in their macroscopic and local properties. The obvious drawback is that to be generated they require precise information on a specific world location and that generating topologies is a computationally intensive process.

The goal of this paper is to understand how we can improve NPART with data coming from `TrueNets` to make NPART more scalable and generic. Ideally, we would like to have pre-computed empirical distributions coming from `TrueNets` that represent generic categories of inhabited places (urban, rural, intermediate and suburban) and that can be used to tailor the behaviour of NPART. As a first step in this direction, we take data coming from 4 real areas, we generate networks with `TrueNets` and we extract the data needed to synthetically generate topologies using NPART. We call this hybrid approach NPART+. Then, we compare the topologies generated by `TrueNets`, NPART and NPART+ and we measure their similarity on the basis of different graph metrics. Our results show that NPART+ produces topologies with macroscopic features that are closer to the realistic ones compared to the original NPART. NPART+ (with two variants we describe later on) is a valuable, generic instrument to research on wireless multi-hop networks. Yet, at the local level there are still significant differences between the topologies generated by `TrueNets` and NPART+, which makes `TrueNets` approach still unique.

Finally we note that even if our work focuses on wireless mesh network, the concepts at the base of a multi-hop mesh network are present also in other emerging fields. As an example, wireless backhauling for 5G wireless networks coupled with software defined networking and network function virtualization is a topic that is raising interest [4], and the same concept using Free Space Optical communications is emerging [5]. In both these examples, realistic topologies to study the system performance are essential, and thus, we believe our results are of generic interest beyond mesh networks.

II. Topology Generators

In this section we summarize the main features of NPART and `TrueNets`, the two topology generators for wireless mesh networks that are at the basis or our proposal.

A. NPART: Node Placement Algorithm for Realistic Topologies

NPART was developed to provide a flexible tool for the generation of mesh network topologies [1]. The main characteristics of NPART are the following ones:

- **Realism:** The generator takes in input parameters measured on real networks and creates networks topologies with properties similar to the original ones.
- **Randomness:** The generator is capable to create new, random networks while preserving the properties of realism.

NPART is a point placement process, i.e., an algorithm that randomly places points in a 2 dimension space and connects them to generate a network topology. This kind of algorithms has been extensively used to simulate mesh networks, ad hoc networks and sensor networks. The random function that places nodes in the space and the function that places links between nodes determines the properties of the graph. In the most simple case, nodes are placed at random in a square and links connect any couple of nodes whose distance is lower than a certain threshold.

NPART is based on the observation of the growth of real networks in which a new node is more likely to join the network if it's in an area that is already populated by other nodes. This is implemented by representing the network on a 2 dimension space that extends with the number of nodes, so the density of nodes per area unit does not necessarily increases with the number of nodes. Once a new candidate node is generated it needs a point of attachment, a node that is already part of the network to create a link with. The point of attachment is not simply any node in the communication range but is chosen in order to maximize a fitness metric. This metric is made in such a way that the degree distribution of the network approaches an user provided empirical degree distribution. In the original work the degree distribution is taken from the observation of two existing networks made of hundreds of nodes.

Even though this is probably the most accurate approach to generate urban mesh networks, it has three main issues. The first one is that the parameters to generate the topologies must be generic enough to avoid over-fitting the original network, however there are not enough networks to have a set of realistic degree sequence to choose from. The second one, which is also analyzed in the original paper, deals with correlated shadowing. Without shadowing, the feasibility of links does not depend on their length, i.e., as long as a link is feasible the distribution of link length is uniform. When shadowing enters into the model instead, a longer link has less probability of being realized, because the probability of having an obstacle between the endpoints increases with their distance. Furthermore correlated shadowing introduces a correlation between the feasibility of links based on the position of nodes. If node j has line of sight with node i, then the probability of node k of having line of sight with node i can be parameterized on the distance from j. Modeling correlated shadowing is essential for cellular networks that have a high density of terminals per squared meter [6], but it is intrinsically limited by the availability of measures in real world scenarios.

The lack of correlated shadowing emerges while analyzing the number of bridges and articulation points in the generated graphs. The authors of NPART observe that the generated graphs show a number of leaf nodes that is larger than what they measure on real networks and heuristically address this issue by introducing an optional pruning of the 20% of the leaves. With this approach the articulation point density is more accurate at the expenses of the degree distribution, and again, a new parameter derived from the observed networks is introduced.

The third main issue with NPART is the fact that modeling networks using only data coming from existing networks intrinsically limits the model to what was previously observed. The data on available mesh networks are generally coming from observation of networks in urban areas [7], [8], [9], [10], [11], and we expect these data to change dramatically if measured on suburban or rural areas.

