Development of network generation model with the properties of real road networks by machine learning

Takumi Mori[a], Hiroe Ando[b,*], Ryuji Kakimoto[b]

[a] Kumamoto University, Graduate School of Science and Technology, Japan
[b] Kumamoto University, Center for Water Cycle, Marine Environment and Disaster Management, Japan
* Corresponding author

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

February 3, 2025

Keywords: network generation, road network, cluster analysis, machine learning, GAN

1 INTRODUCTION

Network science has developed methods for elucidating the structure and connection characteristics of various data, including infrastructure networks. However, most analyses focus on mathematical approaches assuming regular structures, while experimental approaches using real data are limited. Road networks, with their unique structural characteristics like degree restrictions and hierarchy, are not accurately represented by virtual networks such as grids, which fail to capture the complexity of real cities. Having said that, since it is difficult to obtain general results in a small number of real cities, the road network generation with the characteristics of real road networks is required. Hartmann *et al.* (2017) recognized road networks as images and generated a generalized road network using machine learning. This research aims to develop a road network generation model that uses machine learning to represent networks as matrices, making them suitable for network science analysis. The contribution of road network generation model is the applicability of growth and robustness models in network science to road networks, allowing more generalized results to be obtained by using a large number of artificial road networks rather than standard test networks to validate transportation network models such as the Sioux Falls-Anaheim networks.

2 ROAD NETWORK GENERATION MODEL

2.1 Matrix-based Generative Adversarial Networks

This research use a machine learning algorithm called Generative Adversarial Networks (GANs) to generate road networks. GANs are a type of generative conflict network proposed by Goodfellow *et al*. (2014) and consist of two main components: a generator and a discriminator. The generator tries to generate data that is close to the real thing to deceive the discriminator, and the discriminator tries to distinguish between the generated data and the real data. Learning proceeds by minimizing the probability that the generator will deceive the discriminator and maximizing the probability that the discriminator will see through the generated data. This interaction gradually leads the generator to produce more realistic data.

To apply the generated road network to network science analysis that do not require physical location information, this study adopts an adjacency matrix representation consisting of only the connection structure to generate road networks. We improved the generation model by inputting the adjacency matrix of an real road network as training data and outputting the adjacency matrix. The size of the learning and generation adjacency matrices must be equal, so it is necessary to prepare the networks for training with the same number of nodes. In the process of generating the network using training data, some constraints are placed to reflect the characteristics of the real road network. Constrained matrix generation is expected to produce realistic and actionable results that consider road and traffic patterns. We add a constraint ensuring that the network is a “connected graph” as a basic road property. This constraint is imposed because real cities do not have isolated parts, except in special cases such as small islands. In addition to these basic constraints, there are two other types of constraints: the degree distribution and the degree correlation. In the degree distribution constraint, a degree list of the nodes to be generated is created based on the node degree distribution of the real network. Next, a combination of nodes is randomly selected, and links are added so that the specified degree is not exceeded. The degree correlation constraint is an iterative process in which links are randomly added or removed until the degree correlation is the median of the real road network ±0.1. The degree correlation is the correlation between the degrees of two nodes connected by a link, and takes values from -1 to 1. Hereafter, the model constrained by degree distribution and the connected graph is referred to as Model 1, while that by degree correlation and the connected graph is referred to as Model 2. Figure 1 shows the generation flow of network generation. It should be noted that the connected graph constraint is adjusted in the final step, after other constraints have been imposed.

