



OPEN ACCESS

EDITED BY

Susan Mertins,
Leidos Biomedical Research, Inc.,
United States

REVIEWED BY

Samik Ghosh,
Systems Biology Institute, Japan
Sruvan Kumar Miryala,
Northeastern University, United States

*CORRESPONDENCE

Paola Lecca,
✉ Paola.Lecca@unibz.it

RECEIVED 05 June 2023

ACCEPTED 13 November 2023

PUBLISHED 24 November 2023

CITATION

Lecca P, Lombardi G, Latorre RV and Sorio C (2023), How the latent geometry of a biological network provides information on its dynamics: the case of the gene network of chronic myeloid leukaemia.
Front. Cell Dev. Biol. 11:1235116.
doi: 10.3389/fcell.2023.1235116

COPYRIGHT

© 2023 Lecca, Lombardi, Latorre and Sorio. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

How the latent geometry of a biological network provides information on its dynamics: the case of the gene network of chronic myeloid leukaemia

Paola Lecca^{1*}, Giulia Lombardi², Roberta Valeria Latorre³ and Claudio Sorio³

¹Faculty of Engineering, Free University of Bozen-Bolzano, Bolzano, Italy, ²Department of Mathematics, University of Trento, Trento, Italy, ³General Pathology Division, Department of Medicine, University of Verona, Verona, Italy

Background: The concept of the latent geometry of a network that can be represented as a graph has emerged from the classrooms of mathematicians and theoretical physicists to become an indispensable tool for determining the structural and dynamic properties of the network in many application areas, including contact networks, social networks, and especially biological networks. It is precisely latent geometry that we discuss in this article to show how the geometry of the metric space of the graph representing the network can influence its dynamics.

Methods: We considered the transcriptome network of the Chronic Myeloid Leukemia K562 cells. We modelled the gene network as a system of springs using a generalization of the Hooke's law to n -dimension ($n \geq 1$). We embedded the network, described by the matrix of spring's stiffnesses, in Euclidean, hyperbolic, and spherical metric spaces to determine which one of these metric spaces best approximates the network's latent geometry. We found that the gene network has hyperbolic latent geometry, and, based on this result, we proceeded to cluster the nodes according to their radial coordinate, that in this geometry represents the node popularity.

Results: Clustering according to radial coordinate in a hyperbolic metric space when the input to network embedding procedure is the matrix of the stiffnesses of the spring representing the edges, allowed to identify the most popular genes that are also centres of effective spreading and passage of information through the entire network and can therefore be considered the drivers of its dynamics.

Conclusion: The correct identification of the latent geometry of the network leads to experimentally confirmed clusters of genes drivers of the dynamics, and, because of this, it is a trustable mean to unveil important information on the dynamics of the network. Not considering the latent metric space of the network, or the assumption of a Euclidean space when this metric structure is not proven to be relevant to the network, especially for complex networks with hierarchical or modularised structure can lead to unreliable network analysis results.

KEYWORDS

network geometry, graph embedding, dynamical systems, spring systems, chronic myeloid leukaemia, systems biology

1 Introduction

With the emergence of systems biology around the year 2000, the representation of a system of interacting biological entities, such as proteins, molecules, functional complexes, *etc.*, in the form of a network or graph has become preponderant and an unreliable prerequisite of any mathematical model regarding both the static and dynamic properties of the network. This representation of the components of a system as network nodes and their interactions as arcs between the nodes proved to be easy to understand as it is intuitive and also an excellent tool for organising data. However, the immediacy of understanding such a representation comes at the price of its low informational power, its susceptibility to misinterpretation and its use that often takes place under tacit or even unconscious assumptions. Particularly in the graph representation of a network, it is natural to think of the concept of distance between nodes as the number of arcs separating the nodes, or, if the weights of the arcs are known, as the weighted sum of the number of arcs separating the nodes. In doing so, it is implicitly assumed that the distance between two nodes is a Euclidean distance, or, in other words, that the metric space in which the network resides is flat Euclid space. This implicit assumption on a measure as important as the distance between nodes, used in multiple contexts as a measure of the intensity of an interaction between nodes, if not of the propensity of the interaction itself, may not only be reductive or approximate, but may even be incorrect. An erroneous assumption about the metric space that represents the geometry of the network carries serious risks, one of which is that of not being able to grasp the organizational principles of the typology and consequently the dynamics of the network. Indeed, the distribution of widely used centrality metrics like as degree and clustering coefficient reflect the features of the metric space, which defines the network's geometry. For example, heterogeneous degree distributions and significant clustering emerge naturally as reflections of the underlying hyperbolic geometry's negative curvature and hyperbolic metric characteristic (Krioukov et al., 2010). On the opposite, if a network has some metric structure and a heterogeneous degree distribution, the network has an effective hyperbolic geometry below (Krioukov et al., 2010).

It is often said to indicate the metric space of a network that the graph representing the network is "embedded" in a metric space, which is called the latent geometry of the network. The adjective "latent" is justified by the fact that the graph representation of a network does not make visible the characteristics of the metric space in which the coordinates of the nodes are actually defined. The verb "to embed", on the other hand, although commonly used, we condemn somewhat misleadingly, since the network, if endowed with a metric structure, is in fact not embedded in a metric space as if it were a distinct entity that fits into it, but is itself a portion of it, more precisely a discrete version of the continuous metric space that represents it. The use of the verb "to embed" stems from the procedures dedicated to understanding what the latent geometry of the network might be and based on tests in which the network is considered to have metrics of a different nature and then the distortion that the new metric has with respect to the original metric defined by the network's similarity matrix (i.e., weighted adjacency matrix) is assessed.

The latent geometry of a network is an important area of study in network science. We refer the reader to Boguñá et al. (2021); Jhun (2022) or an overview of the studies and fields of application of the study of the latent geometry of a network. In Jhun (2022), it is reported that latent geometry has been used to travel networks efficiently (Kleinberg,

2000; Boguñá et al., 2010) detect missing links (Liben-Nowell and Kleinberg, 2007; Clauset et al., 2008), map the brain (Allard and Serrano, 2020), and analyse proximity network (Papadopoulos and Flores, 2019). Interestingly, it has been shown that the map of contagions of various pandemics develops through paths defined on the latent geometry of the network of contacts and movements of individuals (Taylor et al., 2015). Systems biology has also benefited from the results of latent network geometry analysis, in particular the study of genetic networks and protein-protein interaction networks as reported in (Alanis-Lobato et al., 2018; Härtner et al., 2018; Pio et al., 2019; Klimovskaia et al., 2020; Sun et al., 2021; Lecca and Re, 2022; Lecca, 2023; Seyboldt et al., 2022).

While we can say that the relationship between latent geometry and static topological properties of the network, such as those measured by the centrality indices, is well established, the relationship between latent geometry and network dynamical properties is little investigated. A recent attempt in this direction was made by Rand et al. (2021). In this paper, Rand et al. study embryonic development. From egg to adult, embryonic development results in the reproducible and organised manifestation of complexity. In this process, the activity of gene networks culminates in the sequential differentiation of distinct cell types that construct this complexity, which has been likened by Conrad Waddington metaphor (Fard et al., 2016; Squier, 2017; Sánchez-Romero and Casadesús, 2021) to a flow through a landscape with valleys representing alternative destinies. Geometric approaches enable the formal description of such landscapes and codify the types of behaviours produced by differential equation systems.

With this study of ours, we wish to make a contribution in this still very unexplored field of the relations between latent geometry and the evolution of a network, particularly a biological network. We propose a method to infer the equations governing the dynamics of a network of genes previously identified by the authors (Lombardi et al., 2022) as involved in the development and progression of Chronic Myeloid Leukaemia (CML). The method consists of two steps: i) the determination of the latent geometry of the network through embedding of the network in three models of metric space (Euclidean, hyperbolic, and spherical), and ii) the determination of the dynamic equations describing this metric space. If the result of the step i) is the hyperbolic metric, the parameter of the dynamics of the interactions in the network conceived as a subspace of a hyperbolic space will depend on the hyperbolic distance between the interacting partners. Similarly, if the result of step i) is a spherical metric, the dynamics of the network will be parametrized by distance of the interacting nodes in the spherical space, and finally, if the result of step i) is an Euclidean metric, the network will be a dynamical systems whose parameters will depend on Euclidean distance between the interacting nodes.

In this study, we conceive of a network as a system of springs, in which the nodes constitute the masses and the arcs the springs that connect these masses/nodes. The spring constant represents the transmission efficiency of the interaction between the nodes. The interaction between a node A and a node B is seen as a change in the state of A causing a change in B. In accordance with the spring model, the interaction between nodes is seen as a propagation of the alteration of A's state through the spring to B, which absorbs the alteration of A in turn changing its state. The vibrational states of the networks nodes are governed by a generalization of the Hook law. According to this law, the spring constant is calculated by dividing the force required to stretch or compress a spring by the lengthening or

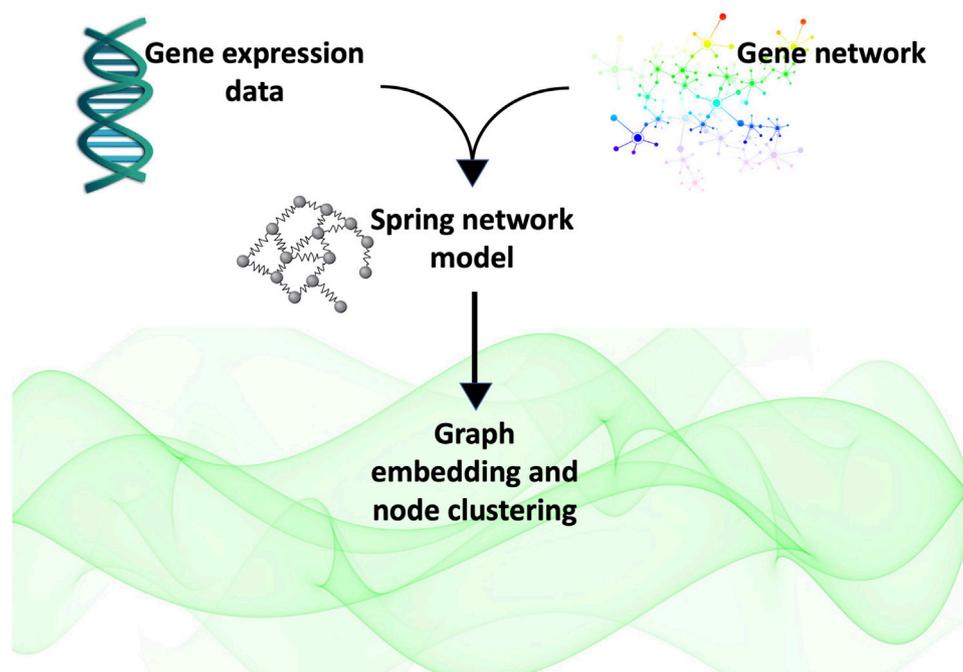


FIGURE 1

In this study, we obtained the gene network of interest by querying the Pathway Commons database with the list of genes of interest. We represented the network as a system of springs whose masses are the expression level of the genes as measured in our experiments in (Lombardi et al., 2022). We calculated the weighted adjacency matrix of the network as that matrix whose entries are given by the spring constant calculated at equilibrium. Finally, we used this matrix to embed the graph into three spaces (flat, positively curved and negatively curved)space) in order to determine which of them best represented the network's latent geometry. Finding the hyperbolic space fits best the latent geometry of the network, we proceeded to cluster the nodes according to their radial coordinate, that representative of the node popularity (Papadopoulos et al., 2012).