B. TrueNets

`TrueNets` is a topology generator that models the growth of a wireless mesh networks in a given, real environment [2]. By taking advantage of very precise (down to one point per squared meter) morphological models derived from LIDAR measurement campaigns it is able to predict with high precision whether between two given points in space there's Line of Sight (LoS). Considering that most mesh networks use ISM bands (around 5GHz or 2.6GHz that need LoS to communicate) `TrueNets` naturally overcomes the correlated shadowing problem and creates realistic topologies. Indeed, `TrueNets` does not generate synthetic network topologies, it generates topologies of networks *that could actually exist*.

`TrueNets` is made of three main components:

1) A PostGIS database
2) libterrain
3) A network growth strategy

The PostGIS database contains the LIDAR measurements of the interested areas represented as raster maps with a precision of one point per squared meter, and the vectorial data of all the buildings in the area.

These data are used by libterrain to emulate the feasibility of a wireless link between two given buildings. In the case the link is feasible libterrain also gives a rough estimation of the signal loss due to obstacles or refraction using a single knife edge approximation [12]. Figure 1 shows a projection in the z-plane of a wireless link with an estimation of LoS and of the Fresnel zone.

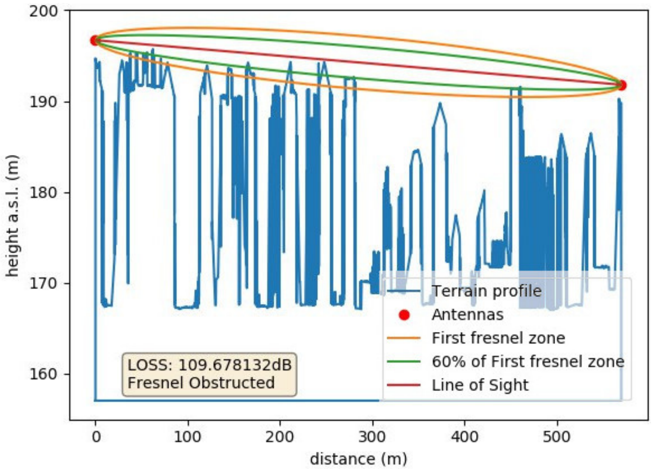

Fig. 1. Graphic representation of the elevation profile with LOS and Fresnel area.

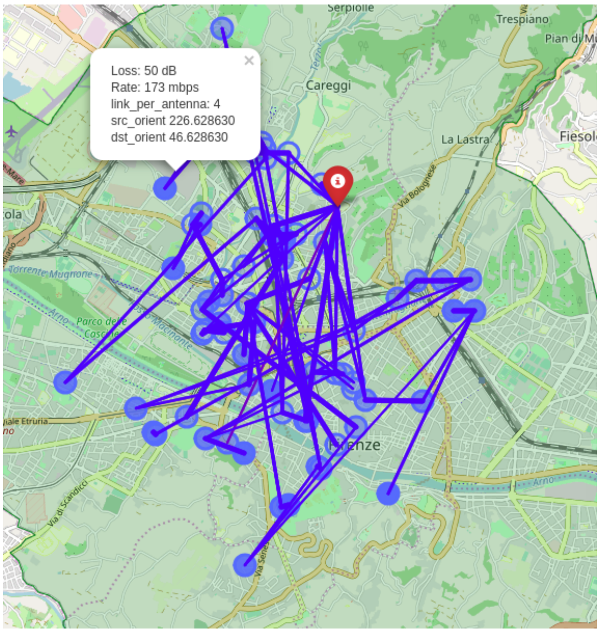

Fig. 2. A network topology generated in the area of Florence, Italy.

Once a LoS estimation is available together with a map of a certain area, it is possible to feed this information to a network growth strategy in order to emulate the growth of a mesh network in a given environment. Figure 2 represents a topology generated using data coming from the city of Florence, Italy.

The topologies generated by `TrueNets` are fully realistic, realizable in the real world, but the drawback of `TrueNets` is two face First, the generation process is resource greedy. In fact, depending on the area of interest, the database containing the LIDAR measurements can take up to hundreds of gigabytes. Moreover, due to the heavy process of checking the LoS between two buildings, the generation of each topology can take hours (currently 50 minutes to generate a 500 nodes topology using 64 cores). Second, even if `TrueNets` generated topologies do not fit a single network, they fit a specific area, and can be not generated for areas for which data is not available.

III. NPART+

The first goal of this work was to generalize NPART to an arbitrary network size removing the dependency on the degree distribution observed in the original paper and replacing it with synthetic degree distributions. In the literature there are observations that suggest that degree distributions of mesh networks may fit a power law distribution with $\alpha = 1.55$ [13]. We heuristically truncate it to a maximum value, as real world networks can not have arbitrarily high degree and we obtain a scalable generator. We refer to this approach as NPART in the rest of the paper. When we apply the suggested pruning of 20% of the leaf nodes instead, we refer to the approach as NPART-LP (NPART with **L**eaf **P**runing). These two strategies represent the base line of our comparisons.

