The Geometry and Topology of Data Analysis

DATA 2014

The Geometry and Topology of Data Analysis

DATA 2014

Herbert ∃delsbrunner   IST ∀ustria

I CLASSIFICATION

II SHAPE

III SCALE

IV APPLICATIONS

I Classification

II Shape

III Scale

IV Applications

I.1 . . . IN BIOLOGY

CAROLUS LINNAEUS (also Carl von Linné) 1707 - 1778

Systema Naturae

1 st edition in 1735
10th edition in 1758

I.1 … in Biology

Carolus Linnaeus (also Carl von Linné) 1707 - 1778

Systema Naturae — 1 st edition in 1735, 10 th edition in 1758

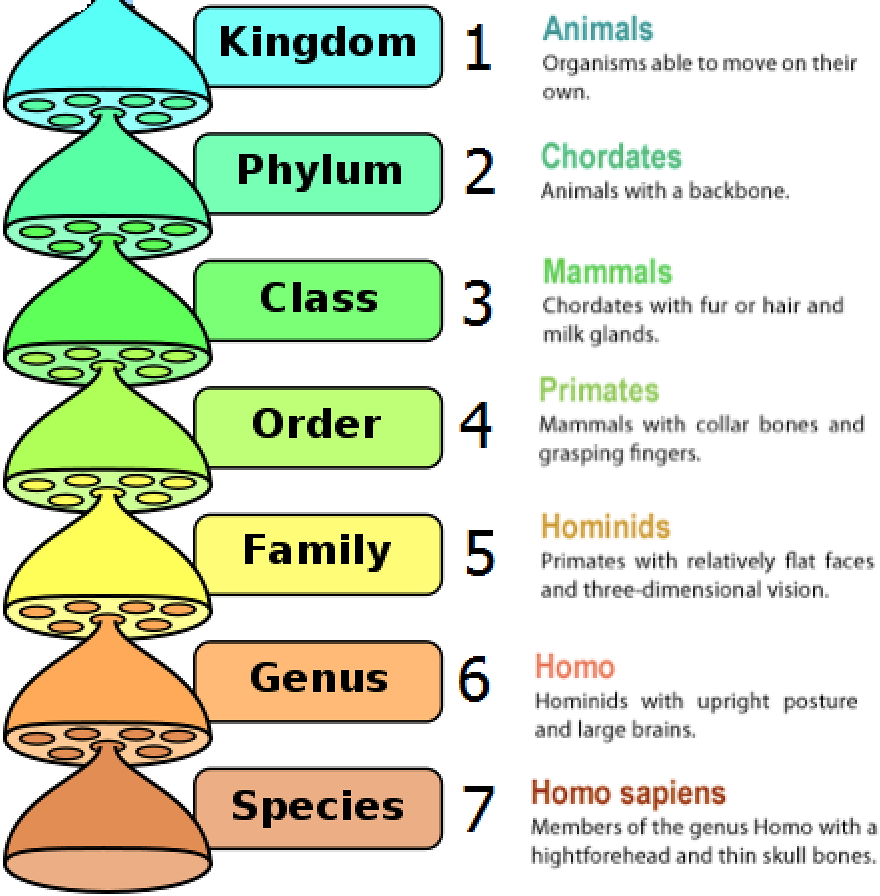
continously adjusted because of new insights, e.g.

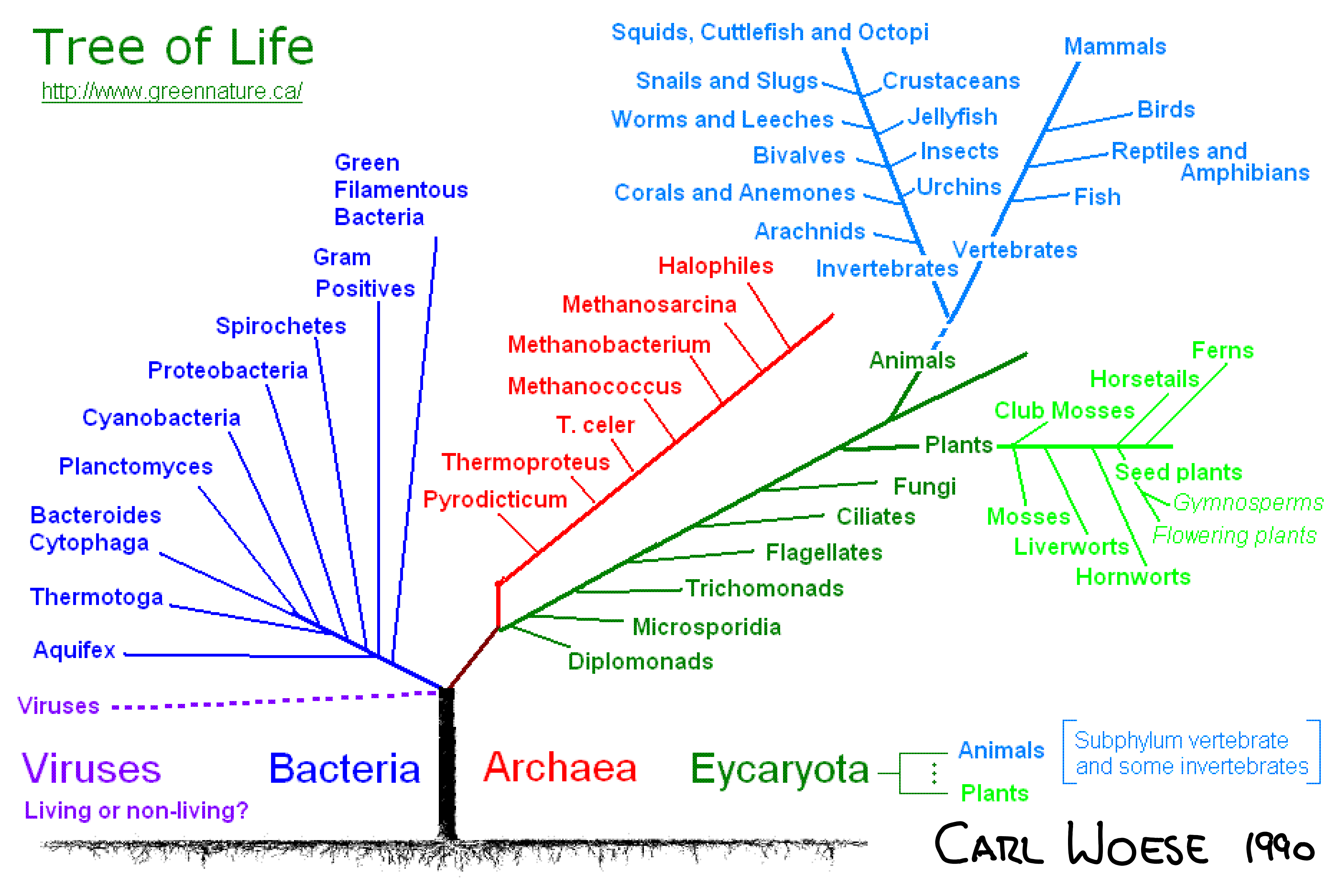
the concept of natural evolution

(On the Origin of Species, Darwin, 1859)

information stored in DNA

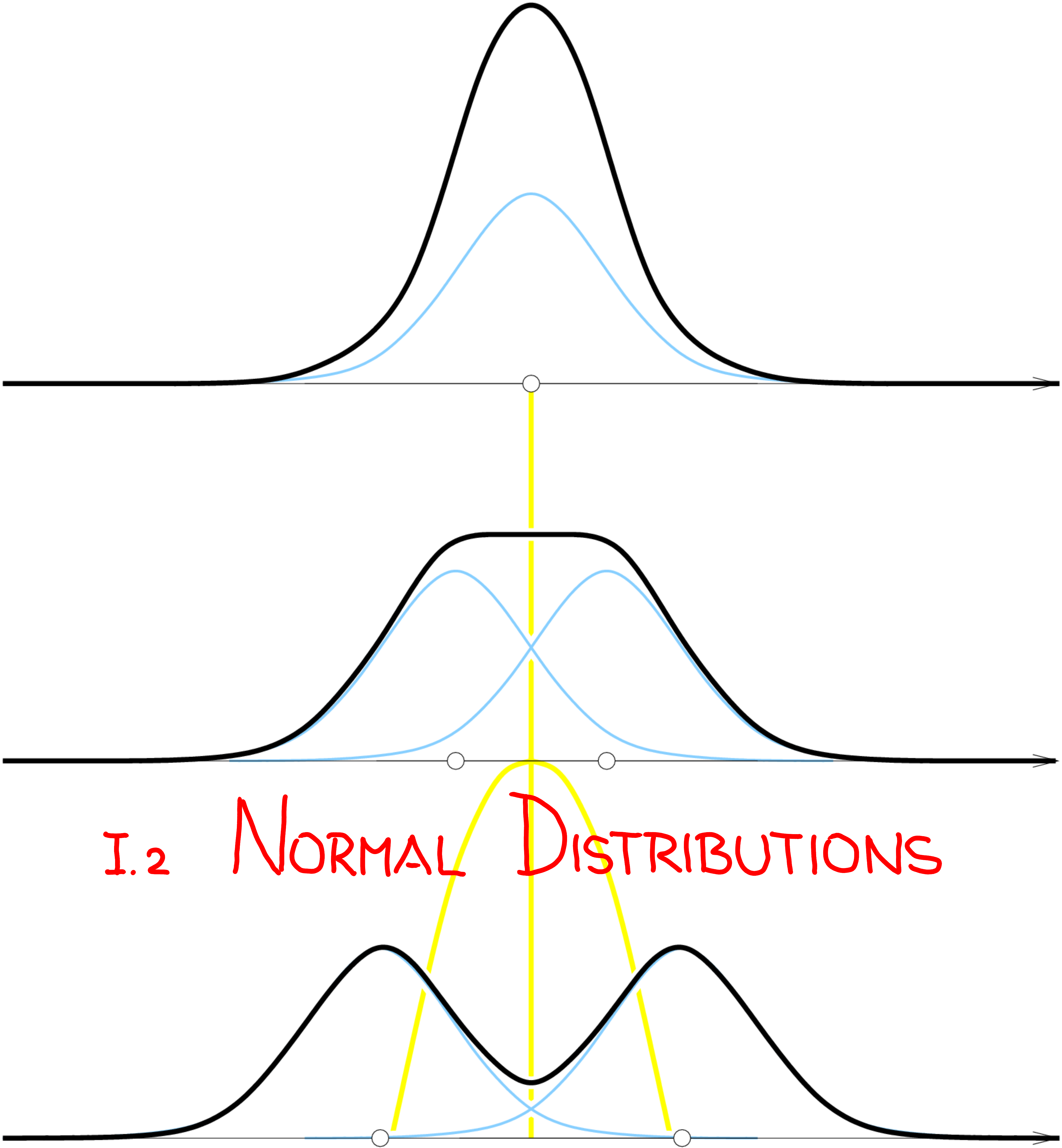
(A structure for DNA, Watson and Crick, 1953)

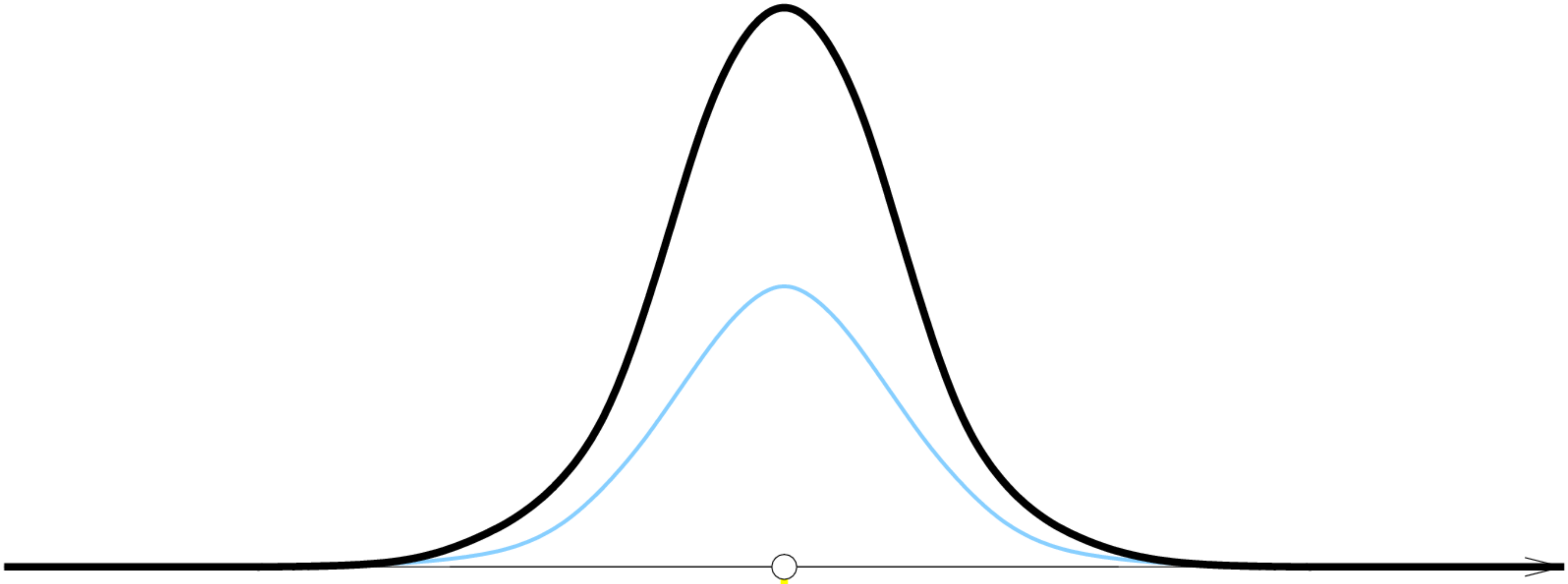

Kingdom
1
Animals
Organisms able to move on their own.
Phylum
2
Chordates
Animals with a backbone.
Class
3
Mammals
Chordates with fur or hair and milk glands.
Order
4
Primates
Mammals with collar bones and grasping fingers.
Family
5
Hominids
Primates with relatively flat faces and three-dimensional vision.
Genus
6
Homo
Hominids with upright posture and large brains.
Species
7
Homo sapiens
Members of the genus Homo with a hightforehead and thin skull bones.



Tree of Life
http://www.greennature.ca/
Squids, Cuttlefish and Octopi
Mammals
Snails and Slugs
Crustaceans
Worms and Leeches
Jellyfish
Birds
Bivalves
Insects
Reptiles and Amphibians
Corals and Anemones
Urchins
Fish
Arachnids
Vertebrates
Invertebrates
Green Filamentous Bacteria
Gram Positives
Halophiles
Methanosarcina
Spirochetes
Methanobacterium
Animals
Ferns
Proteobacteria
Methanococcus
Horsetails
Cyanobacteria
T. celer
Club Mosses
Plants
Planctomyces
Thermoproteus
Seed plants
Fungi
Pyrodicticum
Gymnosperms
Bacteroides Cytophaga
Ciliates
Mosses
Flowering plants
Flagellates
Liverworts
Thermotoga
Trichomonads
Hornworts
Microsporidia
Aquifex
Diplomonads
Viruses
Viruses
Living or non-living?
Bacteria
Archaea
Eycaryota
Animals
Plants
Subphylum vertebrate and some invertebrates
Carl Woese 1990

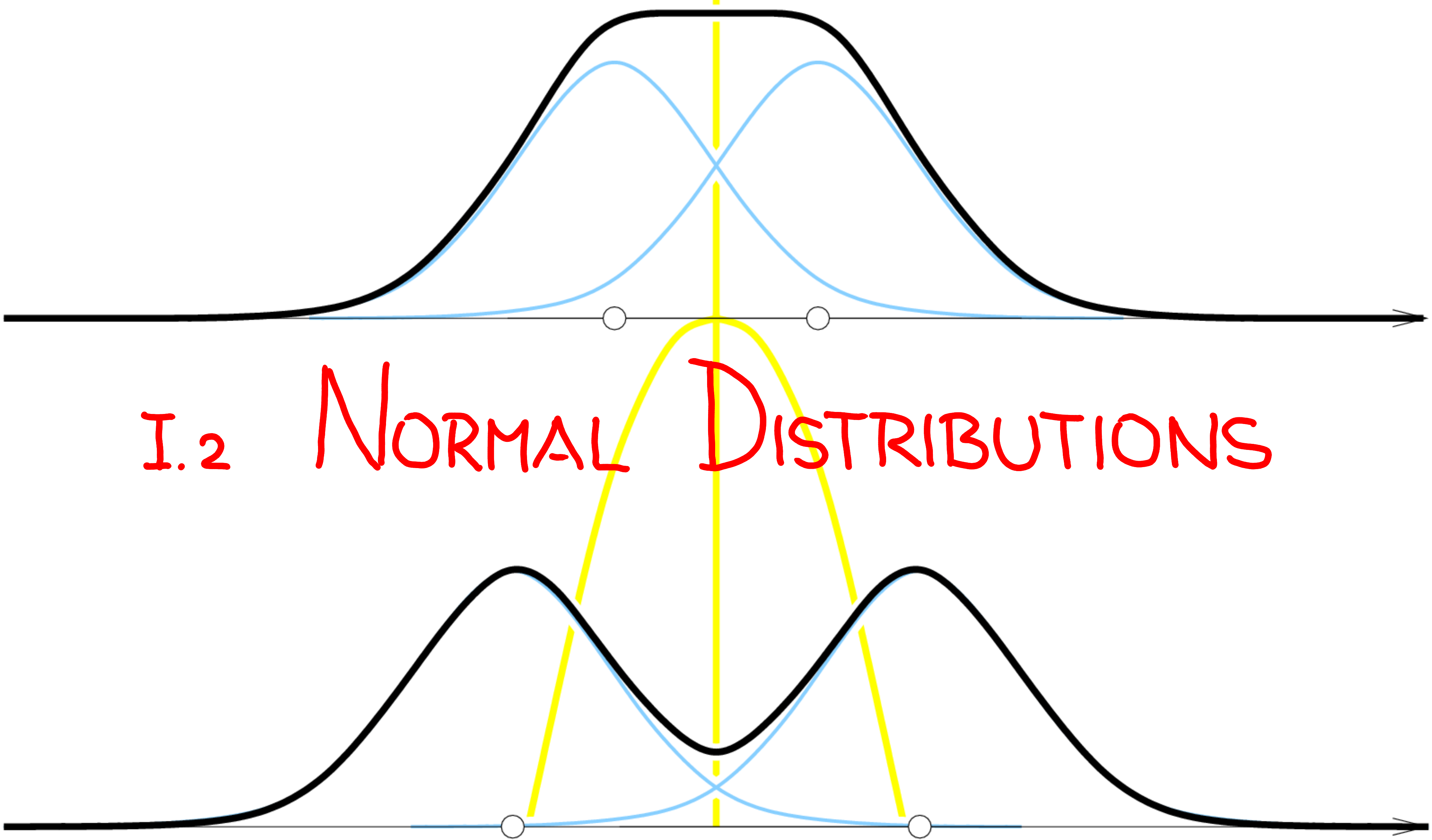
I.2 Normal Distributions





more modes than isotropic kernels in $\mathbb{R}^2$

(Carreira-Perpiñán, Williams 2003)



I.2 Normal Distributions

I.3 SAHN Clustering

(sequential agglomerative hierarchical non-overlapping)

I.3 SAHN Clustering

(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;

repeat find clusters P and Q;

merge P and Q;

until one cluster remains.