shortening of the spring. It is stated mathematically as $k = -F/\Delta x$, where Δx is the displacement of the mass, F is the force applied over x , and k is the spring constant (also known as *spring stiffness*). The propagation velocity of the elastic wave in a spring stressed by a force is directly proportional to the square root of the spring's elastic constant. A stiffer spring has a greater spring stiffness, and *vice versa*. As a consequence, a high spring stiffness is interpreted as high efficiency and thus greater ease in the transmission of interaction between nodes. The elastic constant metaphor, in network metric space, corresponds to a measure of similarity between nodes, such that nodes connected by harder springs are closer nodes in terms of similarity. In the model of network we present here, the elastic constants of the springs are obtained from a generalization of the Hook's law for a system with N masses and E springs (N corresponding to the number of nodes and E corresponding to the number of edges), where the mass of the node is given by its total degree and the change in the position of the node is given by the index of vibrational centrality proposed by Estrada and Hatano (2010). The matrix of elastic constants is used in network embedding procedures in three types of space, Euclidean, hyperbolic, and spherical. The metric space for which the embedding of the network shows a minimum distortion of the values of this matrix is considered as the best approximation for the metric space of the network. The distances of the nodes in this metric space constitute the parameters of the network dynamics, which we describe here in terms of mass action law.

The article is organised as follows: in Section 2 we introduce the three types of isotropic spaces considered in this study and the embedding techniques we used to identify which of the metric

spaces considered best represents the network's latent geometry. In Section 3 we describe the data and the gene network of the case study. In Section 4, we describe the mathematical modelling of the gene network as a system of springs, and finally in Section 5 we report the results obtained. This is followed by some concluding remarks and a recapitulation of the study performed (Section 6). In Figure 1 we illustrate the main steps of the analysis presented in this study.

2 Network geometry and methods of embedding

A graph embedding consists in the determination of the coordinates of the graph node in a given metric space in such a way that the graph similarity matrix is reproduced with as little error as possible. The embedding of a graph thus consists of the problem of finding the coordinates of the nodes in a given metric space from the similarity matrix of the graph, which is a measure of the distances between nodes. In the final analysis, embedding consists of finding coordinates of points given their distance. Isotropic spaces can only be classified as Euclidean (flat), elliptic (having positive curved), or hyperbolic (having negative curvature). In the following sections, we will recall some basic definitions, such as that of inner product and distance for these three types of spaces, and briefly mention the mathematical techniques of graph embedding, of which there are many variants in literature. We

also recall how the latent geometry is related to the structure and organizational principles of the network (e.g., presence of communities, hierarchical organization, etc.).

2.1 Euclidean space

The Euclidean geometry is based on the following five postulates: (i) Any two points can be joined by a straight line segment. (ii) Any portion of a straight line can be stretched forever. (iii) Any straight line segment can be used as the radius of a circle, with one endpoint serving as the centre. (iv) All right angles are congruent. (v) When two lines are drawn so that they intersect a third in a fashion that results in a side where the total of the inner angles is less than two right angles, the two lines will always cross each other if they are extended far enough.

More formally, an Euclidean space, is a real vector space (i.e., a vector space whose field of scalars is \mathbb{R}) E equipped with a positive definite symmetric bilinear form $\varphi: E \times E \rightarrow \mathbb{R}$. The real number $\varphi(x, y)$ is called the *inner product* between the vectors $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$, that is defined as

$$\begin{aligned} \varphi(\mathbf{x}, \mathbf{y}) &= (x_1, x_2, \dots, x_n) \cdot (y_1, y_2, \dots, y_n) \\ &= x_1y_1 + x_2y_2 + \dots + x_ny_n. \end{aligned} \tag{1}$$

Usually the inner product of two vectors \mathbf{x}, \mathbf{y} is denoted with the angular bracket $\langle \mathbf{x}, \mathbf{y} \rangle$. In the Euclidean space \mathbb{R}^n the distance between the points whose coordinates are given by the vectors \mathbf{x} and \mathbf{y} is

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \sqrt{\langle \mathbf{x}, \mathbf{y} \rangle} = \sqrt{\varphi(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x})} \\ &= \sqrt{(y_1 - x_1)(y_2 - x_2) \dots (y_n - x_n)} \equiv \|\mathbf{y} - \mathbf{x}\|. \end{aligned} \tag{2}$$

To embed a graph into an Euclidean space, we used the classical (metric) multidimensional scaling algorithm that, given as an input the pairwise dissimilarities matrix $\{d_{ij}\}$, reconstructs a map that preserves distances. The algorithm implements the following steps.

1. Find a random arrangement of points, for example, by taking a sample from a normal distribution.
2. Determine the distances between the points.
3. Find the best monotonic transformation for the proximity to get the best scaled data.
4. Find a new arrangement of points to reduce the stress between the optimally scaled data and the distances. The stress of the embedding in Euclidean space is defined by the following residual sum of squares

$$\text{Stress}(x_1, x_2, \dots, x_n) = \sqrt{\sum_{i \neq j=1, \dots, n} \left(d_{ij}^{(\text{input})} - d_{ij}^{(\text{embedding})} \right)^2} \tag{3}$$

where, in the case of Euclidean embedding, $d_{ij}^{(\text{embedding})} = \|x_i - x_j\|$.

5. Compare the stress to a certain standard. If the stress is too low, stop the algorithm; otherwise, go back to step 2.

We implemented the embedding in R (R Core Team, 2021), using the function `cmdscale` (Gower, 2015) of the library `stats`.

Theoretical foundations and details about multidimensional scaling techniques can be found in many text books and review paper [see, for example, (Borg and Groenen, 2005; Cox and Cox, 2008; Zhang and Takane, 2010)].

2.2 Hyperbolic geometry and the Poincaré model

Hyperbolic geometry accepts the first four axioms of Euclidean geometry but rejects the fifth, namely, that there exists a line and a point not on the line with at least two parallels to the given line crossing through the provided point. This is equivalent to performing geometry on a surface with a constant negative curvature. This geometry differs greatly from the more conventional Euclidean geometry, and are hard to visualise. The main reason is that by the Hilbert's theorem (Hilbert, 1933) the hyperbolic plane cannot be isometrically embedded into Euclidean 3D-space (isometric means preserving the length of every curve). We must flatten the curvature to display the hyperbolic plane. In doing this, many of the straight lines in hyperbolic space end up being curved as a result. The French mathematician Henri Poincaré is responsible for one of the widely accepted theories for flattening the hyperbolic plane and the n -dimensional ball model (Poincaré disk in 2D) (Anderson, 2005).

The Poincaré n -dimensional ball $\mathbb{B}_{\mathbb{R}}^n$ ($\mathbb{B}_{\mathbb{R}}^n = \{\mathbf{x} \mid \|\mathbf{x}\|^2 < 1\}$) is a model for n -dimensional hyperbolic geometry in which lines are represented by circle diameters or by arcs of a circle with ends perpendicular to the boundary of the ball. (Figure 2). If $n = 2$ the Poincaré model is a unit open disc. We briefly summarize here the method in Conn (2010) to calculate the distances in the unit disc model. Consider the fractional linear transformation S that sends $\infty \mapsto i$ and $\pm 1 \mapsto \pm 1$. S sends the real axis to the boundary of the unit disc and, since fractional linear transformations preserve the orientation of circles, it sends the upper half-plane to the disc's interior. The H_2 -distance between two points a, b in the unit disc is the H_1 -distance between their preimages $S^{-1}(a), S^{-1}(b)$ in the upper half-plane (Conn, 2010), and in this way the unit disc inherits a metric from the metric of the upper half-plane.

Let D^1 denote the interior of the unit disc and suppose

$$\gamma: [0, 1] \rightarrow D^1$$

is a piecewise continuously differentiable curve. If $H_k(\gamma)$ ($k = 1, 2$) denotes the length of the curve γ , then

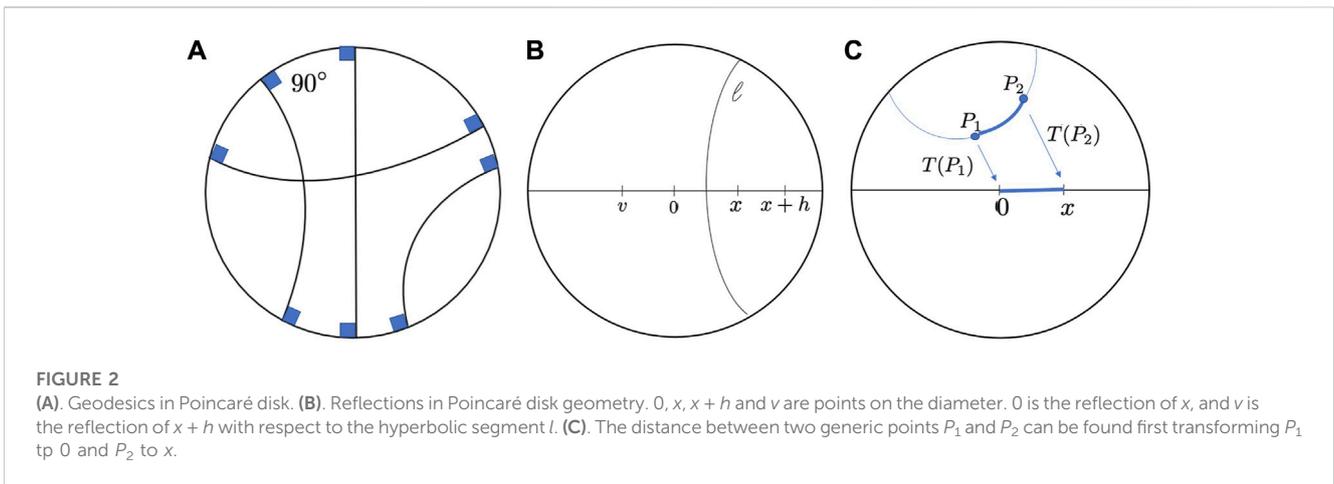
$$H_2(\gamma) = H_1(S^{-1} \circ \gamma). \tag{4}$$

Writing $S^{-1} \equiv T$, we have

$$\begin{aligned} H_1(T \circ \gamma) &= \int_{T \circ \gamma} \frac{1}{\text{Im}(z)} |dz| = \int_0^1 \frac{1}{\text{Im}((T \circ \gamma)(t))} |(T \circ \gamma)'(t)| dt \\ &= \int_0^1 \frac{1}{\text{Im}(T(\gamma(t)))} |T'(\gamma(t))| |\gamma'(t)| dt \\ &= \int_{\gamma} \frac{1}{\text{Im}(T(z))} |T'(z)| |dz|. \end{aligned} \tag{5}$$

