TOPDRIM – Topology Driven Methods for Complex Systems

Call / Grant agreement no: FP7-ICT-2011-8 / 318121

D1.1 Elementary coarse-grained features of datasets
WP1: Topology of Data

WP1 Leader: Mario Rasetti
rasetti@isi.it

TOPDRIM Project Coordinator: Emanuela Merelli
emanuela.merelli@unicam.it

Authors: G.Petri, F.Vaccarino, M. Rasetti
Editor: F. Vaccarino
Period Covered: 1 October 2012 – 30 September 2013
Date: 30 September 2013

Activity: RTD

TOPDRIM Project Officer: Roumen Borissov

Project information

Project acronym:	TODPRIM
Project title:	Topology Driven Methods for Complex Systems
Call / Grant agreement no.:	FP7-ICT-2011-8 / 318121
Funding Scheme:	Collaborative project
Project website:	http://www.topdrim.eu
Periodic Report	1 Year
Reporting Period: From	01-Oct-2012
To	30-Sept-2013
Date of preparation:	30th September 2013
Beneficiary organisation:	ISI Foundation
Project coordinator:	Prof. Emanuela Merelli Associate professor – head of the Computer Science Division. Unicam University of Camerino, via del Bastione, 1 62032 Camerino (MC)
Project coordinator organisation:	University of Camerino, Italy
Project coordinator telephone number:	+39 0737 402564 and +39 338 3990412
Project coordinator email address:	emanuela.merelli@unicam.it
WP Leader:	Mario Rasetti
WP Leader email address:	rasetti@isi.it

Consortium Details

1. UNICAM - UNIVERSITÀ DEGLI STUDI DI CAMERINO
2. SDU – SYDDANSK UNIVERSITET
3. UVA – UNIVERSITEIT VAN AMSTERDAM
4. OU – THE OPEN UNIVERSITY
5. ISI – FONDAZIONE ISTITUTO PER L'INTERSCAMBIO SCIENTIFICO (I.S.I.)
6. AMU – UNIVERSITE D'AIX MARSEILLE

Table of contents

1. Summary

The aim of this deliverable is to describe in a formal topological framework the features of general datasets and extend the domain of applicability of existing methods in order to provide the data-driven founding for future work on joint statistical and topological modelling efforts.
In order to associate a topological space to a general collection of points, usually one relies on the existence of an underlying (real or inferred) metric space. This allows us to build an interpretation of data-points as vertices of a combinatorial graph whose edges are determined by proximity. This type of description, however, completely ignores higher order (many-body) effects that might be playing a role.

This problem can be overcome by considering such graph as the scaffold for a higher-dimensional object obtained by completing the graph to a simplicial complex. Then computational topological techniques can be adopted in order to extract the topological features of the original dataset.
Such construction relies however on the existence of a meaningful metric, which in general is not given by the data structure.

Thus, main contribution toward deliverable D1.1 consisted in:

1. developing a language to address this type of problems, by extending persistence homology to the case of non-metrical spaces and building appropriate computational software (ISI, UNICAM);
2. exploiting these new tools to analyze various complex case-study systems, namely the persistent structure of real-world complex networks and the perturbation to the brains' functional structure caused by selected drugs;
3. checking that our approach is mathematically well-grounded by recasting it in categorial terms. In addition to the activities above, topologically motivated analysis were used also to approach problems belonging to phase transitions in classical Hamiltonian systems and biologically-inspired Markov chains (ISI, AMU).

Current results provide a strong foundation for both the next deliverables and further applications of the tools we developed to new systems. In particular, the results of WP2 on the information geometry of dynamical systems on networks constitute a natural stepping stone for the comparison between homological and dynamical properties of such systems required to fulfill the goals of deliverable D1.2.

2. Partner contribution / collaboration

ISI: A general dataset can be seen as a point cloud in an appropriate space endowed with some relation between the points. Therefore, in general we can describe a dataset as a weighted network where points constitute the nodes and the relation defines the strength of the links among such nodes. Unfortunately, in real datasets the relations are hardly metrical, for example in the case of very heterogeneous datasets (census data, mobile phone calls, etc).
Naturally, it is possible to enforce a metrical structure on such relational networks. However, this generally destroys important information contained within the dataset, because it forces the distances between nodes to obey additional properties. This is clearly shown by the analysis real-world weighted networks [1]: we used a selection of standard network metrics and found that the resulting homological features, although robust to metric changes as expected, were scarcely informative.
This problem can be overcome by using the actual relational structure and its weighting patterns. We devised a filtration mechanism, akin to a stratigraphy of the dataset, which allows us to preserve and extract the complete homological structure without further assumptions.
This is done by considering the set of all filtered networks, ordered by the descending thresholding weight parameter, in the spirit of persistent homology [2]. Studying the changes of the topological structure along such filtration provides a natural measure of robustness for the topological features emerging across different scales. This method allows us to probe multiple layers of organized structure and is also robust to decimation procedures [3]. In particular, in a selection of real-world complex networks, we highlighted two classes, distinguished by their homological features, which were then found to be caused by differences in the higher order interactions. Moreover, this was linked to differences in the spectral properties of the networks, which have been long known to affect the evolution of dynamical processes on networks, with particular reference to the algebraic connectivity and the synchronizability thresholds.

Among the applications, two – relevant for social and infrastructural networks – are the study of the weighted rich club's geometry beyond the aggregate measure, and the generalisation of network embedding models to include homological information. This is ongoing work that will continue also within the context of deliverable D1.2 because of its focus on the relationship between geometrical and topological properties in complex systems (through bounded Gromov spaces).

The mathematical consistency and generality of the method introduced was discussed in [5]: in brief, from within category theory, the introduced weighted graph homology can be shown to be the most general framework for persistent homology. In addition, the work described above resulted in a publicly available Python module, Holes, which streamlines the steps necessary for the complete analysis (pre-processing, persistent homology computation and analysis of the results).

Finally, ISI partnered with the Institute of Psychiatry (King's College London) to apply persistent homology techniques to the discrimination of drug effects on brain functional networks obtained from fMRI data. In particular, ISI performed the experimental design, crafted and performed the analysis on the datasets provided by King's College. Homological features were instrumental in this work, since the differences found between placebo-injected and drug-injected patients were at the level of mesoscopic structures, i.e. functional circuits, and were thus invisible to the standard global or local metrics customarily used in network science.

ISI-UNICAM: The computational complexity involved in preprocessing of large datasets and in calculating the related persistent homology warrants appropriate computational tools. However, despite the promising results of topological data analysis, only separate software

libraries designed for specific purposes exist. In order to allow wide-spread adoption of topological techniques, an efficient and open-source toolbox is necessary for the scientific community at large. The two groups are currently collaborating on the first stages of this project. In particular the first step requires increasing the processing capabilities of ***Holes*** while maintaining its accessibility. This is being done by developing a high-performance Java library, ***jHoles***, which maintains the interoperability with the Python module. A prototype has been finished and is undergoing stress tests. Preliminary tests showed a very large performance increase (>200%) in terms of speed and memory management.

ISI-AMU: One very interesting problem in Hamiltonian dynamics is that of the microscopic origin of phase transitions. This question can be approached by characterizing the changes of the system's configuration manifold in terms of its topological properties. In particular, two theorems guarantee that – for a large class of potential functions – phase transitions can be related to sudden topological changes of the manifold. However, due to the large number of degrees of freedom and the computational complexity, it was impossible to directly verify this.
Persistent homology allows instead direct access to the topology and its shifts of such large-dimensional manifolds, allowing direct confirmation of the theory. Preliminary results of this work have been presented at European Conference on Complex Systems 2013 within the "Topological Methods for Complex Systems" satellite and a publication is in preparation.

3. Outlook (no cost is claimed for this work)

In the outlook, ISI's contribution (weighted graph persistent homology) provides a basis for Task 1.2, in which ISI will check the suitability of Gromov's spaces of bounded geometries to encode the relationship obtained by WP2 between geometric and topological quantities. A crucial goal is to prove that the partition function is well-defined for appropriate control parameters whose values are most likely determined by the torsion invariants associated with an orthogonal representation of the fundamental groups of the set of manifolds. The work ongoing with AMU on the topology of configuration manifold of critical Hamiltonians already constitutes a stepping stone in this direction, in particular toward understanding whether the phase transitions are described by a passage from a simple homotopy type to another.

Work developed so far was dictated by the necessity to define a suitable measure on the space of data, allowing us to construct the statistical functionals (first of all entropy, but indeed any probability distribution and the related moments) necessary to single out and evaluate correlations (patterns ≡ knowledge) in such space. Preliminary results are summarized hereafter. No publication has been completed yet, but active work on the subject is in progress, expected to lead soon to two manuscripts.

It is well known that in the continuous case there are relationships between the heat flow, acting on differential forms on a closed oriented manifold M, and the topology of M. In view of Hodge theory, one can recover the Betti numbers of M from the heat flow. Furthermore, Ray and Singer defined an analytic torsion (nonzero only on odd-dimensional manifolds), smooth invariant of acyclic flat bundles on M, proved by Cheeger and Muller to be equal to the classical Reidemeister torsion. This analytic torsion behaves in some way as an odd-dimensional counterpart of the Euler characteristic.

If M is not simply-connected, then there is a covering space analog of the Betti numbers. Using the heat flow on the universal cover of M, one can define the L^2-Betti numbers of M by taking the trace of the heat kernel not in the ordinary sense, but as an element of a type II von Neumann algebra. This essentially amounts to integrating the local trace of the heat kernel over a fundamental domain in M. The covering space analog of the analytic torsion, T(M), could be considered as well, having the same relation to the L^2-cohomology as the ordinary analytic torsion bears to de Rham cohomology.

We aim to computing T(M) in the case where M is hyperbolic. T(M) is clearly proportional to the hyperbolic measure ('volume'), which is expected to be a topological invariant by Mostow rigidity. The question is whether the constant of proportionality is nonzero. The invariant of M which in the PL case gives the hyperbolic 'volume' is the simplicial volume ||M|| of Gromov. The relationship between ||M|| and T(M) needs to be made more explicit.

4. Appendix: attached papers that are integral part of this deliverable

1: [ISI]. G. Petri, M. Scolamiero, I. Donato, F. Vaccarino. Networks and Cycles: A Persistent Homology Approach to Complex Networks. Proceedings of the European Conference on Complex Systems 2012. 93-99 (2012)

2: [ISI]. I. Donato, G. Petri, M. Scolamiero, L. Rondoni, F. Vaccarino. Decimation of Fast States and Weak Nodes: Topological Variation via Persistent Homology. Proceedings of the European Conference on Complex Systems 2012. 295-301 (2012)

3: [ISI]. G. Petri, M. Scolamiero, I. Donato, F. Vaccarino. Topological Strata of Weighted Complex Networks. PLoS ONE. e66506.doi:10.1371/journal.pone.0066506 (2013)

4: [ISI]. G. Petri, P. Expert, R. Carhart-Harris, P. Hellyer, D. Nutt, F. Turkheimer, F. Vaccarino. Homological backbone of brain functional networks. Submitted (2013).

5: [ISI]. F. Vaccarino, M. Scolamiero, G. Petri. One graph to rule them all. Internal report. (2013).

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

T. Gilbert et al. (eds.), *Proceedings of the European Conference on Complex Systems 2012*, Springer Proceedings in Complexity, DOI 10.1007/978-3-319-00395-5_15,

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

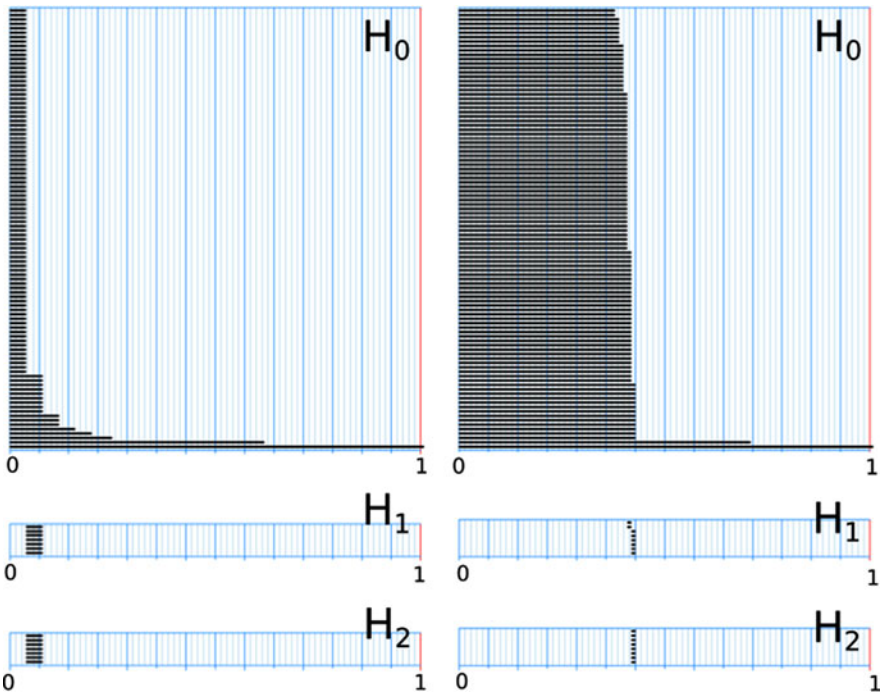


Fig. 15.1 Barcodes for the shortest path metric (*left*, panels (**a**), (**c**) and (**e**)) and the Von Neumann metrics (*right*, panels (**b**), (**d**) and (**f**)) on the C. Elegans brain network. From *top* to *bottom*, we report the intervals of existence of the homological spaces H_0 (panels (**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, $H_2 \ldots$, 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.

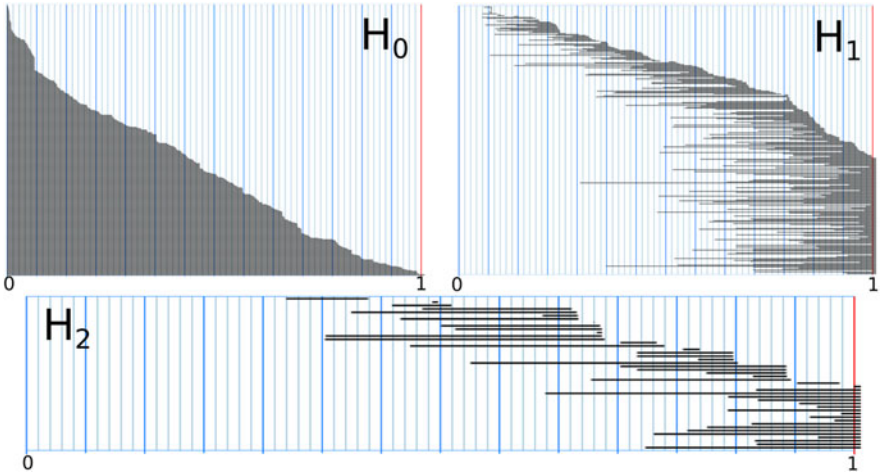


Fig. 15.2 Barcodes obtained from the weighted-clique complex filtration of a network of online contacts for the homology groups H_0 (**a**), H_1 (**b**) and H_2 (**c**). Persistent H_1 and H_2 generators imply that the existence of loops and chains of tethraidra formed by nodes which are weakly interacting 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.

Acknowledgements The authors acknowledge Mario Rasetti for insightful discussions and constant support.

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

Chapter 39
Decimation of Fast States and Weak Nodes: Topological Variation via Persistent Homology

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

Abstract We study the topological variation in Markov processes and networks due to a coarse-graining procedure able to preserve the Markovian property. Such coarse-graining method simplifies master equation by neglecting the *fast* states and significantly reduces the network size by decimating weak nodes.

We use persistent homology to identify the robust topological structure which survive after the coarse-graining.

Keywords Markov processes · Complex networks · Coarse-graining (theory) · Persistent homology · Computational topology

39.1 Introduction

Networks have received large attention over the last decade [5, 6] because of their ability to encode complex behaviours in simple ways, namely through the topology of their connectivity and the type of interactions between their elements. Furthermore, such interactions often evolve according to stochastic rules, characteristic that tightly connects complex networks to Markov processes. However, the salient features of large networks can be hard to identify due to the sheer size of the systems and heterogeneity in linking and weight distributions. In this context, effective techniques to highlight dominant structures within large networks are of extreme importance both for control and understanding [3]. In particular, such techniques have to preserve chosen network properties while reducing the complexity of the system.

We analyse a coarse-graining procedure inspired by the method recently developed for continuous Markov processes [1, 2]. This procedure simplifies the numer-

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

I. Donato (✉) · L. Rondoni · F. Vaccarino
Dipartimento di Scienze Matematiche, Politecnico di Torino, C.so Duca degli Abruzzi n. 24, 10129 Torino, Italy

L. Rondoni
INFN, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy

T. Gilbert et al. (eds.), *Proceedings of the European Conference on Complex Systems 2012*, Springer Proceedings in Complexity, DOI 10.1007/978-3-319-00395-5_39,

ical treatment of chemical and biological processes in terms of master equations reducing the number of variables with a decimation of *fast* states and a subsequent renormalization of the weights of all the surviving states in order to preserve the Markovian property.

In particular, we slightly modify the method of [1] in order to extend it to the case of complex networks. Whereas, in the case of Markov processes, the faster states are removed, in the case of networks we consider as candidates for decimation the nodes displaying smaller strength i.e. weak nodes. For every decimated node i, we introduce edges between all pairs of i's neighbours, effectively substituting the node itself with a *clique* composed by its former neighbours. The renormalization procedure of assigning weights to these new edges is described in the following section.

