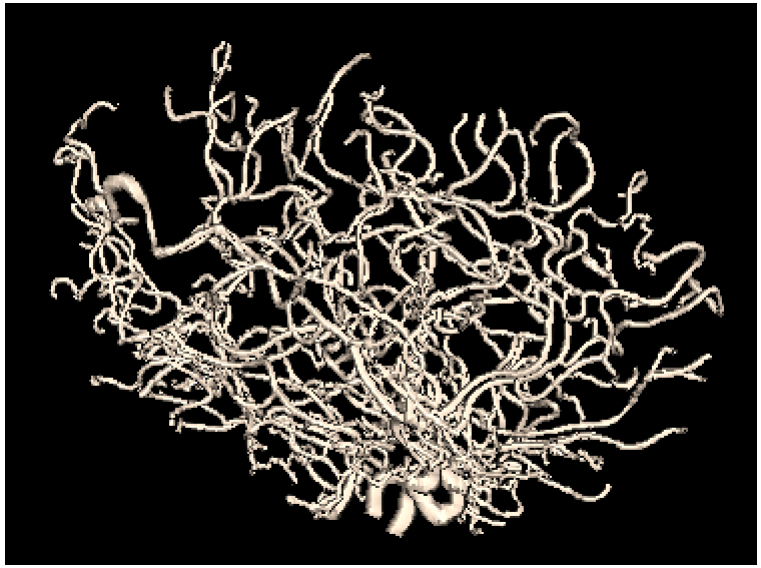
The Annals of Applied Statistics

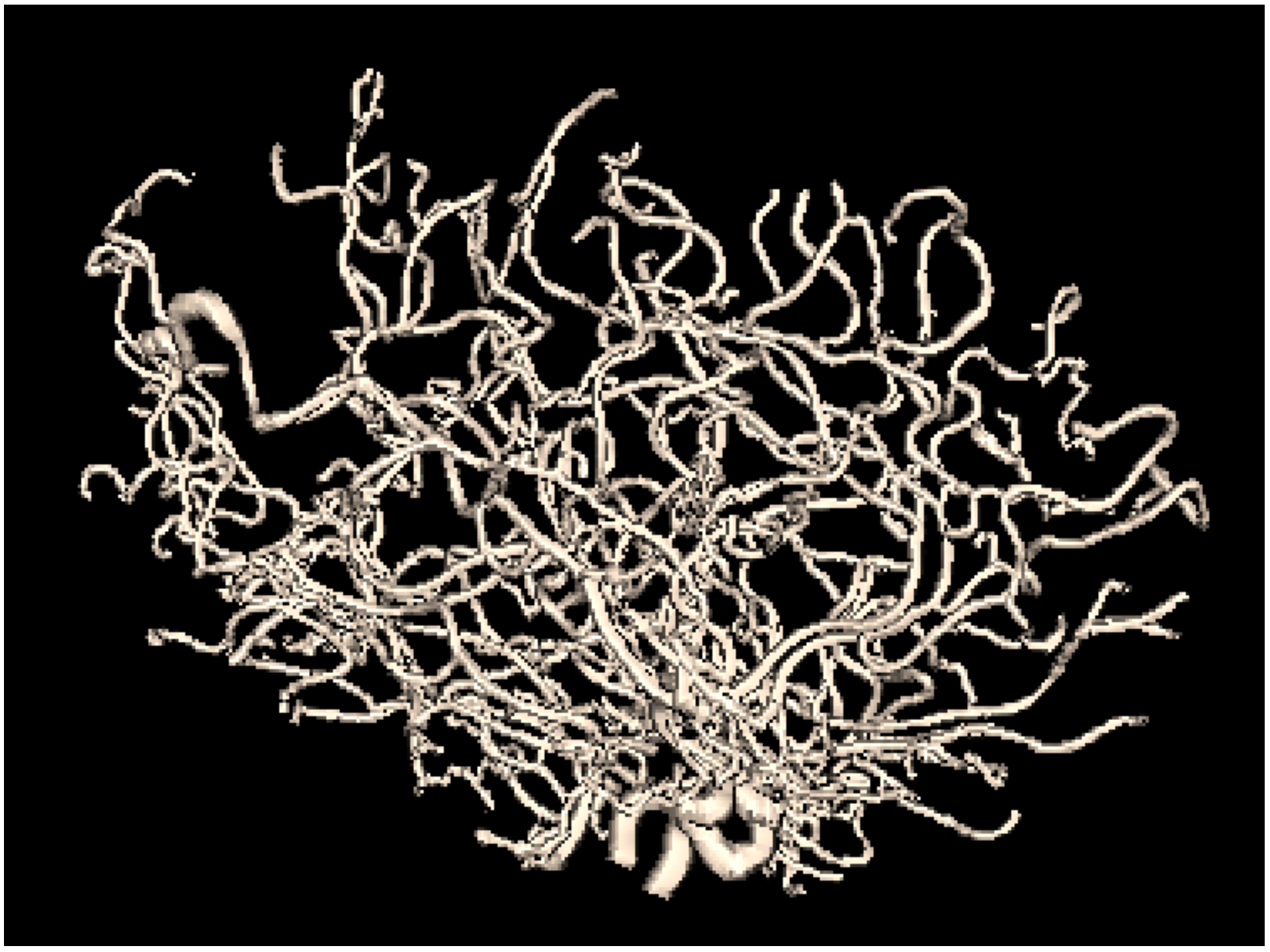
http://projecteuclid.org/euclid.aoas/1458909913

