

Chapter 11

Characterisation of the Idiotypic Immune Network Through Persistent Entropy

Matteo Rucco, Filippo Castiglione, Emanuela Merelli
and Marco Pettini

Abstract In the present work we intend to investigate how to detect the behaviour of the immune system reaction to an external stimulus in terms of phase transitions. The immune model considered follows Jerne's idiotypic network theory. We considered two graph complexity measures—the *connectivity entropy* and the *approximate von Neumann entropy*—and one entropy for topological spaces, the so-called *persistent entropy*. The simplicial complex is obtained enriching the graph structure of the weighted idiotypic network, and it is formally analyzed by persistent homology and persistent entropy. We obtained numerical evidences that *approximate von Neumann entropy* and *persistent entropy* detect the activation of the immune system. In addition, persistent entropy allows also to identify the antibodies involved in the immune memory.

11.1 Introduction

Complex systems are system typically characterised by a number of not identical agents whose aggregate activity is nonlinear and often exhibits hierarchical self-organisation under selective pressures. Although complex systems science is a relatively young research area, it attracts lots of interest from researchers mainly due to the emerging of new techniques in several fields, e.g., physics, mathematics, data analysis and computer sciences [12, 22]. Classical data analysis (both descriptive

M. Rucco (✉) · E. Merelli
School of Science and Technology, University of Camerino, Camerino, Italy
e-mail: matteo.rucco@unicam.it

E. Merelli
e-mail: emanuela.merelli@unicam.it

F. Castiglione
Institute for Applied Mathematics (IAC) CNR, Rome, Italy
e-mail: f.castiglione@iac.cnr.it

M. Pettini
Centre de Physique Théorique, Aix-Marseille University, Marseille, France
e-mail: pettini@cpt.univ-mrs.fr

and exploratory) can not be sufficient for analyzing the huge amount of data that usually characterize a complex system. Persistent homology, a branch of computational topology, is nowadays largely applied for the study of complex systems [5]. Ibekwe et al., used Topological Data Analysis (TDA) for reconstructing the relationship structure of *E. coli* O157, they also prove that the non-O157 is in 32 soils (16 organic and 16 conventionally managed soils) [11]. De Silva used TDA for the analysis of sensor network [4]. TDA has been successfully applied for the study of viral evolution in biological complex systems [2]. Rucco et al., used a set of topological data analysis techniques for improving the pulmonary embolism detection in [21]. Petri et al., used an homological approach for studying the characteristics of functional brain networks at the mesoscopic level [18].

In this paper we intend to study the behaviour of a biological complex system: the idiotypic network of the mammal immune system from an information-theoretic viewpoint. In order to accomplish our aim we used both a classical approach graph-based and an innovative approach topology-based. The rest of this paper is organised as follows. Section 11.2 reports the explanation the theory of our case study and the theoretical background related to our work. In Sect. 11.3 we summarised the analysis of our *in silico* experiment. Section 11.4 provides concluding remarks.

11.2 Background

11.2.1 Case Study: The Immune Network Theory

In 1974 Niels Jerne suggested the *idiotypic network theory* to explain the phenomena of antigenic recognition by the immune cells in terms of a network of interacting cells and antibodies. Jerne's model introduced several features of immune system (I.S.), briefly when the antigen is presented to the organism, the I.S. reacts following two possible pathway: suppression or immunity. The class Ab1 of antibodies, elicited directly by the antigen, elicits the production of new antibodies Ab2 which in turn induces Ab3 and so on. This phenomenon is known as the *idiotypic cascade*. During the onto-genesis the I.S. learns which antibodies should be produced and the system remembers this decisions for a long time. This phenomena, called *immunological memory* is a property of the network of cells as a whole, rather than of the individual cells [9].

