Topological Coding and Visualization Grammar of the Development of C. elegans

A. Krämer
Dept. of Biochemistry
University of Kiel
Olshausenstr. 40 N11,
D - 24098 Kiel, Germany
akraemer@biochem.uni-kiel.de, 4dce.org

Abstract

This paper reports on the development of a method to build a mathematical model for a growing biological organism. A tiny worm called Caenorhabditis elegans is chosen as paradigm. The problem of describing biological development by mathematical means is addressed. After an abstraction process, the organism is embedded into a suitable vector space. For this purpose, a vector spaces chain is build in which topological modules of the object are described. The vector space topology may be characterized by simplex sets. The structure of the simplex sets is normalized and its connectivity is used as a compact description of the development. Development of biological organisms takes place in 4D (3D space and time). The 3D scenes code may be a characteristic pattern and to reconstruct the object. It can be subject to operations and the result may be used to construct a new scene or the next state of development. A bio-application named " Topologizer" is being developed (using VTK) to support the mathematical description and analysis of the organisms growth in a massive parallel and distributed project approach.

1 Introduction

The growth of biological organisms towards their complex morphology is governed by rules. The development of a worm, the nematode Caenorhabditis elegans from the egg to the adult stage serves as a model. Caenorhabditis elegans is a good model organism for this project since this organism takes an extreme position in terms of development characteristics in its species. The adult nematode consists of a constant number of cells (hermaphrodites have 959 somatic nuclei and male have 1031) with invariant cell lineage [SSWT83]. The development for the group of male follows the one pattern, as the development for the group of female (hermaphrodites, i.e male and female in one) follows another pattern. Besides, the chromosomal DNA of this organism has been completely sequenced, –(with a size of $8 \cdot 10^7$ nucleotide pairs which is eight times larger than yeast Saccharomyces and about one half of the fruit fly Drosophila)– which is as well the case for some bacteria and viruses. The human chromosomal DNA is said to be sequences completely, however the methods applied produce an average error of $\approx 3\%$ in the results, which is not the case with the DNA sequence of Caenorhabditis elegans.

The nematode (C. elegans) is a small worm of $\approx$ 1.3 mm length and a diameter of $\approx$ 1/20 of its length. It is a hermaphrodite and produces $\approx$ 300 eggs during its life. It is easy to cultivate in the laboratory on agar dishes inoculated with bacteria (Escherichia coli) as food. An egg has a length of $\approx$ 60 μm and a diameter of$\approx$30μm. For more details, see http://wormbase.org.

1.1 Previous work

The development in three space dimensions and time has been well studied by [Se97] and [RJJB99]. The genetic approach is implemented by [Se97], where cells are tracked in space and time after genetic manipulations are applied. A very promising approach is chosen by [Hea01], where a synthetic development model is constructed by visualizing the fused set of all available data (e.g. geometric as well as chromosomal data are considered). With the developed tool impacts of DNA manipulations can be visualized, sadly the tool is not freely avaiorgasisms lable. However, it is still not known how growth rules and morphological structure is coded on the chromosomal DNA.

1.2 This work

As stated above this paper addresses the problem of describing or coding biological development. It begins with the

data acquisition and ends up at a description of 3D and 4D scene with a normalized incidence matrix. The analysis has therefore been structured into sections:

- A: Geometry. Segmentation of the egg into a growing mosaic of cells is using the easily recognized nucleus of each cell. The position of the center of each cell nucleus is representative of the cell.
- B: Topology. A eigen-vector space with coordinate system is introduced such that the development as a set of algebraic operations on segments is describable. The adjacency of the cell is approximated by a Delaunay triangulation and Voronoi tessellation of the nuclei of the eggs in 3D and 4D.
- C: Pattern. In the above vector space the relation between 4D segments of development are characterized and the attempt is made to understand these as being assembled of lower dimensional topological objects (i.e. simplex sets). On the other hand matrix methods are applied to discover regularities in the development of the structure.