Ann. Appl. Stat.
Volume 10, Number 1 (2016), 198-218.

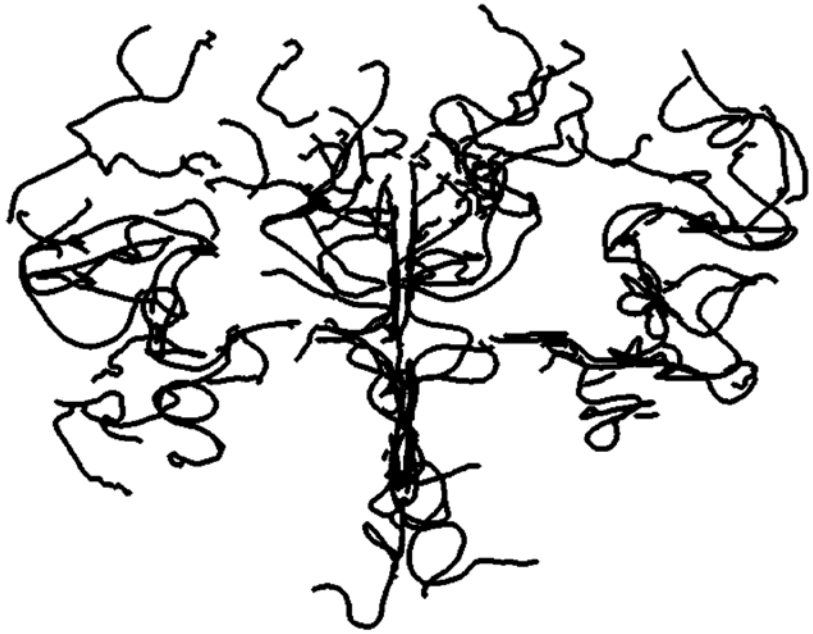
Persistent homology analysis of brain artery trees

Paul Bendich, J. S. Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer





24 years old



68 years old

Each brain corresponds to an (anonymous) human subject. The information retained on each subject (age, gender, handedness) is stored in res/subjectdata.mat

In subjectdata.mat, **handedeness** is encoded as follows:

Right handed: 1
Left handed: 2
Ambidextrous: 3
Left/Ambidextrous: 4
Right/Ambridextrous: 5

Gender is encoded as follows:

Female: 1
Male: 2
Male -> Female: 3
Female -> Male: 4

The brain tree structure of subject N is stored in decorated_tree.aca.CaseN.mat, here are further details.

case-num (obvious)

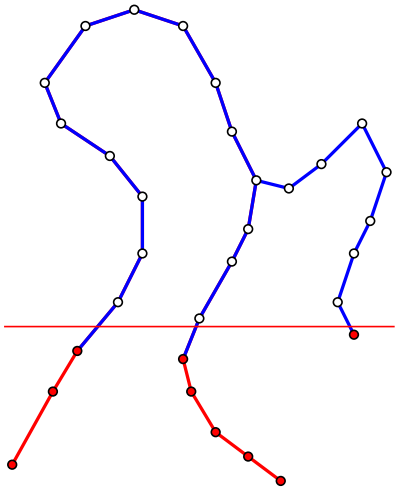
V. This is a 1 X (number of vertices) struct; there are usually **around 120K vertices per tree.**
The location field in V{k,1} stores the **3-D coordinates** of the kth vertex. For details on the rest of the fields in V, please email sskwerer@email.unc.edu; note that none of these fields are relevant at the moment to this project.

