Ricci curvature of random and empirical directed hypernetworks

Check for updates

Wilmer Leal[1,2] ![ORCID], Marzieh Eidi[2] ![ORCID]* and Jürgen Jost[2,3]

---

*Correspondence:
Marzieh.Eidi@mis.mpg.de
[2]Max Planck Insitute for
Mathematics in the Sciences,
Inselstr. 22, 04103 Leipzig, Germany
Full list of author information is
available at the end of the article

Abstract

Relationships in real systems are often not binary, but of a higher order, and therefore cannot be faithfully modelled by graphs, but rather need hypergraphs. In this work, we systematically develop formal tools for analyzing the geometry and the dynamics of hypergraphs. In particular, we show that Ricci curvature concepts, inspired by the corresponding notions of Forman and Ollivier for graphs, are powerful tools for probing the local geometry of hypergraphs. In fact, these two curvature concepts complement each other in the identification of specific connectivity motifs. In order to have a baseline model with which we can compare empirical data, we introduce a random model to generate directed hypergraphs and study properties such as degree of nodes and edge curvature, using numerical simulations. We can then see how our notions of curvature can be used to identify connectivity patterns in the metabolic network of *E. coli* that clearly deviate from those of our random model. Specifically, by applying hypergraph shuffling to this metabolic network we show that the changes in the wiring of a hypergraph can be detected by Forman Ricci and Ollivier Ricci curvatures.

Keywords: Directed hypergraphs, Discrete curvature, Ricci curvature, Forman-Ricci curvature, Ollivier-Ricci curvature, Random models of directed hypergraphs, Metabolic networks

Introduction

Network analysis has placed special emphasis on properties of nodes. Since networks, represented by graphs, are widely used to model discrete systems whose structure is given by relationships among objects, we shall develop tools that allow a complementary analysis of networks focused on properties of edges.

Undirected graphs are only the simplest type of model for relations between discrete entities. Many empirical data expressing such relations have more structure than that of an undirected graph (Spivak 2009). For instance, the relations could be directed and/or weighted. Moreover, a relation could also involve more than two entities, as for instance in coauthorship networks or chemical reactions. Such relations can be modelled by hypergraphs rather than graphs. The hypergraph of coauthorships is undirected, whereas that of chemical reactions is directed, because a reaction may proceed from educts to

products, but not in the reverse direction. We therefore seek quantities that can be evaluated not only for edges of graphs, but also for hyperedges of (possibly weighted and/or directed) hypergraphs.

Recently, notions inspired by Riemannian geometry appeared very promising in this direction. More precisely, it has been discovered that concepts of curvature can be formulated in such a way that they apply naturally not only to smooth Riemannian manifolds, but also to various kinds of discrete spaces (Forman 2003; Saucan 2019), like graphs (Jost and Liu 2014; Ollivier 2007) or hypergraphs (Asoodeh et al. 2018; Banerjee 2020). Much effort has focused on concepts of Ricci curvature in this context, and that is also what we shall explore in this paper, drawing on recent theoretical work from our group, like notions of such Ricci curvature for directed hypergraphs (Eidi and Jost 2020; Leal et al. 2018).

In this paper, after a short history of curvature notions, and in particular Ricci curvature, from the beginnings to some recent advances, we present the results of our generalizations of Forman-Ricci (Leal et al. 2018) and Ollivier-Ricci (Eidi and Jost 2020) curvature notions to directed hypergraphs. Then, we show that they are powerful tools for exploring local properties of directed hypergraph motifs. To conclude, we carry out a curvature-based analysis of the metabolic network of *E. coli*. This article is an invited extended version of the extended abstract published in the book of abstracts of the "The 8th International Conference on Complex Networks and their Applications" (Leal et al. 2019).

Ricci-curvature: from Riemannian geometry to discrete structures

Geometry is a branch of mathematics which is concerned with shapes, sizes, positions and distances. Classically, it is considered as the science of space, and of objects in space. A fundamental step was taken by Gauss who defined the notion of intrinsic curvature of a surface. This curvature can be determined independently of the embedding into Euclidean (Flat) space, that is, by taking measurements only on that surface itself and not from any position outside that surface in space. Inspired by Gauss' discovery, Riemann proposed the concept of a "many fold extended quantity" which today is simply called a "manifold". A manifold is a geometric object which locally looks like Euclidean space of some dimension. Thus, a two-dimensional manifold is locally modelled by the Euclidean plane. But this resemblance is only qualitative, but not quantitative. For instance, the surface of a sphere is locally geometrically different from that of the Euclidean plane, even though both the sphere and the plane are two-dimensional manifolds. This local deviation from being Euclidean is quantified by curvature. In two dimensions, curvature is a single number, but on a general n-dimensional Riemannian manifold, the curvature tensor has $\frac{n(n-1)}{2}$ components. There exist, however, certain averages or contractions of the components of the curvature tensor that already carry significant geometric information. Foremost among those is the Ricci curvature which averages the components of the curvature tensor corresponding to a tangent direction on the manifold.

In geometric terms, Ricci curvature compares the average distance between two close enough balls with the distance of their centres. The Ricci curvature at a point and along a vector (tangent to the manifold at that point) is positive (negative) if the distance between the balls is less (larger) than the distance between their centres. When the Ricci curvature is positive, the manifold is simple in the important sense that its topology is much

constrained, and also stochastic processes on such a manifold are well controlled. For instance, in such a positively curved space, random walkers can be coupled so that they will become closer and closer.

Inspired by these results, there have been different attempts to transfer this curvature notion to more general/non-smooth structures. Mathematicians have looked for a corresponding notion for Ricci curvature which is valid "at a certain scale" and that applies to smooth manifolds as well as to discrete structures such as graphs. For that purpose, one needed a "synthetic" notion for the curvature which does not depend on infinitesimal properties such as continuity and differentiability (Ollivier 2013). Different such notions of generalized Ricci curvature have been proposed which, while they coincide for smooth manifolds, yield rather different results for graphs or other discrete structures. Among the most successful ones are the notions of Ricci curvature proposed by Forman (2003) and Ollivier (2007). In recent years, they have been extensively used for detecting local structures of networks (see for instance Samal et al. 2018).