Since T has the form

$$T(z) = \frac{iz - 1}{-z + i}. \tag{6}$$



we have that

$$\text{Im}(T(z)) = \text{Im}\left(\frac{(iz - 1)(-\bar{z} - i)}{|-z + i|^2}\right) = \frac{1 - |z|^2}{|-z + i|^2}, \tag{7}$$

and

$$|T'(z)| = \frac{2}{|-z + i|^2}, \tag{8}$$

we obtain

$$H_2(y) = \int_y \frac{2}{1 - |z|^2} |dz| \tag{9}$$

that is general formula calculating distances in the Poincaré disc. $2/(1 - |z|^2)dz$ is the element of arc length. Consequently, the distance between two points $a, b \in \mathbb{C}$ on Poincaré disc is

$$d(a, b) = H_2(a, b) = \log \frac{|1 - \bar{a}b| + |b - a|}{|1 - \bar{a}b| - |b - a|}. \tag{10}$$

Any diameter of the unit disc is a geodesic, so if z is a point in the unit disc, then the Euclidean segment from 0 to z is also a hyperbolic segment from 0 to z . We have hence that

$$H_2(0, z) = \int_0^{|z|} \frac{2}{1 - t^2} dt = 2 \tanh^{-1}(|z|) = \log\left(\frac{1 + |z|}{1 - |z|}\right). \tag{11}$$

Complex networks connect different nodes. This diversity indicates that there is at least some element taxonomy, meaning that all nodes can be classified in some way. This classification means that nodes can be separated into large groups that are made up of smaller subgroups that are made up of even smaller sub-subgroups, and so on. The relationships between such groups and subgroups can be approximated by treelike structures, which illustrate hidden hierarchies in networks. Krioukov et al. demonstrated that the metric structures of trees and hyperbolic spaces are equivalent (Krioukov et al., 2010; Kurkofka et al., 2021; Lecca, 2023; Lecca and Re, 2022). It is not necessary for the node classification hierarchy to be exactly a tree, but rather approximately a tree. When a network can be approximated by a tree, its latent geometry is negatively curved (Gromov, 2007).

To perform the embedding into a hyperbolic space (Poincaré model), we used the function `hydraPlus` of the R library `hydra` (HYperbolic Distance Recovery and Approximation) (Keller-Ressel,

2019), that uses a strain-minimizing hyperbolic embedding based on reduced matrix eigendecomposition (Keller-Ressel and Nargang, 2020). The stress of embedding in hyperbolic space is then given by formula (3), where $d^{(\text{embedding})}$ is given by the output of `hydra`.

2.3 Spherical geometry and embedding

Spherical geometry is the geometry of a hypersphere's surface. The hypersphere can be easily immersed in euclidean space; for example, the embedding of a three-dimensional sphere of radius r is well known relation $x^2 + y^2 + z^2 = r^2$, with $\mathbf{x} = (r \sin u \sin v, r \cos u \sin v, r \cos v)^T$. A simple extension of this is the embedding of a $(n - 1)$ -dimensional sphere in n -dimensional space:

$$\sum_{i=1}^n x_i^2 = r^2. \tag{12}$$

There is a constant sectional curvature of $1/r^2$ throughout this curved surface. The length of the shortest curve that lies in the space and connects the two points is the geodesic distance between two points in a curved space. The geodesic on the hypersphere is a perfect circle for a spherical space. The distance is equal to the width of the arc that connects the two locations on the great circle.

If two points in the hypersphere's centre form an angle with θ_{ij} , then the distance between them is

$$d_{ij} = r\theta_{ij}. \tag{13}$$

A point can be represented by a position vector \mathbf{x}_i of length r with the coordinate origin at the origin of the hypersphere. We can also write

$$d_{ij} = r \cosh \frac{\langle \mathbf{x}_i, \mathbf{x}_j \rangle}{r^2} \tag{14}$$

since the inner product is $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = r^2 \cos \theta_{ij}$.

To perform the embedding of the graph in a hypersphere, we used the method proposed by Wilson et al. (2014a), and the Matlab code that this authors made available in Wilson et al. (2014b). We summary briefly the core of embedding method in this way.

Given a dissimilarity matrix D , we want to determine the set of points on a hypersphere that give the same distance matrix. Because the curvature of the space is unknown, we must also determine the

radius of the hypersphere. We have n items of interest, thus we would ordinarily look for a $n - 1$ dimensional Euclidean space. A coordinate system with the origin at the centre of the hypersphere is considered. A matrix \mathbf{X} of point positions vectors is constructed in such a way that

$$\mathbf{X}\mathbf{X}^T = \mathbf{Z} = \{z_{ij}\} = r^2 \cos \frac{d_{ij}}{r} \quad (15)$$

\mathbf{Z} is a $n \times n$ matrix that is positive semi-definite and has rank $n - 1$ since the embedding space has dimension $n - 1$, \mathbf{X} is made up of n points that are located in a space of dimension $n - 1$, and so does the embedding space. This means that the \mathbf{Z} 's eigenvalues are positive, with only one being zero. This observation can be used to calculate the radius of curvature. Then, in order to find r , Wilson et al. (2014a) proposed to create $\mathbf{Z}(r)$ and identify the smallest eigenvalue λ_1 , to calculate then the optima radius of curvature as

$$r^* = \arg \min_r |\lambda_1[\mathbf{Z}(r)]|. \quad (16)$$

The stress of embedding in hyperbolic space is then given by formula (3), where $d^{(\text{embedding})}$ is given by the elements of the matrix \mathbf{Z} .

2.4 How latent geometry influences network dynamics

By the term “dynamic” of a system, we mean the *time and space* evolution of the system as described by differential and/or algebraic equations whose variables are quantitative features of the system's actors, and whose mathematical form model the topological system's organization. The equations of the dynamics are parameterized by the dynamical properties of the system itself (such as frequency of oscillation, if the system is oscillatory, elastic constant, if the system is assimilated to a spring system, specific rate of reaction, etc.) There are interesting studies showing how the geometry of complex networks affects the dynamics. To cite a relevant contribution to the field, we mention the work of Millán et al. (2018) which shows that the latent geometry of a network has a significant impact on the synchronization dynamics. Unlike Millán et al. work, which is more focused on the dynamic properties of the system (i.e., parameters and synchronization laws), here we focus on the influence that latent geometry can have on network organization. And since from the network organization, the dynamics of the network is derived, we can expect latent geometry to influence the dynamics. In particular, the geometry of the network determines the presence or absence of functional modules containing highly cooperative nodes. The identification of these possible functional clusters can be done correctly only if the metric space of the network is identified. In fact, this space defines the distance between nodes, the measure on which clustering algorithms are based. A clustering in Euclidean space may lead to a different result from clustering in hyperbolic space, the distance computed in this space being different from the distance computed in Euclidean space. The correct dynamics is one whose parameters and functional modules are established by the latent geometry for at the network under consideration.

In this study, we conceived a network as a spring system. Through the identification of the most appropriate latent geometry of the network under consideration, i.e., that geometry that most closely reproduces the values of the spring constants of the edges thought of as springs, we were able to identify cluster of gene drivers for the network dynamics. The role of drivers of these genes was validated through functional analysis of them. In the next sections, the data from which we built the network, as well as the model and analysis of the network itself are reported.

3 Data and gene network

We use here the data of gene expression relevant to the landscape of Chronic Myeloid Leukemia K562 cells. We refer the reader to a recent publication by the authors (Lombardi et al., 2022)], where we describe the experimental activity implemented for data measurement and algorithmic procedures for selecting differentially expressed genes. For the reader's convenience we summarise it briefly below.

On an Agilent whole human genome oligo microarray (#G4851A, Agilent Technologies, Palo Alto, CA), the RNAs from the samples were hybridised. This microarray consists of 60,000 distinct human transcripts represented by 60-mer DNA probes created using SurePrint technology. The manufacturer's recommended protocol was followed when one-color gene expression was carried out. In a nutshell, samples were used to extract the total RNA fraction using the Trizol Reagent (Invitrogen). Agilent Technologies' Agilent 2100 Bioanalyzer was used to evaluate the quality of the RNA samples. RNAs with low integrity (RNA integrity number less than 7) were not included in the microarray analysis. Using the Low Input Quick-Amp Labelling Kit, one colour (Agilent Technologies) in the presence of cyanine 3-CTP, labelled cRNA was produced from 100 ng of total RNA. In a revolving oven, hybridizations were carried out for 17 h at 65°C. Agilent's scanner produced images with a 3 μm resolution, and Agilent Technologies' Feature Extraction 10.7.3.1 software was utilised to extract the microarray raw data. The GeneSpring GX 11 programme (Agilent Technologies) was then used to analyse the microarray results. Data transformation was used to normalise all of the data's negative raw values to 1.0 using the 75th percentile. Only the probes expressed in at least one sample (marked as Marginal or Present) were retained using a filter on low gene expression.

The data used in this work come from the aforementioned examination of the CML cell transcriptome (K562) using microarray hybridization under various settings. The cells were transfected with full-length PTPRG and compared to three controls: cells transfected with the empty vector, cells transfected with a PTPRG inactive mutant with a mutation in the catalytic domain (D1028A), and cells treated with Imatinib, which targets the oncogene BCR/ABL1. The complete dataset is publicly available at the GitLab repository. <https://gitlab.inf.unibz.it/Paola.Lecca/chronic-myeloid-leukemia-genes>.

Here, from the entire dataset available at this link, we only considered the gene expression levels of the untreated group (empty vector and inactive mutant domain D1028A) and those of the treatment group expressing full-length PTPRG. We then selected the genes, that, according to the analysis in Lombardi et al. (2022),

result to be differentially expressed between the two groups. To construct the gene network, we queried PathwaysCommons (PathwayCommons.All.hgnc repository) (Cerami et al., 2010a; Cerami et al., 2010b) by providing as input for the search the list of gene names we considered in this study. The obtained gene network is a representation of molecular associations specified through nodes (genes) and edges (molecular interactions or statistical relationships). Among the various format, PathwaysCommons gives as output result of the query the gene networks also in SIF (Simple Interaction format), which is a table providing details on gene-gene interactions. This format offers various levels of detail such as: interaction type, reference data source, Pubmed id, reference pathways, and mediators id. Our analyses focused on the most granular level of information, namely, the interactions between pairs of genes, listed in the SIF table as “Participant A” and “Participant B” (we refer the reader to the public repository of our data to view the data format). The types of interactions included in the network are as follows: interacts-with, in-complex-with, catalysis-precedes, controls-state-change-of, controls-transport-of, controls-transport-of-chemical, controls-expression-of, controls-phosphorylation-of, controls-production-of, chemical-affects, consumption-controlled-by.

As a final result of querying to common pathways and selecting differentially expressed genes on the two groups (treated and untreated), we obtained a network that is a non-planar multi-edge graph with 2,080 nodes and 3,745 edges, that we simplify to a non-multi-edge graph with 2,080 nodes and 3,464 edges.

