Persistent homology to quantify the quality of surface-supported covalent networks

Abraham Gutierrez
a.gutierrez@math.tugraz.at
Institute of Discrete Mathematics
Graz University of Technology
Graz, Austria

Mickaël Buchet
buchet@tugraz.at
Institute of Geometry
Graz University of Technology
Graz, Austria

Sylvain Clair
sylvain.clair@univ-amu.fr
Aix Marseille Univ,
Univ Toulon, CNRS, IM2NP
Marseille, France

Supplementary Information

Contents

1 Robustness Analysis

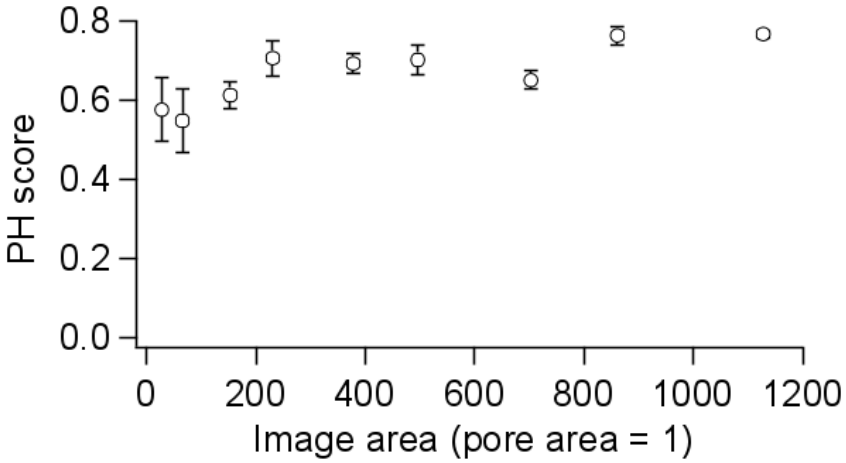

Figure S1: The PH scores are reported for equivalent simulated networks images depending on their size.

To analyze the robustness of the method with respect to the image size, the PH score was computed for a series of images representing simulated networks with honeycomb structure, as reported in Figure S1. The image area was normalized by considering that a single pore has an area of 1. A representative image of such network and its corresponding persistent

diagram are given in Figure 1 of the main manuscript. The images with sizes lower than 100 (normalized area) provided lower scores and high error bars. They should therefore not be considered. The standard deviation of the PH score obtained for image sizes larger than 100 amounts to 0.06, which is of similar order of magnitude as the standard deviation obtained for a fixed image size with 20 different simulated images (0.05).

2 Overview of the PH Scoring Method

The scoring method can be split into three steps. The first two steps consist in creating a persistent diagram from the experimental data. The last step is the scoring computation for the regularity of networks.

First, we treat the input image in order to obtain a binarized version. The result is a black and white image where one of the colors corresponds to the matter. For simplicity, we assume that this is the white part of the image. The binarization process is done using a simple thresholding method, the threshold being decided depending on the quality of contrast of the input image.

Then we compute the persistence diagram of the binarized image. Details on the mathematical background of the method can be found elsewhere (e.g. [S1]). In brief, every pixel is given a numerical value based on its distance to the boundary between the black and white domains. Pixels are considered adjacent if they touch along an edge. White pixels adjacent to a black pixel are given the value 0. Other white pixels are given the value -d where d is the distance to the closest border pixel considering only adjacent pixels to be at distance 1 using the Manhattan distance. Conversely the black pixels are given a positive value equal to the distance to the nearest white pixel. We then consider the sublevel sets of this function on pixel and study the evolutions of holes in the structure. Each hole in the structure has a birth time when the enclosing circle first appears and a death time when the hole is filled. This can be expressed through the persistence diagram which is a multi-set of points in the plane, each point corresponding to the pair (birth, death) for a hole. The difference death-birth is called the lifespan of a hole.

In a last step we compute the PH score. A scale factor s is given. For every point p in the persistent diagram, we count the points located at distance at most its lifespan divided by s. These points correspond to the holes that are deemed similar to p. The score relative to p is then defined as the number of similar points divided by the total number of points in the diagram. The score of the diagram is the maximal such score that can be obtained across the diagram. The score depends on the scale parameter s but is experimentally stable for s values between ~ 3 and 6. Because the method is based on a binarized image, a perfectly flat background is required for the experimental image. In some cases a smoothened image representing the underlying background surface was subtracted to the original image.

