Discrete Curvatures and Network Analysis

Emil Saucan[*1,2], Areejit Samal[†3], Melanie Weber[‡4], Jürgen Jost[§5,6]

[1] *Technion - Israel Institute of Technology, Haifa, Israel*
[2] *Kibutzim College of Education, Tel Aviv, Israel*
[3] *The Institute for Mathematical Sciences, Homi Bhabha National Institute, Chennai, India*
[4] *Princeton University, Princeton, NJ, USA*
[5] *Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany*
[6] *The Santa Fe Institute, Santa Fe, NM, USA*

Abstract

We describe an approach to the analysis of chemical (and other) networks that, in contrast to other schemes, is based on edges rather than vertices, naturally works with directed and weighted edges, extends to higher dimensional structures like simplicial complexes or hypergraphs, and can draw upon a rich body of theoretical insight from geometry. As the approach is motivated by Riemannian geometry, the crucial quantity that we work with is called Ricci curvature, although in the present setting, it is of course not a curvature in the ordinary sense, but rather quantifies the divergence properties of edges. In order to illustrate the method and its potential, we apply it to metabolic and gene co-expression networks and detect some new general features in such networks.

1 Introduction

In chemistry, at various levels of description, one encounters discrete representations of chemical structures or processes. These could be the atoms and their bonds in a molecule, strings of nucleotides or amino acids, the ingredients and products of chemical reactions, metabolic flow networks. Also similarity or descendence relations can be represented

---

[*]semil@ee.technion.ac.il

[†]asamal@imsc.res.in

[‡]mw25@math.princeton.edu

[§]Corresponding author: jost@mis.mpg.de

by trees or other such geometric structures. Moreover, we also find the coauthorship or citation networks of researchers in chemistry and other fields. In mathematics, systematic theories of discrete structures are currently among the most active research topics. In this contribution, we want to describe such an important recent research topic, that of discrete curvatures, and start its application to the analysis of chemical data. The research agenda developed here will then be applied more systematically in future research.

Let us start with perhaps the basic discrete structure relevant in this context, that of a network. Networks are represented by graphs, formed by edges that express relations between pairs of elements (vertices, nodes). These relations could be:

- binary (that is, present/absent, or in formal notation, 1/0)

- weighted

- directed

There could also exist relations between triples or n-tuples of elements. Such higher order relations can no longer be expressed by graphs, and they would lead us to a representation by more general mathematical objects, like simplicial complexes or hypergraphs.

In data analysis, when working with networks (or, similarly, more general such structures), we want to:

- identify qualitative properties of networks in a particular domain, and

- compare different networks, within or between domains,

- through computationally efficient and quick schemes.

In order to achieve this in a systematic and conceptually grounded manner, we take a look at another discipline that has developed tools to study the geometry of configurations, in order to see what we can use and what we need to develop on that basis. That discipline is, perhaps at first somewhat surprisingly in this context, Riemannian geometry, with its fundamental notion of curvature.

A crucial advantage of the approach presented here is that all prescriptions how to define such curvature values are guided by systematic mathematical theory. This is in stark contrast to the prevailing attitude in network analysis to introduce formal quantities and measures in a rather adhoc manner.

2 Curvature

In Riemannian geometry, curvature is defined in terms of first and second derivatives of the metric tensor [8]. There are different notions of curvature

- Sectional curvature, associated to tangent planes

- Ricci curvature, associated to tangent vectors (directions), obtained from averaging sectional curvatures

- Scalar curvature, associated to points, obtained from averaging Ricci curvatures

Sectional curvature is a full invariant, in the sense that it encodes all the (local) information about a Riemannian metric. It measures the divergence or convergence of geodesics. Ricci curvature is weaker, but still surprisingly powerful. It measures the growth of volumes and the coupling properties of random walks. Sectional curvature becomes 0 precisely for (locally) Euclidean spaces. Thus, there is a reference space, Euclidean space, that curvature allows a comparison with. Curvature measures the local deviation of a space from being Euclidean. Spaces of positive curvature have a local geometry similar to spheres, while those of negative curvature are similar to hyperbolic spaces. The situation between positive and negative curvature is not entirely symmetric. Sectional curvature ≤ 0 is well understood. The geometry of negatively curved spaces is very different from that of positively curved ones. Spaces of curvature ≥ 0 (or better $\geq K > 0$) are more difficult. Concerning Ricci curvature, somewhat surprisingly, the situation seems opposite. Ricci curvature ≥ 0 (or better $\geq K > 0$) enjoy strong restrictions, whereas Ricci ≤ 0 has no consequences, because every manifold of dimension ≥ 3 carries some complete metric of negative Ricci curvature, as was shown by Lohkamp [10]. Perhaps that difference between sectional and Ricci curvature inequalities can be intuitively explained by the fact that sectional curvature encodes global properties, whereas Ricci curvature (and also scalar curvature) rather characterize local properties. For sectional curvature, at least in the negative case, one captures its geometric content by zooming in from infinity. In contrast, for the geometric content of Ricci curvature, in particular in the positive case, one investigates local expansion properties. Scalar curvature is a much weaker invariant than Ricci curvature. While it is nevertheless important in Riemannian geometry, for our purposes, it will not play an essential role.

