# POBTOG: A Population-based Topology Generator for Country-Wide Communication Networks

Fabian Poignée*, Frank Loh*, Simon Schardt†, Fabian Lipp†, David Hock† and Tobias Hoßfeld*
*University of Würzburg, Institute of Computer Science, Würzburg, Germany
E-Mail: {fabian.poignee | frank.loh | tobias.hossfeld}@uni-wuerzburg.de
†Infosim GmbH & Co. KG, Würzburg, Germany
E-Mail: {schardt | lipp | hock}@infosim.net

**Abstract**

Network topology data is critical for evaluating and optimizing service provider infrastructure, but real-world topologies are often disclosed or incomplete, limiting research. To address this, we present POBTOG, a population-based network topology generator creating hierarchical topologies comprising core and distribution layers. Our tool uses global population and administrative datasets, combined with a dualtree space-partitioning algorithm and population-weighted k-means clustering to generate topologies for any country worldwide, while only general design information, such as the number of layers and their connections, is required.

## I. Introduction

High-speed networks interconnecting buildings, businesses, cities, and countries drive globalization and economic growth. As a result, networks are growing increasingly complex, with heterogeneous devices serving diverse functions. Each provider develops a unique topology, continuously expanded and optimized. For service providers, overall network performance matters more than its specific structure. Large providers of video streaming, social network, and cloud storage services, such as Amazon, Netflix, Meta, Amazon, or Google, depend on reliable, high-quality networks for their services. These companies either rely on existing infrastructure or partnerships with network providers [1], [2] to assess and improve performance.

Recently, digital twinning has gained traction for network, traffic, and scalability assessment [3]–[5]. It enables extensive testing without affecting live systems and supports multi-layer analysis in high accuracy if input data and system structure is precise. However, service and network providers cannot disclose sensitive internal information for confidentiality and security reasons and digital twins can become inaccurate despite sound methodology.

To avoid testing on real traffic, researchers use traffic models and datasets [6], [7], along with generalized network statistics of households, or within core network components [8], [9]. Performance assessments still require insight into the actual topology and capabilities. Datasets like the Internet Topology Zoo [10] provide references but cover only a small subset of real networks. Provider networks are often heterogeneous, with unique characteristics and advantages not publicly disclosed. This lack of detailed topologies limits network improvements and insights from digital twins.

To mitigate this issue, we propose a population-based and adaptable approach for hierarchical network topology generation on any geographic area, which we implement in POBTOG for country-wide topology generation comprising core and distribution network. With this approach, network providers only need to share general design principles, such as the number of layers in their network together with connection information, e.g., organization as a star, ring, or as a mesh. With this information, network topologies can be generated, eliminating the need to disclose detailed topology information or company internals. Consequently, this is a valuable additional tool in network research as the evaluation of new technologies on large topologies is often a key aspect. Thus, we present our generator design, compare our population-based approach to alternatives, and show an exemplary topology for Germany.

## II. Generator Design

In the following, we present implementation specifics of POBTOG, detail on its population-based space partitioning approach, and assess the methodology of using population information as basis.

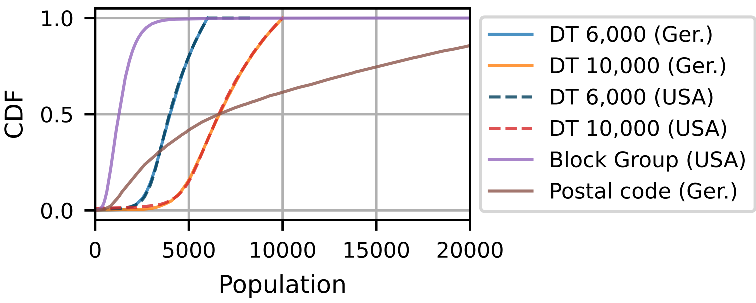


Figure 1: Population distribution for different approaches.

### A. *Implementation Details*

Our topology generation contains five steps explained hereafter.