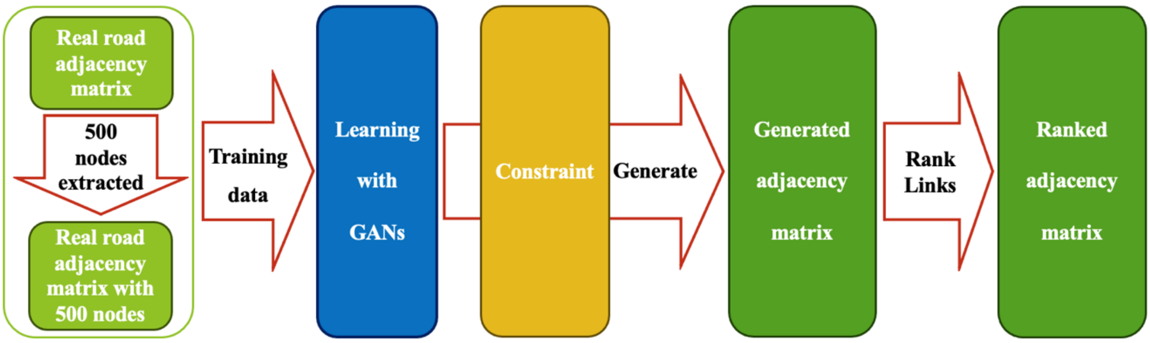

Figure 1 – *Network generation flow*

2.2 Learning and Generation of Meshed Real Road Networks

This study used 10km square mesh units from the ArcGIS Geo Suite Road network data (2023 edition) for the Kyushu region of Japan as training data. To define the characteristics of the road network to be generated, the road network mesh was characterized by clustering analysis using six connectivity structure indicators. The six indicators are as follows: the link density is the total road length per unit area, the node density is the number of nodes per unit area, the average degree is the average number of connections per node, the weighted degree correlation is the strength of the connections between nodes, the fractal dimension is an indicator of the complexity of the network by a box counting method, and the average shortest path length between all nodes. Based on these indicators, the road network mesh in the Kyushu region was clustered using the Ward method, and the optimal number of clusters was determined to be 5 by the elbow method. The 613 meshes was classified into five clusters: urban, suburban, coastal, coastal (without residential areas), and mountainous. 67 meshes included in one cluster that is recognizes as an urban road network with a large population were used as training data for network generation model in this paper. One node is randomly selected from each mesh and the target is extended to the nodes connected to that node, repeatedly until the criteria number of nodes is reached. Here, 500 nodes are used as a criterion, based on the condition that 95% of the nodes in each mesh are included.

Using meshes included in the urban cluster as training data, 100 artificial road networks were generated by Model 1 and 2. The generated networks and real road networks were compared using the squared error loss function in equation (1), where, i is type of index $(i = 1,2,\dots,6)$, G is generated network by matrix-based GANs, R is real road network, $Value$ is the median index i value for a set of networks. The six types of indicators used for comparison are shown in Table.1.

From Table 1, a comparison of Model 1 and 2 reveals the following features: Model 2 has large difference in the number of links, and Model 2 is larger than Model 1 in the average degree and the maximum degree. The minimum degree are both 0 due to the connected graph constraint. In degree correlations, Model 1 is slightly larger than Model 2 with degree correlation constraints. In cluster coefficients, both models have large values and are far from the real roads. From the total loss, Model 1, with only degree distribution constraint, also keeps degree correlation losses small and is evaluated to be a better fit to the real road network.

Comparisons with real roads for both generation models revealed the features and challenges that could be reproduced. In the results of the squared error loss function, Model 2 slightly smaller error, with high reproducibility of the mean, maximum, and minimum degrees. However, there is a high loss in the reproduction of local connections, such as cluster coefficients, which may be due to the emphasis on the overall degree distribution and number of links across the network and the lack of consideration for local connections. Figure 2 shows the network connection structure with one mesh from the training data and one adjacency matrix generated by Model 1 as examples. These are diagrams without location information. It is possible to replicate many network data sets that have similar visual characteristics but do not actually exist. Although this paper generates 100 road networks with 500 nodes, this number can be flexibly changed.

$$y_i = \left(1 - \frac{Value_i^G}{Value_i^R}\right)^2 \tag{1}$$

Table 1 – *The squared error loss function y_i in each generation model*

Model	Num of Links	Average degree	Maximum degree	Minimum degree	Degree Correlation	Cluster coefficient	Total
1	0.013	0.013	0.028	0.000	0.023	0.906	0.983
2	0.044	0.051	0.111	0.000	0.018	0.875	1.099
RRNG	0.013	0.013	0.028	0.000	0.022	1.000	1.075