Forman Ricci curvature

In 2003, Forman defined his notion of Ricci curvature for cell complexes (Forman 2003). While Forman's definition applies to general CW-complexes, for our purposes it suffices to explain it for simplicial complexes. A simplicial complex Σ on a vertex set $V = \{v_1, \ldots, v_n\}$ consists of a collection of simplices, that is, subsets of V. When such a subset α contains $p + 1$-vertices, it is called a p-simplex, because we can think of this combinatorial object also as a p-dimensional geometric simplex. A 0-simplex is simply a vertex, a 1-simplex is also called an edge, and a 2-simplex a triangle. We write $\beta < \alpha$, and say that β is a boundary-simplex of α, when the $(p - 1)$-simplex β is a subset of the p-simplex α. For the definition of a simplicial complex, it is required that whenever the p-simplex α belongs to Σ, then so do all its boundary simplices $\beta < \alpha$. Thus, for instance, when our simplicial complex contains the triangle $\{u, v, w\}$ with vertices u, v, w, it has to contain also the three edges $\{u, v\}$, $\{v, w\}$, and $\{u, w\}$, as well as inductively also the 0-simplices $\{u\}$, $\{v\}$ and $\{w\}$. For the definition of a hypergraph, that is, a formal object of the type considered in this paper, this condition will not be required. Thus, hypergraphs are more general objects than simplicial complexes, and correspondingly more difficult to treat mathematically, but this generality will be needed to adequately model metabolic networks, for instance.

Returning to simplicial complexes for the moment, Forman defines functions

$$F_p : \{p - \text{simplices}\} \longrightarrow \mathbb{R} \tag{1}$$

by putting, for a p-simplex α,

$$F_p(\alpha) := \sharp\{(p+1)\text{-simplices } \beta > \alpha\} + \sharp\{(p-1)\text{-simplices } \gamma < \alpha\} - \sharp\{\text{parallel neighbors of } \alpha\}, \tag{2}$$

where, as mentioned, $\beta > \alpha$ means that the p-simplex α is contained in the boundary of the $(p + 1)$-simplex β, and analogously, $\gamma < \alpha$ means that the $(p - 1)$-simplex γ is contained in the boundary of α. And a parallel neighbor of the p-simplex α is another simplex α' of the same dimension that is disjoint from α, but either contained in the boundary of some $(p + 1)$-simplex that also contains α in its boundary or containt some $(p - 1)$-simplex in its boundary that is also contained in the boundary of α, but not both.

(As mentioned, the definition applies more generally to cell complexes, but that is not needed for our present purposes.)

The concept is perhaps most easily understood when we consider graphs. In fact, an undirected graph is a simplicial complex that has only 0-simplices (the vertices or nodes) and 1-simplices (the edges). The Forman curvature of an edge e with nodes i, j is simply given by $F(e) = 4 - \deg(i) - \deg(j)$. Here, the edge is not contained in any 2-simplex, because there are none in a graph, and hence the first term in (2) is 0. It has 2 0-simplices contained in it, its vertices i, j, and hence the second term is 2. Finally, it has $\deg(i) + \deg(j) - 2$ parallel neighbors, the other edges emanating from the vertices i and j.

Edges connecting nodes with large degree have very negative Forman-Ricci curvature values, and these are typically edges that play a key role in the cohesion of a network. The Ricci curvature of a graph therefore can extract important information about the global structure of the graph from local quantities.

Ollivier Ricci curvature

Ollivier took a different approach to generalize Ricci curvature notions (Ollivier 2007). He compared the distance between two close balls with the distance between their centres in general metric measure spaces. Equivalently, he considered how the distance between two coupled random walkers evolves in time. A ball is given by a distribution of mass, and a random walker also follows a probability distribution. Two such distributions are optimally coupled when their Wasserstein distance is minimized. In formal terms, this works as follows. Let (X, d) be a discrete metric space with a random walk m. Let m_x be the local measure at the point $x \in X$. (All masses here and in the sequel are unit masses, that is, $m_x(X) = 1$.) That is, a random walker starting at x selects his position at the next step randomly according to m_x. Let $y \in X$ be another point. The Ricci curvature of (X, d, m) in the direction (x, y) then is

$$O(x, y) := 1 - \frac{W_1\left(m_x, m_y\right)}{d(x, y)} \tag{3}$$

where W_1 is the 1-Wasserstein distance between m_x and m_y on X,

$$W_1\left(m_x, m_y\right) := \inf_{\mathcal{E} \in \Pi\left(m_x, m_y\right)} \sum_{(u,v) \in X \times X} d(u, v) \mathcal{E}(u, v) \tag{4}$$

and $\Pi(m_x, m_y)$ is the set of measures on $X \times X$ that project to m_x and m_y, resp. Such a measure can be seen as the coupling of the two random walkers starting at x and y, resp.. That means that when the first walker moves to u, the second one moves to v with probability $\mathcal{E}(u, v)$ where u is distributed according to m_x and v according to m_y. The optimal coupling minimizes the average distance in (4).

On an unweighted and undirected graph, the metric distance between two vertices equals the minimal number of edges needed to reach one from the other. The move of a random walker who finds herself at a vertex reaches any of its direct neighbors with equal probability. This is the simplest case, and in the following sections we present our generalizations of these two curvature notions for directed hypergraphs.

The curvature of directed hypergraphs

Forman-Ricci curvature of directed hypergraphs

Here, we recall (Leal et al. 2018) where we have developed Forman-Ricci Curvature for directed hypergraphs. Formally, a directed hypergraph is a couple $H = (V, E)$ where V is a set of vertices and E a set of ordered pairs of subsets of V called hyperedges. Moreover, given a hyperedge $e = (e_i, e_j) \in E$, where $\varnothing \neq e_i \subset V$ and $\varnothing \neq e_j \subset V$ are called the tail and the head of e, respectively. We define the Forman-Ricci curvature of e as

$$F(e) = |e_i| + |e_j| - \sum_{i \in e_i} \text{in-deg}(i) - \sum_{j \in e_j} \text{out-deg}(j) \tag{5}$$

where the in(out)-degree of a vertex i is the number of hyperedges that have i as part of its head (tail).

In fact, in (Leal et al. 2018), different types of Forman Ricci curvature of a directed hyperedge have been introduced. Here we use only the version in (5) to be able to compare it with the other notion of Ricci curvature, namely Ollivier, for a given hyperedge.