The code used for the calculation is available at http://www.geometrie.tugraz.at/buchet/phscore/

3 An Empirical Approach to Digital Images and Persistent Homology

The purpose of this section is to explain in informal terms the *persistent homology* and how to analyze *topological features* of a digital images with it. The *topological features* of a 2D-image are its *holes*[1]. As example, Figure S2 shows a black-and-white image whose black pixels form two holes.

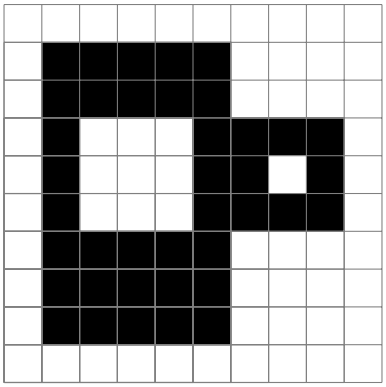

Figure S2: Digital image with 10×10 pixels. The black pixels of the image form a shape with two holes, one "big" on the left-hand side and one "small" on the right-hand side.

[1] Just for the purposes of this informal section.

The *base* of an image is the set of pixels that form the shape we want to study; the rest of the pixels are *background* pixels. For example, in a black-and-white image, we can specify the shape we want to study by refering to its white-base or black-base; in the black-base of Figure S2, we observe two holes. We want an efficient mathematical procedure that "estimates" the number of holes of an image and "measures" how "big" these holes are. In the case of the black-base of Figure S2, this procedure should say in a mathematical way: "there are two holes; one of them is big and the other small". Notice that we are not interested in knowing the location of the holes and the method we describe next wont care about the location neither.

3.1 Extracting Topological Features of an Image via a Filtration

In this subsection, we describe in a informal way how to extract the topological features of an image via a *filtration*. A filtration of a black-and-white image I is a series of equal-sized, black-and-white images $I_{t=i}, I_{t=i+1}, \ldots, I_{t=i+n}$ which are labeled with a "time" parameter t and whose respective sets of black pixels "grow" with the time parameter. In particular, this means that the black pixels of any particular image in the filtration contains all black pixels of all the previous images in the filtration; it also means that the last image in the filtration will be totally black.

The is no unique way to define a filtration, however, a definition that is convenient for the extraction of the topological features of an image is by growing from the "deepest" pixels in the image's base into ever more "distant" pixels[2]. We will refer to this filtration as the Manhattan-filtration of the image I or simply as the filtration of I. To illustrate this, we show the filtration of Figure S2.