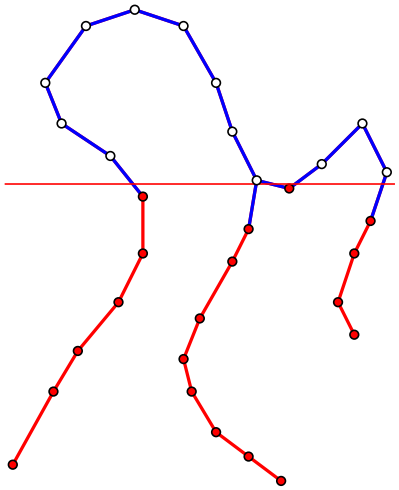
Branch. This is a 1 X (number of branches) struct; there are usually **around 50 branch vertices per tree**. Details are as above for V.

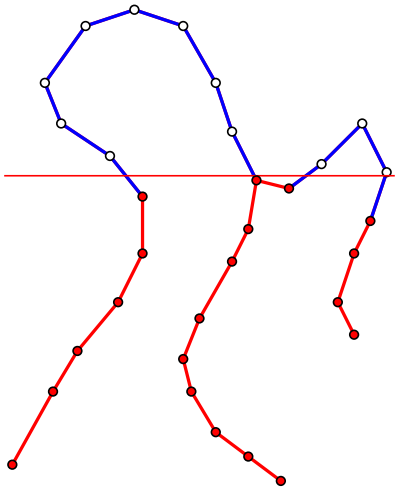
E. This is a (number of edges) X 2 array. This is the edge list of the tree, where the prescence of (k,l) in the array indicates an edge between vertices k and l. Note **that most vertices are degree two, some are higher.**

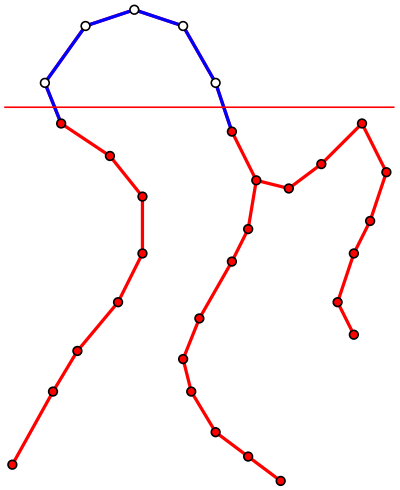
In BasicTreeInfo.mat, Contains a structure with fields:

