A Methodology to Select Topology Generators for Ad Hoc Mesh Network Simulations

Michael O'Sullivan*, Leonardo Aniello†, Vladimiro Sassone‡
Electronics and Computer Science, University of Southampton, Southampton, United Kingdom
Email: *M.O'Sullivan@soton.ac.uk, †L.Aniello@soton.ac.uk, ‡vsassone@soton.ac.uk

Abstract—Many academic and industrial research working on Wireless Communications and Networking rely on simulations, at least in the first stages, to obtain preliminary results to be subsequently validated in real settings. Topology generators (TG) are commonly used to generate the initial placement of nodes in artificial Ad Hoc Mesh Network topologies, where those simulations take place. The significance of these experiments heavily depends on the representativeness of artificial topologies. Indeed, if they were not drawn fairly, obtained results would apply only to a subset of possible configurations, hence they would lack of the appropriate generality required to port them to the real world. Although using many TGs could mitigate this issue by generating topologies in several different ways, which would entail a significant additional effort. Hence, the problem arises of what TGs to choose, among a number of available generators, to maximise the representativeness of generated topologies and reduce the number of TGs to use.

In this paper, we address this problem by investigating the presence of bias in the initial placement of nodes in artificial Ad Hoc Mesh Network topologies produced by different TGs. Indeed, any bias introduced by some TGs would affect the representativeness of generated topologies. We propose a methodology to quantify this bias and select what TGs to employ to minimise it, given the number of TGs to use. Our methodology relies on the extraction of a number of features from generated topologies. We carry out experiments on three well-known TGs, namely BRITE, NPART and GT-ITM. We find that selecting NPART (if one TG needs to be chosen) or BRITE and NPART (if two TGs need to be chosen) allows to minimise the bias index. The impact of our research lies in the provisioning of an empirical methodology to select the TGs to use for Ad Hoc Mesh Network simulations, which is expected to be beneficial for researchers involved in experimental evaluations on this type of networks.

Index Terms—Topology generator, Ad Hoc Mesh Network, BRITE, NPART, GT-ITM

I. INTRODUCTION

An *Ad Hoc Mesh Network* is based on a decentralised topology of devices/nodes that cooperate to implement some routing protocol, i.e. each device forwards its own and other devices' traffic according to a specific algorithm with the aim of reaching the target destination. Ad Hoc Mesh Networks do not rely on any fixed infrastructure and each node can only communicate with those other nodes lying within the transmission range of one another. Ad Hoc Mesh Network applications are wide and significant, ranging from wireless sensor networks to vehicular ad hoc networks (VANETs) to mobile ad hoc networks (MANETs), and they are used in everyday scenarios as well as more critical settings, such as military operations.

Several Ad Hoc Mesh Network aspects are still being investigated by the research community, e.g. routing protocols [2][3] and security [16][17]. For convenience, many academic works heavily rely on simulation to test a proposed solution and obtain preliminary results that are used to validate its effectiveness. Network simulations commonly entail evaluating a given approach on many different Ad Hoc Mesh Network topologies to ensure results are meaningful, i.e. to have evidence that they can apply to a wide variety of networks and are not tied to particular network configurations. Hence, as also suggested by Gunes et al. [4], a key aspect in any network¨ protocol simulation is the design and selection of what test network topologies to consider.

Network *topology generators* (TGs) are usually employed to create a possibly large number of topologies, on the basis of predefined network models, real-world measurements and additional parameters available to tune the generation process. Although any TG is designed and implemented to generate a representative set of topologies, different TGs do not rely on the same models and assumptions, do not follow the same generation approach and thus are likely to produce diverse topologies, which in turn can lead to obtain dissimilar simulation results [5] [9], Hence, we claim that the choice of the TG can affect this type of experiments, i.e. a TG is likely to introduce *bias* in simulations. This holds true for Ad Hoc Mesh Network simulations as well, where TGs are used to generate the initial placement of nodes, which in turn plays an important role in the way an Ad Hoc Mesh Network evolves over time.

