A Short Exposition of Topological Applications to Security Systems

Dimitrios Panagopoulos, Sotirios Hassapis

Abstract

In this article several practical applications of algebraic topology are presented. After a short technical review of the necessary theory applications to sensor networks are presented. A very short reference of applications to data analysis follows.

Keywords: Persistent homology, Euler integration, sensor coverage, data analysis

1 Introduction

For many years algebraic topology has been considered as an abstract mathematical field with none or few practical applications.However many astract mathematical ideas have found unexpected applications in real-life problems. This is also the case with algebraic topology during the past couple of years. Theories and techniques created by mathematicians in order to answer abstract problems are being used to answer problems such as protein docking, image analysis, data analysis and space coverage by sensor networks.

This article focuses, mainly, on the last two cases i.e. applications of algebraic topology to sensor networks and data analysis. A brief introduction to the necessary mathematical background is given in the beginning of the article. In the second section applications to sensor networks are presented while in the third section approaches related to data analysis are discussed.

2 Mathematical background

2.1 Algebraic homology

In this section we give a short introduction to Homology theory. We begin with the definition of a (simplicial) complex.

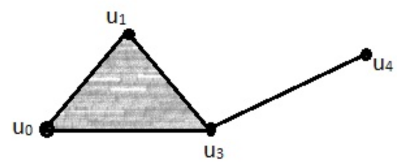

Figure 1: A simple example of a simplicial complex

Definition 2.1. *Let V be a finite set. A collection $\mathcal{K}$ of subsets of V is called a complex if $\alpha \in \mathcal{K}$ and $\beta \subseteq \alpha$ implies $\beta \in \mathcal{K}$.*

A set in $\mathcal{C}$ with $k+1$ elements is called a k-simplex and we define its dimension to be k. If $\beta \subseteq \alpha$ then we call β a face of α. We call β a proper face if in addition $\beta \neq \alpha$.

In this article complexes will, mostly, refer to geometric objects. A 0-simplex will be a point, a 1-simplex an arc, a 2-simplex a triangle and so on. In figure 1 there is an example of a complex which is consisted of a triangle $\{u_0, u_1, u_2\}$ and a line $\{u_2, u_3\}$ (and of course all their faces).

Given a complex $\mathcal{K}$, we define $C_n(K)$ to be the vector space whose base is the set of n-simplices of $\mathcal{K}$ with coefficients over a field. The base element that corresponds to the n-simplex $\{u_0, \dots, u_n\}$ is denoted by $[u_0, \dots, u_n]$.

For example for the complex of figure 1 we have that $C_2(K)$ is a one dimensional vector space with base $[u_0, u_1, u_2]$ while $C_1(K)$ is a four dimensional vector space with base $[u_0, u_1], [u_0, u_2], [u_1, u_2], [u_3, u_4]$. For each n we define the boundary operator $\vartheta_n : C_n(K) \to C_{n-1}(K)$ to be the linear map with

$$\vartheta([u_0, \dots, u_n]) = \sum_{i=0}^{n} (-1)^i [u_0, \dots, \widehat{u_i}, \dots, u_n]$$

where $\widehat{u_i}$ means that u_i is deleted. A straight forward calculation [8, p. 105] verifies that $\vartheta_{n-1} \circ \vartheta_n = 0$. Hence we have a chain complex

$$C_\bullet(K) : \cdots \xrightarrow{\vartheta_{n+2}} C_{n+1}(K) \xrightarrow{\vartheta_{n+1}} C_n(K) \xrightarrow{\vartheta_n} C_{n-1}(K) \xrightarrow{\vartheta_{n-1}} \cdots$$

Definition 2.2. *Let $C_\bullet(K)$ be a chain complex as above. The n-th Homology group $H_n(K)$ is the group quotient $Ker\vartheta_n / Im\vartheta_{n+1}$. The rank of the n-th homology group is called the n-th betti number.*

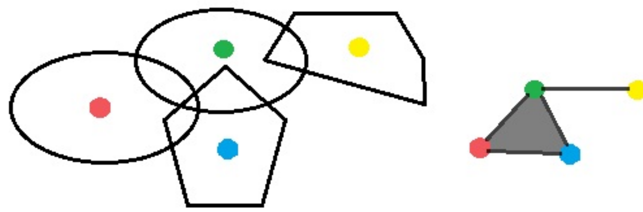

Figure 2: Four sets in the plane and their corresponding nerve

Let K be a complex and A a subcomplex contained in it. Then we can construct the quotient K/A. Geometrically K/A is obtained from K by collapsing A to a single point. The homology of K/A is called the relative homology of K with respect to A and is symbolized by $H_\bullet(K, A)$.