Curvature, as originally defined by Riemann [15], is given in terms of certain combinations of first and second derivatives of the metric tensor and thereby seems to require a differentiable structure. But in modern research, it was found that curvature inequalities, that is, some curvature being smaller or larger than some number K, can be equivalently expressed in terms of certain local or global properties that for their formulation do not require a differentiable structure. These properties could be certain distance relations in triangles, intersection properties of distance balls, or relations between the volumes of such balls of different radii.

3 Network curvature

Curvature can also be applied to networks, when conceived as geometric objects. For instance, we may consider a network as a metric space. When the edges of the network are unweighted and undirected, the distance between two vertices connected by an edge would be 1. And when it takes at least n edges to get from one particular vertex to some other one, their distance would be n.

More abstractly, in whichever way we geometrize networks, the basic idea in defining curvature inequalities for networks consists in taking some of the local or global properties that are equivalent to or implied by curvature inequalities in Riemannian geometry as the definition and exploring their consequences. We start with the following translation scheme.

- Sectional curvature was associated to tangent planes. As a plane is determined by 3 points, sectional curvatures should be assigned to collections of 3 vertices.

- Ricci curvature was associated to tangent vectors, that is, directions. Such a direction is determined by 2 points, that is, an edge. Thus Ricci curvature would be assigned to edges. Like a direction, that edge could be directed, and so, we should get meaningful notions of Ricci curvature also for directed networks.

- Scalar curvature was associated to points. Therefore, in the context of networks, it should be assigned to vertices.

As already mentioned, sectional curvature elucidates global properties, zooming in from infinity. Therefore, when using sectional curvature tools to probe networks, one

should naturally look at triples of vertices that are far apart from each other. In contrast, Ricci curvature was associated with local properties. Therefore, in the network context, we shall look at expansion of edges, or the transportation between vertices along edges. Likewise, scalar curvature was related to local properties. In the network context, these are simply the degrees of vertices. Of course, the vertex degrees are relatively weak invariants, although the distribution of vertex degrees in a network does provide some structural insight.

4   Ricci curvature of networks

While a good definition of sectional curvature for general metric spaces has been developed in [1], this has not yet been systematically explored on empirical networks. Therefore, in this contribution, we shall concentrate on Ricci curvature. There exist two different notions of generalized Ricci curvature that are useful in the present context. We shall now explore those.

4.1   Ricci curvature I: Ollivier

The starting point of Ollivier's notion of generalized Ricci curvature [12, 13] is the *transportation distance* between two points in a metric space. One compares the distance between those points themselves with the distance between two small balls around them. That is, each point in one ball has to be transported to some point in the other ball. The *transportation cost* is given by the distance between those two points. One then tries to arrange these pairs of points in the two balls optimally, in the sense that the total transportation cost between the two balls is as small as possible.

Since in this paper, we consider the Ollivier curvature only for networks, we let (X, d) be a finite graph with some notion of distance d between vertices. As explained, for an unweighted graph, the distance between adjacent vertices is 1. Let m_1, m_2 be two probability measures on X. Note that, for probability measures, $0 \le m_i(A) \le 1, i = 1, 2$, for any subset A of X. A *coupling* (or *transport plan*) of m_1 and m_2 is a measure π on $X \times X$ with *marginals* m_1 and m_2, i.e. such that, for all measurable sets $A, B \subset X$, the following properties hold: $\pi[A \times X] = m_1[A]$ and $\pi[X \times B] = m_2[B]$, that is, $\sum_{x \in X} \pi(x, y) = m_2(y)$, and $\sum_{y \in X} \pi(x, y) = m_1(y)$. Given a (positive) *cost function*[1] $c(x, y)$ on $X \times X$, one can

---

[1]The cost function $c(x, y)$ should be viewed (as, indeed, intended originally by Monge), as the work

define the *Monge-Kantorovich minimization problem*

$$\min \sum_{X \times X} c(x, y)\pi(x, y), \tag{1}$$

where the minimum is taken over all the transport plans. The transport plans attaining the minimum are called *optimal transport plans*.

The L^1-*Wasserstein* (or *transportation*) *distance* between the measures m_1, m_2 is defined as

$$W_1(m_1, m_2) = \min \sum_{X} d(x, y)\pi(x, y), \tag{2}$$

where the minimum is taken over all the transportation plans between m_1 and m_2.