Importantly, we show on a range of different networks that this decimation technique is able to preserve the robust homological properties of the system, obtained through a persistent homology approach, while at the same time reducing significantly the complexity of the computation. We see also that the topological features depend crucially on the variation of coarse-graining level.

We report the example of *C. elegans* brain neural network of the type of results obtained for networks under the procedure that decimates the weak nodes and an asymmetric scale-free graph associated to an irreversible Markov process, as example of the procedure that decimates fast states.

39.2 Decimation of Fast States and Weak Nodes

The coarse-graining procedure, considered here, allows to advance the understanding of a certain Markov process with very different time-scale by neglecting the fast dynamic. Indeed, if the process is described by Master equation (here ω_{ij} indicates the rate of going from state i to state j whereas P_i is the probability of being in state i)

$$\frac{dP_i}{dt} = \sum_{i \neq j} (P_i \omega_{ij} - P_j \omega_{ji}) \tag{39.1}$$

and we are interested in the slow dynamics, it is usually not necessary but computationally demanding to exactly integrate the fast dynamics.

In this method the coarse-graining is parametrized by some threshold i.e. the coarse-graining level. Particularly, we can decimate all states having an average permanence time smaller than a prescribed threshold $\Delta\tau$ (*fast* states), where the time spent in a generic state n is exponentially distributed with average $\tau_n = 1/\omega_n^{out}$ and ω_n^{out} is the sum of the outgoing rates from n.

Similar considerations can be made in the context of network analysis, where typically the main interest is the description of the underlying backbone of the system, which is usually defined in terms of link weights [3] or connectivity [4]. From this point of view, weak nodes, that is nodes with low outgoing strength, are good candidates for removal, since they do not contribute significantly to the main structure

of the network. In particular, given a strength threshold s and a node i, we decimate all the nodes with total outgoing strength $s_{ij} = \sum_{j \in \Gamma_i} \omega_{ij} < s$ (*weak* nodes).

It is important to replace the decimated nodes with effective interactions between the nodes they used to bridge. Namely, this means that a disappearing node i is substituted by links creating a fully connected *clique* among i's former neighbours. The weights of the *clique* edges are setted in such a way to preserve the Markovian propriety that leads to the following renormalization [1]:

$$\tilde{\omega}_i^j = \omega_i^j + \frac{\omega_i^n \omega_n^j}{\omega_n^{out}} \tag{39.2}$$

Note that this procedure corresponds to neglect the time spend in the removed state. We use the same renormalization (39.2) in the case of the network although the decimated nodes are the weak ones.

Interestingly, it was proved that, given a set of candidates for decimation, the result of this procedure does not depend on the order in which the nodes are removed, i.e. the procedure is commutative, widening its applicability and generality.

39.3 Persistent 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. 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. Moreover, the value of the filtration parameter at which a certain generator provides important information about its role within the network.

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 [7]. Here, we exploit the network's weighted clique structure. The clique complex is obtained by associating to each maximal k-clique, a completely connected subgraph formed by k nodes, the $(k-1)$-simplex whose vertices are the clique's nodes. In this way however one produces a single simplicial complex, which

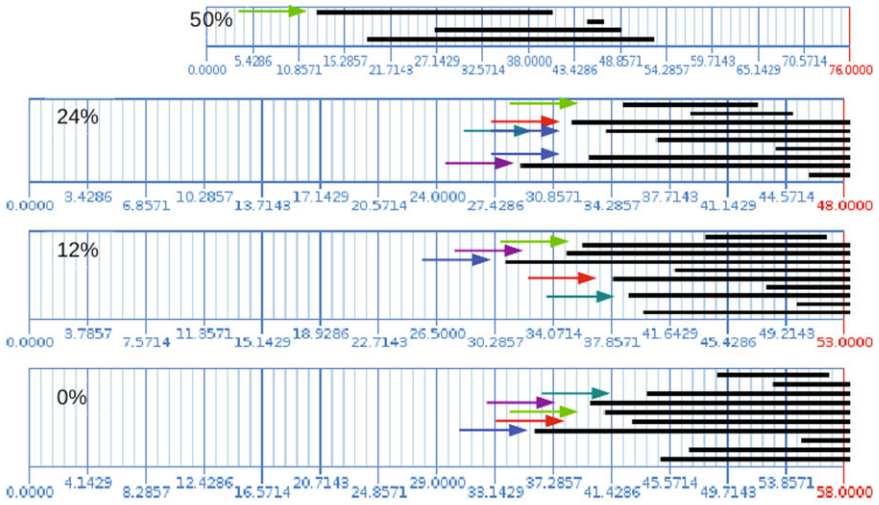


Fig. 39.1 The barcodes are shown with the percentage of the decimated nodes. *The arrows* indicates the persistence of the same loops. *The overlapped arrows* mean that two cycles have merged. We see the disappearance of the persistent cycles only when the decimation becomes massive

does not convey in itself any information about the robustness of the topological features it displays.

It is possible to devise a filtration, i.e. a sequence of simplicial complexes, combining link weights and clique structure. Given the weighted adjacency matrix ω_{ij}, we let ϵ vary in $\big(\min(\omega_{ij}), \max(\omega_{ij})\big)$ 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.

39.4 Topological Variation Due to the Decimation Procedure

We applied the decimation procedure to different weighted graphs representing real networks and than we have studied the relevant topological features to investigate if they are destroyed by the coarse-graining. We report two examples, one is the *C. elegans* neural network, where we decimated *weak* nodes, and the other one is an asymmetric weighted scale-free graph where we decimated *fast* states because of its interpretation as an irreversible Markov process.

We show the barcodes for the H_1 generators at different percentages of nodes removed for both the *C. elegans* neural network (Fig. 39.2) and the scale-free graph (Fig. 39.1). The barcodes summarize the topological information dictated by the dynamic of the network. Indeed, they are a collection of horizontal lines representing the homology generators in an arbitrary order. The starting and final points of these

segment indicate the filtration level where they have appeared and disappeared. In particular, the H_1 generators are the cycles i.e. edges set in circular way with weight stronger then the near ones. The moment in the filtration in which they appeared gives information about the scale of the cycles (or voids if we are taking in consideration the H_2 generators). Indeed, persistent homology is a tool able to see also the structures (cycles, voids etc. ...) in the meso-scale and, furthermore, able to show the edges and nodes constituting such structures. In Figs. 39.2 and 39.1, for example, arrows of the same color highlight cycles that are present both in the initial graph and in the decimated ones. Hence, we can distinguish between graphs displaying the same homology generators distribution or having the same generators.

Decimation and Persistent Homology Applied to a Random Weighted Scale-Free Markov Process We have applied the coarse-graining procedure to a 50 states continuous Markov process, obtained from a directed scale-free graph. The motivation for applying this method to a scale-free graph is that it is recognized to reproduce observed properties of the world-wide web [9] and that this coarse-graining procedure is especially useful to integrate master equation when there are a lot of different scales.

The model used to create the initial data gives a graph that grows with preferential attachment so that the in- and out-degrees distributions follow power laws [8]. After we added random weights to each edge and divided the weights from i to j by the number of neighbours of i in order to prevent that the lifetime of state i would be the shorter the higher is its degree centrality. This step permit also to increase the separation of time scales. The H_1 generators for the full network and the ones corresponding to 12 and 24 percent of the weaker states removed (Fig. 39.1) evidence a good preservation of the relevant cycles. If a node of a certain cycle is removed we observe the shortening of that cycle in the coarse-grained graph. There could also be the merging of two cycles when a node to be decimated belong to both structures.

Importantly, the decimation procedure was also applied to a graph obtained in the same way that the previous one but without dividing the weights by the number of neighbours. We observed in this case that the cycles after the coarse-graining are not so long preserved because of the higher degrees of the fast states.

Decimation and Persistent Homology Applied to C. Elegans Neural Network Fig. 39.2 reports the barcode for the H_1 generators, which correspond to loops composed by strong links, obtained from the weighted clique persistent homology after a series of decimation. The plots refer in particular to respectively 40, 60, 80 and 85 percent of the network nodes being removed. Interestingly, we find that a set of generators, characterized by an early appearance in the barcode and a long persistence, that is emerging at large weight thresholds, survive through the decimation. This is particularly interesting especially when considering the high percentages of node removal, hence highlighting the extreme robustness and importance of the topological network features identified through their persistence. The other smaller cycles instead disappear under the decimation results, confirming their role as topological noise.

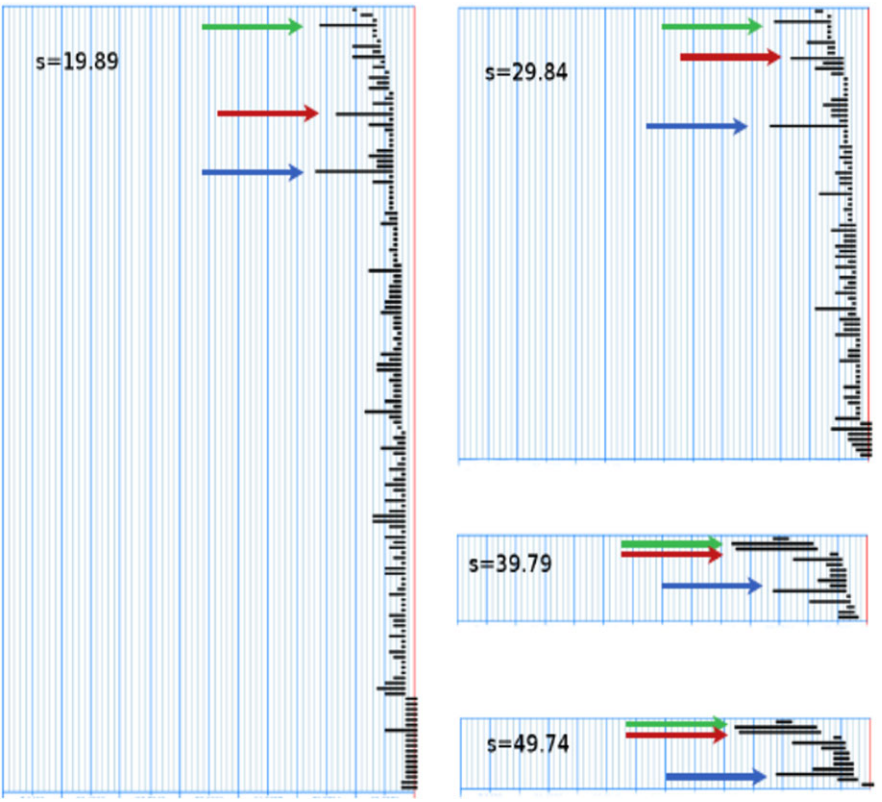


Fig. 39.2 In the first two figures (the ones corresponding to a threshold values $s = 19.89$ and $s = 29.84$) the generators are the same although decreased because of a reduction of the 34 per cent of the nodes between the two figures. The cycles start to change for a threshold value $s = 39.79$, where the reduction of the nodes become very large (more then 50 per cent with respect to *the above figure*), but an accurate study bring us to the conclusion that the cycles are only slightly varied. For example, the first stripe on the figure with $s = 29.84$ corresponds to the persistence cycle between the nodes 71, 74, 216, 207, whereas the first stripe on the figure with $s = 19.89$ corresponds to the cycles 71, 74, 216, 305 i.e. 305 takes the place of 207 that, effectively, was decimated

39.5 Conclusion

We used the persistent homology to explore topological features dictated by the dynamics on networks and Markov processes.

We conclude that the relevant topology features are essentially preserved under a coarse-graining based on decimation of weak nodes and fast states. The dynamics, before and after the decimation procedure, concentrates on the same cycles provided that the nodes even if fast should not be of the higher degree.

Acknowledgements L. Rondoni gratefully acknowledges financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-

2013)/ERC Grant agreement No. 202680. The EC is not liable for any use that can be made on the information contained herein.

References

1. Pigolotti S, Vulpiani A (2008) J Chem Phys 128:154114
2. Puglisi A, Pigolotti S, Rondoni L, Vulpiani A (2010) J Stat Mech 2010:P05015
3. Serrano M, Boguñá M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. Proc Natl Acad Sci USA 106(16):6483
4. Grady D, Thiemann C, Brockmann D (2011) Parameter-free identification of salient features in complex networks. arXiv:1110.3864
5. Albert R, Barabasi A (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47–97
6. Dorogovtsev SN, Goltsev AV, Mendes JFF (2008) Critical phenomena in complex networks. Rev Mod Phys 80(4):1275–1335. doi:10.1103/RevModPhys.80.1275
7. Zomorodian A, Carlsson G (2005) Computing persistent homology. Discrete Comput Geom 33(2):249–274
8. Bollobás B, Borgs C, Chayes J, Riordan O (2003) Directed scale-free graphs. In: Proceedings of the fourteenth annual ACM-SIAM symposium on discrete algorithms, pp 132–139
9. Barabasi A, Albert R (1999) Emergence of scaling in random networks. Scienze 286:509–512. Also in Physica A 281:69–77 (2000)

OPEN ACCESS Freely available online

PLOS ONE

Topological Strata of Weighted Complex Networks

Giovanni Petri[1]*, Martina Scolamiero[1,2], Irene Donato[1,3], Francesco Vaccarino[1,3]

1 ISI Foundation, Torino, Italy, **2** Dipartimento di Ingegneria Gestionale e della Produzione, Politecnico di Torino, Torino, Italy, **3** Dipartimento di Scienze Matematiche, Politecnico di Torino, Torino, Italy

Abstract

The statistical mechanical approach to complex networks is the dominant paradigm in describing natural and societal complex systems. The study of network properties, and their implications on dynamical processes, mostly focus on locally defined quantities of nodes and edges, such as node degrees, edge weights and –more recently– correlations between neighboring nodes. However, statistical methods quickly become cumbersome when dealing with many-body properties and do not capture the precise mesoscopic structure of complex networks. Here we introduce a novel method, based on persistent homology, to detect particular non-local structures, akin to *weighted holes* within the link-weight network fabric, which are invisible to existing methods. Their properties divide weighted networks in two broad classes: one is characterized by small hierarchically nested holes, while the second displays larger and longer living inhomogeneities. These classes cannot be reduced to known local or quasilocal network properties, because of the intrinsic non-locality of homological properties, and thus yield a new classification built on high order coordination patterns. Our results show that topology can provide novel insights relevant for many-body interactions in social and spatial networks. Moreover, this new method creates the first bridge between network theory and algebraic topology, which will allow to import the toolset of algebraic methods to complex systems.

Citation: Petri G, Scolamiero M, Donato I, Vaccarino F (2013) Topological Strata of Weighted Complex Networks. PLoS ONE 8(6): e66506. doi:10.1371/journal.pone.0066506

Editor: Renaud Lambiotte, University of Namur, Belgium

Received January 26, 2013; **Accepted** May 7, 2013; **Published** June 21, 2013

Funding: GP is supported by the TOPDRIM project funded by the Future and Emerging Technologies program of the European Commission under Contract IST-318121. ID and MS are partly supported by Project Lagrange Ph.D. Grant. FV is partially supported by PRIN 2009 "Spazi di Moduli e Teoria di Lie". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: giovanni.petri@isi.it

Introduction

Complex networks have become one of the prominent tools in the study of social, technological and biological systems [1–3]. In particular, weighted networks have been largely used to convey not only the presence but also the intensity of relations between nodes in a network. Real-world networks display however intricate patterns of redundant links with edge weights and node degrees usually ranging over various orders of magnitudes [4,5]. This makes very hard to extract the significant network structure from the background [6–9], especially in the case of very dense networks [10,11]. Alongside topological filtering methods [12,13], the typical approach to this problem is to choose a suitable threshold for the edge weights, e.g. global [10] or local [14], and study the reduced graph composed by only the edges of weight larger (smaller) than the threshold parameter. In any case, some properties of the original graph are inevitably lost under such transformation.

To avoid this pitfall, given a weighted network G we consider the set of all filtered networks, $\mathcal{F}(G)$, ordered by the descending thresholding weight parameter, in the spirit of *persistent homology* [15–18].

Persistent homology is a recent development in computational topology designed for robust shape recognition and data-discovery from high dimensional datasets [19]. It has found successful application in various fields, ranging from biological systems (e.g.brain correlation networks [20] and breast cancer diagnosis [15]), computer vision and sensor network coverage problems [15] all the way to the analysis of large scale cosmological structure [22]. Its central device is the construction of a simplicial *filtration* of the original dataset: data points are usually embedded in a metric space in order to extract from their configuration a sequence of growing simplicial complexes, which approximates with increasing precision the original dataset. Studying the changes of the topological structure along such filtration provides a natural measure of robustness for the topological features emerging across different scales. In analogy to the metric example, we call the set $\mathcal{F}(G)$ *graph filtration*: considering the set of all filtered networks captures the link weights and connectivity structure over all weight scales, without the need to resort to any assumption on an eventual metric structure underlying the graph structure. The graph filtration of a network Ω is built following these steps :

- Rank the weights of links from ω_{max} to ω_{min}: the discrete parameter ϵ_t scans the sequence.
- At each step t of the decreasing edge ranking we consider the thresholded graph $G(\omega_{ij},\epsilon_t)$, i.e. the subgraph of Ω with links of weight larger than ϵ_t.

Figure 1a provides a schematic illustration of the rank filtration. This approach preserves the complete topological and weight information, allowing us to focus on special mesoscopic structures: *weighted network holes*, that relate the network's weight-degree structure to its homological backbone.

