




Tracing patterns and shapes in remittance and migration networks via persistent homology

Paul Samuel P. Ignacio^{1,2*}  and Isabel K. Darcy¹

*Correspondence:

paulsamuel-ignacio@uiowa.edu

¹Department of Mathematics,
University of Iowa, Iowa City, USA

²Department of Mathematics and
Computer Science, University of the
Philippines Baguio, Baguio City,
Philippines

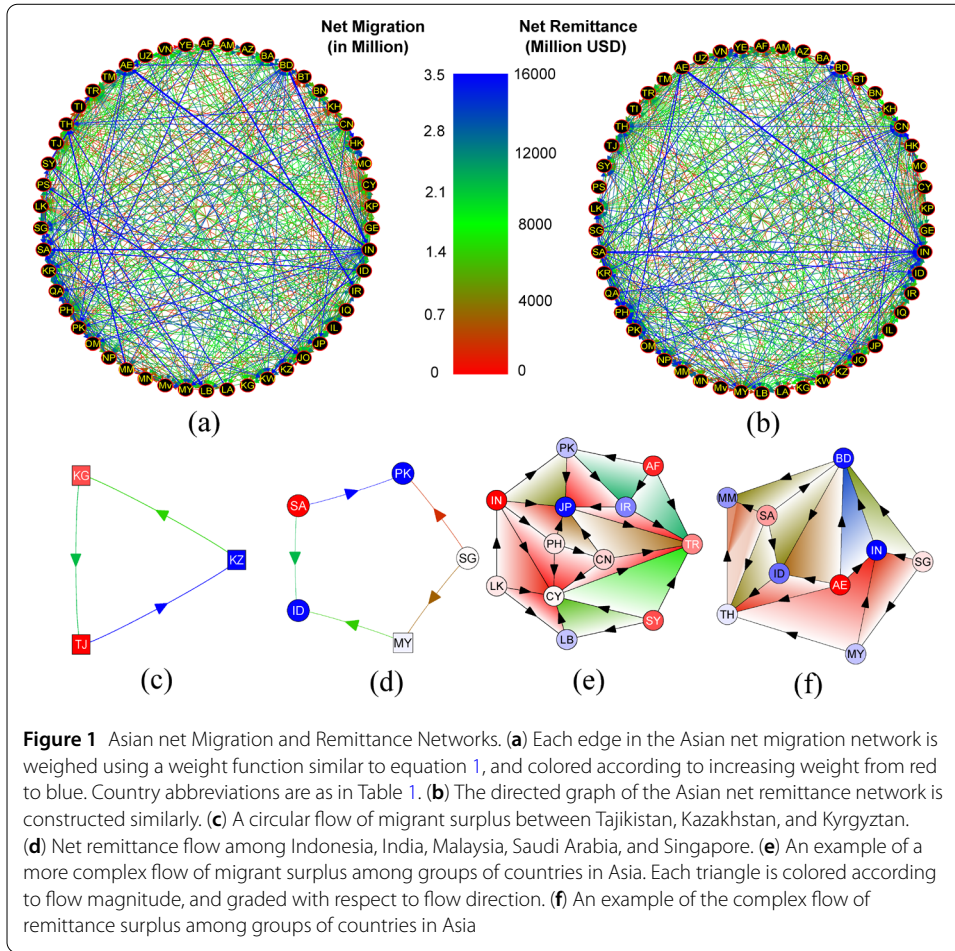
Abstract

Pattern detection in network models provides insights to both global structure and local node interactions. In particular, studying patterns embedded within remittance and migration flow networks can be useful in understanding economic and sociologic trends and phenomena and their implications both in regional and global settings. We illustrate how topo-algebraic methods can be used to detect both local and global patterns that highlight simultaneous interactions among multiple nodes, giving a more holistic perspective on the network fabric and a higher order description of the overall flow structure of directed networks. Using the 2015 Asian net migration and remittance networks, we build and study the associated directed clique complexes whose topological features correspond to specific flow patterns in the networks. We generate diagrams recording the presence, persistence, and perpetuity of patterns and show how these diagrams can be used to make inferences about the characteristics of migrant movement patterns and remittance flows.

Keywords: Migration network; Remittance network; Persistent homology

1 Introduction

Migration and remittances have become important facets of modern human society. The United Nations' Department of Economics and Social Affairs (UNDESA) reports that the total number of international migrants, defined as foreign-born or foreign citizens for the purpose of estimation, has continued to grow rapidly from year 2000 to year 2015 at a rate faster than that of the world's population. Many published reports, such as UNDESA's [1] and World Bank's [2], provide comprehensive statistical descriptions and forecasts on world migration and remittances. These, however, tend to focus on statistics about direct and specific region-region interactions: origin and destination of highest flow, rate of increase or decrease of flow, etc. On the other hand, studying migration and remittances in the network setting allows us to generate a clearer picture of the complex interactions embedded in its structure. Fagiolo and Mastroiello [3] used a complex-network perspective to study the binary and weighted architecture of the international migration network, exploring link and node statistics, assortativity, and clustering among others. In [4, 5], centrality in the migration network was studied in terms of weighted and unweighted vertex indegree^a and outdegree. Simply put, these measure the diaspora and reach of migrants



coming in and going out of a particular country. In addition, community detection via link density was also explored.

In this note, we demonstrate how we can use topo-algebraic methods to identify flow patterns in the Asian net migration and remittance networks resembling cycles involving multiple countries. One advantage of this method over graph theory methods is that it is able to recover higher dimensional patterns, capturing more complex interactions among multiple nodes and offering a more holistic perspective on the topology of the network (see Fig. 1). For a list of countries included in this analysis, see Table 1.

We define net remittance to be the positive difference in the remittances exchanged by two countries. We consider the *net remittance network* $R = (V, E_R)$ whose vertex set V represents countries, and edges weighed according to net remittances with direction towards the country that receives more. Hence, the weight of a directed edge $(a, b) \in E_R \subseteq V \times V$ represents the profit country b gains from exchanging remittances with country a and can be thought of as a measure of the flow of remittance surplus between the two countries. We can store the information in this directed network into an incidence-weight matrix $W = [w_{ab}]$ where

$$w_{ab} = \begin{cases} r_{ab} - r_{ba} & \text{if } r_{ab} > r_{ba}, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

Table 1 Ordered list of Asian countries with abbreviation codes

Order	Country	Abbrev.	Order	Country	Abbrev.
1	Afghanistan	AF	26	Lebanon	LB
2	Armenia	AM	27	Malaysia	MY
3	Azerbaijan	AZ	28	Maldives	MV
4	Bahrain	BA	29	Mongolia	MN
5	Bangladesh	BD	30	Myanmar	MM
6	Bhutan	BT	31	Nepal	NP
7	Brunei Darussalam	BN	32	Oman	OM
8	Cambodia	KH	33	Pakistan	PK
9	China	CN	34	Philippines	PH
10	Hong Kong	HK	35	Qatar	QA
11	Macau	MO	36	Republic of Korea	KR
12	Cyprus	CY	37	Saudi Arabia	SA
13	Dem. People's Rep. of Korea	KP	38	Singapore	SG
14	Georgia	GE	39	Sri Lanka	LK
15	India	IN	40	State of Palestine	PS
16	Indonesia	ID	41	Syria	SY
17	Iran	IR	42	Tajikistan	TJ
18	Iraq	IQ	43	Thailand	TH
19	Israel	IL	44	Timor-Leste	TI
20	Japan	JP	45	Turkey	TR
21	Jordan	JO	46	Turkmenistan	TM
22	Kazakhstan	KZ	47	United Arab Emirates	AE
23	Kuwait	KW	48	Uzbekistan	UZ
24	Kyrgyzstan	KG	49	Vietnam	VN
25	Laos	LA	50	Yemen	YE

and r_{ab} represents the total remittance from a to b . The definition of the weight function yields a simple directed graph admitting no self-loops (see Fig. 1). We define net migration and the migration network $M = (V, E_M)$ in a similar manner. With these definitions, the extracted patterns then reflect the flow dynamics of migrants and remittances within groups of countries, which may provide insights about inter-country relationships and their relative impacts on the group's overall dynamics. The incidence-weight matrix of the net migration and remittance networks used in this study is available upon request from the authors.

Our approach utilizes a relatively new method in data analysis known as *persistent homology* whose early roots trace back to Size Theory in the 90's [6–9] (for connected components) and later generalized to higher dimensions by Edelsbrunner, Letscher, and Zomorodian [10]. We refer the reader to [11–13] for a detailed account of the history and development of the theory. In this note, we apply persistent homology to data viewed as a directed network, from which we build a sequence of mathematical objects each contained in the next. Distinct (non-homologous) generators of topological features that persist significantly along the sequence of mathematical objects are then regarded as encoding the overall shape of the point-cloud. We make precise what we mean by generators in Sect. 3.2. Several works have employed this approach in studying the topology of undirected networks [14–17]. An extension of this approach to directed networks using path complexes has been explored by Chowdhury and Mémoli [18, 19], and via neighborhood complex by Horak et al. [20]. In [17], Petri et al. introduced a filtration that allowed them to compute the persistent homology of weighted undirected networks, and remarked that their methods are amenable to extension to directed networks following the directed clique construction of Palla et al. [21]. Reimann et al. [22] used the same directed clique construction to compute homology of neural networks, but they did not analyze the persis-

tence of features. The contribution of our work is the implementation and development of Petri et al.’s idea to extend persistent homology theory to directed networks via cliques. The integration of the directional component to the theory allows us to talk about *cycle types* and *perpetuity* of topological features of directed networks. We expound on these notions in Sect. 3. We also present a modification to barcodes, the output diagram of persistent homology, by introducing a coloring scheme that distinguishes features of the same dimension.

The rest of the paper is organized as follows. Section 2 describes the data sets and the estimation measures used. In Sect. 3, we discuss the theoretical foundations, highlight the contributions of our work, and present an example using a toy network. Section 4 presents the patterns and shapes extracted from the Asian net migration and remittance networks. Finally, Sect. 5 presents a summary and concluding remarks.

2 Dataset description

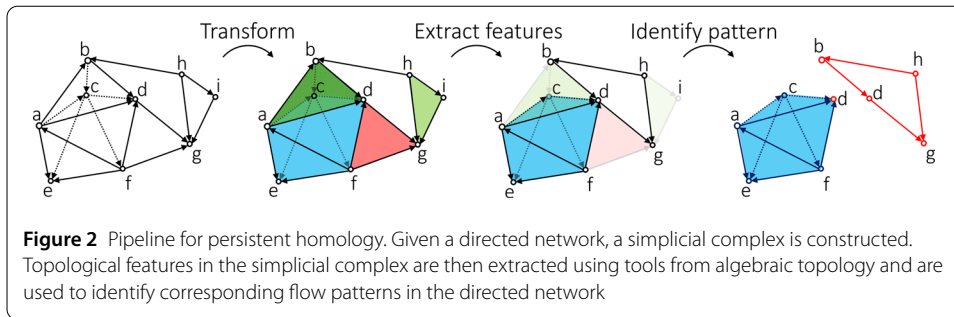
According to UNDESA’s international migration report in 2015 [1], nearly half of all international migrants worldwide were born in Asia, and nearly a third live there. From year 2000 to year 2015, this region had more new international migrants than any other major area.^b Interestingly, Asia also ranks second on the percentage (60%) of people living in a different country but within the same region. These make Asia one of the most highly fluid regions in terms of internal migrant mobility.

We perform our analysis on the 2015 Asian net migration and remittance networks which include 50 countries and states as listed in Table 1. We use data on foreign-born population obtained from the UN Global Migration Database and on bilateral remittances from the World Bank database as reported respectively in [23] and [24].

The migration data uses estimates on foreign-born population based on censuses, registers, and models. These estimates also take into account factors such as the estimated size of the total population in the country of destination, and specific country circumstances such as sudden migrant influx or exodus due to conflict, economic booms or busts, and major changes in migration policies. For lack of data source, estimates for the Democratic People’s Republic of Korea were based on “model” countries chosen according to similarity in criterion for enumerating international migrants, proximity, and migration experience. A complete documentation of the migration data set is available in [25].