In the simplest case, we have an unweighted and undirected graph and consider two neighboring vertices $(x \sim y)$, that is, x and y are connected by an edge and therefore have distance 1. We let m_x and m_y be the measures on the distance balls $B(x, 1) = \{z : z \sim x\}, B(y, 1) = \{w : w \sim y\}$ (for simplicity, we omit the centers x and y, respectively, from the balls, because we know that their distance is 1 anyway) around x and y, respectively, of radius 1. The measure is $m_x = \frac{1}{d_x}$, where d_x is the degree of x, that is, the number of neighbors of x puts equal weight on all neighbors. $W_1(m_x, m_y)$ then is the transportation distance between the two measures m_x and m_y.

The definition of the Ricci curvature then is

$$\kappa(x, y) := 1 - W_1(m_x, m_y). \tag{3}$$

When the transportation distance $W_1(m_x, m_y)$ between the neighborhoods of x and y, as represented by the measures m_x and m_y, is smaller than the distance between their centers, one has positive Ricci curvature in the sense of Ollivier, and when it is larger, the Ricci curvature is negative. The geometric intuition is that the more positive the Ricci curvature is, the more two distance balls centered at nearby points overlap, and therefore, the cheaper it is to transport the mass from one to the other.

Triangles and quadrangles decrease transportation cost, hence increase Ricci curvature. Here, a triangle is a configuration of three vertices x, y, z that are mutually neighbors. Thus, such a z is contained in both $B(x, 1)$ and $B(y, 1)$, and therefore, it need not be transported at all between the two balls. It therefore incurs a transportation cost of 0. Likewise, when we have a quadrangle involving x and y, that is, vertices z, w with $z \sim x, w \sim y$ and $z \sim w$, then we can transport z to w with a cost of 1.

---

necessary (or amount of energy needed) to transport a unit mass from point x to the point y.

For an illustration, based on [9], see Fig. 1.
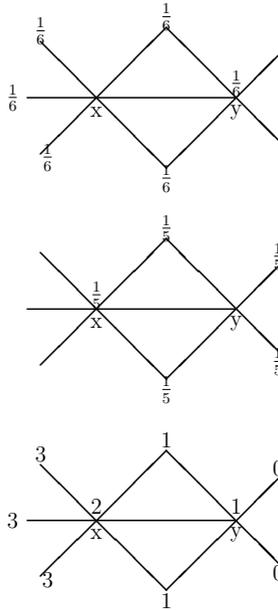
Figure 1. Illustration of optimal mass transportation on a simple graph configuration: *Starting configuration (top)* The vertex x has 6 neighbors, each of them getting mass $\frac{1}{6}$; mass 0 at all other vertices; *Target configuration (middle)* The vertex y has 5 neighbors, each getting mass $\frac{1}{5}$; *The optimal transportation plan (bottom)* Mass is moved from vertices with larger value to those with smaller ones. Since not all the mass from the three vertices denoted x with label 3 can go the single neighbor denoted y with label 2 (x itself), some of it has to be transported to other neighbors y with smaller labels. On the other hand, the two vertices of the triangles with label 1 are neighbors of both x and y, and so, their mass of $\frac{1}{6}$ each need not be transported at all. However, as these vertices have to acquire mass $\frac{1}{5}$ in the target configuration, they still need to get a little mass from other vertices.

Recalling the defining Equation (3), and with the notation

$$\sharp(x,y) := \text{number of triangles which include } x, y \text{ as vertices, for } x \sim y, \qquad (4)$$

we have the inequality of [9], saying that

$$\kappa(x,y) \geq -\left(1 - \frac{1}{d_x} - \frac{1}{d_y} - \frac{\sharp(x,y)}{d_x \wedge d_y}\right)_+ - \left(1 - \frac{1}{d_x} - \frac{1}{d_y} - \frac{\sharp(x,y)}{d_x \vee d_y}\right)_+$$
$$+ \frac{\sharp(x,y)}{d_x \vee d_y}. \qquad (5)$$

where, for any real numbers s, t, we put $s_+ := \max(s, 0), s \vee t := \max(s, t), s \wedge t := \min(s, t)$.

Of course, one can also consider $\kappa(x, y)$ for two vertices that are not neighbors. One simply puts $\kappa(x, y) := 1 - \frac{W_1(m_x, m_y)}{dist(x, y)}$, where, as explained, $dist(x, y)$ is the minimal number of edges needed to get from x to y. For studying graphs and networks, it seems more useful, however, to work with the neighborhood graphs introduced in [2] instead.