Ollivier-Ricci curvature of directed hypergraphs

In (Eidi and Jost 2020) we introduced this curvature notion for directed hypergraphs by using the Wasserstein distance between two probability measures associated to a directed hyperedge. We say that $u \rightarrow e_i$ if there exists a hyperedge $e = (e_k, e_i)$ such that $u \in e_k$. Similarly, $e_j \rightarrow v$ if there exists a hyperedge $e = (e_j, e_k)$ such that $v \in e_k$. Given a hyperedge $e = (e_i, e_j)$, we define two sets, the *masses* $\mathcal{M} = \{u : u \rightarrow e_i\}$ and the *holes* $\mathcal{H} = \{v : e_j \rightarrow v\}$. We assume that we have a probability measure for each of them, denoted by $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{H}}$. The (directed) distance $d(u, v)$ between a mass u and a hole v of a hyperedge e, defined as the minimal number of directed hyperedges connecting them, is at most 3. It is precisely 3 if there is no shorter way to move from u to v than to go through e. It is 0 when $u = v$ is at the same time a mass and a hole of e. Again, formally, we want to solve an optimal transport problem (although in general this can no longer be interpreted as a coupling of random walkers), that is, move the first probability measure to the second one in an optimal way. This amounts to minimizing

$$\sum_{u \rightarrow e_i} \sum_{e_j \rightarrow v} d(u, v) \mathcal{E}(u, v) \tag{6}$$

over the set of all matrices $\mathcal{E}$ (called transport plans) whose entries $\mathcal{E}(u, v)$ represent the amount of mass from $\mu_{\mathcal{M}}(u)$ moved from vertex u to vertex v.

Given an optimal transport plan, if m_δ is the amount of mass that is moved at distance δ, then (by evaluating a formula analogous to (3)) the Ollivier-Ricci curvature O of e is defined as
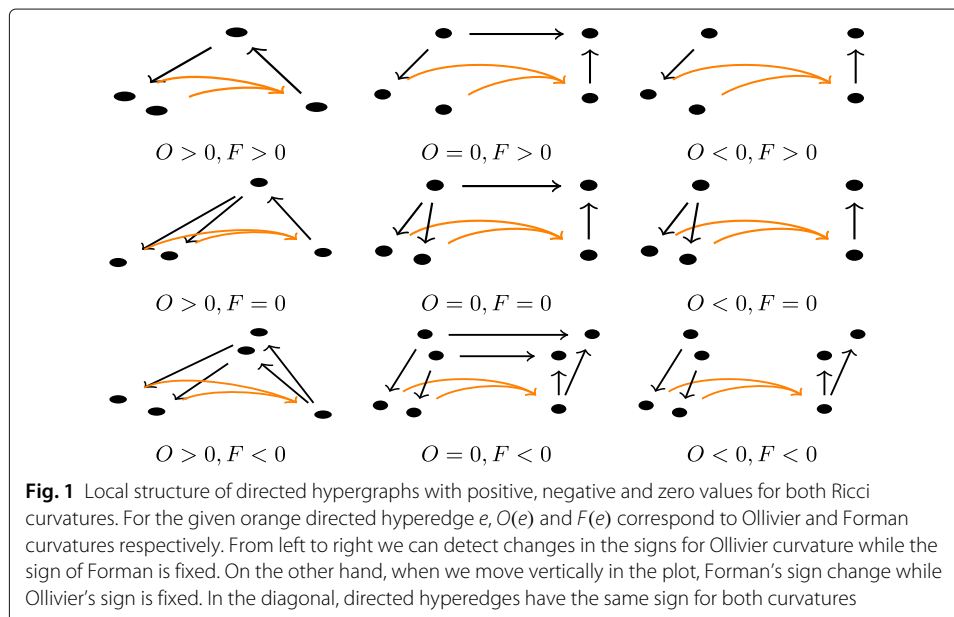
$$O(e) = m_0 - m_2 - 2m_3. \tag{7}$$

It is bounded above by 1 (reached when $m_0 = 1$ i.e. when each mass coincides with a hole of the same size) and below by -2 (reached when $m_3 = 1$ i.e. each mass has to be moved at distance 3).

Forman and Ollivier Ricci curvatures are complementary tools for the local analysis of connectivity patterns

Connectivity motifs

In this section we show how Forman F and Ollivier O Ricci curvatures can be used for the identification of specific connectivity motifs of directed hyperedges. We discuss signs and values of F and O for the orange hyperedges of the nine directed hypergraphs presented in Fig. 1, based upon the connections of their tails and heads. From left to right we can detect changes in the sign of O while the sign of F is fixed. On the other hand, when we move vertically in the plot, the F sign changes while the O sign is fixed. In the diagonal, directed hyperedges have the same sign for both curvatures. In the first column, each orange hyperedge $e = (e_i, e_j)$ has $O(e) > 0$. The reason is that the set of masses and the set of holes associated to e are equal, namely $\mathcal{M} = \mathcal{H}$. This implies that the distance between any mass and any hole is zero. Therefore, $m_0 = 1$ while $m_2 = m_3 = 0$. Thus, from Eq. 7, $O(e) = 1$. In contrast, F decreases when we move vertically in the same column. In particular, $F(e)$ is positive for the uppermost hypergraph because the size of the tail of e is greater than the sum of indegree values of the vertices in e_i, namely, $|e_i| - \sum_{i \in e_i}$ in-deg$(i) = 1$ and, at the same time, the sole vertex in the head of e has only one outgoing arrow, then $|e_j| - \sum_{j \in e_j}$ out-deg$(j) = 0$. When moving down in the same column, $\sum_{i \in e_i}$ in-deg(i) and $\sum_{j \in e_j}$ out-deg(j) increase by one in the second and third hypergraphs, respectively (with values $F(e) = 0$ and $F(e) = -1$).

The arguments above explain the signs of F in the second and third columns. On the other hand, $O(e)$ values decrease from left to right in rows, due to the increase in the distance between each mass and each hole of e. Distances in the second column increase by 1 compared to the first column, which means that $m_0 = m_2 = m_3 = 0$ and $O(e) = 0$. Similarly, for any hypergraph in the third column, the distance from masses to holes becomes 3, and therefore, $m_0 = m_2 = 0, m_3 = 1$ and $O(e) = -2$.