Despite the fact that each TG has its own peculiarities, and that sometimes researchers can select a TG on the basis of the specific mathematical or physical model they need, there are in general several TGs that can be used to create artificial topologies representing the initial placement of nodes in Ad Hoc Mesh Networks. In this context, the best option would be to use all the available TGs to run simulations on the largest possible range of topologies, so as to ensure that obtained results are not biased by the choice of a specific TG, or subset of TGs. On the other hand, using many TGs proves to be really demanding for researchers in terms of required time and effort to delve into the technical issues of each TG. Therefore, a trade-off arises between reducing the effort to spend in setting up the simulations, i.e. minimising how many TGs to use, and maximising the representativeness of the simulations themselves, i.e. *minimising the bias introduced by TG selection*. In this paper, we delve into the initial placement of the nodes by the topology generators and the analysis of the differences between these topologies generated by distinct TGs to help researchers to reduce how many TGs to use while still preserving the representativeness of generated topologies. In particular, *given a fixed number of available TGs*, we address the following research questions.

- RQ1: How to measure the difference between topologies generated by distinct TGs? i.e. how to characterise the

bias introduced by the choice of a specific TG rather than using all the TGs?
- RQ2: how to choose what TG, or TGs, to use to reduce such a bias?

The approach we propose relies on a compact, numeric representation of topologies, based on a number of aspects about how network nodes are placed over the plane (e.g. inter-node distance, clustering) and about how Ad Hoc Mesh Networks work (e.g. nodes can only communicate with other nodes within their transmission range). Each topology is modelled as a vector of numeric features, which enables to compute distance metrics. We consider a fixed number of TGs and propose to interpret the bias as a measure of the differences that arise in generated Ad Hoc Mesh Network topologies when selecting any single TG, or subset of TGS, instead of picking all the available TGs.

We define the *bias index* as the combination of a number of feature regarding node placement, i.e. Inter-node distance, Spatial distribution, Node density, Shared Neighbours Distribution and Clustering coefficient. The bias index considers all these features for the topologies generated by a given TG, and measures the distance of these topologies from the general population, i.e. from all the topologies generated by all the TGs selected for the experiments. We only consider the initial stage of any experiment, before any data is transmitted or node moved.

We tackle RQ1 by focusing on two complementary facets of the distances between topologies. On the one hand, we want to *quantify* the bias by measuring the average distance between topologies generated by distinct TGs. In the specific, we use *Hedges' g* measure of effect size to compute the *bias index*, which measures the *difference between topologies produced by a specific TGs, or subset of TGs, and those created by all the available TGs*. On the other hand, we are also interested in evaluating to what extent existing differences are distinguishing of some TG, i.e. whether such differences allow to determine which TG generated a topology, regardless of the extent of those differences. In this regard, we employ machine learning techniques to compute the *classification accuracy*, i.e. to estimate how precisely we can discover which TG generated a topology.

We answer RQ2 by proposing a simple methodology, based on the bias index, to select what TGs to use to reduce the bias, depending on how many TGs can be picked at most.

We carry out an experimental evaluation using three well-known TGs, i.e. BRITE, NPART and TG-ITM. Obtained results show that using a single TG is likely to introduce bias, and that in this case picking NPART is the best choice to mitigate this issue. If two TGs can be used, BRITE and NPART provide the lowest bias. The experiments on the classification accuracy show that topologies can be correctly classified according to their TGs with high accuracy, i.e. up to almost 78%, and that, in this specific case, four topology features contribute most to distinguishing between different TGs.

To the best of our knowledge, this is the first work in literature that systematically investigates the differences between topologies generated by diverse TGs in the context of Ad Hoc Mesh Network simulation. The contributions of this work are

1) the definition of a *vector-based representation of Ad Hoc Mesh Network topologies*, based on a number of features derived from different aspects of node placement;
2) the definition of a novel metric to assess the differences between TGs, i.e. the *bias index*;
3) a methodology to choose what TG, or TGs, to use among available TGs to minimise the bias;
4) an experimental evaluation on BRITE, NPART and GTITM TGs, showing the presence of bias in picking either a single TG or a pair of TGs.