The data set on bilateral remittances is a composite of the migration data and the total remittance inflows. Let M_{ji} represent the number of migrants from country j living in country i , and R_j the total remittance inflow to country j from all countries, computed as the sum, where data is available, of two components in the International Monetary Fund’s Balance of Payments Statistics: compensation of employees and personal transfers. For some countries, data is obtained from the respective country’s Central Bank and other relevant official sources. A detailed account of this computation can be found in [2]. From M_{ij} and R_i , the bilateral remittances between two countries are estimated using a method devised by Ratha and Shaw [26]. The remittance from country i to country j is the product of the number of migrants M_{ji} , and the average remittance ρ_{ij} sent home by a migrant as modeled by

$$\rho_{ij} = \begin{cases} Y_j & \text{if } Y_i < Y_j, \\ Y_j + (Y_i - Y_j)^\beta & \text{if } Y_i \geq Y_j, \end{cases} \tag{2}$$



where Y_i and Y_j are respectively the average gross national income per capita of the host and home countries, and β is a parameter between 0 and 1. This model assumes that an average migrant sends at least as much as the per capita income from home. For each country, the parameter β is estimated so that the total lateral remittances to country j matches R_j , that is

$$R_j = \sum_i \rho_{ij} M_{ji}. \tag{3}$$

The World Bank [24] states that some issues about these estimates are:

- (a) the data on migrants in various destination countries are incomplete;
- (b) the incomes of migrants abroad and the costs of living are both proxied by per capita incomes in terms of purchasing power parity;
- (c) there is no way to capture remittances flowing through informal, unrecorded channels.

The data for bilateral remittances does not include State of Palestine so we treat these missing values as zero.

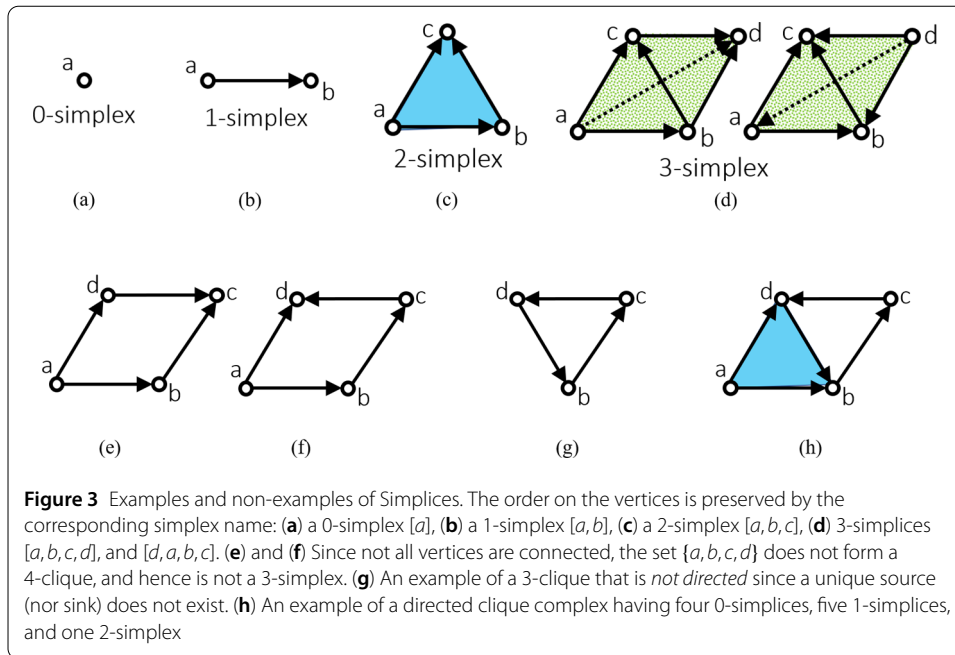
3 Methods

3.1 Directed clique complexes

The central idea is the following: *given a directed network, construct a mathematical object C (called a simplicial complex) in such a way that resulting topological features of C correspond to patterns in the network* (see Fig. 2). One advantage of this approach is that these patterns, which give a higher-order description of the underlying network (i.e. beyond clustering or linking), can be easily extracted via their corresponding topological features using tools from algebraic topology.

A directed network can be viewed as a collection of smaller subnetworks called *directed cliques* [21]. A k -clique $\{v_{i_1}, v_{i_2}, \dots, v_{i_k}\}$, i_j a nonnegative integer for $1 \leq j \leq k$, is a set of k vertices that induce a subnetwork where every pair of vertices is connected by an edge. A *directed k -clique* is a k -clique with a unique *source*^c and a unique *sink* (see Fig. 3). Observe that this definition induces an ordering on the set of k vertices and a maximal directed path $v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_{k-1}$. Thus, this directed k -clique will be denoted by the ordered k -tuple $[v_0, v_1, \dots, v_k]$. This gives a natural way of simplifying network flow structure by adopting, whenever possible, an overall unidirectional flow from the source to the sink in vertex communities having complete local connections.

To transform our directed network into a topological object, it is sufficient to define what the simplices are. The *directed clique complex* is obtained by defining simplices of



dimension k to be directed $k + 1$ cliques. Doing so transfers, in a natural way, the directional information on cliques to spatial information on simplices as the latter can be visually represented by geometric objects: a 0-simplex is a point, a 1-simplex is a directed edge connecting two points, 2- and 3-simplices are respectively filled in triangles and solid tetrahedra whose edges form directed paths from the unique source to the unique sink as in Fig. 3.

We remark here that our definition of simplices via directed cliques is motivated by the flow structure we are concerned with. A k -simplex represents a group of $k + 1$ countries where pairwise interactions exist between any two members, and the overall flow structure can be characterized by the unidirectional flow from a unique source (that sends to every country) to a unique sink (that receives from every country). Masulli and Villa [27] used the same construction to investigate how resulting topological invariants, the Euler characteristic and network degree, can be used to assess specific functional and dynamical properties of directed networks. This definition can be modified to fit an appropriate alternate model. One alternative construction, due to Grigor'yan et al. [28–30] uses paths in a directed network to define simplices. This construction allows more flexibility in the definition of simplices, which in turn gives rise to an increased complexity in simplex variety that makes the approach less intuitive. For example, the network on four vertices in Fig. 3(e) is in fact the sum of two 2-paths, and hence regarded as a two-dimensional simplex generator in a *path complex*. This however, as Chowdhury and Mémoli [19] point out, shows that path homology is able to differentiate this particular motif from other types of cycles. Another construction due to [31] is via ordered tuple complexes where simplices are ordered tuples (v_0, v_1, \dots, v_p) closed under deletion of entries, and where repetition of vertices is allowed.

3.2 Directed clique homology

Now that the directed network has been converted to a simplicial complex, we can extract its topological features by employing a standard method in algebraic topology originally

used to classify topological spaces. Let C denote the simplicial complex associated to the directed network. For each $n = 0, 1, 2, 3$, we consider the set $K_n(C)$ of all n -simplices in C , and the free vector space $C_n = \mathbb{Z}/2\mathbb{Z}[K_n(C)]$ of all formal linear combinations of n -simplices with coefficients either 0 or 1. We can relate consecutive spaces C_n and C_{n-1} by extending linearly the *boundary map* $\partial_n : K_n \rightarrow K_{n-1}$ defined on simplices as

$$\partial_n(\sigma_n) = \sum_{i=0}^n [v_0, v_1, \dots, \hat{v}_i, \dots, v_n], \tag{4}$$

where \hat{v}_i means v_i is omitted. By definition, the boundary of a 0-simplex is zero. For example, if $\sigma_2 = [a, b, c]$ as in Fig. 3,

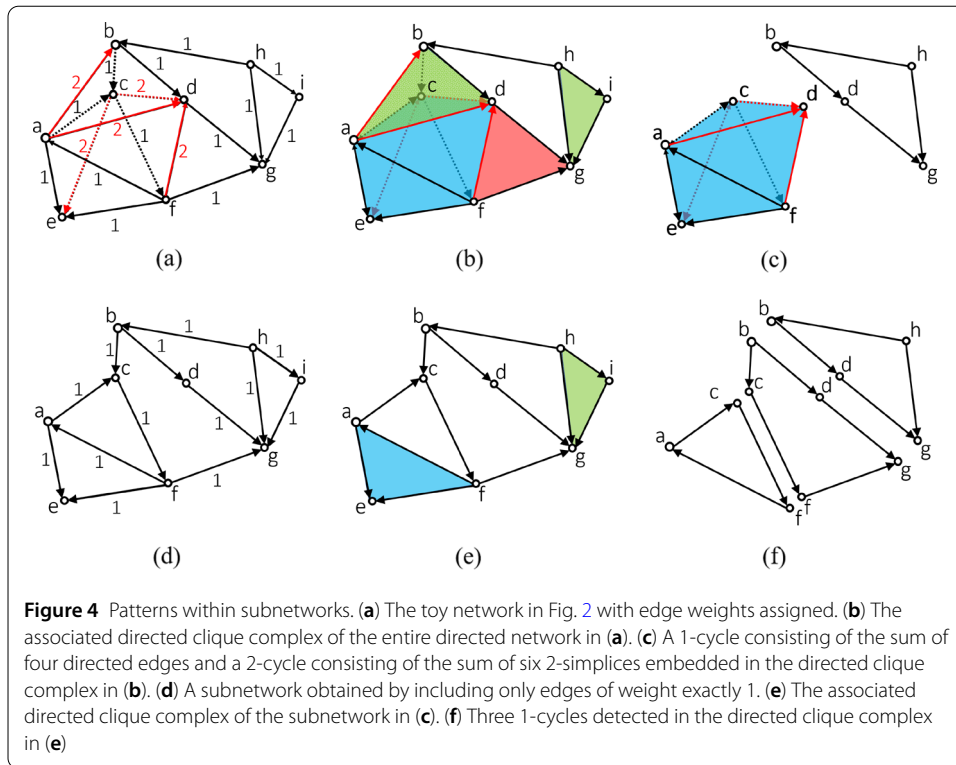
$$\partial_2(\sigma_2) = [b, c] + [a, c] + [a, b].$$

The collection of the outputs of the map ∂_n , which we will call *boundaries*, is denoted by $\text{Im } \partial_n$. It can be verified that the images under ∂_1 for the directed clique complexes in Figs. 3(e), 3(f), and 3(g) are all zero. In general, elements of C_n whose image under ∂_n is zero are called *n-cycles*, and the collection of all such elements is denoted by $\ker \partial_n$. Observe that applying ∂_n to the output of ∂_{n+1} always yields zero. This means that all boundaries of sums of $n + 1$ -simplices are n -cycles in C_n . The *nth homology group* H_n is then defined as the abstract algebraic quotient space

$$H_n = \ker \partial_n / \text{Im } \partial_{n+1}. \tag{5}$$

In this space, any n -cycle that is the boundary of a linear combination of $n+1$ -simplices is treated as zero, and hence regarded algebraically trivial. For example, the cycle $c_1 = [a, d] + [d, b] + [a, b]$ is trivial since it is the boundary of the 2-simplex $[a, d, b]$. As a consequence, a pair of n -cycles whose difference is trivial in the above sense are considered *homologous*. For example, the 1-cycles $[a, b] + [b, c] + [c, d] + [a, d]$ in Fig. 3(f) and $[b, c] + [c, d] + [d, b]$ in Fig. 3(g) are homologous in the directed clique complex in Fig. 3(h) since their difference is precisely c_1 . Alternatively, these two 1-cycles can be viewed as surrounding the same “hole” in the directed clique complex. We then say that these 1-cycles generate the same class in the homology group H_1 . In a similar manner, the generators of the homology group H_n are the representative linear combinations of n -dimensional simplices that surround distinct topological features embedded in the simplicial complex. These topological features capture complex structures and interactions in the network fabric (see Fig. 1(c)–(f)). For instance, the generators of H_0 partition the vertices into clusters in a manner similar to single linkage clustering. We will expound on this in Sect. 3.6.

