Persistent Homology applied to location problems

Driss Bennis^{1,a}, Fouad Gharib^{1,b} and Ghita Lebbar^{1,c}

¹Faculty of Sciences, Research Center CeReMAR, Mohammed V University in Rabat, Morocco

Abstract. Different approaches to solve location problems in transport and logistics have been developed in the literature. This article introduces a new approach using the concept of persistent homology which has been proved to be an efficient method in topological data analysis; and has been served as an alternative new tool in many and various research areas such as image processing, material science and biological systems. Precisely, inspired by the notions of the first homology groups and the persistent homology which mainly describe the behaviour of connectivity relation between elements during a filtration of specific topological spaces; we develop a new method and approach for the treatment of facility location—network design problems.

1 Introduction

The concept of persistence is first introduced by Edelsbrunner, Letscher, and Zomorodian, in [6]; then refined by Carlsson and Zomorodian in [21]. It was used to provide a rigorous response to the following problem: For a parameterized family of spaces, those topological features which persist over a significant parameter range are to be considered as signal with short-lived features as noise (see [7]). Since then, persistent homology has gathered an enormous attention as a fundamental tool in Topological Data Analysis (TDA). It has proved utility and significance to solve different problems in several domains such as biology, image processing and sensor networks (see, for instance, the papers [1,13,16,19]). For a general background on persistent homology, worth to mention the book by Zomorodian [20], survey by Edelsbrunner, Harer [5] and the survey by Ghrist [7]. The mentioned authors offer an introduction to persistent homology which can be useful for mathematical researchers based on topology.

The main interest of our work is related to location decision using persistent homology. Location decision is a critical element in transport planning. In general, decision makers are asked to select a site (location) among several existing sites (cities, district, airports, ...) on a given network in order to satisfy a number of customers while considering some constraints. Several researchers have developed different mathematical models for location decision which are mainly based on the classical optimisation methods. The present paper provides a new treatment of location problems using the

concept of persistent homology. Therefore, for a practical understanding, the second section presents some basic notions and background related to persistent homology. Section 3 highlights the problem related to transport design. In section 4, we will focus on our contribution, based on a location decision issue and the adopted approach using persistent homology to solve the problem.

2 Mathematical background

In this section, we give a short presentation on some basic notions concerning simplicial complexes, simplicial homology and persistent homology (more details can be found in [5,7,20]).

2.1 Simplicial complexes

An abstract simplicial complex or simply a simplicial complex is a set K together with a collection S of subsets of K which satisfies the two following conditions:

- S contains all singletons $\{v\}$ with $v \in K$.
- S is closed under subsets: if $\tau \subseteq \sigma$ and $\sigma \in S$, then $\tau \in S$.

An element σ of S of cardinal a positive integer n will be called an (n-1)-simplex or simply a simplex. So, 0-simplices are the singletons which are usually called points; and 1-simplices consist of two elements of K called edges. There are also the 2-simplices called triangles and so on. In fact, any simplicial complex can be represented in a real affine space using points, edges,

^adriss.bennis @um5.ac.ma; ^bfouad.gharib1994@gmail.com; ^cghitaa.lebbar@gmail.com

triangles and so on. Thus, using the induced topology on this geometric representation, one can talk about the connected components of a simplicial complex. Namely, this notion is of great interest in simplicial homology theory. Next section shows that the number of connected components of a simplicial complex is the dimension of some kind of vector spaces. Several examples of simplicial complex can be found in the literature. The most useful ones are geometric simplicial complexes, complexes constructed from graphs, complexes constructed from a geometric space (Nerve of a covering).