A weighted network hole of weight ω is a loop composed by n nodes $i_0,i_1,i_2,\ldots,i_{n-1}$, where all cyclic edges (i_l,i_{l+1}) (with $i_0 \equiv i_n$)

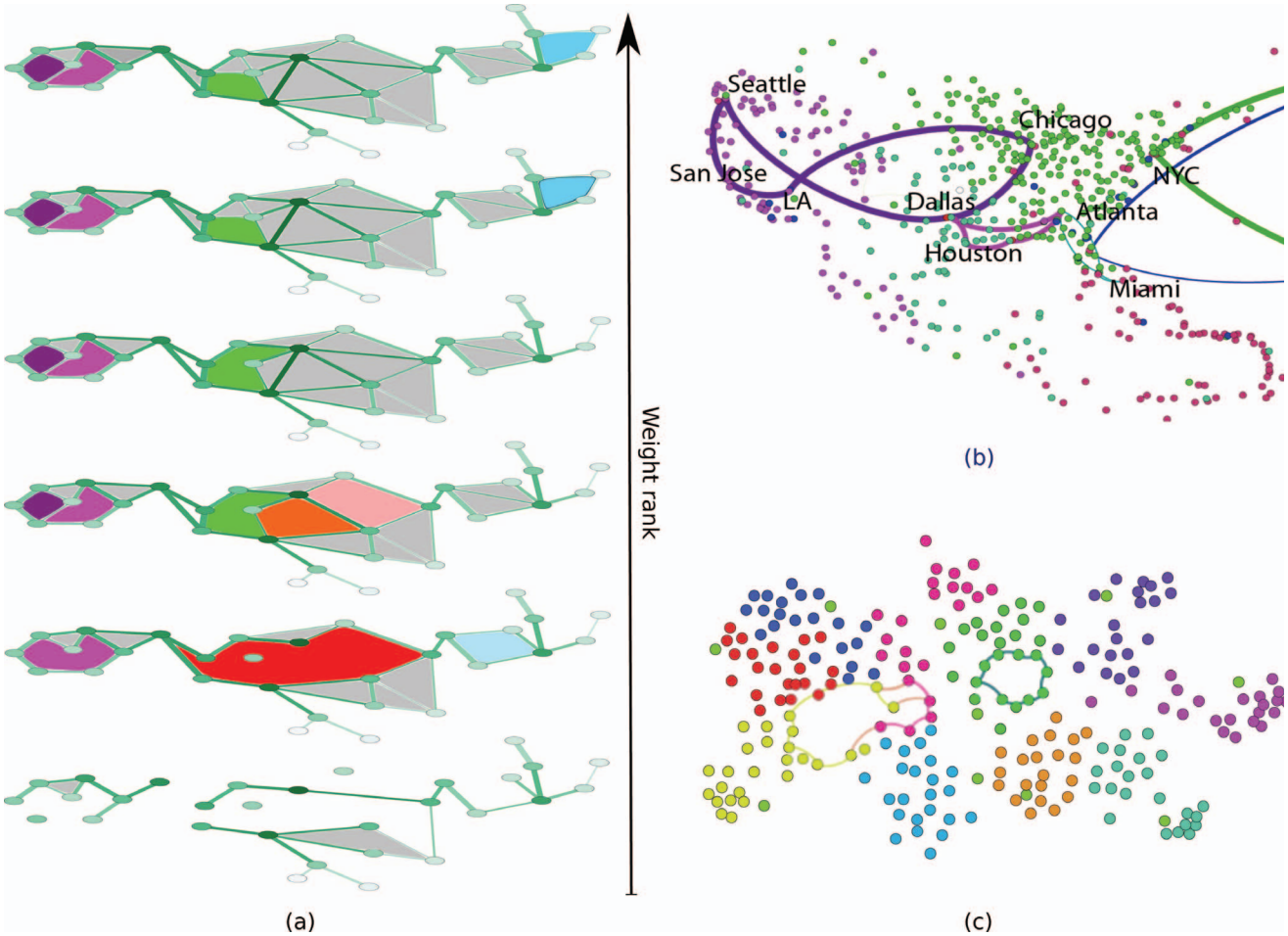


Figure 1. Weight rank clique filtration and homology of networks. *(a)* The weight rank filtration proceeds from the bottom up. Weighted holes (colored) and cliques (gray) appear as links are added. Weighted holes can branch into smaller holes, which have then independent evolution, persisting or dying along the filtration as links close them by 3-cliques. The cartoon shows two very long-persistence holes (violet and purple) appearing quite early and living until the end, while the largest hole (red) branches into three smaller holes, of only one survives to the end of the filtration (green). *(b)* A selection of weighted holes from the US air passenger network (year 2000). The node colors represent the best modularity partition of the entire network. The cycles are all long-persistence one, chosen to represent different behaviors: for example, the Chicago-Los Angeles-San Jose-Seattle cycle spans a large spatial distance, implying weaker connectivity across the cycle and within the region encompassed by the cycle, while the cycle going east from New York connects the east coast to three large European network and its persistence is due to the reduced connectivity due to the Atlantic Ocean. (c) A selection of the strongest cycles in the face-to-face contact network in a primary school (see SI for details on dataset). Node colors represent different classes in the school. Cycles are often found across communities, since by definition they probe the presence of holes among network regions. However, this is not the only information they convey. The cycle contained in a single community (green) testify the presence of peculiar contact geometries even within dense community structures.
doi:10.1371/journal.pone.0066506.g001

have weights $\geq\omega$, while all the other possible edges crossing the loop are strictly weaker than ω. We focus on this special class of subgraphs, because formally such weighted holes represent the generators of the first homology group, H_1, of the clique complex of the graph thresholded by weight ω (see Materials and Methods). The aim of this paper is to characterize the evolution of these generators along the network filtration. As we swipe the network from the largest to the smallest weights, network holes appear and potentially close.

By unearthing their properties, we obtain the main contribution of this paper: the statistical features of weighted network holes yield a classification of real-world networks in two classes, depending on the compatibility or lack thereof with null models generated by graph randomisations. Furthermore, this classification is defined by mesoscopic homological structures that cannot be reconduced to local properties alone.

The method used for the classification itself, which we call *weighted clique rank homology*, is the second novel main contribution of this paper. It allows to recover complete and accurate long-range information from noisy redundant network data, by building on persistent homology [16], a recent theory developed in computational topology [17], which we extend to the case of networks.

Each weighted hole g is characterized by three quantities: its birth index β_g, its persistence p_g and its length λ_g. After ranking links in a descending order according to their weights, the birth index of a hole is the rank t of its weight ω. As we proceed adding links to the filtration in ranking order, it is possible that a link with rank $t'>t$ will appear and cross the hole. We call this closure of the weighted hole, or *death* δ_g. The persistence p_g is the interval between the birth and death of g, $p_g=\delta_g-\beta_g=t'-t$. Finally, the length λ_g is the number of links composing g.

Similarly to stratigraphy, each step of the filtration is a topological stratum of the network, where the edge weight rank plays the role of depth. Intuitively, g can then be thought as an underground cavity, hidden in the link-weight fabric of the network, and β_g, p_g and λ_g as its maximal depth, vertical size and girth respectively.

Results

Homological Network Classes

We applied this analysis to various social, infrastructural and biological networks (see SI for a detailed list). In order to compare datasets, indices are normalized by the corresponding filtration length (maximal rank) T, so that all β_g, δ_g, and thus p_g, vary in the unit interval. In addition, we compared each dataset with two randomized versions, obtained by weight reshuffling and edge-swapping respectively. While both randomisations preserve the weight and degree sequences (and the relative distributions ($p(k)$ and $p(\omega)$), the first one redistributes only the edge weights and is meant to destroy weight correlations, preserving the joint degree distribution $p(k,k')$ and thus the degree assortativity. The second instead randomizes the network through double-edge swaps, preserving $p(k)$ and $p(\omega)$ but destroying both weight and degree correlations [23]. We stress that, as the degree and weight sequences are preserved in the randomisations, they cannot account for the differences in the observed homology.

The statistical distributions obtained for the $\{\beta_g\}$, $\{p_g\}$ and $\{\lambda_g\}$ for H_1 cycles highlight a natural division of the analysed networks in two broad classes (Fig. 2):

Class I networks. cycle distributions are markedly different from the randomized versions (cycles display shorter persistence times, earlier and broader birth distributions and very short lengths as compared to their randomized versions);

Class II networks. cycle distributions are very close to their random versions (late appearance, short persistences, long cycles).

The short cycles of Class I networks nest hierarchically and appear and die over all scales while those in the randomized counterparts are born uniformly along the filtration but are more persistent, producing largely hollow network instances. The implications are twofold. Since cycles represent weaker connectivity regions, this results in class I networks being more *solid* than the randomized versions, while class II networks resemble more closely the randomized instances. Second, since the cycle abundance ratio between real and random instances is the same in the two groups, the differences between class I and II does not depend on cycle abundance, but rather on their properties.

This can be seen easily by compressing the whole information within two scalar metrics which do not depend on the number of generators in a given network filtration. We define the *network hollowness* h_i and the *chain-length normalized hollowness* $\tilde{h}_i$ as:

$$h_k = \frac{1}{N_{g_k}} \sum_{g_k} \frac{p_{g_k}}{T} \tag{1}$$

$$\tilde{h}_k = \frac{1}{N_{g_k}} \sum_{g_k} \frac{\lambda_{g_k}}{N} \frac{p_{g_k}}{T} \tag{2}$$

where $\{g_k\}$ is the set of generators of the k-th homological group H_k and $N_{g_k} = \dim H_k$ their number. The first is a measure of the average generator persistence, while the second weights generators according to both their length and persistence. Table 1 reports the values for h_1 and $\tilde{h}_1$. Class I networks have lower hollowness values as compared to their randomized versions, while class II ones show comparable values.

Interestingly, the hollowness values for the H_2 generators mostly vanish for the randomized instances (Table 1), as opposed to the case of real networks. It appears that, while persistent one-dimensional cycles are more easily generated in the randomized instances, higher forms of network coordination, e.g. H_2 generators (akin to two-dimensional surfaces bounding three-dimensional voids), do not only display different properties in comparison to the real network, but are instead wiped away. These findings hint therefore to the presence of higher order coordination mechanisms in real world networks.

Naturally, the two network classes do not represent a binary taxonomy and should be considered as two extremes of a range over which networks are distributed. For example, we find networks that interpolate between these classes, e.g. the online messages network has short persistence intervals, but also late cycle appearances and short length cycles. However, classes do not appear to display uniform behavior for local and two-body quantities: degree- and weight-distributions and correlations are mixed within the same group and do not provide a direct answer for the nature of the two classes. Similarly, a recently proposed measure of structural organisation, *integrativeness* [24], which measures the neighborhood overlap around strong links, does not provide insights to explain class I, since within the latter one finds both integrative and dispersive networks.

Finally, the classes do not show a consistent pattern in *assortativity*: for example, class I includes the gene network (assortative) and the airport networks (disassortative), while class II includes the assortative co-authorship networks and the disassortative Twitter data. Therefore, assortativity cannot be the discriminating factor between classes.

Higher Order Organization

Because homology is essentially a non-local property, it was expectable that the local measures mentioned would not be able to explain the observed homological patterns. Network homology can be seen in fact as the weighted complement to the *perturbative dK*-series approach [8]: the latter proceeds by successive bottom-up constraints on k-body correlations, rapidly becoming very cumbersome, while our method returns the complete superposition of the network's degree and weight correlation layers in a non-perturbative (top-down) fashion.

A simple artificial network helps illustrating this point: Random Geometric Graphs (RGG) have been recently shown to display long-range many-body correlations [25,26]. We find also that they have homological structures reminding of class I networks (Fig. 2a, b and c) and the same relation to their randomized versions. Class I networks are the result of high-order coordination in a similar way. This is supported also by the presence in real networks and RGGs of higher homology generators, which require elaborate coordination patterns in order to appear. While these cycles almost disappear in randomized versions of real-world networks, they are present in the case of RGGs.

For the latter and the airports, this organisation can be thought as the result of the non-local constraint imposed by the metric of the underlying space [27]. Although spatial constraints are harder to fathom for social and genetic systems, alternative explanations are possible: for example, the homological structure of the observed online communication and gene networks can be thought as stemming from group interactions among people (e.g. mailing lists, multi-user mails) and biological functions (e.g.

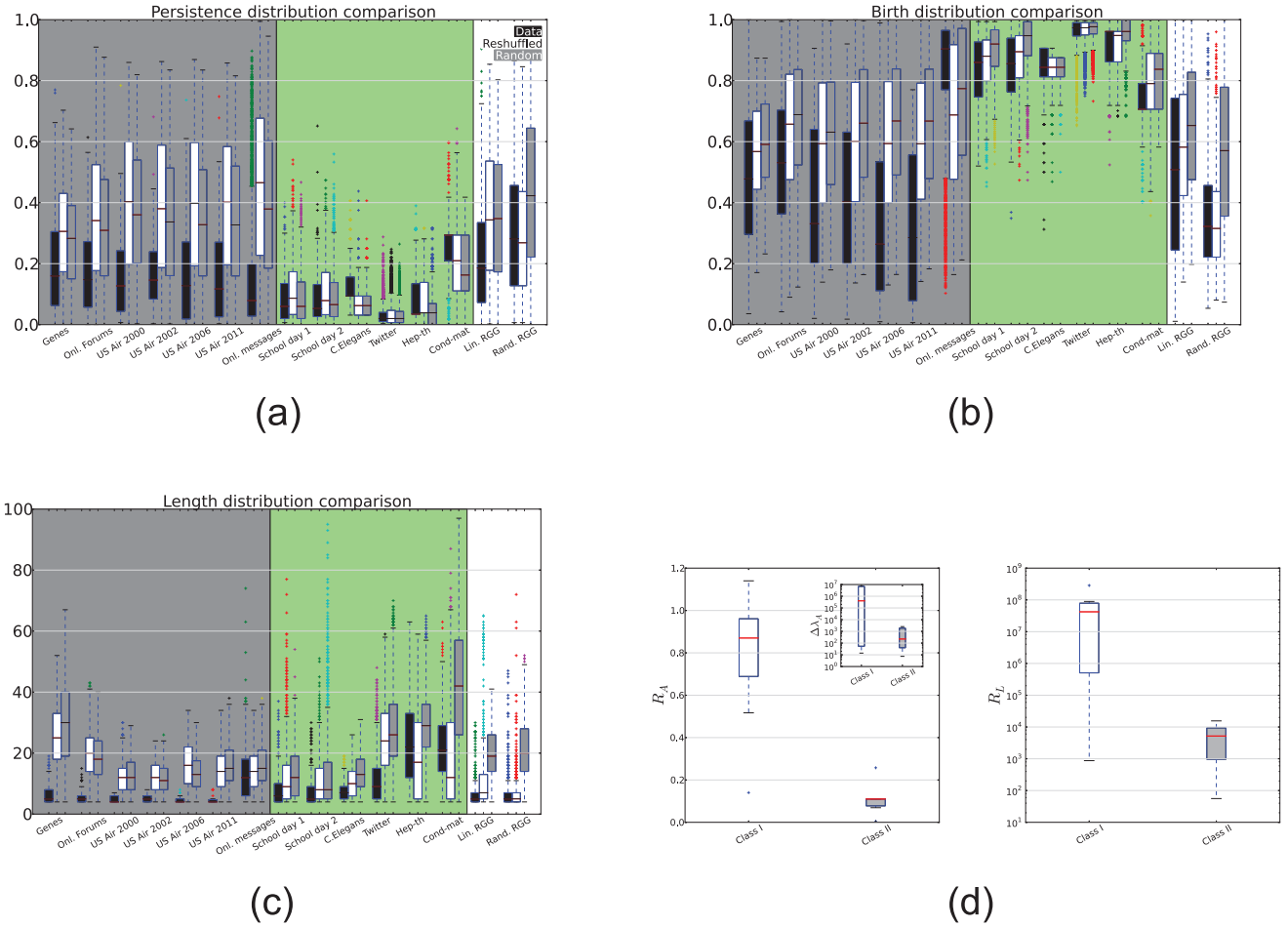


Figure 2. Statistical and spectral properties of H_1 generators. Box plots of the distributions of persistences $\{p_g\}$ (panel *a*)), births $\{\beta_g\}$ (panel *b*)) and lengths $\{\lambda_g\}$ (panel *c*)) for the 1d cycles (H_1 generators) of real networks (*black*), reshuffled (*white*) and randomized (*gray*). The gray and green shaded areas identify the two network classes described in the main text: class I is significantly different from the random expectations, with shorter, less persistent cycles that appear across the entire filtration; class II networks are not significantly different from the random versions, with long cycles and late birth times in the filtration. The characteristics of class I networks imply a stratification of cycles that betrays the presence of large, non-local organisation in the network structure, which is not present in class II networks. For comparison, an example of RGG network (600 nodes in the unitary disk, linking distance 0.01), known to have higher order degree correlations, had edge weights set according to $\omega_{ij} \propto (k_i k_j)^{\theta}$, with $\theta = 1$ (linearly correlated weight RGG) and $\theta = 0$ (random weight RGG). In both cases, the distributions of cycles' properties resemble closely those of class I networks. Panel *d*) finally reports the distribution of adjacency spectral gaps $\Delta\lambda_A$ and R_A (left plot) and the Laplacian eigenratio R_L (right plot). All the quantities show significant ($p < 0.05$) differences between the two classes, implying that the homological structure affect the dynamical properties of networks, e.g. the synchronizability threshold.
doi:10.1371/journal.pone.0066506.g002

pathways) respectively, which provide an underlying non-local mechanism for the emergence of homological patterns.