1.3 Computational geometry

Throughout this work a set of computational geometry method is applied which is common to the ares of pattern recognition and visualization. The Delaunay and Voronoi algorithm is well known and a huge amount of work has been done on e.g. dynamic delaunay algorithms since [ME96] and [ELZ00]. The description of 3D scenes and their development in 4D can be well performed with e.g. the tensor voting method [MTL00]. Tensors and groups of tensors may characterize the objects properties, synthesizing these to evolving segments. Distinct element modeling combined with dynamic delaunay triangulation [FML00] implements a discrete element algorithm with reduced computational cost. [1] These are methods to visualize the property of a scene or to run the simulation of a given model and may be employed as well in the coming work.

Since visualization is not the primary topic of this work but the construction of a model is, we have to overcome some shortcomings of pure descriptive and phenomenological algorithms while we try to reduce the observation of biological developments to the set of related primatives meeting the above models requirements.

[1] It is a numerical technique which solves engineering problems that are modeled as a large system of distinct interacting general shaped (deformable or rigid) bodies or particles that are subject to gross motion.

1.4 Microscopic data

An egg taken from the nematode is placed in a microscope (Nomarski Differential Interference Optic or Laser Scanning Confocal Microscope), ≈1500 times enlarged and scanned in space and time. A time sequence of such (≈1700)3D-stacks form a 4D data set i.e. a movie in 3D format. The data set represents the development of an egg from its zygote (one cell) state until prior to hatching (≈540 cell state).

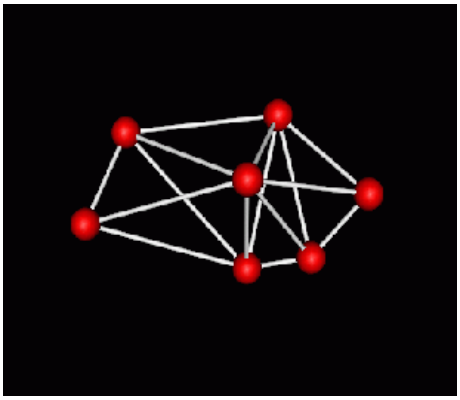
1.5 The cell lineage

The remarkable feature of the C. elegans development is the invariance of its cell lineage and the constant number of cells, of which the worm is made up (for details see “http://elegans.swmed.edu/”). The data of the cell lineage has been gathered in a database and may be displayed as a tree structure by a computer program of SIMI Inc. (“http://www.simi.net”). This program not only allows to display, but also to modify data of the database. The database is so organized that each record holds data of a specific cell, i.e. the name of a cell; the time of birth; a reference time; the array of time position; and other data.

1.6 The adjacency of cells

The cell lineage shows the time or causal sequence in which cells divide and exist during the development of the egg. The time frame in which the cells devision a takes place is followed by a time frame in which (a topological) sorting process of cells is observed. Cells begin to move through the egg to take their final functional position. This phenomena is not a priori encoded in the linage tree, since cells do not remain at the place where they are born, and its as not necessarily covered by ontogeny or phylogeny [2] The nuclei of the cells are clearly visible in the microscope. Their position in space and time of division can be monitored. This allows the definition of a distance matrix between the individual nuclei. For the seven cell state of fig. 1 the distance matrix is given in tab. 2

[2] i.e. the development of the animal embryo and young traces the evolutionary development of the species. The theory was influential and much-popularized earlier but has been of little significance in elucidating either evolution or embryonic growth. Haeckel’s Law (1866): Ontogeny recapitulates phylogeny, i.e., an embryo repeats in its development the evolutionary history of its species as it passes through stages in which it resembles its remote ancestors. (Embryos, however, do not pass through the adult stages of their ancestors; ontogeny does not recapitulate phylogeny (i.e. the history of the evolution of a species or group, especially in reference to lines of descent and relationships among broad groups of organisms.). Rather, ontogeny repeats some ontogeny - some embryonic features of ancestors are present in embryonic development [Wol91])