Finally, for our purposes we will require the following two constructions.

Definition 2.3. *[3, 5] Let $\mathcal{S}$ be a set of sets. We define the nerve $Nrv\mathcal{S}$ of $\mathcal{S}$ to be the set of subsets of $\mathcal{S}$ with no empty intersection:*

$$Nrv\mathcal{S} = \{\mathcal{X} \subseteq \mathcal{S} : \bigcap \mathcal{X} \neq \emptyset\}$$

The nerve is always a complex. In figure 2 an example with four sets in $\mathbb{R}^2$ is given and on the right hand side the corresponding complex is depicted. If $\mathcal{S}$ contains disks, then the corresponding complex is called the Čech complex.

Definition 2.4. *[3] Let $V = \{v_1, \ldots, v_n\}$ be a set of points of a metric space (X, d). Then for a real number $\epsilon > 0$ we define the Rips (or Rips-Vietoris) complex to be the complex whose k-simplices are $\{v_{i_0}, \ldots, v_{i_k}\}$ with $d(v_{i_j}, v_{i_l}) \leq \epsilon$.*

2.2 Euler integral

In this subsection we define the Euler integral. Initially, we need to define the Euler characteristic.

Definition 2.5. *Let X be a complex, the Euler characteristic $\chi(X)$ is the alternating sum $\sum_n (-1)^n c_n$ where c_n is the number of n-simplices.*

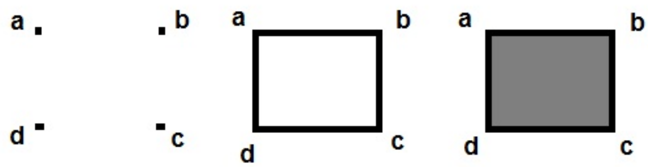

Figure 3: The Rips complexes of four points on the vertices of a unit square. On the left for ϵ in $[0, 1)$, in the middle for values in $[1, \sqrt{2})$ and on the right for values in $[\sqrt{2}, +\infty)$

The definition generalizes the well-known Euler characteristic in two dimensional complexes. The relation of the Euler characteristic to the n-th homology group is given by the following theorem.

Theorem 2.6. *[8, p. 146] Let X be a complex, then*

$$\chi(X) = \sum_n (-1)^n rank H_n(X).$$

From the above theorem it is clear that the Euler characteristic is a homotopy invariant. Hence it can be defined for many topological spaces. For example:

1. for a finite set X the Euler characteristic equals the number of points in X,
2. $\chi(X)$=vertices-edges+faces=2 for plane graphs,
3. for a subset X of $\mathbb{R}^2$ with n holes $\chi(X) = 1 - n$,
4. $\chi(X) = 2 - 2g$ for an orientable surface X of genus g.

A simple argument based on the Mayer-Vietoris sequence [8] gives that for two complexes A, B

$$\chi(A \cup B) = \chi(A) + \chi(B) - \chi(A \cap B).$$

This fact allows us to set up an integration theory using Euler characteristic.

Definition 2.7. *[1, 2] Let X be a complex and $CF(X)$ the abelian group of functions from X to $\mathbb{Z}$ with generators the characteristic functions 1_σ, where*

σ is a closed simplex of X. Then for a function $h = \sum_\alpha c_\alpha 1_\alpha \in CF(X)$ the Euler integral with respect to the Euler characteristic is defined to be

$$\int_X h d\chi = \sum_\alpha c_\alpha.$$

3 Homology and sensor networks

3.1 Hole detection

Perhaps the most important question about a sensor network is whether it covers an entire area or not. Furthermore if there exists a hole in the coverage we should have ways to detect it. This problem is easily solved if the location of each sensor of the network is known. On the other hand there exist scenarios where our sensors, or at least most of them, do not have any information about their location. This might be due to the fact that our sensors are too small to carry positioning systems or because these systems are too expensive. Maybe in the near future swarms of low cost sensors will be spread in an area for collecting data. For example Smart Dust was a research proposal [11] to DARPA to build wireless sensor nodes with a volume of one cubic millimeter. The project led to a working mote smaller than a grain of rice.

