

ISSN 2451-7100

Publicaciones
del IMAL

IMAL Preprints
#2024-0069

On the evolution of topological connectivity by thresholding of affinities. An application to public transport

Hugo Aimar - Carlos Exequiel Arias - Ivana Gómez

 Publication Date: February 27, 2024

2024

Publisher: Instituto de Matemática Aplicada del Litoral IMAL "Dra. Eleonor Harboure"
(CCT CONICET Santa Fe – UNL)



<https://imal.conicet.gov.ar>
<https://imal.conicet.gov.ar/preprints-del-imal>

Publishing Director: Dra. Estefanía Dalmaso



edalmaso@santafe-conicet.gov.ar



I M A L

ON THE EVOLUTION OF TOPOLOGICAL CONNECTIVITY BY THRESHOLDING OF AFFINITIES. AN APPLICATION TO PUBLIC TRANSPORT

HUGO AIMAR, CARLOS EXEQUIEL ARIAS, AND IVANA GÓMEZ

ABSTRACT. In this paper we use the neighborhood topology generated by affinities between pairs of points in a set, in order to explore the underlying dynamics of connectivity by thresholding of the affinity. We apply the method to the connectivity provided by the public transport system in Buenos Aires.

1. INTRODUCTION

The number of connected components of a space is a topological concept which provides some interesting information on the interactions between the points of the set. On the other hand, the evolution of the number of connected components by thresholding of affinities between points or parts of a set, can be taken as a benchmark of the system under analysis. Roughly speaking an affinity can be thought as the reciprocal of a distance. Nevertheless the mathematical notion of metric or distance on a set is very precise and in applications, frequently, we find non metrizable affinities. Before introducing the definitions, let us consider the weighted undirected affinity graph to which we shall apply our results.

Approximately one third of the total population of Argentina is concentrated around Buenos Aires City. The acronym AMBA (Área Metropolitana Buenos Aires) is used to name these 41 cities with a total population of 17 million people. The public transportation of people between these 41 districts has a unified system of electronic tickets, for buses, trains, subways. The acronym of this system is SUBE (Sistema Único de Boleto Electrónico) Unified System for Electronic Tickets. The global public data provided by SUBE, regarding the number of daily transactions, gives a transportation affinity matrix between any pair of the 41 cities in AMBA. As we shall see this affinity is far from being related to the geographic distance of any couple of districts. In [AAGM22] the authors introduce in this setting the diffusive metrics of Coifman and Lafon [CL06] defined by the spectrum of the Laplace operator in the weighted undirected graph of the 41 cities as vertices and a measure given by the SUBE data of the affinity of any two cities in AMBA. We shall consider three of these matrices corresponding to 2020 in three different moments of the evolution of COVID-19 in Argentina: March, April and June and we shall compare their connection dynamics by thresholding of the affinities.

The aim of this paper is to construct non necessarily metrizable topologies on data sets with affinities and to study the evolution of the number of connected components after thresholding of the affinities. Once the theoretical aspects of the proposal are achieved in sections 2 and 3, in Section 4 we apply the results to the SUBE matrices introduced above.

2020 *Mathematics Subject Classification.* Primary. 54H30. Secondary 05C40.

Key words and phrases. connected components; neighborhood topology; public transport.

2. AFFINITIES AND INDUCED TOPOLOGIES ON ABSTRACT SETS

Given a metric on a set X , a topology is immediately determined on X as the family of all open subsets of X , i.e. the sets $G \subseteq X$ such that for every $x \in G$ there exists a positive r for which the metric ball centered at x with radius r is contained in G . The notion of affinity is, in some heuristic sense, reciprocal to the notion of metric. In fact, two points in the abstract metric space X , which could be two vertices of a metric graph, can be considered to have high affinity if they are close in the metric sense. Let us give the precise definition of what we shall call an affinity on an abstract set.

Definition 2.1. Let X be a set. A positive real valued function A defined on the product set $X \times X$ is said to be an **affinity on X** if

- (1) A is symmetric; $A(x_1, x_2) = A(x_2, x_1)$ for every $x_1, x_2 \in X$; and
- (2) $A(x, x) = +\infty$ for every $x \in X$.

We shall also say that (X, A) is an affinity space.