The rest of the paper is organised as follows. Section II describes background and discusses related work. The system model for our investigation is introduced in section III. The methodology we propose is detailed in section IV. The experiments and obtained results are presented in section V. Finally, section VI draws conclusions and outlines possible future work.

II. BACKGROUND AND RELATED WORK

In this paper we focus on TGs that provide the initial placement of nodes over a plane. As we are dealing with Ad Hoc Mesh Networks, we are not interested in how nodes are connected among each other and assume that any node can communicate directly with all the nodes lying within its transmission range.

TGs can differ mainly in how nodes placement is decided [14] and what each node represents [5].

Node placement strategy can be based either on some *predefined model* or on *real-world measurements*. In the former case, a certain probability distribution can be used, such as the *Waxman model* [15], or specific strategies can be enforced to preserve the inter-node distance among nodes placed on a line (*chain node placement*) or to position nodes at the intersections of square cells when the plane is organised as a grid (*grid node placement*). In the latter case, nodes positions are instead determined in compliance with real-world measurements of existing network topologies. Nodes in an artificial topology can represent either autonomous systems (AS), i.e. *AS-level topologies*, or routers, i.e. *router-level topologies*.

Some existing works in literature deal with the investigation of diverse aspects of TGs, e.g. how realistic generated topologies are. Several works [7-9] focus on TGs for Internet topologies by comparing the topologies they generate with available real Internet map topologies, with the aim of assessing to what extent those topologies can be considered realistic. Rossi et al. [13] propose a framework to analyse Internet topologies by using a multi-level approach based on a number of graph measures and existing reference datasets. Their goal is to assess whether Internet TGs comply with their claimed objectives and how realistic generated topologies are. Our work differs from those papers mainly because we do not evaluate whether artificial topologies are realistic, rather we investigate the bias in topologies generated by different TGs. Furthermore, we tackle Ad Hoc Mesh Networks rather than Internet.

Heckmann et al. [5] compare three TGs according to the similarity of generated topologies with an available collection of real-world topologies.

Although all those works, likewise ours, focus on evaluating and comparing existing TGs, the main difference lies in the goal of such a comparison. In fact, while existing literature is interested in measuring how well generated topologies represent real-world networks, we concentrate on an orthogonal aspect by investigating whether picking a certain TG rather than another one, or rather than choosing more TGs, can introduce bias. From this point of view, our contribution is novel and complements existing research on comparing available TGs.

III. SYSTEM MODEL AND METHODOLOGY

We consider a set TG with N^{TG} topology generators (TG), i.e. $|TG| = N^{TG}$. Each TG generates coordinates for the initial placement of nodes, i.e. devices, within a defined square *topology area*, with sides *D units* long. Each TG tg_i generates a set T_i with N^T topologies, where $i=0,...,N^{TG}-1$. The set containing all the topologies generated by all the TGs is referred to as

$$T = \bigcup_{i=0}^{N^{TG}-1} T_i \tag{1}$$

hence $|T| = N^{TG} \cdot N^T$. Each topology $t_j \in T_i$ has N nodes $N_j = \{n_k\}$, where $i=0,...,N^{TG}-1$, $j=0,...,N^T-1$, $k=0,...,N-1$. Each node n_k is identified by its bi-dimensional coordinates (x_k,y_k) in the topology area, where $0 \le x_k,y_k \le D$. Given two nodes n_a and n_b $(a,b=0,...,N-1)$, we define their Euclidean distance as

$$d(n_a,n_b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \tag{2}$$

In Ad Hoc Mesh Networks, any device can establish connections with other devices placed within a specific distance, which we refer to as *radius r*. We consider a number N^R of different radii $R = \{r_i\}$, where $i=0,...,N^R-1$ and $0 < r_j < r_{j+1} < D$ for $j=0,...,N^R-2$.