Further evidence of this behavior can be found by zooming on specific cycles which convey information about underlying constrains hidden in the network weight-link connectivity patterns. For example, the cycle structure of the air passenger network detects the expected reduced connectivity over oceans in the form of strong persistent cycles– and the strong backbone of US airport hubs, which is then filled by the local (intra-community) links (Fig. 1b). Another example can be found in the school children's face-to-face contact network. As expected we find the most significant cycles to link together different school classes (yellow and pink cycles in Fig. 1c). However, we also find that a school class (green nodes), despite being both a network community and 3-clique component [28], is characterized by a strong internal H_1 generator, which might be reflecting peculiar social dynamics coming from same-gender biases, different seating arrangements or schedules for part of the class [29].

Spectral Correlates of Homology Classes

At the opposite extreme of local quantities lie the spectral properties of networks. It is very important therefore to investigate whether it is possible to highlight peculiar spectral signatures of the two classes. Network eigenvalues, especially those of the Laplacian matrix, figure prominently in a number of applications, ranging from spectral clustering [30] to the propensity to synchronize of a set of oscillators distributed on the nodes [31]. Given a graph G, we denote its adjacency matrix $A(G)$ and its Laplacian matrix as $L(G) = D - A(G)$, where $d_{ij} = \delta_{ij} \sum_k a_{ik}$. For a symmetric network with N nodes, $A(G)$ has a set of real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_{N-1} \geq \lambda_N$. The spectral gap $\Delta\lambda_A = \lambda_1 - \lambda_2$, and its normalized version, $R_A = \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_N}$, effectively measure how far the

leading eigenvalue lies in comparison to the bulk of the eigenvalue distribution [32].

Interestingly, we find that class I networks have significantly larger spectral gaps ($p<0.05$ comparing the distributions) than class II networks (in Fig. 2d and Table 2 for information on individual datasets). Despite being somewhat neglected in the complex networks literature, $\Delta\lambda_A$ has been linked to the notion of natural connectivity [33]: it encodes spectral information about network redundancy in terms of the number of closed paths and is defined as $\bar{\lambda}=\log\left[\frac{1}{N}\sum_{i=1}^{N}e^{\lambda_i}\right]$. Rewriting $\bar{\lambda}=\lambda_1+\log\left[\frac{1}{N}(1+\sum_{i=2}^{N}e^{\lambda_i-\lambda_1})\right]$, it is easy to see that for large gaps all the terms in the sum are exponentially suppressed and therefore $\bar{\lambda}$ is essentially dominated by the leading adjacency eigenvalue modulo a size effect, $\bar{\lambda}\sim\lambda_1-\log N$. This result is consistent with the nested cycle structure that we highlighted in class I. More importantly, we find a difference between the two classes in the topological constraints to synchronization processes. For the Laplacian $L(G)$, label the set of eigenvalues $0=\lambda_1<\lambda_2^L\leq\lambda_3^L\leq\ldots\leq\lambda_N^L$ and define the Laplacian eigenratio $R_L=\frac{\lambda_N^L}{\lambda_2^L}$. Barahona and Pecora [34] showed that a set of dynamical systems, placed on the network's nodes and coupled according to the graph adjacency with a global coupling σ, has a linearly stable synchronous state if

$$R_L<\beta \tag{3}$$

where β is a purely dynamical parameter. This inequality implies that networks displaying very large R_L are hard (or impossible) to synchronize. Panel IVb of Fig. 2 shows again a significant difference between the two classes: class I networks have much larger eigenratios, making them hardly synchronizable.

Our results show therefore a deep connection between the homological network structure, the network spectral properties and their implications on network dynamics. Indeed, the role of mesoscopic structures in the stability and evolution of dynamical systems on networks is gradually emerging, as shown for example by recent work based on the concepts of basic symmetric subgraphs and their legacy eigenvalues in the global network spectrum [35], and is indeed being shaped by algebraic methods, well suited to capture the geometric information hidden within the network fabric.

Conclusions

Hitherto, the homological structure of weighted networks could not be systematically studied. Our method, grounded in computational topology, allows to probe multiple layers of organized structure. It highlighted two classes of network distinguished by their homological features, which we interpreted as caused by differences in the higher order networks organisations that are not captured by (quasi)local approaches.

Among the many possible applications, two very relevant ones for social and infrastructural networks are the study of the weighted rich club's geometry beyond the aggregate measure [23,36], and the generalisation of network embedding models to include homological information [37]. Furthermore, the two classes displayed also a marked difference in their spectral gap distributions and in particular in the values of the algebraic connectivity, implying that the different homological structures are correlated with different synchronizability thresholds.

This work therefore provides a stepping stone towards understanding the coupling between network dynamical processes and the network's homology.

Finally, the filtration's construction rule is flexible and can be readily adapted to other problems. Similarly to changing goggles, different edge metrics can be used (e.g. betweenness or salience [38]), the thresholding method varied (e.g. local thresholding [14]) or the filtration promoted to a filtering on two quantities (e.g. edge weight and time in a temporal network) using *multi-persistent* homology [39].

Materials and Methods

Datasets

The dataset analysed in this paper cover a broad range of fields, spanning social, infrastructural and biological networks. Figures S1–S15 in the File S1 report the analysis for the individual datasets as opposed to the class-aggregate of figure 2.

In detail, they are:

US air passenger networks. The networks refer to the years 2000, 2002, 2006 and 2011. The years were chosen to provide snapshots of the air traffic situation at 4–5 years intervals, plus one extra (year 2000) just before the events of 9/11 which significantly affected the air transportation industry. The data used are publicly available from the website of the Bureau of Transportation Statistics (http://www.transtats.bts.gov/). Individual flights between airports were aggregated on routes as defined by origin and destination cities. The weight reported is the yearly aggregated passenger traffic.

C.Elegans. The network is available at http://cdg.columbia.edu/cdg/datasets and reports a weighted, directed representation of the C. Elegans's neuronal network [40]. The network was symmetrized by summing the weights present on edges between the same nodes (given ω_{ij} and ω_{ji}, $\omega_{ij}^{symm}=\omega_{ji}^{symm}=\omega_{ij}+\omega_{ji}$).

Online messages and forums. The online messages network consists of messages in a student online community at University of California [41]. The online forum network refers to the same online community, but focuses on the activity of users in public forums, rather than on private messages [42]. Both networks are publicly available online at Tore Opsahl's website (http://toreopsahl.com/datasets/).

Gene network. The gene interaction network used in the paper is a sampling of the complete human genome dataset available from the University of Florida Sparse Matrix Collection. Each node is an individual gene, while the edges correlates the expression level of a gene with that of the genes (using a NIR score [43]). The node set of the analysed network was obtained by randomly choosing an origin node, then adding its neighborhood to the node set; the neighborhoods of the newly added nodes were then added to the node set recursively until a given number of nodes was obtained (in the case used the target number of nodes was $N=1300$). Then all the edges present in the original network between the nodes in the node set were added, effectively taking a connected subgraph of the original network. To reduce the computational complexity due to the large density of the graph, the weighted clique filtration was stopped at an edge weight of 0.09 (similarly to the choice made in [24]).

Twitter. The dataset consists of a network of mentions and retweet between Twitter users and is available online on the Gephi dataset page (http://wiki.gephi.org/index.php/Datasets). Weights are proportional to the number of interactions between a pair of users.

School face-to-face contact network. The dataset contains two days of recorded face-to-face interactions in a primary school.

Table 1. Summary of hollowness values.

Dataset (class)	h_1	$\tilde{h}_1$	h_1^{sh}	$\tilde{h}_1^{sh}$	h_1^{rnd}	$\tilde{h}_1^{rnd}$	h_2	$\tilde{h}_2$
Genes(I)	0.515	0.003	0.020 ± 0.001	0.0007 ± 0.00001	0.0151 ± 0.0004	0.00023 ± 0.00005	0.35	0.006
Online forums(I)	0.175	0.001	0.355 ± 0.005	0.007 ± 0.001	0.325 ± 0.005	0.007 ± 0.001	0.02	0.0003
US Air 2000(I)	0.160	0.001	0.405 ± 0.005	0.0065 ± 0.0007	0.358 ± 0.006	0.0060 ± 0.0005	0.02	0.0003
US Air 2002(I)	0.186	0.0008	0.39 ± 0.01	0.0037 ± 0.0003	0.34 ± 0.01	0.0034 ± 0.0003	0.23	0.002
US Air 2006 (I)	0.167	0.0005	0.398 ± 0.005	0.0036 ± 0.0005	0.348 ± 0.008	0.0032 ± 0.0003	0.165	0.001
US Air 20011(I)	0.181	0.0006	0.41 ± 0.01	0.0034 ± 0.0002	0.35 ± 0.01	0.0033 ± 0.0003	0.076	0.0007
Online messages(I)	0.21	0.0014	0.190 ± 0.002	0.0017 ± 0.0001	0.185 ± 0.002	0.0015 ± 0.0001	0.02	0.0003
School day 1 (II)	0.088	0.0034	0.113 ± 0.002	0.007 ± 0.001	0.093 ± 0.002	0.006 ± 0.001	0.015	0.0012
School day 2 (II)	0.090	0.0033	0.115 ± 0.002	0.0065 ± 0.0005	0.098 ± 0.003	0.0089 ± 0.0008	0.01412	0.00095
C. elegans (II)	0.0784	0.002	0.0745 ± 0.0017	0.001 ± 0.0001	0.0896 ± 0.0023	0.0041 ± 0.0005	0.058	0.002
Twitter (II)	0.03	0.0001	0.030 ± 0.001	0.0002 ± 0.0001	0.029 ± 0.001	0.0002 ± 0.0001	0.01	0.0001
Hep-th (II)	0.08	0.0002	0.075 ± 0.001	0.0002 ± 0.0001	0.0508 ± 0.0003	0.0002 ± 0.0001	–	–
Cond-mat (II)	0.26	0.0004	0.20 ± 0.003	0.0002 ± 0.0001	0.180 ± 0.002	0.0005 ± 0.0001	–	–
Lin. RGG	0.227	0.003	0.368 ± 0.005	0.006 ± 0.001	0.355 ± 0.002	0.012 ± 0.001	0.28	0.006
Ran. RGG	0.3	0.0041	0.299 ± 0.005	0.0045 ± 0.0002	0.649 ± 0.40	0.015 ± 0.001	0.115	0.003

Summary of hollowness values. For each dataset, we report the values of the *hollowness* h_1 and *cycle-length normalized hollowness* $\tilde{h}_1$ for H_1 cycles for real networks and their randomisations (*sh* and *rnd*). Most networks (class I in particular) show lower values than for their randomized versions. We also report the values of the *hollowness* h_2 and *cycle-length normalized hollowness* $\tilde{h}_2$ for H_2 cycles for real networks. The values for the randomized networks are not reported as –strikingly– the randomisations do not inline any higher homology, while almost all real networks inline positive values of the H_2 hollowness.
doi:10.1371/journal.pone.0066506.t001

Each node represents a child, with the edge weight between two nodes being proportional to the amount of time the two children spent face to face. We analysed the two days separately, yielding two networks. The dataset has been collected by the Sociopattern project (http://www.sociopatterns.org/) and analysed in [29].

Co-authorship networks. The networks analysed are the weighted co-authorship networks of the Condensed Matter E-print Archive between 1995 and 1999 (cond-mat) and the High-Energy Theory E-print Archive between 1995 and 1999 (hep-th) [44].

The graph edgelists used in the paper are available online as part of the code package we developed [45].

Finally, for comparison we use Random Geometric Graphs (RGG) [46,47], which are simple models of spatial networks: a RGG is generated by sprinkling N of nodes randomly on a metric space that acts as a substrate (usually a disk of unitary radius or a square with identified edges), and then linking nodes that are closer than a given linking distance d.

The networks analysed in this article are undirected and weighted, because the weighted clique filtration finds a natural application in such case. However, schemes for directed networks can be easily devised and tailored to specific case studies, e.g. one could adopt the definition used in the directed clique percolation method [48] in order to associate network structures to simplices.

Persistent Homology

The method we use to uncover weighted holes is persistent homology of the weight clique rank filtration. In this section we will briefly explain persistent homology and its realization through the weight rank clique filtration.

Persistent homology is a technique from computational algebraic topology that can be viewed as parametrized version of simplicial homology [49]. The two definitions needed for simplicial homology are those of *simplicial complex* and *homology*. A *simplicial complex* is a non empty family X of finite subsets, called faces, of a vertex set with the two constraints:

- a subset of a face in X is a face in X,
- the intersection of any two faces in X is either a face of both or empty.

We assume that the vertex set is finite and totally ordered. A face of $n+1$ vertices is called $n-$face and denoted by $[p_0, \ldots, p_n]$. The interpretation of low dimensional faces is intuitive: a $0-$face is a vertex, a $1-$face is a segment, a $2-$face is a full triangle, a $3-$face is a full tetrahedron. The dimension of a simplicial complex is the highest dimension of the faces in the complex.

Morphism between simplicial complexes are called simplicial maps. A simplicial map is a map between simplicial complexes with the property that the image of a vertex is a vertex and the image of a $n-$face is face of dimension $\leq n$.

Simplicial Homology with coefficients in a field is a functor from the category of simplicial complexes to the category of vector spaces [49]. Homology of dimension n assigns to each simplicial complex X, the vector space $H_n(X)$ of n-cycles modulo boundaries and to every simplicial map $X \xrightarrow{f} Y$ the linear map $H_n(f) : H_n(X) \to H_n(Y)$.

The construction that leads to the vector space H_n is the following. Given a simplicial complex X of dimension d, consider the vector spaces C_n on the set of $n-$faces in X for $0 \leq n \leq d$. Elements in C_n are called $n-$chains. The linear maps sending a $n-$face to the alternate sum of its $(n-1)-$faces

$$\partial_n : C_n \to C_{n-1}$$

Table 2. Summary of spectral quantities values.

Dataset (class)	R_A	$\Delta\lambda_A$	R_L
Genes(I)	1.14	14.6	873
Online forums(I)	0.5	$4\cdot10^5$	$3.4\cdot10^5$
US Air 2000(I)	0.868	$6.9\cdot10^6$	$6.7\cdot10^6$
US Air 2002(I)	0.872	$6.3\cdot10^6$	$2.8\cdot10^7$
US Air 2006 (I)	0.958	$7.7\cdot10^6$	$4.2\cdot10^7$
US Air 20011(I)	0.941	$6.9\cdot10^6$	$8.9\cdot10^7$
Online messages(I)	0.14	$1.1\cdot10^4$	$6.7\cdot10^4$
School day 1 (II)	0.11	$2.5\cdot10^3$	56
School day 2 (II	0.08	$2.3\cdot10^3$	110
C. elegans (II)	0.25	76	$1.8\cdot10^3$
Twitter (II)	0.11	370	$1.5\cdot10^4$
Hep-th (II)	0.11	7.4	$9.6\cdot10^3$
Cond-mat (II)	0.005	0.24	$5.2\cdot10^3$
Lin. RGG	0.0034	34	836
Ran. RGG	0.018	54	255

Summary of spectral quantities. For each dataset, we report the values of R_A, $\Delta\lambda_A$ and R_L. The two classes inline different spectral properties, with particular reference to R_A which is related to the network expansion property.
doi:10.1371/journal.pone.0066506.t002

$$[p_0,\ldots,p_n]\rightarrow\sum_{i=0}^{n}(-1)^i[p_0,\ldots,p_{i-1},p_{i+1},\ldots,p_n].$$

shares the property $\partial_{n-1}\circ\partial_n=0$.

The subspace $ker\partial_n$ of C_n is called the vector space of $n-$cycles and denoted by Z_n. The subspace $Im\partial_{n+1}$ of C_n, is called the vector space of $n-$boundaries and denoted by B_n. Note that from $\partial_{n-1}\circ\partial_n=0$ it follows that $B_n\subseteq Z_n$ for all n.

The $n-$th simplicial homology group of X, with coefficients in k, is the vector space $H_n:=Z_n/B_n$.

Persistent homology is the homology of a *filtration*, i.e. an increasing sequence of simplicial complexes

$$X_0\subset X_1\subset\ldots\subset X_n=X,$$

as opposed to that of a single simplicial complex.

It assigns to a filtration the homology groups of the simplicial complexes $H_n(X_v)$ and the linear maps $i_{v,w}:H_n(X_v)\rightarrow H_n(X_w)$ induced in homology by the inclusions $X_v\mapsto X_w$ for all $v\leq w$. Note that the linear maps $i_{v,v+1}$ are not always injective, meaning that some homological features can disappear along the filtration. These features are encoded by the persistent homology generators: an element $g\in H_n(X_v)$ such that there is no $h\in H_n(X_w)$ for $w<v$ with the property that $i_{w,v-w}h=g$. Two indices completely determine a generator $g\in H_n(X)$, namely its birth, β_g and its death δ_g. The index β_g traces the first index such that g is in the filtration and δ_g is the index of the simplicial complex in which the cycle becomes a boundary (i.e. disappears homologically). The persistence (lifetime) of a generator is measured by $p_g:=\delta_g-\beta_g$. The length of a cycle, that is the number of faces composing it, is denoted by λ_g.