11.2.2 Graph Entropy Measures

Given a dynamical system and its graph representation there are several measures for its characterization as suggested by Reidys [15]. Even if one needs a global measure for culling the dramatic leap in the dynamics of the system, some classical measures

are not sufficient for detecting whenever the system reacts to a stimulus: e.g., the *density* of a graph and defined as the ratio between the present links and the number of all possible links in the graph can not significantly change during the stimulus. For this reason we argue that *entropies* are more meaningful. In this study we report on the application of *connectivity entropy* and *approximate von Neumann entropy* [16, 17]. Both measures can be interpreted as a complexity measure of a graph, in fact both are depending by the number and the degree of the vertices.

Connectivity entropy has been used by Ortiz et al., for analyzing the structural properties of a social network and then for identifying the set of key players in the network. We repeated the experiment using the idiotypic network instead of the social network [16].

Approximate von Neumann entropy has been used by Han and collaborators on a graph classification and characterization tasks. In our approach we used this entropy measures for distinguish graphs corresponding to the same system but in different conditions, namely before, during and after a stimulus [8].

Consider now the following definitions.

11.2.2.1 Connectivity Entropy

Let $G = (V, E)$ be a graph, with $V = \{v_1, v_2, \dots, v_n\}$ the set of vertices and E the edges. We can define [16]:

Connectivity of a node $v_i \in V$ in a graph such as:

$$\chi(v_i) = \frac{\text{deg}(v_i)}{2N}, N > 0 \quad (11.1)$$

where $\text{deg}(v_i)$ is the number of incident edges to vertex v_i and N the total number of edges in the graph. Because, $0 \leq \chi(v_i) \leq 1$, and $\sum_{i=1}^N \chi(v_i) = 1$, $\chi(v_i)$ is known as *connectivity probability distribution* of the graph.

The Connectivity entropy H_{co} of a graph G is:

$$H_{co}(G) = -\sum \chi(v_i) \log_2 \chi(v_i) \quad (11.2)$$

11.2.2.2 Approximated von Neumann Entropy

Let $G = (V, E)$ be a graph, with $V = \{v_1, v_2, \dots, v_n\}$ the set of vertices and E the edges. Let $\text{deg}(u)$ the degree for the vertex u and defined as the sum of the weight of its incident edges. The *approximate von Neumann entropy* is defined as [8]:

$$H_T^U = \ln|V| - \frac{1}{2|V|} \sum_{(u,v) \in E} \frac{1}{\text{deg}(u)\text{deg}(v)} \quad (11.3)$$

11.2.3 Topological Data Analysis

Topological data analysis (TDA) is based on the concept of computational homology, a tool that transforms *local data* into *global algebraic structure*. Roughly speaking, the idea is to infer a “shape of the data”, building the so-called *simplicial complexes*. A simplicial complex is a nested collection of simplices (vertices, line segment, triangle, etc. ...). Then, homology associates to the simplicial complex a sequence of abelian groups $H_k(\mathbb{X})$, $k \in \mathbb{X}$. The vector containing the ranks of each group is a topological invariant, the so-called **Betti numbers**. Betti numbers are the numbers of *holes* in a space with different dimensions. *Persistent homology* is one of the most used techniques for computing the topological invariants of a topological space. It returns a parametrized version of the Betti numbers: the Betti barcodes (see for example Fig. 11.1) [5]. The barcodes are equipped with the *generators* of the topological feature (connected components, holes, voids, etc.). Generators are the set of nested simplices forming the topological features.

11.2.4 Simplicial Complexes from Graph

Given a graph directed or undirected, it is possible to construct a simplicial complex from it following several approaches [10, 13]. In this paper we apply the *clique weight rank persistent homology* (CWRPH) [19]. This innovative technique is based on the concept of a flag complex: given a graph G , the simplices of a clique complex $C(G)$ are the complete subgraphs of G and the 0-simplices of $C(G)$ are the vertices of G , (i.e., the complete subgraph complex). The maximal simplices are given by the collection of vertices that make up the cliques of G . In the literature, a clique complex is also referred to as a flag complex. CWRPH describes a formal procedure for building a $C(G)$ but taking into account the weights of the links of G .