In [AG18] two of the authors show that a natural metric determined by the affinity A can be defined on X if and only if the affinity A satisfies a transitivity property. This transitivity property reflects an heuristic idea, which is not always true in some interesting models as we shall see in our application in the last section.

This heuristic idea, which sometimes is not valid in concrete situations, is the following: $A(x_1, x_2)$ large and $A(x_2, x_3)$ large implies that $A(x_1, x_3)$ is large. Or more quantitatively $A(x_1, x_2) > \lambda > 0$ and $A(x_2, x_3) > \lambda$, then $A(x_1, x_3) > \lambda/2$. Actually the precise result is that under some extra transitivity condition in A there exists a metric d on X related to A by $A(x_1, x_2) = \varphi(d(x_1, x_2))$ with $\varphi(0) = +\infty$, $\varphi(\infty) = 0$ and φ decreasing. (See [AG18]).

The aim of this section is to construct a topology associated to a given non transitive affinity without going through metric structures on the given set. The main tool is the construction of a topology on X by neighborhood systems defined on X . See Kelley [Kel62]. Let us recall that a topology on X is a subfamily τ of subsets of X that contains both \emptyset and X and is closed under finite intersections and arbitrary unions. With $\mathcal{P}(X)$ we denote, as usual, the set of all subsets of X .

Proposition 2.1. Let X be a set and let $\mathcal{N} : X \rightarrow \mathcal{P}(\mathcal{P}(X))$ be a function that to each x assigns a nonempty family \mathcal{N}_x of subsets of X satisfying the following properties,

- (i) if $U \in \mathcal{N}_x$, then $x \in U$;
- (ii) if U and V belong to \mathcal{N}_x , then $U \cap V \in \mathcal{N}_x$;
- (iii) if $U \in \mathcal{N}_x$ and $V \supset U$, then $V \in \mathcal{N}_x$.

Then the family

$$\tau = \{U : U \in \mathcal{N}_x \text{ for every } x \in U\}$$

is a topology in X .

Proof. Trivially \emptyset and X , both belong to τ , since, by (iii) $X \in \mathcal{N}_x$ for every $x \in X$. If U and $V \in \tau$, then, by (ii), $U \cap V \in \tau$. From (iii) it follows also that $\cup_{\alpha \in \Gamma} U_\alpha \in \tau$ if each U_α belongs to τ . \square

Connectivity is a topological property. As such it can be defined precisely in any topological space (X, τ) .

Definition 2.2. Let (X, τ) be a topological space, i.e. a set X with a topology τ . We say that (X, τ) is **connected** if X can not be decomposed as the union of two nonempty

and disjoint open sets. In other words (X, τ) is connected if do not exist U and V in τ with

- (C1) $U \neq \emptyset, V \neq \emptyset$;
- (C2) $U \cap V = \emptyset$; and,
- (C3) $U \cup V = X$.

Given a subset S of X when (X, τ) is a topological space, the inherited topology in S is given by $\tau_S = \{U \cap S : U \in \tau\}$. We say that a subset S of X is connected if the topological space (S, τ_S) is connected with the above definition. This can also be rephrased with no reference to the hereditary topology as follows. If (X, τ) is a topological space, a set $S \subset X$ is connected if and only if do not exist two open sets U and V in τ such that $S \subset U \cup V, U \cap V = \emptyset, U \cap S \neq \emptyset$ and $V \cap S \neq \emptyset$.

Lemma 2.2. *Let (X, τ) be a topological space and let $\{K_\alpha : \alpha \in \Lambda\}$ be a family of connected sets in X , such that $\bigcap_{\alpha \in \Lambda} K_\alpha \neq \emptyset$. Then $\bigcup_{\alpha \in \Lambda} K_\alpha$ is connected.*

Proof. Take $x \in \bigcap_{\alpha \in \Lambda} K_\alpha$. If $K = \bigcup_{\alpha \in \Lambda} K_\alpha$ were not connected, we would have U and V open and disjoint members of τ such that $U \cap K \neq \emptyset, V \cap K \neq \emptyset$ and $K \subset U \cup V$. Since $x \in K$ and $U \cap V = \emptyset$, then x belongs to one and only one of them. Assume that $x \in U$, so $x \notin V$. Since $V \cap K \neq \emptyset$ then $V \cap K_\alpha \neq \emptyset$ for some $\alpha \in \Lambda$. Hence, $K_\alpha \subset K \subset U \cup V$, but $x \in U \cap K_\alpha$, so that $U \cap K_\alpha \neq \emptyset$. This implies that K_α is not connected. \square