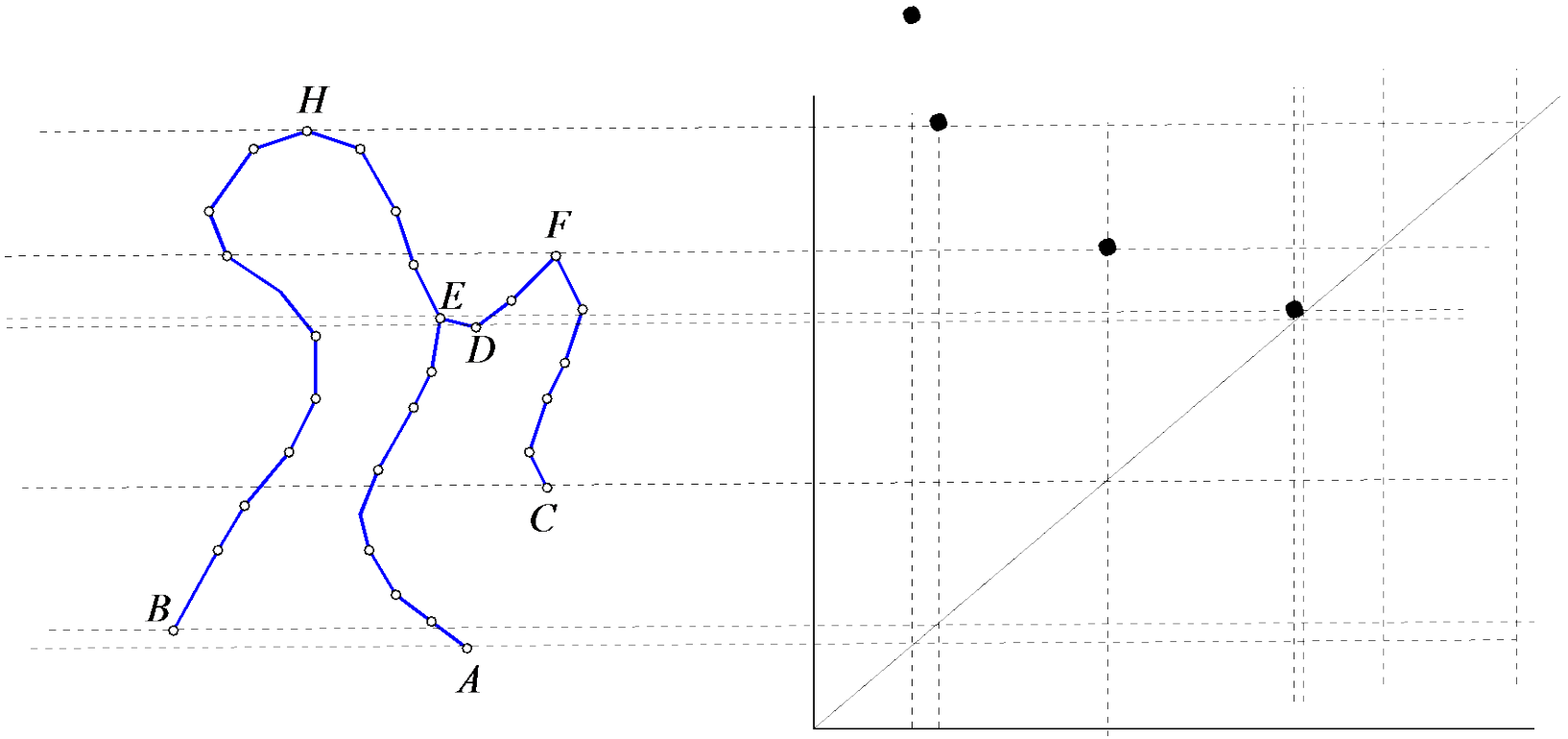
subjectID: obvious
totalLength: the total vessel length in the brain artery tree