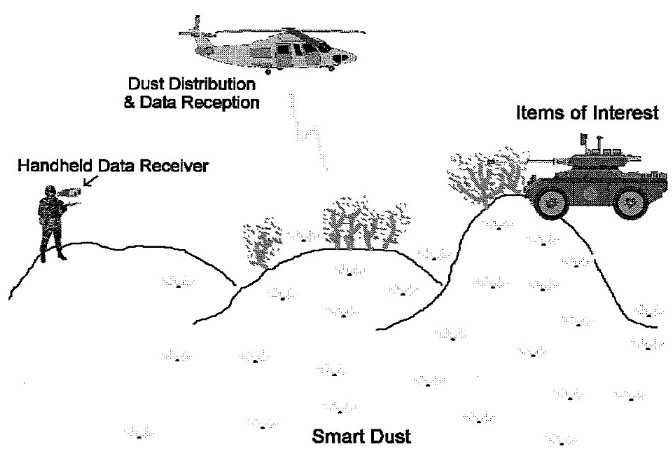

Figure 4: A smart dust - like network whose sensor nodes are delivered by a helicopter and data received by a handheld device [11].

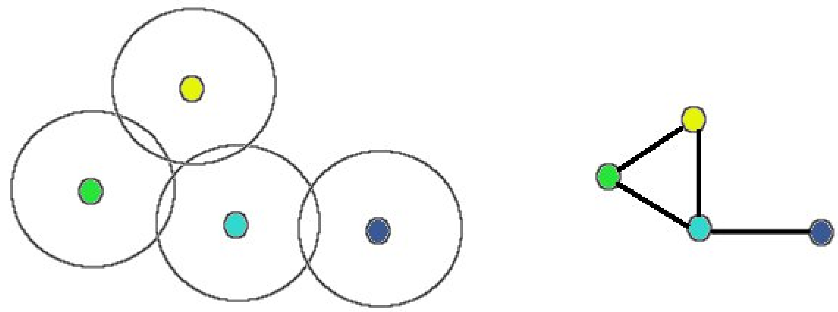

Figure 5: Left: a simple sensor network of coverage disks. Right: the corresponding Čech complex. Note that the triangle is not filled since the three circles on the left do not have a common intersection

The first step in using homology is to create a complex from the sensor network. Assuming that each sensor covers a disk of radius r_c, and of course that it can detect the presence and the identity of any other sensors in that disk, we can create the Čech complex C of the coverage disks (see fig. 5). For the Čech complex of a set $\{U_\alpha\}$ of sets the following theorem [12, 13] holds:

Theorem 3.1. *Let $\{U_\alpha\}$ be a collection of disks and C the corresponding Čech complex, then the union $\bigcup_\alpha U_\alpha$ has the homotopy type of C.*

Thus, if $H_1(C) \neq 0$ we can deduce that our network contains a hole. Unfortunately, there exists no way of constructing the Čech complex for a sensor network. That is the main reason for introducing the Rips complex. Although the Rips complex does not capture the topology of the cover, it gives a good approximation. Furthermore, the Čech complex is nested between two Rips complexes as described in the following theorem [3, 13]:

Theorem 3.2. *Let X be a set of points in $\mathbb{R}^2$ and C_ϵ the Čech complex of the cover of X by balls of radius $\epsilon/2$ and $R_{\epsilon'}, R_\epsilon$ the Rips complexes for constants ϵ', ϵ. Then $R_{\epsilon'} \subseteq C_\epsilon \subseteq R_\epsilon$ whenever $\frac{\epsilon}{\epsilon'} \geq \frac{2}{\sqrt{3}}$.*

Remark 3.3. *For $\epsilon' < \epsilon$ we have $R_{\epsilon'} \subseteq R_\epsilon$ and hence the inclusion maps $i : R_{\epsilon'} \to R_\epsilon$ defines a map between homology groups $i_* : H_\bullet(R_{\epsilon'}) \to H_\bullet(R_\epsilon)$. The study of these maps will prove useful later in 3.4. It is the beginning of a promising theory called persistent homology.*

In [4] a divide and conquer method is introduced that utilizes the homology of the Rips complex to detect not only the presence of holes in the network's coverage but also their location.