To create NPART+ we need to derive empirical degree distributions using `TrueNets`, in order to feed them to NPART. In this work we set a target network size of 500 nodes, but the same process can be repeated for any arbitrary network size.

A. *Generating Topologies with* `TrueNets`

We use `TrueNets` to create topologies using data from the region of Tuscany (Italy), in which we selected 4 areas that can be classified as urban, suburban, intermediate and rural and we use a growth strategy that we briefly review in the rest of this section (see the original work for details on the data and the process [2]).

The strategy starts by picking one of the different buildings we have chosen before-hand for each specific area, which will be called seed. The seed is chosen manually in order to avoid pathological cases of buildings with zero visibility. Then, in a given radius around the seed a random building is chosen and its LoS to the seed is verified, if it's not connectable another one is chosen until a connectable one is found. Every time a new building is added to the network the radius is recomputed around the area occupied by the network and new buildings become eligible to be connected to the network. When in the network there is more than one node, the LoS of the new buildings is tested against all the nodes in the network, and the one that offers the best path loss is chosen. In order to have realistic conditions, an upper bound of 4 devices per building has been set. This aims to model the fact that most wireless mesh nodes are operated on private building's roofs that cannot house large trails to host more than 4 devices. Note that each device is the faithful description of existing devices with their antenna aperture, and that each device can be used to create more than one link. Finally, when the network reaches the size of 500 nodes its growth stops and the obtained topology is saved on disk. Using this setup, 55 different topologies have been generated for 4 areas, resulting in a total of 220 topologies.

From these topologies we derive empirical distribution for degree and edge-to-node ratios as follows. Let E be an element in the set of areas:

$$E \in \{\text{URBAN}, \text{SUBURBAN}, \text{INTERMEDIATE}, \text{RURAL}\}$$

For each area in E we create a set of undirected graphs G_E, we call the union of all sets $\mathcal{G}$. Let d_G be the empirical degree

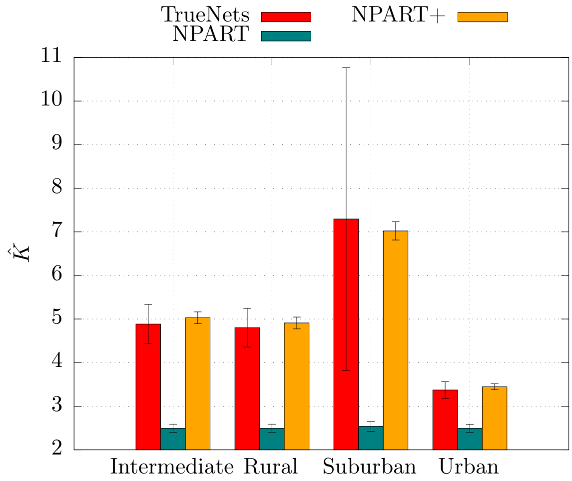

Fig. 3. Vertex to leaves ratio

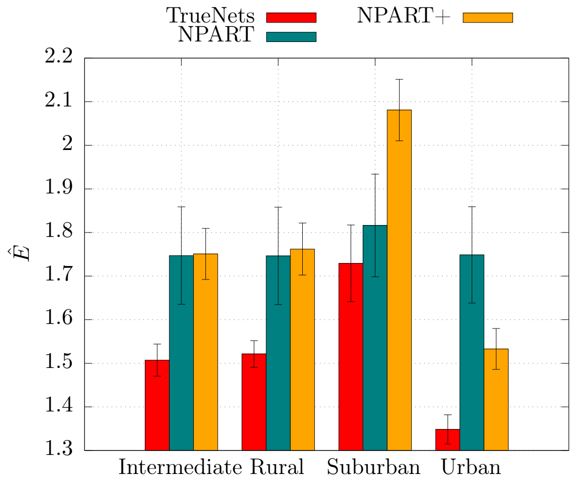

Fig. 4. Edge to vertex ratio

distribution of the graph G. In practice d is a vector of size equal to the maximum degree of the graph and $d(i)$ is the number of nodes with degree i. Then we can define $\hat{d}_E$ as the vector mean of these empirical distribution among an area

$$\hat{d}_E = \frac{1}{|G_E|} \sum_{G \in G_E} d_G$$

We use $\hat{d}_E$ as input to the NPART algorithm, and we call this new approach NPART+. Figures 3 and 4 show some initial comparison of the graphs generated with the three mentioned approach in the four areas. We note that the NPART+ matches quite precisely the ratio of leaves to edges, which is a natural consequence of using the degree distribution from `TrueNets`, but still fails to capture the measured edges to vertex ratio.