Ollivier's notion of Ricci curvature can be extended to weighted and directed networks. When the vertices v have weights w_v, the mass is distributed among the neighbors z of x according $\frac{w_z}{\sum_{v \sim x} w_v}$, and accordingly for y. When the edges have weights, the transportation cost should get multiplied by the inverse of the edge weights, because edges of smaller weight have less carrying capacity. And when the network is directed, we can only use directed edge paths going from a neighbor of x to some neighbor of y. Then the transportation distance from x to y will be different in general from that from y to x, and therefore in general $\kappa(x, y) \neq \kappa(y, x)$.

The Ollivier-Ricci has many nice properties. In particular, it controls eigenvalues of the graph Laplacian [12, 13], with improved estimates via neighborhood graphs [2]. This is important because those eigenvalues yield much insight into the global and local structure of a network and therefore constitute important tools in network analysis. But instead of exploring that aspect, we now turn to the second notion of generalized Ricci curvature.

4.2 Ricci curvature II: Forman

The idea behind Forman's definition [7] of generalized Ricci curvature is that Ricci curvature measures how fast the volume of distance ball grows. For networks, that means how fast edges spread in different directions. In particular, edges with very negative curvature

should play a special role for the spreading out and hence for, e.g., information dispersal in a network.

Forman's curvature [17] for an undirected edge e of weight w_e with endpoints v_1 and v_2 with weights w_{v_1}, w_{v_2} is

$$\text{Ric}(e) = w_e \left(\frac{w_{v_1}}{w_e} + \frac{w_{v_2}}{w_e} - \sum_{e_{v_1}, e_{v_2} \sim e} \left[\frac{w_{v_1}}{\sqrt{w_e w_{e_{v_1}}}} + \frac{w_{v_2}}{\sqrt{w_e w_{e_{v_2}}}} \right] \right) \tag{6}$$

where e_{v_1}, e_{v_2} are the edges connected to nodes v_1 and v_2. (Note that here the second sum should not be interpreted as a double sum, but rather like the sum over all the edges incident to e.) When all nonzero weights are 1, or equivalently, when the network is unweighted, this reduces to

$$\text{Ric}(e) = 4 - \deg(v_1) - \deg(v_2) = \sum_{i=1}^{2} (2 - \deg(v_i)). \tag{7}$$

This quantity can be easily understood. It measures the spreading or dispersion at the vertices of the edge e. Thus

$$\text{Ric}(e) \ll 0 \text{ if both vertices have high degree.} \tag{8}$$

In contrast,

$$\text{Ric}(e) = 0 \text{ if both vertices have degree 2,} \tag{9}$$

and so, a graph has a cycle has $\text{Ric} \equiv 0$. (The other only configuration having $\text{Ric} \equiv 0$ is, as one can easily check, a tripod.) This is in contrast to the Ollivier curvature, where the length of the cycle affects the magnitude of the curvature. Thus, typically, the Forman-Ricci curvature of an edge is negative, in contrast to the Ollivier-Ricci curvature which is frequently positive. Therefore, we might expect that these two notions of Ricci curvature capture different properties of a network. Nevertheless, it turns out that in empirical networks, they are usually highly correlated. This is useful, because the Forman curvature is much easier to compute than the Ollivier curvature [16].

As, we have seen in the preceding simple examples, Forman's Ricci curvature can infer global (i.e. topological) facts about a network solely from local information (namely curvature). Below, we shall illustrate this for metabolic flow networks. Since those networks are directed, we shall now describe our generalization of the Forman-Ricci curvature to directed networks [18]. For this purpose, we separate the contributions of the two vertices

involved,

$$\mathrm{Ric}(e) = w_e \left(\frac{w_{v_1}}{w_e} - \sum_{e_{v_1} \sim e} \frac{w_{v_1}}{\sqrt{w_e w_{e_{v_1}}}} \right) + w_e \left(\frac{w_{v_2}}{w_e} - \sum_{e_{v_2} \sim e} \frac{w_{v_2}}{\sqrt{w_e w_{e_{v_2}}}} \right) \qquad (10)$$

and define the curvature of a directed edge by only using the term involving its initial vertex, or alternatively that for its terminal vertex. While computing the Forman curvature of a directed edge $e = \overrightarrow{v_1 v_2}$ that originates from node v_1 and terminates at node v_2, we take into account only those directed edges that either terminate at node v_1 or originate at node v_2 [18]. Alternatively, for a node in a directed network, we can distinguish its incoming and outgoing edges. Denote the set of *incoming* and *outgoing* edges for a node v by $E_{I,v}$ and $E_{O,v}$. We then define the *In Forman curvature* $\mathrm{Ric}_I(v)$ and the *Out Forman curvature* $\mathrm{Ric}_O(v)$ by

$$\mathrm{Ric}_I(v) \;\; = \;\; \sum_{e \in E_{I,v}} \mathrm{Ric}(e_v) \qquad (11)$$

$$\mathrm{Ric}_O(v) \;\; = \;\; \sum_{e \in E_{O,v}} \mathrm{Ric}(e_v)\,, \qquad (12)$$

summing over only the incoming or outgoing edges, respectively.