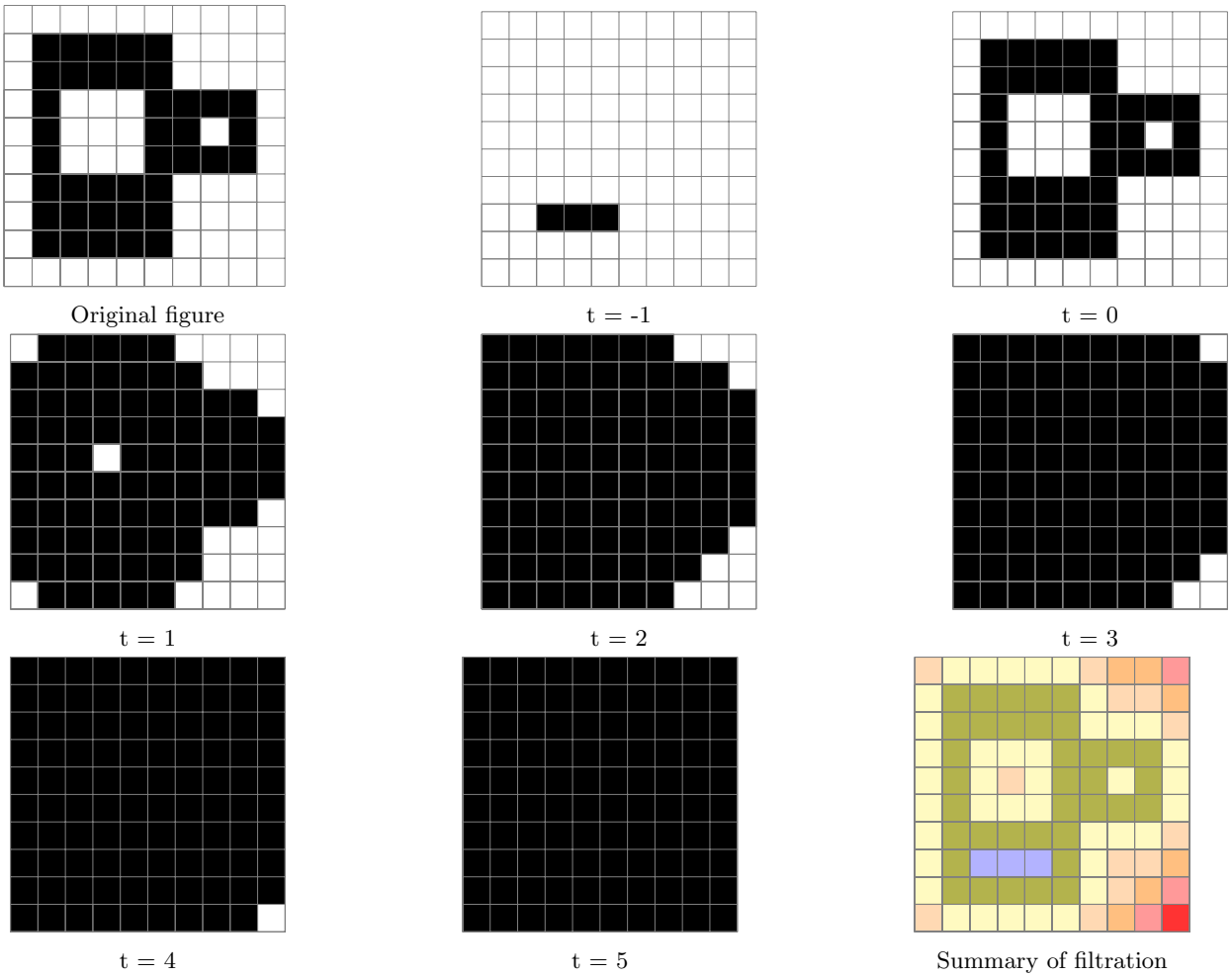

Figure S3: Filtration of the black-base of Figure S2.

We now focus on how to extract the topological-features of an image with its Manhattan-filtration: we observe in increasing

[2]This process is formally described in the next subsection.

order the images of the filtration and at each time, we write down if a hole appears or disappears in the current filtration image; at the end, we have a multi-set[3] of ordered pairs $\{(x_1, y_2), \ldots, (x_m, y_m)\}$. Every pair (x_i, y_i) refers to a exactly one hole that appeared at time $t = x_i$ of the filtration and died at time $t = y_i$. We will exemplify this process with the filtration shown in Figure S3

• $t = -1$: no holes are destroyed or created;
• $t = 0$: two new holes appear. We write two tuples (one for each hole) plus its birth times in the first coordinates: $(0, -), (0, -)$;
• $t = 1$: one of the holes dies [4]. We write down in its corresponding tuple, the time of its death: $(0, -), (0, 1)$;
• $t = 2$: one hole dies. We write its death time in its corresponding tuple: $(0, 2), (0, 1)$;
• $t = 3$: no holes are destroyed or created;
• $t = 4$: no holes are destroyed or created;
• $t = 5$: no holes are destroyed or created.

We return the multi-set of tuples that was left at the end

$$\{(0, 2), (0, 1)\}.$$

This multi-set of tuples has a close relation with the topological features of the base.

The *lifespan* of a topological-feature (hole) in a filtration is the subtraction "DeathTime - BirthTime". This number quantifies how long a hole remains "alive" during the filtration. It is natural to expect that a "big hole" lives longer than a "small" one.

The lifespan of a hole with tuple (x_i, y_i) is given by $y_i - x_i$. In the multi-set given in the previous example, we see that the lifespan of the hole corresponding to the first tuple is $2 - 0 = 2$ and the lifespan of hole corresponding to the second tuple is $1 - 0 = 1$. From here, we conclude that the hole of the first tuple is bigger than the hole of the second tuple.

One usual way to visualize the tuples is by plotting them in the plane. The resulting plot is call the *persistence diagram* of the image with base X (this refers to the set of pixels that form the shape we want to study). In the case of the tuples of Figure S2, we get

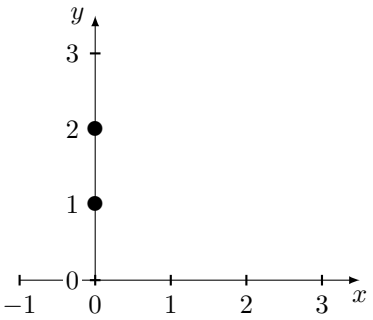

Figure S4: Persistence diagram of the black-base of Figure S2.