2.2 Simplicial homology

Simplicial homology is a powerful tool for description and characterization of some topological features such as connectivity and the existence of holes. Given K as a simplicial complex. The $\mathbb{Z}/2\mathbb{Z}$ vector space generated by the p-dimensional simplices of K is denoted $C_p(K)$. It consists of all p-chains, which are formal sums

$$c = \sum_{j} \gamma_{j} \sigma_{j} \tag{1}$$

where the γ_j are 0 or 1 (in $\mathbb{Z}/2\mathbb{Z}$) and the σ_j are p-simplices in K. For a positive integer p, consider the linear map $\partial\colon C_p(K)\to C_{p-1}(K)$, called boundary map, defined on n-simplices as follows: For every p-simplex σ , $\partial(\sigma)$ is the formal sum of the (p-1)-dimensional faces (i.e., subsets of σ of cardinal p). Thus, the boundary of the chain c (i.e., $\partial(c)$) is obtained by extending ∂ linearly, and so

$$\partial(\mathbf{c}) = \sum_{i} \gamma_{i} \partial(\sigma_{i}), \tag{2}$$

where we understand that the addition between coefficients is modulo 2, *i.e.*

$$1 + 1 = 0.$$
 (3)

It is not difficult to check that:

$$\partial \circ \partial = \partial^2 = 0 \tag{4}$$

The p-chains that have boundary 0 are called p-cycles. They form a subspace Z_p of C_p . The p-chains which are the boundary of (p+1)-chains are called p-boundaries and form a subspace B_p of C_p . The fact that:

$$\partial^2 = 0 \tag{5}$$

implies that $B_p \subseteq Z_p$. The quotient group

$$H_p(K) := Z_p/B_p \tag{6}$$

is the p-th simplicial homology group of K with $\mathbb{Z}/2\mathbb{Z}$ coefficients. The dimension of $H_p(K)$ is called the p-th Betti number or simply the p-Betti number of K and it is denoted $\beta_p(K)$ or just β_p .

Overall, the homology groups describe spaces through their Betti numbers β_p . Precisely, β_0 refers to the number of connected components of the given simplicial complex, β_1 refers to the number of holes of the simplicial complex.

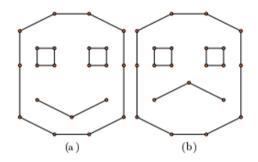


Figure 1: Two simplicial complexes with similar Betti numbers: β_0 =4 and β_1 =3.

2.3. Persistent homology

Persistent homology is considered as a powerful tool in topological shape analysis which describes the changes in homology that occur to an object with respect to a given parameter.

For a simplicial complex K, a filtration F of K is a finite sequence of complexes:

$$F: \emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m = K \tag{7}$$

The idea of persistent homology consists of, instead of examining the homology $H_i(K;F)$ of each individual term K_i of the filtration F, the iterated inclusions are examined:

h:
$$H_i(K;F) \rightarrow H_i(K;F)$$
 for all $i < j$. (8)

These maps reveal which features persist. For a clear understanding, let us give a simple example visualised in the above figures:

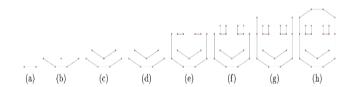


Figure 2: Filtration F₁ of Happy face P₁

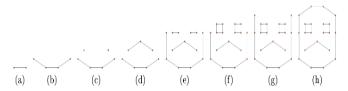


Figure 3: Filtration F₂ of Sad face P₂

Although both pictures P_1 and P_2 are different, they have identical betti numbers. Persistent homology reveals the changes in homologies in the given filtration and prove the difference between both objects. Usually, the parameter intervals arising from the basis for $H_*(K;F)$ are visualised in the form of a barcode. Precisely, a barcode is a graphical representation of $H_k(K;F)$ as a collection of horizontal line segments in a plane whose horizontal axis corresponds to the parameter and whose vertical axis represents an arbitrary) ordering of homology generators (i.e., generators of $H_*(K;F)$). For example, the barcode corresponding to the filtration F_1 above is of the form:

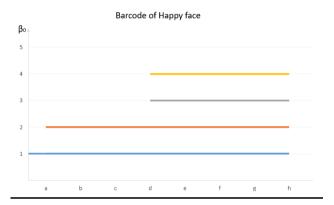


Figure 4: Barcode of happy face

And the barcode corresponding to the filtration F₂ is of the form:

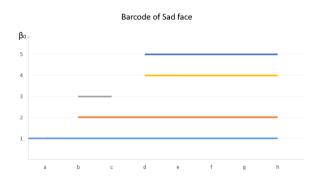


Figure 5: Barcode of sad face

The two barcodes illustrate clearly the difference.