IV. Methodology to analyse the bias of topology Generators

In general, a topology generator (TG) introduces bias if the topologies it generates are not representative enough of some target application, such as analysing routing protocols. It is not trivial to decide whether a given set of topologies can be considered representative enough of a certain application, let alone it is possible to provide general criteria to evaluate the representativeness of a group of topologies regardless of what they are intended to be used for. However, if we consider the universe set TU, containing all the possible topologies, and a subset of it S⊂TU, we can investigate to what extent S is representative of TU by inspecting the differences between topologies in S and topologies in TU. We propose to use those differences to analyse the bias of using topologies in S only, i.e. the *larger* and *sharper* such differences, the higher the bias.

Although we cannot have in practice a set like TU, we do have a number of available TGs, TG (see section III), which can be used to generate a set of topologies T. While we do not know how much T is representative of TU, we claim that T is the best approximation of TU we can aim for from a pragmatic point of view. Hence, to measure the bias introduced by a TG $tg_i \in$ TG, we can examine the differences between the topologies it generates, T_i, and the topologies in T.

We propose a two-steps methodology to analyse the bias of TGs. The first step is modelling topologies by extracting a number of characteristic features, which will be used to have a compact, numeric representation of topologies and enable to measure the differences between them. The second step is indeed computing a metric to quantify the dissimilarities between topologies generated by different TGs. The extent of those differences, i.e. how large they are, provides an objective scale of the bias. The approach we propose consists in computing the average distance between the topologies generated by a TG and all the topologies in T. By mapping topologies into the space generated by the chosen features, we use the Hedges' g [6], measure of effect size, to quantify the difference between two populations: the topologies generated by a specific TG and the topologies in T. We refer to such a difference as *bias index*.

We used the following types of features to characterise topologies: *inter-node distance* (statistics on distances between nodes), *spatial distribution* (statistics on how nodes are distributed over the topology area), *node density* (how many nodes are within the transmission range, on average, for each radius in R), *shared neighbours distribution* (how many nodes are within the transmission range of each node pair, on average, for each radius in R) and *cluster coefficient* (how many neighbour node pair are within transmission range, on average, for each radius in R). For a detailed explanation of how those features are computed, refer to [12]. We detail how we compute the bias index and the TG selection approach in sections IV-A and IV-B, respectively.

A. Bias Index

The bias index of a TG $tg_i \in$ TG with respect to all the TGs in TG is measured as the distance between the topologies T_i generated by tg_i and all the topologies generated T. This distance is computed on the basis of the following feature-based representation of a topology t_j

$$t_j = \langle f_0^j, \ldots, f_{F-1}^j \rangle \tag{3}$$

where f_k^j is the value of the k-th feature of t_j ($k = 0,...,F-1$) and F is the number of used features, equal[1] to $12+3N^R$.

Hedges' g [6] is used to estimate the standardised mean difference between two populations, i.e. the average distance between the elements of two different populations, measured in standard deviations. Although in its original form it can be applied to single-dimension elements only, we propose to extend Hedges' g to F dimensions to quantify the difference between topologies in T_i and in T.

We first detail how to apply Hedges' g to a single feature f_k, where $k = 0,...,F-1$. We define T^k and T_i^k as the projections of T and T_i to feature f_k, respectively, as follows

$$T^k = \{f_k^j \vee t_j = \langle f_0^j, \dots, f_{F-1}^j \rangle \in T\} \quad (4)$$

$$T_i^k = \{f_k^j \vee t_j = \langle f_0^j, \dots, f_{F-1}^j \rangle \in T_i\} \quad (5)$$

Let m^k and s^k be the mean and standard deviation of T^k, respectively. Let m^k_i and s^k_i be the mean and standard deviation of T_i^k, respectively. In compliance with the original formulation, we define Hedges' g for a single feature f_k as

$$g_i^k = \frac{m^k - m_i^k}{s_i^{*k}} \quad (6)$$

where $s^*{}_i^k$ is the pooled standard deviation for T^k and T_i^k, computed as follows

$$s_i^{*k} = \sqrt{\frac{(|T^k| - 1) \cdot (s^k)^2 + (|T_i^k| - 1) \cdot (s_i^k)^2}{|T^k| + |T_i^k| - 2}} \quad (7)$$