For a given topological space (X, τ) and a given point $x \in X$ the family $C_x = \{S \subset X : x \in S \text{ and } S \text{ is connected}\}$ is nonempty. From Lemma 2.2, we have that $C(x) = \bigcup_{\{C \text{ is connected and } x \in C\}} C$ is connected for every $x \in X$. $C(x)$ is named the **connected component** of (X, τ) containing x . The whole space is connected if and only if there is one and only one connected component. The number of connected components of (X, τ) shall be an important parameter for our later analysis. We denote it by $\kappa(X, \tau)$ or κ when the topological context is clear.

Let us now use Proposition 2.1 in order to construct a topology on X starting from an affinity in X .

Proposition 2.3. *Let (X, A) be an affinity space. For $\alpha > 0$ and $x \in X$, set $E(x, \alpha) = \{y \in X : A(x, y) > \alpha\}$. Let \mathcal{N} be the function that to each $x \in X$ assigns the family*

$$\mathcal{N}_x = \{U \subset X : E(x, \alpha) \subset U \text{ for some } \alpha > 0\}$$

of parts of X . Then, the function \mathcal{N} satisfies (i), (ii) and (iii) in Proposition 2.1. Hence the family

$\tau_A = \{U \subset X : \text{there is a function } \alpha : U \rightarrow \mathbb{R}^+ \text{ such that } E(x, \alpha(x)) \subset U \text{ for every } x \in U\}$, *is a topology on X .*

Proof. It is clear from Proposition 2.1 that we only need to check that \mathcal{N} satisfies (i), (ii) and (iii). Since $A(x, x) = +\infty$, we have that $x \in E(x, \alpha)$ for every $\alpha > 0$. Hence if $U \in \mathcal{N}_x$, we have that $x \in U$ because $E(x, \alpha) \subset U$ for some $\alpha > 0$. This proves (i). Take now U and V two sets in \mathcal{N}_x , then, there exists α_U and α_V two positive real numbers with $E(x, \alpha_U) \subset U$ and $E(x, \alpha_V) \subset V$. Hence, with $\alpha = \max\{\alpha_U, \alpha_V\}$ we have that $E(x, \alpha) \subset U \cap V$ and $U \cap V \in \mathcal{N}_x$. So (ii) is proved. Property (iii) is immediate. \square

Some examples are in order. Notice first that $A_1 \equiv +\infty$ is an affinity in any set X , then the induced topology τ_{A_1} is the trivial $\{\emptyset, X\}$. Hence, the topologies provided by Proposition 2.3 need not to be metrizable. On the other extreme case, if $A_2(x, x) = +\infty$ and $A_2(x_1, x_2) = 1$ for $x_1 \neq x_2$, since $\{x\} = E(x, 2)$ for every $x \in X$, then $\{x\}$ is open

(belongs to τ_{A_2}) for every $x \in X$. So that $\tau_{A_2} = \mathcal{P}(X)$. Notice also that while (X, τ_{A_1}) is connected, (X, τ_{A_2}) is totally disconnected in the sense that the only connected subsets of X are those that contain only one point. It is easy to check that when (X, d) is a metric space, then $A(x_1, x_2) = d^{-1}(x_1, x_2)$ is an affinity on X and the topology τ_A is the metric topology in X . Our example A_1 above shows that not every affinity induced topology is metrizable.

3. THRESHOLDING OF AFFINITIES

Let (X, A) be a given affinity space. For λ positive define the thresholding of A at λ , by

$$A^\lambda(x, y) = \begin{cases} +\infty & \text{for } A(x, y) > \lambda \\ A(x, y) & \text{for } A(x, y) \leq \lambda. \end{cases}$$

Some elementary properties of the family of affinity spaces (X, A^λ) are contained in the next statement.

Proposition 3.1.