11.2.4.1 Improved jHoles

jHoles is the first Java high-performance implementation of the CWRPH algorithm [1]. The first release of *jHoles* implemented the following standard clique weight rank descending persistent homology:

1. Extract the descending (ascending) list W of all weights w_t indexed by the discrete parameter t ;
2. List all maximal cliques of each connected component in G with the Bron-Kerbosch algorithm [23];

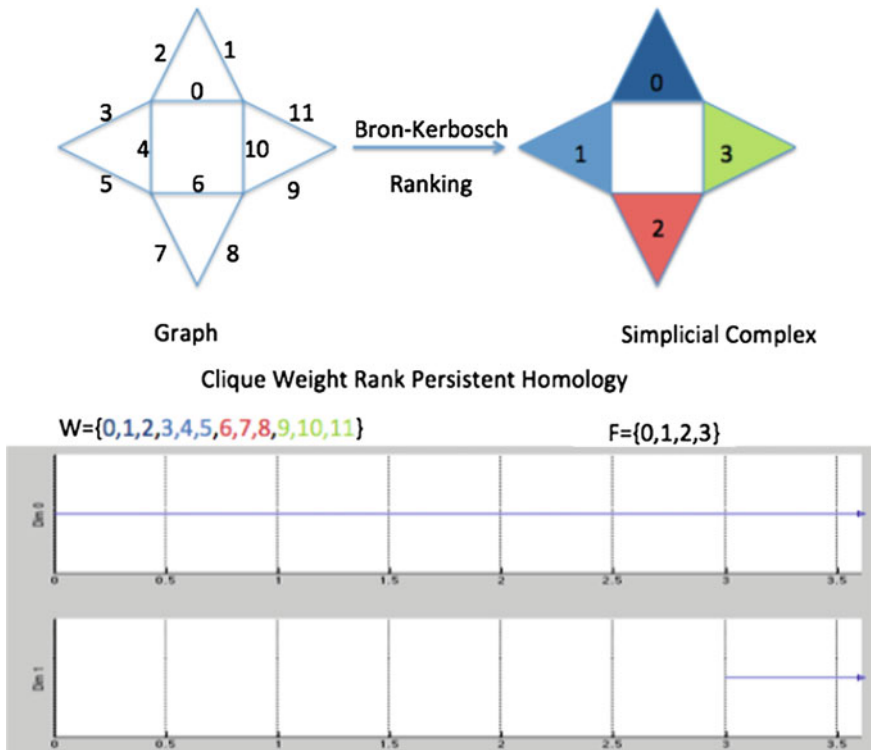


Fig. 11.1 jHoles application example. Construction of a topological space from an undirected weighted graph G (up). The Betti barcode representing the evolution of topological invariants (down). Where W is the set of weights for the graph G , and F denotes the set of filter values used during the computation of persistent homology for the simplicial complex. The resulting simplicial complex is characterized by the tuple $\beta_0 = 1$, and $\beta_1 = 1$. $\beta_0 = 1$ indicates that there is only 1 connected components formed by the four 2-simplices (filled triangles), while $\beta_1 = 1$ indicates that the simplices are arranged in a circular motif that corresponds to a persistent homological loop of dimension 1

3. Find all the combinations of each clique: a n – clique, with $n \geq 3$ must be tessellated with a set of 3 – cliques, because a simplex is just the generalization of a n -dimensional triangle;
4. For each combination and clique, rank it according to the index t of the minimum (maximum) weight of the face;
5. The resulting structure is a clique simplicial complex over which persistent homology can be computed; output barcodes, intervals and generators.

Here we propose an *improved* version of *jHoles* that implements the following algorithm:

1. Extract the descending (ascending) list W of all weights w_t , indexed by the discrete parameter t ;

2. (Parallel) List all maximal cliques of each connected component in G with the Bron-Kerbosch algorithm [23];
3. (Parallel) Find recursively all permutations of each clique (clique tessellation with a set of 3-cliques);
4. (Parallel) For each permutation and clique, rank it according to the index t of the minimum (maximum) weight of the face;
5. The resulting structure is a clique simplicial complex over which persistent homology can be computed; output barcodes, intervals and generators.