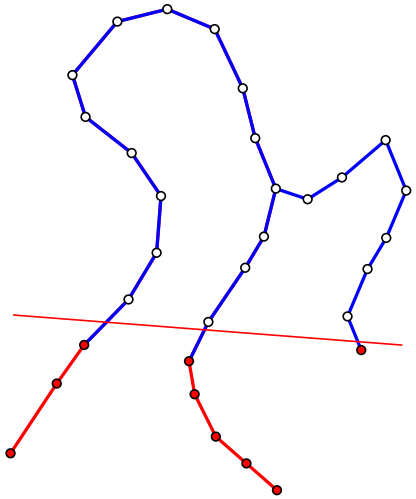


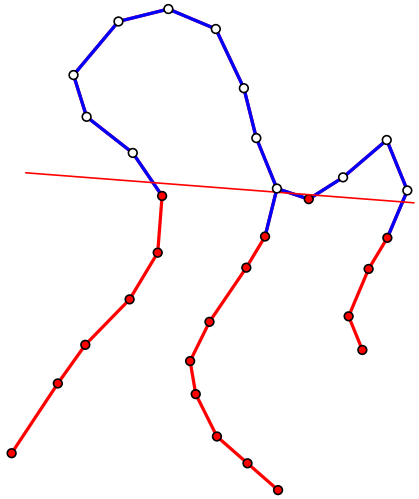