3.1 Graph embedding in presence of noise of input data: Some remarks

The presence of noise on the data in the adjacency matrix used as input to the graph embedding procedures could be a vexing problem if embedding stresses in different metric spaces are to be compared to identify which metric space is best representative of the latent geometry of the network. Noise, for example, may not allow weak edges to be distinguished from the absence of nodes and may affect the reliability of the measurement of even the most robust arcs (i.e., those with the greatest weight). Data analysis frequently faces the challenge of distinguishing between real weak edges and noise-induced low-weight edges. To solve this issue, noise is typically either eliminated or studied in the absence of data.

In the specific case of our study, the experimental data from which we start to construct the weighted adjacency matrix of the graph are very accurate. Our dataset was validated comparing the outcome of the cDNA microarray with the analysis of a specific set of genes chosen for being informative and for being predicted up and downregulated. Validation was performed in triplicate with quantitative PCR on a new, independent, preparation of cDNA derived from the same cell lines, thus ensuring that the results present in our dataset represent a true variation in mRNA levels. Notably the analysis permitted to predict a shift to erythroid differentiation of the cells that was confirmed also at protein level. All supporting data are reported on the publication (Lombardi et al., 2022).

Interesting and noteworthy works elucidating the role and the influence of noise in graph embedding has been done recently by

Maddalena et al. (2022) and Okuno and Shimodaira (2019). The treatment of the presence of noise is in fact so complex that it deserves the implementation of a focused study and consequently the writing of a separate article. It is out of the scope of this study, give the high quality of the data we used here, but it is our intention to explore this issue further in a forthcoming study.

To the best of our knowledge at present, we find of particular interest the study of Blevins et al. (2021). Instead, by analysing the structure of noisy, weak edges that have been artificially added to model networks, the authors explored how noise and data coexist in this work. They discovered that there are qualitative classifications of noise structure that arise, and that these noisy edges can be used to categorise the model networks. The authors state that the structure of low-weight, noisy edges varies depending on the topology of the model network to which they are added. Interestingly, Blevins et al. showed that noise is a complex, topology-dependent, and even valuable phenomenon in characterising higher-order network interactions rather than a monolithic annoyance.

4 Mathematical model of the gene network

To estimate the weights of the network arcs, we conceptualise the network as a system of masses (representing the nodes) and springs (representing the edges), as in Figure 3. Estrada and Hatano (2010) has brought a remarkable contribution to spring-like network models. In a complex network, Estrada and Hatano suggested a new metric for measuring node vulnerability. The metric is based on an analogy where the network's nodes are represented by masses and its edges by springs. They defined the measure as the node displacement, or the amplitude of vibration of each node, under variation caused by the thermal bath in which the network is intended to be immersed, and that represents the environment from which stimuli may possibly come. The Estrada index for the vibrational centrality of the node i is defined as the node displacement $(\Delta x)_{ii}$

$$(\Delta x)_{ii} = \sqrt{\frac{T}{k} \mathbf{L}_{ii}^+}, \quad (17)$$

where T is the temperature of the external bath, and k is the spring stiffness. Estrada and Hatano assumed that the network edges are identified with springs with a common spring stiffness k .

Instead in the network spring model of Lecca and Re (2020), the authors postulated that weights of the arcs are given by the stiffness of the springs representing the arcs, so each arc may have a different stiffness/weight. The harder the spring, the more efficiently the signal is transmitted from node to node; the softer the spring, the less quickly the signal is transmitted from node to node. According to this metaphor, edges characterised by high values of the stiffness of the hypothetical spring joining them are nodes that interact more effectively than nodes whose spring stiffness joining them is lower. The stiffness of the spring is thus interpreted as the efficiency of the interaction. Next, we briefly summarize the computational method developed by Lecca and Re (2020), and used in this study, to calculate the stiffnesses of the springs.

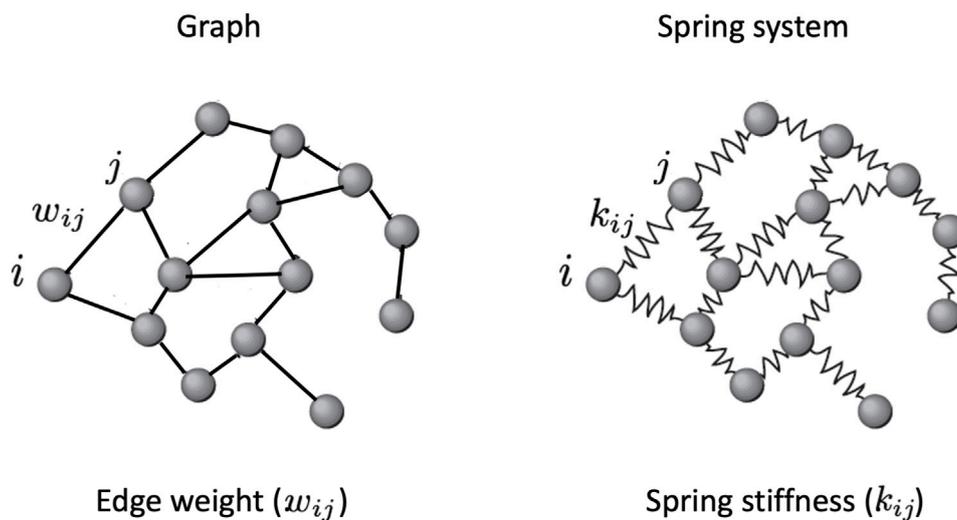


FIGURE 3
 A spring system, also known as a spring network, is a model of physics used in engineering and physics that is represented as a graph with a mass at each vertex and a spring with a specific stiffness and length along each edge. Extending the Hooke's law to higher dimensions (see the Section ?), it is possible to calculate the spring stiffnesses that in this model represent the edge weights.

The elastic force applied to the nodes by the springs according to the generalised Hooke's law for a system of N springs is

$$F_{\text{elastic}} = -\mathbf{K}\Delta\mathbf{x}, \tag{18}$$

where \mathbf{K} is the matrix providing the stiffnesses of all of the springs, and the elements of $\Delta\mathbf{x}$ are the vibrational centralities of the nodes. We obtain the force on the nodes by multiplying F_{elastic} by the transpose of the graph weighted incidence matrix \mathbf{C}^T , where, in general, the weights are given by the node masses, i.e.,

$$\mathbf{C} = \mathbf{A}\mathbf{M} \tag{19}$$

where \mathbf{A} is the unweighted incidence matrix. We should remark that weighting the incidence matrix with node mass values means taking into consideration the nodes' inertia to the propagation of the elastic force through the springs incident to them (Lecca and Re, 2020). In this study, the nodes' masses are given by the nodes' total degree.

The force on node is then defined by

$$F_{\text{nodes}} = -\mathbf{C}^T\mathbf{K}\Delta\mathbf{x} \tag{20}$$

At the equilibrium $F_{\text{nodes}} = \mathbf{0}$, i.e.,

$$\mathbf{C}^T\mathbf{K}\Delta\mathbf{x} = \mathbf{0}, \tag{21}$$

where \mathbf{K} is obtained as the nullspace (or kernel) of \mathbf{C}^T , in formula:

$$\mathbf{K} = \text{Ker}(\mathbf{C}^T). \tag{22}$$

Indeed, all vectors \mathbf{K} that have the properties that $\mathbf{C}^T\mathbf{K} = \mathbf{0}$ and \mathbf{K} are not zero make up the null space of any matrix \mathbf{C}^T .

Once \mathbf{K} is obtained, we construct the dissimilarity matrix of the graph, which is then used as input for the embedding algorithms, as follows

$$d_{ij} = \frac{1}{1 + k_{ij}} \tag{23}$$

where k_{ij} are the elements of the matrix \mathbf{K} . Thus, nodes connected by a spring with a high value of the elastic constant have a lower dissimilarity value than nodes connected by a spring with a low value of the elastic constant. This reflects the situation where the propagation speed of the interaction along a spring with high stiffness is higher than along a spring with low stiffness.