For each homology group, the information about the filtration is collected in a barcode: the set of intervals $[\beta_g;\delta_g]$ for all generators $g\in H_n$, which constitutes a handy complete invariant of H_n [16]. An alternative way to represent the persistent homology of a filtration is through persistence diagrams [16,50], which we use extensively in the SI. A persistence diagram is a set of points in the plane counted with multiplicity. It can be recovered from the barcode considering the points $(\beta_g,\delta_g)\in\mathbb{R}^2$ with multiplicity given by the number of generators with the same persistence interval. In the SI, the reader can find H_1 persistent diagrams of the real world datasets examined for the classification, together with the explicit comparison to the results for their relevant randomized versions.

Filtrations

In classical applications, the filtration is obtained from a point cloud using the Rips-Vietoris complex and persistent homology used to uncover robust topological features of the point cloud. We instead use the clique weight rank filtration to uncover properties deriving from the topology and weighted structure of weighted networks.

Recalling that an $n-$clique is a complete subgraph on $n+1$ vertices, the *clique complex* is a simplicial complex built from the cliques of a graph. Namely there is a $n-$face in the simplicial complex for every $(n+1)-$clique in the graph. The compatibility relations are satisfied because subsets of cliques and intersection of cliques are cliques themselves.

The *Weight Rank Clique filtration* on a weighted network Ω combines the clique complex construction with a thresholding on weights following three main steps.

- Rank the weights of links from ω_{max} to ω_{min}: the discrete parameter ϵ_t indexes the sequence.
- At each step t of the decreasing edge ranking we consider the thresholded graph $G(\omega_{ij},\epsilon_t)$, i.e. the subgraph of Ω with links of weight larger than ϵ_t.
- For each graph $G(\omega_{ij},\epsilon_t)$ we build the clique complex $K(G,\epsilon_t)$.

The clique complexes are nested along the growth of t and determine the weight rank clique filtration. Note that this construction is in fact the clique complex of each element in the graph filtration.

In particular, persistent one dimensional cycles in the weight rank clique filtration represent weighted loops with much weaker internal links.

There is a conceptual difference in interpreting H_1 persistent homology of data with the Rips-Vietoris filtration and H_1 persistent homology of weighted networks with the weight rank clique filtration. While in the first case persistent generators are relevant and considered features of the data, short cycles are more interesting for networks. This is because random networks, or randomisations of real networks, display one dimensional persistent generators at all scales, while short lived generators testify the presence of local organisation properties on different scales.

Computational Complexity

Computing the filtration of a large dataset can be extremely demanding computationally. The identification of the maximal cliques requires in general exponential time, although algorithms exists for special cases that allow solutions to be obtained in polynomial time. In addition, the javaPlex library [51] requires the explicit enumeration of the simplicial facets appearing at each filtration step, which implies the need for large memory resources in order to calculate the persistent homology. However, there are a number of simplifications and improvements to the brute force approach that provide a significant reduction of the problem's complexity. In the metrical case, this is usually done by constructing a smaller complex, the *witness complex* [52],

which approximates with controlled precision [52] the homology of the original data.

In the case of non-metrical discrete spaces, for example networks, one cannot easily construct a witness complex through a controlled sub-sampling of the network. Luckily, it is still possible to reduce the computational complexity in different ways: first, one can limit the analysis to the first s homology groups, which amounts to restricting the clique detection and storage to cliques up to size $s+2$, which reduces the problem to polynomial in time and memory; second, it is possible to parallelize the computation of persistent homology [53]; finally, the more elegant solution is to calculate the homology of an homologically equivalent but much smaller filtration (see the tidy set construction [54]). With respect to the standard clique complex case, the tidy set in particular was shown to reduce the number of simplices along the filtration of various orders of magnitude number of simplices and of one order of magnitude the total memory required. Therefore, a combination of the techniques mentioned above allows to scale up dataset sizes to large-scale networks.

Supporting Information

File S1
(PDF)

Acknowledgments

The authors acknowledge M. Rasetti for stimulating discussions.

Author Contributions

Conceived and designed the experiments: GP FV. Performed the experiments: GP MS ID FV. Analyzed the data: GP MS ID FV. Contributed reagents/materials/analysis tools: GP MS ID FV. Wrote the paper: GP MS ID FV.

References

1. Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45: 167–256.
2. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DH (2006) Complex networks: Structure and dynamics. Phys Rep 424: 175–308.
3. Dorogovtsev SN, Goltsev AV, Mendes JFF (2008) Critical phenomena in complex networks. Rev Mod Phys : 1275–1335.
4. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proc Natl Acad Sci USA 101: 3747–3752.
5. Barabási A (1999) Emergence of scaling in random networks. Science 286: 509–512.
6. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: Simple building blocks of complex networks. Science 298: 824–827.
7. Vázquez A, Dobrin R, Sergi D, Eckmann JP, Oltvai ZN, et al. Proc Nat Acad Sci USA 101: 17940–17945.
8. Mahadevan P, Krioukov D, Fall K, Vahdat A (2006) Systematic topology analysis and generation using degree correlations. ACM SIGCOMM 36: 135–146.
9. Conradi C, Flockerzi D, Raisch J, Stelling J (2007) Subnetwork analysis reveals dynamic features of complex (bio)chemical networks 104: 19175–19180.
10. Eguíluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian AV (2005) Scale-free brain functional networks. Phys Rev Lett 92: 028102.
11. Song WM, Di Matteo T, Aste T (2012) Hierarchical information clustering by means of topologi3 cally embedded graphs. PloS One 7, e31929.
12. Tumminello M, Aste T, Di Matteo T, Mantegna RN (2005) A tool for filtering information in complex systems. Proc Natl Acad Sci USA 102: 10421–10426.
13. Chalupa J, Leath PL, Reich GR (1979) Bootstrap percolation on a bethe lattice. J Phys C.
14. Serrano M, Boguñá M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. Proc Nat Acad Sci USA 106: 6483.
15. Ghrist R (2008) Barcodes: The persistent topology of data. B AM Math Soc 45.
16. Carlsson G, Zomorodian A (2005) Persistent homology - a survey. Discrete Comput Geom 33: 249–274.
17. Carlsson G (2009) Topology and data. B Am Math Soc 46: 255–308.
18. Petri G, Scolamiero M, Donato I, Vaccarino F (2013) Metric and weighted clique persistent ho3 mology for complex networks. In: Proceedings of the European Conference on Complex Systems 2012.
19. Lum P, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, et al. Scientific Reports.
20. Hyekyoung L, Chung M, Hyejin K, Bung-Nyun K, L DS (2011) Discriminative persistent homol4 ogy of brain networks. In: Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on. 841–844. doi:10.1109/ISBI.2011.5872535.
21. Nicolau M, Levine A, Carlsson G (2011) Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proceedings of the National Academy of Sciences 108: 7265–7270.
22. Weygaert R, Vegter G, Edelsbrunner H, Jones B, Pranav P, et al. (2011) Alpha, betti and the megaparsec universe: On the topology of the cosmic web. In: Gavrilova M, Tan C, Mostafavi M, editors, Transactions on Computational Science XIV, Springer Berlin Heidelberg, volume 6970 of Lecture Notes in Computer Science. 60–101. doi:10.1007/978-3-642-25249-5 3. URL http://dx.doi.org/10.1007/978-3-642-25249-5_3.
23. Opsahl T, Colizza V, Panzarasa P, Ramasco JJ (2008) Prominence and control: The weighted rich-club effect. Phys Rev Lett 101: 168702.
24. Pajevic D, Plenz S (2012) The organization of strong links in complex networks. Nat Phys 8: 429–436.
25. Barthelemy M (2011) Spatial networks. Phys Rep 499: 1–101.
26. Antonioni A, Tomassini M (2012) Degree correlations in random geometric graphs. Phys Rev E 86: 037101.
27. Barrat A, Barthelemy M, Vespignani A (2005) The effects of spatial constraints on the evolution of weighted complex networks. J Stat Mech 05: P05003.
28. Palla G, Derényi I, Farkas I, Vicsek T (2005) The effects of spatial constraints on the evolution of weighted complex networks.
29. Stehlé J, Voirin N, Barrat M, Cattuto C, Isella L, et al. (2011) High-resolution measurements of face-to-face contact patterns in a primary school. PLoS One 6: e23176.
30. Gfeller D, De Los Rios P (2007) Spectral coarse graining of complex networks. Phys Rev Lett 99: 38701.
31. Chavez M, Hwang DU, Amann A, Hentschel HGE, Boccaletti S (2005) Synchronization is enhanced in weighted complex networks. Phys Rev Lett 94: 218701.
32. Farkas IJ, Derenyi I, Barabási AL, Vicsek T (2001) Spectra of \real-world" graphs: Beyond the semicircle law. Phys Rev E 64.
33. Jun WU, Barahona M, Yue-Jin T, Hong-Zhong D (2010) Natural connectivity of complex networks. Chin Phys Lett 27: 078902.
34. Barahona M, Pecora LM (2002) Synchronization in small-world systems. Phys Rev Lett 89: 054101.
35. MacArthur BD, Sánchez-Garca RJ (2009) Spectral characteristics of network redundancy. Phys Rev E 80: 026117.
36. Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. Nat Phys 2: 110–115.
37. Boguñá M PF, Krioukov D (2010) Sustaining the internet with hyperbolic mapping. Nat Comms 1: 1–8.
38. Grady D, Thiemann C, Brockmann D (2012) Robust classification of salient links in complex networks. Nat Comm 3: 864.
39. Carlsson G, Zomorodian A (2009) Theory of multidimensional persistence. Discr Comput Geom 42: 71–93.
40. Watts D, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393: 440–442.
41. Opsahl T, Panzarasa P (2009) Clustering in weighted networks. Soc Net 31: 155–163.
42. Opsahl T (2010) Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. Soc Net.
43. Gardner T, di Bernardo D, Lorenz D, Collins J (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. Science 301: 102–105.
44. Newman MEJ (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci USA 98: 404–409.
45. Petri G (2013). Holes - python package for persistent homology calculations. URL http://lordgrilo.github.com/Holes/.
46. Barthélemy M (2011) Spatial networks. Phys Rep 499.
47. Penrose M (2003) Random Geometric Graphs. Oxford, UK: Oxford University Press.
48. Palla Gea (2007) Directed network modules. New J Phys 9.
49. Munkres JR (1984) Elements of Algebraic Topology. 2725 Sand Hill Road Menlo Park, California 94025: Addison-Wesley Publishing Company.
50. Steiner DC, Edelsbrunner H, Harer J (2007) Stability of persistence diagrams. Discrete Comput Geom 37: 103–120.
51. Tausz A, Vejdemo-Johansson M, Adams H (2011). Javaplex: A research software package for persistent (co)homology. "Software available at http://code.google.com/javaplex".
52. de Silva V, Carlsson G (2004) Topological estimation using witness complexes. Symp Point- Based Graphics, ETH Zurich.
53. Bauer U, Kerber M, Reininghaus J (2013) Clear and compress: Computing persistent homology in chunks. arXiv: 13030477.
54. Zomorodian A (2010) The tidy set: a minimal simplicial set for computing homology of clique complexes. Proceedings of the 2010 annual symposium on Computational geometry : 257–266.

topdrim

Deliverable D1.1

Homological backbones of brain functional networks

G. Petri
ISI Foundation, Via Alassio 11/c, 10126 Torino - Italy

P. Expert and F. Turkheimer
Centre for Neuroimaging Sciences, Institute of Psychiatry, De Crespigny Park, Kings College London, London SE5 8AF, UK

R. Carhart-Harris and D. Nutt
Centre for Neuropsychopharmacology, Imperial College London, London W12 0NN, UK

P. J. Hellyer
Computational, Cognitive and Clinical Neuroimaging Laboratory, Division of Brain Sciences, Imperial College London, London W12 0NN, UK

F. Vaccarino
ISI Foundation, Via Alassio 11/c, 10126 Torino - Italy and Dipartimento di Scienze Matematiche, Politecnico di Torino, C.so Duca degli Abruzzi n.24, Torino, 10129, Italy

Networks, as efficient representations of complex systems, have appealed to scientists for a long time and now permeate many areas of science, including neuroimaging [1]. Traditionally, the structure of complex networks has been studied through their statistical properties and metrics concerned with node and link properties, e.g. degree-distribution, node centrality and modularity. Here we study the characteristics of functional brain networks at the mesoscopic level, focusing on their homological properties, more specifically the first homological group, that uncovers cycles. In this context, we introduce the homological backbones, a new set of objects designed to represent compactly the homological features of the links supporting the cycles and make their homological properties amenable to networks theoretical methods. We apply these tools to compare resting-state functional brain activity in fifteen healthy volunteers after intravenous infusion of placebo and psilocybin, the main psychoactive component of magic mushrooms. The results suggest that the brain's network structure undergoes a dramatic change post-psilocybin, revealing many transient structures of low stability but also a small number that have an especially high stability. These cycles appear to traverse different network communities indicating that the brain becomes functionally more intergrated under psilocybin compared to normal resting-state. We interpret these findings in the context of unconstrained, hyperassociative cognition in the psychedelic state.

Keywords: brain functional networks, fMRI, persistent homology, psilocybin

I. INTRODUCTION

The understanding of global brain organisation and its large scale integration remains a challenge for modern neurosciences. Recently, network theory has been increasingly used to analyse neuroimaging data (fMRI, EEG, MEG) in order to address these questions [1, 2], and has even shown potential for clinical application [3, 4].

A natural way of approaching the analysis of large-scale complex systems is to devise a meaningful measure of dynamical similarity between the microscopic constituents and interpret it as the strength of the link between those elements. Commonly used similarity measures are for example (partial) correlations or coherence [5–7]. The resulting matrix can be viewed as a fully connected, weighted and possibly signed adjacency matrix representing the network of interactions between the system's constituents. Unfortunately, the analysis of such weighted networks is not straightforward and renders the use of usual network metrics ill-defined, begetting the need to produce a sparser network representation. Sparsity may be achieved by thresholding the adjacency matrix following a criterion aimed at minimising spurious similarities. However this approach invariably leads to loss of information as thresholds are established either empirically or according to general purpose statistical criteria which may neglect weaker but still functionally meaningful interactions. Importantly, non-dominant interactions can be crucial to understand how a system integrates [8]. In addition to the observations above, a general limitation of standard network approaches stems from their focus either on local or global quantities that forfeit the intermediate, mesoscopic scales.

In this paper, we present an alternative route to analyse weighted networks and more specifically to analytically compare neuroimaging datasets: *weighted graph persistent homology*. The value added of this method over con-

ventional network techniques lie in its capability to describe mesoscopic structures within the system that coexist over different spatial and intensity scales, making it a good candidate for describing brain functional networks [9]. Moreover, this approach uses the whole information, without any ad-hoc filtering.
This is particularly relevant in the case of fMRI data as homological information allows one to uncover functional circuits and their relative importance within the structure of the fMRI correlation matrices.
The method was introduced in [10] and it relies on an extension of the metrical persistent homology theory originally introduced by [11, 12].
Standard simplicial homology details the multidimensional hole structure of a topological space X and is widely used in applications ranging from sensor networks coverage [13] to shape recognition in computer vision [14].
This is because topological information is independent of the metrical properties of the space and thus it provides a robust indicator in noisy or incomplete datasets.
Persistent homology [15] refines the standard homological analysis by focusing on the homology of an inclusion chain of progressively coarser simplicial complexes, a *filtration*, obtained from the data. Informally, simplicial complexes can be thought as a high-dimensional polygonal foam approximating the shape of the original dataset in an appropriate space. Just like in the intuitive case of lower dimensionality, this foam can be studied to obtain information about the dataset's features.
In particular, we focus on the evolution of each homological feature (n-dimensional hole) along the inclusion chain of complexes providing a natural measure of significance for the detected features, which can then be used to discern the meaningful ones from topological noise.
Weighted graph persistent homology takes a step further by adapting the persistent homology machinery to the case of general weighted graphs. The key point here is that this methodology permits to analyse dataset spaces where no metrical structure is given a priori, as in the case of general similarity measures.
We build the inclusion chain through a process called *weight clique filtration* (see sect. II C and II B), which is conceptually akin to a stratigraphy within the link-weight fabric of a network. The filtration swipes across all weight and distance scales within the correlation network identifying strongly correlated units. The circuits among these units constitutes mesoscopic regions of reduced functional connectivity. Weights on edges are ranked in descending order and a series of binary graph snapshot $G_i,\ i \in 0, 1, ...M$ is built by thresholding the network along the weight ranking. Here M is the maximal weight rank. To each graph snapshot G_i, the corresponding clique complex $K(G_i)$ is canonically associated: k-cliques are mapped to $(k-1)$-simplices, inheriting their configuration from the underlying network structure. Since the filtration proceeds over all weights, it preserves all the information contained in the correlation structure of the data. Moreover, it also highlights how the network properties evolve along the filtration providing insights about which and when circuits emerge: each circuit is associated to a corresponding generator of a homology group and the information about its importance is encoded in the form of "time-stamps" recording the birth and death of the generators g_j^k of the homology groups H_k.

We applied this methodology to a prototypical dataset, an fMRI resting-state time-series recorded in subjects after intravenous injection of a placebo or, on a separte ocassion one week later, psilocybin [16]. The order of scans we balanced so that 7 subjects had psilocybin in the first scan and 8 placebo. Psilocybin is a naturally occurring psychedelic compound found in the psilocybe genus of mushrooms, which grow across the world. Persistent homology has been previously applied to neuroimaging data before [17, 18]; however, the focus in these papers was on H_0, the first homological group which contains information about the emergence of connected components.
Here we focus on the second homological group H_1, which characterises the cycles present in the simplicial complexes created along the filtration, i.e. sets (or chains) of regions that are more correlated along the circuit than across.