Notice that a variety of patterns may arise from a given feature. For example, a 1-cycle in the directed clique complex may correspond to any of the patterns in Figs. 3(e)–3(g). For differentiation, we will refer to 1-cycles whose corresponding pattern in the directed network has no source nor sink, as in Fig. 3(g), as *circuits*. This pattern depicts a circular flow in the network. In addition, 1-cycles whose pattern has a unique source and a unique sink will be called *Type 1* 1-cycles (see Figs. 3(e) and 3(f)), while 1-cycles with multiple sources and sinks will be referred to as *Type 2* 1-cycles.

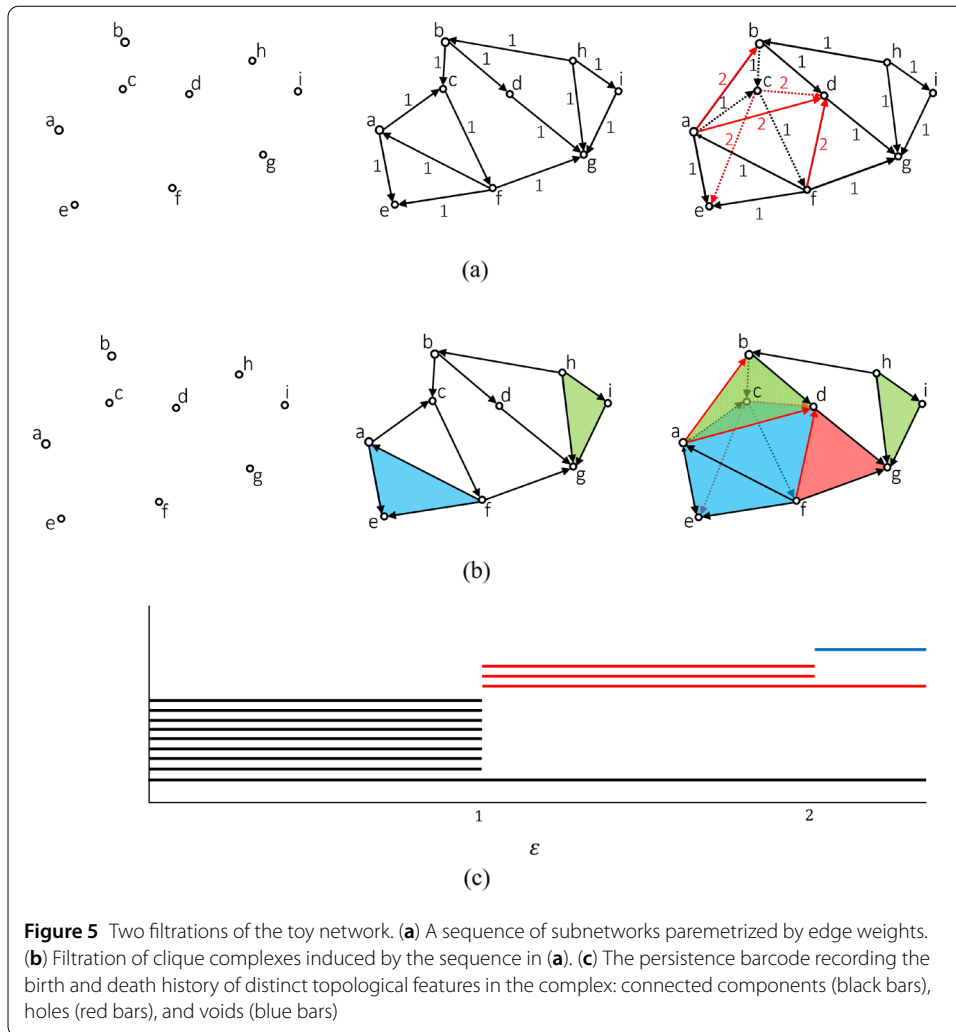


3.3 Filtrations, persistence, and perpetuity

Let us assign weights to our toy network (see Fig. 2). Examining the entire directed network, we see in Fig. 4(c) that it contains a Type 1 1-cycle $([h, b] + [b, d] + [d, g] + [h, g])$ and a 2-cycle $([a, c, d] + [c, f, d] + [f, a, d] + [a, c, e] + [c, f, e] + [f, a, e])$. Now, suppose we filter the network by including only edges of weight exactly 1, then build its corresponding directed clique complex (see Fig. 4(d)–4(f)). In this case, we detect three 1-cycles and no 2-cycles. This shows that *filtering* the network via edge weight may reveal more patterns invisible to homology when all edges are included. For this reason, we will consider a sequence of directed networks (each a subnetwork of the next) parametrized by edge weight: for each threshold $\epsilon \geq 0$ we construct the subnetwork consisting of all vertices and all directed edges of weight at most ϵ (see Fig. 5(a)). This in turn induces an associated filtration of directed clique complexes (see Fig. 5(b)).

By examining the features of each subcomplex and keeping track of their representative generators, we can identify topological features that persist significantly across the development of the complex. Persistent cycles are regarded as significant patterns in the directed network. This is the main idea of persistent homology due to Edelsbrunner, Letscher, and Zomorodian [10] who originally applied it over arbitrary filtrations of simplicial complexes. Many adaptations and refinements have been proposed to make persistent homology computationally efficient [32, 33] and suited for a wide variety of data sets. In particular, Petri et al. [16, 17] applied persistent homology on the filtered clique complex for weighted *undirected* networks. As suggested in their paper, we modify their approach by adopting the directed k -clique construction of Palla et al. [21].

Throughout the network development, cycles can be trivialized by becoming the boundary of higher dimensional simplices. Cycles can also become homologous to older cycles. In these cases, we say that the cycle *dies*. Otherwise, we say it *persists*. This birth and



death history of homological cycles throughout the filtration can be transcribed in a *barcode*, a diagram consisting of bars representing the birth and death history of features of all dimensions. In a barcode, the length of the bars record the *persistence* of the features, and measure how long a pattern survives before it gets killed off. Figure 5(c) shows the persistence barcode of the filtered directed clique complex in Fig. 5(b): black bars represent connected components, red bars represent 1-cycles, and the blue bar represents the lone 2-cycle. The long black bar signifies that connectivity in the directed network is characterized by a single component. Similarly, the longest red bar detects the single persistent 1-cycle ($[h, b] + [b, d] + [d, g] + [h, g]$), and the blue bar detects the 2-cycle in Fig. 4(c).

It must be noted that the intrinsic structure of the directed network and the filtration we use significantly influence how the resulting persistence barcode is read. In the usual undirected persistent homology on finite metric spaces, since the threshold filters only through *distances between vertices*, higher dimensional simplices are always created as the threshold scans through all distances, and holes of any dimension eventually get filled in. This means that all but one 0-dimensional bar in the barcode must die. In contrast, our construction of the simplicial complex limits the simplices to maximally connected subnetworks built up from *actual directed edges connecting vertices*. Hence, 1-dimensional sim-

plices are introduced only when corresponding directed edges are present, and once the threshold reaches the maximum edge weight, no additional simplices can be constructed. This then means that in the barcode, the topological features of the entire directed network are encoded by the bars that never die. We will refer to these bars as *perpetual*. We note that perpetual bars indeed represent the Betti numbers after computing homology at the end of the filtration. As we are using a similar filtration (see Sect. 3.4) implemented by JHOLES [34] to compute persistent homology for weighted *undirected* networks, the Betti numbers computed by JHOLES and the perpetual bars in our barcode have the same interpretation.

We compute persistent directed clique homology using the standard persistence algorithm that appears in [35] adapted for directed clique complexes using code written from scratch.

3.4 Max-to-min weight filtration

Given a network with incidence-weight matrix $[w_{ij}]$, we can transform nonzero weights w_{ij} using the function

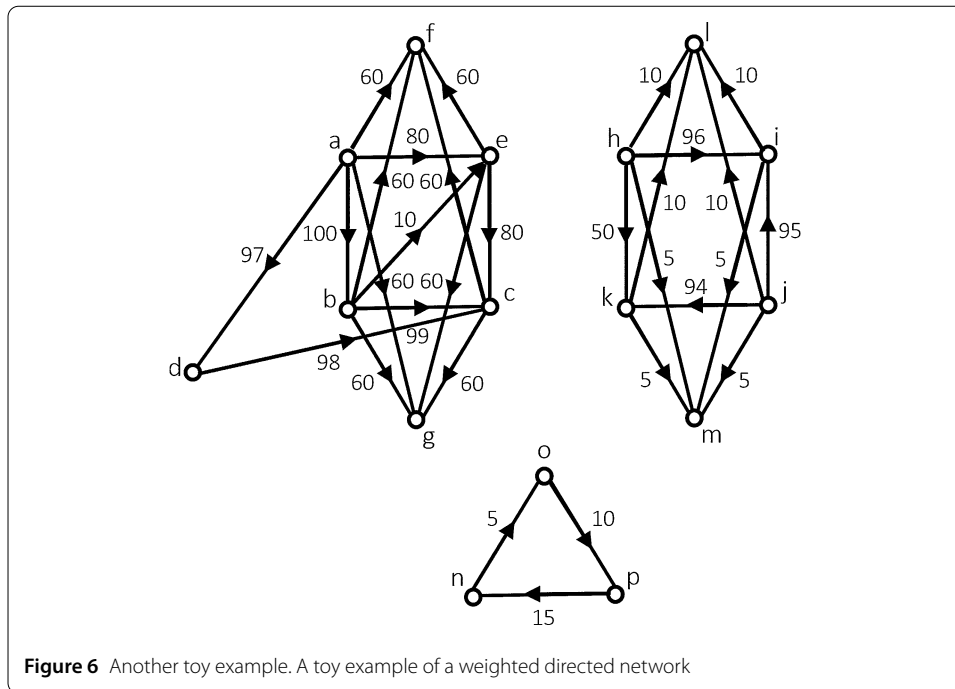
$$f(w_{ij}) = (\max([w_{ij}]) + 1) - w_{ij} \quad (6)$$

and induce the filtration of clique complexes using the transformed values. This transformation amounts to thresholding from the largest weight in the network to the smallest, and is done so that persistence picks up the most significant (magnitude-wise) flow patterns. With this transformation, cycles having large weights (i.e. significantly large flows) are born early, and since persistence depends on birth and death of features, cycles that are born early but get killed off by edges having significantly smaller weights still have large persistence. This approach is essentially the *weight rank clique filtration* introduced in [17] with the threshold reparametrized to reflect increasing differences from the maximum weight, and is different from most other filtrations as thresholding is via decreasing dissimilarities: two 0-simplices will be connected early if the directed flow between them is large.

We will employ this approach in computing persistent features in both net migration and remittance networks.

3.5 Variation on persistence barcodes

Since homological cycles correspond to *actual* subnetworks, we can encode additional information in the 1- and 2-dimensional barcodes by coloring the bars according to the standard deviation of the weights in the cycles they represent. We adhere to the elder rule and use the oldest generators to represent homology classes, that is, we compute the standard deviation of all the edge weights present in the oldest linear combination of simplices that surround a topological feature. This allows differentiation between cycles of the same dimension according to the variation in the edge weights, and is most useful for distinguishing cycles that are born at a later stage in the filtration as illustrated in an example in the following section. The introduction of this coloring scheme supplements the information encoded within barcodes, and could potentially be combined with other tools, such as persistence landscapes [36, 37], homological scaffolds [38], and persistent entropy [39, 40], that enrich the information extrapolated from barcodes.