(1)

	n_1	n_2	n_3	n_4	n_5	n_6	n_7
n_1	0	0.1	0.2	0.2	0.2	0.2	0.3
n_2	0.1	0	0.1	0.1	0.2	0.2	0.3
n_3	0.2	0.2	0	0.1	0.1	0.1	0.2
n_4	0.2	0.1	0.1	0	0.1	0.1	0.1
n_5	0.2	0.2	0.1	0.1	0	0.1	0.1
n_6	0.2	0.2	0.1	0.2	0.1	0	0.1
n_7	0.3	0.3	0.1	0.1	0.1	0.1	0

(2)

A (top panel): Movie of a Ball-stick model in the 7 cell stage. The ball are at the position of the nuclei. The sticks connect balls of cell neighbors as obtained by Delaunay triangulation (mouse click on the picture runs the movie) . B(bottom table) Distance matrix of A

The square of Euclidean distance is $d^2_{n_i,n_j} = (x_{n_i} - x_{n_j})^2 + (y_{n_i} - y_{n_j})^2 + (z_{n_i} - z_{n_j})^2$ between nucleus n_i and n_j. The value d_{n_i,n_j} is entered at the element (i, j) in the distance matrix table tab. 2. This matrix is symmetric with zeros at the diagonal. The distance matrix gives the distances between all nuclei. Of particular interest for cells are their neighboring cells. The mathematical construction of Delaunay triangulation produces a relationship between the nearest neighbors (see [GR97]). It connects the nuclei such that a net of triangles (in the plane) and of tetrahedron (in 3D-space) results (see fig. 1). No nuclei lies within a triangle or tetrahedron, and even no nuclei lies within the enclosing circles or sphere through the points (or nuclei) forming the triangle or the tetrahedron. The matrix table 2

	n_1	n_2	n_3	n_4	n_5	n_6	n_7
n_1	1	1	1	0	0	1	0
n_2	1	1	1	1	0	1	0
n_3	1	1	1	1	1	1	1
n_4	0	1	1	1	1	1	1
n_5	0	0	1	1	1	1	1
n_6	1	1	1	1	1	1	0
n_7	0	0	1	1	1	0	1

(3)

Connectivity matrix for figure 1

may be simplified in this case to give the adjacency or incidence matrix. (see table 3). The entry of this matrix are given by the rule: “if two nuclei (points) are connected in the triangulation net, a 1 else a 0 is entered as a matrix element”. Also this matrix is symmetric with ones the diagonal.

This adjacency matrix may not reflect the real neighbors of the cell harboring the corresponding nuclei since the nucleus may not be the center of cell nature. The triangulation only approximates the generation of cells in nature. Nevertheless, if there is no more information available, then the position of the cell nuclei it is a good approximation to reality. For the establishment of the real neighborhood, the hulls (membranes) of the cells should be known to touch.

1.7 Representation of structures in a vector space

To meet the models requirements for the description of patterns in the development, a “vector space chain” with coordinate system is introduced. The abstracted organism is embedded and described in this space. And its development can be described with the boundary operator.