- (a) For each $\lambda > 0$, (X, A^λ) is an affinity space;
- (b) for $0 < \lambda_1 < \lambda_2$, $A^{\lambda_1}(x, y) \geq A^{\lambda_2}(x, y) \geq A(x, y)$.

In this way, given an affinity A on the set X , we produce a one parameter family of affinities A^λ on X for $\lambda > 0$. Applying Proposition 2.3 for each $\lambda > 0$ we obtain a corresponding topology τ_λ on X determined by the affinity space (X, A^λ) . Let us denote with τ_A the topology induced on X by A . With the notation introduced in Section 2, for each topological space (X, τ) , $\kappa(X, \tau) = \#\{C : C \text{ is a connected component of } (X, \tau)\}$, where $\#$ denotes the cardinal. With the above family of topologies on X given by the thresholding of A , we obtain a function $\kappa : \mathbb{R}^+ \rightarrow [1, \#(X)]$ given by $\kappa(\lambda) = \kappa(X, \tau_\lambda)$. Set also $\kappa_A = \kappa(X, \tau_A)$.

Proposition 3.2. With the above notation, the following properties hold

- (a) if $0 < \lambda_1 < \lambda_2$, then $\tau_{\lambda_1} \subseteq \tau_{\lambda_2} \subseteq \tau_A$;
- (b) $\kappa(\lambda)$ is nondecreasing and bounded above by κ_A ;
- (c) when X is finite, for λ large enough, we have that $\tau_\lambda = \tau_A$ and $\kappa(\lambda) = \kappa_A$.

Proof. (a) Take $0 < \lambda_1 < \lambda_2$, $U \in \tau_{\lambda_1}$ and $x \in U$. Then, there exists $\alpha > 0$ such that $E_{\lambda_1}(x, \alpha) = \{y \in X : A^{\lambda_1}(x, y) > \alpha\} \subset U$. On the other hand, since from (b) in Proposition 3.1, $A^{\lambda_1}(x, y) \geq A^{\lambda_2}(x, y)$, we have that $E_{\lambda_2}(x, \alpha) \subseteq E_{\lambda_1}(x, \alpha)$. So that $E_{\lambda_2}(x, \alpha) \subset U$ and $U \in \tau_{\lambda_2}$. The same argument shows that $\tau_{\lambda_2} \subseteq \tau_A$ for every $\lambda_2 > 0$. (b) Take again $0 < \lambda_1 < \lambda_2$. Then, since $\tau_{\lambda_1} \subseteq \tau_{\lambda_2}$, every τ_{λ_2} connected set is also τ_{λ_1} connected. In fact, if K is τ_{λ_2} connected, then K can not be written as $K = (K \cap U) \cup (K \cap V)$ with $U, V \in \tau_{\lambda_2}$, $U \cap V = \emptyset$ and $K \cap U \neq \emptyset \neq K \cap V$. Since every U in τ_{λ_1} is also in τ_{λ_2} , then K can not be τ_{λ_2} connected. Take now $C_2(x)$ the connected component in τ_{λ_2} containing $x \in X$, hence $C_2(x)$ is also τ_{λ_1} connected and contains x . Thus $C_1(x) = \bigcup_{\{C: \tau_{\lambda_1} \text{ connected and } x \in C\}} C \supseteq C_2(x)$. Then we have a function assigning to each τ_{λ_2} connected component C_2 a τ_{λ_1} connected component C_1 which is onto. Hence $\kappa(\lambda_2) \geq \kappa(\lambda_1)$. The same argument shows that $\kappa(\lambda) \leq \kappa_A$ for every $\lambda > 0$. (c) Notice that, being $\{A(x, y) : A(x, y) < \infty\}$ a finite set of positive real numbers, it has a maximum M . Hence, for $\lambda > M$ we get $A^\lambda = A$. So that $\tau_\lambda = \tau_A$ and $\kappa(\lambda) = \kappa_A$. \square

Let A be a given affinity on a finite set $X = \{1, 2, \dots, n\}$. In this case A can be seen as an $n \times n$ symmetric matrix with positives entries such that $A_{ii} = +\infty$ for every

$i = 1, \dots, n$. Of course A_{ij} can take the value $+\infty$ in several other pairs (i, j) outside the diagonal as far as the symmetry is preserved. We proceed to construct a simple graph \mathcal{G}_A associated to A as follows. Take $\mathcal{V} = X$ as the set of vertices and $\mathcal{E}_A = \{\{i, j\} : A_{ij} = A_{ji} = +\infty\}$ as the set of edges; $\mathcal{G}_A = (\mathcal{V}, \mathcal{E}_A)$.