To improve this metric we call $\hat{r}_E$ the scalar mean of the edge-to-vertex ratio computed on all the graphs in a certain area

$$\hat{r}_E = \frac{1}{|G_E|} \sum_{G \in G_E} \frac{|E_G|}{|V_G|}$$

We modified NPART to accept the additional parameter $\hat{r}_E$ as follows. At the end of the generation of the network the process iteratively removes non-bridge edges from the network (i.e., edges that do not partition the network), until the target ratio between vertexes and edges is reached. We name the version of NPART using $\hat{d}_E$ and $\hat{r}_E$ NPART+EP (NPART+ with **E**dge **P**runing).

IV. Evaluation Metrics

In order to evaluate the differences between the generated graphs we generated 550 topologies for each area, for each approach (NPART, NPART-LP, NPART+ and NPART+EP). We use the graphs generated with `TrueNets` as the ground truth and we compare the results obtained with the other approaches to assess their similarity. For this purpose we introduce several graph metrics, which we divide in two groups. In the first group we include metrics that address general features of graphs, in the second group we include metrics that have a direct impact on the network performance. All the metrics we review are adaptations of known graph evaluation metrics to our specific context. In the rest of the paper we refer to a graph G with vertices in V_G and edges in E_G. For the sake of readability we use the $|\cdot|$ operator to express a set size but also, with a small abuse of notation, the number of nodes when applied to a graph, so that $|V_G| = |G|$.

A. Analytical Metrics

a) Vertex to leaves ratio: Let $L_G \subseteq V_G$ be the set of vertexes with degree 1, then:

$$\hat{K} = \frac{1}{|G_E|} \sum_{G \in G_E} \frac{|V_G|}{|L_G|}$$

b) Edge to vertex ratio: Edge to vertex ratio, or graph density, is simply the ratio between the number of edges and vertices in G

$$\hat{E} = \frac{1}{|G_E|} \sum_{G \in G_E} \frac{|E_G|}{|V_G|}$$

c) Vertex to articulation-point ratio: An articulation point is a node that, if removed, partitions the network in two disconnected component. Let $AP_G \subseteq V_G$ be the set of articulation points, then:

$$\hat{W} = \frac{1}{|G_E|} \sum_{G \in G_E} \frac{|V_G|}{|AP_G|}$$

d) Size of the Largest Bi-Connected Component (BCC): A bi-connected component of a graph G is a sub graph that does not contain articulation points. If G is bi-connected the network is robust to the failure of one edge. In general, the size of the largest BCC of a graph is a rough measure of connectivity of a graph. Assuming $bcc(G)$ returns the list of the sizes of all BCCs in a graph G then:

$$\hat{L} = \frac{1}{|G_E|} \sum_{G \in G_E} |max(bcc(G))|$$

e) Average clustering coefficent: The clustering coefficient is a metric of local density of G. Let e_{ij} be the undirected edge that connects the vertex v_i with the vertex v_j and N_{v_i} the set of neighbors of v_i

$$N_{v_i} = \{v_j : \langle e_{ji} \rangle \in E\}$$

Let the local clustering coefficient for an undirected graph G and a vertex v_i be the number of closed triangles incident on a vertex v_i over the number of all the possible closed triangles

$$C(G, v_i) = \frac{2 \cdot |\{e_{jk} : v_j \in N_{v_i}, v_k \in N_{v_i}, e_{jk} \in E\}|}{|N_{v_i}|(|N_{v_i}| - 1)}$$

Then the Average clustering coefficient **C** is defined as

$$C(G) = \frac{1}{|V_G|} \sum_{v \in V_G} C(G, v)$$

and the average over all the graph of an area E is:

$$\hat{C} = \frac{1}{|G_E|} \sum_{G \in G_E} C(G)$$

B. Impact Metrics

a) Robustness to Random Failures: One way of estimating the robustness of a graph is to remove a random set of vertices and observe if the remaining part of the graph is still connected. Ideally, a robust graph is a graph for which we can remove some vertices, and the rest of the vertices can still communicate one with the other. In practice if some of the removed vertices are articulation points, then removing them makes the graph disconnected. The largest is the fraction of vertices we remove, the higher is the chance of removing some articulation points. Note however that not all articulation points are equal: removing an articulation point that divides the network in two disconnected sub graphs of equal size has an impact that is much higher than removing a cut point that isolates only one node. Thus, the number of cut-points is not sufficient to estimate robustness and we need to introduce a better metric.

Let $\overset{v}{-} V : G'$ be the operator which yields the graph G' by removing from G the vertexes from set V, $LC(G)$ be the largest connected component of G and K_r be a random ordering (with seed r) of the graph vertexes V_G. We split K_r in 100 disjoint subsets of size $\delta = \frac{|V_g|}{100}$ and we call $S_{j,r}$ the $j$esim subset.