First, we find that the homological structure of the two conditions are different as seen in the persistence diagrams 1, the psilocybin group having shorter lived cycles compared to the placebo group.

Secondly, to understand better the structure of the cycles that appear along the filtration, and more specifically the roles of the edges forming those cycles, and to make the persistent homology results amenable to classical network tools, we introduce two secondary networks: the persistence $\mathscr{H}^p$ (Eq. 1) and frequency $\mathscr{H}^f$ (Eq. 2) homological backbones. In these networks the edges are weighted, respectively, with the number of cycles an edge belongs to and the total persistence of the cycles an edge belongs to. The homological backbones encode the specific role of the links between the different brain regions, and enable, for example, the discrimination between edges that have a long total persistence but belong to many cycles (hub links) or a few long lived, highly stable cycles (restricted links).

Remarkably, the methodology presented in this paper allows one to transform a signed weighted network, the partial-correlation matrix, into a sparser weighted network while preserving the information contained in the original data.

II. METHODS

A. Dataset

A sample of 15 Healthy control subjects was scanned using fMRI [16]. Each subject underwent an T1-Weighted anatomical scan followed by an eyes-closed resting-state fMRI scan, which lasted 12 min. Participants were scanned on two occasions, 14 days apart. Placebo was given on one occasion and psilocybin on the other in a balanced order. Infusions of psilocybin or placebo began 6 min after the start of the scan. The fMRI data were acquired using a gradient-echo EPI sequence, TR/TE 3000/35 ms, field-of-view = 192 mm, 64 64 acquisition matrix, parallel acceleration factor = 2, 90 flip angle. Fifty-three oblique- axial slices were acquired in an interleaved fashion, each 3 mm thick with zero slice gap ($3 \times 3 \times 3$-mm voxels). A total of 240 volumes were acquired (120 normal resting-state, 120 after injection of placebo/psilocybin). The data were high-pass filtered with a cutoff of 300 seconds. Correction of motion related image distortion and esitmation of a 6-dimension motion model for each functional run were estimated using MCFLIRT from the FMRIB Software Library [19].

For each subject, the T1 image was segmented into n=194 cortical and subcortical regions, including white matter and CSF compartments using Freesurfer (http://surfer.nmr.mgh.harvard.edu/), according to the Destrieux anatomical Atlas [20]. Segmented brain images were registered to the middle volume of the motion corrected functional imaging data using boundary based registration [21] and mean timecourses for voxels within each each region were extracted in naitive image space. The time-series were used to calculate partial correlations between all regions couples (i, j) and motion correction parameters, thus covariating white matter signal, CSF signal, movement and modelling the effect of the placebo/drug injection.

B. Filtrations

In classical applications, the filtration is obtained from a point cloud in a metric space using the Rips-Vietoris complex and persistent homology used to uncover robust topological features of the point cloud.
This is done by considering the neighbourhood $\Gamma_r(x)$ of radius r around each point x in the point cloud. Given a set of points $X = \{x_0, \ldots, x_m\}$, if $\Gamma_r(x_i) \cap \Gamma_r(x_j) \neq$ for all pairs $(x_i, x_j) \in X \times X$, one adds the simplex generated by X to the filtration at the step corresponding to r.
For the general case of similarities without a metric structure, one can instead use the clique weight rank filtration to uncover the properties deriving from the topology and weight structure of networks.
Recalling that an $n-$clique is a complete subgraph on $n+1$ vertices, the *clique complex* is a simplicial complex built from the cliques of a graph. Namely there is a $n-$face in the simplicial complex for every $(n+1)-$clique in the graph. The compatibility relations are satisfied because subsets of cliques and intersection of cliques are cliques themselves.

The *Weight Rank Clique filtration* on a weighted network Ω combines the clique complex construction with a thresholding on weights following three main steps.

- Rank the weights of links from ω_{max} to ω_{min}: the discrete parameter ϵ_t indexes the sequence.
- At each step t of the decreasing edge ranking we consider the thresholded graph $G(\omega_{ij}, \epsilon_t)$, i.e. the subgraph of Ω with links of weight larger than ϵ_t.
- For each graph $G(\omega_{ij}, \epsilon_t)$ we build the clique complex $K(G, \epsilon_t)$.

The clique complexes are nested along the growth of t and determine the weight rank clique filtration. Note that this construction is in fact the clique complex of each element in the graph filtration.
In particular, persistent one dimensional cycles in the weight rank clique filtration represent weighted loops with much weaker internal links.
There is a conceptual difference in interpreting H_1 persistent homology of data with the Rips-Vietoris filtration and H_1 persistent homology of weighted networks with the weight rank clique filtration. While in the first case persistent generators are relevant and considered features of the data, short cycles are more interesting for networks. This is because random networks, or randomisations of real networks, display one dimensional persistent generators at all scales, while short lived generators testify the presence of local organisation properties on different scales.

C. Persistent homology

In this section we give a brief definition of persistent homology .
The two definitions needed for simplicial homology are those of *simplicial complex* and *homology*.

Definition 1. *A simplicial complex is a non empty family X of finite subsets,faces, of a vertex set with the constraints:*

- *a subset of a face in X is a face in X,*

- *the intersection of any two faces in X is either a face of both or empty.*

In the case of real datasets, we can assume without loss of generality that the vertex set is finite and totally ordered. A face of $n+1$ vertices is called $n-$face and denoted by $[p_0, \ldots, p_n]$. Low dimensional faces can easily be imagined: a $0-$face is a point, a $1-$face a line or segment, a $2-$face a full triangle, a $3-$face is a filled tetrahedron. The dimension of a simplicial complex is defined as that of the face with the highest dimension in the complex.

Definition 2. *A simplicial map is a map between simplicial complexes with the property that the image of a vertex is a vertex and the image of a $n-$face is face of dimension $\leq n$.*

Definition 3. *Simplicial Homology with coefficients in a field is a functor from the category of simplicial complexes to the category of vector spaces [15].*

The homology of dimension n assigns to a simplicial complex X, the vector space $H_n(X)$ of n-cycles modulo boundaries and to every simplicial map $X \xrightarrow{f} Y$ the linear map $H_n(f) : H_n(X) \to H_n(Y)$.
The vector space H_n are constructed as follows. Given a simplicial complex X of dimension d, consider the vector spaces C_n on the set of $n-$faces in X for $0 \leq n \leq d$. The elements of C_n are called $n-$chains. Consider also the linear map, *boundary map* ∂_n, which sends a $n-$face to the alternate sum of its $(n-1)-$faces:

$$\begin{aligned} \partial_n : C_n &\longrightarrow C_{n-1} \\ [p_0, \ldots, p_n] &\to \sum_{i=0}^{n} (-1)^i [p_0, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n]. \end{aligned}$$

It is easy to see that ∂_n has the property $\partial_{n-1} \circ \partial_n = 0$.
We refer to the $ker\, \partial_n$ of C_n as to the vector space of $n-$cycles and is denoted by Z_n. The subspace $Im\, \partial_{n+1}$ of C_n is instead the vector space of $n-$boundaries and we denote by B_n. Note that the property $\partial_{n-1} \circ \partial_n = 0$ implies that $B_n \subseteq Z_n$ for all n, which means that the elements of H_n are cycles that are not boundaries, therefore identify $(n+1)$-dimensional holes within the topological space.
The $n-$th simplicial homology group of X, with coefficients in k, is the vector space $H_n := Z_n / B_n$.
Persistent homology is the homology of a *filtration*, i.e. a sequence of inclusions between successive simplicial complexes

$$X_0 \subset X_1 \subset \ldots \subset X_n = X.$$

It equips a filtration with the homology groups $H_n(X_v)$ and the linear maps $i_{v,w} : H_n(X_v) \to H_n(X_w)$ induced by the inclusions $X_v \hookrightarrow X_w$ for all $v \leq w$. A generator of $H_n(X_v)$ is an element $g \in H_n(X_v)$ such that there is no $h \in H_n(X_w)$ for $w < v$ with the property that $i_{w,v-w} h = g$. The linear maps $i_{v,v+1}$ are in general non-injective, which means that the generators can vanish at some point along the filtration. The information about the birth β_g and death δ_g of the homological features along the filtration is completely described by the corresponding generator g of the homology groups. The persistence of g then given is measured by $\pi_g := \delta_g - \beta_g$. One of the standard ways to represent the features of each H_n is through the corresponding persistence diagram [11, 12]: a persistence diagram is a set of points in $\mathbb{R}^2$ counted with multiplicity, where each point represents a generator g and has coordinates (β_g, δ_g). Hence, for each subject in the two groups we have therefore a set of persistence diagrams relative to the persistent homology groups H_n. In this paper, we focused on H_1 and used the H_1 persistence diagrams of each group to construct the corresponding persistence probability densities for H_1 cycles. The two-dimensional distributions of Figure 1 show the probability densities for the two groups.

D. Graph homological backbone

Given the generators g_i^k of the persistent homology group H_k, one can build network representation of the original data exploiting the homological features. Each generator g_i^k is associated to a cycle path on the original network and has a persistence π_{g_i}. We exploit this to define the *persistence homological backbone* $\mathscr{H}_G^p$ of graph G as the network

composed by all the cycle paths weighted with their persistences. If an edge e belongs to multiple cycles $g_0^1, g_1^1, \ldots, g_s^1$, its weight is defined as the sum of the generators' persistences:

$$\omega_e^\pi = \sum_{g_i^1 | e \in g_i^1} \pi_{g_i}. \tag{1}$$

Similarly, we can define a second network, the *frequency homological backbone* $\mathscr{H}_G^f$ of graph G, where the weight of an edge e is the number of different cycles it belongs to:

$$\omega_e^f = \sum_{g_i^1} 1_{e \in g_i^1}. \tag{2}$$

By definition, the two backbones have the same links, although differently weighted.
The construction of these two backbones therefore highlights the role of links which are part of many and/or long persistence cycles, isolating the different roles of edges within the functional connectivity network. For example, the persistence backbone encodes the overall persistence of a link through the filtration process aggregated on all patients, while the second highlights the number of cycles it belongs to. The combined information given by the two backbones then enables to decipher the nature of the homological role of different links. Note also that the construction is able to transform a signed network into a positive-weighted one, maintaining all the homological information. Finally, we remark that the definition of backbones we gave depends on the choice of a specific basis of the homology group. In this case, we decided to choose the basis composed by the algebraic cycles whose underlying graph cycle is maximal. The choice is motivated by our interest mesoscopic structures.

III. RESULTS

Starting from the processed (see section II A) fMRI time-series, the linear correlations between regional time-series were calculated after covarying out the variance due to all other regions and the residual motion variance represented by the 6 rigid motion parameters obtained from the pre-processing, yielding a partial-correlation matrix χ^α for each subject. The matrices χ^α were then analysed through the *weighted clique rank filtration* (see section II C and II B for more details). This procedure finds the generators g_i of the first homological group H_1 at each step of the filtration process. Each generator g_i of the homological group H_1 identifies a lack of mesoscopic connectivity in the form a one dimensional cycle which is associated with its birth and death times along the filtration; hence each g_i is associated with a point in R^2. The set of those points for all subjects in each group can be used to estimate a two dimensional probability density function, the persistence diagram. The persistence diagram then represents the persistence behaviour of H_1.

Grouping together the persistence diagrams of each group, we defined a persistence probability density for the group (Fig. 1) that constitutes the statistical signature of the group's H_1 features. In particular, although the number of cycles in the groups are comparable, the two probability densities strongly differ ($p < 10^{-10}$). The placebo group displays generators appearing and persisting over a limited interval of the filtration. On the contrary, most of the generators for the psilocybin group are situated in a well-defined peak at small birth indices, indicating a general shorter cycle persistence. The psilocybin distribution is also endowed with a "fat tail" on the significance of which we will come back later. The difference in behaviour of the two groups is made explicit when looking at the probability distribution functions of the persistence of the generators (Fig. 2), which are found to be statistically significantly different (KS=0.65, p-val=$5.28 \cdot 10^{-33}$).

In order to better interpret and understand the differences between the two groups, we introduce two secondary networks: the frequency and persistence homological backbones, $\mathscr{H}^f$ and $\mathscr{H}^p$ (see section II D), which inform about the edges supporting the cycles. The weight of the edges in these secondary networks is proportional to the total number of cycles an edge is part of and the total persistence of those cycles respectively. They complement the information given by the persistence density distribution, where the focus is on the entire cycle's behaviour, with information on the single link. In fact, individual edges belonging to many and long persistence cycles represent functionally stable 'hub' links. As with the persistence density distribution, the backbones are obtained at a group level by aggregating the information about all subjects in each group. These networks are slightly sparser than the

original complete χ^{α} networks:

$$\rho(\mathscr{H}^{f,p}_{Pla}) = \frac{2m(\mathscr{H}^{p}_{Pla})}{n(n-1)} = 0.88 \tag{3}$$

$$\rho(\mathscr{H}^{f,p}_{Psi}) = \frac{2m(\mathscr{H}^{p}_{Psi})}{n(n-1)} = 0.89 \tag{4}$$

and they also show very similar densities. The probability distribution functions for the edges' weights are shown in Fig. 3 (a). While edges belong statistically to the same number of cycles, Fig. 3 (a-inset) ($p = 0.8$), the weights distribution of the persistence homological backbone are different, the psilocybin showing a fatter tail ($p < 10^{-39}$), Fig. 3 (a). The difference between the two groups homological backbones is immediately evident when looking at the scatter plot of the edges frequency versus persistence Fig. 3 (b). The placebo group has a linear relationship between the two quantities, while the psilocybin group has a much larger dispersions, showing that, on average, edges in $\mathscr{H}^{f,p}_{Psi}$ are longer-lived but still appear in the same number of cycles.

Gathering together the results from the persistent homology analysis, the insights provided by the homological backbones imply that although the mesoscopic structures, i.e. cycles, in the psilocybin condition are more versatile than in the placebo group (i.e. many of them are unstable), their constituent edges are more stable.

IV. DISCUSSION

Two main results are presented in this paper. Firstly, the stability of mesoscopic association cycles is reduced by the action of psilocybin, as shown by the difference in the probability density function of the generators of the second homology group H_1. Figure 1 shows the comparison of the probability densities obtained for the birth-death pairs of cycles in the two groups. Most of the cycles are concentrated in the bottom-left segment (white area) representing cycles of low stability.

A simple reading of this result would be that the effect of psilocybin is to relax constraints on brain function - ascribing cognition a flexible quality. This could be explained by the reduced activity in subcortical (e.g. the basal ganglia and thalamus) and cortical hubs (e.g. the posterior cingulate cortex, PCC) induced by psilocybin [16] that presumably function to constrain information processing. The thalamus is known to be involved in maintaining cortical oscillatory rhythms [22] and oscillatory synchrony is a classic mechanism for conferring constraint on a system's behaviour. Broadband oscillatory power is decreased after psilocybin and this phenomenon is especially pronounced in the PCC [23]. The PCC is a functional heterogenous structure that contains spatial segments that functionally connect to a range of different brain networks [24, 25]. The PCC also appears to be involved in mediating global brain function [26] as might be expected by its uniquely high metabolism [27] and connectivity [28]. Thus, a principal effect of psilocybin may be to disorganise activity in strategically important regions of the brain such as the PCC and as a corollary, this will render functional important mesoscopic circuits unstable.

The second aspect of our results is the insight that the homological backbones give us, as they allow to go beyond the picture given by persistent homology by gaining access to information about the edges constituting the cycles found along the filtration process. The results of this analysis tell us that although edges are statistically part of the same number of cycles in the two conditions (i.e. there are not more cycles in the psychedelic state), there is a set of edges that is predominant in the psilocybin group compared to the placebo group Fig. 3. That is, these edges or links support cycles that are especially stable. This implies that the brain does not simply become a random system after psilocybin but retains some organisational features. Further work is required to identify the functional significance of these edges but qualitively, they traverse functional network communities, implying that psilocybin increases communication between modules that are usually relatively discrete as shown in 4.

One possible by-product of this greater communication across the whole brain would be the phenomenon of synesthesia which is often reported in conjunction with the psychedelic state. Synesthesia is described as an inducer-concurrent pairing, where the inducer could be a grapheme or a visual stimulus, that generates a secondary sensory output, like a colour for example. Many different types of synesthesia have been reported [29], however genuine synesthesia is very stable in the inducer-concurrent association and often subjects only have one specific type of association. Drug-induced synesthesia instead leads to chain associations, pointing to dynamic causes rather than fixed structural ones as may be the case for acquired synesthesia [30]. Synesthesia is also consistent with the the phenomenology of the sensory deprived state, where there is decreased gain on incoming sensory information and a reciprocal increase in the gain on spontanous activity - including activity in high-level multisensory regions [31]. Broadly consistent with this, it has been reported that subjects under the influence of magic mushrooms, have objectively worse colour perception performance despite subjectively intensified colour experience [32].

To summarise, we present a new method to analyse fully connected, weighted and signed networks and apply it to a unique fMRI dataset of subjects under the influence of magic mushrooms. We find that the psychedelic state is associated to a less constrained and more integrated brain function, which is consistent with reported psychedelic experiences.

Funding statement GP and FV are supported by the TOPDRIM project funded by the Future and Emerging Technologies program of the European Commission under Contract IST-318121. I.D. PE and FT are supported by a PET Methodology Program Grant from the Medical Research Council UK (Ref G1100809/1).

