Chapter 15
Networks and Cycles: A Persistent Homology Approach to Complex Networks

Giovanni Petri, Martina Scolamiero, Irene Donato, and Francesco Vaccarino

Abstract Persistent homology is an emerging tool to identify robust topological features underlying the structure of high-dimensional data and complex dynamical systems (such as brain dynamics, molecular folding, distributed sensing).

Its central device, the filtration, embodies this by casting the analysis of the system in terms of long-lived (*persistent*) topological properties under the change of a scale parameter.

In the classical case of data clouds in high-dimensional metric spaces, such filtration is uniquely defined by the metric structure of the point space. On networks instead, multiple ways exists to associate a filtration. Far from being a limit, this allows to tailor the construction to the specific analysis, providing multiple perspectives on the same system.

In this work, we introduce and discuss three kinds of network filtrations, based respectively on the intrinsic network metric structure, the hierarchical structure of its cliques and—for weighted networks—the topological properties of the link weights. We show that persistent homology is robust against different choices of network metrics. Moreover, the clique complex on its own turns out to contain little information content about the underlying network. For weighted networks we propose a filtration method based on a progressive thresholding on the link weights, showing that it uncovers a richer structure than the metrical and clique complex approaches.

Keywords Complex networks · Persistent homology · Metrics · Computational topology

15.1 Introduction

Over the last decade complex networks have become one of the prominent tools in the study of social, technological and biological systems. By virtue of their sheer

G. Petri (✉) · F. Vaccarino
ISI Foundation, Via Alassio 11/c, 10126 Torino, Italy

M. Scolamiero · I. Donato · F. Vaccarino
Dipartimento di Scienze Matematiche, Politecnico di Torino, C.so Duca degli Abruzzi n. 24, 10129 Torino, Italy

sizes and complex interactions, they cannot be meaningfully described and controlled through classical reductionist approaches.

Within this framework, the study of the topology of complex networks, and its implications for dynamical processes on them, has most often focused on the statistical properties of nodes and edges and therefore found a natural and effective description in terms of statistical mechanical models of graph ensembles [1, 2]. These models rely for their formulations on local interactions and become quickly hard to manage when higher correlations are included or one-step approximations are not sufficient, as Schaub et al. [3] pointed out for the case of community detection algorithms for example.

The last few years saw a new perspective emerge that focuses on the very *geometry* of complex network. It was promoted by a large availability of new (typically geosocial) data coming from spatial networks [4], but also by analytical and numerical results on the relations between geometrical properties and global features of complex networks, e.g. the hyperbolic embedding of the Internet with the resulting increased efficiency of greedy routing algorithms [5], stationarity conditions for chemical networks [6] and brain cortex dynamics [7].

In this work, we take on this perspective and study the geometrical properties of networks through the goggles of *persistent homology*, a technique originally introduced by [8, 9] to uncover robust topological information from noisy high-dimensional point clouds. Persistent homology works by extracting from a dataset a growing sequence of simplicial complexes (called *filtration*), indexed by a parameter ϵ, and studying the associated homology groups, which encode the geometrical information (for example, the holes of an n-torus). The robustness of each topological feature is then obtained from the *persistence* of the corresponding generator along the filtration,

For example, in the case of the torus, there will be two persistent generators associated to the two non-equivalent loops on its surface.

Persistent homology has received some attention in the context of networks [10], but there has been no systematic study on its efficiency and sensibility for networks yet. This is of particular importance since, in contrast with the unique natural metric available for point cloud datasets, networks allow various rules to generate the filtration.

Our results will show that the salient features of the homology do not change significantly under different metrics and that there exist a metric scale ϵ_c at which the filtration displays the richest structure.

We will then study a second method to create the filtration, relying only on the network clique structure. Unfortunately, this will turn out to yield little additional information.

In the case of weighted networks it is possible to devise a refined filtration based on the clique structure of the network thresholded by ϵ, which yield a much richer picture than the simple clique complex method.

The rest of this work is organised as follows. In the next section a minimal introduction to homology and its persistent sister is given. The following section will present selected results of simulations and datasets under different choices of metrics for the network filtration.

We conclude then presenting the procedure for the filtration built with the link-thresholded clique structure and briefly discuss the results and implications for future research. In particular, we have discuss the possibility of expanding the method by considering multi-filtrations, that is filtrations indexed by more than one parameter.