The total amount of flow through a node v then is

$$\mathrm{Ric}_{I/O}(v) = \mathrm{Ric}_I(v) - \mathrm{Ric}_O(v)\,. \qquad (13)$$

Similarly, for a directed edge, we can take the difference inflow $-$ outflow at the initial vertex and outflow $-$ inflow at the terminal vertex.

We underline the importance of the In and Out Forman curvatures in the study of directed (hyper-)networks arising in Chemistry in Section 5.1 (see, in particular, Figure 3).

With Forman's curvature notion, one can also construct a Ricci flow [19, 20] to smoothen a network or to identify sensitive regions

$$\tilde{\gamma}(e) - \gamma(e) = -\mathrm{Ric}_F(\gamma(e)) \cdot \gamma(e)\,, \qquad (14)$$

where $\tilde{\gamma}(e)$ denotes the new (updated) weighting scheme with $\gamma(e)$ being the initial (given) one.

Summarizing some of the preceding leads us to a comparison between Ollivier-Ricci and Forman-Ricci, involving the following items

- They emphasize different aspects: clustering vs. dispersion.

- Forman is simpler to compute.

- Ollivier is related to network coherence (via eigenvalue estimates).

- There is a natural Ricci flow for Forman.

- Empirically, they are strongly correlated.

Remark As already mentioned in the introduction, not all discrete structures arising in chemistry can be fully captured by networks, that is, graphs. When we have also higher order relations, we are naturally lead to simplicial complexes or hypergraphs. For a simplicial complex, we have vertices $v_1, \ldots, v_N$. A k-dimensional simplex is given by a relation among $k+1$ distinct vertices, say, $v_{i_0}, \ldots, v_{i_k}$, and it is required that whenever we have such a simplex, also all its subsimplices are members of the complex. That means that whenever we have such a k-simplex, and $\{j_0, \ldots, j_\ell\} \subset \{i_0, \ldots, i_k\}$ (all indices in a set are assumed to be distinct, and so $\ell \leq k$), then also the ℓ-simplex with vertices $v_{j_0}, \ldots, v_{j_\ell}$ belongs to the complex. When this condition is not satisfied, we speak of a hypergraph instead of a simplicial complex.

Forman's notion of Ricci curvature naturally extends to weighted simplicial complexes. According to this formula, for a higher dimensional simplicial complex, we also get positive contributions to the curvature of an edge from two-dimensional faces containing the edge in question. That is, whenever an edge is contained in a triangle, this increases its curvature. Thus, when, for instance, we insert triangles to express triple relations, we increase the Ricci curvature [21].

5 Application of Forman-Ricci curvature to metabolic and gene co-expression networks

5.1 Metabolic networks

Metabolism plays a central role in living organisms. Biochemical reactions in the metabolic network of a cell are responsible for converting nutrient metabolites into key metabolites required for growth and maintenance of an organism. As an example of application of Forman-Ricci curvature to chemical networks, we here present results from an analysis of

metabolic networks inside two well-studied models, *Escherichia coli* and *Saccharomyces cerevisiae*.
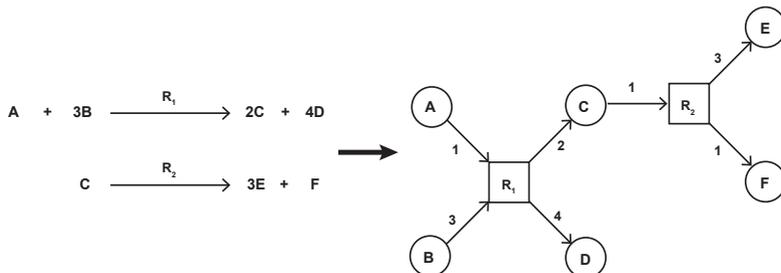
Figure 2. Construction of directed bipartite graph for metabolic networks. In this figure, we have shown the construction of the directed bipartite graph for two biochemical reactions. Here, metabolites in the graph are depicted as circles while reactions are depicted as rectangles.