Fig. 1 Local structure of directed hypergraphs with positive, negative and zero values for both Ricci curvatures. For the given orange directed hyperedge e, $O(e)$ and $F(e)$ correspond to Ollivier and Forman curvatures respectively. From left to right we can detect changes in the signs for Ollivier curvature while the sign of Forman is fixed. On the other hand, when we move vertically in the plot, Forman's sign change while Ollivier's sign is fixed. In the diagonal, directed hyperedges have the same sign for both curvatures

Hyperloops and their curvature

The definition of a loop in a hypergraph depends on the definition of a path. The strictest version requires that the head and tail of the hyperedge $e = (e_i, e_j)$ coincide, i.e. $e_i = e_j$. If, moreover, the hyperedge is isolated, then $F(e) = 0$ while $O(e) = 1$ (see the top right hyperedge of Fig. 2). If the hyperedge is not isolated, both curvature notions might change. As an example, if we move down in column one of Fig. 2, we see $F(e)$ and $O(e)$ decrease, as the result of the incoming neighbors added to the tail of e and the outgoing neighbors added to its head.

A second, more flexible scenario, requires one of the following two cases to occur, $e_i \subset e_j$ or $e_j \subset e_i$. In this case, if the hyperedge is isolated, then $F(e) > 0$ and $O(e) < 1$. Again, both might change if the nodes of e have further connections. For instance, if we move down in column two of Fig. 2, $F(e)$ and $O(e)$ decrease due to the addition of hyperedges incident to e.

Finally, the most flexible version of a hyperloop simply requires the tail and head of e to have non-empty intersection, namely, $e_i \cap e_j \neq \varnothing$. Naturally, any hyperedge of the two types described before represents a particular case of this version. All hyperloops in

Fig. 2 Hyperloops and their curvature. The following are the values from left to right in rows. First row: $F(e) = 0, 1, 2$, while $O(e) = 1.0, 0.4444, 0.25$. Second row: $F(e) = -1, 0, 1$, and $O(e) = 0.7222, 0.1111, -0.25$. Third row: $F(e) = -3, -2, -1$, and $O(e) = 0.2777, -0.3888, -1.0$
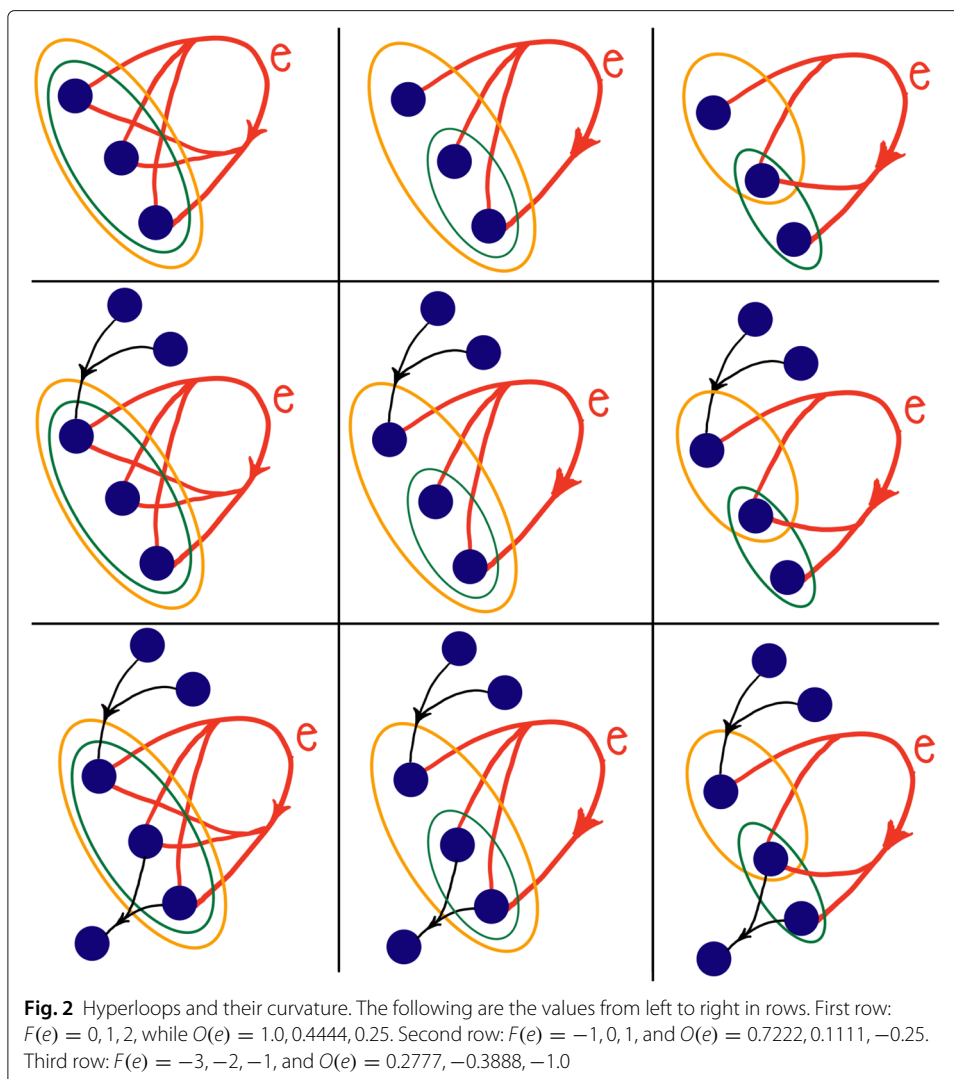
Fig. 2 are instances of this definition, showing that they can be flat, positively or negatively curved, for both F and O.

A random model of directed hypergraphs and its curvature fingerprint

In this section we introduce a random model of directed hypergraphs and explore the distribution of the degrees of nodes, the sizes of tails and heads, and the Forman and Ollivier Ricci curvatures. The model serves as a baseline to understand how curvature distributions look when connections (hyperedges) are made randomly following simple rules. It might be regarded as a hypergraph version of the Erdös-Rényi (ER) model. Here the size of a hyperedge e, can be any number between 1 and $|V|$. However, we decide not to choose the tail and head sizes of hyperedges randomly (with the obvious restriction that their sum should equal the size of the hyperedge), but rather let them match the statistical properties of the empirical networks that we want to analyze. The process is explicitly described in Algorithm 1.