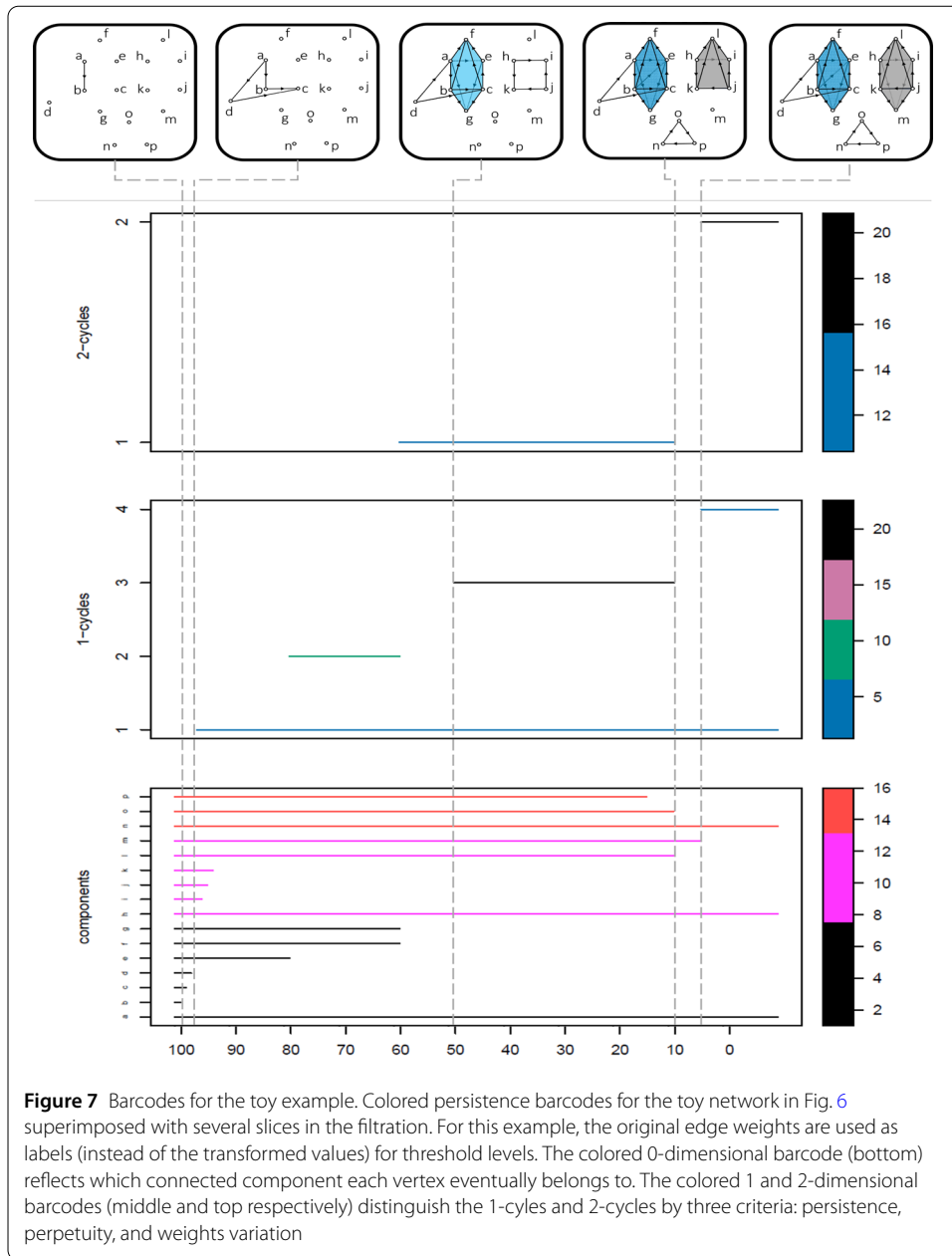


3.6 Another toy example

In this example, we will use the same techniques introduced in the beginning of this chapter. The difference from the previous example is that we are going to use the transformed values to filter the weights in the directed network so that along the filtration, edges having large weights are born first as described in Sect. 3.4.

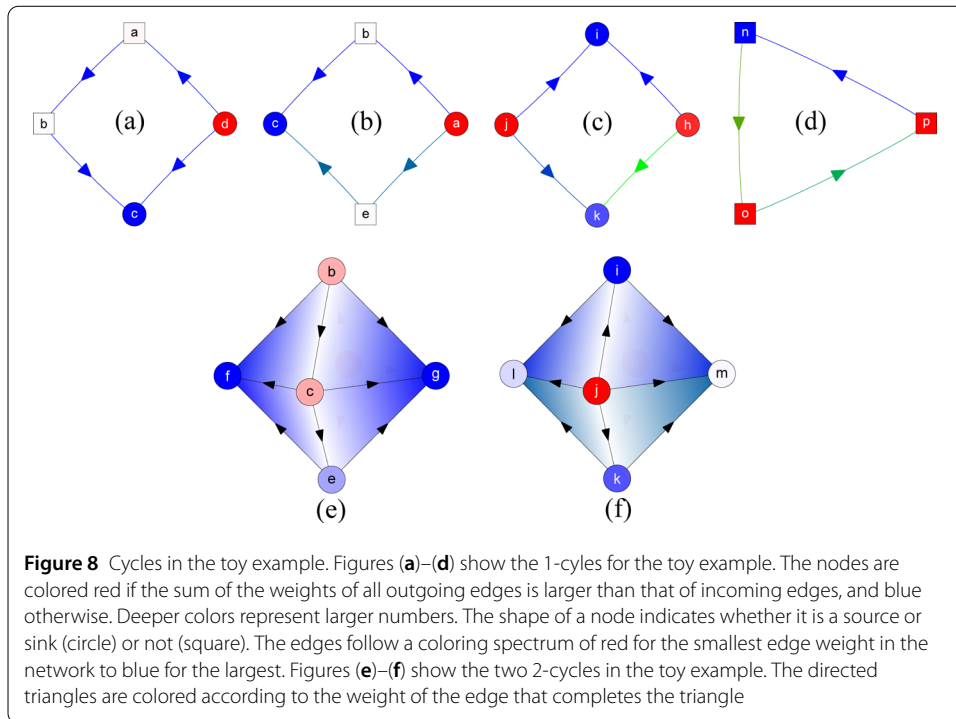
Consider the weighted directed network given in Fig. 6. Several slices of the filtration are superimposed on the persistence barcode shown in Fig. 7. The three persistent and perpetual bars in the 0-dimensional barcode (bottom) detects the three connected components in the directed network. Each bar in this barcode represents a vertex in the network, and is colored according to the component it eventually belongs to. In general, the connected components are formed in a manner similar to how clusters are formed via single linkage hierarchical clustering: a vertex is connected to a component at threshold ϵ if the maximum weight of the edges connecting the vertex to any vertex in the component is at least ϵ . Hence, just as the single linkage dendrogram, the 0-dimensional barcode records the number of connected components in the underlying undirected graph formed at any threshold level.

The 1-dimensional barcode (middle of Fig. 7) detects the four 1-cycles in Figs. 8(a) to 8(d). Let us examine the information each bar provides. The early birth and small standard deviation of the first persistent and perpetual bar shows that the cycle it represents has large and almost equal edge weights. Its persistence and perpetuity suggest that this cycle is an actual 1-dimensional feature of the entire directed network. The second bar starts relatively early, suggesting that the corresponding cycle has relatively large weights. The increased but small standard deviation reveals that the edge weights are slightly spread, and its small persistence reveals that this cycle gets killed off shortly after birth. The third bar born midway through the diagram reveals that the smallest edge weight in the cycle it represents is half of the largest edge weight in the directed network. The extremely high standard deviation of this bar shows that the edge weights in the cycle are considerably



spread, and the relatively high persistence shows that this pattern survives for a while before getting trivialized. The last bar being born late reveals that the edge weight completing the cycle is very small. Its small standard deviation suggests that the other edge weights in the cycle are also small. Its perpetuity reveals that this cycle is also a 1-dimensional feature of the directed network.

The 2-dimensional barcode detects the two spheres in Figs. 8(e) and 8(f). The first persistent bar being born before the halfway mark in the diagram shows that all edges in the 2-cycle have weights bigger than half of the largest weight in the network. Although not extremely spread, the large range of values covered by its assigned color suggests that the weight distribution in the 2-cycle is somewhat spread. The death of the bar signifies that

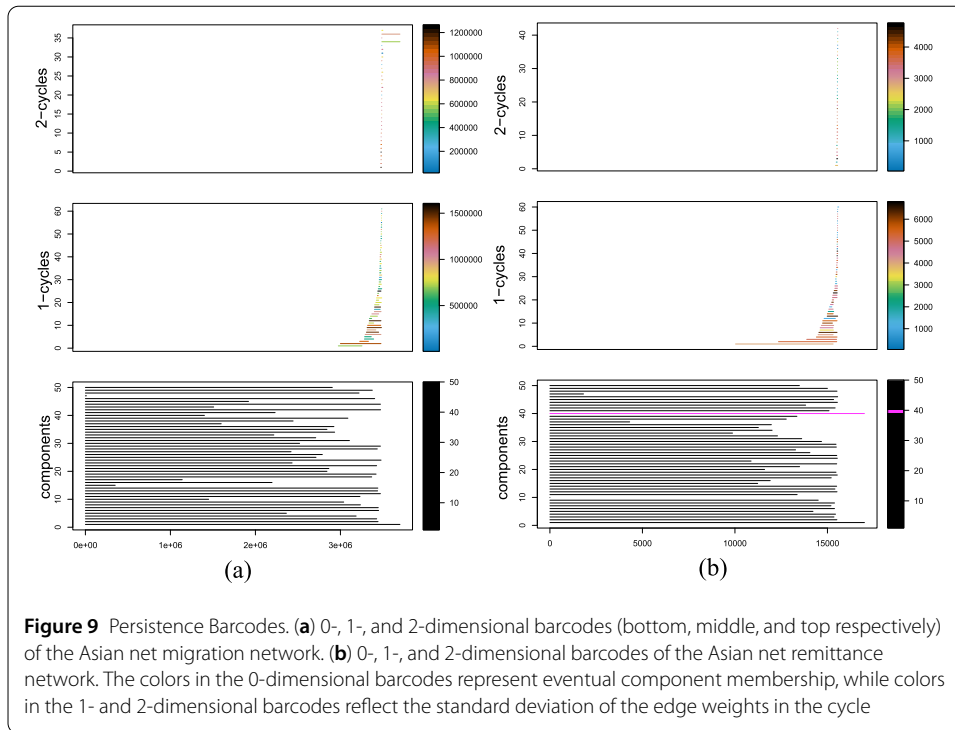


this 2-cycle is eventually trivialized. The perpetual black bar shows that a 2-cycle with highly variable edge weights is a 2-dimensional feature of the entire directed network.

4 Patterns and shapes in Asian net migration and remittance networks

The barcodes in Fig. 9 show the persistence of the homological cycles of dimensions $n = 0, 1, 2$ for the 2015 Asian net migration (9(a)) and remittance (9(b)) networks. The bars in the 0-dimensional barcode represent the Asian countries arranged as in order given in Table 1 starting from the bottom. For the net migration network (bottom of Fig. 9(a)), the single perpetual bar indicates that every country in Asia had an unequal exchange of migrants with at least one other Asian country in 2015. For the net remittance network, the two persistent and perpetual bars represent the distinct perpetual connected components in the network. The bottom perpetual bar represents the connectedness of 49 out of the 50 states in the Asian remittance network, while the other perpetual bar represents State of Palestine whose remittance data is missing. In both the migration and remittance networks, the relatively long bars in the 0-dimensional barcode suggest, in accordance with the reparametrization described in equation (6), that the net migration and remittances between countries in Asia are mostly small relative to the largest net migrant flow (3,487,351 people from IN to AE) and remittance (15,558.59 million USD from HK to CN).

The distinct 1- and 2-cycle generators for the Asian migration network appear in Figs. 10–11 and Figs. 12–13 respectively arranged beginning from the generator that corresponds to the lowest bar in the appropriate barcode. The edges of 1-cycles are colored according to weight. These 1-cycles reveal patterns that highlight both intra- and inter-regional migrant movement, and the role that each country plays in these groups. For example, while IN tends to be a dominant source of migrants within cycles, and SA a dominant sink, they do become transitory countries in Figs. 10.16 and 11.39 respectively.