Finally, to obtain the bias index g_i for tg_i, we combine all the F values g_i^k by considering each of them as a distance along one dimension, as follows

$$g_i = \sqrt{\sum_{k=0}^{F-1} (g_i^k)^2} \quad (8)$$

B. TG Selection

The bias index can be used to choose what TG to pick to reduce the possible bias. Selecting the TG with the lowest bias index would correspond to using the set of topologies with the lowest distance, on average, from the whole set T of available topologies. According to the methodology approach introduced at the beginning of this section, this in turn means choosing the most representative subset of topologies available, if a single TG has to be selected.

If more than one TG can be picked, say p out of N^{TG}, then the same strategy can be used by considering the possible $\binom{N^{TG}}{p}$ subsets of T, each in the form

$$T_{i_0,\dots,i_{p-1}} = j = 0p - 1T_{ij} \quad (9)$$

with $T_{ij} \subset T$, $0 \leq i_0 < \cdots < i_{p-1} < N^{TG}$, $0 < p < N^{TG}$, and computing the corresponding bias index. Again, the subset with the lowest bias index is the most representative of T. We refer to $g_{i0,\dots,i_{p-1}}$ as the bias index of $T_{i0,\dots,i_{p-1}}$.

V. EXPERIMENTAL EVALUATION

We apply the proposed methodology to a number of well-known TGs, namely BRITE [10], NPART [11] and GTITM [1]. The parameters we choose to instantiate the model (see section III) are reported in section V-A. The experiments on bias index and obtained results are described in section V-B.

A. Evaluation Settings

With reference to the system model defined in section III, we consider the $N^{TG} = 3$ TGs described in the previous section, i.e. TG = {BRITE, NPART, GT-ITM}, and generate $N^T = 1000$ topologies for each TG. Each topology has $N = 1000$ nodes. The reference topology area has sides $D = 1000$ units long. We evaluate the following $N^R = 8$ radii: R = {5,10,20,30,40,60,80,100}.

B. Bias Index Evaluation

We compute the bias index g_i for each TG and $g_{i,j}$ for each pair of TGs, as described in section III-A. The results are reported in table I. As can be noted, BRITE topologies seem to be significantly different from those generated by NPART and GT-ITM, and vice-versa, which suggests that using either TG alone would provide a set of topologies significantly different from the set including all the topologies. However, if only one TG has to be selected, NPART proves to generate topologies that are less different on average from those generated by all available TGs. If two TGs can be chosen, BRITE and NPART show to be the best pair to consider.

It should be noted that the bias index of a single TG can be vastly different from the index when combined with one of the other 2 TG's. This is because the 'Bias Index' is a measure of how different each TG or combination of TG's is against the general population, the more TG's in the bias index, the nearer to the general population, so a combination of all the TG's would give a bias index of 0.

TABLE I

BIAS INDEX OF THE CONSIDERED TGS.

Topology Generator(s)	Bias index
NPART	1.890
GT-ITM	2.145
BRITE	4.282
BRITE + NPART	0.908
BRITE + GT-ITM	0.976
NPART + GT-ITM	2.430

1 There are 7 features for inter-node distances, 5 features for spatial distribution and as many features as the number of radii N^R for (i) node density, (ii) shared neighbours distribution and (iii) clustering coefficient.

VI. CONCLUSION

In this paper, we investigate the presence of bias in node placement in Ad Hoc Mesh Networks simulations due the choice of what TG, or TGs, to use among a fixed number of available TGs. This is of importance in the design of simulation used in the testing and analysis of network protocols and signal processing.

In particular, we explore these research questions: (a) how to measure the difference between topologies generated by distinct TGs? and (b) how to choose what TG, or TGs, to use to reduce such a bias?

To answer the first question, we propose a metric, namely the bias index, based on measurements of a number of characteristic features of generated topologies. We use these measurements to calculate the distance between topologies generated by a single TG and topologies produced by all available TGs. To answer the second question, we propose a methodology to select the TG, or TGs, that minimise the bias index. We present an experimental evaluation where we compute the bias index for three well-known TGs: BRITE, NPART and GT-ITM. Obtained results prove that topologies generated by a single TG are different from those created by all the three TGs.