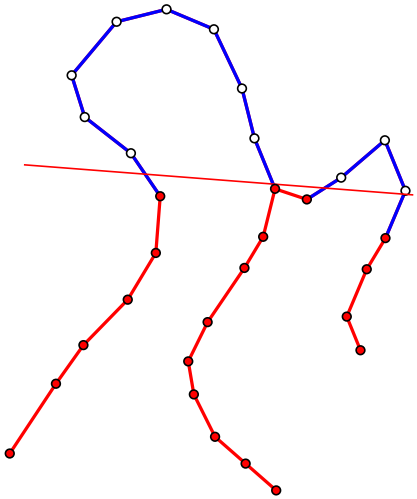


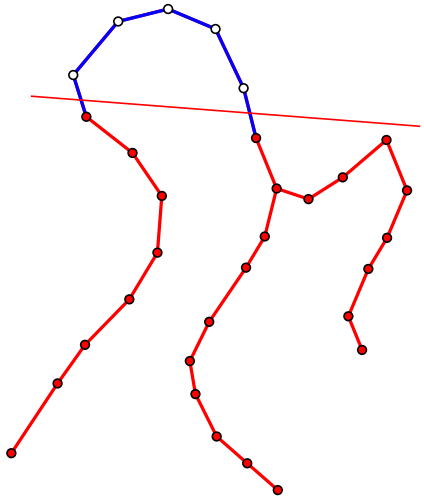