The adjacency matrix of this graph is $\mathfrak{A}_{ij} = 1$ if $\{i, j\} \in \mathcal{E}_A$ and $\mathfrak{A}_{ij} = 0$ when $\{i, j\} \notin \mathcal{E}_A$. In other words, \mathfrak{A} has ones where A takes the value $+\infty$ and \mathfrak{A} has zeros where A has finite numbers.

Given a simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the idea of a path joining two vertices i, j is simple and well known. We say that there is a **path joining i and j** if there exist an integer $k > 1$ and a sequence $i = i_1, i_2, \dots, i_k = j$ of vertices in \mathcal{V} such that $\{i_l, i_{l+1}\}$ belongs to \mathcal{E} for every $l = 1, \dots, k - 1$. A subset \mathcal{V}' of \mathcal{V} is said to be **path connected** with respect to \mathcal{G} if for every pair i and j of vertices in \mathcal{V}' there exists a path in \mathcal{V}' joining i and j . A **path connected component** is a maximal path connected subset of \mathcal{V} . In other words \mathcal{V}^* is a path connected component if \mathcal{V}^* is path connected and if $\tilde{\mathcal{V}} \supsetneq \mathcal{V}^*$, then $\tilde{\mathcal{V}}$ is not path connected.

In order to use Python algorithms, such as `number_connected_components`, the following result will be useful.

Proposition 3.3. *Let A be a given affinity on the set $X = \{1, 2, \dots, n\}$ and let \mathcal{G}_A be the induced simple graph. Then*

- (i) *a subset \tilde{X} of X is connected in the sense of Definition 2.2 with respect to the neighborhood topology τ_A in X if and only if \tilde{X} is path connected with respect to \mathcal{G}_A ;*
- (ii) *the number of topological connected components in (X, τ_A) coincides with the number of path connected components with respect to \mathcal{G}_A .*

Proof. Notice first that (ii) follows from (i). On the other hand, to prove (i) we only need to show that the entry $\mathfrak{A}_{ij} = 1$ if and only if any τ_A open set containing i also contains j . But $\mathfrak{A}_{ij} = 1$ if and only if $A_{ij} = +\infty$. Take U a τ_A open set such that $i \in U$. Then there exists $\alpha > 0$ such that $E(i, \alpha) \subset U$. Since $E(i, \alpha) = \{k : A_{ik} > \alpha\}$ and $A_{ij} = +\infty$ we have that $j \in E(i, \alpha) \subset U$. So that $j \in U$. On the other hand, if j belongs to any open set containing i , then $j \in E(i, \alpha)$ for every $\alpha > 0$. This implies $A_{ij} = +\infty$ and $\mathfrak{A}_{ij} = 1$. \square

Before introducing, in the next section, the application of the above analysis, let us illustrate some simple examples in the real line in order to get some insight regarding the relation between the shape of the curve $\kappa(\lambda)$ as a function of the threshold parameter and the distribution of the data points. The four examples are subsets of real numbers and the affinity is the metric one, that is $A(x, y) = \frac{1}{|x-y|}$. For the first example, see Figure 1, consider the set $X_1 = \{f_1(i) = \log_2 i : i = 1, 2, \dots, 20\}$. For the second, we take the set $X_2 = \{f_2(i) = \sqrt{i-1} : i = 1, 2, \dots, 20\}$, for the third $X_3 = \{f_3(i) = 20(1 - \frac{1}{i}) : i = 1, 2, \dots, 20\}$ and for the last one we take $X_4 = \{f_4(i) = 20(1 - (\frac{5}{6})^{i-1}) : i = 1, 2, \dots, 20\}$. Let us notice that the concave shapes of $\kappa(\lambda)$ correspond to convergence of the generating sequences as in cases X_3 and X_4 .

4. APPLICATION TO THE PUBLIC TRANSPORT IN BUENOS AIRES

In Figure 2, we show schematically the relative geographic locations of the 41 cities in AMBA. The numbers in each one of the districts will be used as the indices $i = 1, \dots, 41$ that identify the vertices of the graph and the affinities that we shall consider.