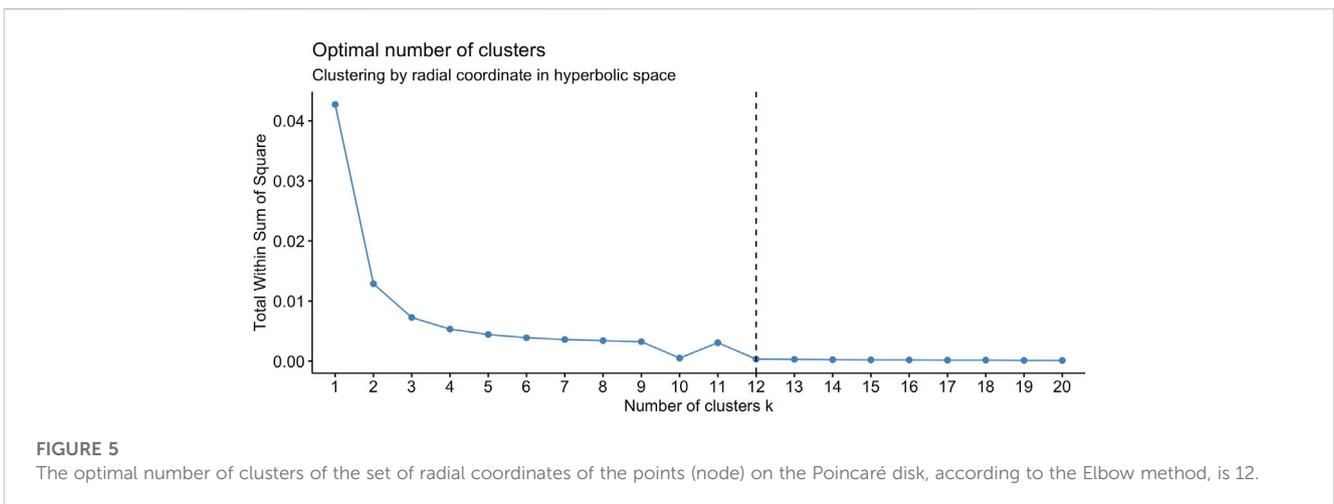
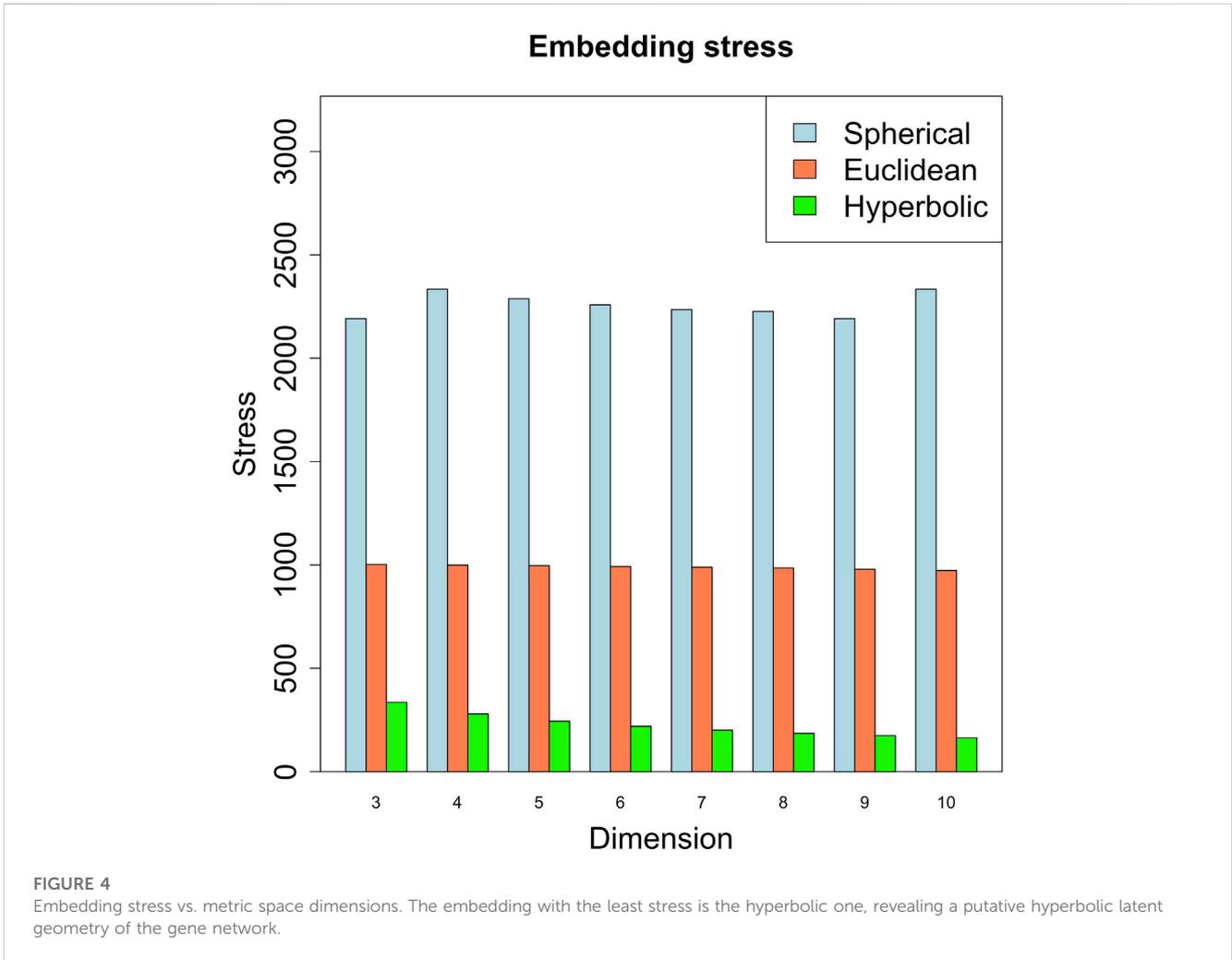
Of particular interest is in case the system is not in equilibrium. In fact, \mathbf{K} is independent on $\Delta\mathbf{x}$ only when the system is at equilibrium, i.e., when $F_{\text{nodes}} = \mathbf{0}$ and Eq. 21. In non-equilibrium conditions, we have instead that $F_{\text{nodes}} = \mathbf{C}^T\mathbf{K}\Delta\mathbf{x} \neq \mathbf{0}$. Suppose that we know the forces F_{nodes} acting on nodes. For example, this could be the case in which perturbation experiments are implemented to measure and analyse the responsiveness of the network nodes to stimuli and/or stresses, or, assimilating forces on nodes to white noise distributed over all nodes, noise always present in biological systems at the micro-scale given their inherent stochastic dynamics). To calculate the matrix \mathbf{K} , in this case, the requirements are that the matrices \mathbf{C}^T and $\Delta\mathbf{x}$ are invertible, so that

$$\mathbf{K} = (\mathbf{C}^T)^{-1}F_{\text{nodes}}(\Delta\mathbf{x})^{-1}. \tag{24}$$

Note, $\Delta\mathbf{x}$ is invertible if and only if all the entries on its main diagonal are non-zero, which means that little to much all nodes have a significantly non-zero response to stress.

5 Results

We embedded the gene network in the three metric spaces considered by considering different dimension values. We started with dimension 3, since the network is not planar. As shown in Figure 4, the embedding that produces the least amount of stress on the dissimilarity matrix - obtained as in Section 4 - is the hyperbolic embedding. The network is then



characterized by a power-law degree, and by a hierarchical structure reflected also in the presence of clusters in the radial coordinates of the points, that is known to represent the node popularity (Papadopoulos et al., 2012; Yang and Rideout, 2020; Kovács and Palla, 2021). By nodes having high

popularity, we mean nodes that are related to the majority of the other nodes in the graph [see also (Lecca et al., 2023) for a short review of the Papadopoulos et al. definition of node popularity]. These nodes can aid in the efficient spreading of information throughout the network.

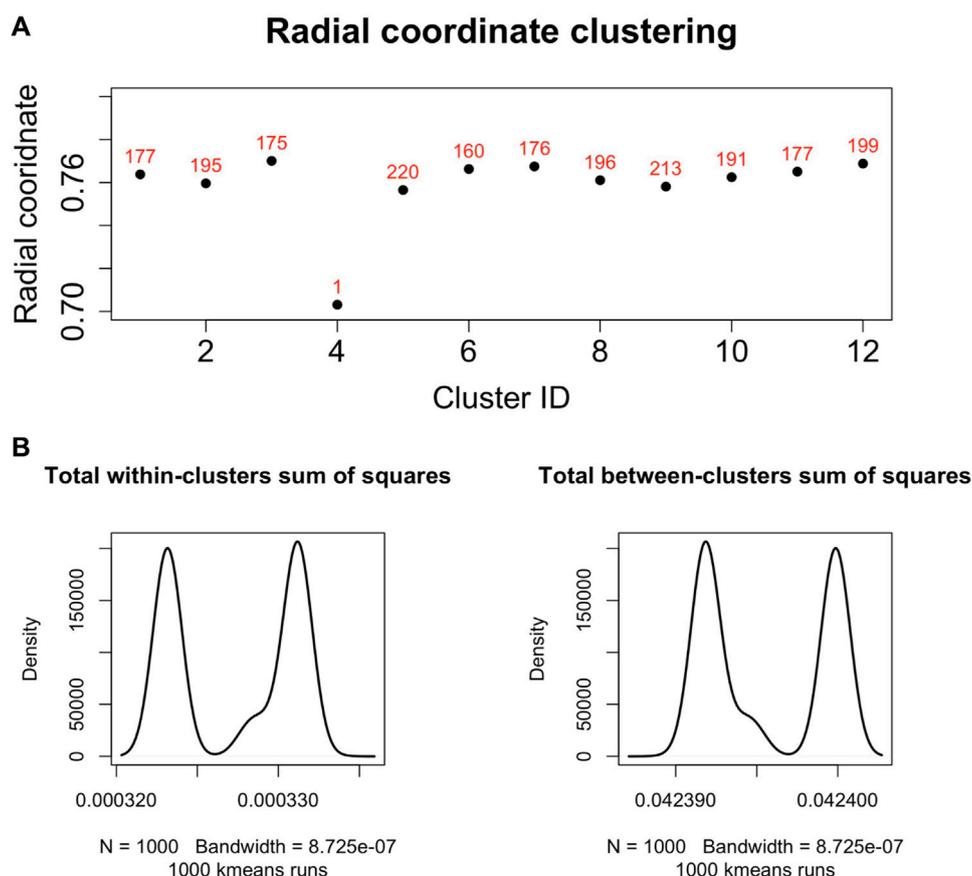
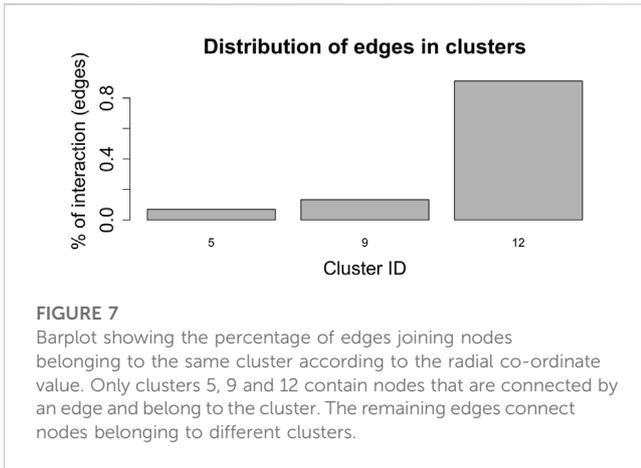


FIGURE 6 (A). Centroids of the radial coordinate clusters versus cluster identifier in a single run of k-means algorithm. The number of genes belonging to each cluster is shown in red. The gene belonging to the cluster number four is ZRANB1, a gene characterized by low immune cell specificity, and belong to NK-cells immune cell expression cluster (Uhlén et al., 2022b). (B). We performed 1,000 runs of the k-means for the clustering of the radial co-ordinate of the nodes of the network in hyperbolic space and drew the distributions of the within- and between-clusters sum of squares. The within-cluster sum of squares quantifies the internal cohesion inside each cluster. The between-cluster sum of squares quantifies the external separation between clusters. The figure shows that for the k-means clustering of the radial coordinates the between-clusters sum of squares is two orders of magnitude greater than the within-cluster sum of squares, revealing the accuracy and then reliability of the clustering results.

We found that the set of radial coordinates, whose values range in [0.7036133, 0.7709305], is characterized by 12 clusters as determined by the Elbow method (Umargono et al., 2020) (see Figure 5). The range of radial coordinate values is. In Figure 6A we report the cluster ID and the size of the 12 clusters of the radial coordinates as obtained by a single run of the k-means algorithm. We found that the gene with the smallest popularity (i.e., with the smallest radial coordinate) is ZRANB1. This gene allows for K63-linked polyubiquitin modification-dependent protein binding and thiol-dependent deubiquitinase activity. Involved in a variety of functions, including the positive control of the Wnt signalling pathway, protein deubiquitination, and cell morphogenesis regulation (NIH, 2023). However, according to the data in The Human Atlas of Proteins is a low immune cell specificity gene (Pontén et al., 2008; Uhlén et al., 2017; Uhlén et al., 2022a; Uhlén et al., 2022b). To assess the stability and the quality of the clustering, we repeated the k-means 1,000 times and graphed the distributions of the within- and between-sum of squares (see Figure 6B) that show that the first is two order of magnitude smaller than the second. The skewness of the distributions and the disproportion between within-

and between-cluster sum of squares indicate the stability and accuracy of the clustering, respectively.

