ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

# Using Degree Constrained Gravity Null-Models to understand the structure of journeys' networks in Bicycle Sharing Systems.

Remy Cazabet[1] and Pierre Borgnat [2] and Pablo Jensen[2] *

1- Sorbonne Universites, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606,
Paris, France

2- Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique,
Lyon, France

**Abstract**. Bicycle Sharing Systems are now ubiquitous in large cities around the world. In most of these systems, journeys' data can be extracted, providing rich information to better understand it. Recent works have used network based-machine learning, and in particular space-corrected node clustering, to analyse such datasets. In this paper, we show that spatial-null models used in previous methods have a systematic bias, and we propose a degree-contrained null-model to improve the results. We finally apply the proposed method on the BSS of a city.

## 1 Introduction

Bike Sharing Systems (BSS) are now ubiquitous in many cities all over the world. They offer a new type of public transportation system, often considered complementary to more traditional Public Transport. Because most BSS systems are composed of fully automated electronic docking stations, large datasets describing the usage of the system are usually accessible.

These systems can naturally be represented as oriented, weighted networks: BSS stations correspond to nodes, and the number of journeys from a station to another correspond to the weight of the corresponding oriented edge.

Among the many tools of machine learning on graphs, *community detection* (or *node clustering*) has attracted a lot of attention in recent years. Its goal is to find relevant sets of nodes corresponding to a modular organisation of the network at the mesoscopic scale. These sets of nodes are characterised by a high intern density, and a low extern density (more edges inside clusters than between them). One can refer to [1] for an introduction to the topic.

Although community detection can be applied directly on the BSS trip network, and find relevant spatial organisation of cities, a variant of this approach have been proposed recently. The idea is to level out the impact of distance between nodes, in order to discover communities corresponding to a modular

*Acknowledgements. This work is funded in part by the ANR Vel'Innov: Vel'Innov ANR-12-SOIN-0001-02, European Commission H2020 FETPROACT 2016-2017 program under grant 732942 (ODYCCEUS), by the ANR (French National Agency of Research) under grants ANR-15-CE38-0001 (AlgoDiv) and ANR-13-CORD-0017-01 (CODDDE), by the French program "PIA - Usages, services et contenus innovants" under grant O18062-44430 (REQUEST), and by the Ile-de-France program FUI21 under grant 16010629 (iTRAC).

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

organization independent from the spatial properties. In the rest of this paper, we call this process *space-corrected community detection*.

### 1.1 Related Works

Two recent papers have proposed methods to find space-corrected communities. Both of them are extensions of the most widely used method for conventional community detection, the Louvain method [2]. This method is a fast, greedy optimisation of a quality function called the Modularity [3]. Modularity is a measure of the quality of a partition defined as the difference between the fraction of intern-edges in the network and the expected fraction of intern-edges according to a null-model. In the original version, the null-model used is a randomized version of the network preserving the degree of nodes. For space-corrected community detection, the solution consists in choosing a null-model that takes into account the effect of distance between nodes.

In [4], the authors propose to use a gravity-based null-model. We detail this model in section 2. They apply their approach on a network of phone calls in Belgium, and successfully uncover a linguisitic partition, hidden for conventional community detection. Several articles [5, 6] applied this approach on different case studies.

In [7], the authors propose to replace the gravity-based spatial null-model by a radiation-based one [8]. Although radiation-based models have attracted a lot of attention recently in transport modeling, it is not adapted to BSS, because it implies that the closer a pair (source,destination) is, the higher the probability of observing journeys between them. In BSS datasets, few rides are oberved between stations located at less than a kilometer from each other, i.e the deterrence function is not monotically decreasing.

In [6], the gravity-based method presented in [4] was applied on several BSS networks in north america and in London. This community analysis, together with other network analysis tools, was used to demonstrate similarities in the aggregate properties of these systems.

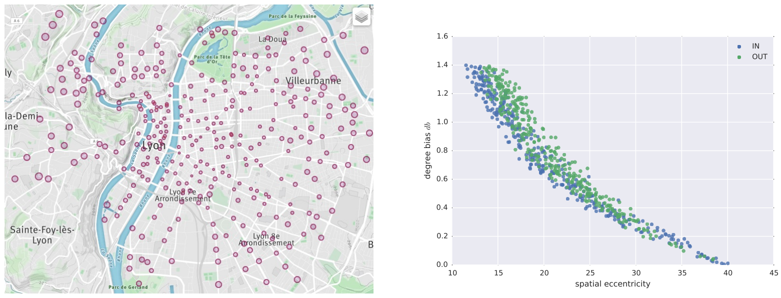
## 2 Definition of a Degree Constrained Gravity-Based Model

The null-model introduced in [4] is based on works coming from the transportation domain, where gravity models have long been used to model the repartition of trips among areas such as cities, countries or districts. It can be expressed as [9]:

$$P_{ij}^{Gra} = n_i n_j f(d_{ij}) \tag{1}$$

with $n_i$ a notion of the *intrinsic strength* of node $i$ (for instance, depending on application cases, its population, number of jobs, parking lot, etc.), $d_{ij}$ is the distance between nodes $i$ and $j$, and $f(d)$ any deterrence function. Instead of being decided *a priori*, the deterrence function can be learned from the data as

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.