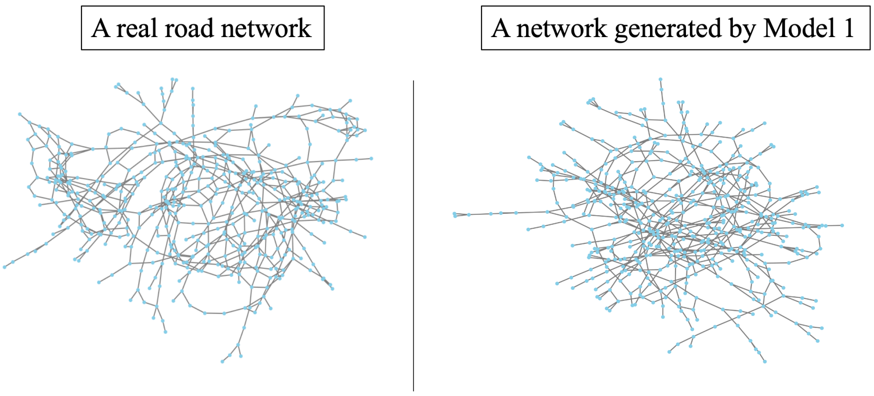

Figure 2 – *Examples of each road network connection structure*

3 RANKING LINKS BY ROAD ATTRIBUTES

3.1 Assigning road attributes and the relationships between attributes

The road network generation model up to here is based on an adjacency matrix that does not consider road attributes. It is important to consider the rank of the roads, as the real road links has attributes.

Road rank also relates to the hierarchical nature of roads, which is an inherent characteristic of road networks. Therefore, we propose a method to set the road rank based on the real road attributes for Model 1 which had better reproducibility. The elements of matrix generated by Model 1 are assigned road ranks corresponding to the speed limit from rank 1 (30km/h) to rank 6 (80km/h). The percentage of road ranks for each link corresponds to the percentage of real roads in the urban cluster (30km/h: 5.1%, 40km/h: 45.9%, 50km/h: 39.5%, 60km/h: 8.8%, 70km/h: 0.4%, 80km/h: 0.3%). In addition, the difference in rank between connecting links is limited to a maximum of 4, a restriction to avoid connections from local roads to highways. This model that adds rank connection constraints to Model 1 is called a ranked road network generation model (RRNG).

3.2 Generated road networks with attributes

100 adjacency matrices with road ranks were generated by RRNG using the real road network of the urban cluster as training data. Table 1 shows the squared error loss functions of RRNG is 1.075. The average, maximum and minimum degree are small, and the characteristics are close to the real road. However, the cluster coefficient is 1, which is not like the real road. The number of links, the average, maximum and minimum degree are the same values because the network topology was generated using GANs with the same procedure as in Model 1. The ranked constraint slightly reduced the degree correlation, however increased the difference in the clustering coefficient, resulting in a slight increase in the total. Although Model 1 was superior in terms of the sum of the squared error loss function than RRNG, the results of the degree-related indicators show that it was possible to rank the road network while maintaining the accuracy of reproducibility of topology characteristics.

4 CONCLUSIONS

This study proposed a road network generation model to replicate artificial road network of the same size with road attributes by using the real road network data belonging to the urban cluster of the Kyushu region of Japan as training data. The proposed model uses GANs which is one of the machine learning methods, to learn and generate adjacency matrices. The model that imposed degree distribution and connected graph constraints showed results close to the characteristics of the real road networks. By providing road ranks based on speed limits and connection constraints between road ranks, we generated a ranked adjacency matrix that maintains the topology specific to the road. This provides a large number of road networks with homogeneous connection structures of the required size for the validation of the mathematical model. Network Science analysis requiring hundreds of test networks also can be applied.

In the future, there is a need to introduce new constraints to improve local reproducibility, which was poor in the proposed model. It is also necessary to verify the possibility of generating road networks that are included in other clusters, such as suburban or mountainous areas, rather than the urban cluster. Furthermore, an evaluation method is needed to confirm the reproducibility of the allocation of road attributes, in addition to maintaining the network topology.

References

Hartmann, Stefan, Weinmann, Machael, Wessel, Raoul & Klein, Reinhard, 2017. StreetGAN: towards road network synthesis with generative adversarial networks. *Proceedings of International Conference on Computer Graphics, Visualization and Computer Vision*, **25**, pp. 133-142.

Goodfellow, Ian J., Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron & Bengio, Yoshua: Generative Adversarial Nets, 2014. *Advances in neural information processing systems*, **27**.