Each permutation is a simplex belonging to the complex, while each maximal clique is a largest simplex. Steps from 2 to 4 are tasks that can be executed in parallel respectively on each connected component or face and are the core of the filtration. Step 4 ranks each face according to the index of the minimum (maximum) weight of its edges for the standard descending (ascending) filtration. The use of permutations instead of combinations in step 3 significantly improves algorithm performances and memory usage (the number of permutations of a set is strictly smaller than the number of its combinations).

Roughly speaking, the persistent homology algorithm, spans over the set F of filter values and at each iteration it introduces the simplex ranked with the actual filter value and then computes the homology of the new topological space (see Fig. 11.1) [6].

11.2.5 Persistent Entropy

Diaz et al., defined an entropy based on the persistent barcode (Definition 3 of [3]). The aim of their paper is an algorithm entropy-driven for finding the best filtration of a set of simplices. We argue that when the filtration is given their entropy can be easily extended without loosing the interpretation à la Shannon. Here we propose to use the maximum of the filtration value of a persistent barcode plus one as upper bound, let call this quantity m .

Definition 1 (*Persistent entropy*) Given a filtered topological space equipped with an ascending filtration algorithm, the set of filtration value F and the corresponding persistence barcode $B = [a_j ; b_j] : j \in J$. A persistence line in a barcode is conventionally represented as $[a_j ; \infty)$ here it is substitute with $[a_j ; m)$ where $m = (\max\{F\} + 1)$.

$$E(F) = - \sum_{j \in J} p_j \log(p_j) \quad (11.4)$$

where $p_j = l_j/L$, $l_j = b_j - a_j$, and $L = \sum_{j \in J} l_j$

11.3 Analysis of the Idiotypic Network

11.3.1 Simulation Results

In this section, we provide a detailed study of the simulation of the *idiotypic network* obtained with *C-ImmSim* that is the agent-based simulator simulation of the immune system [20]. The computational ABM employed is discrete in mathematical terms because it represents biological entities as individuals in a heterogeneous population of cells and molecules. In particular, the major classes of cells of the lymphoid lineage and some of the myeloid lineage. Moreover the model accounts for various interleukins and messengers. This “discreteness” confers the model the character of being “easily scalable” in terms of introducing new biological complexity, at variance with corresponding equation based models. The model is stochastic, meaning it can naturally display biological “controlled” variability: for example, it is possible to separate the sorting of repertoire specificities from the random occurrences (encounters, binding, cell death, cell replacement, diffusion, cell division) during the running of the response. In other words, each repertoire expresses a private specificity, and by repeating runs with random events, the impact of different repertoires can be compared and their variations statistically determined, at the same time increasing the significance of results. The ABM model, is a polyclonal model, since all lymphocytes are equipped with a receptor represented as a binary string. This allows for a number of immunological features such as expressed and potential repertoire definition, specific recognition/binding, antigens peptides presentation, specific clonal memory, hypermutation, etc. In our configuration a simulation has a lifespan of 2190 ticks, where a tick = 8h, and a repertoire of at most 10^{12} antibodies and antigen volume equal to $V = 10 \mu\text{L}$. The results are the average over 100 runs. In the simulator each idioype (both antigens and antibodies) is represented with a bit-string, in our case of 12 bit length. An idioype interacts with each other if and only if their Hamming distance is $11 \leq d(A_j, A_k) \leq 12$. The pair-wise distances are stored in a matrix, the so-called *Affinity matrix*: $J_{i,k}$. We defined a weight function for the idiotypic network, the *coexistence function* between antibodies (see 11.5):

$$C_{Ab_{j,k}}(t) = \frac{d(Ab_j(t), Ab_k(t)) * [Ab_j(t)] * [Ab_k(t)]}{\sum_{l=1}^n [Ab_l(t)]} \quad (11.5)$$

where $[Ab_l(t)]$ is the concentration of the l th antibody Ab_l at tick t . The meaning of (11.5) is that *for lower values of affinity the concentration must be more significant because the match between antibodies is less probable* (Fig. 11.2).