15.2 Homology

Formally, homology is an algebraic invariant converting local geometric information of a space into a global descriptor. There are many homology theories, but simplicial homology is the most amenable for computational purposes thanks to its combinatorial structure.

This kind of homology is applied to simplicial complexes, that are combinations of vertices, segments, triangles and higher dimensional analogues, joined according to specific compatibility relations. As we will see in the following, simplicial complexes can be constructed from discrete spaces or networks. Low dimensional homology groups have an intuitive interpretation. Given a simplicial complex X, $H_0(X)$ is the free group generated by the connected components of X, $H_1(X)$ is the free group generated by the cycles in X, $H_2(X)$ is the free group generated by voids—holes bounded by two-dimensional faces. The Betti numbers count the number of generators of such homology groups.

The standard tool to encode this information is the so-called *barcode*, which is a collection of intervals representing the lifespans of such generators. Long-lived topological features can be distinguished in this way from short-lived ones, which can be considered as topological noise. There are various ways of building persistence modules out of a given dataset. The most known are the Rips-Vietoris complex, the Cech complex and the clique complex [8]. The first two require a metric space for the data and are generated by inflating spheres of the same radius around points (or nodes in a network) and associating set of points to simplices according to the overlap of the corresponding spheres. They can also be used to create a filtration out of general network, once a metrical structure is given on the network itself (shortest-path, commute time distance, etc). Besides these two methods, there exist a few methods pertaining to networks only [8], the best known being the *clique complex*, which is generated by associating to each maximal clique the simplex generated by the vertices of the clique.

15.3 Robustness Against Metric Change

Network metrics have been well studied, especially in the context of clustering algorithms [11] and Markov Chain models [12]. In addition to the shortest path and commute time metrics, it is possible to define kernel matrices as functions of the
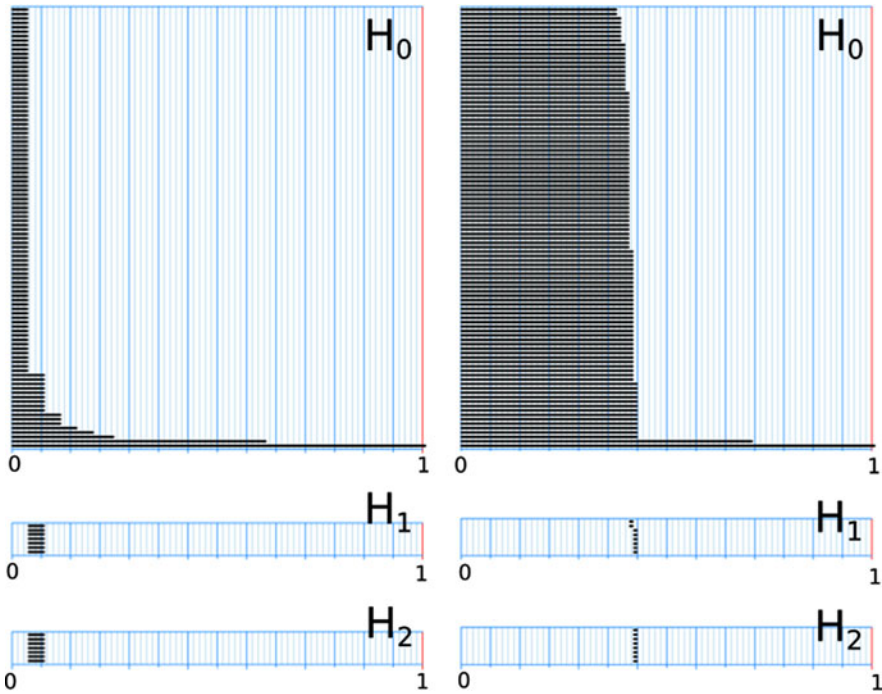
Fig. 15.1 Barcodes for the shortest path metric (*left*, panels (**a**), (**c**) and (**e**)) and the Von Neumann metrics (*right*, panels (**b**), (**d**) and (**f**)) on the C. Elegans brain network. From *top* to *bottom*, we report the intervals of existence of the homological spaces H_0 (panels (**a**) and (**b**)), H_1 (panels (**c**) and (**d**)), H_2 (panels (**e**) and (**f**)). The parameter $\epsilon \in [0, 1]$ increases from *left* to *right*. Each *horizontal line* corresponds to the intervals of existence of a generator of the corresponding homology space. In both cases, the higher homology space are non-trivial only in the vicinity of the merging of a large number of connected components, as highlighted by the drastic reduction in the number of generators of H_0

network's adjacency and Laplacian matrices. From such kernels one obtains a well-defined distance, which effectively turns the network into a metric space.