In the following figure, we show a real image with white-base and its persistence diagram

[3] A multi-set is a list of elements in which an element is allowed to be repeated.

[4] The corners do not count as holes. A hole needs to be completely surrounded by base-pixels.

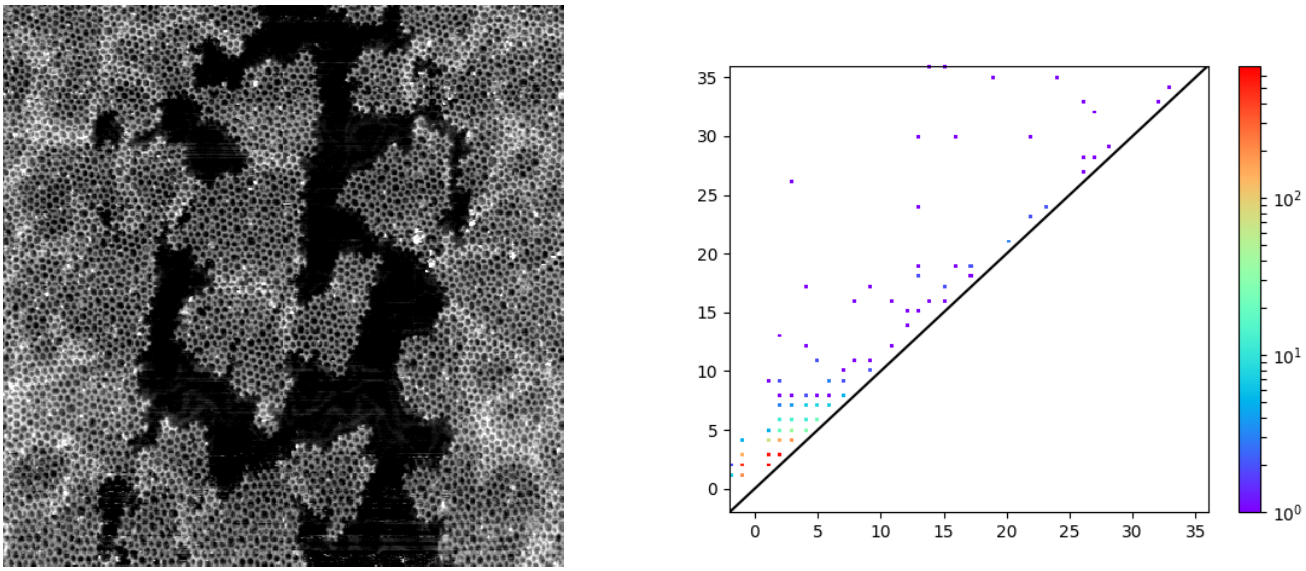

Figure S5: Image with white-base and its persistence diagram

In the persistence diagram presented in Figure S5, the points have colors. The color of a point specifies its multiplicity, i.e. how many times it appears in the multi-set generated by the filtration algorithm. Color red means many repetitions and color blue not so many; this is quantified in the vertical bar on the right-hand side of the image. Every tuple has the form

$$(\text{BirthTime}, \text{DeathTime})$$

and therefore each point appears above the diagonal $x = y$. This simply reflects the fact that

$$\text{BirthTime} < \text{DeathTime}$$

i.e. a hole can not disappear before it has been created. Lastly, the lifespan of a point is the vertical distance to the diagonal i.e. if we take a point in the diagram and move it vertically downwards, its lifespan is the distance covered until it reaches the diagonal. The holes with tuples that are close to the diagonal ("small" lifespan) are normally "small" holes. On the other hand, the holes which tuples are far from the diagonal ("large" lifespan), are "big" holes. To illustrate this, lets observe the persistence diagram shown in Figure S5: there are plenty of tuples in the plane region $[-2, 5] \times [0, 5]$ (red points). These holes are born between $t = -2$ and $t = +5$ and they die before $t = 5$. This reflects the fact that there are plenty of small holes on the shape formed by the white pixels.

3.2 Formal Construction of the Manhanttan-Filtration of a Digital Image

Lets recall that the filtration of a black-and-white image with black base is a series of black-and-white, time-labeled images whose respective sets of black pixels grow with respect to a time parameter. Given a digital black-and-white image, the construction process of its filtration consist of two steps: first, label each pixel of the image with an integer time parameter t [5]; then, the black pixels of the filtration image I_{t_1} are all the pixels that have label less or equal to t_1. For example, the black pixels of the filtration image I_3 are all pixels that have one of the following labels $3, 2, 1, -1, -2, ...$, all other pixels in I_3 are white. This process constructs all the images in the filtration.