ER model description: The input of Algorithm 1 is the number of vertices n and the number of directed hyperedges m. The output is a directed hypergraph $D = (V, E)$ where $V = \{1, \ldots, n\}$, and E is a multiset of directed hyperedges. Each hyperedge e is an ordered pair (e_i, e_j) of subsets of V. First, we associate a random integer between 2 and n to the size of e, namely, $|e|$. Then, $|e|$ nodes are randomly sampled from V without replacement, and, roughly $|e|/2$ are assigned to the tail e_i and roughly $|e|/2$ to the head e_j (see Algorithm 1). Notice that the model produces non-empty tails and heads. The new formed directed hyperedge $e = (e_i, e_j)$ is then added to E. The process is independently carried out m times.

Algorithm 1: Erdős Rényi model for directed hypergraphs (ER)

    input : Number of nodes n; number of directed hyperedges m
    output: Directed hypergraph $D = (V, E)$, where $V = \{1, \ldots, n\}$

1 E starts as an empty list;
2 **for** $i \leftarrow 1$ **to** m **do**
3     $|e|$ = random integer from $\{2, \ldots, n\}$;
4     $v(e)$ = random subset of V with $|e|$ nodes;
5   2 vertices, i and j, are randomly chosen from $v(e)$;
6     $e_i = \{i\} \cup S$, where S is a random subset of $v(e) \setminus \{i, j\}$ of size $(|e| - 2)/2 + r$, r being an integer randomly chosen from $\{0, 1\}$;
7     $e_j = v(e) \setminus e_i$; // notice that j is assign to e_j
8     $e = (e_i, e_j)$ is added to E
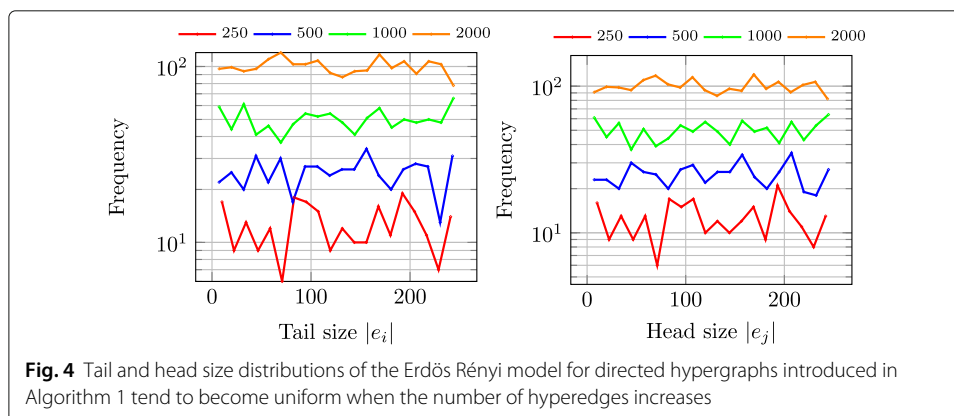9 **end**
10 **return** V, E

Numerical approach to ER properties: We use Algorithm 1 to compute four random directed hypergraphs over a fixed set of 500 vertices. The number of hyperedges are 250, 500, 1000, and 2000, respectively. Figure 3 shows the in-degree and out-degree distributions of nodes and suggests a normal behaviour. Moreover, since the size of tails and heads is an integer chosen randomly from a fixed interval, both distributions, tail sizes and head

Fig. 3 In-degree and out-degree distributions of nodes of the Erdös Rényi model for directed hypergraphs introduced in Algorithm 1 suggest a normal behavior
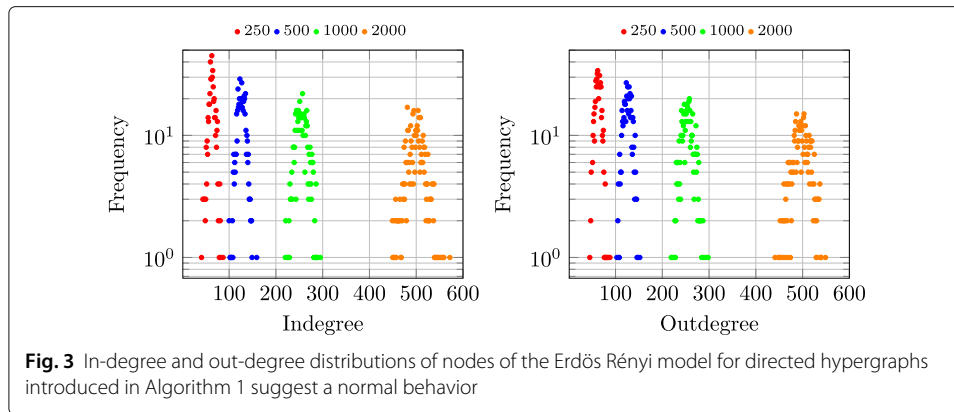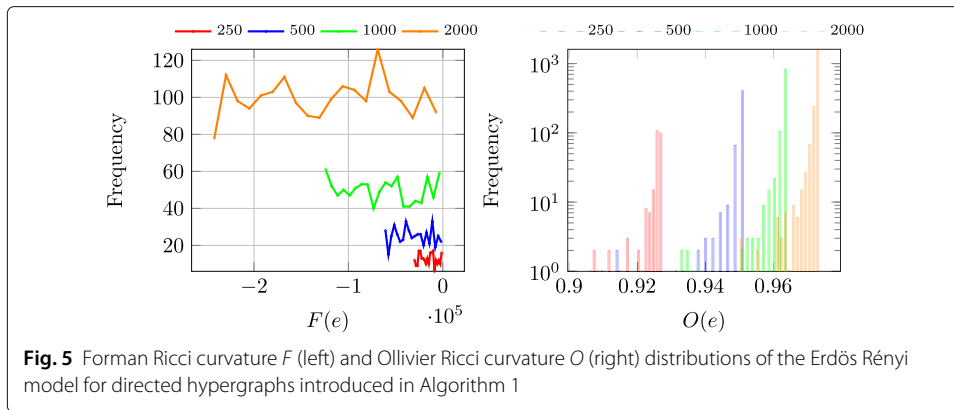
sizes, are expected to be uniform, explaining the shapes in Fig. 4. $F(e)$ is the difference between hyperedge size and the number of incoming and outgoing hyperedges of nodes in the tail and head, respectively. In this model, $F(e)$ values are uniformly distributed over an interval of the form $[x, 0]$ where x decreases with the number of hyperedges m, as shown in Fig. 5 (left). Similarly, $O(e)$ takes positive real values within the interval $[0.9, x]$ where x approaches to 1 as m increases, as shown in Fig. 5 (right). This result is triggered by the presence of arbitrary large hyperedges and the randomness of the wiring process of the ER model which, combined, produce hyperedges with sets of masses and holes as big as the entire set of vertices, as Fig. 6 shows.