3 Location decision

Location decisions have been analysed since 1909 by Weber [18] who formulated a theory to locate an industry with a minimum transportation cost of raw and final products.

Mainly, the acquisition of a new facility within a network is a costly, time-sensitive target. Hence, experts on transport design focus on solving the following issues:

- The optimal choice of nodes to locate facilities
- Minimize the travel cost
- Reduce travel time
- Optimise networks: Infrastructure (roads, bridges, tunnels...), telecommunication (cables) water pipes, electricity cables.

Hence, it can be understood that the choice of location is not arbitrary but requires strategic and expert planners which are challenged by several factors. As an aid for making decisions, the domain of facility location provides tools and methods for finding optimal locations using mathematical models with respect to quantifiable factors. O'Kelly [11] is one of the researchers who proposed a model for hub facility location based on the optimisation method.

4 Persistent Homology as a tool for facility location

This paper reports a new formulation for facility location by developing algorithms inspired from persistent homology. We treat the problem given in [17] through our method which can be applied in different contexts.

The starting point will be a given network consisting of n districts annotated by $(A_1, A_2, A_3, ..., A_n)$ (nodes) in an area of a city with information about distances $d_{i,j}$ between districts A_i and A_j , and numbers of population n_i in each node A_i . A chain of supermarkets decided to open a facility within the districts and they need to examine the most optimum location to serve the whole area.

4.1 Method

The proposed method will follow a procedure which will be explained as follows:

First, it is necessary to fix some nodes as candidates where the facility can be located. The choices are based on transport design criteria which are adopted by planning experts. Then, for a fixed candidate A_i , ponderate each other node A_j by the following coefficient which reflect the "importance" of the node A_j with respect to A_i (see below for the explanation):

$$X_j^i = \frac{1}{n_i} \frac{d_{i,j}}{n_j} \tag{9}$$

In fact, fixing two candidates A_m and A_n, and

consider two nodes A_{e} and $A_{\rm f}$. We discuss the following particular situations:

- If we suppose that our network consists only of three nodes A_m , A_n and A_e . Then, A_m is the optimal solution if $X_e^m \leq X_e^n$. Assume, for example, $d_{e,m} = d_{e,n}$, the inequality $\frac{1}{n_m} \leq \frac{1}{n_n}$ can be explained by the phrase : the bigger is n_m , the more attractive is A_m . On the other hand, if $n_m = n_n$, $d_{e,m} \leq d_{e,n}$ explain the chose of A_m as the optimal solution.
- Now suppose that our network consists only of four nodes A_m , A_n , A_e and A_f . Then, A_m will be the optimal solution if:

$$\frac{X_e^m + X_f^m}{2} \le \frac{X_e^n + X_f^n}{2}$$

For example, if $n_m = n_n$ and $n_e = n_f$, the means above can be reduced to the means of distances. So clearly the optimal solution corresponds to the smallest mean. Also, the ratios $\frac{1}{n_e}$ and $\frac{1}{n_f}$ are there since the number of clients has to affect the number in some sense.

Basing on the above observations we construct a complex $K_{m,r}$ associated to a candidate A_m and for a fixed strictly positive real r as follows (we focus only 0-simplices and 1-simplices):

- The 0-simplices are all the points A_j which are different from A_m .
- Two nodes A_i and A_j is related by an edge (i.e., $\{A_i, A_j\}$ is a 1-simplex) if $\frac{X_i^m + X_j^m}{2} \le r$.

Clearly, for r < r', $K_{m,r} \subseteq K_{m,r'}$. This means that we can construct a filtration $K_{m,r1} \subseteq \cdots \subseteq K_{m,rs}$, where r1,..., rs is an increasing sequence of strictly positive real numbers. We choose such a sequence in the way that at the last real all complexes (for all candidates) have one connected component (i.e., all nodes are connected). In fact, passing from a complex to the next one in a filtration we obtain more edges (the nodes will be "more" connected) and then less connected component (the provided sequence of 0-Betti numbers decreases). Then, we compare between the sequences of 0-Betti numbers. The one which reaches 1 at the earliest time is the most attractive between all the other ones and then can be adopted as a solution following this approach. This can be seen easily when comparing the provided barcodes.