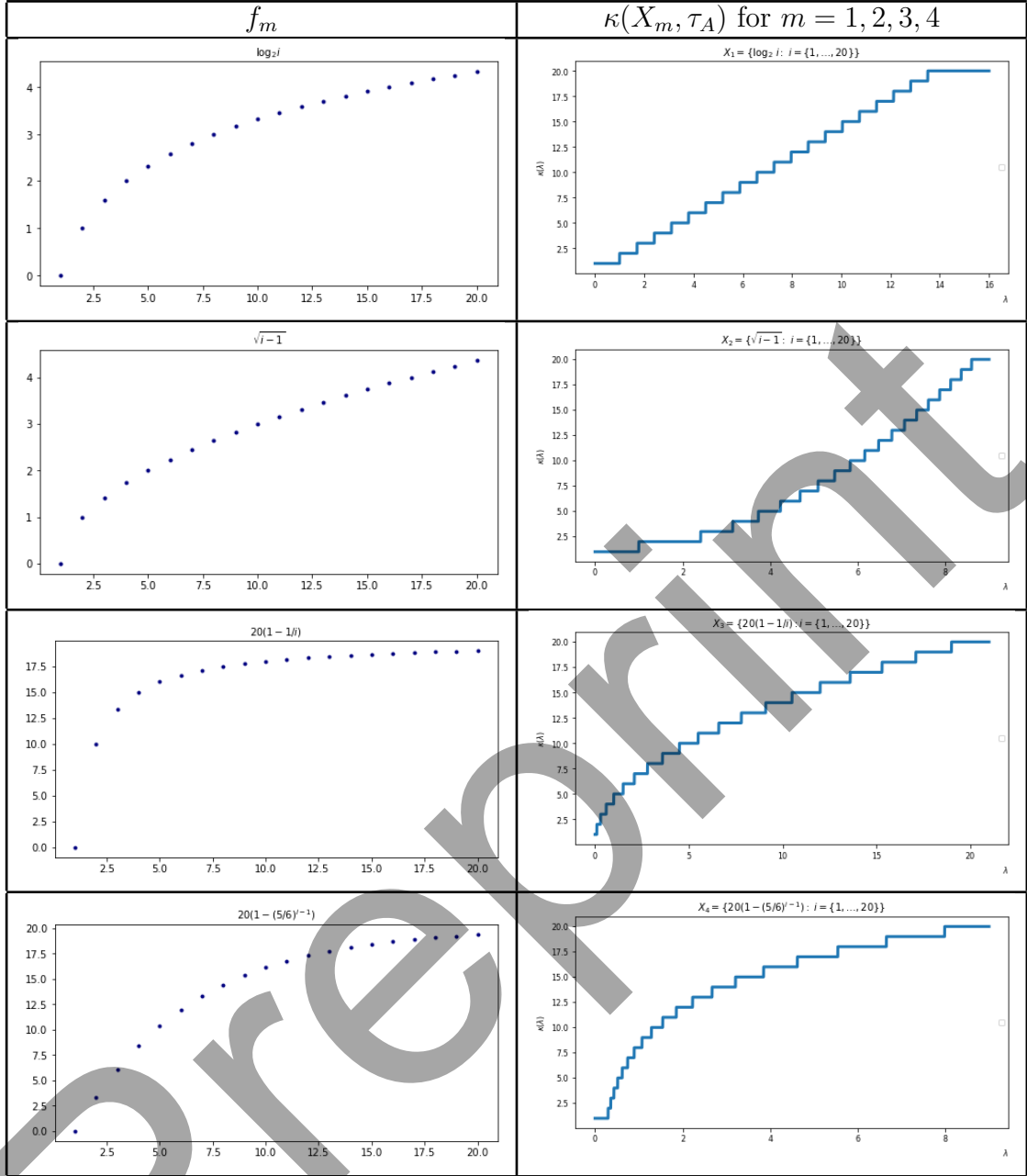


FIGURE 1. $\kappa(\lambda)$ (on the right) for different distribution data points (on the left).