We analysed the metrics associated to: the shortest paths, the commute time between nodes, exponential diffusion [13] and exponential Laplacian diffusion [11, 14], which emerge as solutions of diffusion processes on the corresponding network, the von Neumann kernel [15],which generalises the hub-authority measures, Markov diffusion [16] and random walks with restart.

For each metric, the filtration was generated and the persistent homology calculated. The analysis was repeated on a range of different networks, spanning different network topologies, sizes and origins (biological, social, technological).

For brevity, in this paper, we show only the comparison of the barcodes obtained using the shortest path and the exponential diffusion (with $\alpha = 0.01$) distances for the *C. elegans* neuronal network (Fig. 15.1). In order to compare the results both metrics have been mapped to the interval [0, 1]. Surprisingly, we found that the

higher homology spaces (H_1, H_2 ..., bottom plots in Fig. 15.1) are trivial for most values of the filtration parameter. They do however show the appearance of generators of higher homology groups in the vicinity of the value of ϵ at which a significant number of connected components merges into few, as shown by the decrease in the number of generators of H_0.

In this respect, our results suggest the existence of a particular value ϵ_c, a *metric scale*, at which one observes the most structure in the metrical representation of the network under study. The same behaviour was found in a number of other networks, ranging from the US air passenger network to the human gene regulatory one. Note moreover that, in general, ϵ_c is different from the average distance between the nodes (in terms of the chosen metric) and therefore cannot be explained as a mere effect of the distances distribution. Moreover, if the appearance of non trivial higher homology groups was only due to the merging of small connected components into a giant component, one would expect to observe the same phenomenon also for the merging of smaller components. However, we did not see any of these signatures, supporting the existence of a characteristic scale ϵ_c.

15.4 Clique Complex and Link Weights Thresholding

Another natural filtration of a network is generated by considering its clique structure. The clique complex is obtained by associated to each maximal k-clique, a completely connected subgraph formed by k nodes, the $(k-1)$-simplex whose vertices are the nodes of the clique. The natural parameter for this filtration is the clique dimension k. Recent work [10] tried to uncover specific signatures of modular and cluster structures in complex networks by making use of this filtration. In our analysis the filtration obtained in this way did not show interesting features in addition to the clique structure itself, which however can be investigated without recurring to homological concepts. However, if we consider weighted networks, it is possible to devise a filtration which combines link weights and clique structure. Given the weighted adjacency matrix ω_{ij}, we let ϵ vary in $(\min \omega_{ij}, \max \omega_{ij})$ and consider a sequence of networks, such that the network at step ϵ contains all links (i, j) with $\omega_{ij} > \epsilon$. As we decrease ϵ from its maximum allowed value, we go from the empty network to the original one. For each step, we build the corresponding clique complex and study the persistent homology of the resulting filtration. Figure 15.2 shows the results of this filtration on a large Facebook-like network of online contacts. It is immediately evident that a very rich topological information is present. Long persistent intervals appear both for some generators of H_1 and H_2. The first implies the existence of chains composed by edges with large weights, whose nodes though are not strongly connected across the chain itself, but only with their two neighbours along the chain. The same reasoning applies to the case of H_2 where the building blocks are not segments but triangles. The presence of long persistent H_2 generators is a signpost for higher ordering in the structure of the online contacts. This means that strong pair interactions organise in long loops without significant triadic closure.
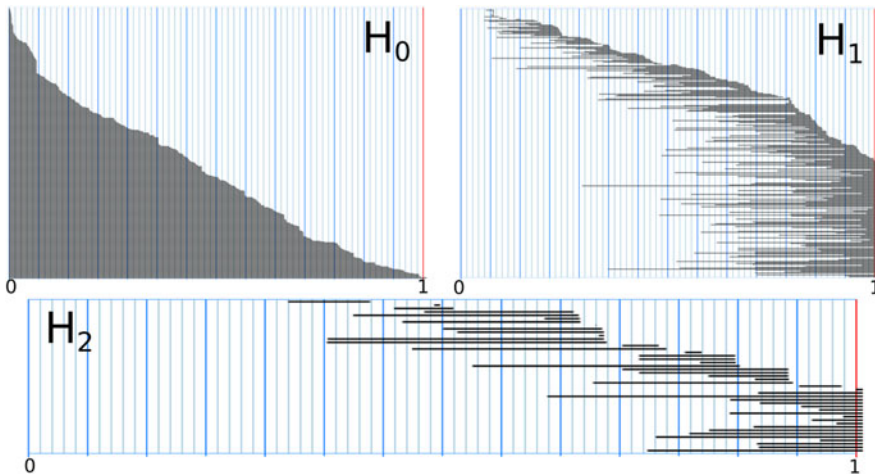
Fig. 15.2 Barcodes obtained from the weighted-clique complex filtration of a network of online contacts for the homology groups H_0 (**a**), H_1 (**b**) and H_2 (**c**). Persistent H_1 and H_2 generators imply that the existence of loops and chains of tethraidra formed by nodes which are weakly interacting with their neighbours in the chain, with the exception of the one directly adjacent *along* the chain. In the case of human contacts, this means that strong pair interactions organise in long loops without significant triadic closure