I.3 SAHN Clustering

(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;

repeat find clusters P and Q;

merge P and Q;

until one cluster remains.

e.g. single-linkage:

$$d(P,Q) = \min_{p \in P,\, q \in Q} d(p,q)$$

I.3 SAHN Clustering

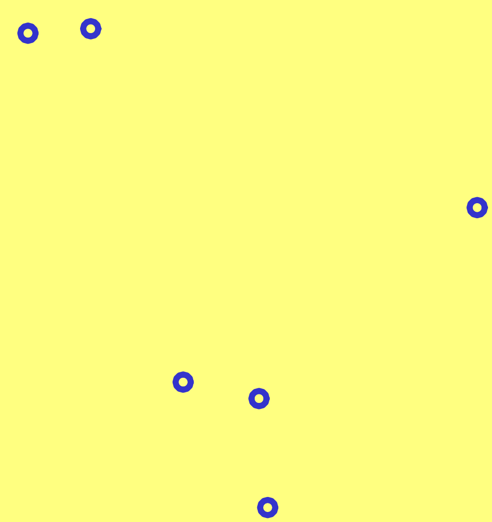
(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;

repeat find clusters P and Q;

merge P and Q;

until one cluster remains.



e.g. single-linkage:

$$d(P, Q) = \min_{p \in P,\, q \in Q} d(p, q)$$

I.3 SAHN Clustering

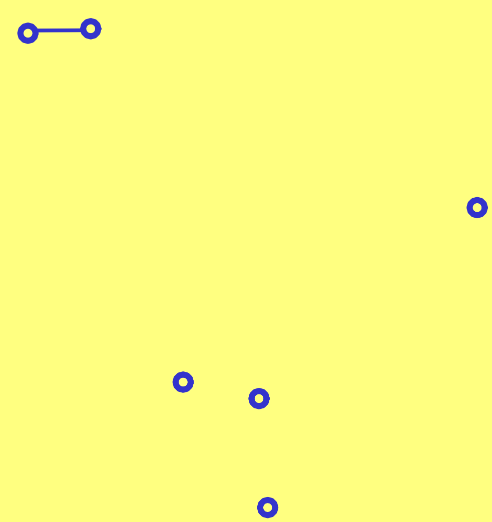
(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;

repeat find clusters P and Q;

merge P and Q;

until one cluster remains.



e.g. single-linkage:

$$d(P,Q) = \min_{p \in P,\ q \in Q} d(p,q)$$

I.3 SAHN Clustering

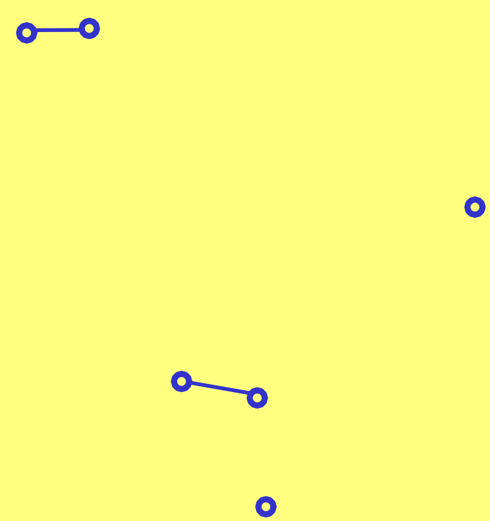
(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;

repeat find clusters P and Q;

merge P and Q;

until one cluster remains.



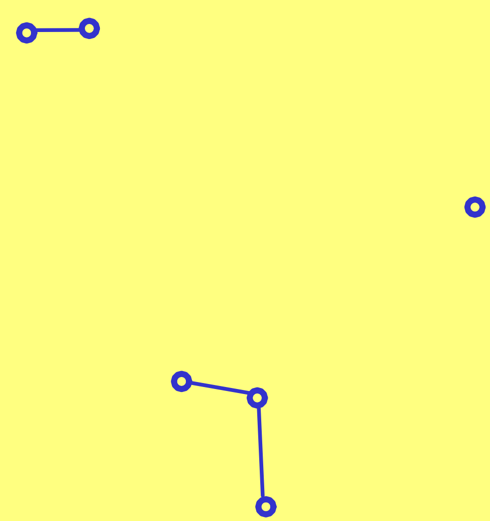
e.g. single-linkage:

$$d(P,Q) = \min_{p \in P,\, q \in Q} d(p,q)$$

I.3 SAHN Clustering

(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;
repeat find clusters P and Q;
merge P and Q;
until one cluster remains.



e.g. single-linkage:

$$d(P,Q) = \min_{p \in P,\, q \in Q} d(p,q)$$

I.3 SAHN Clustering

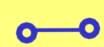
(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;

repeat find clusters P and Q;

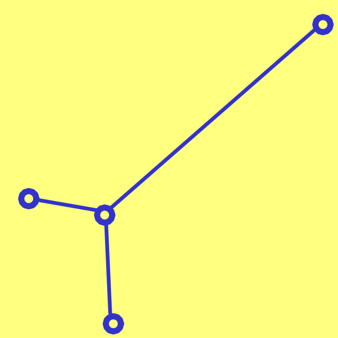
merge P and Q;

until one cluster remains.



e.g. single linkage:

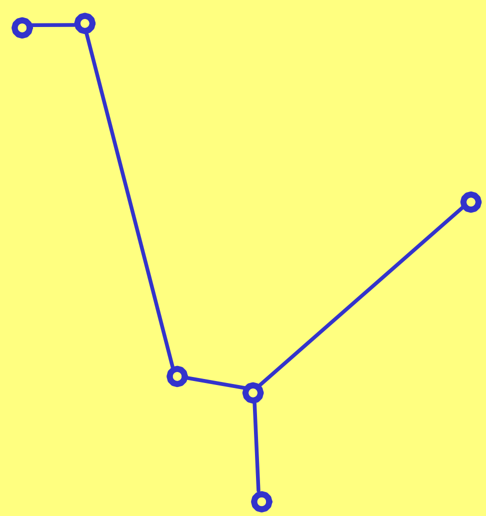
$$d(P,Q) = \min_{p \in P,\, q \in Q} d(p,q)$$



I.3 SAHN Clustering

(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;
repeat find clusters P and Q;
    merge P and Q;
until one cluster remains.



e.g. single-linkage:

$$d(P,Q) = \min_{p\in P,\, q\in Q} d(p,q)$$

I.3 SAHN Clustering

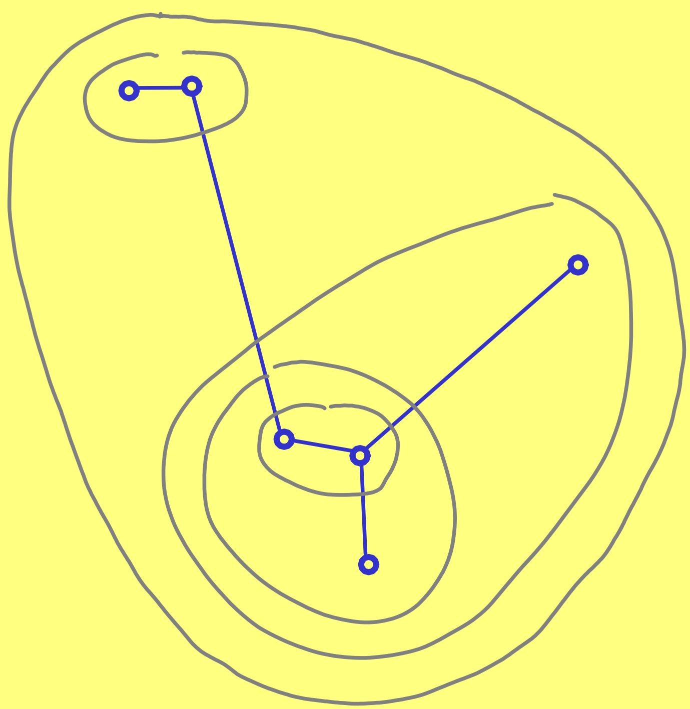
(sequential agglomerative hierarchical non-overlapping)

begin with singleton clusters;

repeat find clusters P and Q;

merge P and Q;

until one cluster remains.



e.g. single-linkage:

$$d(P,Q) = \min_{p\in P,\, q\in Q} d(p,q)$$

I Classification

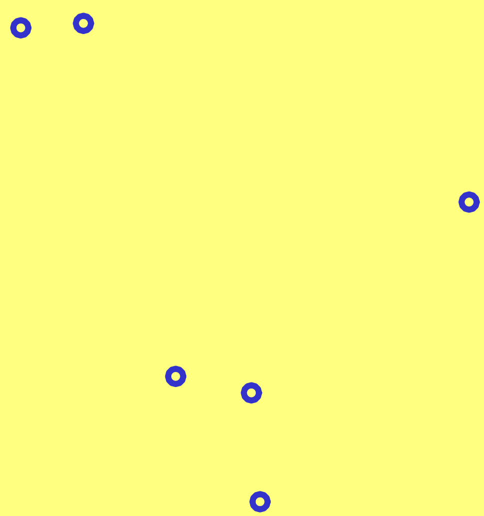
II Shape

III Scale

IV Applications

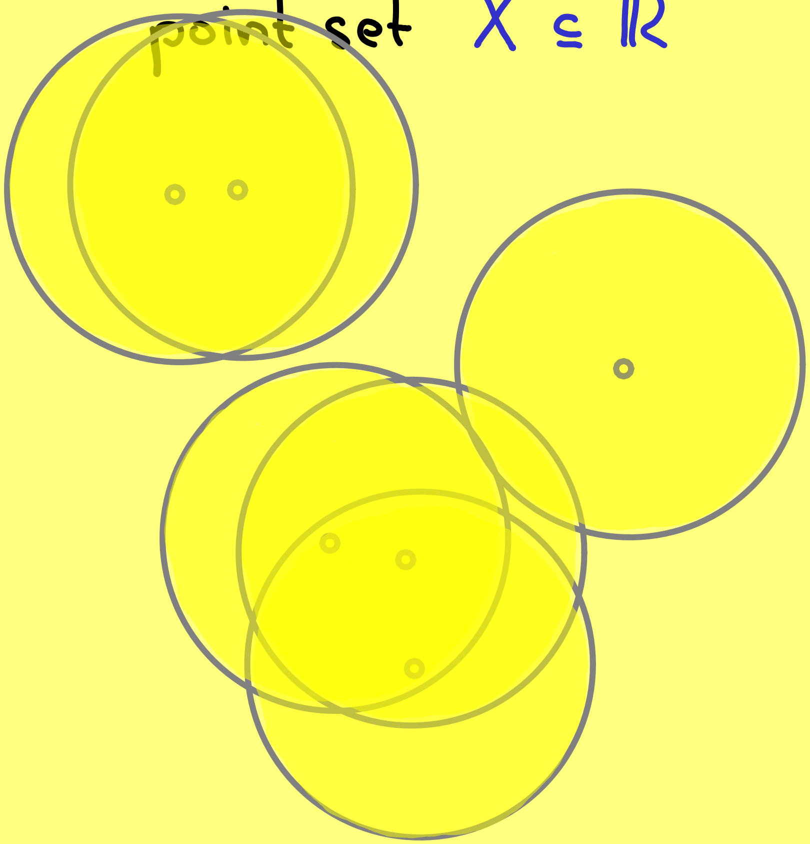
II.1 Čech Complex

point set $X \subseteq \mathbb{R}^n$



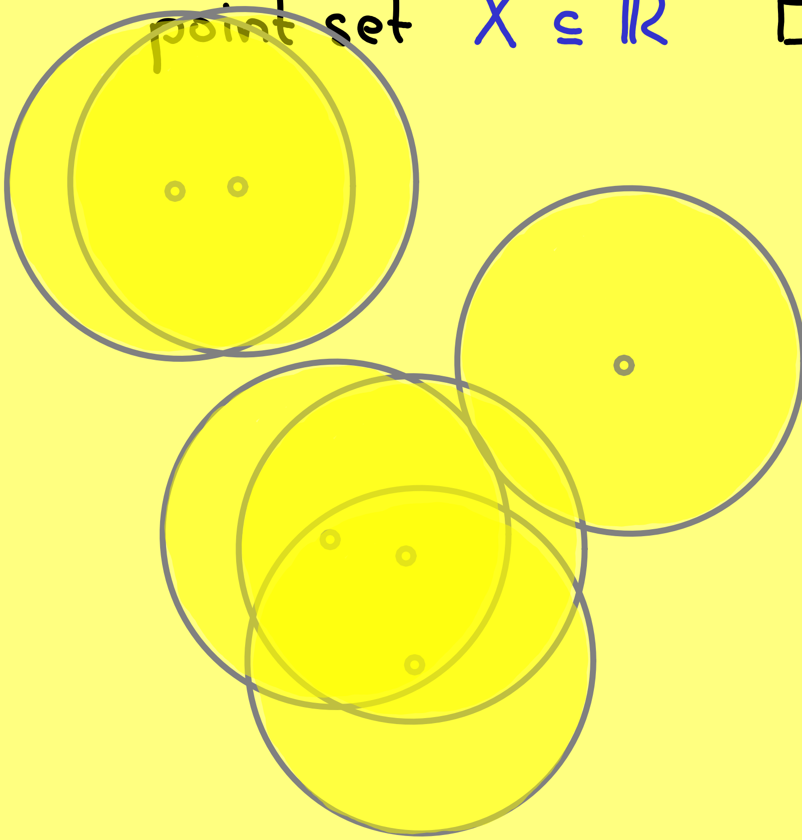
II.1 Čech Complex

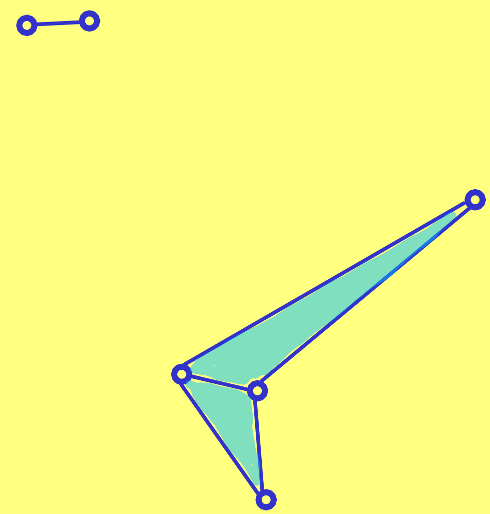
point set $X \subseteq \mathbb{R}^n$



II.1 Čech Complex

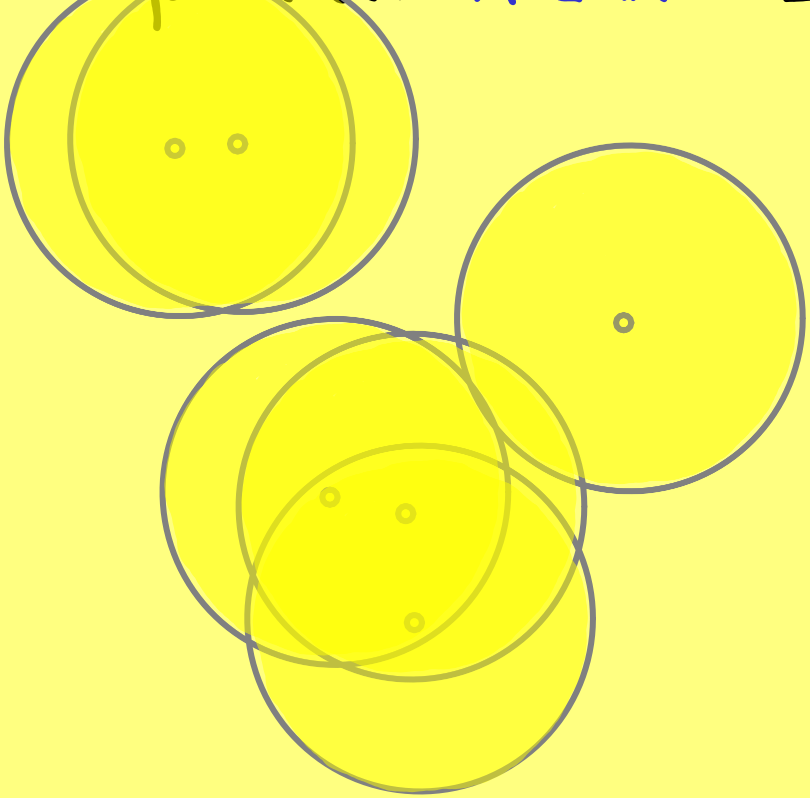
point set $X \subseteq \mathbb{R}^n$ $\Longrightarrow$ simplicial complex $C_r(X)$

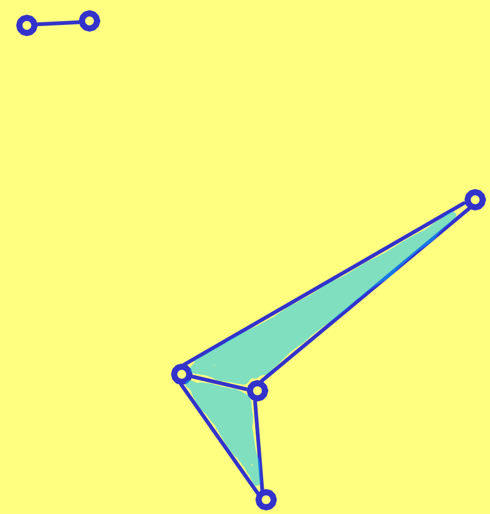




II.1 Čech Complex

point set $X \subseteq \mathbb{R}^n$ $\Longrightarrow$ simplicial complex $C_r(X)$

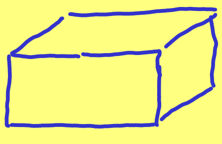




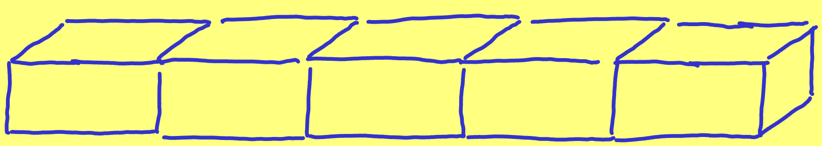
Čech complex for $r \geq 0$ is $\check{C}_r(X) = \{P \subseteq X \mid \bigcap_{x \in P} B_r(x) \neq \emptyset\}$.