As noted above, there exist cases where the Rips complex fails to detect the holes in the network's coverage. This problem can be resolved by strengthening our assumptions. In particular, suppose that the following assumptions hold:

A1. Sensors broadcast their unique IDs. Each sensor can detect the identity of any sensor within a range r_s via a strong signal and within a larger range r_w via a weak signal.

A2. Each sensor covers a disk of radius r_c, where $r_c \geq r_b/\sqrt{3}$.

A3. r_c, r_s, r_w satisfy $r_c \geq r_s\sqrt{1/3}$ and $r_w \geq r_s\sqrt{13/3}$.

A4. The sensors are placed on a bounded subset D of the plane. Moreover sensors can detect the presence(but not the location or direction) of the boundary within a fixed fence detection radius $r_f \geq 0$. ϑD is connected and piecewise linear.

A5. The restricted domain $D - C$ is connected, where $C = \{x \in D : ||x - \vartheta D|| \leq r_f + \frac{1}{2}r_s\}$.

A6. The curve(s) $\{x \in D : ||x - \vartheta D|| = r_f\}$ have internal injectivity radius at least $r_s/\sqrt{2}$ and external injectivity radius at least r_s.

In [13] the following theorem is proved:

Theorem 3.4. *For a network satisfying A1-A6 let F_s, F_w be the subcomplexes of R_s, R_w respectively whose vertices correspond to the sensors that lie within the fence detection radius. The region $D-C$ is contained in the cover if there is a homology class in $H_2(R_s, F_s)$ which is nonzero in $H_2(R_w, F_w)$.*

3.2 Target enumeration

In this section we outline the use of the Euler integral to the target enumeration problem of a sensor network. In particular, given a network of sensors which are only able to detect the presence of an other sensor or of certain "targets" within a disk of radius r_c and where relaying of messages is done between sensors within distance r_c, we want to enumerate the targets in the network's coverage. It should be emphasized that each sensor counts only the number of targets in it's covering disk. It can not identify the targets, detect the direction of the other sensors or the targets. Hence, if a target is contained in the intersection of the covers of two sensors it is counted twice.

Assuming that the sensors cover an entire subspace X of $\mathbb{R}^2$ and for every sensor α the subset U_α of X containing all the sensors to which detect the sensors, then we can define a function $h : X \to \mathbb{N}$, where $h(x)$ is the number of targets α which can be detected by the sensor located at x ($h(x) = \#\{\alpha : x \in U_\alpha\}$). Then the following theorem states that we can use the Euler integral to enumerate the targets in X.

Theorem 3.5. *[2, 1] Given $h : X \to \mathbb{N}$ the counting function of compact target supports in X satisfying $\chi(U_\alpha) = n \neq 0$ for all targets α. Then $\#\alpha = \frac{1}{n}\int_X h d\chi$.*

Naturally, the assumption that at every point of X a sensor is located is not realistic. In practice, we assume that the sensors are located on the vertices of a triangulation of X. In that case the values of h are known only for the vertices and we can integrate the piecewise linear interpolation of h [2].

3.3 Further applications

We should point out that there more applications of algebraic homology to sensor networks. Let the following assumptions hold:

B1. the sensors broadcast their unique IDs and each node can detect the ID of any node within radius r_b,

B2. sensors cover a disk of radius $r_c \geq r_b\sqrt{3}$,

B3. sensors lie in a compact connected subset of the plane whose boundary is connected and piece-wise linear,

B4. every sensor on the boundary knows the IDs of its two adjacent boundary nodes which both lie within distance r_b.

Then [12] the sensor cover contains D if there exists $[a] \in H_2(R, \vartheta D)$ such that $\vartheta a \neq 0$. This fact enables us to detect reductant sensors. (The non-reductant sensors are part of a minimal generator of $H_2(R, \vartheta D)$). Thus, the network can for example, conserve energy by placing in sleeping mode the reductant sensors.