Finally, we can conclude that this method is able to identify mesoscopic and long-range structures which are present in networks, but would otherwise pass undetected with standard methods, and assigns to them also a measure of robustness in the form of the persistence intervals.

References

1. Newman M (2010) Networks: an introduction. Oxford University Press, New York
2. Albert R, Barabasi A (2002) Statistical mechanics of complex networks. Reviews of Modern Physics 74(1):47–97
3. Schaub MT, Delvenne JC, Yaliraki SN, Barahona M (2012) Markov dynamics as a zooming lens for multiscale community detection: non clique-like communities and the field-of-view limit. PloS One 7(2):e32210
4. Barthlemy M (2011) Spatial networks. Physics Reports 499(13):1–101
5. Boguna M, Papadopoulos F, Krioukov D (2010) Sustaining the internet with hyperbolic mapping. Nature Communications 1:62
6. Conradi C, Flockerzi D, Raisch J, Stelling J (2007) Subnetwork analysis reveals dynamic features of complex (bio)chemical networks. Proceedings of the National Academy of Sciences 104(49):19175–19180
7. Henderson JA, Robinson PA (2011) Geometric effects on complex network structure in the cortex. Phys Rev Lett 107:018102

8. Zomorodian A, Carlsson G (2005) Computing persistent homology. Discrete Comput Geom 33(2):249–274
9. Carlsson G (2009) Topology and data. Bulletin of the American Mathematical Society 46(2):255–308
10. Horak D, Maletic S, Rajkovic M (2009) Persistent homology of complex networks. Journal of Statistical Mechanics: Theory and Experiment 2009(03):P03034
11. Fouss F, Yen L, Pirotte A, Saerens M (2006) An experimental investigation of graph kernels on a collaborative recommendation task. In: Sixth international conference on data mining (ICDM'06), pp 863–868
12. Bolch G, Greiner S, de Meer H, Trivedi KS (1998) Queueing networks and Markov chains: modeling and performance evaluation with computer science applications. Wiley-Interscience, New York
13. Kondor R, Lafferty J (2002) Diffusion kernels on graphs and other discrete input spaces. In: Proceedings of the nineteenth international conference on machine learning (ICML'02), pp 315–322
14. Smola A, Kondor R (2003) Kernels and regularization on graphs. In: Learning theory and kernel machines. Lecture notes in computer science, vol 2777, pp 144–158
15. Kandola J, Shawe-Taylor J, Cristianini N (2002) Learning semantic similarity. Advances in neural information processing systems 15:657–666
16. Yen L, Fouss F, Decaestecker C, Francq P, Saerens M (2007) Graph nodes clustering based on the commute-time kernel. In: Zhou ZH, Li H, Yang Q (eds) Advances in knowledge discovery and data mining. Lecture notes in computer science, vol 4426. Springer, Berlin, pp 1037–1045