II.3 Chains, Boundaries, Cycles

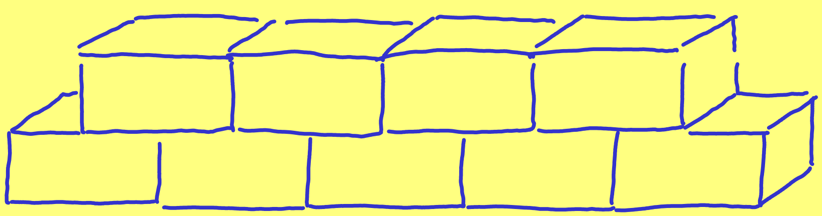
II.3 Chains, Boundaries, Cycles



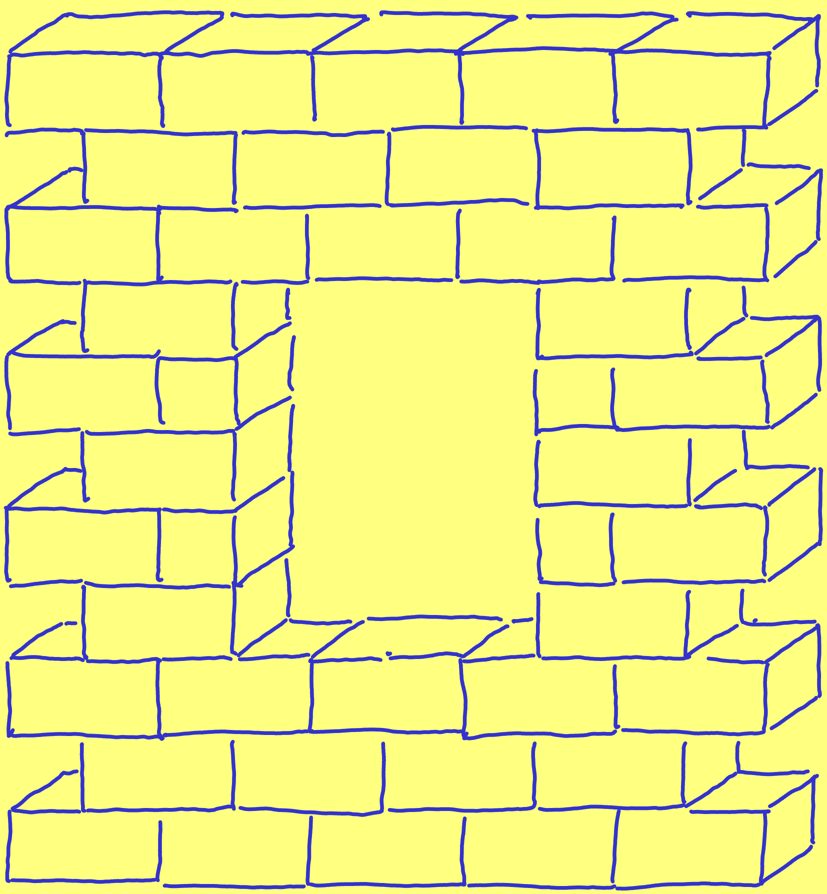
II.3 Chains, Boundaries, Cycles



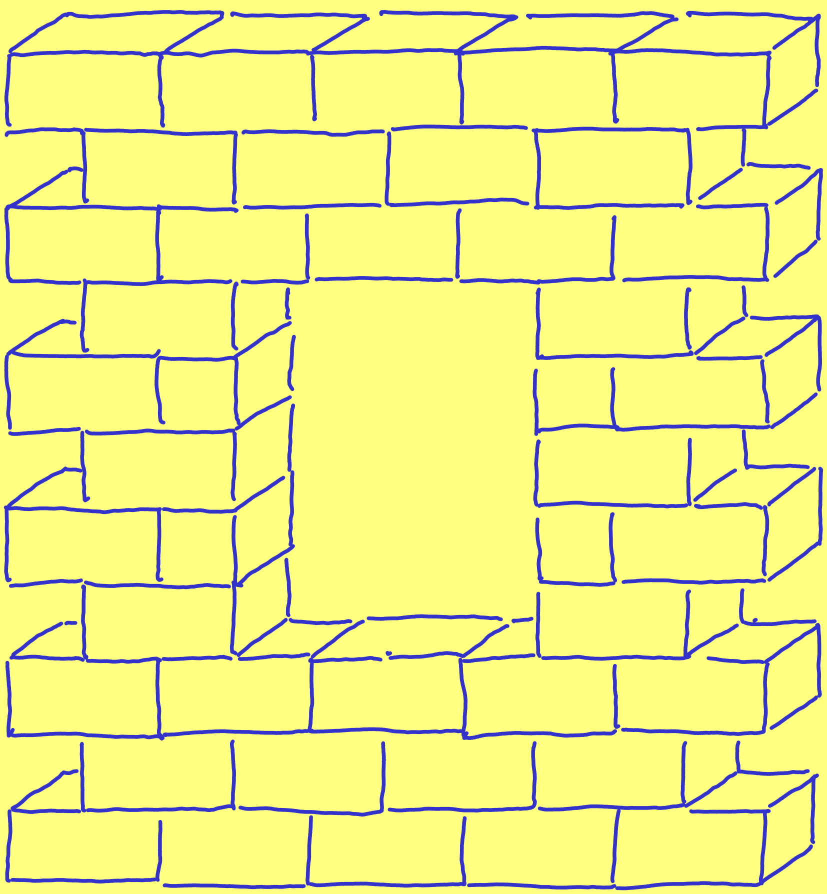
II.3 Chains, Boundaries, Cycles



II.3 Chains, Boundaries, Cycles

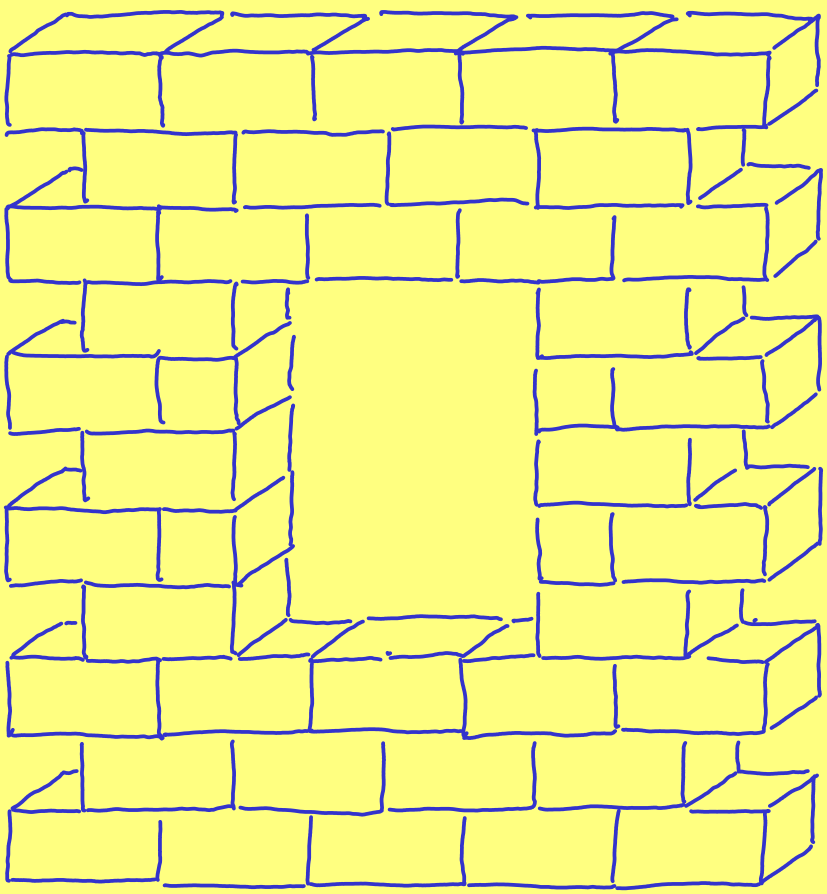


II.3 Chains, Boundaries, Cycles



k-chain is sum of k-dim. bricks

II.3 CHAINS, BOUNDARIES, CYCLES

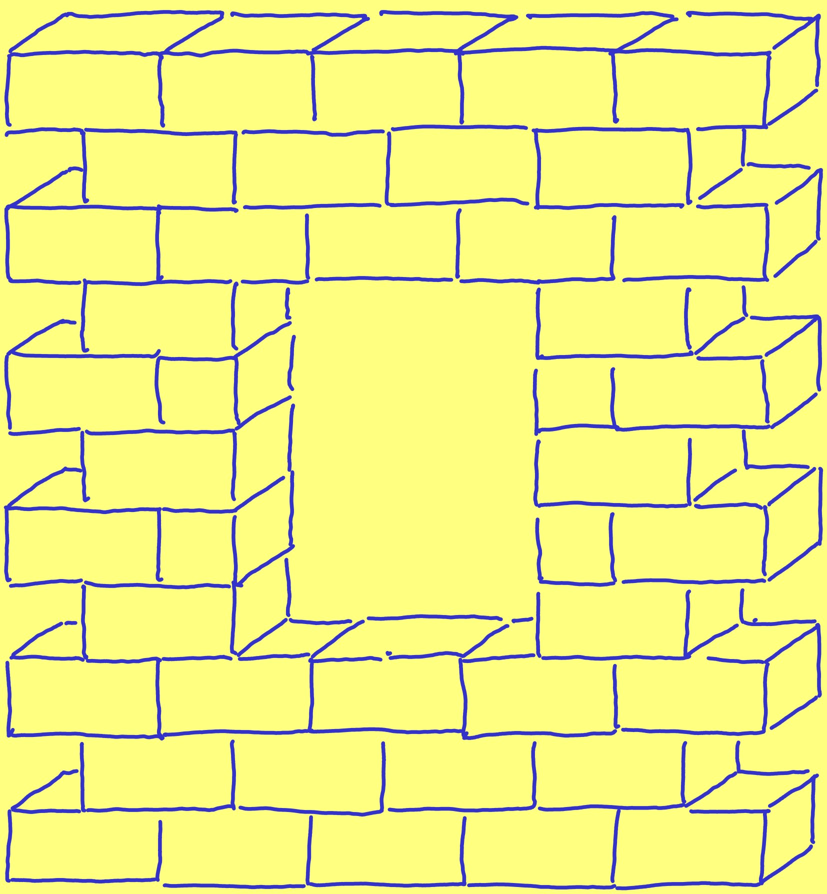


k-chain is sum of k-dim. bricks

(k-1)-boundary $= \partial$ k-chain

$=$ sum of ∂bricks

II.3 Chains, Boundaries, Cycles



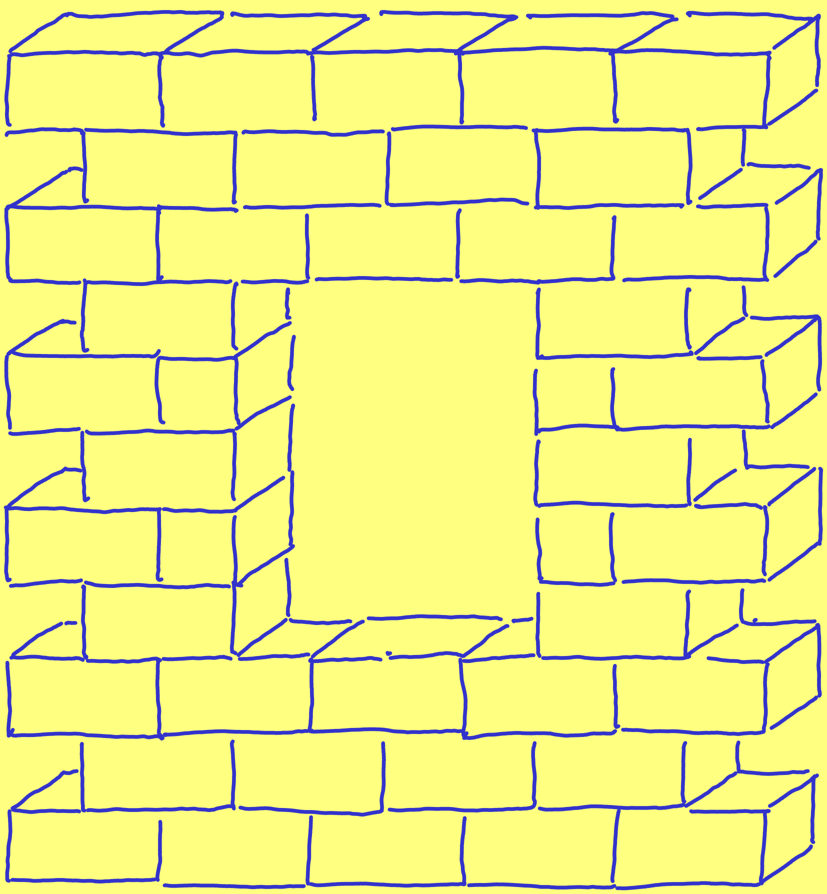
k-chain is sum of k-dim. bricks

(k-1)-boundary = ∂ k-chain

= sum of ∂bricks

∂ k-cycle = 0.

II.3 CHAINS, BOUNDARIES, CYCLES



k-chain is sum of k-dim. bricks

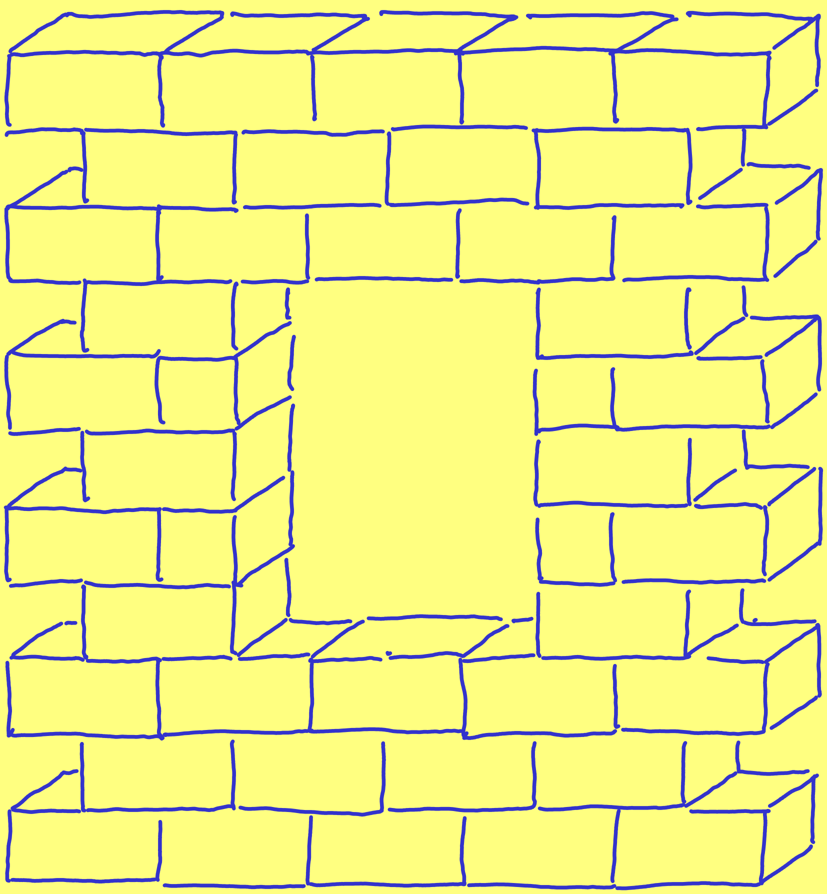
(k-1)-boundary $= \partial$ k-chain

$= $ sum of ∂bricks

∂ k-cycle $= 0$.