Nodes with similar radial co-ordinate have similar popularity, so clustering according to the radial co-ordinate identifies communities of nodes with similar popularity. However, the radial co-ordinate, in addition to representing the popularity of a node, i.e., its degree of connectivity with other nodes in the network, identifies the distance from the origin in the Poincaré ball. In a network with hyperbolic latent geometry, in its representation in the Poincaré ball, the mean degree of a node is a negative exponential function of the node's radial coordinate (Krioukov et al., 2010). Thus, the average degree of a node decreases exponentially with increasing distance of the node from the origin of the Poincaré ball, or, in other terms, the higher the radial co-ordinate of a node, the lower its degree on average. The area inside the unit ball represents the infinite hyperbolic plane, and, consequently, nodes with radial co-ordinate equal 1 are points at infinity. Clustering according to the radial co-ordinate thus identifies bands of points (nodes) that are concentric on the Poincaré disc and that have a decreasing degree of connectivity as one moves away from the origin. This is why we say that



clustering according to radial co-ordinates allows clusters of strongly interconnected nodes to be identified (if any). The cluster of nodes closest to the origin identifies not only nodes with high connectivity, but also nodes that are close to each other (this second characteristic also applies to clusters far from the origin). The coexistence of two characteristics such as high degree and small distance between nodes is typical of a cluster of nodes with efficient interactivity and greater inertia to perturbations induced by external stimuli, such as variation of expression level, interactions with drugs, etc. The short inter-node distance reflects the high efficiency of communications, the high connectivity may be responsible of the node robustness.

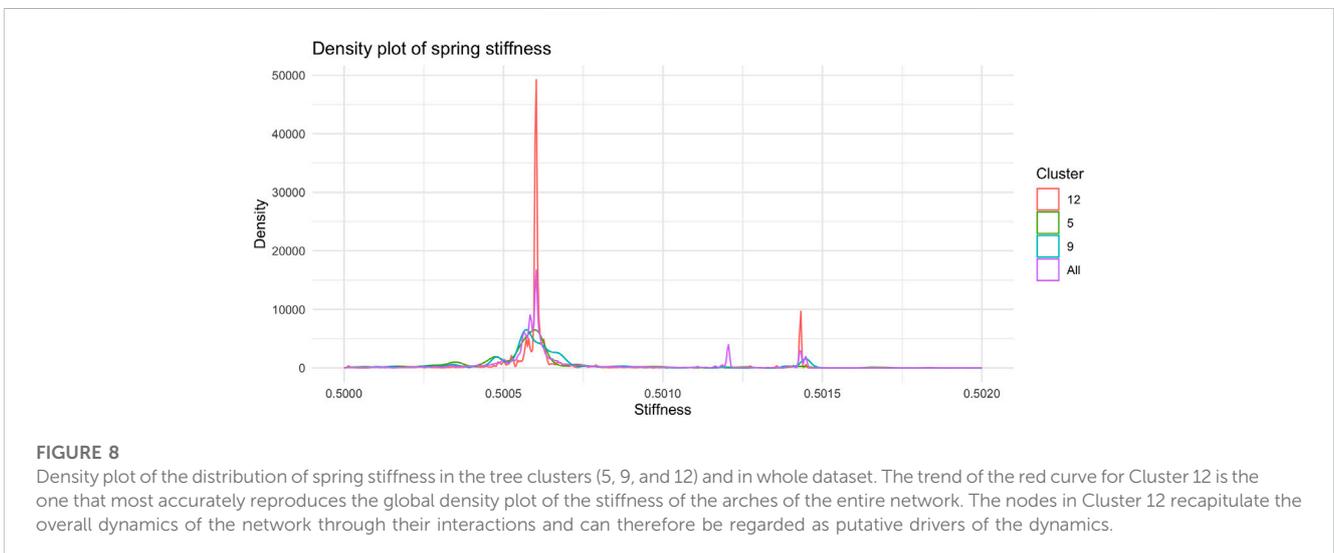
Nodes that are thus highly interconnected and close in network metric space are potential drivers of network dynamics. This conjecture is demonstrated in the case where the distribution of the stiffness of the arcs in the cluster to which these nodes belong is similar to the distribution of the stiffness of the arcs in the overall network. Stiffness is in fact a dynamic property of the system. The cluster of nodes and arcs with dynamic properties that are reflected in the dynamic properties of the entire network can thus be considered a cluster of driver nodes, a characteristic that

designates it as a prime candidate for further wet experiments. In the case of our study, experiments and data from the literature support the hypotheses formulated by the computational analysis, as we shall see below. Indeed, the results we report below are intended to demonstrate these statements.

Figure 7, shows the barplot of the percentage of edges connecting nodes belonging to the same cluster of radial coordinates. Of the 2,162,160 total edges of the graph 1,512 belong to cluster 5, 2,877 to cluster 9, and 19,701 to cluster 12. The remaining 2,138,069 arcs connect nodes belonging to different clusters. In order to understand whether and, if so, how clustering according to radial co-ordinate is reflected in the distribution of spring stiffness, we produced the graph in Figure 8, showing the density plots of the spring stiffness of the interactions between node within the three clusters (5, 9 and 12) compared with the density plot of all spring stiffness of the network. To make the results easier to read and understand, we rescaled the spring stiffness values obtained by formula (22) within a range between 0 and 1 and applied formula (23) to the values obtained in this range.

Of interest we find as shown in this Figure 8 the two peaks of the density plot in red colour corresponding to the stiffness of the interactions between the nodes belonging to cluster number 12. Of the three clusters of radial node distance, number 12 is the one that best reflects the density plot of total spring stiffness. The interactions between nodes belonging to cluster 12 are markedly clustered as is the distribution of stiffnesses across all the arcs of the graph. We interpret this result as the fact that cluster 12 contains nodes that share similar popularity values and are involved in driver interactions of the network dynamics, since the distribution of spring stiffnesses of the arcs of these nodes reproduce the distribution of spring stiffnesses of the entire network.

Cluster 12 contains 199 genes, which a functional analysis implemented with the `enrichGO` function of R library `clusterProfiler` for the Gene Ontology (GO) Enrichment Analysis (Yu, 2012; Yu et al., 2012) finds to have the molecular functions shown in barplot of Figure 9 and the ontologies of the cellular compartments as in Figure 10. The list of the gene names of



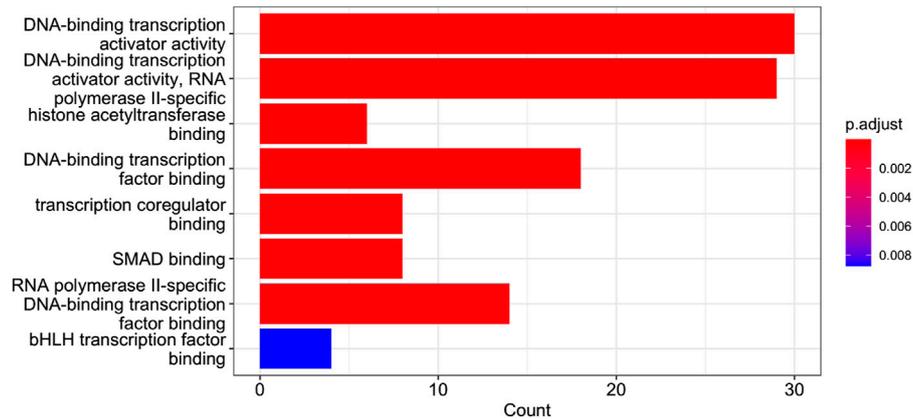


FIGURE 9
GO Enrichment Analysis of the gene set of Cluster 12. The barplot shows the enrichment GO categories of molecular functions after false discovery rate control. See also the verbose tabular output in Cluster_12_GSEA_results_EnrichGO_MF.xlsx provided in [Supplementary Material](#).

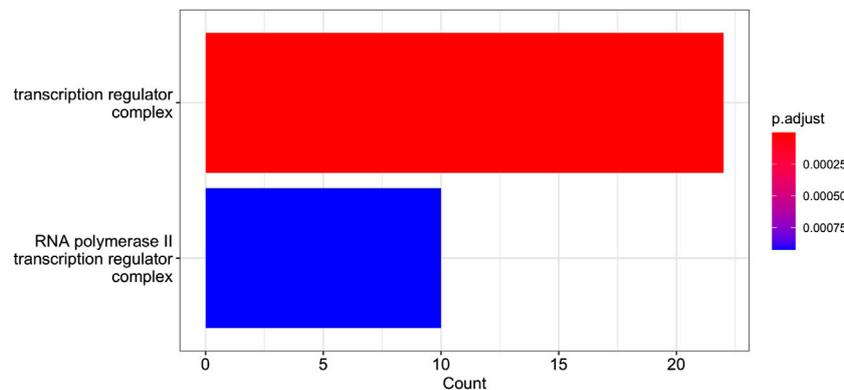


FIGURE 10
GO Enrichment Analysis of the gene set of Cluster 12 (obtained with the R function `enrichGO`). The barplot shows the enrichment GO categories of cellular compartment after false discovery rate control. See also the verbose tabular output in Cluster_12_GSEA_results_EnrichGO_CC.xlsx provided in [Supplementary Material](#).

Cluster 12, as well as the summary of `enrichGO` and of `gost` function of the R library `gprofiler2` (Raudvere et al., 2019; Kolberg et al., 2020; Raudvere et al., 2023) are available in the [Supplementary Material](#). To give a more complete view of the results of the gene set enrichment analysis of Cluster 12, we show in [Figure 11](#) the Manhattan-plots of the gene set enrichment analysis, of which we also give an interactive version in the [Supplementary Material](#).

Of interest is a result shown in [Figure 9](#), namely, the presence in Cluster 12 of genes co-involved in the molecular processes of “SMAD binding”. Smad proteins, are central mediators of the signal transduction of TGF- β family members were identified in the dataset analysed. A Cross-talk between TGF- β /Smad pathway and Wnt/ β -catenin pathway in pathological scar formation has been described suggesting a complicated interaction between the two signal pathways in pathological scar formation (both synergy and antagonism) (Sun et al., 2015). More recently TGF- β /SMAD, Hippo/YAP/TAZ, and Wnt/ β -catenin signalling pathways, major inducers of transcriptional reprogramming, were shown to converge

at several levels and were all required for a proliferative-to-invasive phenotype switch in melanoma development (Lüönd et al., 2021). We already described in a previous study the involvement of Wnt/ β -catenin signalling pathway in the tumour suppressor effect driven by PTPRG in CML (Tomasello et al., 2020) and the current data reporting the involvement of SMAD pathway is in line with a complex cellular reprogramming induced by PTPRG expression whose key role in the haematopoietic differentiation program was already described (Sorio et al., 1997). This complex reprogramming is supported by the large number pathways involved in DNA binding/transcription reported on Cluster 12 GSEA. In particular, in our previous study (Lombardi et al., 2022), we validated the SMAD1 gene. Specifically, qRT-PCR was used to assess gene mRNA levels, and the relative fold changes were calculated between K562 expressing PTPRG and the untreated control group (control and D1028A). The endogenous control was GAPDH. We found that the fold change of SMAD1 is markedly greater in the case of the control [see [Figure 3](#) of Lombardi et al. (2022)].