Let $C_{i,r}$ be the union of such subsets up to index i, with $C_{i,r} \subset C_{i+1,r}$ defined as:

$$C_{i,r} = \bigcup_{j=0}^{i} S_{j,r}$$

and let $G'_{i,r}$ be the graph obtained from G by removing the set of vertex contained in the set $C_{i,r}$

$$G'_{i,r} = G \overset{v}{-} C_{i,r}.$$

Let $P_{i,r}(G)$ be the robustness metric defined as:

$$P_{i,r}(G) = \frac{|LC(G'_{i,r})|}{|G'_{i,r}|}$$

$P_{i,r}(G)$ is a robustness metric that must be averaged using a sufficient number of seeds. If we use 100 random seeds to generate 100 different orderings, we can compute the average $P_i(G)$ on a set of 100 random seeds as:

$$P_i(G) = \frac{1}{100} \sum_{r=0}^{99} P_{i,r}$$

As said, ideally, we would like $P_i(G)$ to be always 1, meaning that in all 100 attempts of removing the set of C_i nodes from G, the remaining part of G is connected. In general this does not happen, as $P_i(G)$ will gradually decrease before we remove all nodes from G (when $i = 99$, G is empty and $P_i(G)$ is arbitrarily set to 0). Yet, the decreasing trend of $P_i(G)$ is an estimation of the robustness of the graph, and we can express it as the area subtended by the line connecting the points $P_i(G)$. The robustness of graph G to the random failure of $\delta \times i$ nodes can then be expressed as $R_i(G)$:

$$R_i(G) = \int_{j=0}^{i} P_j(G) = \sum_{j=0}^{i} \frac{P_j(G) + P_{j-1}(G)}{2}$$

where we applied the trapezoidal rule to the integral. $R_{100}(G)$ is a metric to evaluate the robustness of G, and we can apply it to all the graphs generated with a given approach, that is, given a set of graphs $\mathcal{G}$ with $G \in \mathcal{G}$ then:

$$R_i = \frac{1}{|\mathcal{G}|} \sum_{G \in \mathcal{G}} R_i(G); \ \ P_i = \frac{1}{|\mathcal{G}|} \sum_{G \in \mathcal{G}} P_i(G)$$

To help the reader to visualize this metric we report in Figure 7 an example plot of $P_i(G)$ for one random graph G generated with `TrueNets` (the red curve). As a comparison, we generated a set $\mathcal{G}$ of 50 graphs using NPART, NPART+, NPART+EP (the latter two use d extracted from G). We report in the graph the curve P_i with an envelope that represents the standard deviation around the average on all graphs in $\mathcal{G}$. The value of R_i for each version of the generator is the area subtended by the related curve. We will comment these results in the next section.

b) Robustness to Targeted Attacks: While R expresses the robustness to (correlated) random failures, another interesting robustness metric is the one that expresses robustness to targeted attacks, in which an attacker is able to choose the removed nodes with the goal of maximizing the potential damage.

One convenient approach for the attacker is to target nodes that maximize some centrality metric, for instance betweenness centrality. Let $BC(G)$ be the list of vertices in G ordered by their centrality. Again we split $BC(G)$ in 100 disjoint sets of size $\delta = \frac{|V_g|}{100}$, we are interested only in the first set, named $S_0(G)$.

We generate a list of graphs G'_i with a recursive function defined as:

$$G'_0 = G \overset{v}{-} S_0(G)$$

$$G'_i = G'_{i-1} \overset{v}{-} S_0(G'_{i-1})$$

Similarly to the random case we can define $P_i^T(G)$ as the robustness of G against targeted attacks, as:

$$P_i^T(G) = \frac{|LC(G'_i)|}{|G'_i|}$$

And again derive

$$R_i^T(G) = \sum_{j=0}^{i} \frac{P_j^T(G) + P_{j-1}^T(G)}{2};$$

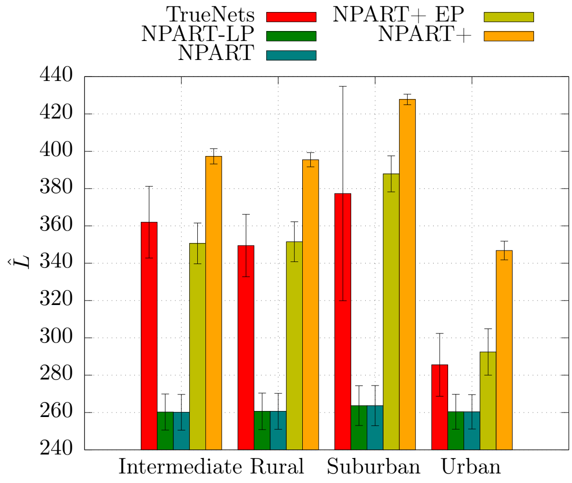

Fig. 5. Size of the Largest BCC.