$$B_k \subseteq Z_k \subseteq C_k$$

II.3 Chains, Boundaries, Cycles



k-chain is sum of k-dim. bricks

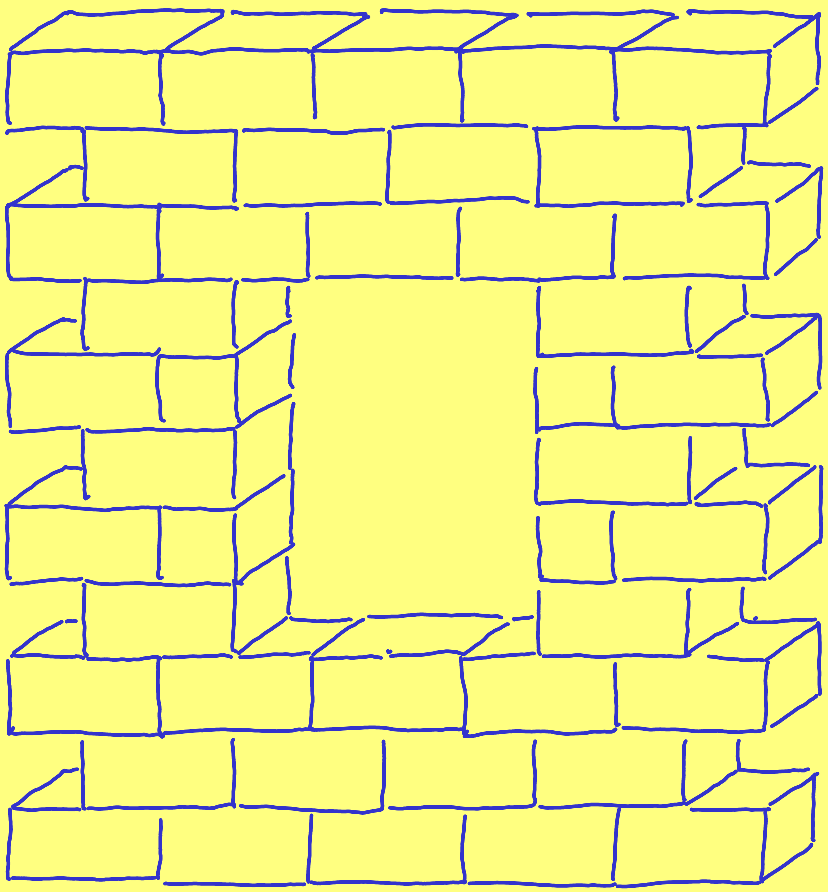
(k-1)-boundary = ∂ k-chain

= sum of ∂bricks

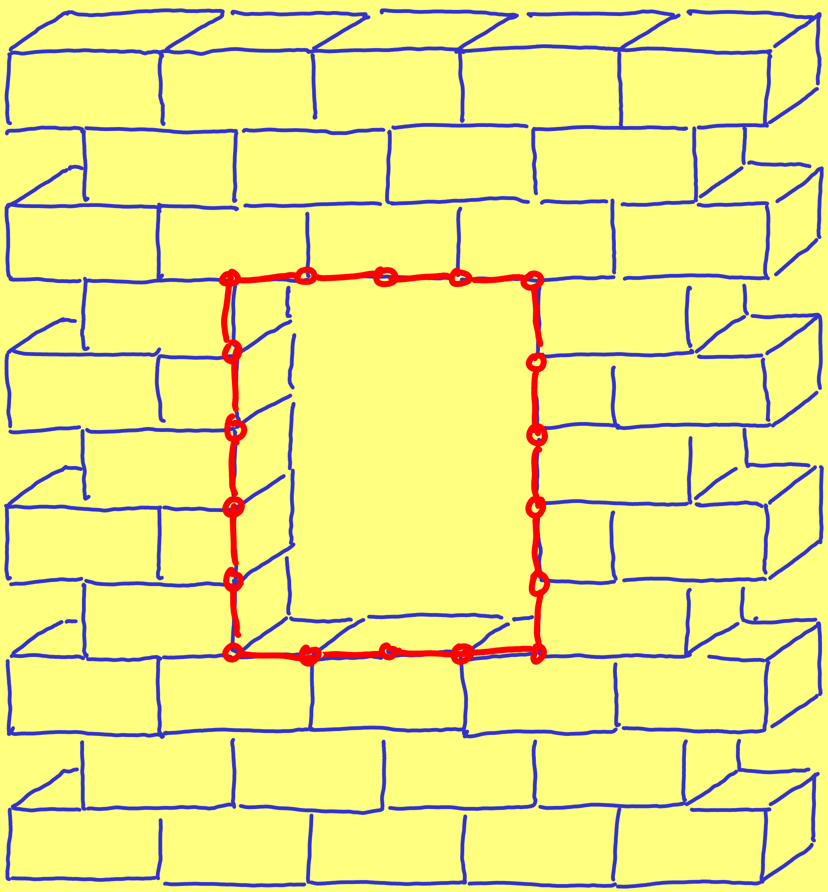
∂ k-cycle = 0.

$$\begin{array}{ccccc} B_k & \subseteq & Z_k & \subseteq & C_k \\ & \swarrow^{\partial} & & & \\ B_{k-1} & \subseteq & Z_{k-1} & \subseteq & C_{k-1} \end{array}$$

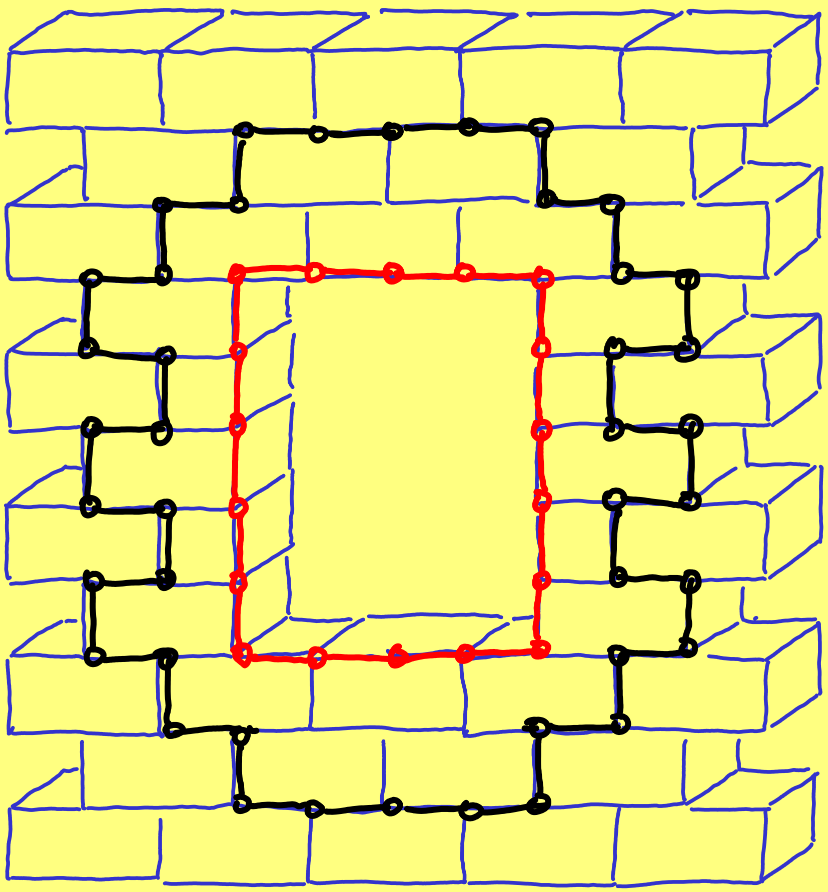
II.4 HOMOLOGY



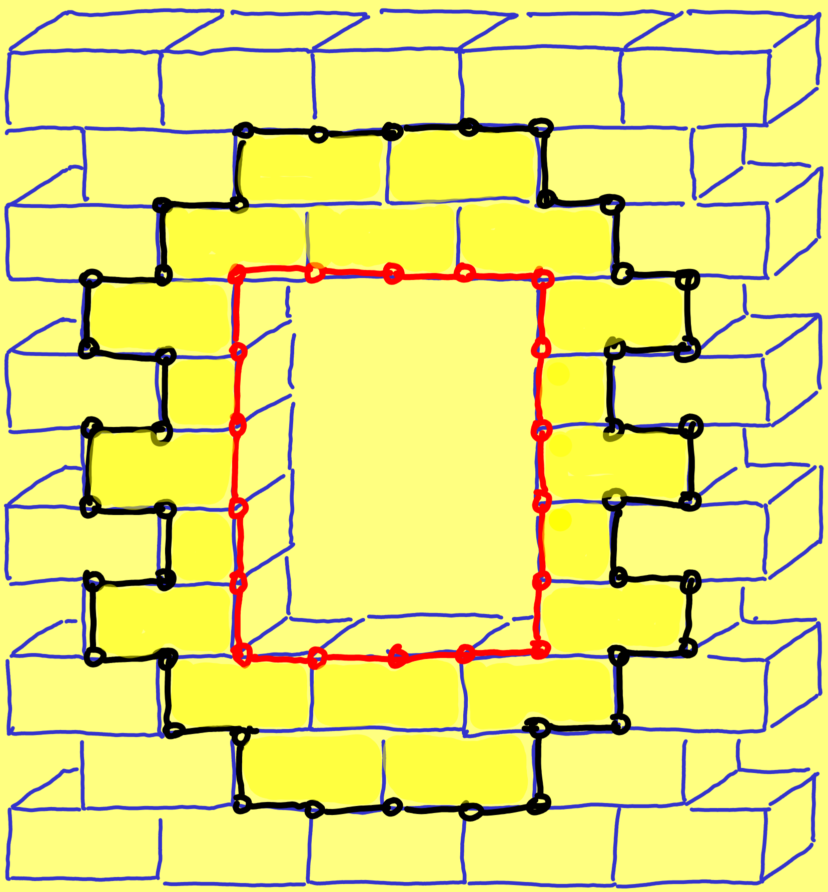
II.4 Homology



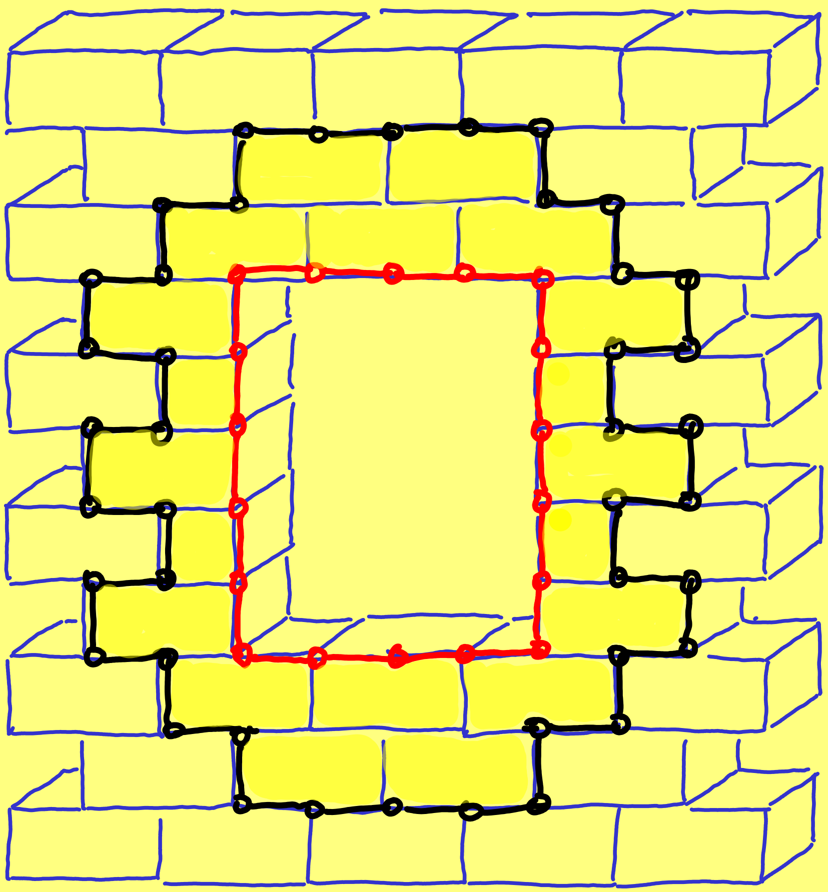
II.4 Homology



II.4 Homology



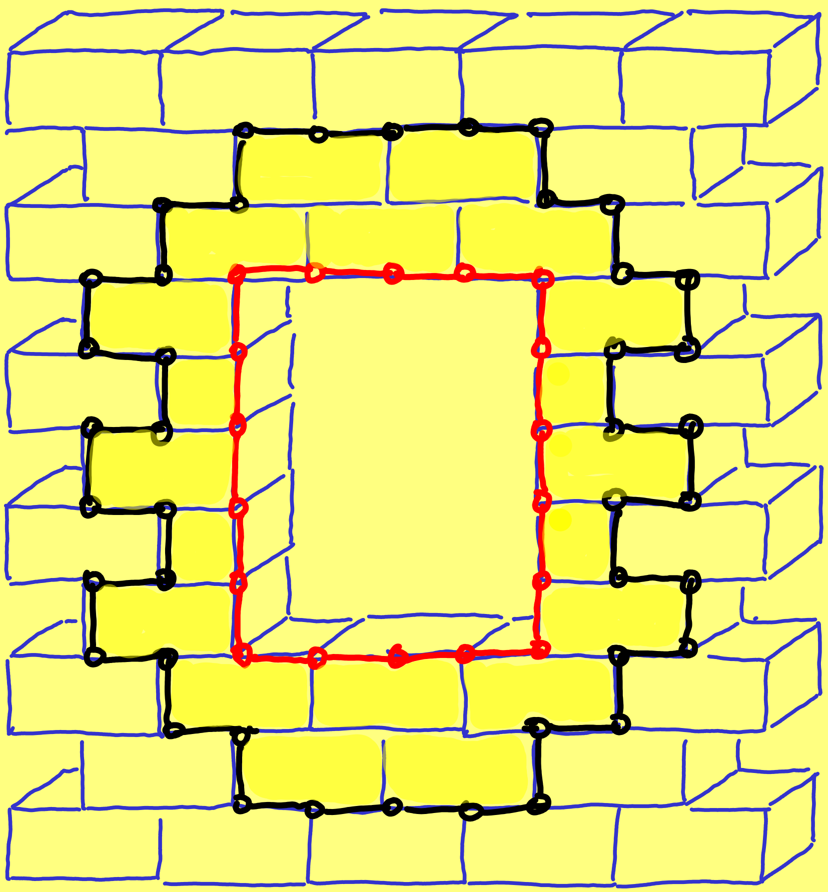
II.4 Homology



k-th homology group is

$$H_k = Z_k / B_k$$

II.4 HOMOLOGY



k-th homology group is

$$H_k = Z_k / B_k$$

k-th Betti number is

$$\beta_k = \text{rank } H_k$$

II.5 INCREMENTAL ALGORITHM

BETTI #s IN $\mathbb{R}^3$:

β_0 = #components

β_1 = #tunnels

β_2 = #voids

II.5 INCREMENTAL ALGORITHM

BETTI #s IN $\mathbb{R}^3$:

β_0 = # components
β_1 = # tunnels
β_2 = # voids

VERTEX $\bullet$ $\beta_0{+}{+}$

II.5 Incremental Algorithm

Betti #s in $\mathbb{R}^3$:

β_0 = #components

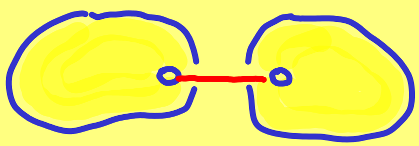
β_1 = #tunnels

β_2 = #voids

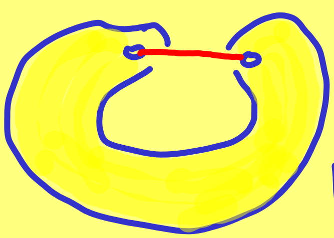
Vertex

$\beta_0{++}$

Edge



$\beta_0{--}$



$\beta_1{++}$

II.5 INCREMENTAL ALGORITHM

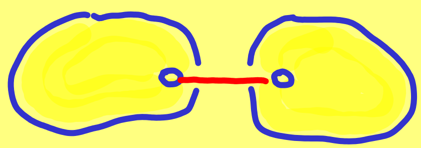
BETTI #s IN $\mathbb{R}^3$:

$\beta_0 =$ #components
$\beta_1 =$ #tunnels
$\beta_2 =$ #voids

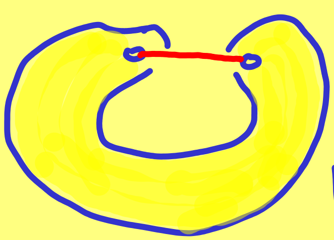
VERTEX

$\beta_0{++}$

EDGE

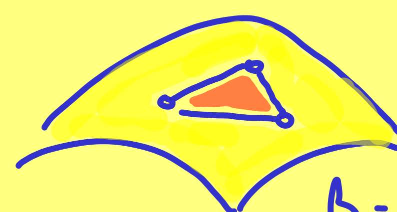


$\beta_0{--}$

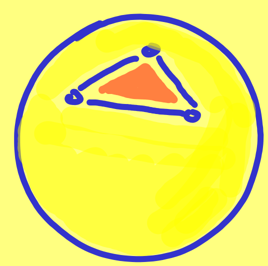


$\beta_1{++}$

TRIANGLE



$\beta_1{--}$



$\beta_2{++}$

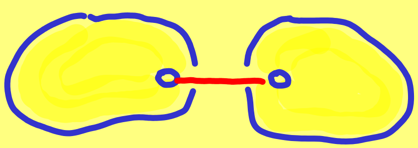
II.5 Incremental Algorithm

Betti #s in $\mathbb{R}^3$:

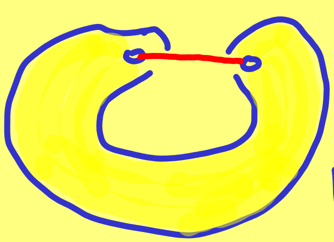
β_0 = #components
β_1 = #tunnels
β_2 = #voids

Vertex β_0++

Edge

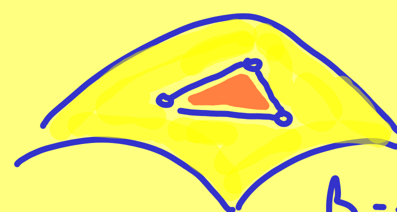


β_0--



β_1++

Triangle



β_1--

β_2++

Tetrahedron β_2--

II.5 INCREMENTAL ALGORITHM

$\beta_0 = \beta_1 = \ldots = \beta_n = 0;$

for $i = 1$ **to** m **do**

  $k = \dim P_i$;

  if $P_i \in k\text{-cycle}$ **then** $\beta_i{++}$

  else $\beta_{i-1}{--}$

  endif

endfor

II.5 INCREMENTAL ALGORITHM

$\beta_0 = \beta_1 = \ldots = \beta_n = 0;$

for $i = 1$ **to** m **do**

$\quad k = \dim P_i;$

$\quad$ **if** $P_i \in k\text{-cycle}$ **then** $\beta_i{++}$ (birth)

$\quad\quad$ **else** $\beta_{i-1}{--}$ (death)