[1] Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neuro.* **10**:186-198 (doi:10.1038/nrn2618).

[2] Meunier D, Lambiotte R, Bullmore E (2010) Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience* **4**:200 (doi: 10.3389/fnins.2010.00200).

[3] Pandit AS, Expert P, Lambiotte R, *et al.* (2013) Traumatic brain injury impairs small-world topology (doi: 10.1212/WNL.0b013e3182929f38 Neurology 10.1212/WNL.0b013e3182929f38). *Neurology* **80**:1826-1833.

[4] Lord LD, Allen P, Expert P, *et al.* (2012) Functional brain networks before the onset of psychosis: A prospective fMRI study with graph theoretical analysis. *NeuroImage:Clinical* **1**:91-98 (doi.org/10.1016/j.nicl.2012.09.008).

[5] Lord LD, Allen P, Expert P, *et al.* (2011) Characterization of the anterior cingulate's role in the at-risk mental state using graph theory. *NeuroImage* **56**:1531-1539 (doi: 10.1016/j.neuroimage.2011.02.012).

[6] Zalesky A, Fornito A, Bullmore E (2012) On the use of correlation as a measure of network connectivity. *NeuroImage* **60**:2096-2106 (doi: 10.1016/j.neuroimage.2012.02.001).

[7] Achard S, Salvador R, Whitcher B, *et al.* (2006) A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs. *Jour. of Neuroscience* **21**:63-72 (doi: 10.1523/JNEUROSCI.3874-05.2006).

[8] Granovetter MS (1973) The strenght of weak ties. *Am. Journ. Soc.* **78**:1360-1380

[9] Simon HA (1962) The architecture of complexity. *Proc. Am. Philo. Soc.* **106**:467-482.

[10] Petri G, Scolamiero M, Donato I, Vaccarino F (2013) Topological strata of weighted complex networks. *PLos One* **8**:e66506 (doi.10.1371/journal.pone.0066506).

[11] Carlsson G & Zomorodian A Computing persistent homology. (2005) *Discrete Comput. Geom.* **33**:249-274 (doi:10.1007/s00454-004-1146-y).

[12] Steiner DC, Edelsbrunner H & Harer J Stability of persistence diagrams (2007) *Discrete Comput. Geom.* **37**:103-120 (doi:10.1007/s00454-006-1276-5).

[13] De Silva V, Ghrist R. (2007) Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology* **7**:339-358

[14] Carlsson G, Ishkhanov T, Silva V, Zomorodian A. On the local behaviour of spaces of natural images. (2008) *Int. J. Comput. Vision* **76**:1-12 (doi:10.1007/s11263-007-0056-x).

[15] Munkres JR (1984) *Elements of Algebraic Topology* Addison-Wesley Publishing Company.

[16] Carhart-Harris RL, Erritzoea D, Williams T, *et al.* (2012) Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proc. Nat. Ac. Sci.* **109**:2138-2143 (doi: 10.1073/pnas.1119598109).

[17] Lee H, Chung MK, Kang H, *et al.* (2011) Discriminative persistent homology of brain networks. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (doi.10.1109/ISBI.2011.5872535).

[18] Lee H, Kang H, Chung MK, Kim BN, Lee DS. (2012) Persistent brain network homology from the perspective of dendrogram. *IEEE Trans Med Imaging.* **31**:2267-77 (doi: 10.1109/TMI.2012.2219590).

[19] Smith SM, Jenkinson M, Woolrich MW, *et al.* (2004) Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**:208-219

[20] Destrieux C, Fischl B, Dale A, & Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **15**:1-15 (doi: 10.1016/j.neuroimage.2010.06.010).

[21] Greve DN & Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* **48**:63-72 (doi:10.1016/j.neuroimage.2009.06.060)

[22] Steriade M. (2006) Grouping of brain rhythms in corticothalamic systems. *Neuroscience* **137**:1087-106.

[23] Carhart-Harris RL, *et al.*, (2013) To appear in *J. Neurosc.*

[24] Leech R, Braga R & Sharp DJ. (2012) Echoes of the brain within the posterior cingulate cortex. *J. Neurosci.* **32**:215-22 (doi: 10.1523/JNEUROSCI.3689-11.2012).

[25] Braga RM, Sharp DJ, Leeson C, Wise RJ, Leech R.J. (2013) Echoes of the Brain within Default Mode, Association, and Heteromodal Cortices *J. Neurosci.* **33**:14031-9 (doi: 10.1523/JNEUROSCI.0570-13.2013).

[26] de Pasquale F., Della Penna S., Snyder AZ., *et al.* (2012) A cortical core for dynamic integration of functional networks in the resting human brain. *Neuron* **74**:753-64 (doi: 10.1016/j.neuron.2012.03.031).

[27] Raichle ME & Snyder AZ (2007) A default mode of brain function: a brief history of an evolving idea. *NeuroImage* **37**:1083-90.

[28] Honey CJ., Sporns O., Cammoun L., *et al.* (2009) Predicting human resting-state functional connectivity from structural connectivity. *Proc. Nat. Ac. Sci.* **106**:2035-2040 (doi: 10.1073/pnas.0811168106).

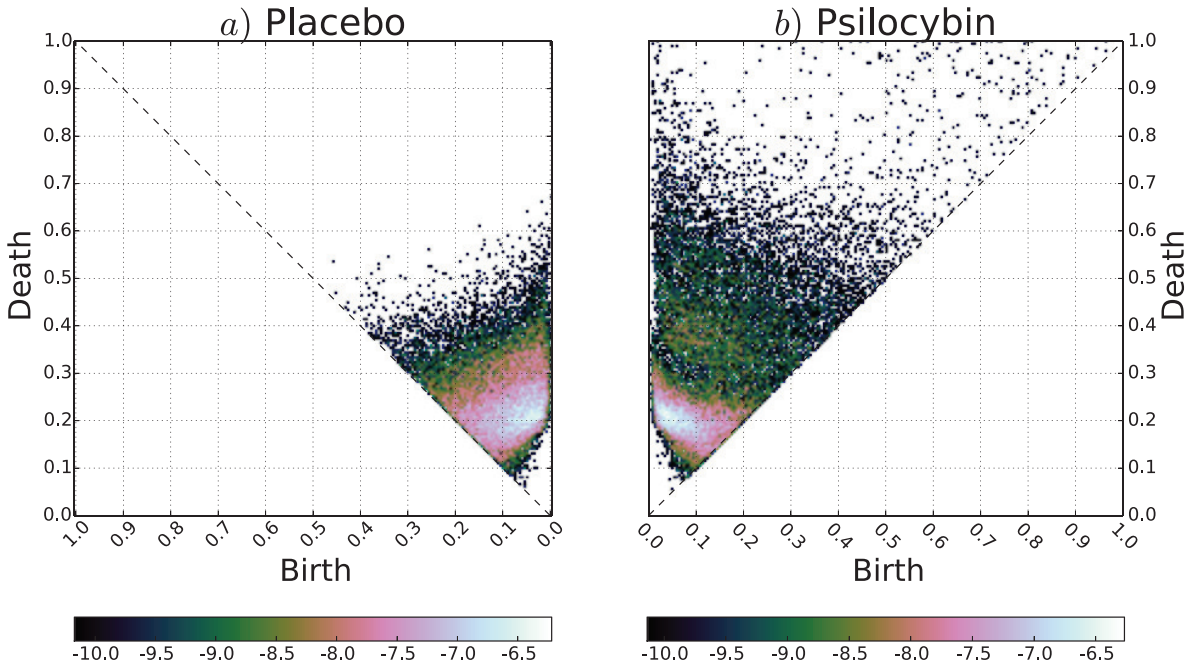


FIG. 1: **Probability densities for the H_1 generators**. Panel *a*) reports the (log-)probability density for the placebo group, while panel *b*) refers to the psilocybin group. The placebo display a uniform broad distribution of values for the births-deaths of H_1 generators, while the pot for the psilocybin is very peaked at small values with a fat tail. These heterogeneity is evident also in the persistence distribution and find explanation in the different functional integration schemes in placebo and drugged brains.

[29] Ward J (2013) Synesthesia. *Annu. Rev. Psychol.* **64**:49-75 (doi: 10.1146/annurev-psych-113011-143840).
[30] Sinke C, Halpern JH, Zedler M, *et al.* (2012) Genuine and drug-induced synesthesia: A comparison. *Consciousness and Cognition* **21**:1419-1434 (doi: 10.1016/j.concog.2012.03.009).
[31] Evart EV (1958) Decreased post-synaptic geniculate transmission after LSD. *Prog. Neurobiol.* **3**:173-194
[32] Hartman & Holister (1963) Effect of mescaline, lysergic acid diethylamide and psylocibin on colour perception *Psychomarmacolgia* **4**:441-451
[33] Newman MEJ. (2006) Modularity and community structure in networks. *Proc. Nat. Ac. Sci.* **103**:8577-8582 (doi: 10.1073/pnas.0601602103).
[34] Blondel VD., Guillaume JL., Lambiotte R. & Lefebvre E (2008) Fast unfolding of communities in large networks. *J. Stat. Mech.: Theory and Experiment* P10008 (doi:10.1088/1742-5468/2008/10/P10008).

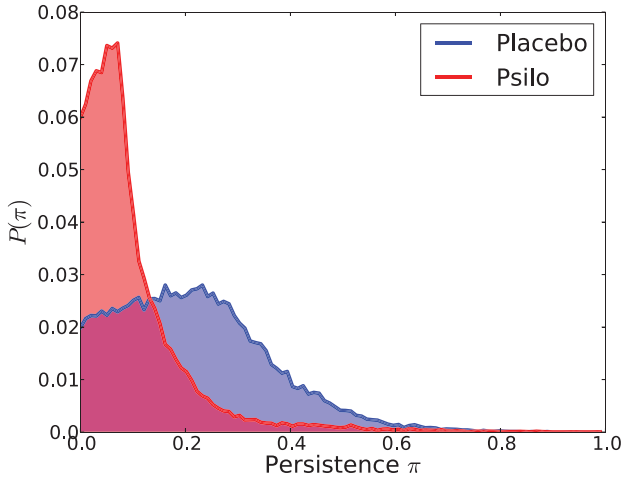
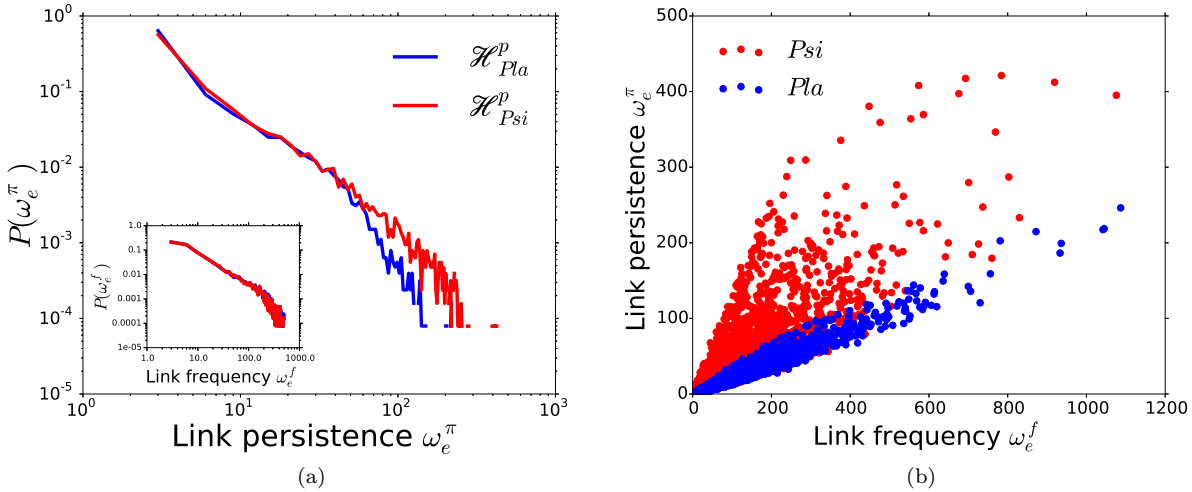
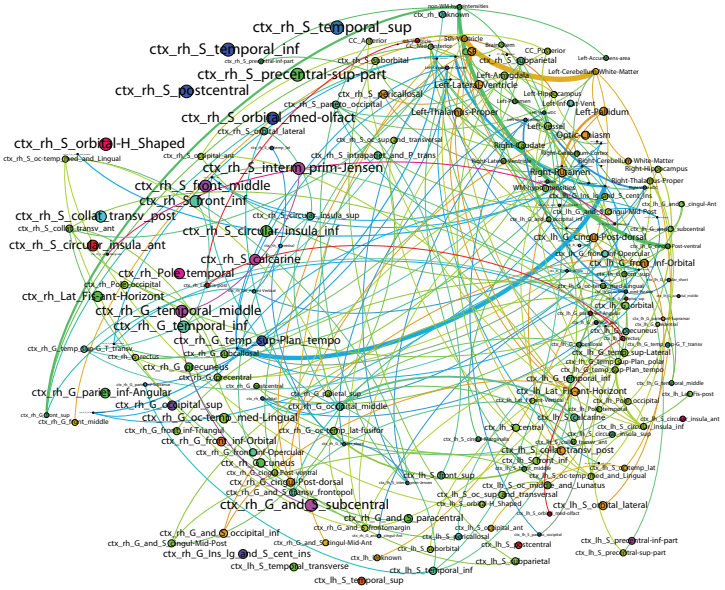


FIG. 2: **Comparison of persistence distributions.** The plot reports the H_1 generators' persistence distributions for the placebo group (blue line) and psilo group (red line). The cycles' persistence distributions are obtained by all the generators identified in a group. The distributions illustrate clearly the difference in the persistence distributions between the two groups.

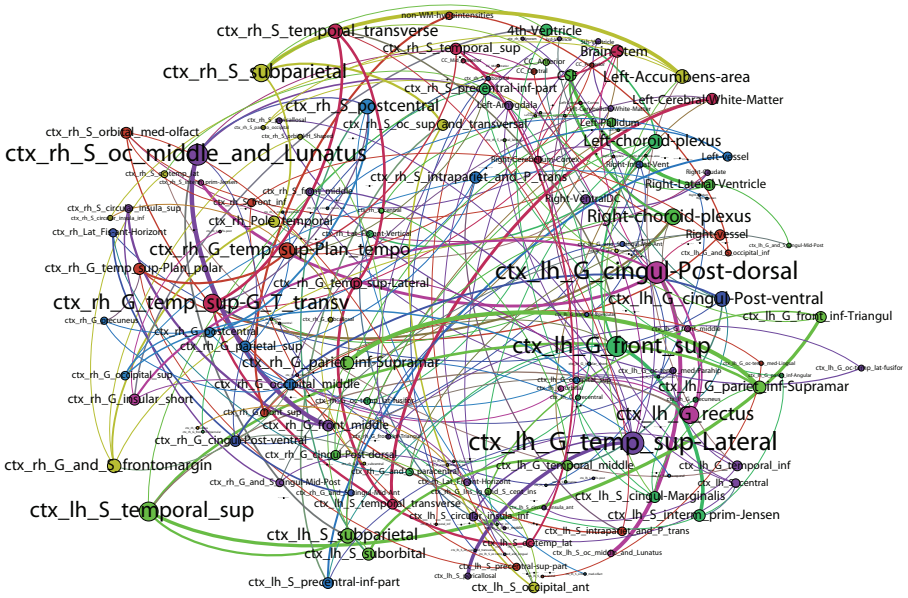


(a) (b)

FIG. 3: **Statistical features of group homological backbones.** Panel (a) reports the probability distributions for the edge weights in the persistence homological backbone (main plot) and the frequency homological backbone (inset). Panel (b) shows the scatter plot of the edge frequency versus total persistence. The placebo backbone display a clean linear relationship. On the contray, the psilocybin backbone is characterised by an heterogeneous relationship between the frequency and total persistence of individual edges, hinting to a different local functional structure within the functional network of the drugged brains.



(a)



(b)

FIG. 4: **Visualisation of the persistence homological backbones.**. The persistence homological backbones $\mathscr{H}^p_{Pla}$ (panel *a*)) and $\mathscr{H}^p_{Psi}$ (panel *b*)) are shown for comparison. In order to highlight the differences in the structure of strong (persistent) edges between the two backbones, only edges in the tail of the weight distributions with edge weight greater than $\omega_{bif} = 20$ are shown. The value of ω_{bif} is chosen as to be slightly smaller than the bifurcation point of the weight distributions in Fig. 3a. Colours represent communities obtained by modularity [33] optimization on the complete backbones using the Louvain method [34]. The width of the links is proportional to their weight in the persistence homological backbone and the size of the nodes are proportionnal to their strength. The proportion of heavy links between communities is much higher in the psilocybin group, showing greater integration.

topdrim

Deliverable D1.1

One graph to rule them all: persistence is graph filtration

Francesco Vaccarino [*] [†], Martina Scolamiero [‡] [†] Giovanni Petri [*] Correspondence and requests for materials should be addressed to G.P. (email: giovanni.petri@isi.it).

[*]DISMA POLITO,[†]ISI Foundation, Via Alassio 11/c, 10126 Torino - Italy, and [‡]KTH - Dept.Math.

Submitted to Proceedings of the National Academy of Sciences of the United States of America

bla bla

Computational topology, complex networks, persistent homology

Introduction

@@@@@@@@@@@@@@@@@@SEZIONE PROVVISORIA MOFOLTOFO PROFOVVIFISOFORIFIAFA

(Multi-)Persistent homology has been inotrduced by GIOVANNI and has become a pervasive tool in computational topology and a well studied topic in pure and applied mathematics, see [**?**] for a recent and well documented survey.