For each simulation we computed the coexistence function (see 11.5) and we used the weighted idiotypic network to calculate the graph entropies (see 11.2, and 11.3) as input for the persistent homology computation. In this work we used the *clique weight rank persistent homology* (CWRPH) algorithm implemented in the new version of *jHoles*, and described in par.: Sect. 11.2.4. The output of the persistent homology has been used for computing both the *persistent entropy* (see 11.4) and for identifying

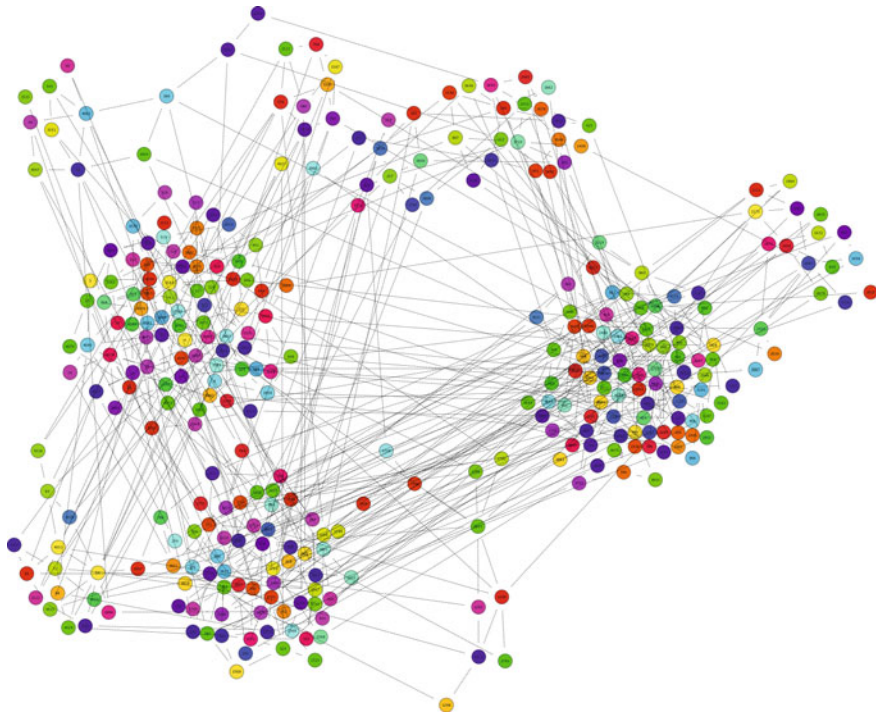


Fig. 11.2 Example of immune network at the end of the simulation (tick = 2190). The nodes represent the antibodies, a link exists if and only if two antibodies are affine. The node color represents the antibodies classes

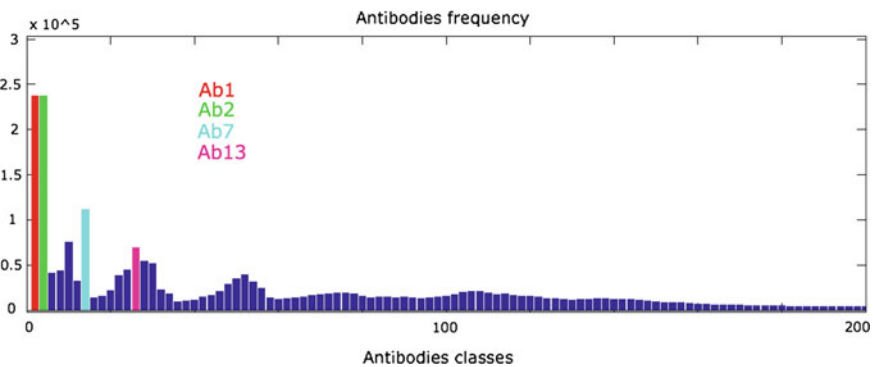


Fig. 11.3 Antibodies frequency

the persistent holes and their generators, namely the persistent antibodies that govern the evolution of the idiotypic network during the virgin state, the activation and the *immune memory* (see Fig. 11.3).