Metabolism of *E.coli*

Metabolic networks are clear examples of directed hypergraphs. They consist of a set of chemical reactions involving a fixed set of chemical species. Each reaction, $e_i \rightarrow e_j$, entails the change of chemical identity of an initial set e_i of metabolites, called reactants, to yield a second set e_j called products. This directed relationship can be captured naturally by the notion of hyperedge. In this section, we use our geometric tools to investigate the structure of the directed hypergraph that underlies the *E. coli* metabolic hypernetwork. In particular, we present the curvature distribution fingerprints of F and O, along with the quantities involved in their computation.

Fig. 4 Tail and head size distributions of the Erdös Rényi model for directed hypergraphs introduced in Algorithm 1 tend to become uniform when the number of hyperedges increases

Fig. 5 Forman Ricci curvature *F* (left) and Ollivier Ricci curvature *O* (right) distributions of the Erdös Rényi model for directed hypergraphs introduced in Algorithm 1
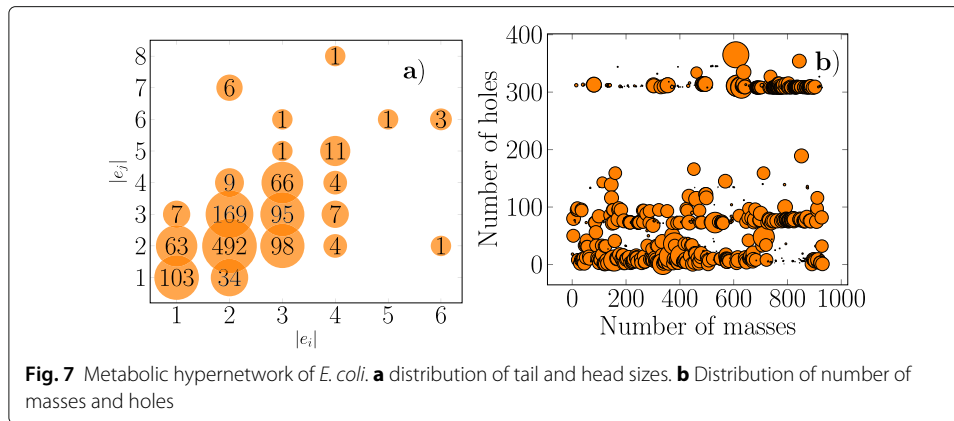
The directed hypergraph of *E. coli*: the metabolism of this bacterium, reported in Reed et al. (2003) (K-12 (iJR904GSM/GPR)), is modeled as a directed hypergraph. Vertices represent chemical substances (metabolites) and hyperarcs stand for metabolic reactions. To each reaction $e_i \rightarrow e_j$ we associate a directed hyperedge $e = (e_i, e_j)$. There are $|V| = 625$ metabolites, 245 reversible reactions and 686 non-reversible ones. Each reversible reaction $e_i \leftrightarrow e_j$ was split into two, the forward and the reverse reaction, yielding two hyperarcs, $e_i \rightarrow e_j$ and $e_j \rightarrow e_i$. Therefore, $|E| = 1,176$.

Now turning to Figs. 7, 12 and 13 we see that the distributions for *E.coli* are very different from those of random hypergraphs. Figure 7a) shows the number of metabolic reactions with $|e_i|$ reactants and $|e_j|$ products. 90% of chemical reactions have at most three reactants and three products (also observed for the whole Chemical Space (Llanos et al. 2019)), which, according to Eq. 5, indicates that frequent curvature values in Fig. 12 (left) are ruled by the accumulated in- and out-degree. In particular, frequent values of curvature were found to distinguish bottle neck and redundant reactions in the metabolic network (Leal et al. 2018). On the other hand, when considering the number of incoming neighbors of reactants and of outgoing neighbors of products for every reaction, frequencies are of the order of hundreds and, for some reactions, almost the whole substrate set, as shown in Fig 7 b). The question that arises is how close those masses and holes are in the metabolic network. The Ollivier-Ricci curvature distribution in Fig. 13 (left) shows