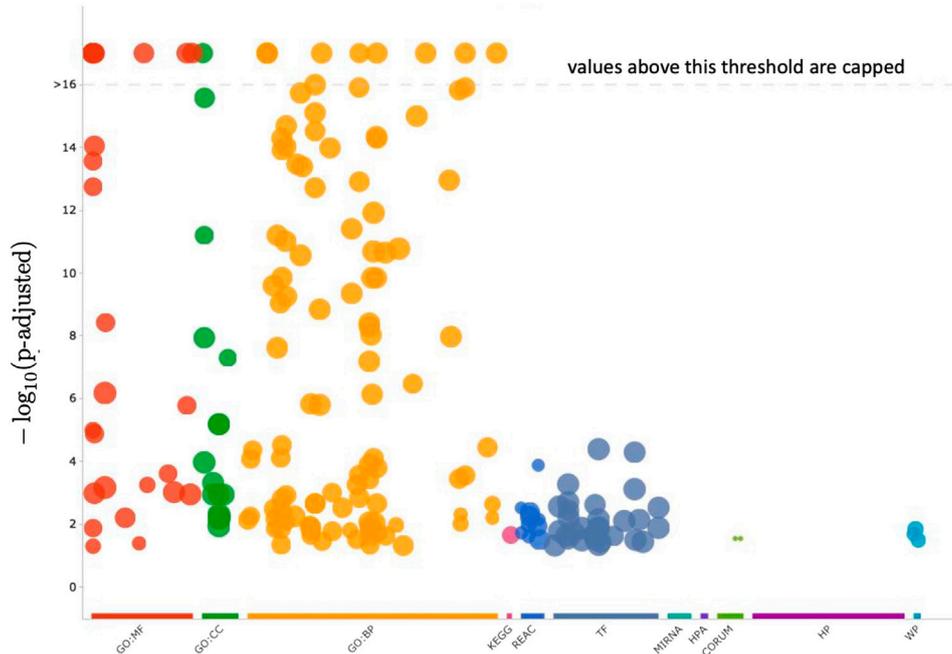


FIGURE 11

In this figure, the enrichment results of the gene set of Cluster 12 are visualized with a Manhattan-like-plot using the function `gostplot` (Raudvere et al., 2023). The x-axis depicts functional terms that are colour-coded and categorised according to data sources and positioned in the fixed “source_order.” The order is set up so that terms that are close together in the source hierarchy are also close together in the Manhattan plot. The modified p -values are displayed on the y-axis in negative \log_{10} scale. Every circle represents one phrase and is proportional to the term size, i.e., larger terms have larger circles. The Supplementary Material includes an interactive version of this plot (`Manhattan_plot_GSEA_Gost.html`). Hovering over the circle in the interactive plot will display the appropriate information. If the $-\log_{10}(p\text{-values})$ exceed 16, they are capped at 16. This adjusts the y-axis scale to keep Manhattan plots from different queries similar, and it is also intuitive because statistically, p -values less than that can all be summarized as highly significant.

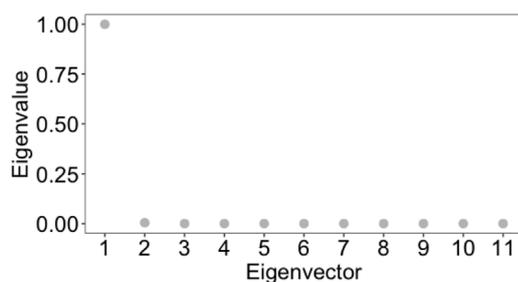


FIGURE 12

Eigengap heuristic: the optimal number of clusters, k , that maximises the eigengap (difference between consecutive eigenvalues of the Laplacian matrix of the graph). The optimal number of clusters is that k such that λ_{k+1} is reasonably large but all other eigenvalues, $\lambda_1, \dots, \lambda_k$, are very small. The closer the eigenvectors of the ideal case are, and hence the better spectral clustering performs, the wider this eigengap is.

5.1 Comparison with spectral clustering

We compared the results of the clustering by radial coordinate in hyperbolic space with the spectral clustering method. This method is widely used to identify communities of nodes in a network by examining the edges that connect them, i.e., taking as an input the weighted adjacency matrix of the

graph. It is a well-established method with theoretical foundations in graph theory (we refer the reader to JingMao and YanXia (2015); von Luxburg (2007) for a review and a tutorial on this popular spectral clustering methods). Processing directly the weighted adjacency matrix of the graph, that is the same input as our embedding procedure, we consider spectral clustering to be the most appropriate method to deal with, compared to clustering methods based on graph centrality measures, or on statistical correlation measures between nodes, who do not into account directly distance measures between nodes. Before applying spectral clustering, we estimated the optimal number of clusters with eigengap heuristics [appropriate procedure for estimating the number of clusters for spectral clustering methods (von Luxburg, 2007)], obtaining that the optimal number of clusters is 2 (see Figure 12). Cluster 1 contains 1,731 nodes and cluster 2 contains 349 nodes. Using the R script `Spectral_clustering.R` to implement spectral clustering - available in GitLab repository, we found that the within cluster sum of squares by cluster is 1.1079409 and 0.2348562, whereas the between sum of squares is 0.6502382. As a consequence, we conclude that the results of the spectral clustering are not reliable. This result highlights how taking into account the latent geometry of the network and with it the clustering according to the spatial co-ordinate of the nodes/points of the network resulted in a much better quality of clustering, compared to a clustering which, as in our

approach, processes the weighted adjacency matrix, but does not consider the latent geometry of the network expressed by the position of the points in the optimal embedding space and the distance defined by the metric in this space.

That various clustering methods are not appropriate for graphs with geometry has also been pointed out by Avrachenkov et al. (2021) that states that while it has been demonstrated that spectral clustering is consistent in some geometric graphs, a cut-based technique (such as spectral clustering) can also be significantly hindered by the geometric structure. It is possible to divide space into regions in such a way that there is relatively little interaction between nodes in two different regions. Therefore, the Fiedler vector of a geometric graph may only be linked to a geometric arrangement and contain no information regarding the labelling of the latent community. Furthermore, because the regions of space can include a balanced number of nodes, the widely used regularisation strategy (Zhang and Rohe, 2018), which seeks to penalise small size communities in order to bring back the vector associated with the community structure in the second position, would not function in geometric graphs.

6 Conclusion

In this study, we modelled the transcriptome network of the of Chronic Myeloid Leukemia K562 cells overexpressing the tumour suppressor gene PTPRG, as a physical system of springs and then deduced the spring constant from topological properties of the nodes, such as total degree. To represent the network, we considered the dissimilarity matrix consisting of the values of the spring's elastic constant, which in our model quantifies the efficiency of information transmission between nodes. Through network embedding procedures that processed the dissimilarity matrix to derive the coordinates of the nodes in a metric space we determined the optimal latent geometry of the network is hyperbolic. This important information made it possible to proceed with the classification of nodes according to radial coordinates (which is the geometric equivalent of the 'physical' concept of node popularity) and to identify a set of candidate driver genes for network dynamics.

This methodology aimed at analysing a network without ignoring the existence of its metric space with a geometry other than the Euclidean one usually imposed or taken for granted, shows how latent geometry can determine a classification of nodes according to their relevance in the network's evolutionary processes, ultimately its dynamics. In the particular case study presented here we obtained that the network has hyperbolic latent geometry, and based on this we proceeded to utilise the concept that in this type of geometry the radial coordinate is a fundamental variable for clustering nodes. Geometries other than hyperbolic are characterised by other spatial variables that can be considered discriminating for the purpose of identifying driver nodes of the dynamics. What is presented in the paper, besides being a concrete result on a specific case

study, is also a proposal for a method of analysing a network in order to reveal information about the dynamics of the network itself.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

Author contributions

PL conceived the paper, the physical and mathematical modelling of the network analysed in the study, and implemented the R code for the graph embedding and node clustering. GL performed the differential expression analysis and produced the network that includes the differentially expressed genes found. RL and CS produced the experimental data. All authors contributed to the article and approved the submitted version.

Funding

This study has been supported by the fund of the DAQETA-CML (Detecting and quantifying (side-)effects of recent experimental therapies against Chronic Myeloid Leukemia) Interdisciplinary Project 2020, earmarked by the Free University of Bozen/Bolzano.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2023.1235116/full#supplementary-material>