Construction of a vector space chain with a boundary operator: The Delaunay triangulation forms a vector space set where all tetrahedron form the base of one vector space, all triangle (i.e. the surface of tetrahedron) form a basis of another vector space and all lines connecting the nuclei form the basis of a third vector space. These vector spaces are transformed into each other by boundary operations. The boundary (or surface) of the tetrahedron consists of triangles (i.e. the facets are basis of the next vector space). The edges (lines) of the circumference triangles are the basis of their next vector space. The endpoints (surface) of the line finally lead to an end of this sequence by contracting to points an basis of a vector space chain. For details of their mathematical construction see [OD71]. Mathematically, each vector space has an existing co-vector space. For the Delaunay triangulation the Voronoi tessellation is the dual or co-vector space. The dimension of the vector space and its co-space sum in the geometrical representation of the d-dimensional space to d. In the example of the egg the co-space to the nuclei (points) is the biological cell as a 3D volume. Each line of connection between nuclei in the Delaunay space corresponds to the area of a cell surface (membrane) which both cells (or nuclei) share in the co-space of Voronoi. Each triangle in the Delaunay space corresponds to a line in the space of Voronoi, that is part of the border of shared cell surfaces (facets of connected cells). Each tetrahedron of the Delaunay space produces in the co-space of Voronoi an endpoint of the lines just mentioned. In this picture, the visible cell with its wall (membrane) is part

of the co-space of Voronoi tessellation where the inner construction of the nuclei and skeleton of cell may be assigned to the space of Delaunay triangulation.

The coefficients in the vector space: Now the cell structure in the egg can be represented as a vector in each of these spaces. The vector space of lines (Delaunay) has as components (coefficients) of the vector the length of the lines (distances at Delaunay triangulation). The triangles have their areas and the tetrahedron their volumes as components. For the Voronoi tessellation the co-space of cells has the volume of each cell, the co-space of contact surfaces has the area of the surface and the co-space of lines has the length of each line of the polygons surrounding the contact face of the cell as components. All these vector components are dependent (i.e. interrelated), which may be verified since the Delaunay as well as the Voronoi structures may be reconstructed from the distances (see tab. 2). This opens the possibility to set conditions not only for line components (as distances) but for the whole set of coefficients used to describe the structure or phenomenom in the developing egg.

1.8 The development in 4D

The representation in Delaunay triangulation and Voronoi tessellation may be straight forward extended into four dimensions (space and time) or even higher. There are several kinds of characterization of the development of the egg in 4D by coordinates of points (i.e. the point of nuclei center). One way is to assign each cell (nuclei), "as time", the division time and as space the coordinates. The space coordinates of center of nuclei just prior to division are causal development states. That is a rough approximation, but allows to get and check the properties derived for the cell lineage and the cell structure in 3D space. The development may subsequently be refined to display intermediate states. The 4D representation is hard to visualize but advantageous in a systematic treatment as e.g. an abstraction, a reconstruction or to find regularities as e.g. growth rules. These 4D regularities may be understood as development structures assembled of lower dimensional simplex sets.

Another possibility to describe the development is to abandon the time concept and use a parameterized function of boundary operations. To implement one step in development from state $A \rightarrow B$ (i.e. a topological operation is applied to the object) we do simply observe the boundaries of the two states development. Comparing these facet sets results in a function for facet operation. More exact the facet set of $\delta = \overline{A \cap B}$ is the difference between state A and B. Since A and B are entities in 3D, their boundary facet set δ is that what represents the development in 4D. This function can later be derived when the normalized connectivity matrices are dealt with. The parameterized facet operation function is the topological counterpart of the above mentioned negative entropy production density. Further it appears to be sensible to consider in the following structured or ordered facet end simplex sets. Algebraic set operations can that easily formulate complex scenes.

1.9 Coding the topology

The objects data consists of a geometric (metric) part (i.e. point coordinates or distances) and a topological part, which describes the connectivity within the object. The geometrical data are represented usually by real numbers (point coordinates), where the topological data are given by relations between abstract "names" (i.e. points are numbered). Now, the connection even in the 4D can be represented as a sequence of integers for lines, poly-lines and polygons (closed poly-lines) and also for triangles, triangle strips (poly triangles) and closed triangle strips (which are an equivalent to polygons). This may be extended to the tetrahedron, tetrahedron strips and closed tetrahedron strips in the 3D space. And further to pentahedron, pentahedron strips and closed pentahedron strips in 4D space. In the 4D example of the development of the egg, the pentahedron strip representation of the Delaunay triangulation gives a very compact representation. The connectivity of pentahedron structures may be coded in a matrix with simplices as index.

