9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Generating Mobility Networks with Generative Adversarial Networks

Keywords: Human Mobility, Flow Generation, Mobility Networks, GANs

Extended Abstract

The mobility network generation problem consists in generating a city's entire mobility network, i.e., a weighted directed graph where nodes are geographic locations and weighted edges represent people's movements between locations, thus describing the entire set of mobility flows. A mobility network can be represented as a weighted adjacency matrix $\mathscr{A}_{n\times n}$, where $n = |V|$ is the number of nodes and an element $a_{i,j} \in \mathscr{A}$ is the number of people moving from node i to node j, with $i, j \in V$.

To solve mobility network generation, we design MoGAN (Mobility Generative Adversarial Network), a deep learning architecture based on Deep Convolutional Generative Adversarial Networks (DC-GANs), a particular type of Generative Adversarial Networks (GANs) [1]. MoGAN consists of a generator G, which learns how to produce new synthetic mobility networks, and a discriminator D, which has the task of distinguishing between real and fake (artificial) mobility networks. G and D are trained in an adversarial manner: D maximizes the probability of correctly classifying real and fake mobility networks; G maximizes the probability of fooling D, i.e., to produce fake mobility networks classified by D as real. Both D and G are Convolutional Neural Networks (CNNs), which are proven effective in capturing spatial patterns in the data. Figure 1 schematizes and describes MoGAN's architecture.

We compare MoGAN with two classical approaches for mobility flows' generation: the Gravity and the Radiation models [2]. We use the implementations available in the library scikit-mobility [3]. We also compare MoGAN with a Random Weighted model (RW) that creates a mobility network where the weight of each edge is randomly chosen from the distribution of weights for that edge in the training set. We use four public datasets describing trips with taxis and bikes in New York City and Chicago during 2018 and 2019 (730 daily networks). Two datasets contain daily information regarding bike-sharing services: the City Bike Dataset for New York City and the Divvy Bike Dataset for Chicago. Each record describes the coordinates of each ride's starting and ending stations and the starting and ending times. We define locations (nodes) based on a squared spatial tessellation and compute the flows (edges) as the number of people moving between pairs of these tiles.

We developed a tailored approach to evaluate the realism of the generated mobility networks. We construct a mobility network for each dataset for each day, obtaining 730 real mobility networks. We split the 730 networks into a training set (584 networks) and a test set (146 networks). We train MoGAN on the training set and generate 146 synthetic mobility networks (synthetic set). We then compute the difference between each network in the synthetic set and each network in the test set, so obtaining $146 \times 146 = 21,316$ values. If the generated mobility networks are realistic, they should differ from the real networks to the same extent real networks differ between themselves. To stress this aspect, we create a set of 146 mobility networks (mixed set), in which half of them are chosen uniformly at random from the test set, and the other half is chosen uniformly at random from the synthetic set. Finally, we compute the pairwise difference between any possible pair of mobility networks in the mixed set.

In our experiments, we calculate several error metrics: such as Normalized Root Mean Square Error (NRMSE), Common Part of Commuters (CPC), and Cut Distance (CD), but also metrics based on the topology of the networks like the Jensen–Shannon (JS) divergence among the weight distributions and several others. Figure 2 shows the distribution of the CPC in the four datasets' test (red), synthetic (blue), and mixed sets (green) for MoGAN (left), the Gravity model (center), and the Radiation model (right). MoGAN's CPC distributions overlap almost entirely in all four datasets, meaning that MoGAN generates mobility networks that are indistinguishable from real ones and way more realistic than those generated by the baselines. CPC results are consistent with the results for the other measures (NRMSE, CD, and JS) [4].

We provide another visualization sample of our model's performances in Figure 3. MoGAN is way better than the Gravity model, the best baseline model, at predicting flows between close tiles. The two models reach a similar performance for flows regarding tiles that are very distant to each other.

An important aspect to investigate as future work is also to what extent MoGAN is geographically transferable [5], i.e., it can be trained on a specific city and then used to generate mobility networks in a different city effectively. Another promising future direction is developing a GAN to generate a realistic mobility network for a specific condition (e.g., a rainy day or a day with some public events in the city). In the meantime, our study demonstrates the great potential of artificial intelligence to improve solutions to crucial problems in human mobility, such as the generation of realistic mobility networks. MoGAN can synthesize aggregated movements within a city into a realistic generator, which can be used for data augmentation, simulations, and what-if analysis. Given the flexibility of the training phase, our model can be easily extended to synthesize specific types of mobility, such as aggregated movements during workdays, weekends, specific periods of the year, or in the presence of pandemic-driven mobility restrictions, events, and natural disasters.