$\quad$ **endif**

endfor

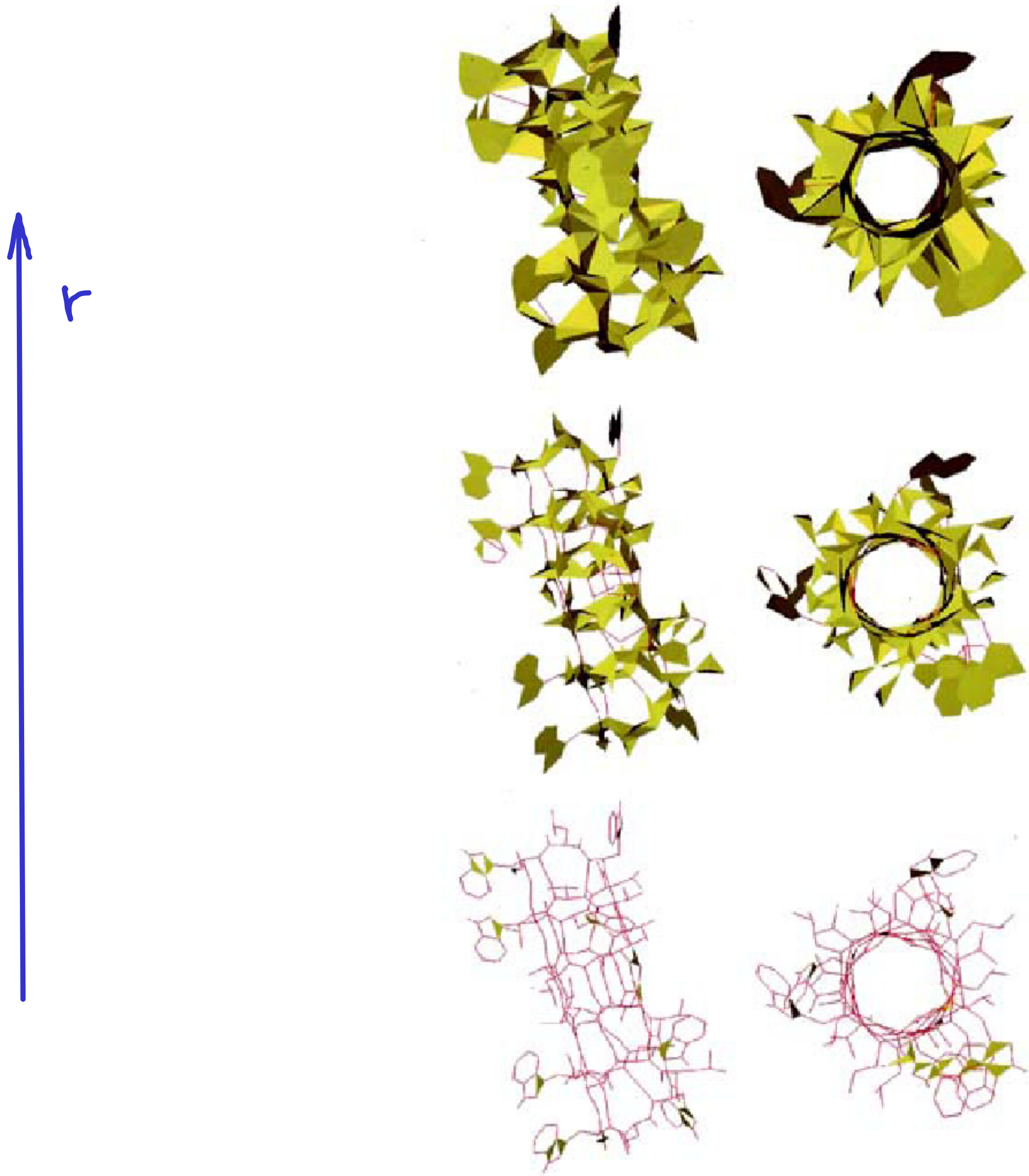
I Classification

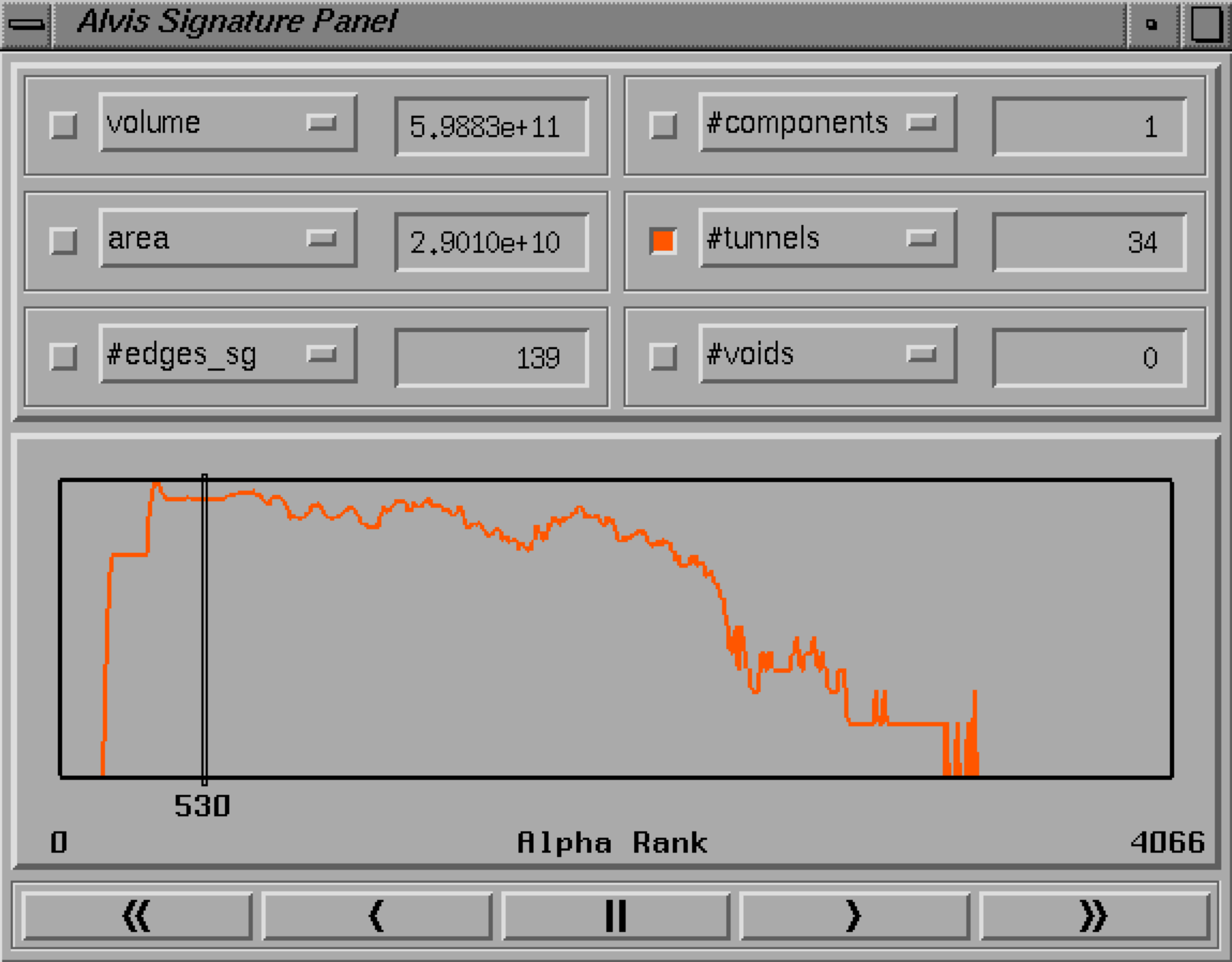
II Shape

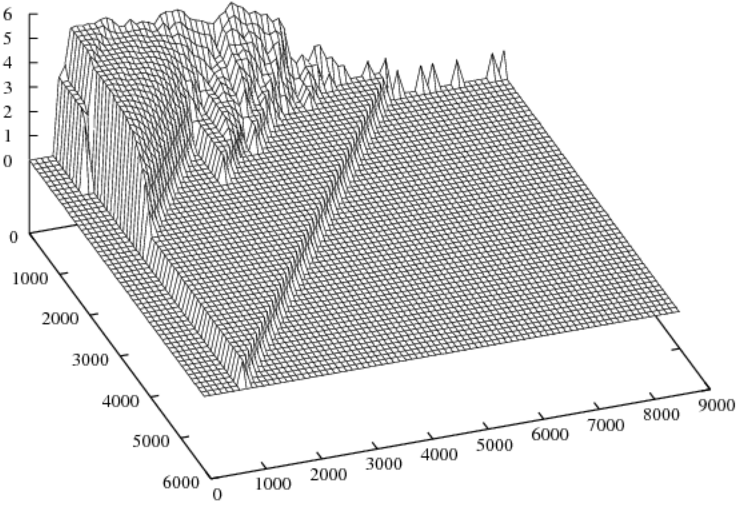
III Scale

IV Applications

III.1 Number of Tunnels




Alvis Signature Panel
volume
5.9883e+11
#components
1
area
2.9010e+10
#tunnels
34
#edges_sg
139
#voids
0
530
0
Alpha Rank
4066



6
5
4
3
2
1
0
0
1000
2000
3000
4000
5000
6000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

III.2 FILTRATION

$$\ldots \subseteq C_{i-1} \subseteq C_i \subseteq \ldots \subseteq C_{j-1} \subseteq C_j \subseteq \ldots$$

III.2 FILTRATION

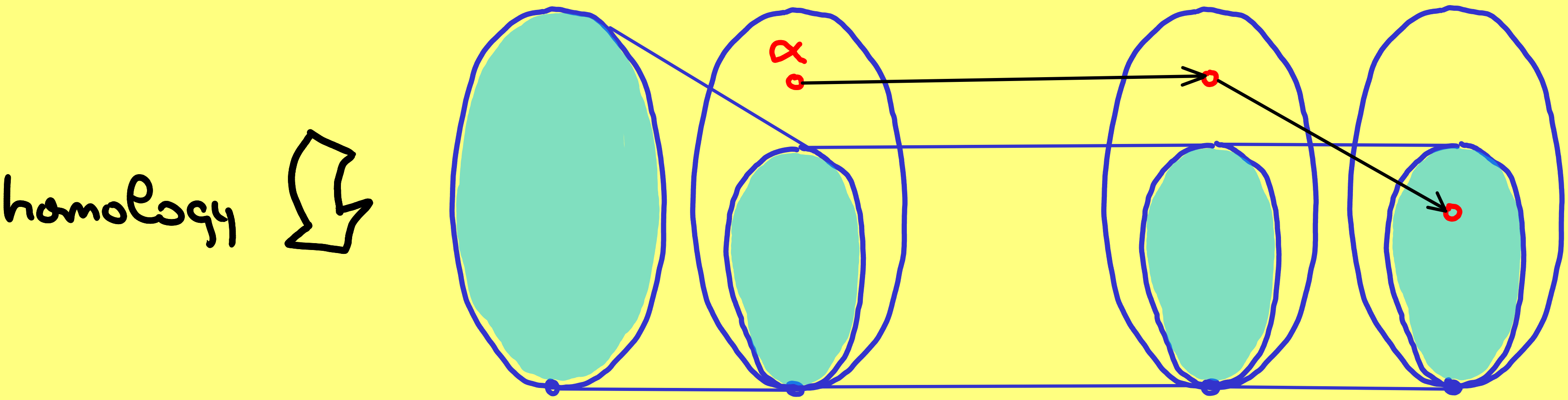
$$\ldots \subseteq C_{i-1} \subseteq C_i \subseteq \ldots \subseteq C_{j-1} \subseteq C_j \subseteq \ldots$$

homology $\Downarrow$

$$\ldots \to H_{i-1} \to H_i \to \ldots \to H_{j-1} \to H_j \to \ldots$$

III.2 FILTRATION

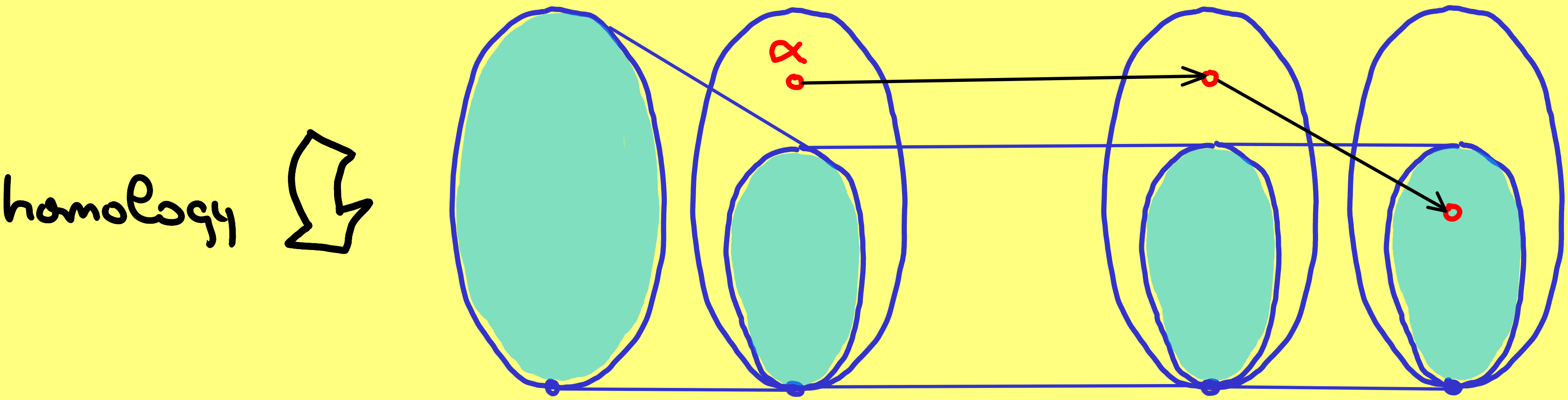
$$\ldots \subseteq C_{i-1} \subseteq C_i \subseteq \ldots \subseteq C_{j-1} \subseteq C_j \subseteq \ldots$$



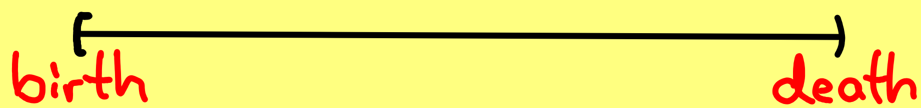
$$\ldots \to H_{i-1} \to H_i \to \ldots \to H_{j-1} \to H_j \to \ldots$$

III.2 Filtration

$$\ldots \subseteq C_{i-1} \subseteq C_i \subseteq \ldots \subseteq C_{j-1} \subseteq C_j \subseteq \ldots$$



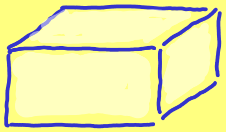
$$\ldots \to H_{i-1} \to H_i \to \ldots \to H_{j-1} \to H_j \to \ldots$$