1.10 Matrix Coding of Simplices

The table 3 gives an example for the connectivity in the seven cell stage. Cells that touch (connect) at a time (stage) have a 1 as matrix element. The matrix is symmetric since the assignment of the endpoints to the row and column is arbitrary. A 2D-simplex (triangle) may be represented in that the identifying numbers of three points are taken as the indices of a matrix with 3 indices. Also in this case all permutation of the three indices result in the same simplex entered as a 1. In the similar a manner the connectivity for higher dimensions may lead to a matrix representation. This representation will be normalized by analyzing the structure of permutation groups which the above type of connectivity matrix is assigned to.

1.11 Connectivity Matrix of Simplices

The method of depicting the connectivity between points by matrices may be extended to higher dimensional simplices. The simplices (e.g. a line) may be sequentially numbered as in the former case the vertices (nuclei). The numbers

assigned to the n-dimensional simplices may be used as indices to a matrix. If a simplex (e.g. a line) is connected to another simplex (i.e. a neighboring line), an entry is made in the connectivity matrix. As in the former case the connectivity structure of n-dimensional simplices can be visualized as dot plot. Analogously three 1-simplices build one 3-simplex (a triangle). The triangle can be represented in a matrix with 3 indices of the lines and so for the connectivity of the higher dimensional simplices. This allows to visualize connectivity structures of higher dimensional spaces of simplices in a lower dimensional graph.

1.12 Dotplot visualization of 4D pentahedron connectivity

The matrix representation allows the visualization of the connectivity by tools developed in matrix computer program packages. The dot diagram in which each non zero entry in the 2D matrix is printed by a dot at the position given by the matrix indeces. By this plot the matrix structures become clearly visible. Figure 1 shows such a matrix structure of the pentahedron connectivity of C.elegans.

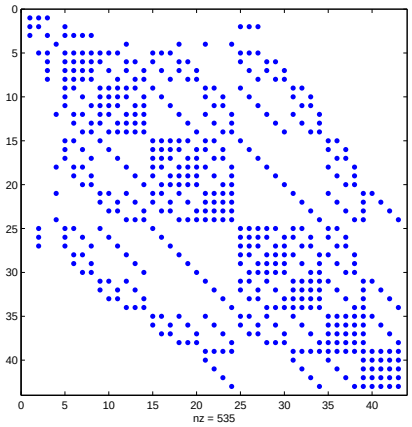


Figure 1: 4D pentahedron connectivity simplex matrix for the development of C.elegans. The matrix has 45 pentahedron as indices and one dot is assigned when two pentahedron are connected. This representation of development is not jet unique.

1.13 Incidence matrix normalized by TDO

The algorithm "Tactical Decomposition by Ordering" (TDO) permutes rows and columns of the incidence matrix so that a normalized matrix results (see for details D.Betten [BBT01] and [BB99]). By this method the variety of homeomorphic matrices (and their geometric representations is groups) are projected to a unique form. For a systematic representation or classification of connectivity in a network such normalized forms are needed. Since the incidence matrix becomes large for increasing numbers of nodes (or cells) we do only consider the 4-cell-, 6-cell-, 8-cell-, 12-cell-stages here (see fig.). A part of the development of the organism is expressed in the transformation from the 4-cell-stage to the 15-cell-stage my mapping the consecutive matrices to the next stage of development (i.e. the geometry it represents). Since the matrices (matroid i.e. matrices only with 0 or 1 entries) may be converted in a geometry, the maps describe a development of a geometric shape in space. In particular in 4D space (time and 3D) a tdo normalize matroid describes the complete development of shape. Also in 3D the sequence of maps may give information about the rules that govern the developing shape of an organism. This as well opens up an application to search in the above 4D incidence matrix for a certain pattern or edit the pattern for re- or for construction of the organism. The experimentally applied biochemical modification to the organism, that results in a modified development would have here a unique "pattern food print" and thus as well in the associated equation of development. The matrices are now vertically (row) indexed with simplex names, in this case 1D points numbered 1 2 3 and so on. Horizontally the matrices are (column) indexed by their connectivity elements, i.e those facets which implement the connectivity.