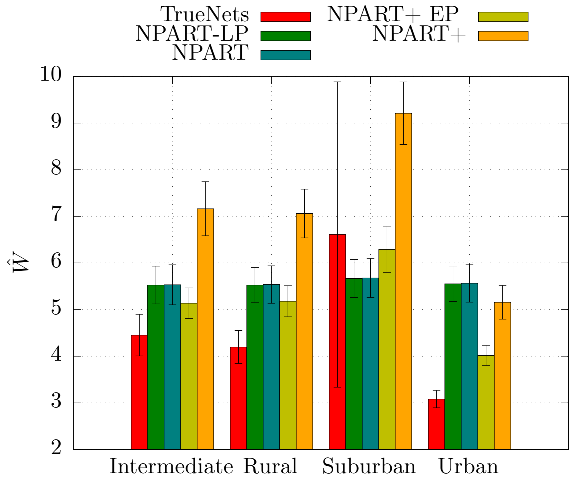

Fig. 6. Vertex to Articulation Point Ratio.

$$R_i^T = \frac{1}{|\mathcal{G}|}\sum_{G\in\mathcal{G}} R_i^T(G);\ \ P_i^T = \frac{1}{|\mathcal{G}|}\sum_{G\in\mathcal{G}} P_i^T(G)$$

Again fig. 10 reports an example of the $P_i^T(G)$ and P_i^T for a random graph G generated with `TrueNets` and a set of random graphs generated using NPART versions.

c) Average shortest path: The average shortest path is computed on G using Dijkstra's algorithm, referred to as $d(v_1, v_2)$

$$\hat{D} = \frac{1}{|G_E|}\sum_{G\in G_E}\frac{1}{|V_G|(|V_G|\text{-}1)}\sum_{\substack{v_1,v_2\in V_g\\ v_1\neq v_2}} d(v_1, v_2)$$

V. Results

We divide the discussion of results in two sections, one dealing with density measures and robustness, and another dealing with clustering coefficient and path length.

A. Density and Robustness

Figure 5 reports the histogram of the size of the largest bi connected component for all the approaches in all the areas (with standard deviation). NPART/NPART-LP always underestimate the size of the largest BCC, so there are more nodes that could be detached by the network breaking only one node. NPART+ tends to slightly overestimate this metric while NPART+EP is the one that seems to best fit `TrueNets`.

Figure 6 reports the density of articulation points, which shows that `TrueNets` has a higher average number of articulation points compared to the ones measured with all variants of NPART. As an exception, the suburban area behaves differently, with `TrueNets` having less articulation points than the others (excluding NPART+). In this case we must note that the suburban area has a much larger standard deviation, due to the fact that the geographical area is large and sparsely inhabited, so that relatively small graphs can be quite different one from the other. Again we can observe NPART+EP is the approach that most resembles `TrueNets`.

These metrics have a direct impact on the graph robustness: looking at Figure 7 we see that the robustness to random failures is lower for `TrueNets` than for other approaches. This

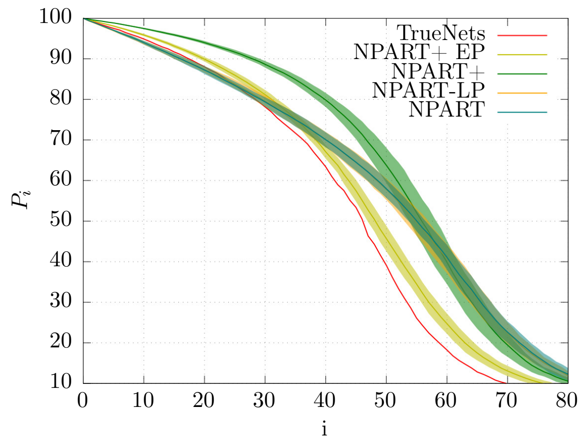

Fig. 7. Robustness to Random Failures (average on all areas): P_i.

single observation is generalized by the value of R_{100} in fig. 8 which confirms a lower robustness, due to the higher number of articulation points. NPART+ is the approach that shows the largest difference from `TrueNets`. Once more NPART+EP is the approach that better resembles `TrueNets`.

The robustness against targeted attacks has a pretty different trend, as figs. 9 and 10 show (note that we use R_{30} instead

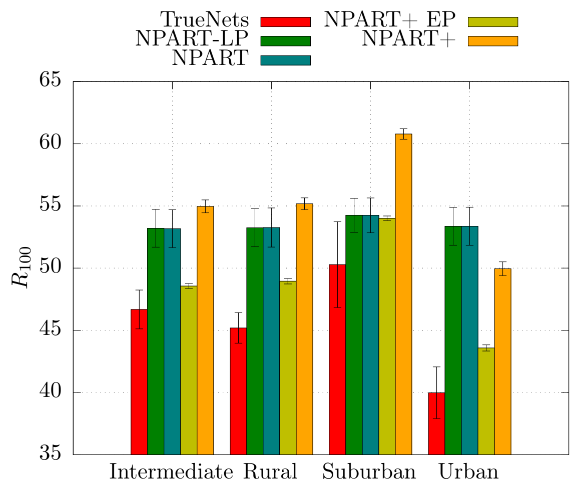

Fig. 8. Robustness to Random Failures: R_{100}.