**Step 1 - Generator Input:** The input for POBTOG is the country to generate the topology in, the number of core and distribution layers, the number of nodes per layer, the local topology type, and the redundancy level for each layer. Furthermore, two external datasets are used: The Global Human Settlement Layer (GHSL) [11], a raster dataset representing population per cell in a $100\,\mathrm{m}$ Mollweide and 3 arcseconds WGS84 resolution. To isolate a country's cells, we use the Global Administrative Areas (GADM) database. It uses ISO 3166-1 alpha-3 codes for 249 regions [12] to define national boundaries, enabling data masking for cells fully or partially within a country and to obtain its population. Population is used as a proxy for network endpoints, such as routers or base stations, which typically serve a set number of people. This input can be adjusted to meet future user needs.

**Step 2 - Space Partitioning:** We apply a dualtree (DT) space-partitioning algorithm to the cells used as input. Like a quadtree [13], which partitions space hierarchically, DT recursively divides the area until each subspace's population falls below a set threshold. Unlike a quadtree's four-way split, DT splits along the longer axis into two subspaces with equal population.

**Step 3 - Node Hierarchy Generation:** Then, we assign node locations in the highest layer using the geographic centroids of the resulting regions. Lower layers are built iteratively using population-weighted k-means clustering on the previous layer's nodes. Cluster centroids become node locations and parent nodes. If a centroid falls outside land, it is shifted to the nearest land point. This process forms a hierarchical tree of all nodes in the topology.

**Step 4 - Network Layer Generation:** To get a final topology, sub-topologies are generated according to the input configuration. For each layer, different configurations are possible. Supported topology types are star, with the parent as center, ring, and full mesh. Since computing the shortest ring is NP-hard, we use the Christofides algorithm [14], which yields a path which is $50\,\%$ longer than the optimum in the worst case. Link capacities follow layer-specific settings. Inter-layer connections take the lower of the two layers' capacities. Other assumptions, such as the physical route of an edge are not incorporated, as information like link delay can be estimated from the geolocation of neighboring nodes [15]. Link speeds do not necessarily have to match common link types, but focus on the total capacity between two locations. This can be achieved by using multiple physical links in parallel. For the same reason we neglect information about the switching device capabilities, such as the number of ports. Finally, since redundancy is crucial to guarantee fault tolerance, redundant nodes are added according to the configured redundancy level of each layer.

**Step 5 - Generator Output:** The final topology is visualized as an interactive HTML file, allowing exploration on a geographical map with options to toggle nodes and edges by layer, as well as their redundant parts. The hierarchical structure of the nodes can be exported to CSV and the final topology with node IDs, locations, links, and capacities as JSON. The latter is also compatible with the topology library in [16], containing a topology generator for industrial networks which allows the definition of network devices. With information about the number of end devices and industrial networks for a network, the approaches can be combined to achieve an end-to-end network topology. In total, POBTOG can efficiently create a comprehensive hierarchical network topology for any country in the world by a single mouse click.

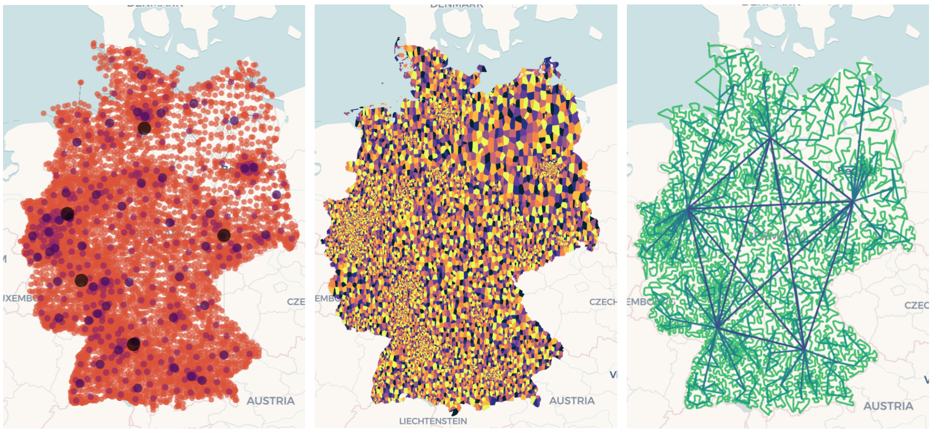


Figure 2: Exemplary nodes (left) and edges (right) within Germany and Voronoi cells for the uppermost layer (mid).