The normalized affinity matrices that we shall consider correspond to three different days of 2020 at the beginning of the circulation of COVID-19 in Argentina. The first A^1 , corresponds to March 4th 2020 before the detection of the spread of COVID-19, with the public transport system working normally. The second, A^2 , corresponds to April 8th 2020, with the beginning of the circulation restrictions imposed by the health administration. The third, A^3 , is constructed with the data corresponding to June 3rd 2020, when the circulation restrictions started to weaken. In all these cases we shall apply the algorithm introduced in the previous sections and we shall obtain the corresponding functions $\kappa(\lambda)$ that describe the evolution of the number of connected components as a function of the threshold parameter λ in the three situations described.

The matrix A^1 , corresponding to March 2020, is the 41×41 normalized matrix with A^1_{ij} measuring the intensity (number of passengers) going from city i to city j and viceversa

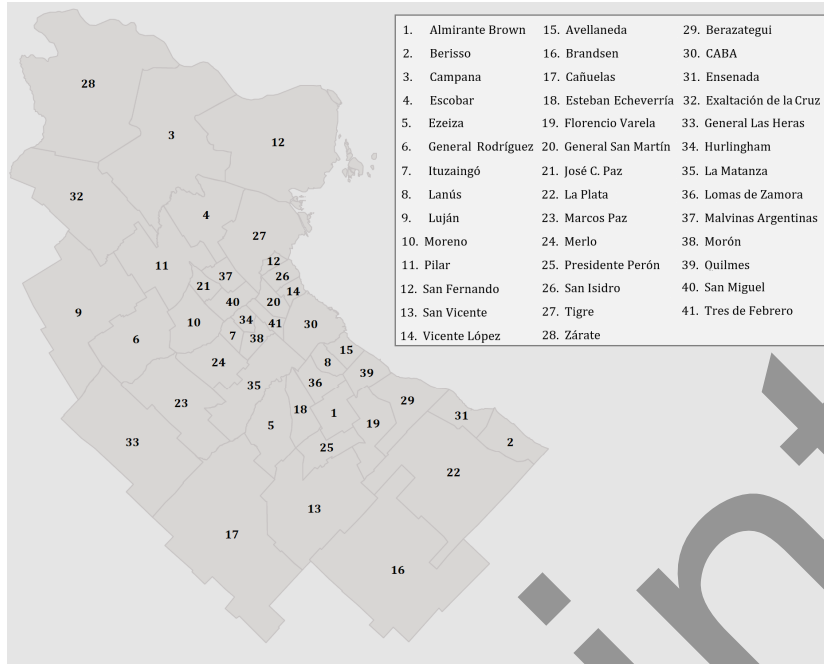


FIGURE 2. Map of the 41 cities of AMBA

by March 4th 2020, is given by Figure 3. The matrix A^2 , Figure 5, is constructed in the same way with the data corresponding to April 8th 2020. The third A^3 , Figure 7, is based on the SUBE data by June 3rd 2020. Figure 4, Figure 6 and Figure 8 show the shapes of the connectivity curves $\kappa_i(\lambda)$, $i = 1, 2, 3$. The drastic changes of the number of passengers between the three different dates, as could be expected do not have a great impact in the shapes of these three connectivity dynamics.

Among the four distributions of data points shown in Figure 1 the one that better fits the shapes of κ_i is that of X_3 . See Figure 9.

Let us finally observe that if we use affinities built only on geographic data, the behavior of the curves $\kappa(\lambda)$ instead of concave look roughly convex. In Figure 10 the affinity between districts i and j is determined by the length of the shared boundaries. In Figure 11, the affinity $A_{ij} = \frac{1}{d_{ij}}$, with d_{ij} the Euclidean distance of the geographical centers of districts i and j .