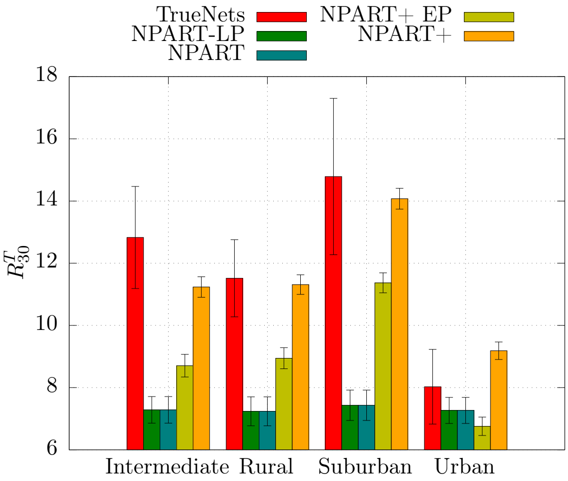

Fig. 9. Robustness to Targeted Attacks: R_{30}^{T}.

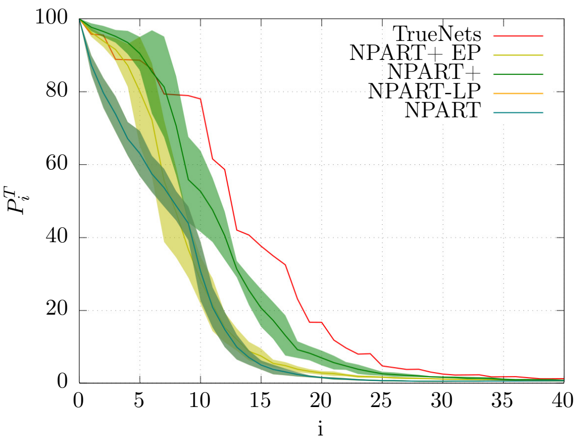

Fig. 10. Robustness to Targeted Attacks (average on all areas): P_i^T.

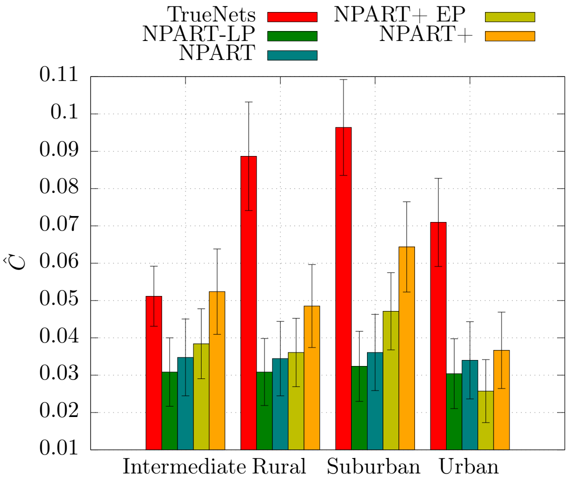

Fig. 11. Clustering coefficient

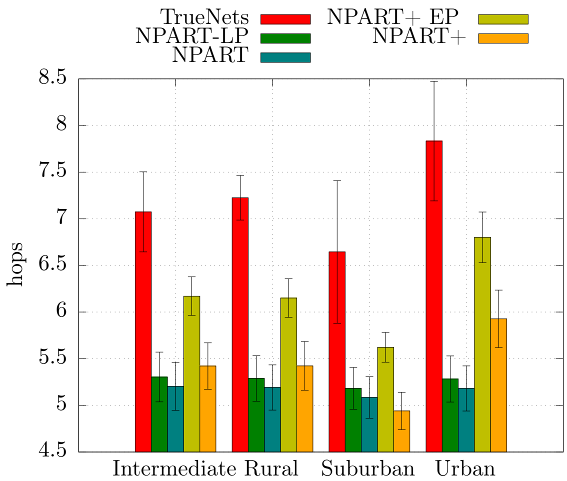

Fig. 12. Average shortest path

of R_{100} as all curves quickly approach zero). In this case in average `TrueNets` graphs seem to be more resistant to failures. This is not surprising as such difference is an effect known in the literature for other kinds of graphs [14]. In general, robustness to random failures depends on the average characteristics of the graph, robustness to targeted attack depends on the presence of a few extremely important nodes that if removed make the whole structure quickly collapse. It is interesting to note how under this metric NPART+ has closer performance to `TrueNets` compared to NPART+EP.