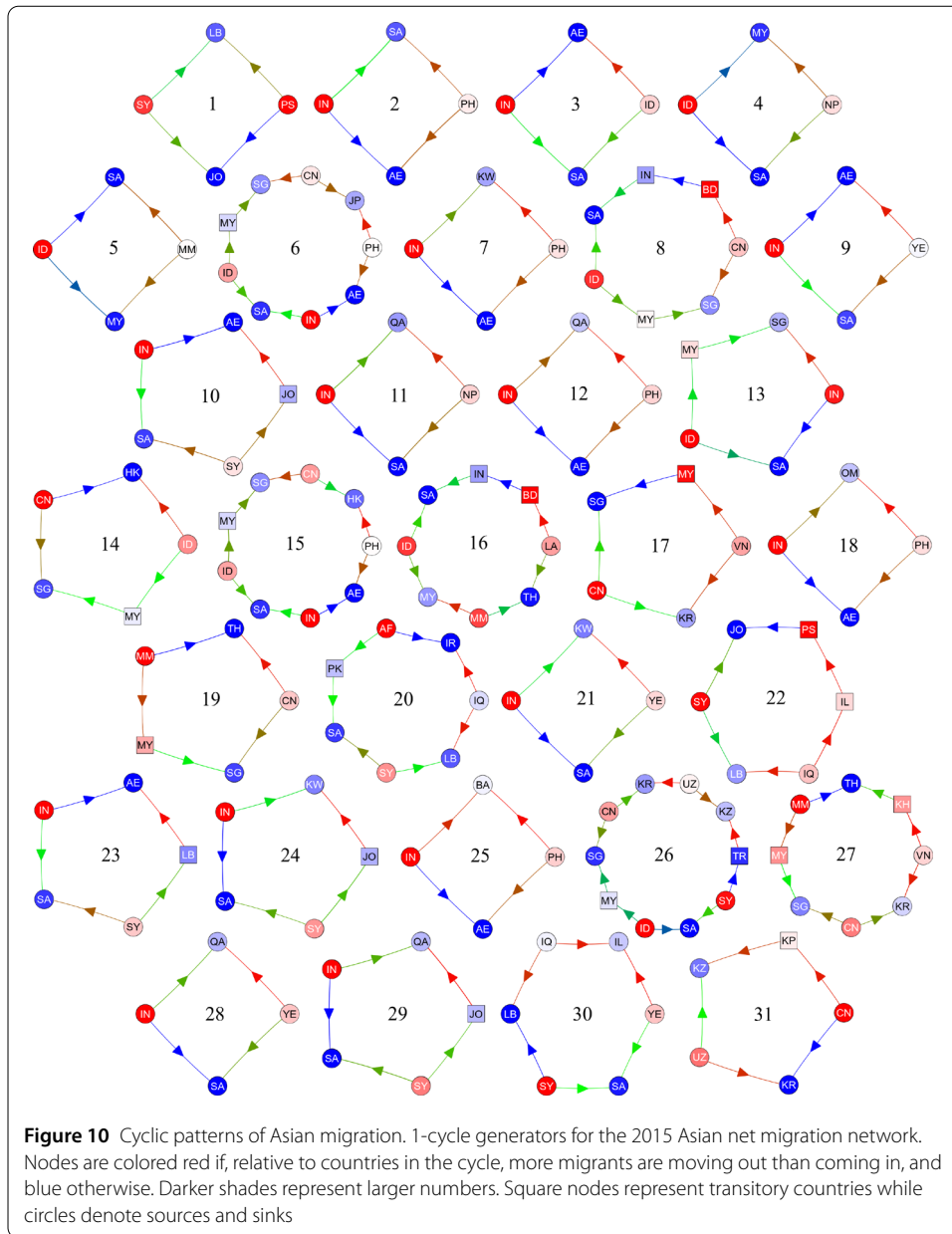


Similarly, while MY tends to be a transitory country in most cycles, it does become a sink in several instances.

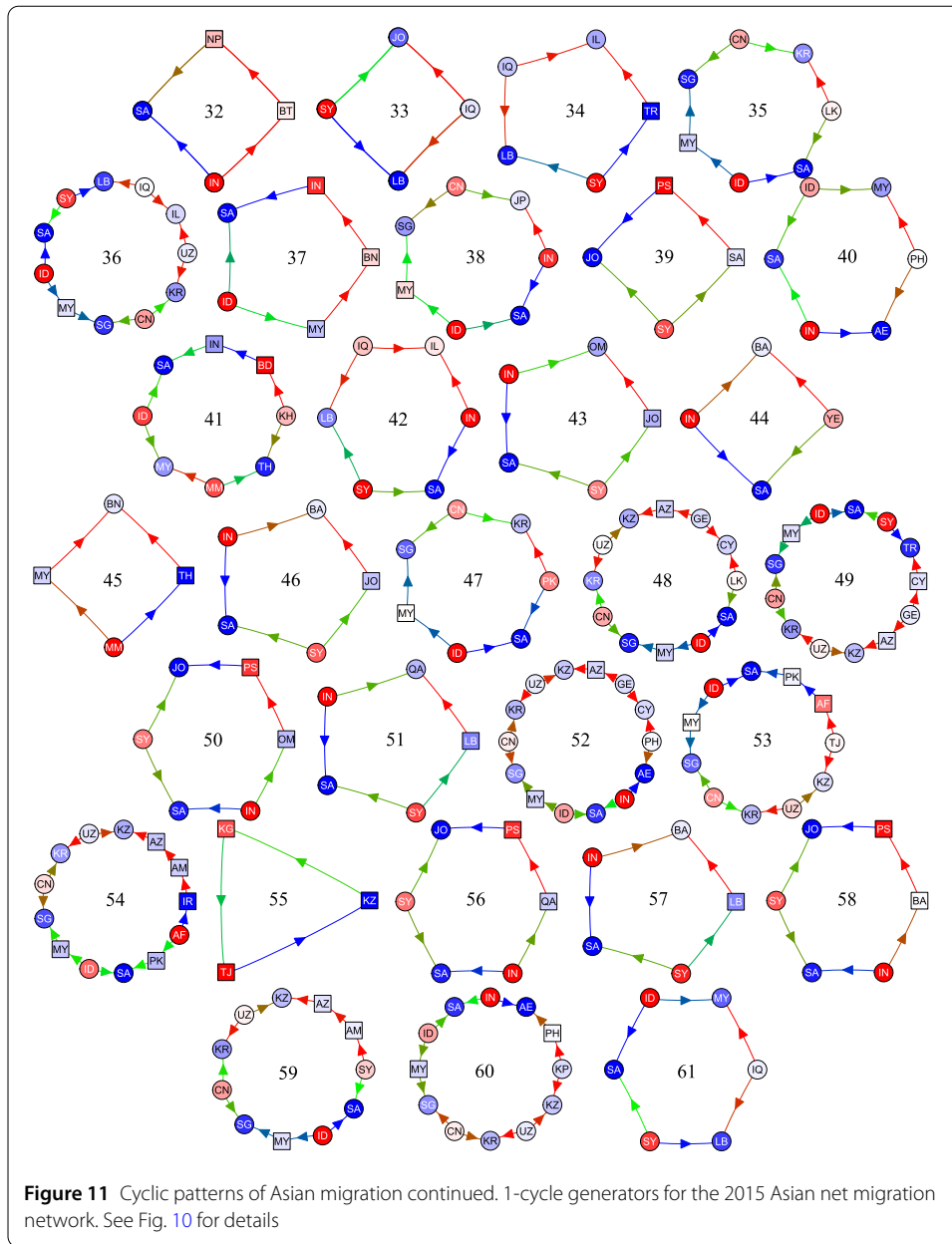
The 2-cycles illuminate more complex migrant and remittance movements and the effect of a unified 3-country flow in a local setting. Each face (directed 2-clique depicting a 3-country unit) in a 2-cycle is colored according to the threshold at which the 2-simplex is born, and graded towards the sink of this directed clique, representing the lower bound for the flow from the source to the sink. Both coloring schemes for the 1- and 2-cycles follow the appropriate color spectrum in Fig. 1. It is worth noting that regional circuits tend have the same sinks (Figs. 12.M4-M7 for the migrant network), and sources (Figs. 14.R4-R5, R9-R12, R13-R17, Figs. 14.R23–15.R24, and Figs. 14.R29-R32 for the remittance network) foreshadowing the position of these countries in a regional level.

The two most persistent 1-cycles are both of Type 2 and appear in Figs. 16(a) and 16(b). The coloring of the corresponding bars in the 1-dimensional barcode (middle of Fig. 9(a)) reveals medium to high variability in the net migrant flow from the source countries PS and SY (Fig. 16(a)), and IN and PH (Fig. 16(b)) to respective sink countries JO and LB, and AE and SA. In particular, the extremely high cycle weights variability for the most persistent 1-cycle (Fig. 16(b)) reflect the significantly higher flow of migrants to AE and SA from IN than from PH. It is worth noting that while PS and SY, and IN and PH play the role of sources in these persistent cycles, none of them are in fact true sources when viewed as nodes in the entire directed network. This shows that homology can find these important cycles illustrating persistent complex flow patterns which would not have been found via standard graph theory techniques. The same can be said of the sinks JO and LB, and AE and SA in these persistent Type 2 1-cycles.

Of the 61 distinct 1-cycle generators throughout the Asian net migration network evolution, only 1 is a circuit (Fig. 11.55) detecting the circular flow of migrant surplus along

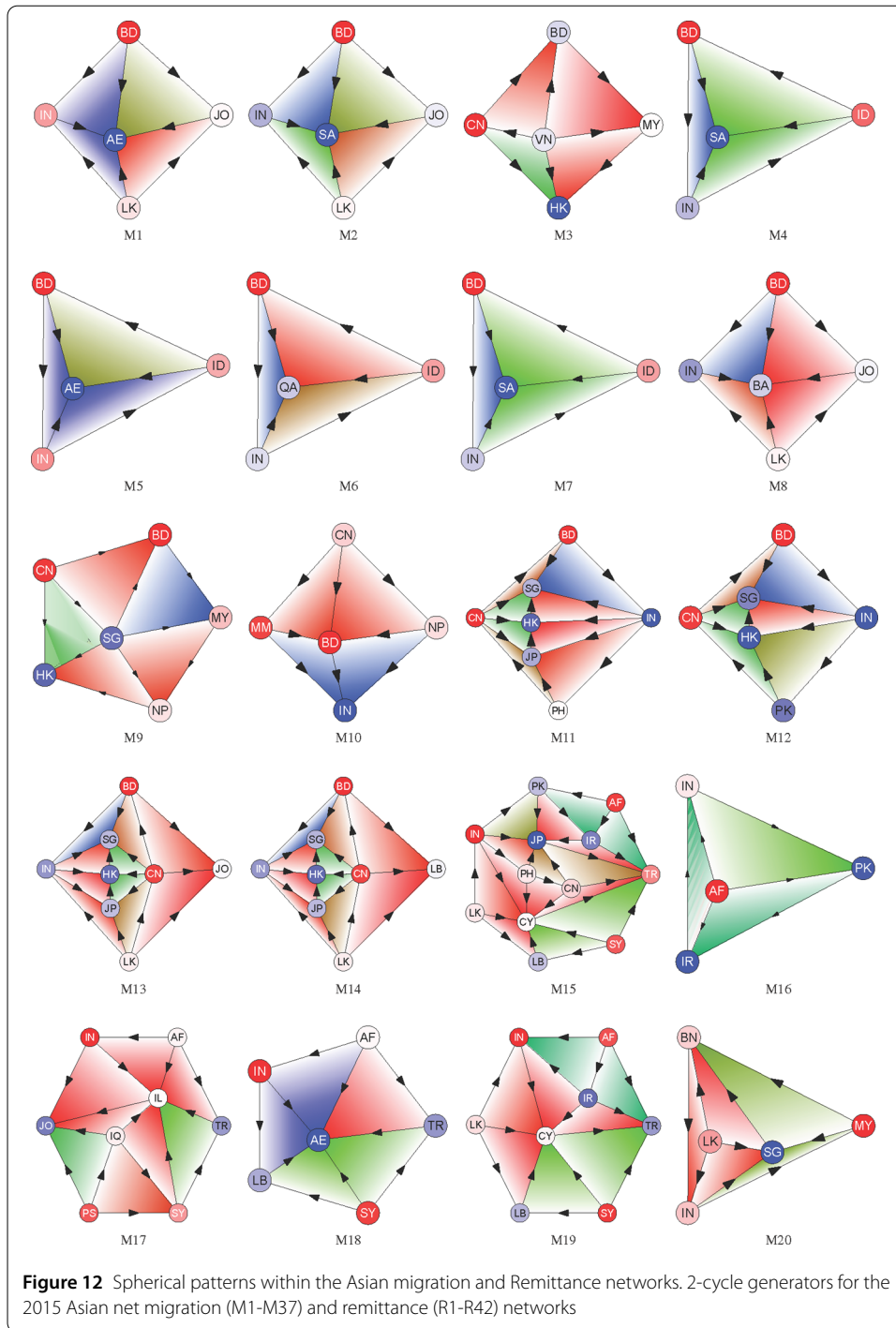


Kazakhstan → Kyrgyztan → Tajikistan → Kazakhstan. Within this circle of migrant movement, Kazakhstan receives a net migrant flow (inflow minus outflow) of 8746 people, while Kyrgyztan and Tajikistan both lose migrants by 3256 people and 5486 people respectively. In addition, only four 1-cycles of Type 1 are detected, having respective sources and sinks IN and SA (Fig. 11.32), ID and SA (Fig. 11.37), SY and JO (Fig. 11.39), and MM and BN (Fig. 11.45). The rest of the detected 1-cycles are of Type 2 having developing states as sources and high income states as sinks. Moreover, while algebraically trivial circuits and Type 1 1-cycles abound as meridians in many 2-cycles, the grading shows that migrant flows are directed towards an external (high income) sink node. In addition, all 1-cycles are born late in the filtration, and the mean standard deviation of 1-cycle edge weights is very high at 81,6704 people. All these suggest a familiar scene: that migrant movement within Asia is characterized by highly unbalanced exodus of migrants from developing to