Fig. 6 Distribution of mass and hole sizes

Fig. 7 Metabolic hypernetwork of *E. coli*. **a** distribution of tail and head sizes. **b** Distribution of number of masses and holes
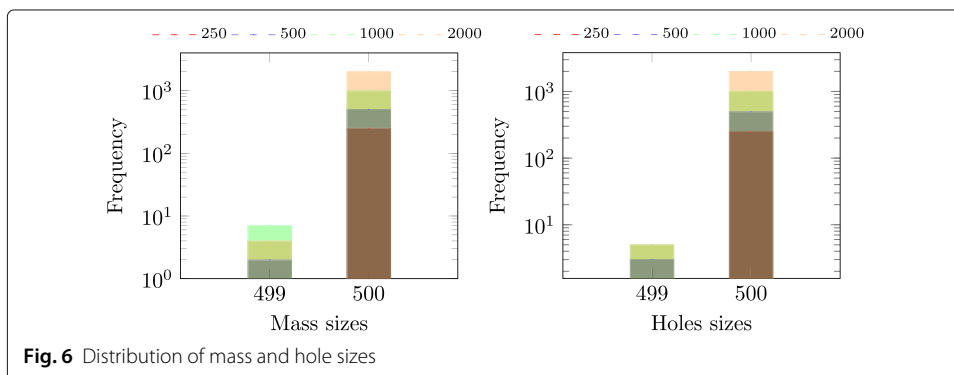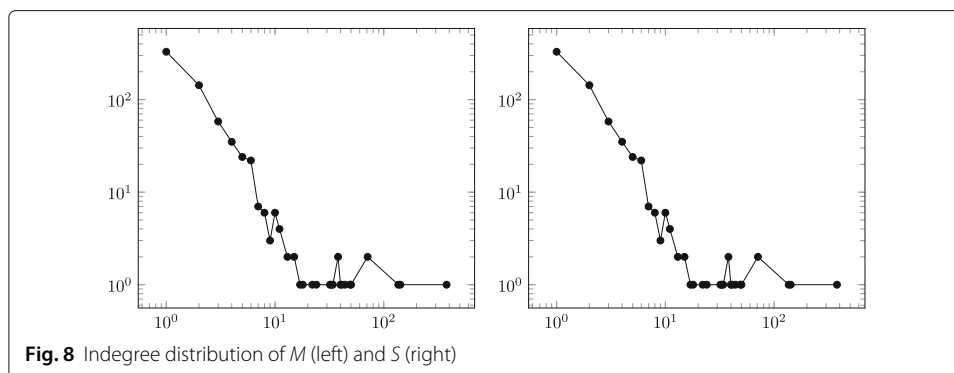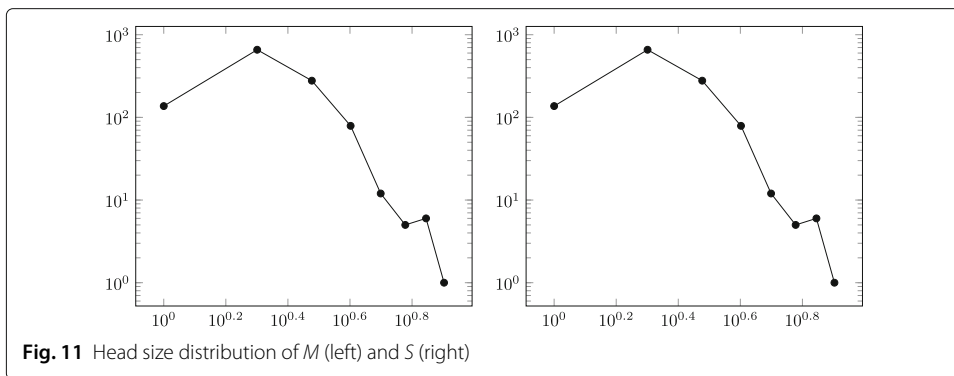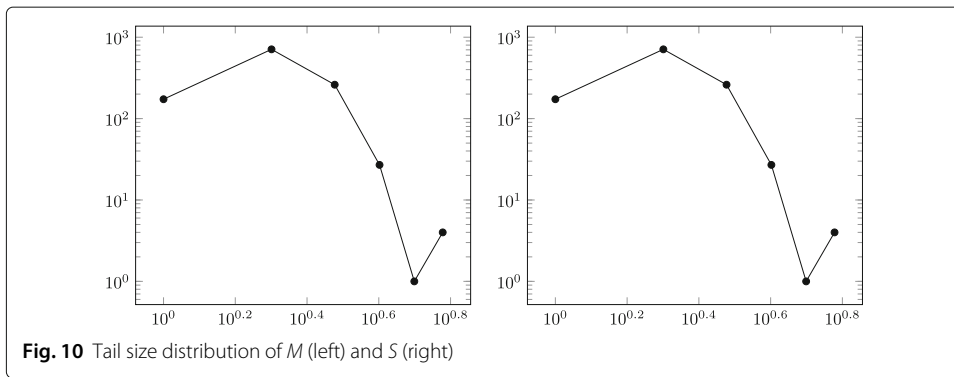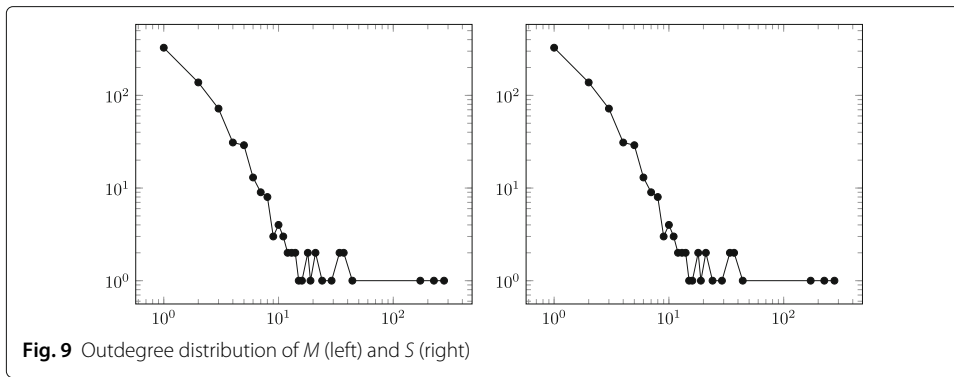
that most masses and holes are at distance less than 3, since the vast majority has curvature greater than -0.5. Less than 10% of incoming and outgoing neighbors are at distance 3. Only four reactions have curvature -2, indicating that their masses are at least three reactions away from their holes.
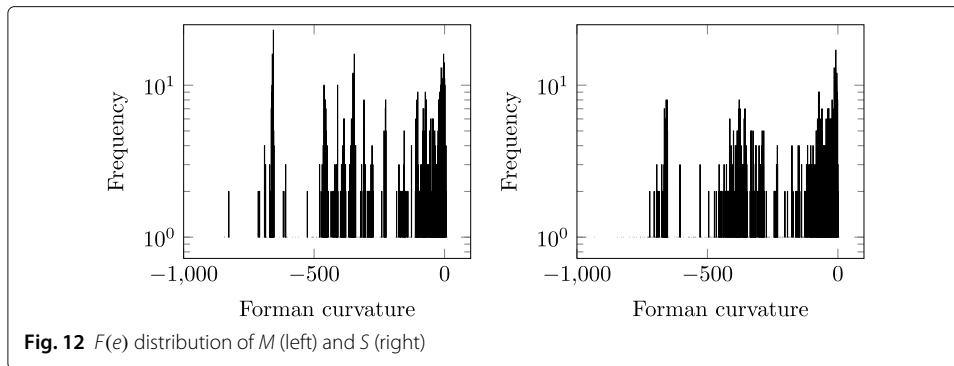
A *shuffled* **directed hypergraph of *E. coli*:** In the shuffling experiment, we start with the metabolic network *M* and end up with a directed hypergraph *S* of the same size (tail and head size distributions) and with the same degree sequence. Our goal is to investigate if our two notions of curvature can capture the wiring changes brought about by the shuffling of the hypernetwork. The distributions of the quantities involved are presented in Figs. 8-13. Figures 8 and 9 show that our shuffling algorithm does preserve the degree sequence i.e. the degree sequences of *M* and *S* are the same. Similarity, Figs. 10 and 11 show that the distributions of tail and head sizes of *M* and *S* are the same. Importantly, *F* and *O* distributions of *M* and *S* are different, as shown in Figs. 12 and 13. Confirming that our notions of curvature detect differences that the degree cannot.