Another example [12] has to do with the problem of determining whether an evader can avoid detection in a network whose sensors are on the move or come online and offline at various times. Under some reasonable assumptions there exists an affirmative answer to this problem.

Finally, in [4] a method for detecting wormhole attacks[1] on a network is presented.

4 Data analysis

4.1 General methods - A first approach

Another unexpected application of topology is to data analysis. The main advantage of using topological methods is that these methods are noise tolerant.

Quite naturally the first application is to use topology to classify different data sets. A dataset may be viewed as a set of points of some n-dimensional real space $\mathbb{R}^n$ (this set is often called cloud set). From the dataset we can construct a complex (i.e. a Rips complex) and calculate it's homology groups and the corresponding betti numbers. It is then possible to try to classify different datasets by examining their respective betti numbers. Several researchers have conducted tests attempting to classify data coming from a great variety of sources, from images to texts [3, 6, 14, 16].

4.2 More sophisticated techniques

The study of the evolution of the homology of a Rips complex R_ϵ obtained from a dataset for various values of ϵ can be used in various applications. For example, it can be used for identifying the main topological characteristics of an object that is sampled. The main topological characteristics are those that are persistent while ϵ changes. The persistence of homology has several practical applications from protein docking and image analysis [5, 15] to fine tuning machine learning algorithms [10].

5 Conclusion

The authors would like to point out that this is a very short exposition. Many details are omitted while a large number of applications is left out. We only hope that this article will motivate the readers to search for more detailed information. The field is relatively new and will certainly welcome

[1] In this kind of attack involving two malicious/infected sensors, a signal received by one sensor is transmitted over a low-latency link and replayed by the other. If the two sensors are located far apart this can cause several problems to the network [9]

researchers from diverse backgrounds ranging from mathematics to software engineering.

References

[1] Y. Baryshnikov, R. Ghrist, *Target enumeration via Euler characteristic integrals*, SIAM J. Appl. Math., 70(3), 2009, 825-844.

[2] Y. Baryshnikov, R. Ghrist, *Target enumeration via integration over planar sensor networks*, Proc. Robotics: Science and Systems, 2008.

[3] G. Carlsson, *Topology and Data*, Bull. of the AMS Vol. 46, N. 2, 2009, 255-308.

[4] H. Chintakunta, H. Krim, *Topological Fidelity in Sensor Networks*, arXiv:1106.6069.

[5] H. Edelsbrunner, J. L. Harer, *Computational Homology*, AMS, 2010.

[6] D. Freedman, C. Chen, *Algebraic topology for computer vision*, HP Laboratories HPL-2009-375, 2009.

[7] R. Ghrist, A. Muhammad, *Coverage and hole-detection in sensor networks via homology*, Proceedings of the 4th international symposium on Information processing in sensor networks,IEEE Press, 2005.

[8] A. Hatcher, *Algebraic Topology*, Cambridge Un. Press, 2002, also available at http://www.math.cornell.edu/∼hatcher/AT/ATpage.html.

[9] M. Y. Malik, *An Outline of Security in Wireless Sensor Networks: Threats, Countermeasures and Implementations*, arXiv:1301.3022.

[10] D. Panagopoulos, *Homology and Lazy Learning*, 1 National conference of HMS and HSOR 2011, 2011.

[11] Kristofer S.J. Pister, *Smart Dust*, BAA97-43 Proposal Abstract

[12] V. de Silva, R. Ghrist, *Homological Sensor Networks*, Notices of the AMS Vol. 54, N. 1, 2007, 10-17.

[13] V. de Silva, R. Ghrist,A. Muhammad, *Blind Swarms for Coverage in 2-D*, Conference: Robotics: Science and Systems - RSS , 2005, 335-342.

[14] G. Singh, F. Mémoli, G. Carlsson, *Topological Methods for the analysis of high dimensional data sets and 3D object recognition*, Point Based Graphics 2007, Prague Slides. Sep 2007.

[15] A. J. Zomorodian, *Topology for computing*, Cambridge Un. Press, 2005.

[16] H. Wagner, *Computational topology in text mining*, ATMCS 2012, 2012.