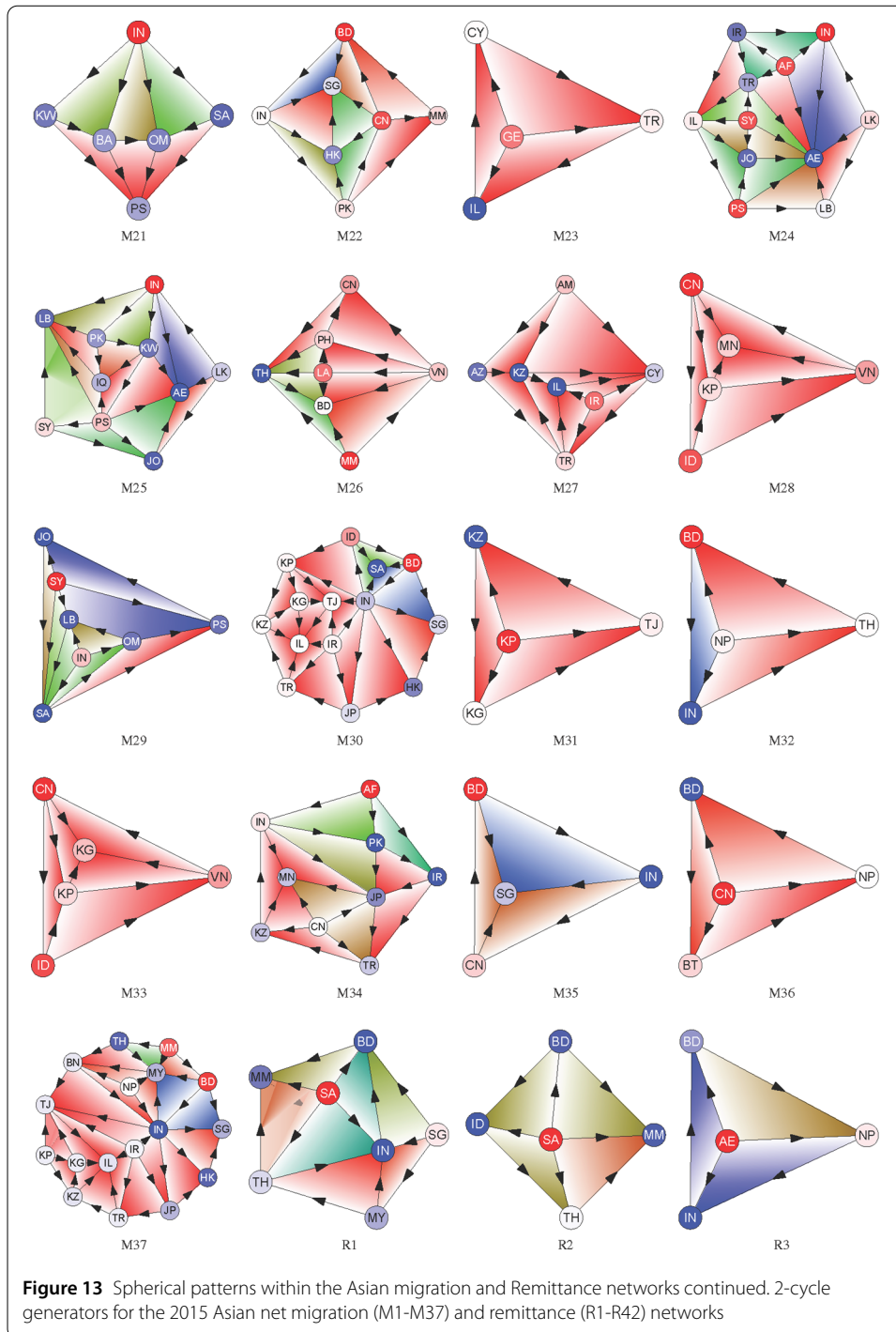
high income states. That all bars in the 1-dimensional barcode die at the end of the filtration reflects that there is in fact very small migration flow across states within the 1-cycles other than the flow described by the 1-cycle itself.

There are two perpetual 2-cycles in the Asian net migration network, shown in Figs. 16(c) and 16(d). These two spheres reveal the complex flow dynamics among the countries included in the 2-cycles. In particular, Fig. 16(c) shows that in this group of countries, AF and CN are sources of migrant flows while IL and MN are attractors. On the other hand, Fig. 16(d) reveals that there is movement of migrants from CN to LK via a transitory circuit that runs along $BD \rightarrow BT \rightarrow NP \rightarrow BD$. As there is also one perpetual connected component, and no perpetual 1-cycle, we see that the overall topology of



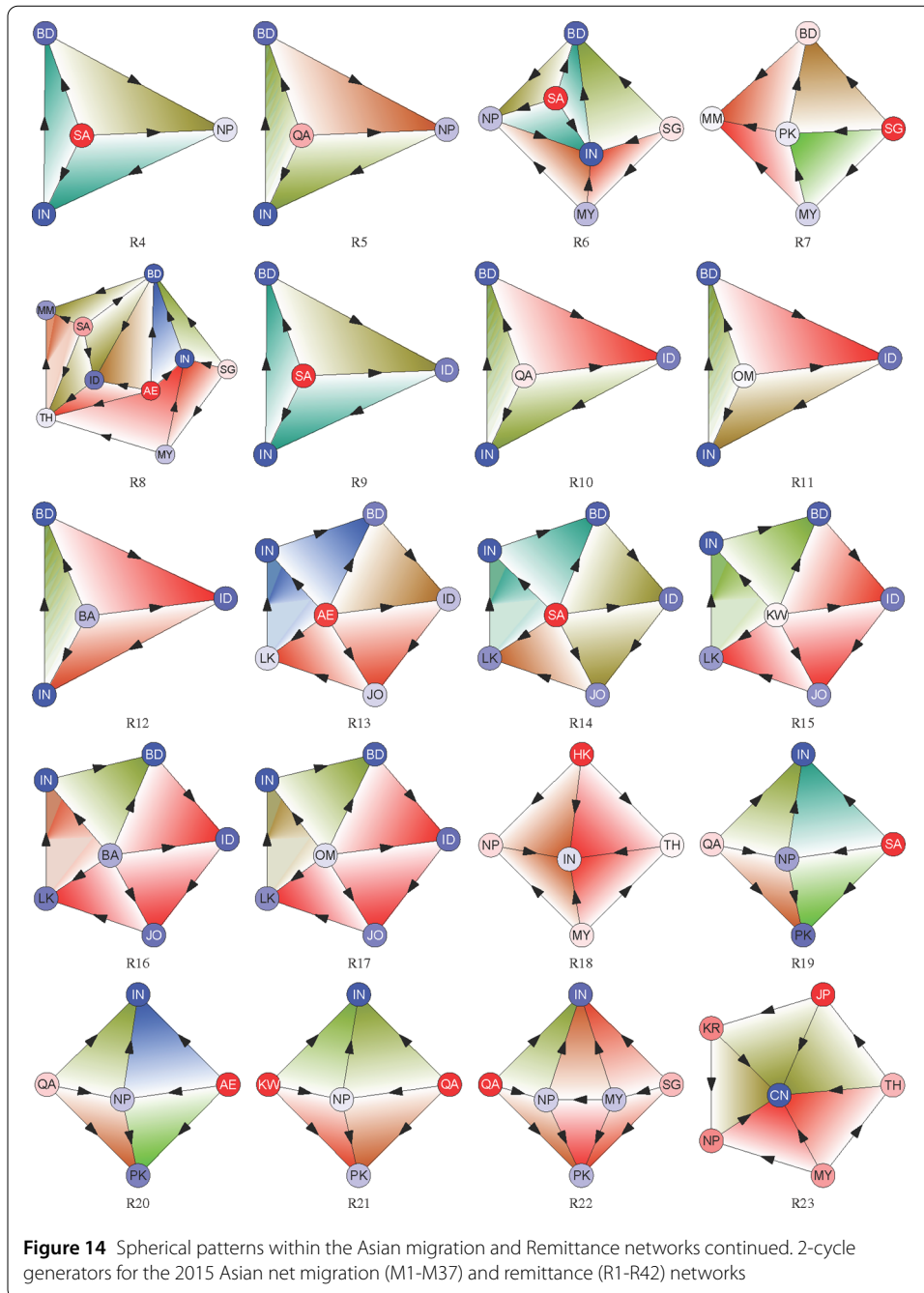
the entire Asian migration network is characterized by two spheres (Fig. 16(c) and 16(d)) joined at a source node (CN).

For the Asian net remittance network, the distinct 1- and 2-cycle generators are shown in Figs. 17–18 and Figs. 13–15 respectively. The two most persistent 1-cycles are also both of Type 2 as shown in Fig. 19. The most persistent 1-cycle (Fig. 19(a)) being born significantly earlier than the rest with relatively low cycle standard deviation suggests that the



flows of net remittances among the countries in the cycle are relatively uniform and are all significantly larger than other flows in the network.