Let p be a pixel in an image with base B. The label of p is the Manhattan distance of p to the *boundary* of the base B if p does not belong to the base and the negative of this distance if p belongs to the base. The boundary of the base consist of all pixels on the base that share at least one of their sides with a pixel outside the base. The Manhattan distance of a pixel p to the boundary B is the minimal amount of jumps (up, down, left, right) needed to land on a pixel of the boundary; in particular, this means that the label of all boundary pixels is zero. Figure S6 illustrates the labeling process.

[5] This means that t is a number in the set $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$

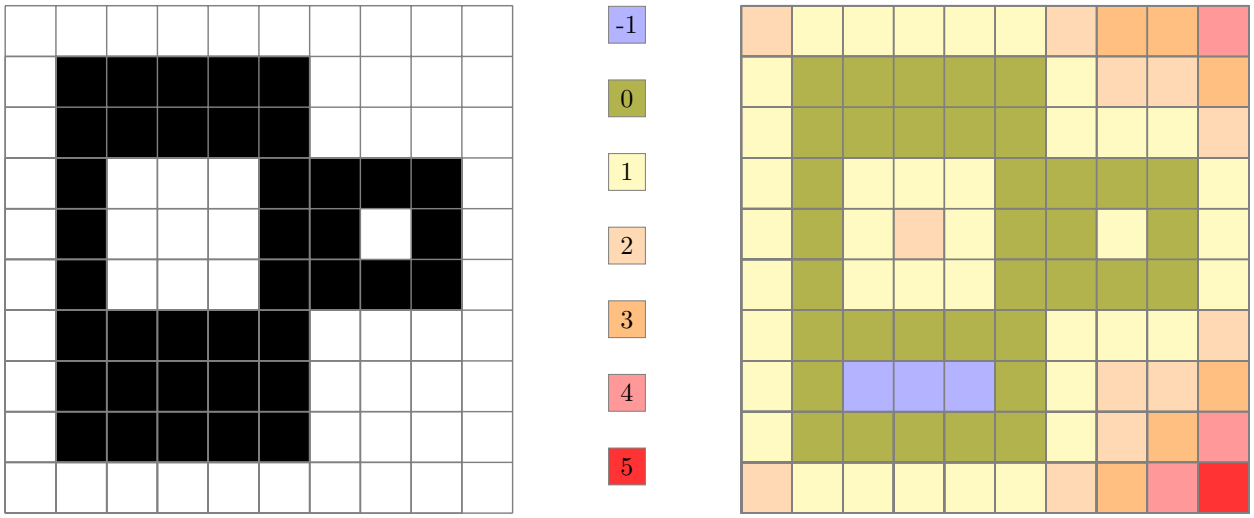

Figure S6: An image in black-base and the labels for each pixel.

3.3 The PH Score of an Image

In this subsection, we explain a how the persistent diagram of a image can be used to measure its regularity: assume we are given some real images of $2D$ networks (e.g. Figure S5), we need to decide if the network is of high quality or not; we are told that if the network is highly regular (e.g. a regular hexagonal grid) the material is good and not so good otherwise. We therefore need to quantify the "deviation" of a network from the ideal regular grid; in our case, the regular hexagonal grid. We do this by scoring the persistence diagrams of the images of the networks.

We propose a way to score the persistence diagram of a network image in order to quantify the "deviation" of the network from the hexagonal regular grid. The Persistent Homology score (PH score) is a value between 0 and 1 that is assign to a persistent diagram; 1 is the best score and 0 is the worst score. For example, the perfect hexagonal grid will have a PH score of 1.

Let be the multi-set obtained by the filtration algorithm applied to the base of an image. Given the user defined parameter s, we define the s-score of a point (x, y) as the proportion of all topological features that were born between $x \pm \frac{y-x}{s}$ and died between $y \pm \frac{y-x}{s}$. The PH score with parameter s is the the maximum over all these s-scores.

The parameter s adjusts the discriminative power between similar holes. A small s means that almost all of the holes are considered similar; for a large s, only holes with an identical tuple (x, y) are considered similar.

References

[S1] Buchet, M., Hiraoka, Y. & Obayashi, I. Persistent Homology and Materials Informatics, in Nanoinformatics (ed Isao Tanaka) 75-95 (Springer Singapore, 2018).