For our analysis, we have used the *E. coli* metabolic network iJR904 [14] which contains 931 reactions involving 761 metabolites, and the *S. cerevisiae* metabolic network iND750 [5] which contains 1149 reactions involving 1061 metabolites. Starting from the list of biochemical reactions in the *E. coli* and *S. cerevisiae* metabolic networks, we have constructed a directed bipartite graph of reactions and metabolites as follows. As a first step, we have converted each reversible reaction in the network into two irreversible reactions corresponding to forward and backward reactions. Subsequently, we have converted each irreversible reaction into directed edges which connect either the substrate metabolites to the reaction or the reaction to product metabolites (Fig. 2). Moreover, we assign a weight to each directed edge in the bipartite graph which corresponds to the stoichiometry of the involved metabolite (substrate or product) in the reaction under consideration (Fig. 2). For example, a reaction $R1$ where 1 molecule of metabolite A combines with 3 molecules of metabolite B to produce 2 molecules of metabolite C and 4 molecules of metabolite D leads to the following 4 edges, $A \rightarrow R1$ with weight 1, $B \rightarrow R1$ with weight 3, $R1 \rightarrow C$ with weight 2, and $R1 \rightarrow D$ with weight 4, in the bipartite graph (Fig. 2).

Using the constructed directed bipartite graph for the metabolic networks of *E. coli* and *S. cerevisiae*, we have computed the distributions of the Forman curvature of a directed edge [18] in the metabolic networks of the two organisms which are shown in Fig.

3. From Fig. 3, it is seen that the distribution of Forman curvature is broad in the metabolic networks of the two organisms. Moreover, we find that there are some perspicuous secondary peaks in the Forman curvature distribution of the two organisms in the range from –400 to –200 with frequency greater than 50, apart from the main peak near 0 (Fig. 3). Specifically, there are 3 such peaks in the Forman curvature distribution of the *E. coli* network and 2 such peaks in the distribution of the *S. cerevisiae* network. We then extracted the edges in the directed bipartite graph that contribute to these secondary peaks in the Forman curvature distribution for the metabolic networks of the two organisms. Interestingly, we found that all the edges that give rise to these secondary peaks in the Forman curvature distribution of the two metabolic networks are between a single metabolite (H^+ or proton) and different reactions in the network. Thus, these secondary peaks in the curvature range from –400 to –200 correspond to the connections of high degree metabolite H^+ which is ubiquitous in any metabolic network. Thus, our analysis can detect substructures inside a network that belong to particular classes of vertices or edges. Once they are identified, one can then analyze their chemical role in detail. In the present example, this is rather straightforward, but in our second case study, where a similar phenomenon will be seen, this will already be more subtle.

We have also computed the In and Out Forman curvatures both of edges and of nodes. As can be seen in Fig. 3, the distributions of these directed versions are, with the exception of a number of sparse, isolated peaks, strongly concentrated towards (and, in fact, even at) 0. However, secondary peaks ("humps") appear in the distribution of the In and Out Forman curvatures of reaction nodes. In contrast, such peaks are absent in the distributions of In Forman curvature of metabolite nodes and Out Forman curvature of metabolite nodes. Moreover, the distributions of In Forman curvature of metabolite nodes and Out Forman curvature of metabolite nodes is broad in comparison to the distributions of In Forman curvature of reaction nodes and Out Forman curvature of reaction nodes (see Fig. 3, (b) and (e) vs. (c) and (f)).
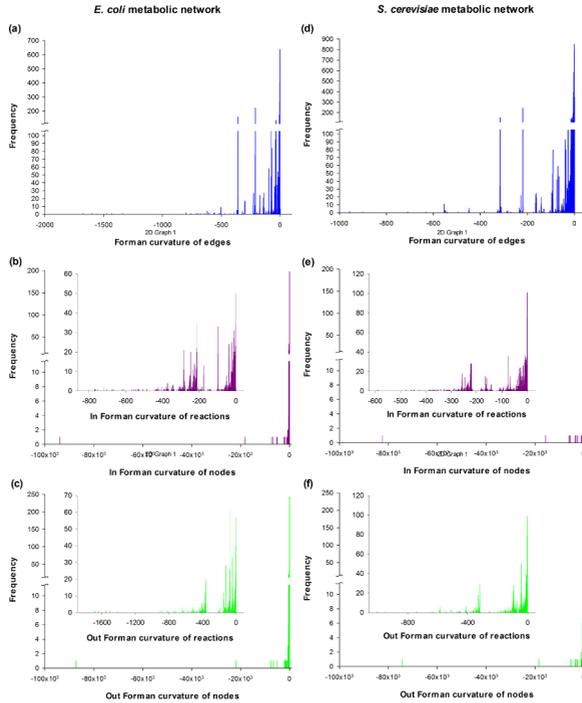
Figure 3. Distribution of Forman curvature of edges, In Forman curavature of nodes and Out Forman curvature of nodes in metabolic networks. (a-c) *E. coli and (d-f) S. cerevisiae. In the bipartite graph associated with the metabolic networks, there are two types of nodes: metabolites and reactions. In (b) and (c), insets show the distributions of In Forman curvature of reaction nodes and Out Forman curvature of reaction nodes in E. coli, and in (e) and (f), insets show the distributions of In Forman curvature of reaction nodes and Out Forman curvature of reaction nodes in S. cerevisiae. We find that the distributions of In Forman curvature of reaction nodes and Out Forman curvature of reaction nodes show secondary peaks while such peaks are absent in the distributions of In Forman curvature of metabolite nodes and Out Forman curvature of metabolite nodes. Moreover, the distributions of In Forman curvature of metabolite nodes and Out Forman curvature of metabolite nodes is broader compared to the distributions of In Forman curvature of reaction nodes and Out Forman curvature of reaction nodes.*