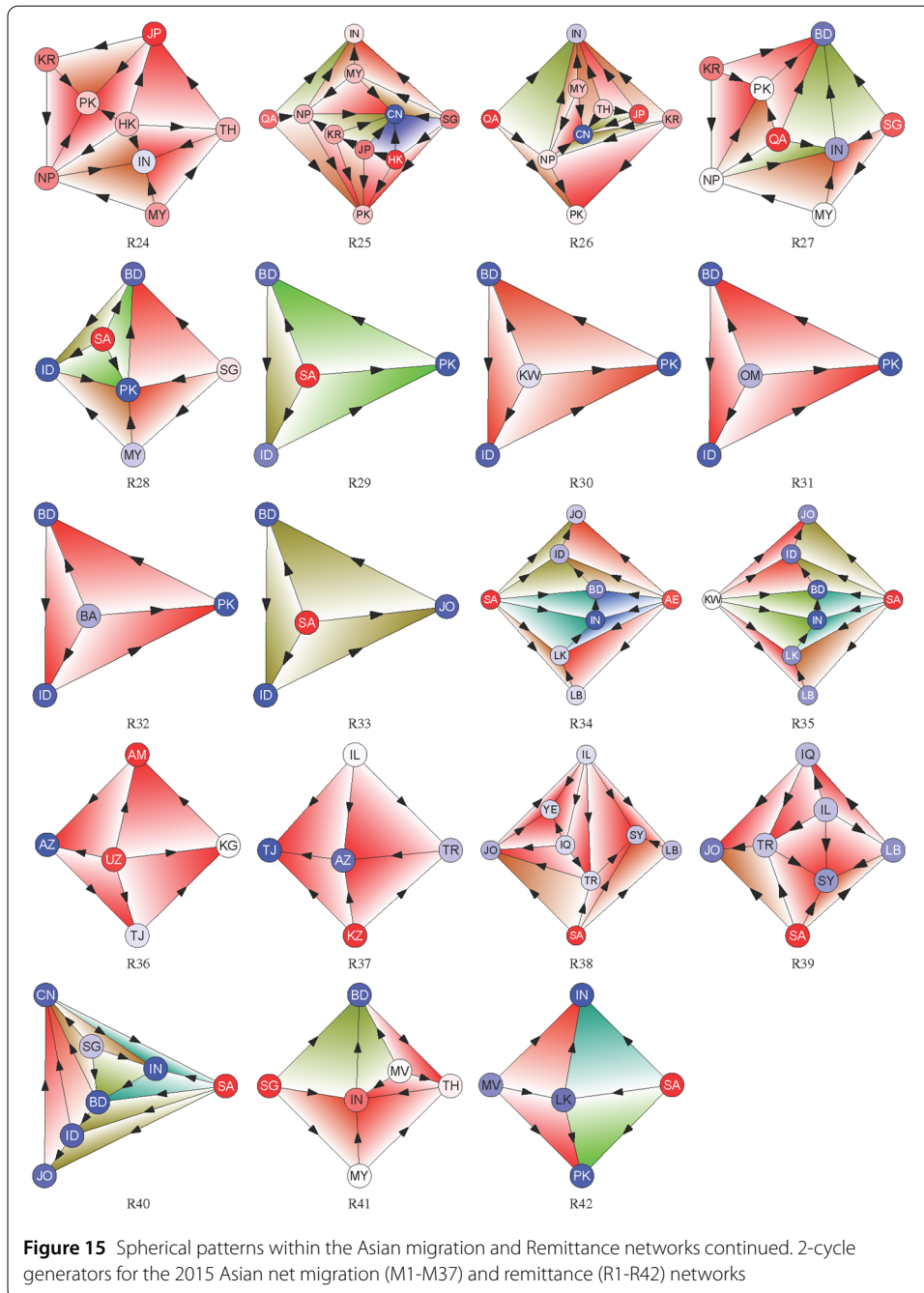
There is only one 1-cycle generator of Type 1 (cycle 59 in Fig. 18) having source SA and sink JO, and the rest are of Type 2 with sources mostly high income states, and sinks developing states. Similar observations as those in the migration network suggest that highly unbalanced flows of remittances from high income to developing countries abound in Asia. An interesting observation is that the second most persistent 1-cycle (Fig. 19(b))



in the net remittance barcode is a copy of the most persistent 1-cycle for the net migration network with all arrows reversed.

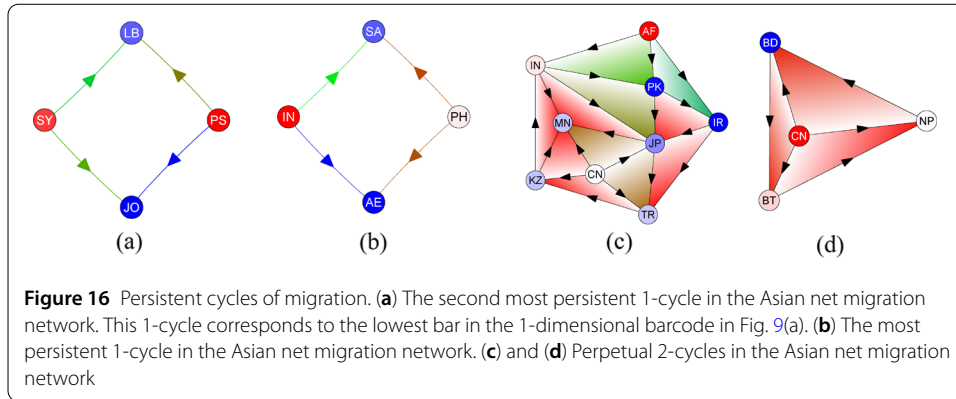
5 Summary and conclusion

In this note, we showed how persistent directed clique homology can be used to extract embedded patterns in weighted directed networks that reveal complex and simultaneous interactions among multiple vertices. Detection of these patterns affords a deeper probe into the local and global topological features of directed network models. By adopting Palla et al.'s [21] directed clique construction, we applied persistent homology to the associated



directed clique complex of a weighted directed network using a filtration technique similar to the weight rank clique filtration introduced by Petri et al. [17]. The definition of directed cliques allowed us to simplify network flow structure by adopting an overall unidirectional flow in vertex communities having complete local connections.

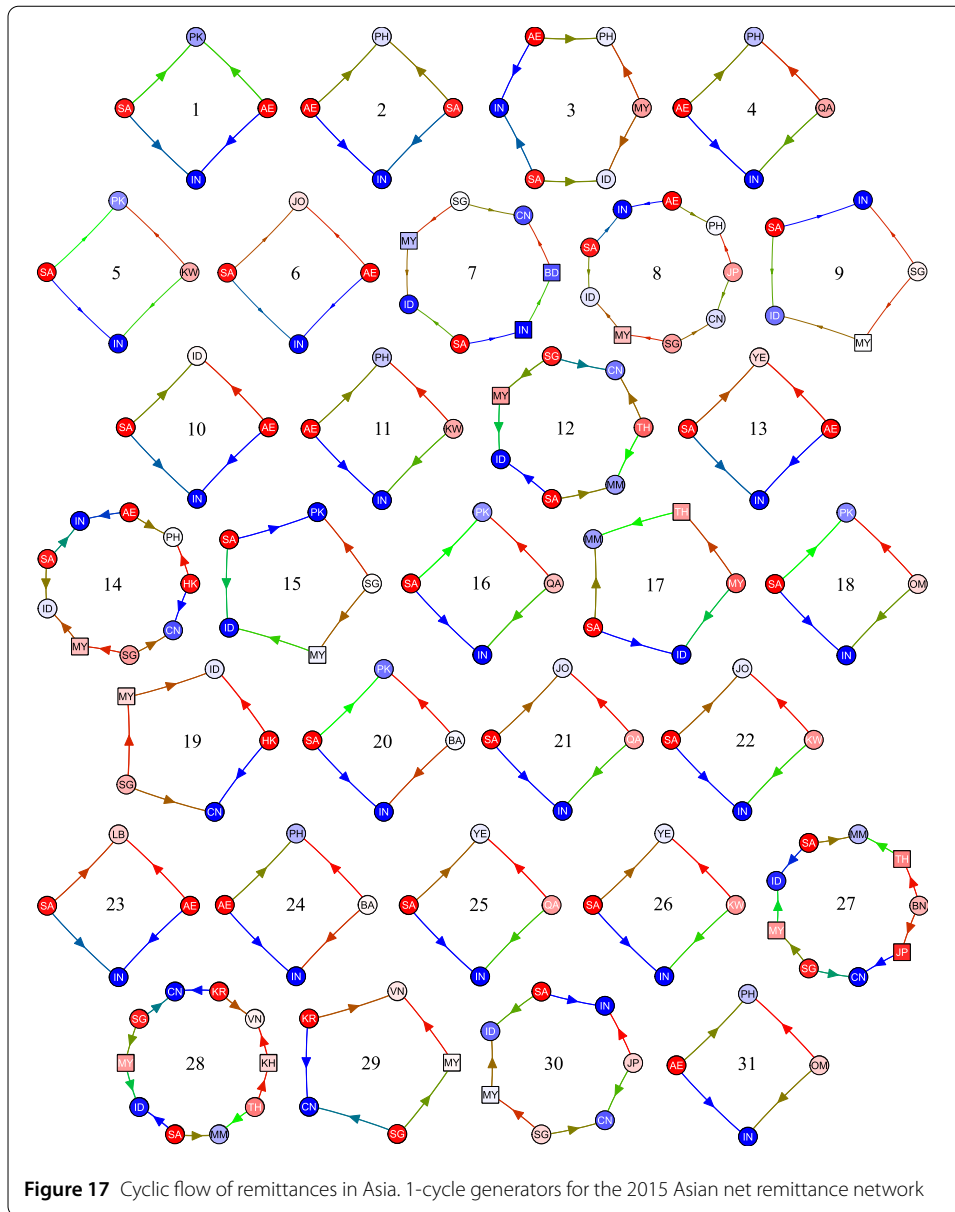
We used tools from algebraic topology to extract topological features from the simplicial model in order to detect specific flow patterns woven in the network fabric. The goal of homology is to understand the shape of a graph by identifying cycles. For 0-dimensional homology, every vertex is a cycle. By modding out by boundaries, we recover weakly connected components. At the 1-dimensional level, we detect circuits as well as Type 1 cycles



(with unique sink and unique source) and Type 2 cycles (with multiple sinks and sources). Note that we observed many more Type 2 cycles than both circuits and Type 1 cycles in both the migrant and remittance networks, illustrating more complex flow pattern than what is generally found via traditional graph analysis. We also recovered higher dimensional topological features such as sphere-like voids, illuminating the complex migrant and remittance flow dynamics within groups of countries in a regional setting. The advantage of persistent homology over homology at a fixed threshold is that one can analyze how the simplicial model changes as the threshold varies. Other contributions of this paper include:

- Classifying 1-cycles as circuits (no sink nor source), Type 1 (with unique sink and unique source), or Type 2 (with multiple sinks and sources). These different types of 1-cycles encode different flow patterns. The definition of the directed clique also allows for the detection of the circuit with three vertices—a feature not commonly detected in other TDA methods.
- Characterizing topological features of directed network by *perpetuity*. While persistent homology on undirected networks tracks long-lived topological features along a filtration, persistent directed clique homology in addition detects topological features of the entire directed network by tracking cycle generators that never get killed off.
- Distinguishing topological features of the same dimension by introducing coloring schemes for the barcodes. For the 0-dimensional barcode, coloring the bars according to eventual connected component membership encodes the clustering of the vertices at the end of the filtration similar to the final clustering output of single linkage hierarchical clustering. For dimensions of at least 1, coloring the bars according to variation in the edge weights within the cycles captures information on similarity or disparity among internal flows in the cycle. It is worth noting that persistence is agnostic of such characteristic in detected cycles.

We obtained the colored barcodes after computing persistent homology from the directed clique complexes of the Asian net migration and remittance networks. For the Asian net migration network, connectivity is characterized by a single perpetual component with most linkages forming at the end of the filtration. An analysis of the sixty-one detected 1-cycles shows that most are of Type 2, including the two that are most persistent. This reflects the abundance of highly unbalanced movements of migrants from groups of low income countries to high income states. Moreover, thirty-seven 2-cycles are detected,