4.2 Algorithms

There are softwares which deal with the

computation of persistent homology in various context. In our context we made our own programs and so all calculation is carried out using the C++ programming by developing the following algorithms. Indeed, to determine the 0-Betti number of a complex associated to one of the candidates, we follow a procedure of three steps as follows:

Fixe a candidate Ai and a real number r.

We need at first an algorithm to calculate all X_i^i .

Algorithm 1

For j=1, j≠i to n do
$$X_j^i = \frac{1}{n_i} \frac{\text{dij}}{n_j};$$

Then, we construct the simplicial complex $K_{i,r}$ associated to r using the following algorithm.

Algorithm 2

1.
$$\mathscr{C} = \{\{A_j\}/j \in \{1,...,n\} \text{ and } j \neq i\}$$

2. For
$$j=1$$
, $j \neq i$ to n do

3. For
$$j'=1$$
, $j' \neq i$ to n do

4. If
$$(\frac{1}{C_{jj'}^i} < r) \in \mathcal{C} \cup \{\{A_j, A_{j'}\}\};$$
 where $C_{jj'}^i = \frac{X_j^i + X_{j'}^i}{2}$

5. End for

6. End for

Finally, we determine and calculate the number of connected components of $K_{i,r}$ using the algorithm beneath.

Algorithm 3

1. Initialize a matrix $M=(m_{ij})_{1 \le j \le n, 1 \le i \le n}$ and $m_{11}=A_1$;

2. i = 0;

3. While $(n\neq 0)$

4. j=0, $k_1=0$;

5. While $(m_{ik_1} \neq \emptyset)$

6. $k_2=0, A=m_{ik_1}$;

7. While $(k_2 \neq n, \text{ and } A_{k_2} \notin M)$

8. if $(\{A_{k_2},A\})$ is 1-simplex

then

9. j++;

 $m_{ij} = A_{k_2} ;$

11. end if

- $\begin{array}{ccc} 12. & & k_2++\;;\\ 13. & & \text{end While}\\ 14. & & k_1++\;;\\ \end{array}$
- 15. end While
- 16. s =the number of boxes filled in the i^{th} line of the Matrix M;
- 17. n = n-s;
- 18. End While
- 19. β_0 = The number of lines containing at least one point;

one point; 20. Return β_0 ;

4.2 Example

Suppose that we have a network consisting of 8 districts annotated by $(A_1, A_2, A_3,..., A_8)$ (eight nodes; see [17]):

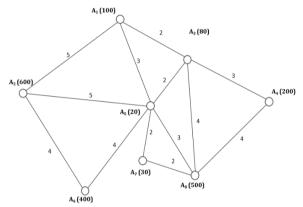
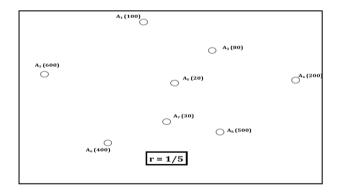
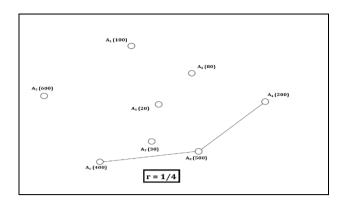
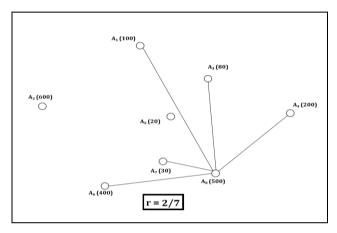


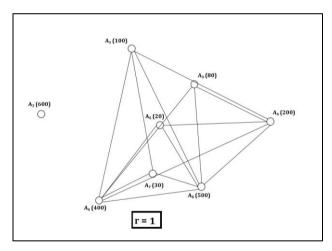
Figure 6: Network of 8 districts

We Fix three candidates A_2 , A_5 and A_8 . Then, we construct the associated barcode for each candidate. Below, the obtained complexes in a filtration associated to A_2 . We used the following sequence of real numbers: $r_1=1/5$; $r_2=1/4$; $r_3=2/7$; $r_4=1$.