References

- Alanis-Lobato, G., Mier, P., and Andrade-Navarro, M. (2018). The latent geometry of the human protein interaction network. *Bioinformatics* 34, 2826–2834. doi:10.1093/bioinformatics/bty206
- Allard, A., and Serrano, M. Á. (2020). Navigable maps of structural brain networks across species. *PLOS Comput. Biol.* 16, e1007584. doi:10.1371/journal.pcbi.1007584
- Anderson, J. W. (2005). *Hyperbolic geometry*. 2 edn. London, England: Springer.
- Avrachenkov, K., Bobu, A., and Drevetov, M. (2021). Higher-order spectral clustering for geometric graphs. *J. Fourier Analysis Appl.* 27, 22. doi:10.1007/s00041-021-09825-2
- Blevins, A. S., Kim, J. Z., and Bassett, D. S. (2021). Variability in higher order structure of noise added to weighted networks. *Commun. Phys.* 4, 233. doi:10.1038/s42005-021-00725-x
- Boguñá, M., Bonamassa, I., Domenico, M. D., Havlin, S., Krioukov, D., and Serrano, M. Á. (2021). Network geometry. *Nat. Rev. Phys.* 3, 114–135. doi:10.1038/s42254-020-00264-4
- Boguñá, M., Papadopoulos, F., and Krioukov, D. (2010). Sustaining the internet with hyperbolic mapping. *Nat. Commun.* 1, 62. doi:10.1038/ncomms1063
- Borg, I., and Groenen, P. J. F. (2005). *Modern multidimensional scaling*. 2 edn. New York, NY: Springer.
- Cerami, E. G., Gross, B. E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., et al. (2010a). Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res.* 39, D685–D690. doi:10.1093/nar/gkq1039
- Cerami, E. G., Gross, B. E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., et al. (2010b). Pathway commons: a resource for biological pathway analysis — pathwaycommons.org. Available at: <https://www.pathwaycommons.org/> (Accessed May 27, 2023).
- Clauset, A., Moore, C., and Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature* 453, 98–101. doi:10.1038/nature06830
- Conn, Z. (2010). Distances in the hyperbolic plane and the hyperbolic Pythagorean theorem. Available at: <http://www.zachconn.com/>.
- Cox, M. A. A., and Cox, T. F. (2008). *Multidimensional scaling*. Berlin, Heidelberg: Springer Berlin Heidelberg, 315–347. doi:10.1007/978-3-540-33037-0_14
- Estrada, E., and Hatano, N. (2010). A vibrational approach to node centrality and vulnerability in complex networks. *Phys. A Stat. Mech. its Appl.* 389, 3648–3660. doi:10.1016/j.physa.2010.03.030
- Fard, A. T., Srihari, S., Mar, J. C., and Ragan, M. A. (2016). Not just a colourful metaphor: modelling the landscape of cellular development using hopfield networks. *npj Syst. Biol. Appl.* 2, 16001. doi:10.1038/npjbsa.2016.1
- Gower, J. C. (2015). “Principal coordinates analysis,” in *Wiley StatsRef: Statistics Reference Online* (Wiley), 1–7. doi:10.1002/9781118445112.stat05670.pub2
- Gromov, M. (2007). *Metric structures for Riemannian and non-Riemannian spaces*. Boston: Birkhäuser. doi:10.1007/978-0-8176-4583-0
- Härtner, F., Andrade-Navarro, M. A., and Alanis-Lobato, G. (2018). Geometric characterisation of disease modules. *Appl. Netw. Sci.* 3, 10. doi:10.1007/s41109-018-0066-3
- Hilbert, D. (1933). “Über flächen von konstanter gaußscher krümmung,” in *Algebra invariantentheorie geometrie* (Berlin Heidelberg: Springer), 437–448. doi:10.1007/978-3-642-52012-9_30
- Jhun, B. (2022). Topological analysis of the latent geometry of a complex network. *Chaos Interdiscip. J. Nonlinear Sci.* 32, 013116. doi:10.1063/5.0073107
- JingMao, Z., and YanXia, S. (2015). “Review on spectral methods for clustering,” in 2015 34th Chinese Control Conference (CCC), Hangzhou, China, 28–30 July 2015 (IEEE), 3791–3796. doi:10.1109/ChiCC.2015.7260226
- Keller-Ressel, M. (2019). hydra: hyperbolic Embedding — cran.r-project.org. Available at: <https://cran.r-project.org/web/packages/hydra/index.html> (Accessed May 27, 2023).
- Keller-Ressel, M., and Nargang, S. (2020). Hydra: a method for strain-minimizing hyperbolic embedding of network- and distance-based data. *J. Complex Netw.* 8. doi:10.1093/comnet/cnaa002
- Kleinberg, J. M. (2000). Navigation in a small world. *Nature* 406, 845. doi:10.1038/35022643
- Klimovskaia, A., Lopez-Paz, D., Bottou, L., and Nickel, M. (2020). Poincaré maps for analyzing complex hierarchies in single-cell data. *Nat. Commun.* 11, 2966. doi:10.1038/s41467-020-16822-4
- Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an r package for gene list functional enrichment analysis and namespace conversion toolset gprofiler. *F1000Research* 9, ELIXIR-709. doi:10.12688/f1000research.24956.2
- Kovács, B., and Palla, G. (2021). The inherent community structure of hyperbolic networks. *Sci. Rep.* 11, 16050. doi:10.1038/s41598-021-93921-2
- Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., and Boguñá, M. (2010). Hyperbolic geometry of complex networks. *Phys. Rev. E* 82, 036106. doi:10.1103/PhysRevE.82.036106
- Kurkofka, J., Melcher, R., and Pitz, M. (2021). Approximating infinite graphs by normal trees. *J. Comb. Theory, Ser. B* 148, 173–183. doi:10.1016/j.jctb.2020.12.007
- Lecca, P. (2023). Uncovering the geometry of protein interaction network: The case of SARS-CoV-2 protein interactome. *AIP Conf. Proc.* 2872 (1), 030008. doi:10.1063/5.0163052
- Lecca, P., and Re, A. (2020). “Stiffness estimate of information propagation in biological systems modelled as spring networks,” in 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea (South), 16–19 December 2020 (IEEE). doi:10.1109/bibm49941.2020.9313294
- Lecca, P., and Re, A. (2022). “Checking for non-euclidean latent geometry of biological networks,” in 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, 06–08 December 2022 (IEEE). doi:10.1109/bibm55620.2022.9995274
- Lecca, P., Re, A., Lombardi, G., Latorre, R. V., and Sorio, C. (2023). “Graph embedding of chronic myeloid leukaemia k562 cells gene network reveals a hyperbolic latent geometry,” in *Proceedings of eighth international congress on information and communication technology* (Singapore: Springer Nature), 979–991. doi:10.1007/978-981-99-3091-3_80
- Liben-Nowell, D., and Kleinberg, J. (2007). The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* 58, 1019–1031. doi:10.1002/asi.20591
- Lombardi, G., Latorre, R. V., Mosca, A., Calvanese, D., Tomasello, L., Boni, C., et al. (2022). Gene expression landscape of chronic myeloid leukemia k562 cells overexpressing the tumor suppressor gene PTPRG. *Int. J. Mol. Sci.* 23, 9899. doi:10.3390/ijms23179899
- Lüönd, F., Pirkl, M., Hisano, M., Prestigiacomo, V., Kalathur, R. K., Beerenwinkel, N., et al. (2021). Hierarchy of TGFβ/SMAD, hippo/YAP/TAZ, and wnt/β-catenin signaling in melanoma phenotype switching. *Life Sci. Alliance* 5, e202101010. doi:10.26508/lsa.202101010
- Maddalena, L., Giordano, M., Manzo, M., and Guarracino, M. R. (2022). “Whole-graph embedding and adversarial attacks for life sciences,” in *Trends in biomathematics: stability and oscillations in environmental, social, and biological models. BIOMAT 2021*. Editor R. P. Mondaini (Springer, Cham). doi:10.1007/978-3-031-12515-7_1
- Millán, A. P., Torres, J. J., and Bianconi, G. (2018). Complex network geometry and frustrated synchronization. *Sci. Rep.* 8, 9910. doi:10.1038/s41598-018-28236-w
- NIH (2023). ZRANB1 zinc finger RANBP2-type containing 1 [*Homo sapiens* (human)]. Available at: <https://www.ncbi.nlm.nih.gov/gene/54764> (Accessed May 30, 2023).
- Okuno, A., and Shimodaira, H. (2019). “Robust graph embedding with noisy link weights,” in *The 22nd international conference on artificial intelligence and statistics, AISTATS 2019*. Editors K. Chaudhuri and M. Sugiyama (Naha, Okinawa, Japan: PMLR), 664–673.
- Papadopoulos, F., and Flores, M. A. R. (2019). Latent geometry and dynamics of proximity networks. *Phys. Rev. E* 100, 052313. doi:10.1103/PhysRevE.100.052313
- Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M., and Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature* 489, 537–540. doi:10.1038/nature11459
- Pio, G., Ceci, M., Prisciandaro, F., and Malerba, D. (2019). Exploiting causality in gene network reconstruction based on graph embedding. *Mach. Learn.* 109, 1231–1279. doi:10.1007/s10994-019-05861-8
- Pontén, F., Jirstrom, K., and Uhlen, M. (2008). The human protein atlas—a tool for pathology. *J. Pathology* 216, 387–393. doi:10.1002/path.2440
- Rand, D. A., Raju, A., Sáez, M., Corson, F., and Siggia, E. D. (2021). Geometry of gene regulatory dynamics. *Proc. Natl. Acad. Sci.* 118, e2109729118. doi:10.1073/pnas.2109729118
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g: profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi:10.1093/nar/gkz369
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2023). Gene list functional enrichment analysis and namespace conversion with gprofiler2. Available at: <https://cran.r-project.org/web/packages/gprofiler2/vignettes/gprofiler2.html> (Accessed June 02, 2023).
- R Core Team (2021). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Sánchez-Romero, M. A., and Casadesús, J. (2021). Waddington’s landscapes in the bacterial world. *Front. Microbiol.* 12, 685080. doi:10.3389/fmicb.2021.685080
- Seyboldt, R., Lavoie, J., Henry, A., Vanaret, J., Petkova, M. D., Gregor, T., et al. (2022). Latent space of a small genetic network: geometry of dynamics and information. *Proc. Natl. Acad. Sci.* 119, e2113651119. doi:10.1073/pnas.2113651119

- Sorio, C., Melotti, P., D'Arcangelo, D., Mendrola, J., Calabretta, B., Croce, C. M., et al. (1997). Receptor protein tyrosine phosphatase gamma, ptp gamma, regulates hematopoietic differentiation. *Blood* 90, 49–57. doi:10.1182/blood.v90.1.49
- Squier, S. M. (2017). *Epigenetic landscapes: drawings as metaphor*. Durham, NC, United States: Duke University Press.
- Sun, N., Pei, S., He, L., Yin, C., He, R. L., and Yau, S. S. T. (2021). Geometric construction of viral genome space and its applications. *Comput. Struct. Biotechnol. J.* 19, 4226–4234. doi:10.1016/j.csbj.2021.07.028
- Sun, Q., Guo, S., Wang, C. C., Sun, X., Wang, D., Xu, N., et al. (2015). Cross-talk between TGF- β /Smad pathway and Wnt/ β -catenin pathway in pathological scar formation. *Int. J. Clin. Exp. Pathol.* 8, 7631–7639.
- Taylor, D., Klimm, F., Harrington, H. A., Kramár, M., Mischaikow, K., Porter, M. A., et al. (2015). Topological data analysis of contagion maps for examining spreading processes on networks. *Nat. Commun.* 6, 7723. doi:10.1038/ncomms8723
- Tomasello, L., Vezzalini, M., Boni, C., Bonifacio, M., Scaffidi, L., Yassin, M., et al. (2020). Regulative loop between β -catenin and protein tyrosine receptor type γ in chronic myeloid leukemia. *Int. J. Mol. Sci.* 21, 2298. doi:10.3390/ijms21072298
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2022a). The human protein atlas — proteomics.org. Available at: <https://www.proteinatlas.org/> (Accessed May 30, 2023).
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2022b). ZRANB1 protein expression summary - the Human Protein Atlas — proteomics.org. Available at: <https://www.proteinatlas.org/ENSG0000019995-ZRANB1> (Accessed May 30, 2023).
- Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357, eaan2507. doi:10.1126/science.aan2507
- Umargono, E., Suseno, J. E., and Vincensius Gunawan, S. K. (2020). “K-means clustering optimization using the elbow method and early centroid determination based-on mean and median,” in Proceedings of the International Conferences on Information System and Technology – CONRIST, Yogyakarta, Indonesia (Setúbal, Portugal: SciTePress) 1, 234–240. doi:10.5220/0009908402340240
- von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics Comput.* 17, 395–416. doi:10.1007/s11222-007-9033-z
- Wilson, R. C., Hancock, E. R., Pekalska, E., and Duin, R. P. (2014a). Spherical and hyperbolic embeddings of data. *IEEE Trans. Pattern Analysis Mach. Intell.* 36, 2255–2269. doi:10.1109/TPAMI.2014.2316836
- Wilson, R. C., Pekalska, E. R. H. E., and Duin, R. P. (2014b). Classical (metric) multidimensional scaling. Available at: <https://www.cs.york.ac.uk/cvpr/post/sphericaembedding/> (Accessed May 23, 2023).
- Yang, W., and Rideout, D. (2020). High dimensional hyperbolic geometry of complex networks. *Mathematics* 8, 1861. doi:10.3390/math8111861
- Yu, G. (2012). enrichGO function - RDocumentation — rdocumentation.org. Available at: <https://www.rdocumentation.org/packages/clusterProfiler/versions/3.0.4/topics/enrichGO> (Accessed June 02, 2023).
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118
- Zhang, Y., and Rohe, K. (2018). “Understanding regularized spectral clustering via graph conductance,” in *Proceedings of the 32nd international conference on neural information processing systems* (Red Hook, NY, USA: Curran Associates Inc.), 10654–10663.
- Zhang, Z., and Takane, Y. (2010). “Multidimensional scaling,” in *International encyclopedia of education*. Editors P. Peterson, E. Baker, and B. McGaw (Oxford: Elsevier), 304–311. doi:10.1016/B978-0-08-044894-7.01348-8