11.4 Concluding Remarks

The charts for each simulations highlight interesting emerging features. In all cases the peak corresponding to the *activation* of the *immune response* has been identified. However the *connectivity entropy* (see Fig. 11.4) does not distinguish between the activation and the *immune memory* states. The *connectivity entropy* highlighted that after time step 199 some new higher-degree antibodies are involved in the dynamics of the system. While, both the *approximated von Neumann entropy* (see Fig. 11.5) and the *persistent entropy* (see Fig. 11.6) are able to recognize the activation of the immune system: the peaks in the charts point out the *immune activation* that is following by a transient that represents the *immune response*. During the immune response the antibodies play a dual role: they can simultaneously elicit and suppress each other. After this transient there is a plateau that represents the persistent *immune network activation* corresponding to the immune memory. Persistent entropy is directly computed from the result of *persistent homology*: the Betti numbers. The analysis of the generators of the homological classes allows to identify the real number of antibodies that have been used: 203 instead of 4096 (see Fig. 11.3). The analysis of the persistent Betti numbers reveals that there is a subset of antibodies arranged in a 1-dimensional hole that is present both in the *activation state* and in the *memory state*. This 1-dimensional hole is formed by the antibodies: Ab_1 , Ab_2 , Ab_7 , Ab_{13} . This hole is formed by the most active antibodies, see the histogram in Fig. 11.3. The removal of this 1-dimensional hole from the barcodes will flatten the entropy, that means this cycle is formed by the most specialized antibodies for the antigen that has been injected. Both the *approximated von Neumann entropy* and the *persistent entropy* can be thought as *complexity measures* for graphs or for simplicial complexes. The reason is evident in their mathematical definitions: von Neumann entropy depends on the total number of vertices and the degree of linked vertices, while persistent entropy depends by the *topological noise* and by the *persistent topological features*.

To conclude, we suggest that *persistent entropy* and in general *topological data analysis* are useful tools for the analysis of dynamical complex systems. Topological data analysis and persistent entropy can be used for discovering hidden patterns among antibodies. The transformation of a graph into a simplicial complex of dimension greater than 1 allows to discover new patterns and than it allows to extract new knowledge that otherwise can not be captured. The dimension of these patterns is equivalent to the dimension of the relation among antibodies. A relation of dimension 2, typically expressed by a graph, represents a classical 2-bodies problem while a n-ary relation represents a n-(anti)bodies problem that makes sense if and only if all the (anti)bodies are communicating simultaneously. Generally, a n-ary relation can not be decomposed in a set of 2-bodies problems: e.g. a filled triangle, that is a simplicial complex of dimension 2, represents a 3-body problems, it exists if and only if simultaneously are present the three vertices and three edges, otherwise it is represented as an empty triangle (Fig. 11.1).