As future work, we plan to carry out additional evaluations to investigate how the bias index is linked to variance in the results of same experiments performed on different TGs. A number of reference algorithms can be chosen, e.g. routing protocols, and executed on available generated topologies to verify whether lower values for bias index can actually lead to reduced variance of obtained results, with respect to the experimental outcomes that would be achieved by using all the available TGs. An additional, significant future work concerns the sensitivity analysis on both system model parameters and TGs configurations, to assess to what extent such a tuning affects computed values for bias index.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Michael O'Sullivan conducted the research, analysed the data and wrote the paper; Leonardo Aniello contributed to problem formulation, manuscript organisation and writing; Vladimiro Sassone supervised the work from an academic point of view; all authors had approved the final version.

REFERENCES

[1] Kenneth L Calvert, Matthew B Doar, and Ellen W Zegura. Modeling internet topology. *IEEE Communications magazine*, 35(6):160–163, 1997.

[2] Bo Cheng and G. Hancke. Energy Efficient Scalable Video Manycast in Wireless Ad-hoc NETworks. In *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, pages 6216–6221, Oct 2016.

[3] C. Cheng and S. Lin. A Hole-Bypassing Routing Algorithm for WANETs. In *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*, pages 547–550, Oct 2017.

[4] M. H. Gunes and M. B. Akg¨ un.¨ Link-level network topology generation. In *Proceedings of 31st International Conference on Distributed Computing Systems Workshops (ICDCSW)*, 2011.

[5] Oliver Heckmann, Michael Piringer, Jens Schmitt, and Ralf Steinmetz. On realistic network topologies for simulation. In *Proceedings of the ACM SIGCOMM workshop on Models, methods and tools for reproducible network research*, pages 28–32. ACM, 2003.

[6] L. V. Hedges. Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational and Behavioral Statistics*, 1981.

[7] D. Magoni and J. Pansiot. Evaluation of internet topology generators by power law and distance indicators. In *Proceedings 10th IEEE International Conference on Networks (ICON 2002). Towards Network Superiority (Cat. No.02EX588)*, pages 401–406, 2002.

[8] Damien Magoni and J J Pansiot. Analysis and Comparison of Internet Topology Generators. *NETWORKING 2002: Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications*, 2345:364– 375, 2006.

[9] Damien Magoni and Jean-Jacques Pansiot. Influence of network topology on protocol simulation. In *International Conference on Networking*, pages 762–770. Springer, 2001.

[10] Alberto Medina, Anukool Lakhina, Ibrahim Matta, and John Byers. BRITE: Universal Topology Generation from a User's Perspective. Technical report, Boston, MA, USA, 2001.

[11] Bratislav Milic and Miroslaw Malek. NPART-node placement algorithm for realistic topologies in wireless multihop network simulation. In *Proceedings of the 2nd international conference on simulation tools and techniques*, 2009.

[12] Michael O'Sullivan, Leonardo Aniello, and Vladimiro Sassone. A Methodology to Select Topology Generators for WANET Simulations (Extended Version), 2019.

[13] R Rossi, S Fahmy, and N Talukder. A Multi-Level Approach for Evaluating Internet Topology Generators. *2013 IFIP Networking Conference*, pages 9 pp.–9 pp., 2013.

[14] M L Sanni, A A Hashim, F Anwar, G S M Ahmed, and S Ali. How to model wireless mesh networks topology. *IOP Conference Series: Materials Science and Engineering*, 53(1):012037, 2013.

[15] B. M. Waxman. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, Dec 1988.

[16] Y. Xu, J. Liu, Y. Shen, X. Jiang, and T. Taleb. Security/qos-aware route selection in multi-hop wireless ad hoc networks. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6, May 2016.

[17] Y. Xu, J. Liu, O. Takahashi, N. Shiratori, and X. Jiang. Soqr: Secure optimal qos routing in wireless ad hoc networks. In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2017.