The code to train/test MoGAN and reproduce our analyses, mainly conduct and the links to the datasets used in our experiments, can be found at `https://github.com/jonpappalord/GAN-flow`.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[2] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

[3] Luca Pappalardo, Filippo Simini, Gianni Barlacchi, and Roberto Pellungrini. scikit-mobility: A python library for the analysis, generation, and risk assessment of mobility data. *Journal of Statistical Software*, 103(4):1–38, 2022.

[4] Giovanni Mauro, Massimiliano Luca, Antonio Longa, Bruno Lepri, and Luca Pappalardo. Generating mobility networks with generative adversarial networks. *EPJ data science*, 11(1):58, 2022.

[5] Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)*, 55(1):1–44, 2021.

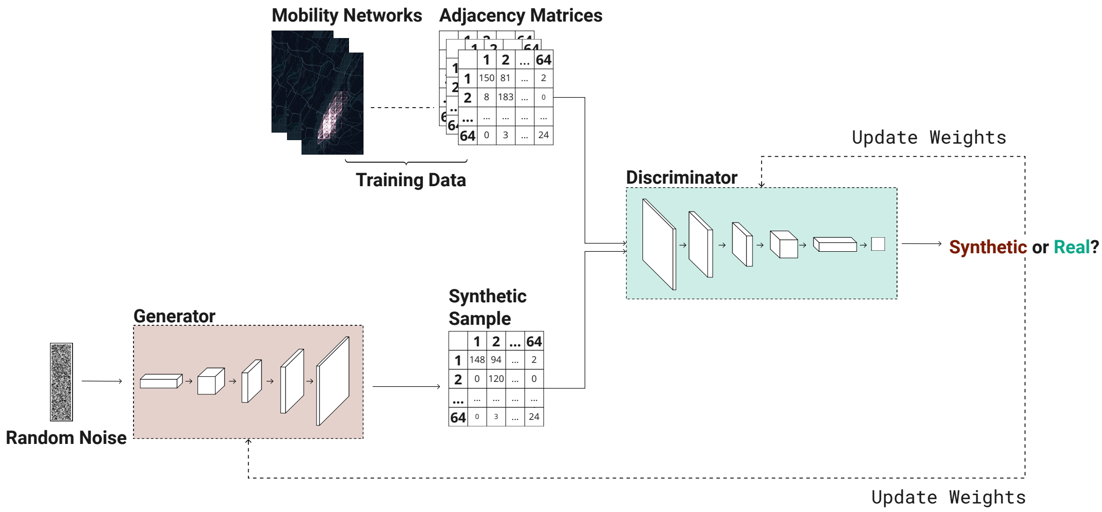

Figure 1: **Architecture of MoGAN.** The generator G (a Convolutional Neural Network or CNN) performs transposed convolution operations that upsample the input random noise vector, transforming it into a 64×64 adjacency matrix representing a mobility network. The discriminator D (a CNN) takes as input both the generated mobility networks and the real ones from the training set and performs a series of convolutional operations that end up with a probability, for each sample, to be fake or real. D's and G's weights are then backpropagated.