**Step 6 - Data Augmentation for Real Usage:** Topologies can also be enriched with external datasets. For example, using the Global Integrated Power Tracker (GIPT) [17], we assign each node an environmental score based on nearby power plants via a weighted k-nearest neighbor approach. The score reflects the carbon intensity of the local power mix (e.g., coal, gas, nuclear, solar), enabling regional estimates of environmental impact. Additional datasets, such as critical infrastructure, can be integrated similarly to support detailed planning and resilience analysis, e.g., to avoid events like Spain's national power outage on April 28, 2025.

### *B. Granularity Investigation and Comparison*

Our approach assumes that higher population density correlates with more end devices and greater network demand. As a result, areas with a larger population require more network nodes, leading to denser networks in urban regions, reflecting current real-world patterns. The lack of real accessible real topologies limits direct validation. Therefore, we rely on the expertise of our industrial partners to qualitatively compare the results to known networks. Additionally, we compare it to similar approaches from social sciences, which often use administrative regions, e.g., census blocks or postal codes. However, these vary significantly in size and population, even within the same country. This inconsistency can skew network modeling. Figure 1 shows the population for USA Census Block Groups (CBGs), German postal codes, and our DT-based partitions for Germany and the USA as cumulative distribution function (CDF) with thresholds of 6000 and 10,000. The x-axis is capped at 20,000 for better visibility. Most CBGs have a low population, though some exceed 40,000 inhabitants. Similarly, German postal codes are also inconsistent, peaking at 58,782. On the other hand, results of our DT approach are more robust since it separates space into areas with a configurable maximum capacity and delivers comparable results for different countries, shown at the examples of Germany and the USA.

## III. Example Topology and Future Work

To demonstrate POBTOG, we use a configuration for Germany, representing an exemplary nation-wide Internet service provider. The code and further visualization for this and other countries is available in [18] (made publicly available on acceptance). The topology uses information from our industry partners and consists of four layers, shown in Table I. A population threshold for the DT of 10,000 for the uppermost layer implicitly results in 12,425 nodes. In reality, those nodes would be points of presence (POPs), for example realized as a main distribution

Table I: Parameter for exemplary four-layer topology.

| | Layer 0 | Layer 1 | Layer 2 | Layer 3 |
|---|---|---|---|---|
| nodes | 5 | 50 | 400 | (12,425) |
| connect | fully | star | ring | ring |
| link | $10\,\mathrm{Tbit\,s^{-1}}$ | $400\,\mathrm{Gbit\,s^{-1}}$ | $100\,\mathrm{Gbit\,s^{-1}}$ | $40\,\mathrm{Gbit\,s^{-1}}$ |
| R | 2 | 2 | 2 | 2 |

frame (MDF). Figure 2 shows the generated nodes on the left, where larger and darker nodes indicate lower layers. Especially the north-east of Germany is sparsely populated, shown by the orange nodes on the highest layer. The area around Berlin stands out in that area with a higher population density. This effect propagates to the next layers, which are more sparse, leading to less purple nodes. However, in a highly urbanized area such as the Ruhr area in the west, a high node density is visible. The middle of Figure 2 shows a Voronoi diagram for the uppermost node layer. Each Voronoi cell can be interpreted as the area of responsibility for the local POP. More nodes in densely populated areas result in smaller Voronoi cells. On the right, the edges of the topology are depicted. The mesh topology of the five core nodes can be seen in dark blue and the rings of the upper layers shades of green.

In the future, we will incorporate Internet exchange points and submarine cable data to interconnect our augmented country-level topologies, facilitating development of more global and sustainable networks and network management strategies.

## IV. Demo Setup Requirements

For our demonstration, we plan to demonstrate the generators capabilities. Conference participants will be able to generate their own topology, e.g., for their own country by choosing a custom configuration on a graphical interface. The computation is done on a remote cluster at the University of Würzburg, Germany and the results will be transferred to a notebook at the conference site. Thus, the following equipment is required for our live demonstration: a table to place a notebook, a poster stand or similar to present the poster, a stable Internet connection. Optionally, the demonstration would benefit from a provided monitor to extend the notebooks' screen so that a larger audience can follow the demonstration at the same time.