Conclusions

The generalizations of Forman and Ollivier Ricci curvatures introduced in Leal et al. (2018) and Eidi and Jost (2020) detect complementary aspects of the connectivity patterns of directed hyperedges. The former detects the difference of the flow in the direction of the hyperedge under consideration and its size. This curvature can be used to define

Fig. 8 Indegree distribution of *M* (left) and *S* (right)

Fig. 9 Outdegree distribution of *M* (left) and *S* (right)

Fig. 10 Tail size distribution of *M* (left) and *S* (right)

Fig. 11 Head size distribution of *M* (left) and *S* (right)

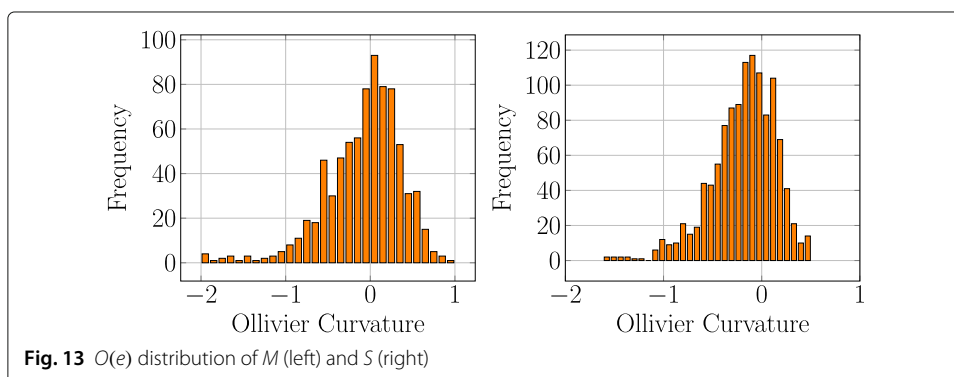**Fig. 12** *F(e)* distribution of *M* (left) and *S* (right)

assortativity in directed hypergraphs. The second informs on the existence of shorter alternative paths from incoming to outgoing neighbors of a given hyperedge, and in particular, measures the overlap of these two sets.

A random model for directed hypergraphs was introduced. The node degree distribution behaves normally, the hyperedge size and Forman Ricci values follow uniform distributions. The model produces directed hyperedges positively curved in the sense of Ollivier, due to the presence of huge hyperedges. Since these parameters, key for the analysis of networks, bear a clear fingerprint, our ER random model is a suitable baseline for the study of empirical hypernetworks.

Importantly, when we evaluate the curvatures on an empirical network, the metabolism of *E.coli*, the results look very different. This points to key insights into the global structure of such metabolic networks.

Changes in the wiring of a hypergraph, which cannot be detected with the degree sequence of the vertices or with the degree sequence of the edges (edge size), can be captured by changes in geometric features of the hyperedges. In particular, Forman and Ollivier Ricci curvature, provide the appropriate resolution to detect changes in the wiring of a hypernetwork produced by vertex degree and edge size preserving processes, as suggested by the hyperedge shuffling presented here and applied to the metabolic network of *E. coli*.

Fig. 13 *O(e)* distribution of *M* (left) and *S* (right)

Author details
[1]Bioinformatics group, Leipzig University, Härtelstr. 16-18, D-04107 Leipzig, Germany. [2]Max Planck Insitute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany. [3]Santa Fe Institute, 1399 Hyde Park Rd, 87501 Santa Fe, United States.

References
Asoodeh S, Gao T, Evans J (2018) Curvature of hypergraphs via multi-marginal optimal transport. In: 2018 IEEE Conference on Decision and Control (CDC). IEEE, Piscataway. pp 1180–1185. https://doi.org/10.1109/CDC.2018.8619706
Banerjee A (2020) On the spectrum of hypergraphs. Linear Algebra Appl. In Press. https://doi.org/10.1016/j.laa.2020.01.012
Eidi M, Jost J (2020) Ollivier ricci curvature of directed hypergraphs. Sci Rep 10:12466. https://doi.org/10.1038/s41598-020-68619-6 1907.04727
Forman R (2003) Bochner's method for cell complexes and combinatorial ricci curvature. Discret Comput Geom 29:323–374
Jost J, Liu S (2014) Ollivier's ricci curvature, local clustering and curvature-dimension inequalities on graphs. Discret Comput Geom 51(2):300–322. https://doi.org/10.1007/s00454-013-9558-1
Leal W, Eidi M, Jost J (2019) Curvature-based analysis of directed hypernetworks. In: Cherifi H (ed). Complex networks 2019 : the 8th international conference on complex networks and their applications; December 10 - 12, 2019 Lisbon, Portugal ; book of abstract. International Conference on Complex Networks & Their Applications. pp 32–34
Leal W, Restrepo G, Stadler PF, Jost J (2018) Forman-ricci curvature for hypergraphs. arXiv:1811–07825. https://doi.org/10.13140/RG.2.2.27347.84001
Llanos EJ, Leal W, Luu DH, Jost J, Stadler PF, Restrepo G (2019) Exploration of the chemical space and its three historical regimes. Proc Natl Acad Sci 116(26):12660–12665. https://doi.org/10.1073/pnas.1816039116 https://www.pnas.org/content/116/26/12660.full.pdf
Ollivier Y (2007) Ricci curvature of Markov chains on metric spaces. math/0701886. Accessed 1 Jan 2020
Ollivier Y (2013) A visual introduction to Riemannian curvatures and some discrete generalizations. In: Analysis and geometry of metric measure spaces. CRM Proc. Lecture Notes, 56, Amer. Math. Soc., Providence. pp 197–220
Reed JL, Vo TD, Schilling CH, Palsson BØ (2003) An expanded genome-scale model of escherichia coli k-12 (ijr904 gsm/gpr). Genome Biol 4:54–15412
Samal A, Sreejith RP, Gu J, Liu S, Saucan E, Jost J (2018) Comparative analysis of two discretizations of Ricci curvature for complex networks. Sci Rep 8:8650. https://doi.org/10.1038/s41598-018-27001-3 1712.07600
Saucan E (2019) Metric curvatures and their applications 2: metric ricci curvature and flow. 1902.03438. Accessed 1 Jan 2020
Spivak DI (2009) Higher-dimensional models of networks. 0909.4314. Accessed 1 Jan 2020

Publisher's Note

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