5.2   Gene co-expression networks

In a second study, we analyze a gene co-expression network with the Forman-Ricci formalism. The analysis of gene co-expressions is a popular method in genomics. Pairwise correlations of gene expression levels are computed across a set of samples to identify genes with highly correlated expression profiles. Such pairs of *co-expressed* genes are likely to have functional and regulatory commonalities making co-expression analysis a widely used tool for studying relationships among genes. Aiming to identify such functional substructures, we evaluate the curvature distribution across a co-expression network.

In this study, we analyze a gene co-expression network built from gene expression measurements in human brain tissue. We use publicly available gene expression data from the *Brainspan* atlas [3, 11] containing measurements for 17290 genes across 508 tissue samples. The construction of co-expression networks from experimental data is shown in Fig. 4. With a significance threshold for pairwise co-expression of ~ 0.8 the resulting network has ~ 1 Million edges.
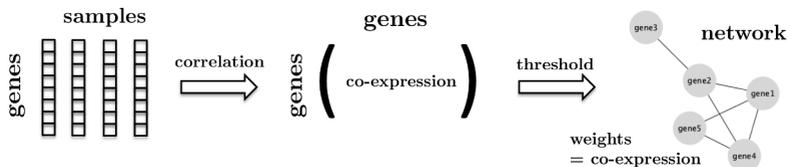
Figure 4. Construction of co-expression networks. In a first step, we compute the correlations of N gene expression levels across a set of M samples. The resulting $N \times N$ matrix contains the pairwise co-expression among all genes. We apply a threshold to maintain only significant co-expression pairs, the resulting sparse matrix is the adjacency matrix of the corresponding co-expression network. The network is weighted; edge weights encode the co-expression value of the respective gene pair.

The computation of Forman-Ricci curvature across the network revealed the distribution shown in Fig. 5. We identify a major hump at curvature-values of ~ -200 and smaller humps at ~ -80 and ~ -30. Starting from this observation, we performed a closer analysis of the genes associated with the subnetworks underlying the humps.
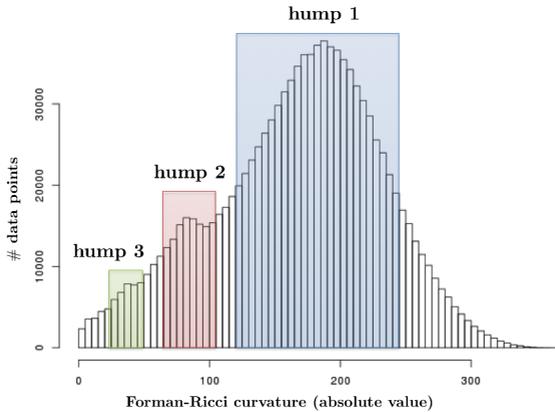
Figure 5. Distribution of Forman-Ricci curvature in gene co-expression networks. The analysis indicates the presence of three substructures, corresponding to a major hump at curvature-values of ~ 200 and smaller humps at ~ 80 and ~ 30.

For this, we analyzed the frequency of housekeeping genes in each of the three subnetworks with the following protocol: We identified the edges that contribute to each of the three humps and their vertices. From the set of vertices for each subnetwork, we selected the most frequent 100vertices. With the help of the *Ensembl* database [4] we identified the standard gene names corresponding to these vertices and compared them with a list of housekeeping genes by Eisenberg and Levanon [6]. This gave the following results:

- hump 1: 26/100 are housekeeping genes (26%)

- hump 2: 19/100 are housekeeping genes (19%)

- hump 3: 16/100 are housekeeping genes (16%)

In contrast, overall we find only 8.9% housekeeping genes with the same protocol, indicating that housekeeping genes occur with higher frequency in the three subnetworks.

These results provide further evidence for the conclusions in our first study, namely that curvature detects structurally important vertices corresponding to frequent elements in the underlying system. Moreover, the results indicate that curvature highlights the connectivity among those elements; a structural property described as the backbone effect [19].

References

[1] M. Bačak, B. B. Hua, J. Jost, M. Kell, A. Schikorra, A notion of nonpositive curvature for general metric spaces, *Diff. Geom. Appl.* **38** (2015) 22–32.