Typically, the starting point is filtration $\mathfrak{F} = X_1 \subset X_2 \subset \cdots \subset X$ of a topological space X i.e. a nested sequence of its subspaces, whose union is the whole X. In this case we say that X is filtered.

Given a filtration $\mathfrak{F} = \{X_t\}_{t\in\mathbb{N}}$ of a topological space and a commutative ring of coefficients k one has a family of homology groups $H_*(X_t, k)$. Since H_* is a functor these groups come together $k-$linear maps $x_{ts} : H_*(X_t, k) \to H_*(X_s, k)$, for all t, s with $t \leq s$. The direct sum $H_*(\mathfrak{F}) := \bigoplus_t H_*(\mathfrak{F})$ is then a $k[x]-$module via $x_{ts} \mapsto x^{s-t}$. We will call $H_i(\mathfrak{F})$ the $i-$th persistent homology group of $\mathfrak{F}$.

A more general approach is obtained by considering filtrations over partially ordered sets (posets) as observed by Carlsson in Bull.AMS.

In the standard case $\mathcal{P} = \mathbb{N}$ and α is the category of topological spaces. Then persistence modules are functorially obtained via the homology functor.

$\mathcal{P}-$persistence objects in $\mathcal{A}$ form a category $\mathcal{A}^{\mathcal{P}}$ and are sometimes called diagram in categories and there are several results on this topic due, as an example, by ALLIEVO KASHDAN, CARLSSON BUBENIK. One can consider the incidence algebra $\mathfrak{P}$ of $\mathcal{P}$ and it is easy to show that the category $\mathcal{A}^{\mathcal{P}}$ is equivalent to the category of $\mathfrak{P}-$modules (see Kashdan). Here we do not have a nice decomposition theorem like in the standard persistence case, but some results can be found in (Carlsson, Nostri...). Nevertheless this point of view allow us to introduce a way of defining persistence, that, in a sense which will be made clear in this paper, is the more general we are permitted to use.

Usually, and especially in application, the space X is finite and built out of some dataset, the *point cloud*.

MAIN

Before going through homology and filtrations let us start with X be a finite topological space, which we can imagine to be given as some sampling taken from a given data set. If the space X is not T_0, then we can always take the Kolmogorov quotient of X, thus making it a T_0-space which is homotopically equivalent to the starting space. Finite T_0-spaces are in bijective correspondence with finite partially ordered sets, briefly posets, via $x \leq y$ if and only if $U_x = cap_{U \text{ closed }: x\in U} \subseteq U_y$. To every finite poset $(X, \leq)$ one can functorially associate a simplicial complex $O(X, \leq)$, the order complex, in the following way: the underlying set is X and $\{x_0, \ldots, x_n\}$ is a face if and only if $x_0 < x_1 < \cdots < x_n$ in $(X, \leq)$. By a result of McCord [**?**], there is a weak homotopy equivalence between a finite T_0-space X and $O(X, \leq)$.

Given a (simple) graph $G = (V, E)$ there is a simplicial complex $\mathrm{Cl}(G)$ associated to it: the clique (or flag) complex. The points $\mathrm{Cl}(G)$ are the vertices of G and $\{v_0, \ldots, v_n\}$ is a face if and only if $(v_i, v_j) \in E$, for all i, j. It is well know that any finite simplicial complex is homeomorphic to the clique (or flag) complex of a graph, namely to the clique complex of the $1-$skeleton of its barycentric subdivision. Furthermore the order complex is the clique complex of its $1-$skeleton.

Therefore, to study the homology of finite topological spaces it is equivalent to study the homology of clique complexes of finite graphs.

If the finite space X is T_1 then it has the discrete topology and the process just described is not very informative because in this case the poset $(X, \leq)$ is discrete and the associated order complex is has vertices the elements of X and it is zero dimensional.

Cech complex

Definition 1. *A graph is is a pair $G = (V, E)$, where V is a finite set whose elements are called vertices and $E \subset \binom{V}{2}$, the family of subsets of V of cardinality two. A morphism of graphs $\varphi : G \to G'$ is given by map $\varphi : V \to V'$ such that $(\varphi(x), \varphi(y)) \in E'$ for all $(x, y) \in E$. The category of graphs will be denoted by $\mathcal{G}$.*

Definition 2. *A simplicial complex is a non empty family X of finite subsets, called faces, of a vertex set with the two constraints:*

- *a subset of a face in X is a face in X,*
- *the intersection of any two faces in X is a face of both.*

We assume that the vertex set is finite and totally ordered. A face of $n+1$ vertices is called $n-$face and denoted by $[p_0, \ldots, p_n]$. A $0-$face is a vertex, a $1-$face is a segment, a $2-$face is a full triangle, a $3-$face is a full tetrahedron. The dimension of a simplicial complex is the highest dimension of the faces in the complex. A morphism of simplicial complex is called simplicial map and is given by a map on vertices such that the image of a face is again a face. We denote by $\mathcal{S}$ the category of (finite) simplicial complex.

Reserved for Publication Footnotes

Given a graph $G = (V, E)$ the clique complex $\mathrm{Cl}(G)$ is a simplicial complex having the same vertices of G and such that $[p_0, \dots, p_n] \in \mathrm{Cl}(G)$ if and only if $(p_i, p_j) \in E$ for all i, j. There is a $n-$face in $\mathrm{Cl}(G)$ for every $(n+1)-$clique in the graph, i.e. a complete subgraph on $n+1$ vertices. It is easy to see that the assignment $G \to \mathrm{Cl}(G)$ gives a covariant functor $\mathrm{Cl} : \mathcal{G} \to \mathcal{S}$.

Fixed a field k, in the following by vector space we intend a $k-$vector space. Given a simplicial complex X of dimension d, consider the vector spaces C_n on the set of $n-$faces in X for $0 \leq n \leq d$. Elements in C_n are called $n-$chains. The linear maps sending a $n-$face to the alternate sum of it's $(n-1)-$faces.

$$\begin{aligned} \partial_n : C_n &\longrightarrow C_{n-1} \\ [p_0, \dots, p_n] &\to \sum_{i=0}^{n} (-1)^i [p_0, \dots, p_{i-1}, p_{i+1}, \dots, p_n]. \end{aligned}$$

shares the property $\partial_{n-1} \circ \partial_n = 0$.

The subspace $ker\, \partial_n$ of C_n is called the vector space of $n-$cycles and denoted by Z_n. The subspace $Im\, \partial_{n+1}$ of C_n, is called the vector space of $n-$boundaries and denoted by B_n. Note that from $\partial_{n-1} \circ \partial_n = 0$ it follows that $B_n \subseteq Z_n$ for all n.

The $n-$th simplicial homology group of X, with coefficients in k, is the vector space $H_n := Z_n / B_n$. The rank of H_n is called the n-th Betti number of X.

The first Betti numbers of X have an easy intuitive meaning: the 0-th Betti number is the number of connected components of X, the first Betti number is the number of two dimensional (poligonal) holes, the third Betti number is the number of three dimensional holes (convex polyhedron).

It is fundamental to note that homology is a functor, this implies the following proposition.

Let X and Y be two simplicial complexes, a simplicial map $f : X \to Y$ determines a linear map between the homology groups $H_i(f) : H_i(X) \to H_i(Y)$ for all i.

Definition 3. *Let $G \in \mathcal{G}$ we*

$\mathcal{P}$-persistence.

Definition 4. *Let $\mathcal{P}$ be a poset and let $\mathcal{A}$ an arbitrary category. We regard $\mathcal{P}$ as a category in they usual way. A $\mathcal{P}-$persistence object in $\mathbf{A}$ is a functor $\varphi : \mathcal{P} \to \mathcal{A}$. As usual we denote the category of these functors and their natural transformation by $\mathcal{A}^{\mathcal{P}}$.*

Definition 5. *Let $\mathcal{P}$ be a poset and let $\mathcal{A}$ an arbitrary category. We regard $\mathcal{P}$ as a category in they usual way. A $\mathcal{P}-$persistence object in $\mathbf{A}$ is a functor $\varphi : \mathcal{P} \to \mathcal{A}$. As usual we denote the category of these functors and their natural transformation by $\mathcal{A}^{\mathcal{P}}$.*

The starting point in persistent homology is a filtration. As in [8], we call a simplicial complex X filtered if we are given a family of subspaces $\{X_v\}$ parametrized by $\mathbb{N}$, such that $X_v \subseteq X_w$ whenever $v \leq w$. The family $\{X_v\}$ is called a *filtration*. There are many ways to construct a filtration from a point cloud or a network, some relevant ones are explained in section II.

The *persistent homology module* of a filtration is given by the homology groups of the simplicial complexes $H_n(X_v)$ and the linear maps $i_{v,w} : H_n(X_v) \to H_n(X_w)$ induced in homology by the inclusions $X_v \hookrightarrow X_w$ for all $v \leq w$.

Following [8], this system is called a module because the vector space $H_n = \oplus_v H_n(X_v)$ can actually be endowed with a $k[t]-$module structure, defining $t \cdot m := i_{v,v+1}(m)$ for $m \in H_n(X_v)$. Note that the linear maps $i_{v,v+1}$ are not always injective. A persistent homology generator is a generator of H_n according to the $k[t]-$structure, i.e an element $g \in H_n(X_v)$ such that there is no $h \in H_n(X_w)$ for $w < v$ with the property that $t^{v-w} h = g$. By the structure theorem on modules over PID, $k[t]-$modules are completely determined by the degree of each generator g (birth of the generator β_g) and the degree in which the generator is annihilated by the module action (death of the generator δ_g). The persistence (lifetime) of a generator is measured by $p_g := \delta_g - \beta_g$. The length of a cycle, number of faces composing it, is denoted by λ_g.

The barcode of a filtration is the set of intervals $[\beta_g; \delta_g]$ for all generators $g \in H_n$, this is a handy complete invariant of H_n, [8]. By persistent topological features we intend generators of H_n such that the interval $[\beta_g; \delta_g]$ is large with respect to the filtration length.

An alternative way to represent persistent homology modules is the persistence diagram [**?**], [**?**]. A persistence diagram is a set of points in the plane counted with multiplicity, it can be recovered from the barcode considering the points $(\beta_g, \delta_g) \in \mathbb{R}^2$ with multiplicity given by the number of generators with the same persistence interval.

Filtrations. In this section we will go through some basic constructions that generate a filtration starting from a point cloud or a complex network.

The most popular filtration for data analysis is the *Rips-Vietoris filtration* [8]. The Rips-Vietoris complex is a simplicial complex associated to a set of points in a metric space in the following way: every point p is the center of a radius ϵ ball $D(p, \epsilon)$ and $n+1$ points $\{p_0, \dots, p_n\}$ determine a $n-$face in the Rips-Vietoris complex if the corresponding radius ϵ balls intersect two by two, i.e $D(p_i, \epsilon) \cap D(p_j, \epsilon) \neq \emptyset$ for all $i \neq j \in \{0 \dots n\}$. Clearly the Rips-Vietoris complex depends on the parameter ϵ and if $\epsilon_1 < \epsilon_2$ the complex with ϵ_1 radius balls is contained in the complex with ϵ_2 radius balls. To the growth of ϵ we obtain an increasing sequence of simplicial complexes, a filtration, the Rips-Vietoris filtration. In this context persistent topological features of the filtration are considered as features of the point cloud.

For unweighted networks, the *Clique filtration* is used in [19] to analyse the difference between the barcodes of random networks, networks with exponential connectivity distribution and scale-free networks. The $k-$skeleton X_k of a simplicial complex X is the subcomplex of X containing all the faces of dimension smaller or equal to k. Consider a complex network and the corresponding clique complex X, the clique filtration is obtained by filtering the clique complex according to the dimension of the skeleton:

$$X_0 \subseteq X_1 \subseteq X_2 \subseteq \dots \subseteq X.$$

Note that persistent features of the Clique filtration are generators of the homology groups of the clique complex. These generators can be directly calculated from the clique complex of the graph, thus the filtration gives no extra information. This is not the case for the following filtration we have introduced for weighted networks in which persistent features cannot be determined from a single simplicial complex in the family but instead reveal the intricate multiscale relation between weights and links in a weighted indirect network.

The *Weight Rank Clique filtration* on a weighted network Ω combines the clique complex construction with a thresholding on weights. The fist step is to rank the weights of links from ω_{max} to ω_{min}: the discrete parameter ϵ_t scans the

sequence. At each step t of the decreasing edge ranking we consider the thresholded graph $G(\omega_{ij}, \epsilon_t)$, i.e. the subgraph of Ω with links of weight larger than ϵ_t. For each graph $G(\omega_{ij}, \epsilon_t)$ we build the clique complex $K(G, \epsilon_t)$. The clique complexes are nested to the growth of t and determine the weight rank clique filtration. Persistent one dimensional cycles represent weighted loops with much weaker internal links.

Results

1. Newman, M. E. J. The Structure and Function of Complex Networks SIAM Rev., 45(2), 167256. (2003).
2. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. H., Complex networks: Structure and dynamics. Phys. Rep. 424,(4-5), 175-308 (2006).
3. Dorogovtsev, S. N., Goltsev, A. V.,& Mendes, J. F. F. Critical phenomena in complex networks. Rev. Mod. Phys., 80(4), 1275–1335. (2008)
4. Milo,R. , Shen-Orr,S. , Itzkovitz,S. Kashtan, N., Chklovskii,D. & Alon, U. Network Motifs: Simple Building Blocks of Complex Networks Science 298 (5594): 824-827 (2002)
5. Vazquez, A., et al. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. Proc. Nat. Acad. Sci. USA, 101(52), 17940–17945, (2004)
6. Mahadevan, P., Krioukov, D., Fall, K., & Vahdat, A. Systematic topology analysis and generation using degree correlations. ACM SIGCOMM, 36(4), 135146, (2006);
7. Conradi, C., Flockerzi, D., Raisch, J., & Stelling, J. Subnetwork analysis reveals dynamic features of complex (bio)chemical networks. Proc. Natl. Acad. of Sci. USA, 104(49), 19175–19180, (2007).
8. Carlsson,G. & Zomorodian,A., Computing Persistent Homology. Discrete Comput. Geom. 33(2), 249-274 (2005).
9. Carlsson, G., Topology and Data. Bulletin of the American Mathematical Society. 46(2), 255-308 (2009).
10. Barrat, A., Barthlemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. Proc. Natl. Acad. Sci. USA 101, 3747–3752, (2004).
11. Barabási, A. Emergence of Scaling in Random Networks. Science 286,509–512 (1999).
12. Eguiluz V.M., Chialvo D.R., Cecchi G.A., Baliki M. & Apkarian A.V. Scale-free brain functional networks. Phys. Rev. Lett. 92:028102, (2005)
13. Song, W. M., Di Matteo, T., & Aste, T. Hierarchical information clustering by means of topologically embedded graphs. PloS One, 7(3), e31929, (2012)
14. Chalupa J, Leath PL, Reich GR Bootstrap percolation on a Bethe lattice. J. Phys. C 12:L31, (1979)
15. Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. A tool for filtering information in complex systems. Proc. Natl. Acad. Sci.USA, 102(30), 10421, (2005).
16. Serrano, M., Boguñá, M., & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. Proc. Nat. Acad. Sci. USA, 106(16), 6483, (2009).
17. Ghrist, R. Barcodes: The persistent topology of data. B. AM. Math. Soc. 45, 61 (2008).
18. Evans, T. S. Clique graphs and overlapping communities. J. Stat. Mech. 2010, P12037 (2010).
19. Horak, D., Maletić, S. & Rajković, M. Persistent homology of complex networks. J. Stat. Mech. 2009, P03034 (2009).
20. Opsahl, T., Colizza, V., Panzarasa, P., Ramasco, J. J., Prominence and control: The weighted rich-club effect. Phys. Rev. Lett., 101 (168702), (2008)
21. Pajevic, S. & Plenz, D. The organization of strong links in complex networks. Nat. Phys. 8, 429–436 (2012).
22. Barthelemy, M., Spatial Networks Phys. Rep. 499:1–101 (2011)
23. Antonioni, A., & Tomassini, M. Degree correlations in random geometric graphs. Phys. Rev. E, 86(3), 037101 (2012);
24. Barrat, A., Barthelemy, M., & Vespignani, A. The effects of spatial constraints on the evolution of weighted complex networks. J. Stat. Mech., 2005(05), P05003.
25. Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814–818 (2005).
26. Colizza, V., Flammini, A., Serrano, M. A., & Vespignani, A. Detecting rich-club ordering in complex networks Nat. Phys. 2(2), 110–115. (2006)
27. Boguñá, M., Papadopoulos, F., & Krioukov, D. Sustaining the Internet with hyperbolic mapping. Nat. Comms., 1(6), 1–8 (2010)
28. Grady, D., Thiemann, C., & Brockmann, D. Robust classification of salient links in complex networks. Nat. Comm., 3, 864, (2012)
29. Carlsson,G., Zomorodian,A. The Theory of Multidimensional Persistence. Discr. Comput. Geom., 42(1), 71-93 (2009)

ACKNOWLEDGMENTS. G.P. is supported by the TOPDRIM project funded by the Future and Emerging Technologies program of the European Commission under Contract IST-318121. I.D. and M.S. are partly supported by Project Lagrange Ph.D. Grant. F.V. is partially supported by PRIN 2009 "Spazi di Moduli e Teoria di Lie".