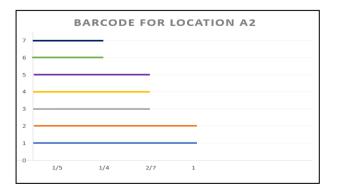


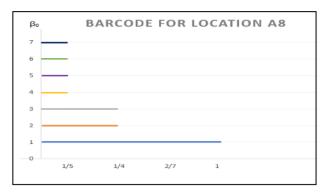






Then we obtain the following barcodes:







Finally, one can observe clearly that the A_5 is the desired solution.

5 Conclusion

We have applied the proposed method to an example of eight districts; which has provided a successful location decision. Thus, this article is the first significant step made to solve location problems based on the concept of persistent homology. It will certainly attract the attention of several researchers. We expect further research to develop the solution for more complex situations with multiple parameters and for different types of location problems.

References

- P. Bendich, J. S. Marron, E.Miller, A. Pieloch and S. Skwerer. "Persistent homology analysis of brain artery trees." Ann. Appl. Stat. 10 (2016): 198-218.
- 2. C. Cocking, *Solutions to facility location–network design problems*; Doctoral thesis at University of Heidelberg (2008).
- 3. Z. Drezner, G.O. Wesolowsky. "Network design: selection and design of links and facility location." Transp. Res. **37** (2003): 241–256.
- 4. H. Edelsbrunner, J. Harer, *Computational Topology An introduction*; American Mathematical Society **26** (2009).
- 5. H. Edelsbrunner, J. Harer. "Persistent homology a survey." Surveys on Discrete and Computational

- *Geometry. Twenty Years Later*, Contemporary Mathematics **453** (2008): 257–282.
- H. Edelsbrunner, D. Letscher, A. Zomorodian. "Topological persistence and simplification." Discrete Comput. Geom. 28 (2002): 511-533.
- R. Ghrist. "Barcodes: the persistent topology of data." Bull. Amer. Math. Soc. (N.S.), 45 (2008): 61–75.
- 8. A.A. Kuehn, M.J. Hamburger. "A heuristic program for locating warehouses." Manag. Sci. **9** (1963).
- 9. T.L. Magnanti, R.T. Wong. "Network design and transportation planning: models and algorithms." Transport. Sci. **18** (1984): 643–666.
- M.T. Melo, S. Nickel, F. Saldanha-da-Gama. "Facility location and supply chain management

 A review." Eur. J. Oper. Res. 196 (2009) 92-106.
- 11. M.E. O'Kelly. "The location of interacting hub lacilities." Transport Sci. **20** (1986): 92-106.
- S.H. Owen, M.S. Daskin. "Strategic facility location: a review." Eur. J. Oper. Res. 111 (1998): 423–447.
- 13. J.A. Perea, J. Harer. "Sliding windows and persistence: An application of topological methods to signal analysis." Found. Comput. Math. (2013): 1–40.
- R. Rahmaniani, A. Ghaderi. "A combined facility location and network design problem with multi-type of capacitated links." Appl Math. Model. 37 (2013): 6400–6414.
- C. S. ReVelle, A. H. Eiselt. "Location analysis: A synthesis and survey." Eur. J. Oper. Res. 165 (2005): 1–19.
- 16. G. Singh, F. Memoli, G. Carlsson. "Topological methods for the analysis of high dimensional data sets and 3D object recognition." Eurographics Symposium on PointBased Graphics, eds Botsch M, Pajarola R (Eurographics Association, Geneva), (2007) 91-100.
- J. P. Škrinjar, N.Brnjac, D. Žgaljić. "Location problems in transport network." J. Mar. Res. 28 (2014): 94-98.
- 18. C. J. Friedrich, Alfred Weber's Theory of the Location of Industries. University of Chicago Press (1929).
- 19. K. Xia, Z. Li, L. Mu. « Multiscale persistent functions for biomolecular structure characterization." Perprint, available at arXiv:1612.08311 (2016).
- 20. A.J. Zomorodian, *Topology for computing*, Cambridge University Press (2005).
- 21. A.J. Zomorodian, G. Carlsson. "Computing Persistent Homology." Discrete Comput. Geom. **33** (2005): 249–274.