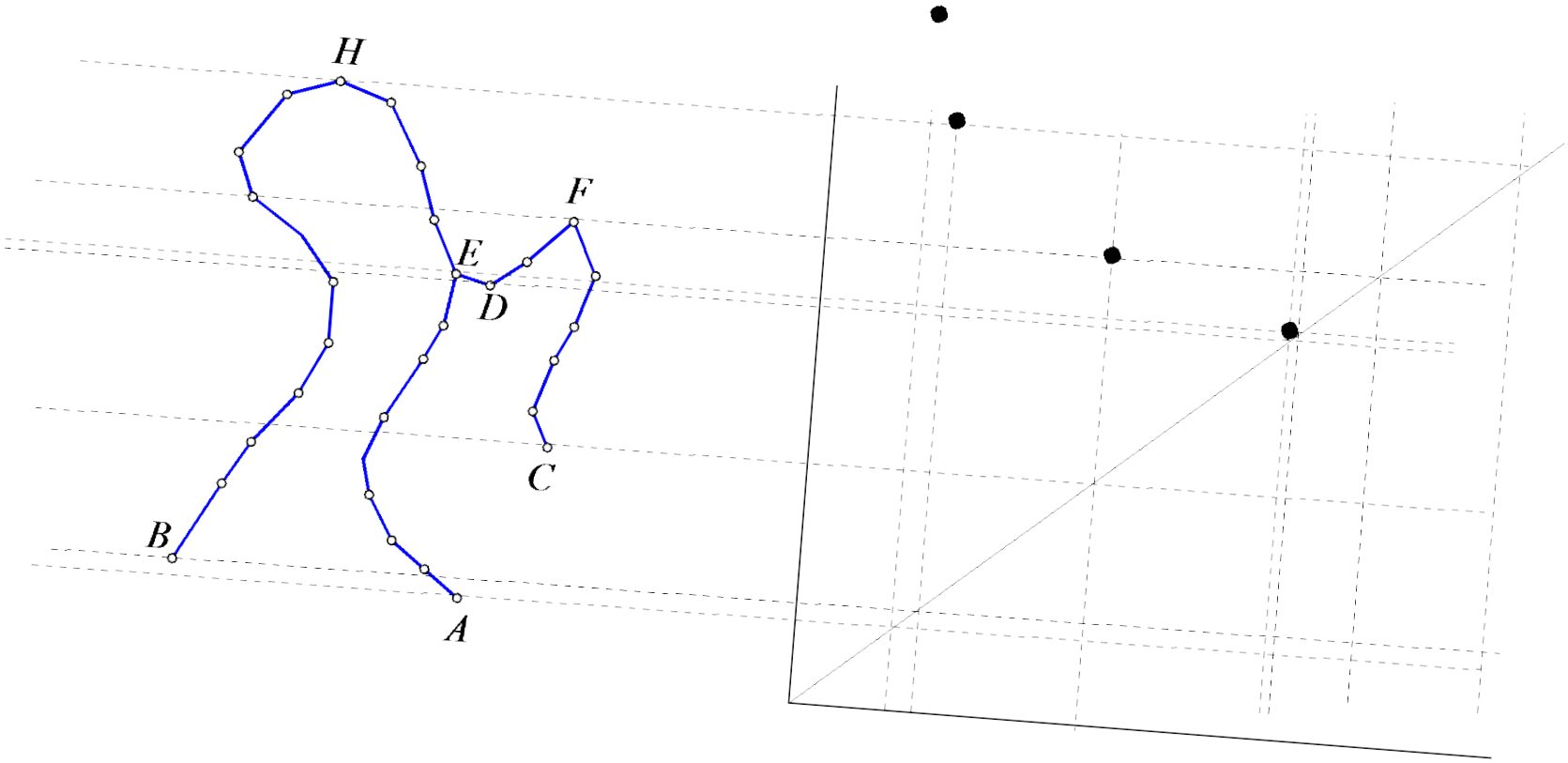

H
F
E
D
C
B
A

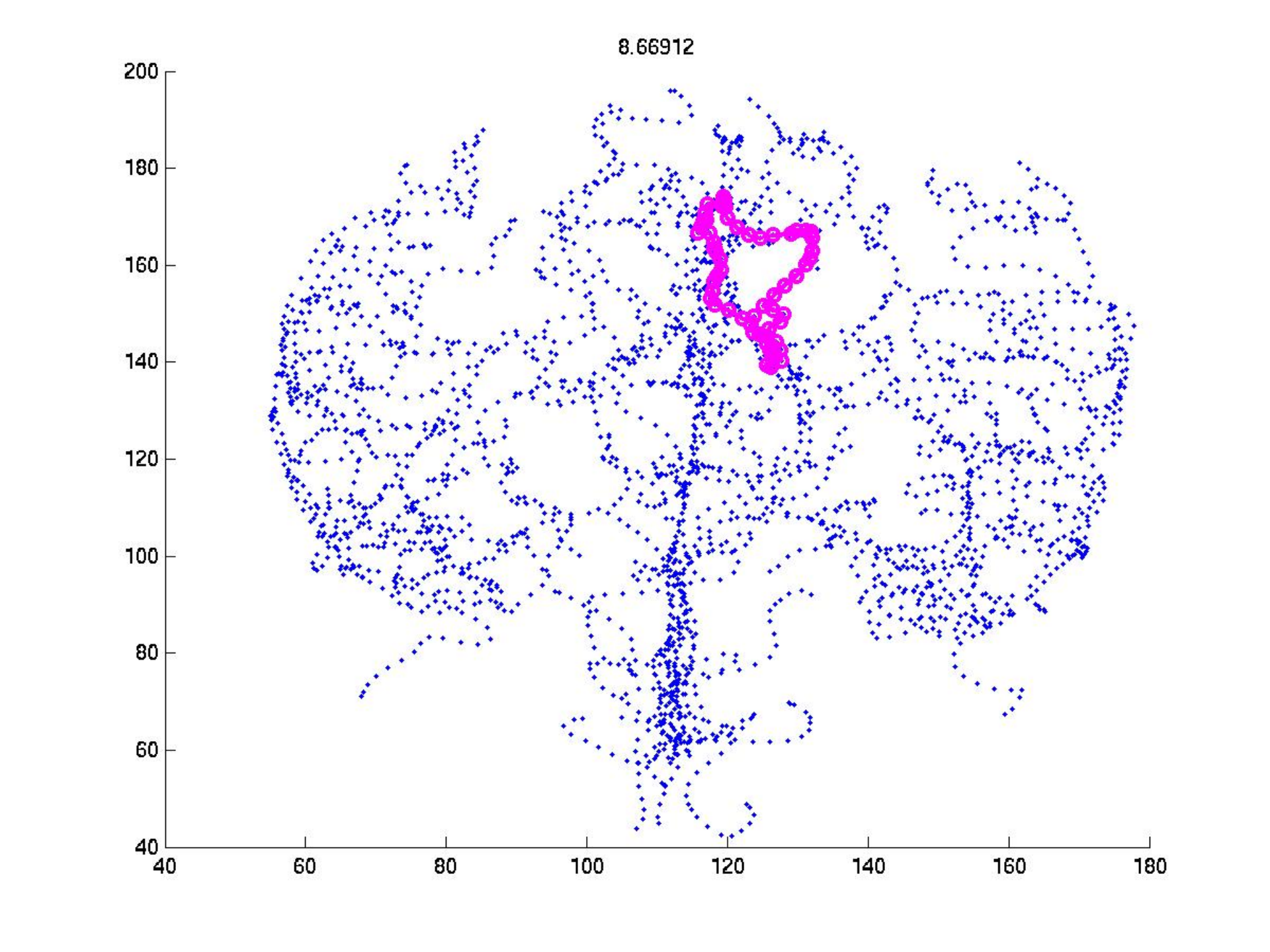