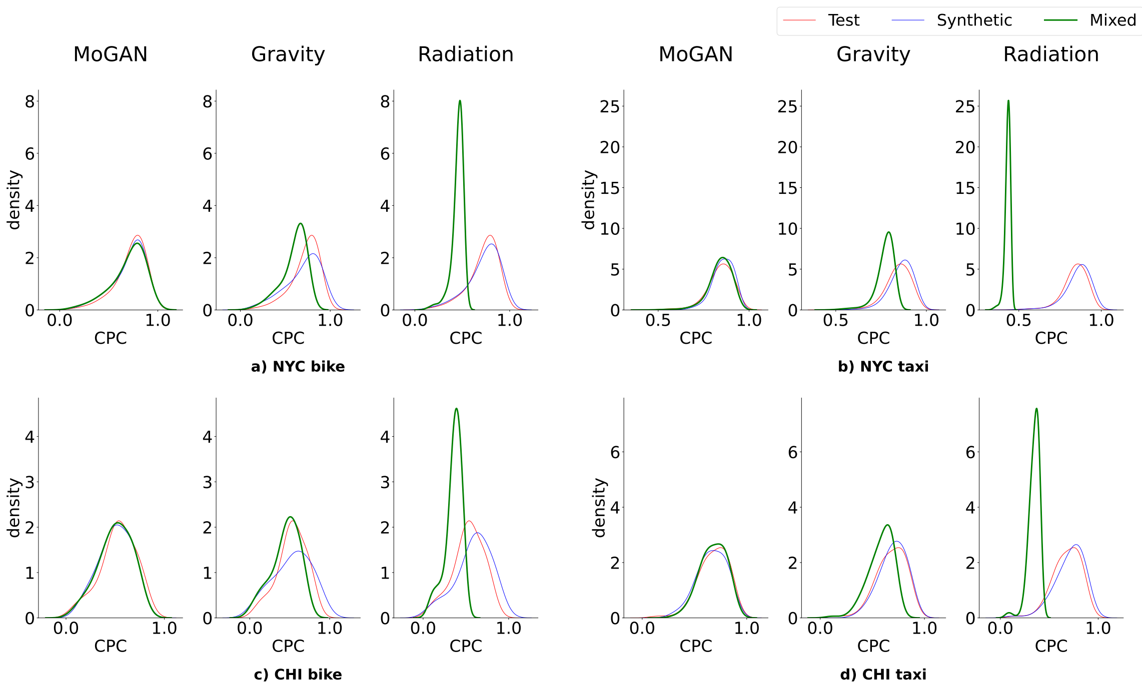

Figure 2: **Results for the CPC.** Distributions of the pairwise CPC distances between mobility networks in the test set (red), synthetic set (blue), and mixed set (green), for the four datasets. For each dataset, we compare the overlap of the distributions of MoGAN and the two baselines (Gravity and Radiation). For both the Gravity model and the Radiation model, the three distributions are significantly different, especially for the latter.

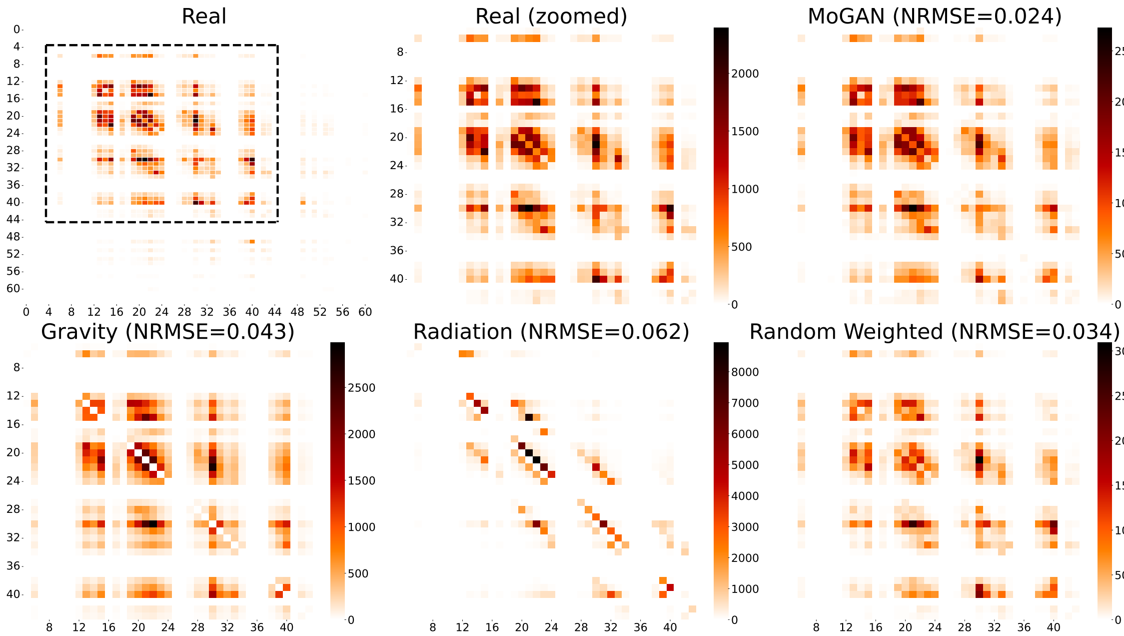

Figure 3: **Visual comparison of the adjacency matrices of the Mobility Networks.** Visualization of the more dense part of the mobility networks of NYC Bikes having the maximum sum of flows observed in the Test Set (Real Zoomed) and of the Mobility Networks having the maximum sum of flows observed in the fake sets produced by all of the other models. Per each generated matrix, we reported the RMSE with respect to the Real matrix. In the top left panel, we show the full 64×64 mobility network and highlight the most dense zones, on which we focus in the other plots of the figure.