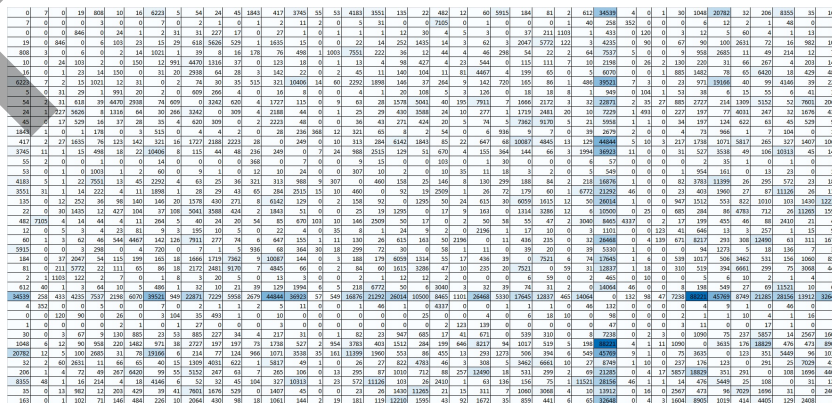


FIGURE 3. Unnormalized affinity matrix A^1 corresponding to March 2020

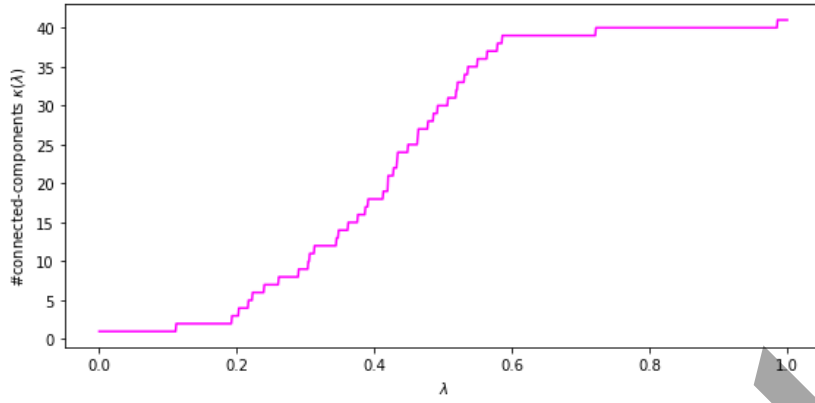


FIGURE 10. Connected components for matrix $A_{ij} = l_{ij}$ where l_{ij} the length of the boundaries shared by the districts i and j of AMBA.

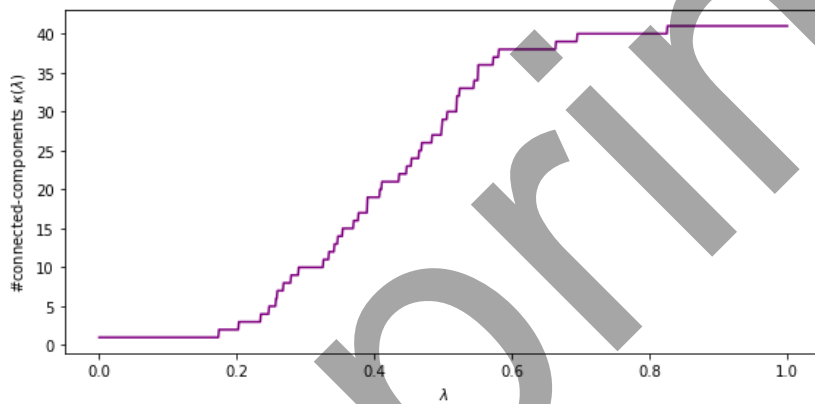


FIGURE 11. Connected components for matrix $A_{ij} = \frac{1}{d_{ij}}$ where d_{ij} is the Euclidean distance between the geographical centers of the districts i y j in AMBA.

STATEMENTS AND DECLARATIONS

Funding. Consejo Nacional de Investigaciones Científicas y Técnicas, grant PIP-2021-2023-11220200101940CO.

Conflicts of interest/Competing interests. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Availability of data and material. Not applicable.

Acknowledgements. This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas-CONICET and Universidad Nacional del Litoral-UNL, in Argentina.

Affiliation. INSTITUTO DE MATEMÁTICA APLICADA DEL LITORAL “DRA. ELEONOR HARBOURE”, CONICET, UNL.

Address. IMAL, Streets F. Leloir and A.P. Calderón, CCT CONICET Santa Fe, Predio “Alberto Casano”, Colectora Ruta Nac. 168 km 0, Paraje El Pozo, S3007ABA Santa Fe, Argentina.

E-mail: haimar@santafe-conicet.gov.ar

carias@santafe-conicet.gov.ar

ivanagomez@santafe-conicet.gov.ar (corresponding author)