(a) Spatial Eccentricity (b) Degree bias

Fig. 1: Illustration of computed spatial eccentricity and degree bias for Lyon's BSS dataset and typical gravity null model.

follows [7]:

$$f(d) = \frac{\sum_{i,j|d_{ij}=d} A_{ij}}{\sum_{i,j|d_{ij}=d} n_i n_j} \tag{2}$$

with $A_{ij}$ the observed flow (number of trips, communications, etc.) between nodes $i$ and $j$.

We can note that in the particular case where the distance has no effect, the deterrence function is a constant function, and the gravity-based model becomes exactly the configuration model

### 2.1 Limits of the gravity-based model

There is a bias when computing directly the gravity-based null-model on a collected spatial network, as it has been done in [6,7] on BSS or any other dataset: the observed *strength* of nodes (number of incoming/outgoing trips) is chosen as a proxy for the *intrinsic strength*. Because the observed strength of a node in a network generated according to the gravity null-model depends both on its intrinsic strength and on its distance to other nodes, this result systematically underestimates the intrinsic strength of nodes with few nodes around (those located at the periphery) and overestimate the strength of those located in the centre. This bias can be checked on any dataset, as we illustrate in Fig.1, by computing the spatial eccentricity of nodes, defined as the average distance to all other stations, and the degree bias *db* for in/out degrees, defined as :

$$db(i) = \frac{deg^{GM}(i)}{deg^{D}(i)} \tag{3}$$

with $deg^{GM}$ the degree according to the gravity model and $deg^{D}$ the degree observed in original data.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

### 2.2 Degree Constrained Gravity-Based model

The null model we propose is searching for a degree constrained solution, i.e a spatial null-model preserving the degrees of nodes. To do so, we take inspiration from the doubly constrained gravity model [10], and adapt it to the case of spatial networks with estimated deterrence function. The intuition is that we are searching for which values of intrinsic attraction of nodes would best explain the observed degree distribution.

The method consists in iteratively estimating the new values for *Incoming estimated intrinsic strength* ($n^{Ieis}$) and *Outgoing estimated intrinsic strength* ($n^{Oeis}$) that satisfies the observed indegrees ($deg^{in}$) and outdegrees ($deg^{out}$) constraints.

We can define them recusively as :

$$n^{Ieis} = \frac{deg^{out}(i)}{\sum_i n^{Oeis} f(d_{ij})}, n^{Oeis} = \frac{deg^{in}(i)}{\sum_i n^{Ieis} f(d_{ij})} \tag{4}$$

And the corresponding Degree Constrained gravity model is:

$$P_{ij}^{DCgrav} = n^{Oeis} n^{Ieis} f(d_{ij}) \tag{5}$$

Starting with initial values $n^{Oeis} = deg^{out}$ and $n^{Ieis} = deg^{in}$, we first compute all values for $n^{Oeis}$, then all values for $n^{Ieis}$, and so on and so forth until the degrees obtained in the gravity model defined in Eq. 5 are close enough to the target network. This process is known to converge [10], and reach a 1% error threashold in a few steps in most cases. For space contraint reasons, we do not discuss this point in this article, we fix a number of iterations, $i = 5$ which is enough to reach convergence in the studied BSS system and similar networks.

### 2.3 Recomputation of the deterrence function

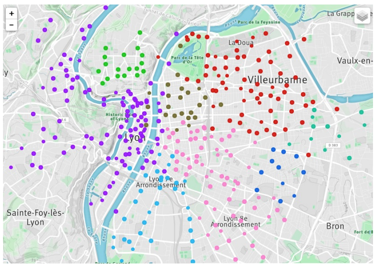
Because the computed deterrence function depends on the *intrinsic strength* of nodes, estimating it using observed degrees as a proxy leads to a biased approximation. By recomputing the deterrence function after each iteration of the algorithm, we can in part correct this bias.