B. Clustering and Path Length

Figure 11 shows the average clustering coefficient generated by all the approaches and shows that in almost all areas there is a relevant difference between graphs generated by `TrueNets` and all the others. We believe this is related to correlated shadowing, which makes short links more likely than long links. Short links connect nodes "*near by*" and thus, there is a higher probability of being connected to nodes that are physically close. If the neighbors of node i are physically close to i, they are most likely also physically close, and thus, the probability of being neighbors is also high.

A prevalence of short links has an impact on the average path length, which is always larger for `TrueNets` than for any other approach, with NPART+EP that is the closest of the other options (see fig. 12). Note that NPART+EP has the same average edges to vertex ratio by design, and it is straightforward to note that graph density has a strong impact on the average path length. Yet, even if in this sense the macroscopic features of the graph are similar, the local properties of `TrueNets` make the average path length always at least one hop longer.

VI. Conclusions

Recent advances in wireless networks lead to a future in which the wireless media will be used not only for the connection to the user terminal, but also in the backhaul network. 5G is one key example as it will create complex network topologies to dynamically support network slicing applications. The need for realistic topologies to test protocols and applications is going to raise since many technologies (like network function chaining and network embedding) are based on heuristics that may have extremely different performance on different topologies.

In this paper we merged two approaches for the generation of synthetic graphs, one that can create extremely realistic topologies but requires the availability of real world data, another one that is very simple to implement but with results that are hard to generalize. We analyzed the produced graphs using several different metrics, and our results can be summarized in three points. A first important one is that realistic topologies from `TrueNets` show that graphs in different geographical areas are actually very different. All the metrics oscillate when passing from urban to rural areas in a way that is hard to predict, and thus, we can conclude that there is no *one size fits all* solution. Researchers need to test their technologies on graphs that represent different areas or they risk to over fit one area and disregard the characteristics of others. The second point, which is consequential, is that NPART+ variants have performances that can be close to `TrueNets` in some aspects but no single version can fit realistic topologies in all aspects. For instance, NPART+EP seems to be the one that performs better in the majority of the cases with the exception of the robustness to targeted attacks, in which NPART+ is closer to `TrueNets`. The original versions of NPART performs worse than NPART+ in all the cases. Researchers should test their technologies using different generators depending on the specific properties that they want to analyze. Finally, the problem of correlated shadowing still persists and consistently generates differences in the clustering coefficient of networks. While correlated shadowing has been addressed to analyze interference at the receiver in cellular networks, little has been done to model backhaul wireless networks, we need more research to generate models based on stochastic geometry that can fill this gap.

REFERENCES

[1] B. Milic and M. Malek, "NPART - node placement algorithm for realistic topologies in wireless multihop network simulation," in *International Conference on Simulation Tools and Techniques (SIMUTools)*, 2009.

[2] L. Maccari, G. Gemmi, R. Lo Cigno, M. Karaliopoulos, and L. Navarro, "Towards scalable community networks topologies," *Ad Hoc Networks*, vol. 94, 2019.

[3] L. Cerdà-Alabern, "On the topology characterization of guifi.net," in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2012.

[4] R. Riggio, A. Bradai, D. Harutyunyan, T. Rasheed, and T. Ahmed, "Scheduling wireless virtual networks functions," *IEEE Transactions on Network and Service Management*, vol. 13, no. 2, 2016.

[5] 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, 2016.

[6] J. Lee and F. Baccelli, "On the effect of shadowing correlation on wireless network performance," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2018.

[7] L. Cerdà-Alabern, A. Neumann, and P. Escrich, "Experimental evaluation of a wireless community mesh network," in *ACM International Conference on Modeling, Analysis & Simulation of Wireless and Mobile Systems (MSWiM)*, 2013.

[8] K. La Curts and H. Balakrishnan, "Measurement and analysis of real-world 802.11 mesh networks," in *ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2010.

[9] V. Brik, S. Rayanchu, S. Saha, S. Sen, V. Shrivastava, and S. Banerjee, "A measurement study of a commercial-grade urban wifi mesh," in *ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2008.

[10] L. Cerdá-Alabern, A. Neumann, and L. Maccari, "Experimental evaluation of BMX6 routing metrics in a 802.11n wireless-community mesh network," in *International Workshop on Community Networks and Bottom-up-Broadband (CNBuB)*, 2015.

[11] 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.

[12] C. Sommer, S. Joerer, M. Segata, O. Tonguz, R. Lo Cigno, and F. Dressler, "How shadowing hurts vehicular communications and how dynamic beaconing can help," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2013.

[13] L. Maccari, "On the technical and social structure of community networks," in *IFIP Networking Conference and Workshops*, 2016.

[14] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, 2000.