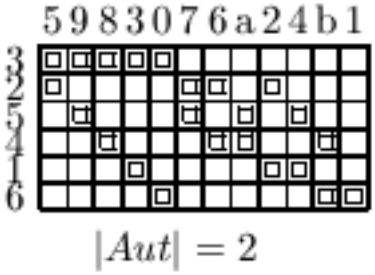


Figure 2: TDO normalized incidence matrices represent the connectivity between cells and the development of the 6–cell stage. —Aut— gives the possible automorphism. The sequence TDO normalized incidence matrices, where each state is characterized by one matrix describes the development. Row indices are cells, column indices are cell contacts. The number of rows gives the cell stage in the egg. The digits and letters at the top of the matroid are a symbolic ascii sequence for the connectivity elements starting with 1. Given is the row and column permutation of the original matroid order. A square in a mesh of grid means the cell is the row is involved in the contact of the column.

2. Summary and Conclusions

We like to show how to described the development of a growing biological organism in a consecutive sequence of vector spaces. The method employs matrix operations to formulate (or code) the topological characteristics of examined developing structures in the vector spaces sequence.

Operations on normalized incidence matrices lead to a visualization of building blocks (or modules), where one building block may capture geometric objects of common function.

A software is being developed to analyze the C. elegans embryos growth process in real-time. It is possible to describe embryos biological development and to compare development with that of other embryos. Therefore a control criteria for biological development is provided. This control criteria may be used to manage an experiment, to estimate its development or to abort an experiment in case of an C.elegans embryo running out of control.

References

[BB99] *Anton Betten and Dieter Betten. The proper linear spaces on 17 points.* Discrete Applied Mathematics*, 1999.*

[BBT01] *Anton Betten, Dieter Betten, and Vladimir D. Tonchev. Unitals and codes.* Preprint submitted to Elsevier Science*, 2001.*

[ELZ00] *Herber Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. Technical report, Department of Computer Science at Duke University, 2000.*

[FML00] *J.-A. Ferrez, D. Mller, and Th.M. Liebling. Dynamic triangulations for granular media...* Lecture Notes in Physics, Springer*, 2000.*

[GR97] Handbook of Discrete and Computational Geometry*. CRC Press New York, 1997.*

[Hea01] *Paul J. Heid and et. al. A computer-assisted system for reconstructing ...* 2001 International Worm Meeting abstract 841*, 2001.*

[ME96] *Facello M. and H. Edelsbrunner.* Geometric techniques for molecular shape analysis*. PhD thesis, University of Illinois, 1996.*

[MTL00] *G. Medioni, C.-K. Tang, and M.-S. Lee. Tensor voting : Theory and applications.* 12eme Congres Francophone AFRIF-AFIA ...(RFIA)*, February 2000.*

[OD71] *G.F. Oster and C.A. Desors. Tellegen's theorem and thermodynamic inequalities.* Journal of theoretical Biology*, 1971.*

[RJJB99] *Hill RJ, Priess JR, Pitt JN, and Leung BH. A java based 4d lineage analysis application.* International C. elegans Meeting 1999 676*, 1999.*

[Se97] *R. Schnabel and et.al. Accessing normal embryogenesis in caenorhabditis ...* Developmental Biology*, 1997.*

[SSWT83] *J E Sulston, E Schierenberg, J G White, and J.N. Thomson. The embryonic cell lineage in the nematode caenorhabditis elegans.* Developmental Biology*, 1983.*

[Wol91] *L. Wolpert. The triumph of embryo.* Oxford University Press*, 1991.*