[2] F. Bauer, J. Jost, S. Liu, Ollivier-Ricci curvature and the spectrum of the normalized graph Laplace operator, *Math. Res. Lett.* **19** (2012) 1185-1205.

[3] *BrainSpan: Atlas of the Developing Human Brain*, 2011. Available from: http://developinghumanbrain.org. Funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01.

[4] F. Cunningham, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. Garca Girn, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, A. K. Khri, S. Keenan, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H. Singh Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, B. L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S. M. J. Searle, G. Spudich, S. J. Trevanion, A. Yates, D. R. Zerbino, P. Flicek, Ensembl 2015, *Nucleic Acids Res.* **43** (2015) D662–D669.

[5] N. C. Duarte, M. J. Herrgard, B. O. Palsson BO, Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14** (2004) 1298–1309.

[6] E. Eisenberg, E. Y. Levanon, Human housekeeping genes, revisited. *Trends Genet.* **29** (2013) 569–574.

[7] R. Forman, Bochner's method for cell complexes and combinatorial Ricci curvature, *Lect. Notes Comput. Sc.* **29** (2003) 323–374.

[8] J. Jost, *Riemannian Geometry and Geometric Analysis*, Springer, 2011.

[9] J. Jost, S. Liu, Ollivier's Ricci curvature, local clustering and curvature-dimension inequalities on graphs, *Lect. Notes Comput. Sci.* **51**(2) (2014) 300–322.

[10] J. Lohkamp, Metrics of negative Ricci curvature, *Ann. Math.* **140** (1994) 655–683.

[11] J. A. Miller, S. L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, A. Ebbert, Z. L. Riley, J. J. Royall, K. Aiona, J. M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S.

Butler, S. Caldejon, A. Carey, C. Cuhaciyan, R. A. Dalley, N. Dee, T. A. Dolbeare, B. A. C. Facer, D. Feng, T. P. Fliss, G. Gee, J. Goldy, L. Gourley, B. W. Gregor, G. Gu, R. E. Howard, J. M. Jochim, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, T. A. Lemon, P. Lesnar, B. McMurray, N. Mastan, N. Mosqueda, T. Naluai-Cecchini, N.-K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman, M. Sarreal, S. Shapouri, N. V. Shapovalova, E. H. Shen, N. Sjoquist, C. R. Slaughterbeck, M. Smith, A. J. Sodt, D. Williams, L. Zöllei, B. Fischl, M. B. Gerstein, D. H. Geschwind, I. A. Glass, M. J. Hawrylycz, R. F. Hevner, H. Huang, A. R. Jones, J. A. Knowles, P. Levitt, J. W. Phillips, N. Šestan, P. Wohnoutka, C. Dang, A. Bernard, J. G. Hohmann, E. S. Lein, Transcriptional landscape of the prenatal human brain, *Nature* **508** (2014) 199–206.

[12] Y. Ollivier, Ricci curvature of Markov chains on metric spaces, *J. Funct. Anal.* **256**(3) (2009) 81-864.

[13] Y. Ollivier, A survey of Ricci curvature for metric spaces and Markov chains, *Adv. Stu. P. M.* **57** (2010) 343–381.

[14] J. L. Reed, T. D. Vo, C. H. Schilling, B. O. Palsson, An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **4** (2003) R54.1–R54.12.

[15] B. Riemann, *Ueber die Hypothesen, welche der Geometrie zu Grunde liegen.* Edited with a commentary by J.Jost, Klassische Texte der Wissenschaft, Springer, Berlin, 2013; English version: B. Riemann, *On the hypotheses which lie at the bases of geometry.* Translated by W. K. Clifford, Edited with a commentary by J. Jost, Classic Texts in the Sciences, Birkhäuser, 2016.

[16] R. P. Sreejith, J. Gu, S. Liu, J. Jost, E. Saucan, A. Samal, Ollivier's vs. Forman's Ricci curvature for networks – a comparison, preprint, 2017.

[17] R. P. Sreejith, K. Mohanraj, J. Jost, E. Saucan, A. Samal, Forman curvature for complex networks, *J. Stat. Mech.* (2016) #P063206.

[18] R. P. Sreejith, J. Jost, E. Saucan, A. Samal, Forman curvature for directed networks, *arXiv:1605.04662.*

[19] M. Weber, E. Saucan, J. Jost, Characterizing complex networks with Forman-Ricci curvature and associated geometric flows, *J. Complex Net.* **5** (2017) 527-550.

[20] M. Weber, J. Jost, E. Saucan, Forman-Ricci flow for change detection in large dynamical data sets, *Axioms* **5** (2016) 26.

[21] M. Weber, E. Saucan, J. Jost, Coarse geometry of evolving networks, to appear in *J. Complex Net.* (2017).