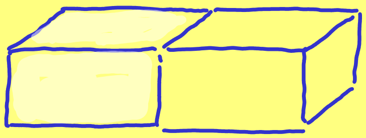
pers = death − birth

III.3 Birth and Death

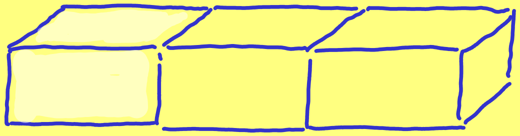
III.3 Birth and Death



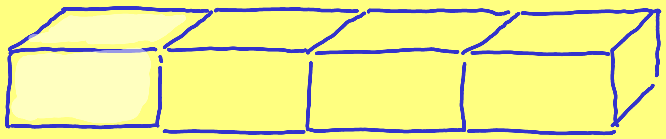
III.3 Birth and Death



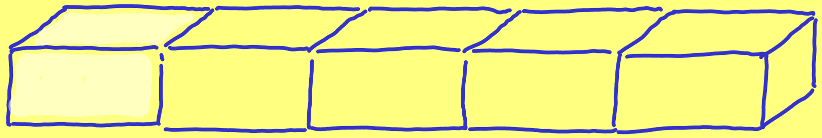
III.3 Birth and Death



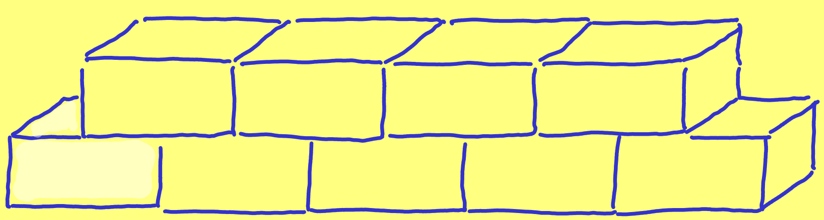
III.3 Birth and Death



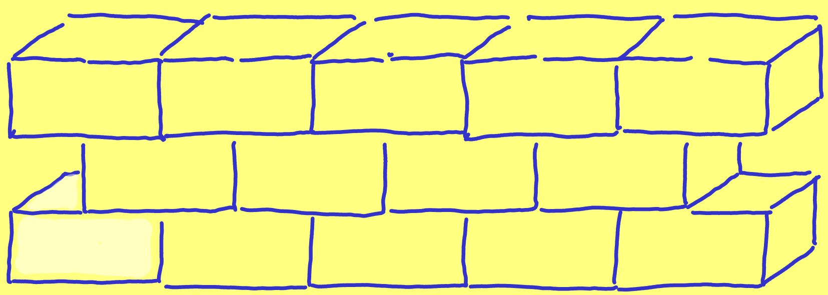
III.3 Birth and Death



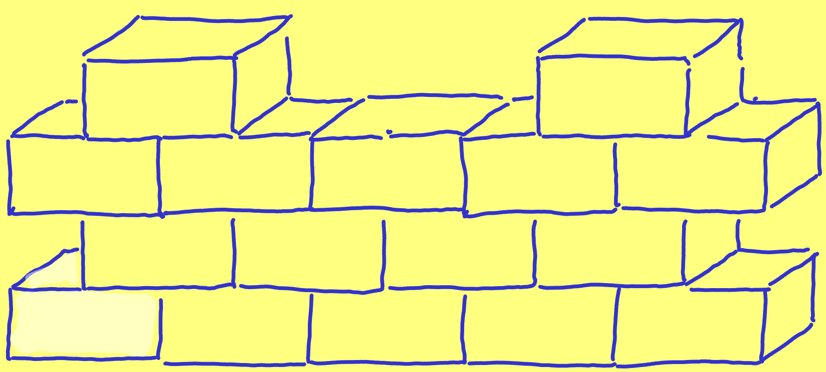
III.3 Birth and Death



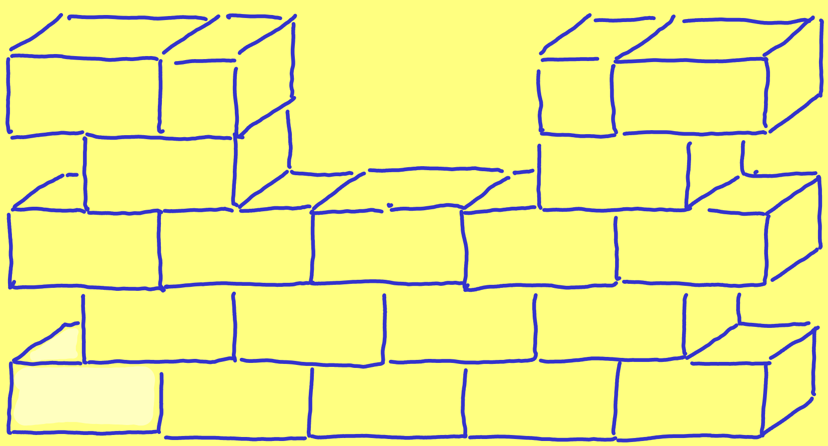
III.3 Birth and Death



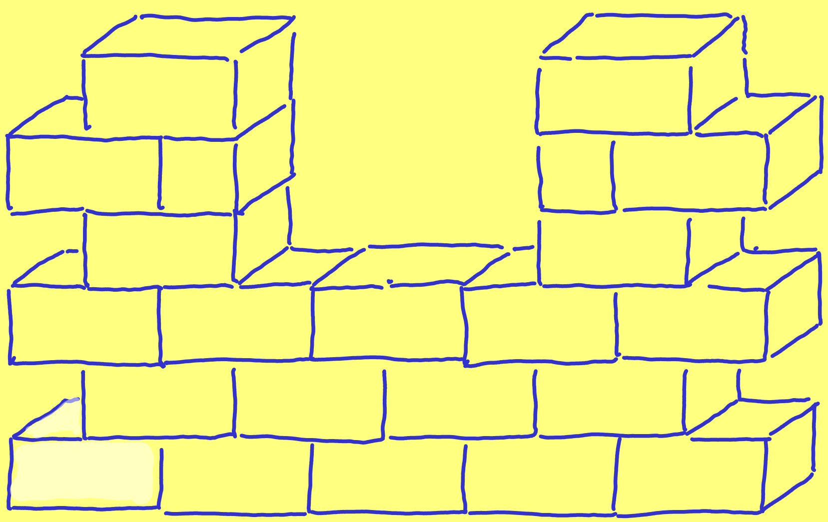
III.3 Birth and Death



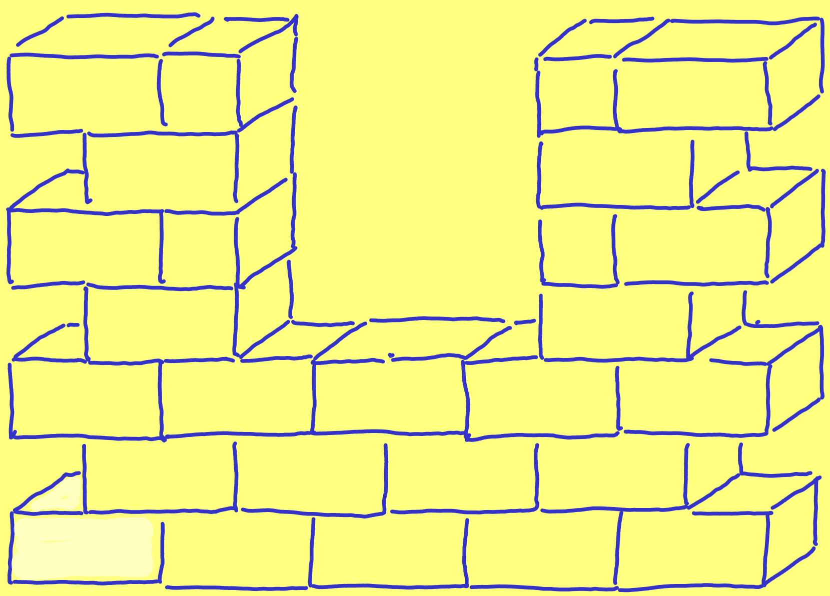
III.3 Birth and Death



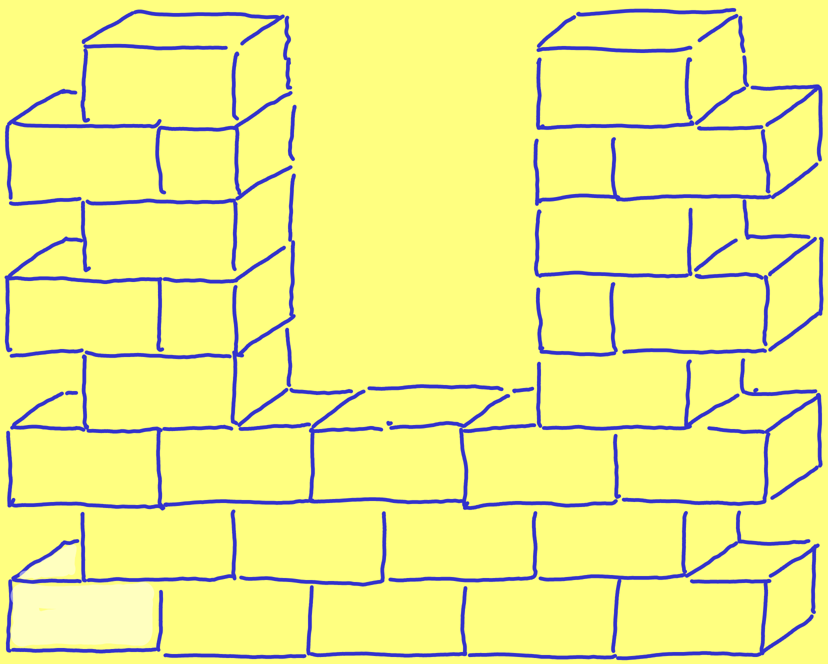