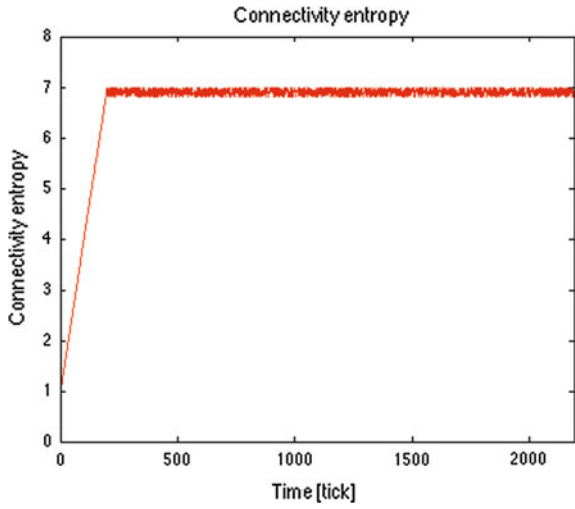


Fig. 11.4 Connectivity entropy. The maximum is reached at tick = 199

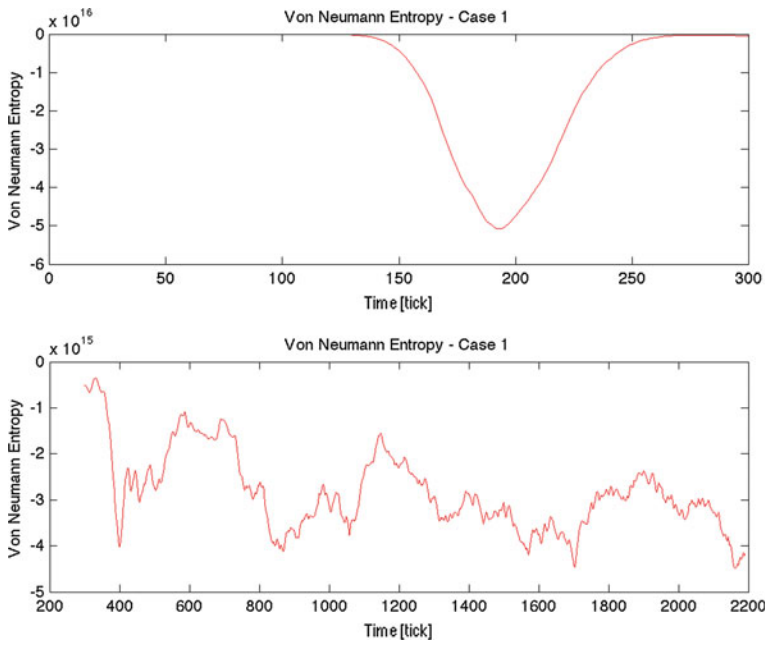


Fig. 11.5 Approximate von Neumann entropy. The minimum is reached at tick = 199

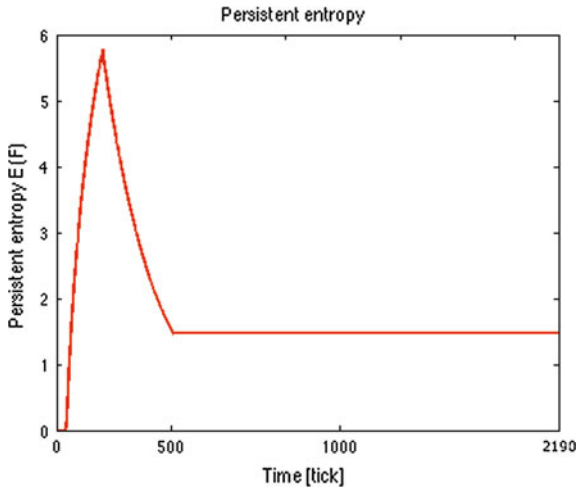


Fig. 11.6 Persistent entropy. The maximum is reached at tick = 199

Future investigation will be focusing on the use of the *persistent entropy* for characterising the S[B] systems [14], and on the use of other entropy measures, like the ones proposed by Felice [7].

Acknowledgments We acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme (FP7) for Research of the European Commission, under the FP7 FET-Proactive Call 8—DyMCS, Grant Agreement TOPDRIM, number FP7-ICT-318121.