## 3 Application on BSS datasets

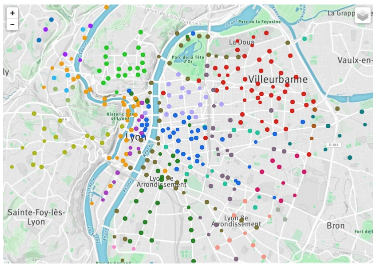
We use a BSS dataset presented in [11], it corresponds to all bicycle trips done in 2011 in the city of Lyon, France, using the bicycle-sharing system called Velo'v [1]. Each trip has a specific origin from a static station in the city, and a destination station which can be any other. It can be studied as a network as in [12], where each station corresponds to a node. The weight on edge $(i, j)$ corresponds to the number of trips done from $i$ to $j$ on a studied period. In this article, we consider all trips done in 2011, week-end and week days together, without filtering hours. The network consists of more than 6 Million edges (trips)

[1] The authors thank JCDecaux (Cyclocity) for having provided access to the Velo'v dataset

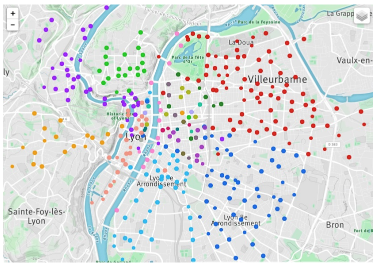
ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.



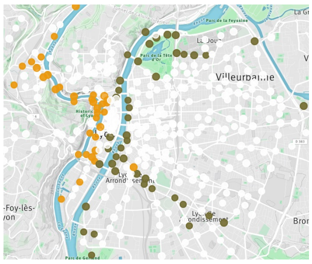
(a) Configuration Model (NG)



(b) DCgravity



(c) Gravity



(d) Details DCgravity

Fig. 2: Communities found on the Lyon BSS dataset, using different null models.

and 343 nodes (stations). We use the great circle distance between stations to learn the deterrence function, although the difference with euclidean distance is negligeable for such short distances.

In Fig. 2, we can observe the communities discovered using the Louvain algorithms and different null-models. Using the usual configuration model, communities correspond to geographical areas of the city, matching more or less *arrondissements* (city districts) of Lyon. Results obtained using Gravity and Degree Constrained Gravity are comparable, but the DC ones are even less spacially constrained. The most remarkable ones, highlighted in Fig. 2(d), correspond to convenient and enjoyable routes along banks of the rivers and parcs. These clusters were only partially discovered using the usual gravity null-model, and arguably correspond to typical usage patterns of Lyon's BSS.

## 4 Conclusion

In this article, we have shown the interested of using a degree-corrected null-model, by focusing on community detection. Such a null-model has many other potential applications: it can be used by bike sharing planners as a model of trip prediction, and as such, can help to predict the activity impact on the global activity of adding or removing stations. It could also be used to estimate the interest of users toward a station, independently of its relative position to others, or to estimate more accurately the influence of distance.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

## References

[1] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.

[2] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[3] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[4] P. Expert, T. Evans, V. Blondel, and R. Lambiotte, "Uncovering space-independent communities in spatial networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 19, pp. 7663–7668, 2011.

[5] Y. Liu, Z. Sui, C. Kang, and Y. Gao, "Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data," *PloS one*, vol. 9, no. 1, p. e86026, 2014.

[6] M. Z. Austwick, O. OBrien, E. Strano, and M. Viana, "The structure of spatial networks and communities in bicycle sharing systems," *PloS one*, vol. 8, no. 9, p. e74685, 2013.

[7] M. Sarzynska, E. A. Leicht, G. Chowell, and M. A. Porter, "Null models for community detection in spatially embedded, temporal networks," *Journal of Complex Networks*, p. cnv027, 2015.

[8] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, no. 7392, pp. 96–100, 2012.

[9] M. Lenormand, A. Bassolas, and J. J. Ramasco, "Systematic comparison of trip distribution laws and models," *Journal of Transport Geography*, vol. 51, pp. 158–169, 2016.

[10] I. Williams, "A comparison of some calibration techniques for doubly constrained models with an exponential cost function," *Transportation Research*, vol. 10, no. 2, pp. 91–104, 1976.

[11] M. Vogel, R. Hamon, G. Lozenguez, L. Merchez, P. Abry, J. Barnier, P. Borgnat, P. Flandrin, I. Mallon, and C. Robardet, "From bicycle sharing system movements to users: a typology of vélo'v cyclists in lyon based on large-scale behavioural dataset," *Journal of Transport Geography*, vol. 41, pp. 280–291, 2014.

[12] P. Borgnat, C. Robardet, P. Abry, P. Flandrin, J.-B. Rouquier, and N. Tremblay, "A dynamical network view of lyon's velo'v shared bicycle system," in *Dynamics On and Of Complex Networks, Volume 2*. Springer, 2013, pp. 267–284.