III.3 Birth and Death



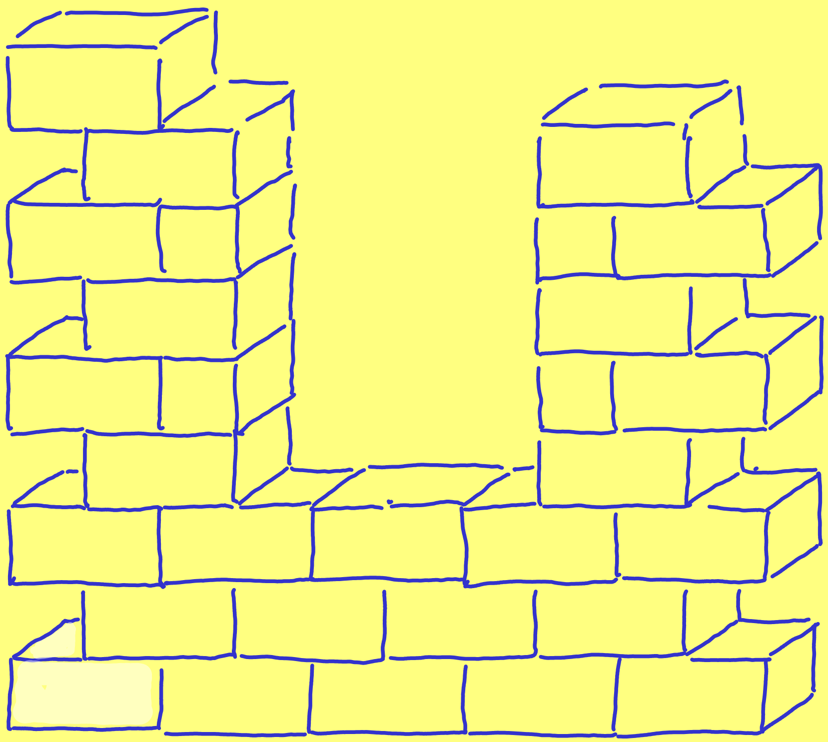
III.3 Birth and Death



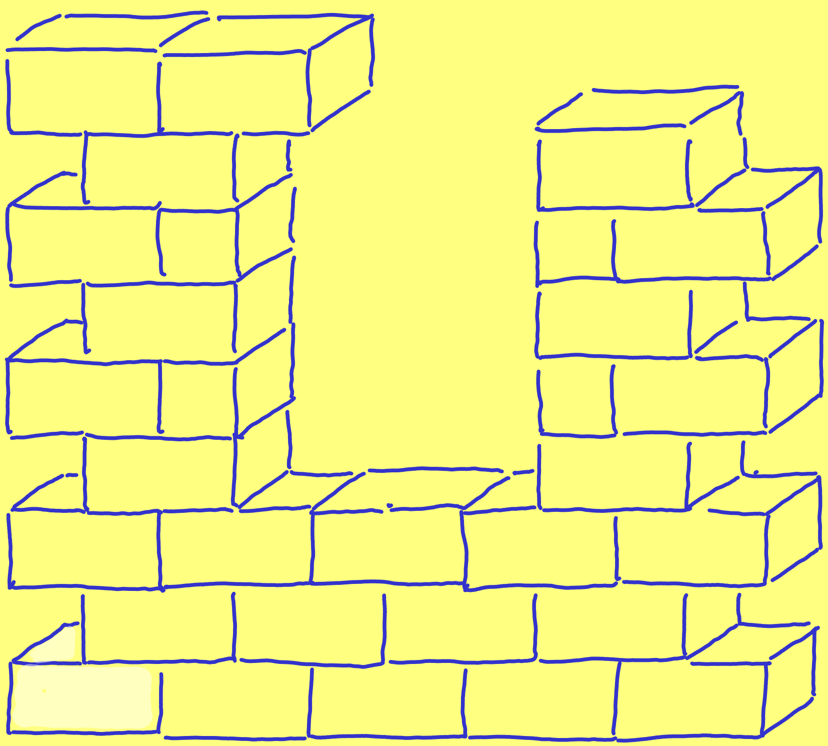
III.3 Birth and Death



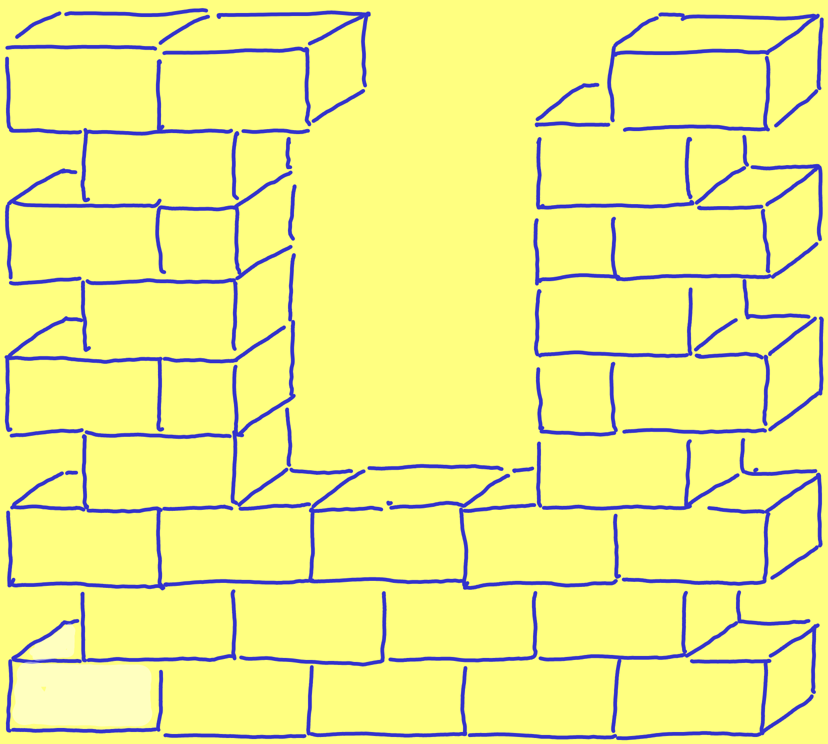
III.3 Birth and Death



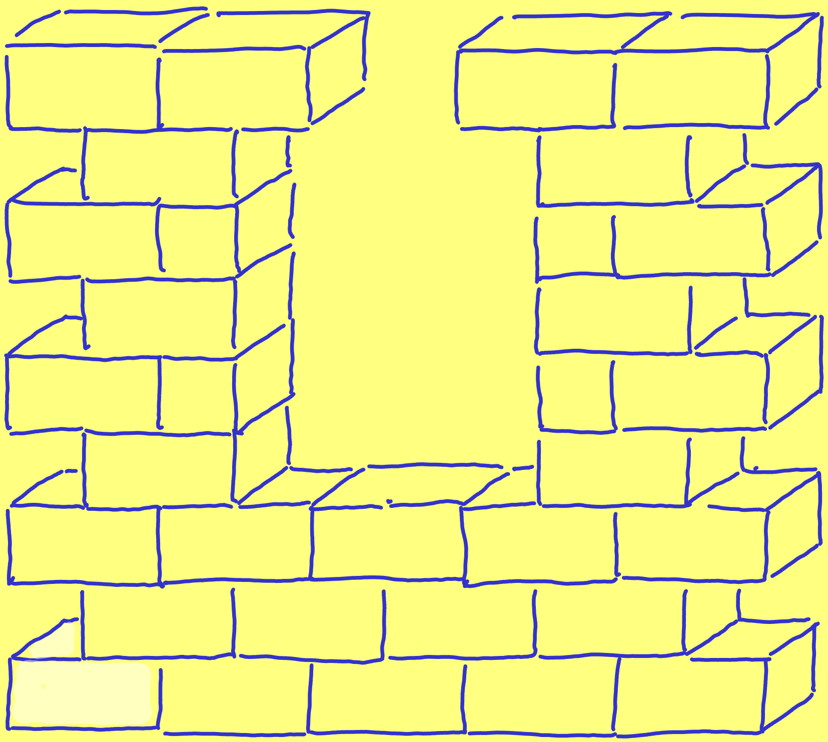
III.3 Birth and Death



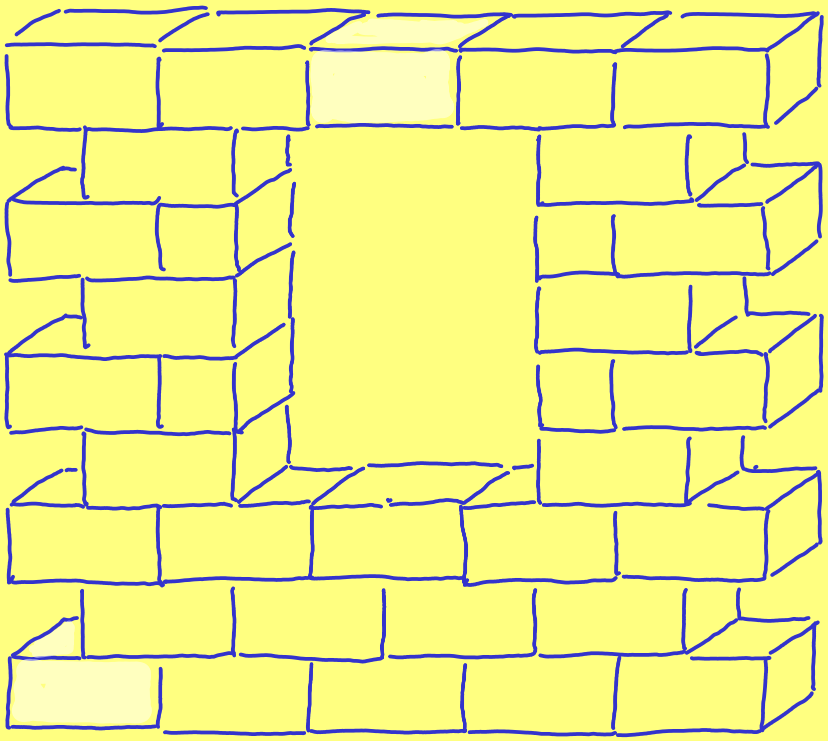
III.3 Birth and Death



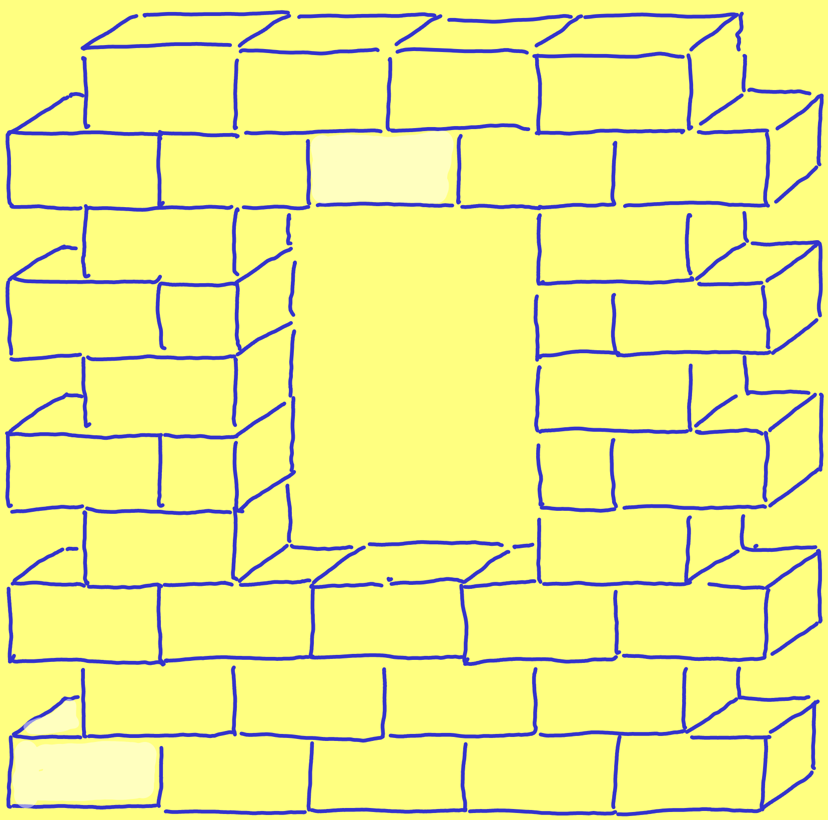
III.3 Birth and Death



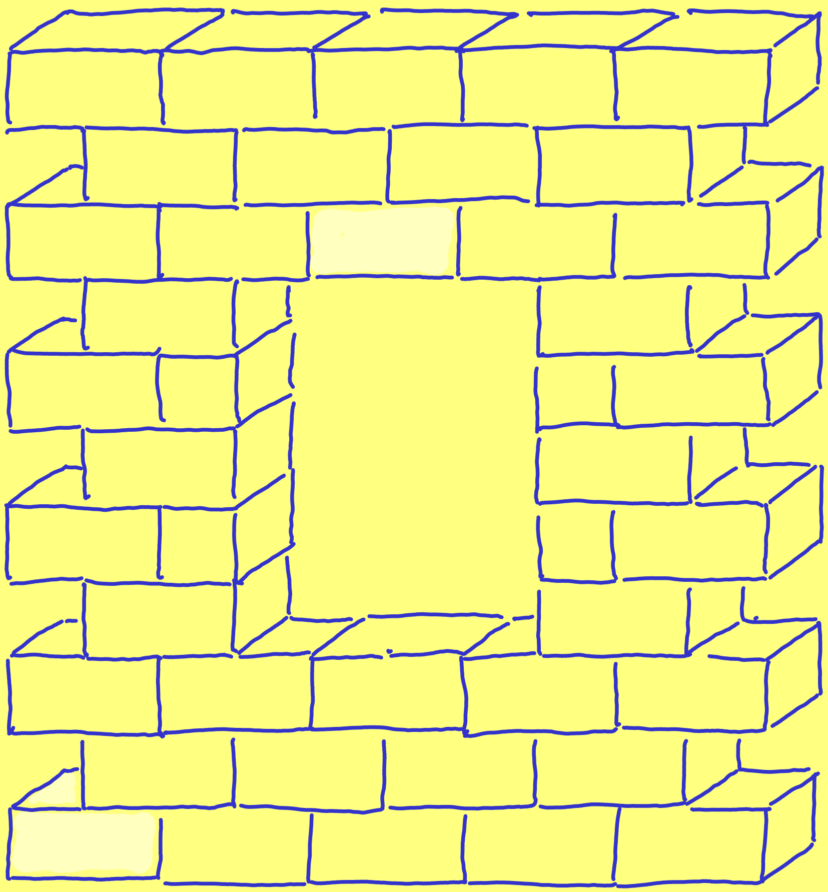
III.3 Birth and Death



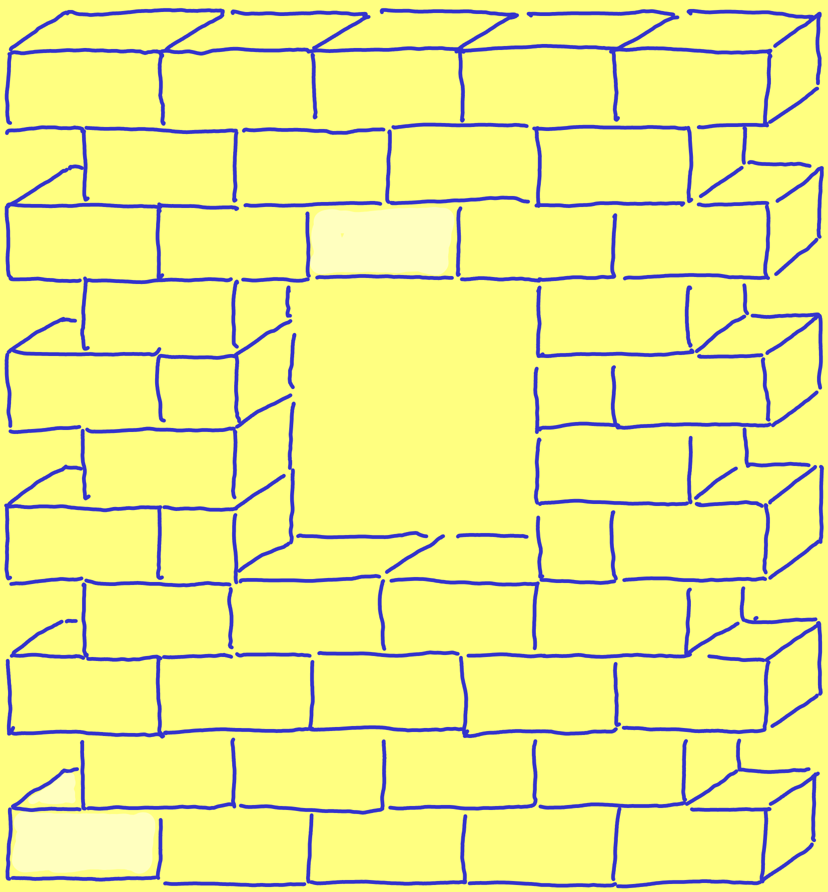
III.3 Birth and Death



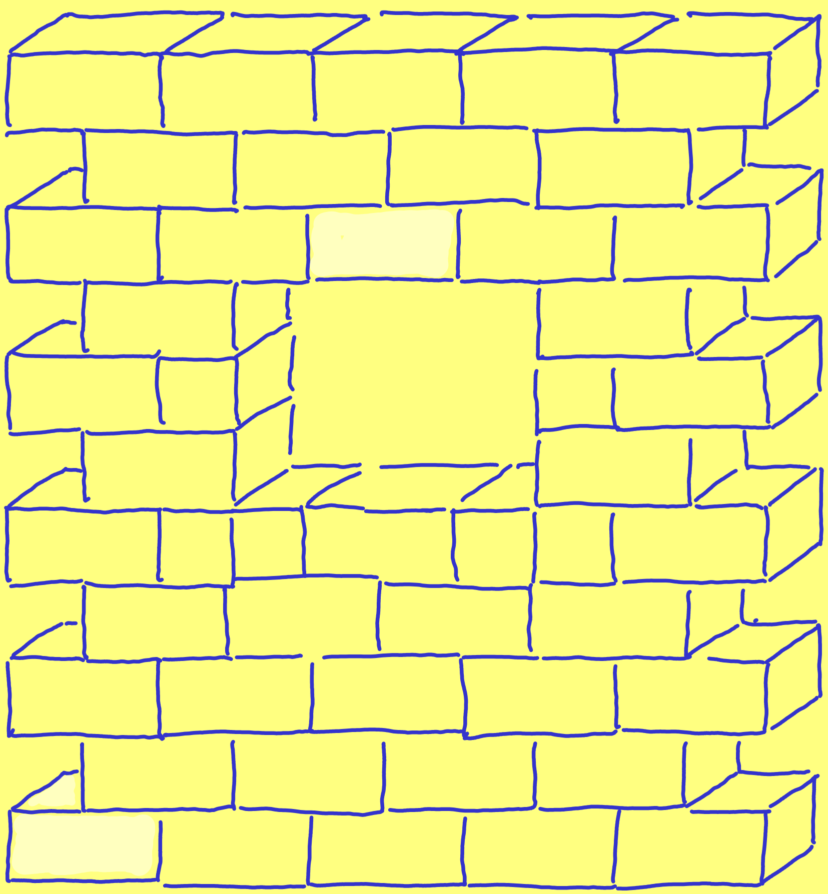
III.3 Birth and Death



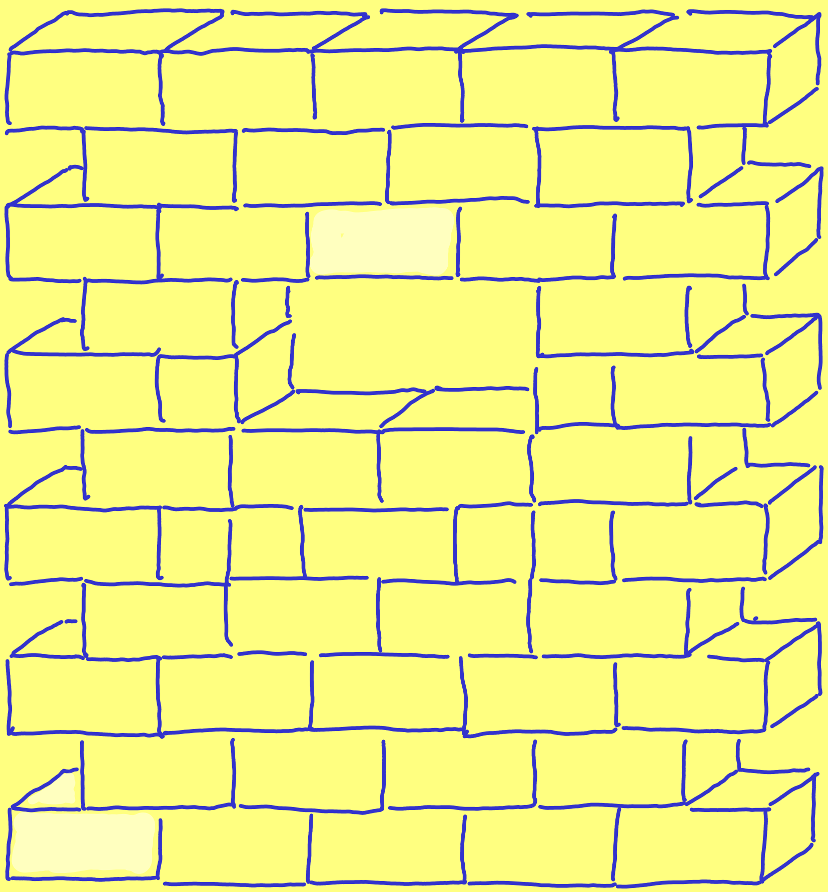
III.3 Birth and Death