References

1. Binchi, J., Merelli, E., Rucco, M., Petri, G., Vaccarino, F.: jHoles: A tool for understanding biological complex networks via clique weight rank persistent homology. *Elect. Notes Theoret. Comput. Sci.* **306**, 5–18 (2014)
2. Chan, J.M., Carlsson, G., Rabadan, R.: Topology of viral evolution. *Proc. Natl. Acad. Sci.* **110**(46), 18566–18571 (2013)
3. Chintakunta, H., Gentimis, T., Gonzalez-Diaz, R., Jimenez, M.-J., Krim, H.: An entropy-based persistence barcode. *Pattern Recogn.* **48**(2), 391–401 (2015)
4. de Silva, V., Ghrist, R.: Coverage in sensor networks via persistent homology. *Algebraic Geom. Topol.* **7**(339–358), 24 (2007)
5. Edelsbrunner, H., Harer, J.: *Computational Topology: An Introduction*. American Mathematical Soc. (2010)
6. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. *Discrete Comput. Geom.* **28**(4), 511–533 (2002)
7. Felice, D., Mancini, S., Pettini, M.: Quantifying networks complexity from information geometry viewpoint. *arXiv preprint arXiv:1310.7825* (2013)
8. Han, L., Escolano, F., Hancock, E.R., Wilson, R.C.: Graph characterizations from von neumann entropy. *Pattern Recogn. Lett.* **33**(15), 1958–1967 (2012)

9. Hoffmann, G.W.: A theory of regulation and self-nonsel self discrimination in an immune network. *Eur. J. Immunol.* **5**(9), 638–647 (1975)
10. Horak, D., Maletić, S., Rajković, M.: Persistent homology of complex networks. *J. Stat. Mech.: Theory Exp.* **2009**(03), P03034 (2009)
11. Ibekwe, A.M., Ma, J., Crowley D.E., Yang, C.-H., Johnson, A.M., Petrossian, T.C., Lum, P.Y.: Topological data analysis of *Escherichia coli* O157: H7 and non-O157 survival in soils. *Frontiers Cell. Infect. Microbiol.* **4** (2014)
12. Jankowski, A., Skowron, A.: *Practical Issues of Complex Systems Engineering: Wisdom Technology Approach* (2014)
13. Jonsson, J.: *Simplicial Complexes of Graphs*, vol. 1928. Springer (2008)
14. Merelli, E., Pettini, M., Rasetti, M.: Topology driven modeling: the IS metaphor. *Nat. Comput.* 1–10 (2014)
15. Mortveit, H., Reidys, C.: *An Introduction to Sequential Dynamical Systems*. Springer Science & Business Media (2007)
16. Ortiz-Arroyo, D., Akbar Hussain, D.M.: An information theory approach to identify sets of key players. In: *Intelligence and Security Informatics*, pp. 15–26. Springer (2008)
17. Passerini, F., Severini, S.: The von Neumann entropy of networks. *arXiv preprint [arXiv:0812.2597](https://arxiv.org/abs/0812.2597)* (2008)
18. Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P.J., Vaccarino, F.: Homological scaffolds of brain functional networks. *J. R. Soc. Interface* **11**(101), 20140873 (2014)
19. Petri, G., Scolamiero, M., Donato, I., Vaccarino, F.: Topological strata of weighted complex networks. *PloS ONE* **8**(6), e66506 (2013)
20. Rapin, N., Lund, O., Castiglione, F.: Immune system simulation online. *Bioinformatics* **27**(14), 2013–2014 (2011)
21. Rucco, M., Falsetti, L., Herman, D., Petrossian, T., Merelli, E., Nitti, C., Salvi, A.: Using Topological Data Analysis for diagnosis pulmonary embolism. *arXiv preprint [arXiv:1409.5020](https://arxiv.org/abs/1409.5020)* (2014)
22. Stein, D.L., Newman, C.M.: Nature versus nurture in complex and not-so-complex systems. In: *ISCS 2013: Interdisciplinary Symposium on Complex Systems*, pp. 57–63. Springer (2014)
23. Tomita, E., Tanaka, A., Takahashi, H.: The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoret. Comput. Sci.* **363**(1), 28–42 (2006)