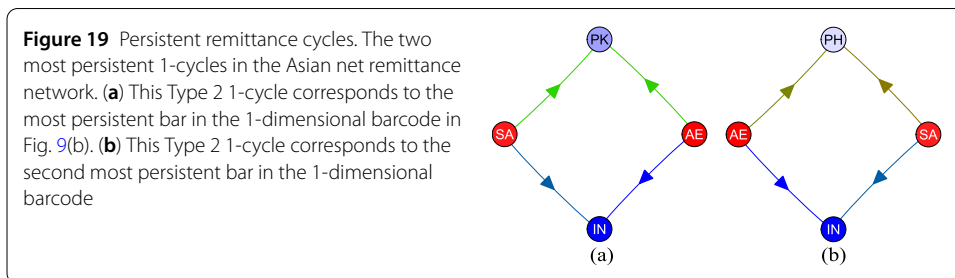
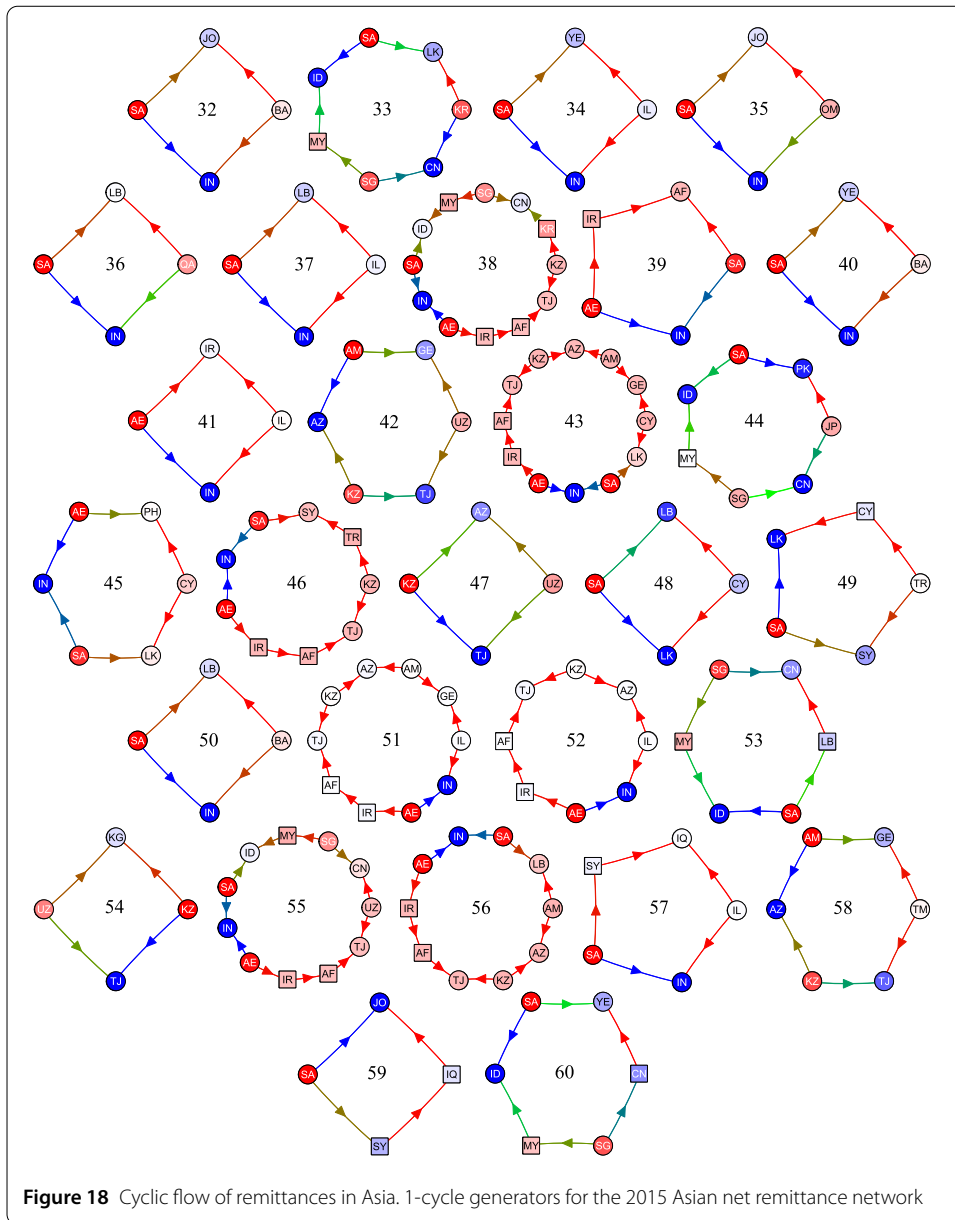
all of which are spheres and two are perpetual, encoding the complex migrant flow structure among multiple countries in Asia. On the other hand, sixty 1-cycles and forty-two 2-cycles are detected in the net remittance network. Most 1-cycles are also of Type 2, two of which have significantly larger persistence, and all 2-cycles are spheres. We generate figures for all detected 1-cycles (Figs. 10–11 and 17–18) and 2-cycles (Figs. 12–15) and present them in the text.

Funding

Paul Samuel Ignacio was supported by the University of Iowa Graduate College post-comprehensive research award for the duration of the study.

Abbreviations

UN, United Nations; UNDESA, United Nations’ Department of Economics and Social Affairs; USD, U.S. Dollars. Country abbreviations with their corresponding code numbers for this work are provided in Table 1.



Availability of data and materials

All data sets used in this study are available from the corresponding databases mentioned in the text. The net migration and remittance matrix is available upon request.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The main idea of this paper was proposed by PSI. PSI prepared the manuscript initially, wrote all code, and prepared all figures. ID contributed ideas and to the writing of the manuscript. Both authors thoroughly read and approved the final manuscript.

Endnotes

- ^a The indegree (resp. outdegree) of a vertex is the number of its incoming (outgoing) directed edges.
- ^b The regions identified by UNDESA are Africa, Asia, Europe, Latin America and the Caribbean, North America, and Oceania.
- ^c A source (respectively sink) is a vertex with all incident edges pointing outward (respectively inward).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 6 February 2018 Accepted: 20 December 2018 Published online: 05 January 2019

References

1. United Nations Department of Economic & Social Affairs PD (2016) International Migration Report 2015: Highlights (ST/ESA/SER.A/375)
2. Ratha D, De S, Plaza S, Schuettler K, Shaw W, Wyss H, Yi S (2016) Migration and Remittances—Recent Developments and Outlook. *Migr. Dev. Brief* 26. <https://doi.org/10.1596/978-1-4648-0913-2>
3. Fagiolo G, Mastrolillo M (2013) International migration network: topology and modeling 88:012812
4. Aleskerov F, Meshcheryakova N, Rezyapova A, Shvydun S (2017) Network analysis of international migration. In: Kalyagin VA, Nikolaev AI, Pardalos PM, Prokopyev OA (eds) Springer, Cham, pp 177–185. <https://doi.org/10.1007/978-3-319-56829-4-13>
5. Tranos E, Gheasi M, Nijkamp P (2015) International migration: a global complex network. *Environ Plan B, Plan Des* 42(1):4–22. <https://doi.org/10.1068/b39042>
6. Frosini P (1992) Discrete computation of size functions. *J Comb Inf Syst Sci* 17(3–4):232–250
7. Verri A, Uras C, Frosini P, Ferri M (1993) On the use of size functions for shape analysis. *Biol Cybern* 70:99–107
8. Verri A, Uras C (1997) Metric-topological approach to shape representation and recognition. *Image Vis Comput* 14:189–207
9. Verri A, Uras C (1997) Computing size functions from edge maps. *Int J Comput Vis* 23(2):169–183
10. Edelsbrunner H, Letscher D, Zomorodian A (2002) Topological persistence and simplification. *Discrete Comput Geom*. <https://doi.org/10.1007/s00454-002-2885-2>
11. Biasotti S, De Floriani L, Falcidieno B, Frosini P, Giorgi D, Landi C, Papaleo L, Spagnuolo M (2008) Describing shapes by geometrical-topological properties of real functions. *ACM Comput Surv* 40(4):12:1–12:87
12. Edelsbrunner H, Morozov D (2012) Persistent homology: theory and practice. European Congress of Mathematics, Krakow, 2–7 July, 2012, Europ Math Soc, 31–50
13. Ghrist R (2008) Barcodes: the persistent topology of data. *Bull Am Math Soc (NS)* 45(1):61–75
14. Carstens CJ, Horadam KJ (2013) Persistent homology of collaboration networks. *Mathematical problems in engineering*. <https://doi.org/10.1155/2013/815035>
15. Patania A, Petri G, Vaccarino F (2017) The shape of collaborations. *EPJ Data Sci* 6(1):18. <https://doi.org/10.1140/epjds/s13688-017-0114-8>
16. Petri G, Scolamiero M, Donato I, Vaccarino F (2013) Networks and cycles: A persistent homology approach to complex networks. In: Gilbert T, Kirkilionis M, Nicolis G (eds) Springer, Cham, pp 93–99. <https://doi.org/10.1007/978-3-319-00395-5-15>
17. Petri G, Scolamiero M, Donato I, Vaccarino F (2013) Topological strata of weighted complex networks. *PLoS ONE* 8(6):1–8. <https://doi.org/10.1371/journal.pone.0066506>
18. Chowdhury S, Mémoli F (2016) Persistent homology of directed networks. In: The 50th asilomar conference on signals, systems, and computers IEEE
19. Chowdhury S, Mémoli F (2017) Persistent Path Homology of Directed Networks. arXiv preprint. [arXiv:1701.00565](https://arxiv.org/abs/1701.00565)
20. Horak D, Maletić S, Milan R (2009) Persistent Homology of Complex Networks. *J Stat Mech*
21. Palla G (2007) Directed Network Modules. *New J Phys* 9. <https://doi.org/10.1088/1367-2630/9/6/186>
22. Reimann MW, Nolte M, Scolamiero M, Turner K, Perin R, Chindemi G, Dlotko P, Levi R, Hess K, Markram H (2017) Cliques of neurons bound into cavities provide a missing link between structure and function. *Front Comput Neurosci* 11:48. <https://doi.org/10.3389/fncom.2017.00048>
23. United Nations Department of Economic & Social Affairs PD Trends in International Migrant Stock: migrants by Destination and Origin (United Nations database, POP/DB/MIG/Stock/Rev.2015)
24. The World Bank Migration and Remittances Data. <http://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data>
25. United Nations Department of Economic & Social Affairs PD (2015) Trends in International Migrant Stock: the 2015 Revision. (United Nations database, POP/DB/MIG/Stock/Rev.2015)
26. Ratha DK, Shaw W (2007) South-South Migration and Remittances. Development Prospects Group, World Bank
27. Masulli P, Villa AEP (2016) The topology of the directed clique complex as a network invariant. *SpringerPlus* 5(1):388. <https://doi.org/10.1186/s40064-016-2022-y>

28. Grigor'yan A, Lin Y, Muranov Y, Yau S-T (2012) Homologies of Path Complexes and Digraphs. arXiv preprint. [arXiv:1207.2834](https://arxiv.org/abs/1207.2834)
29. Grigor'yan A, Lin Y, Muranov Y, Yau S-T (2014) Homotopy Theory for Digraphs. arXiv preprint. [arXiv:1407.0234](https://arxiv.org/abs/1407.0234)
30. Grigor'yan A, Muranov Y, Yau S-T (2015) Homologies of digraphs and the Künneth formula.
31. Turner K (2016) Rips filtrations for quasi-metric spaces and asymmetric functions with stability results. arXiv preprint. [arXiv:1608.00365](https://arxiv.org/abs/1608.00365)
32. Mischaikow K, Nanda V (2013) Morse theory for filtrations and efficient computation of persistent homology. *Discrete Comput Geom* 50(2):330–353. <https://doi.org/10.1007/s00454-013-9529-6>
33. Zomorodian A, Carlsson G (2005) Computing persistent homology. *Discrete Comput Geom* 33(2):249–274. <https://doi.org/10.1007/s00454-004-1146-y>
34. Binchi J, Merelli E, Rucco M, Petri G, Vaccarino F (2014) jHoles: A tool for understanding biological complex networks via clique weight rank persistent homology. *Electron Notes Theor Comput Sci* 306:5–18
35. Edelsbrunner H, Harer J (2010) *Computational Topology: an Introduction*. ISBN 978-0-8218-4925-5
36. Bubenik P (2015) Statistical topological data analysis using persistence landscapes. *J Mach Learn Res* 16(1):77–102.
37. Chazal F, Fasy BT, Lecci F, Rinaldo A, Wasserman L (2014) Stochastic convergence of persistence landscapes and silhouettes. In: *Proceedings of the thirtieth annual symposium on computational geometry. SOCG'14*. ACM, New York, pp 474:474–474:483. <https://doi.org/10.1145/2582112.2582128>
38. Petri G, Expert P, Turkheimer F, Carhart-Harris R, Nutt D, Hellyer P, Vaccarino F (2014) Homological scaffolds of brain functional networks. *J R Soc Interface* 11:20140873. <https://doi.org/10.1098/rsif.2014.0873>
39. Piangerelli M, Rucco M, Tesei L, Merelli E (2018) Topological classifier for detecting the emergence of epileptic seizures. *BMC Research Notes* 11. <https://doi.org/10.1186/s13104-018-3482-7>
40. Merelli E, Piangerelli M, Rucco M, Toller D (2016) A topological approach for multivariate time series characterization: the epileptic brain. In: *Proceedings of the 9th EAI international conference on bio-inspired information and communications technologies (formerly BIONETICS). BICT'15. ICST (institute for computer sciences, social-informatics and telecommunications engineering)*. ICST, Brussels, pp 201–204. <https://doi.org/10.4108/eai.3-12-2015.2262525>

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