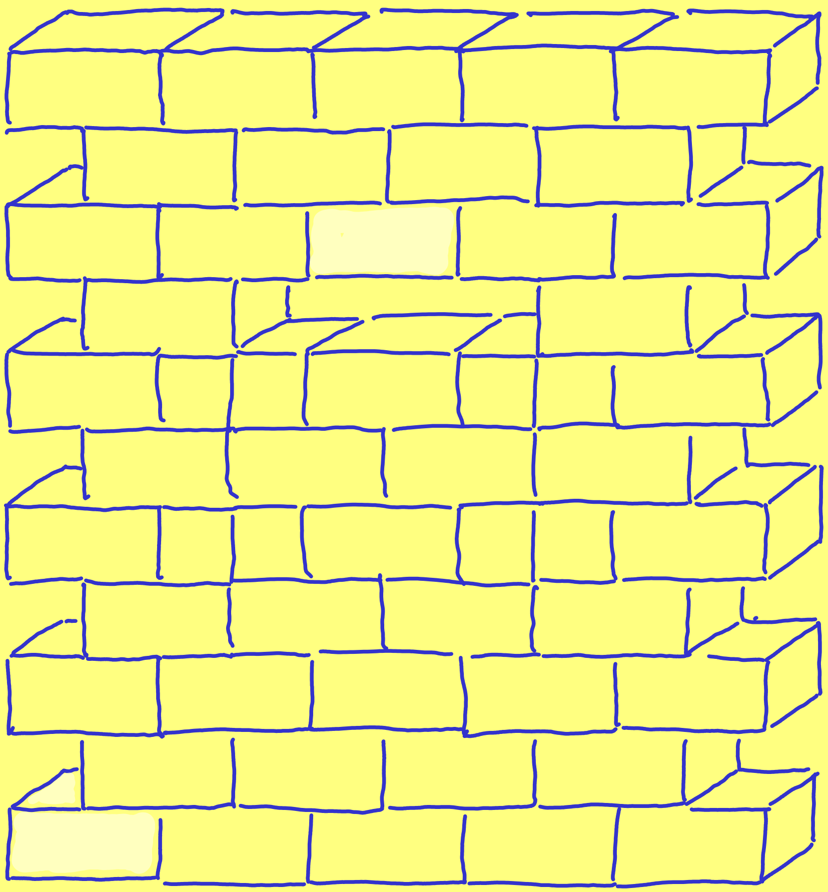
III.3 Birth and Death



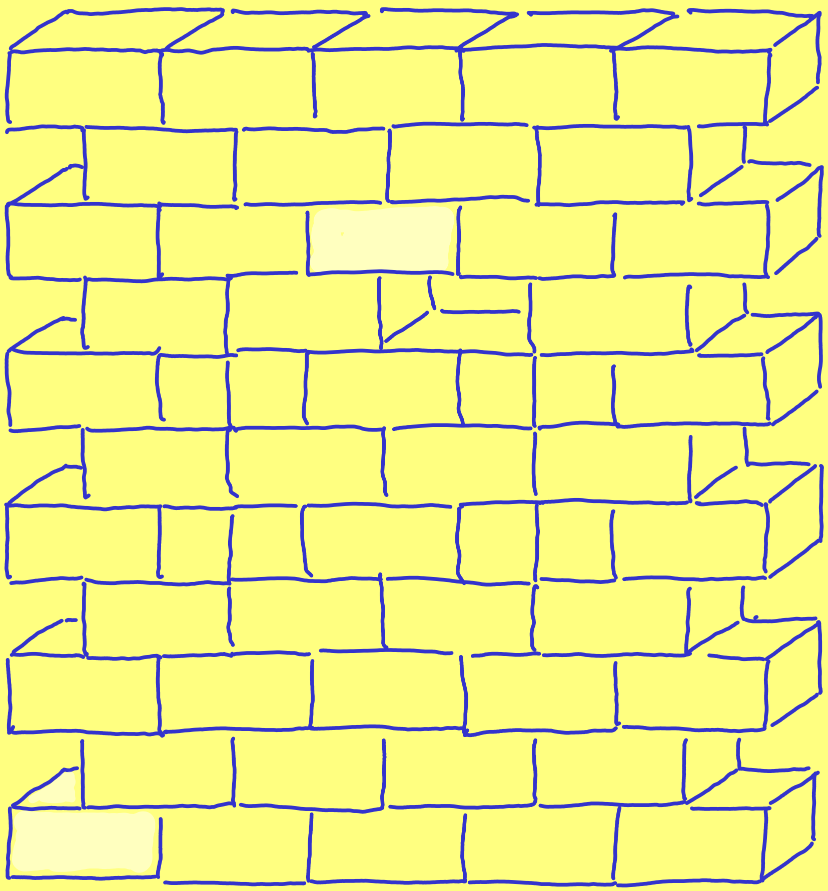
III.3 Birth and Death



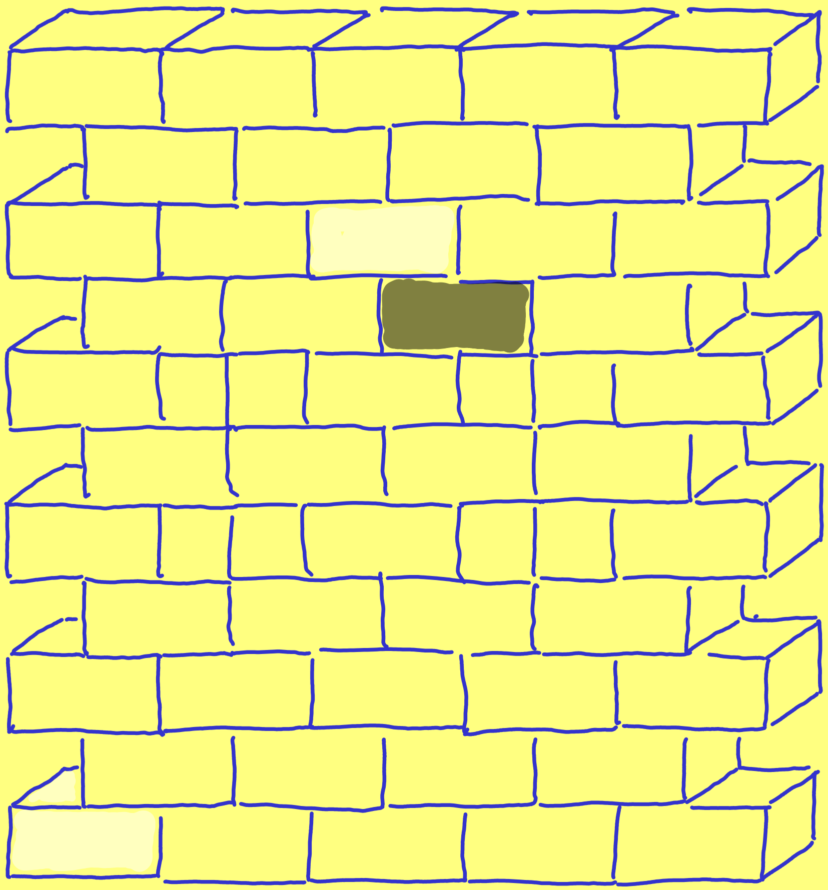
III.3 Birth and Death



III.3 Birth and Death



III.3 Birth and Death

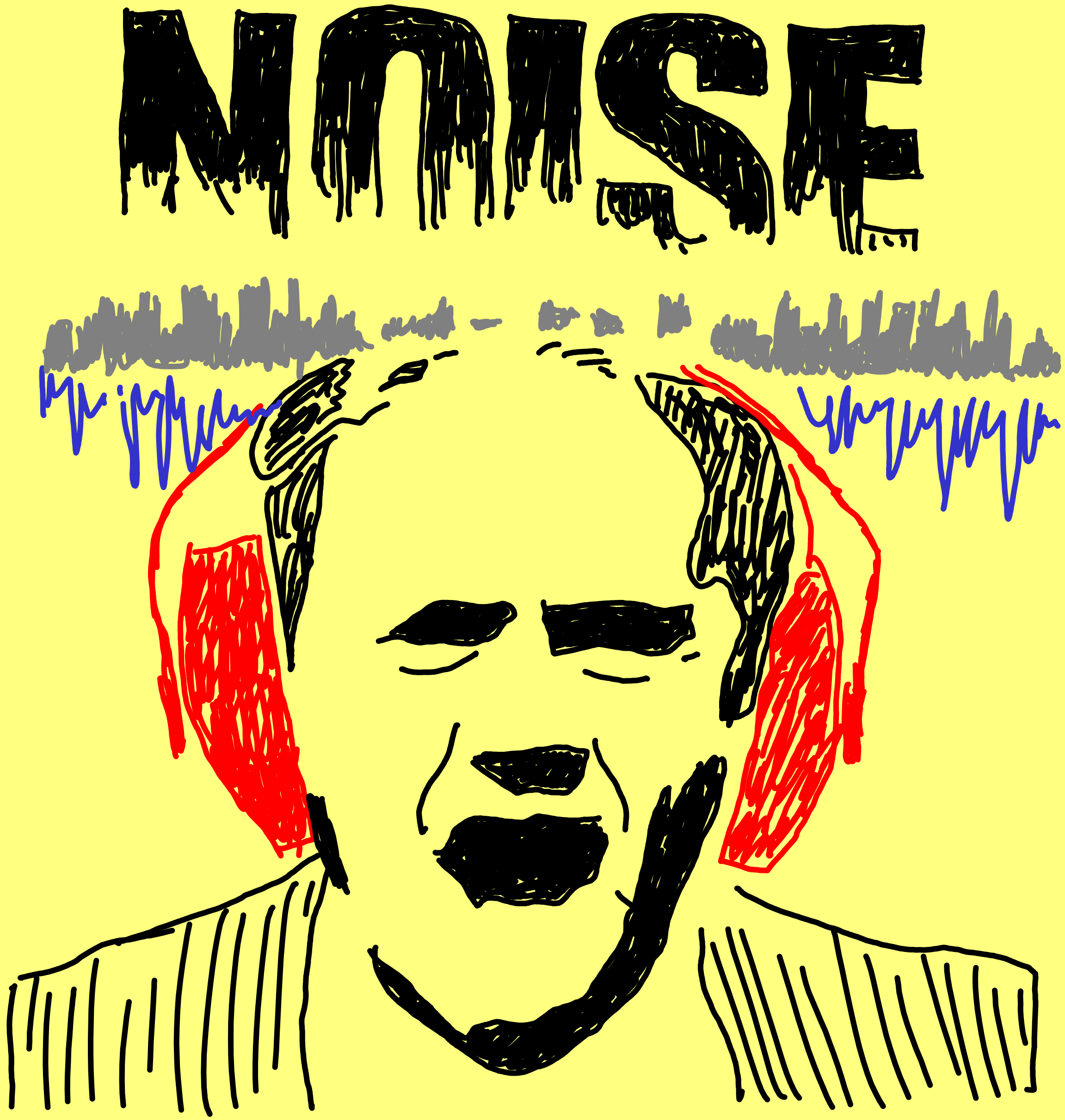


I Classification

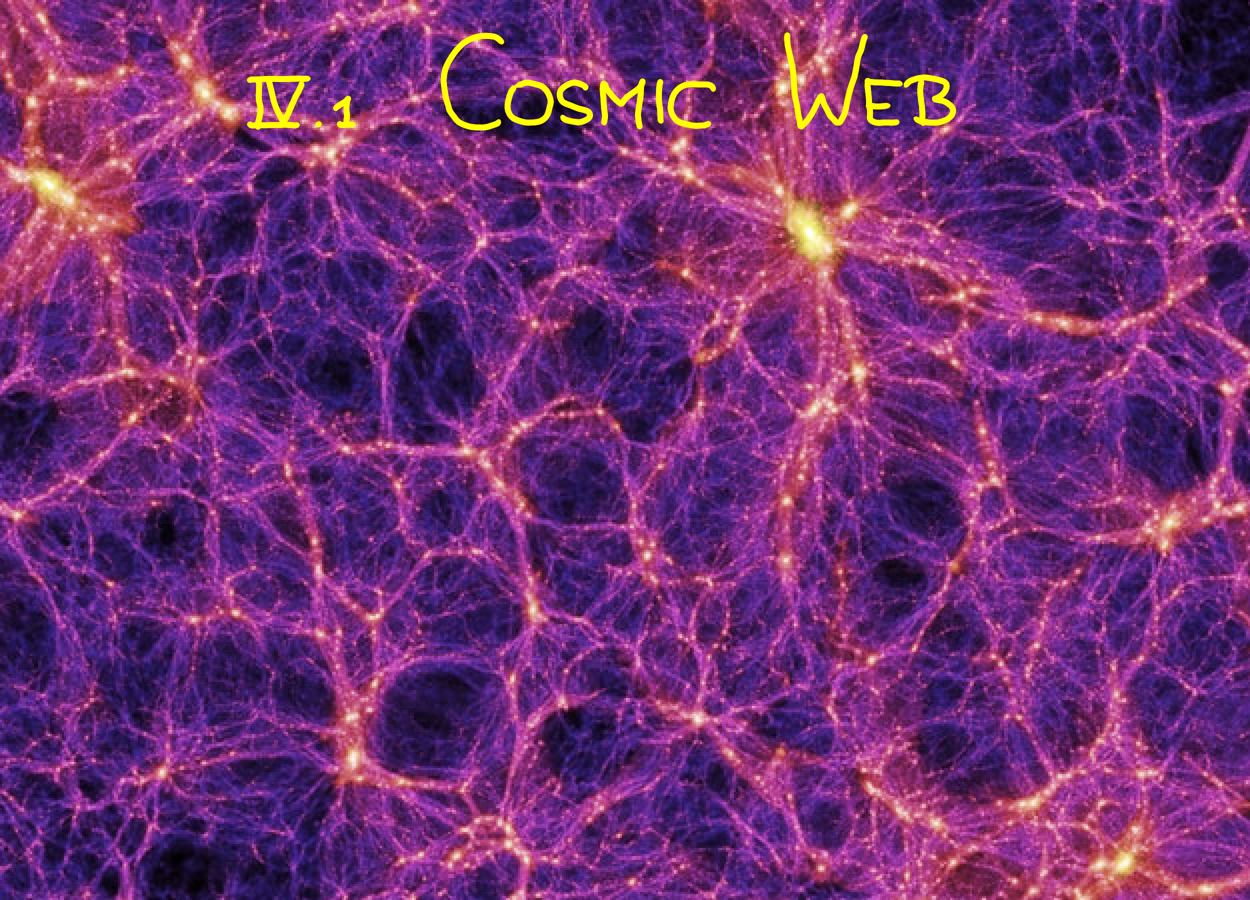
II Shape

III Scale

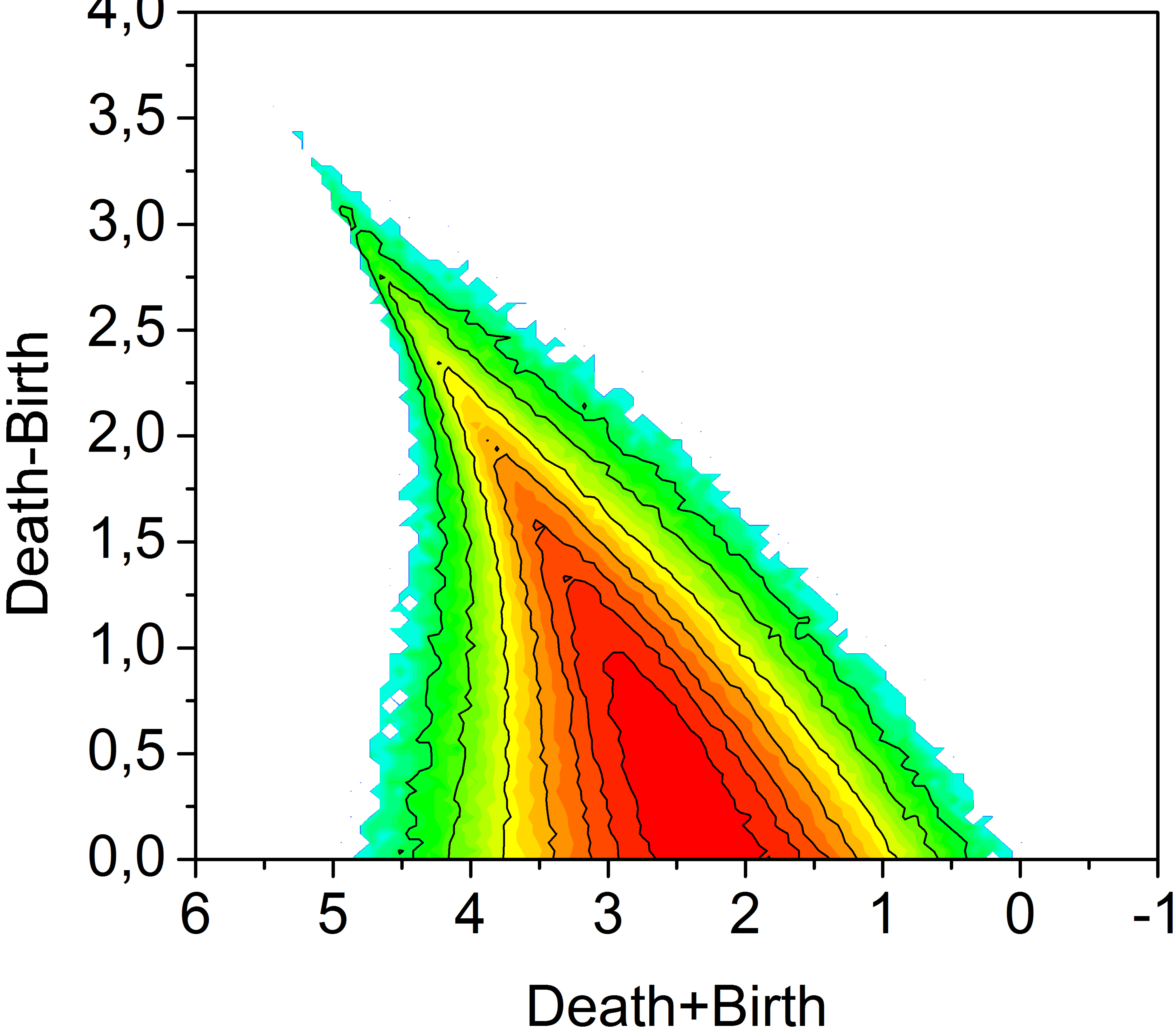
IV Applications



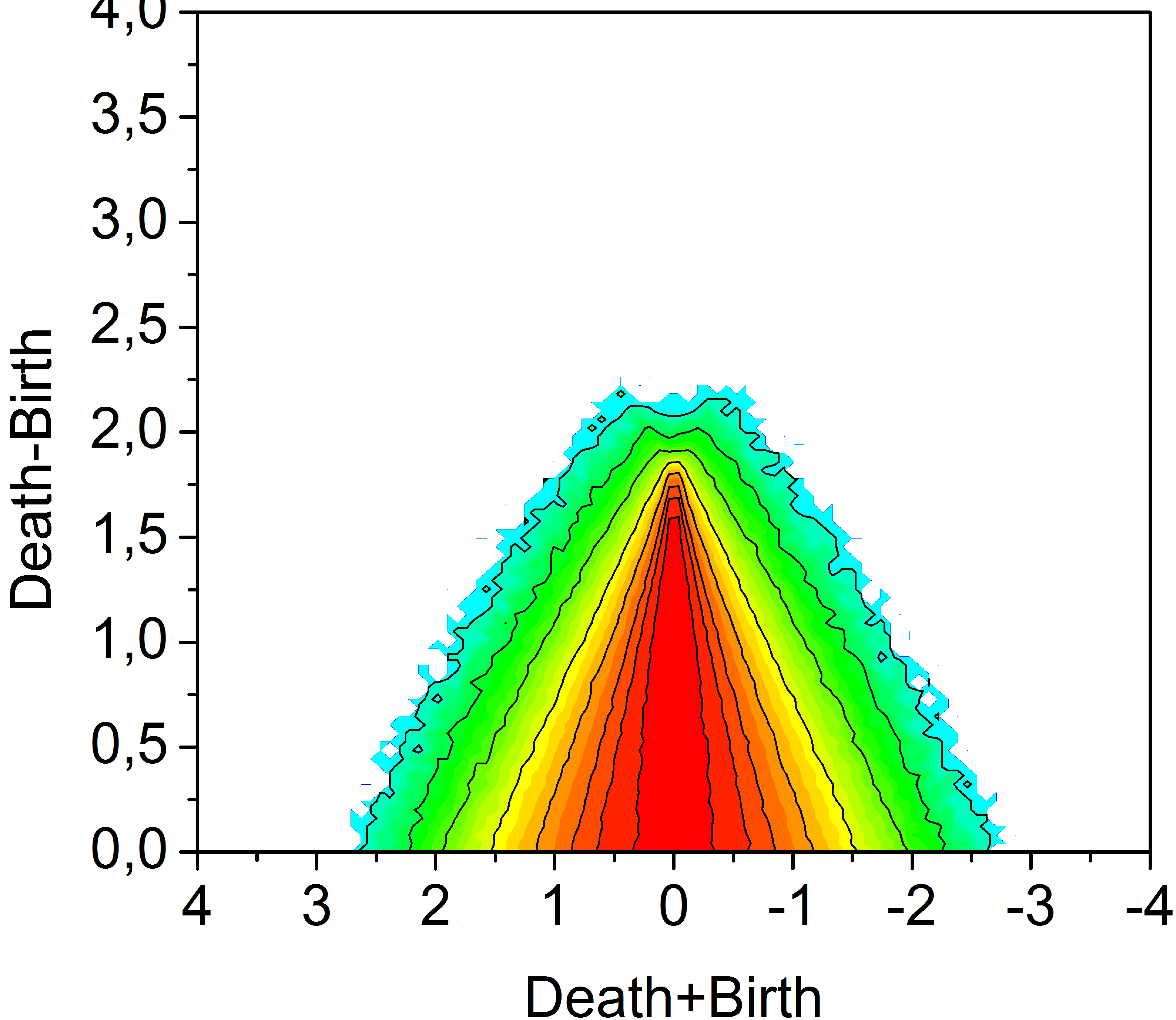
EAV: "Evil is Always and Everywhere"


IV.1 Cosmic Web

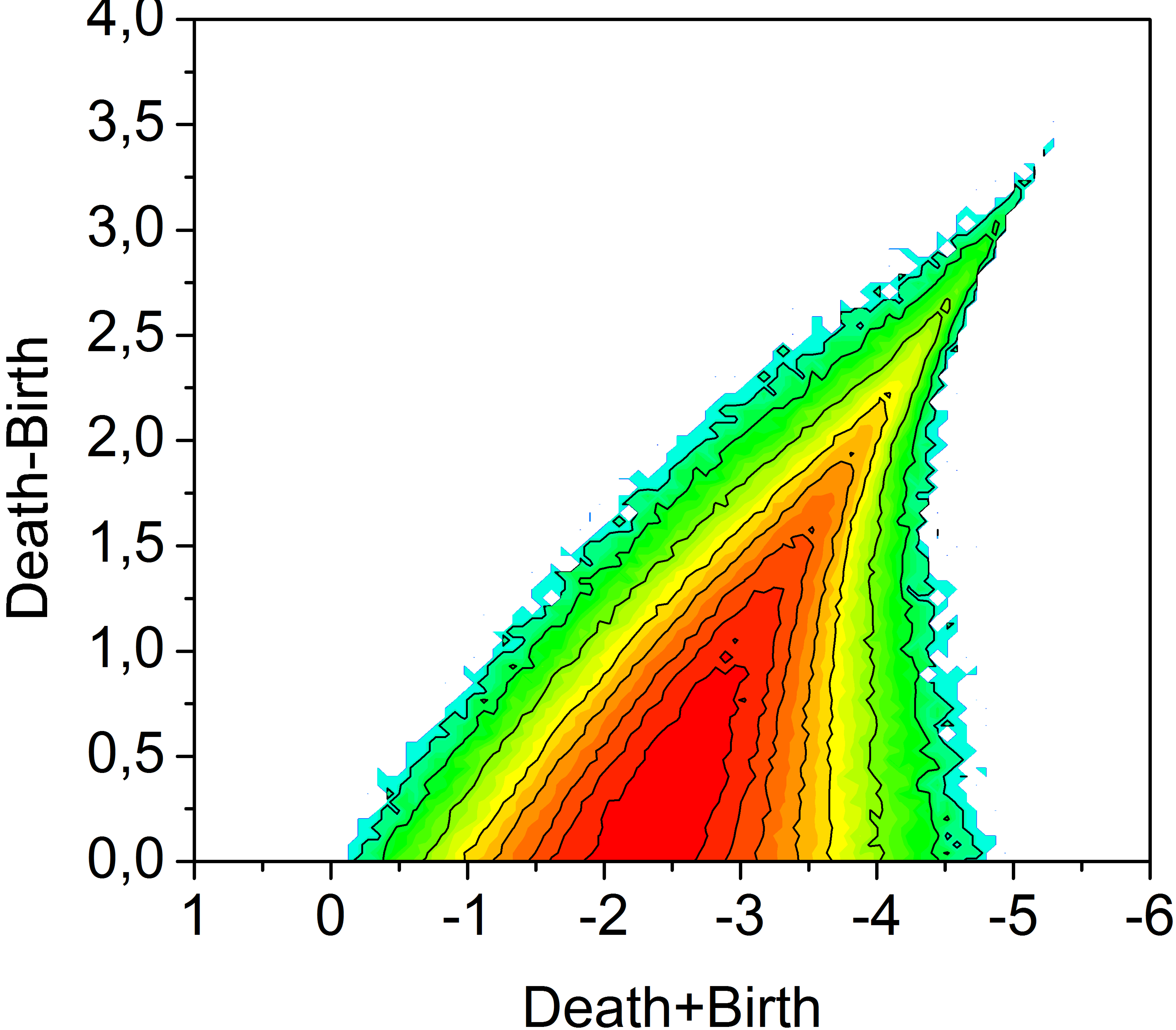
IV.2 Clusters



IV.2 FILAMENTS



IV.2 WALLS



IV.3 BIO-SPHERE

"What is the multi-scale shape of organic matter on Earth?"

THANK YOU