## References


[1] T-Mobile, "T-Mobile and Google Cloud Join Forces on 5G Advanced Network Solutions," 2023, accessed: 2025-06-26. [Online]. Available: https://www.t-mobile.com/news/business/t-mobile-and-google-cloud-join-5g-advanced-network-solutions

[2] ——, "Netflix on us," 2025, accessed: 2025-06-26. [Online]. Available: https://www.t-mobile.com/tv-streaming/netflix-on-us

[3] Y. Wu, K. Zhang, and Y. Zhang, "Digital Twin Networks: A Survey," *IEEE Internet of Things Journal*, 2021.

[4] P. Almasan *et al.*, "Network Digital Twin: Context, Enabling Technologies, and Opportunities," *IEEE Communications Magazine*, 2022.

[5] H. Hong, Q. Wu, F. Dong, W. Song, R. Sun, T. Han, C. Zhou, and H. Yang, "Netgraph: An Intelligent Operated Digital Twin Platform for Data Center Networks," in *ACM SIGCOMM 2021 Workshop on Network-Application Integration*, 2021.

[6] F. Poignée, A. Seufert, F. Loh, M. Seufert, and T. Hoßfeld, "Fitting the Puzzle: Towards Source Traffic Modeling For Mobile Instant Messaging," in *2024 15th International Conference on Network of the Future (NoF)*. IEEE, 2024.

[7] F. Loh, F. Wamser, F. Poignée, S. Geißler, and T. Hoßfeld, "Youtube Dataset on Mobile Streaming for Internet Traffic Modeling and Streaming Analysis," *Scientific Data*, 2022.

[8] DE CIX, "Frankfurt Traffic Statistics," 2025, accessed: 2025-06-26. [Online]. Available: https://www.de-cix.net/en/locations/frankfurt/statistics

[9] Sandvine, "Internet Phenomena," 2025, accessed: 2025-06-26. [Online]. Available: https://www.applogicnetworks.com/phenomena

[10] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The Internet Topology Zoo," *IEEE Journal on Selected Areas in Communications*, 2011.

[11] M. Pesaresi, M. Schiavina, P. Politis, S. Freire, K. Krasnodębska, J. H. Uhl, A. Carioli, C. Corbane, L. Dijkstra, P. Florio, H. K. Friedrich, J. Gao, S. Leyk, L. Lu, L. Maffenini, I. Mari-Rivero, M. Melchiorri, V. Syrris, J. V. D. Hoek, and T. K. and, "Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data," *International Journal of Digital Earth*, 2024. [Online]. Available: https://doi.org/10.1080/17538947.2024.2390454

[12] *ISO 3166-1:2020 Codes for the representation of names of countries and their subdivisions – Part 1: Country codes*, International Organization for Standardization Std., 2020, https://www.iso.org/standard/72482.html.

[13] H. Samet, "The quadtree and related hierarchical data structures," *ACM Computing Surveys (CSUR)*, 1984.

[14] N. Christofides, "Worst-Case Analysis of a New Heuristic for the Travelling Salesman Problem," in *Operations Research Forum*. Springer, 2022.

[15] S. Kaune, K. Pussep, C. Leng, A. Kovacevic, G. Tyson, and R. Steinmetz, "Modelling the Internet Delay Space Based on Geographical Locations," in *2009 17th Euromicro International Conference on Parallel, Distributed and Network-based Processing*. IEEE, 2009.

[16] A. Grigorjew, "Tsn problem generator," https://github.com/lsinfo3/tsn_problem_generator, 2023, accessed: 2025-06-26.

[17] Global Energy Monitor, "Global Integrated Power Tracker," 2025, april 2025 release. [Online]. Available: https://globalenergymonitor.org/projects/global-integrated-power-tracker/

[18] F. Poignée, "POBTOG," https://github.com/lsinfo3/pobtog, 2025.