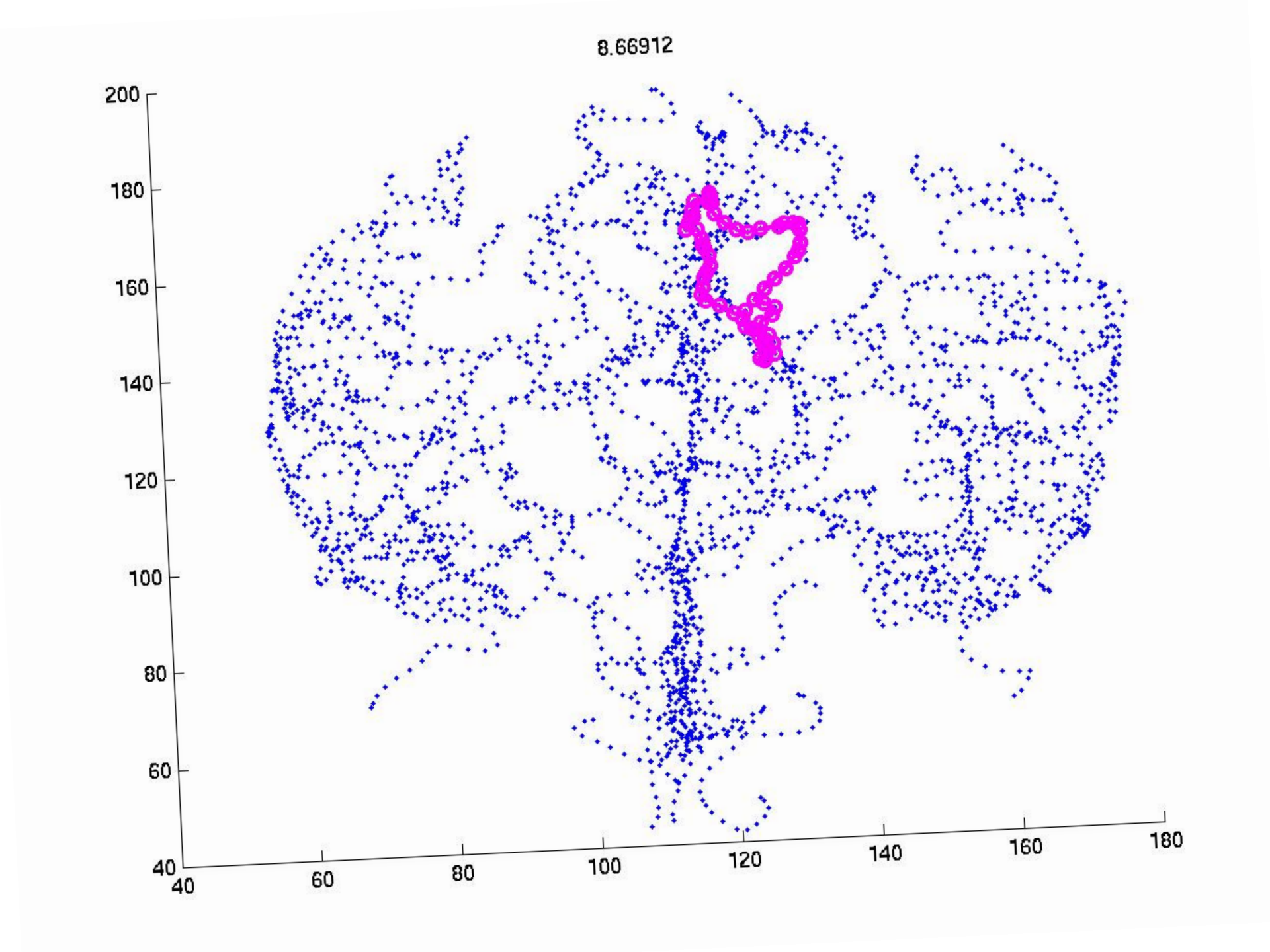


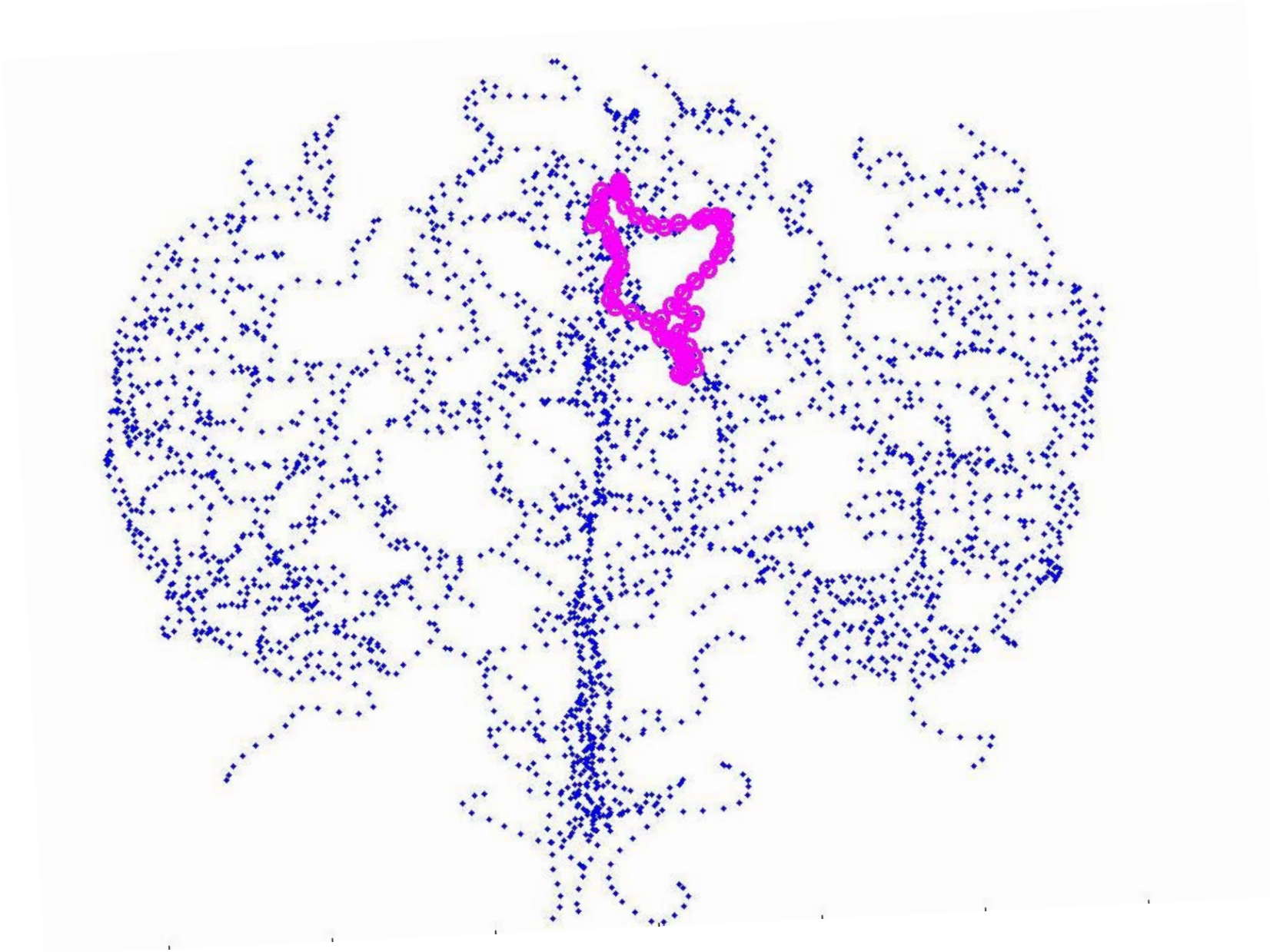


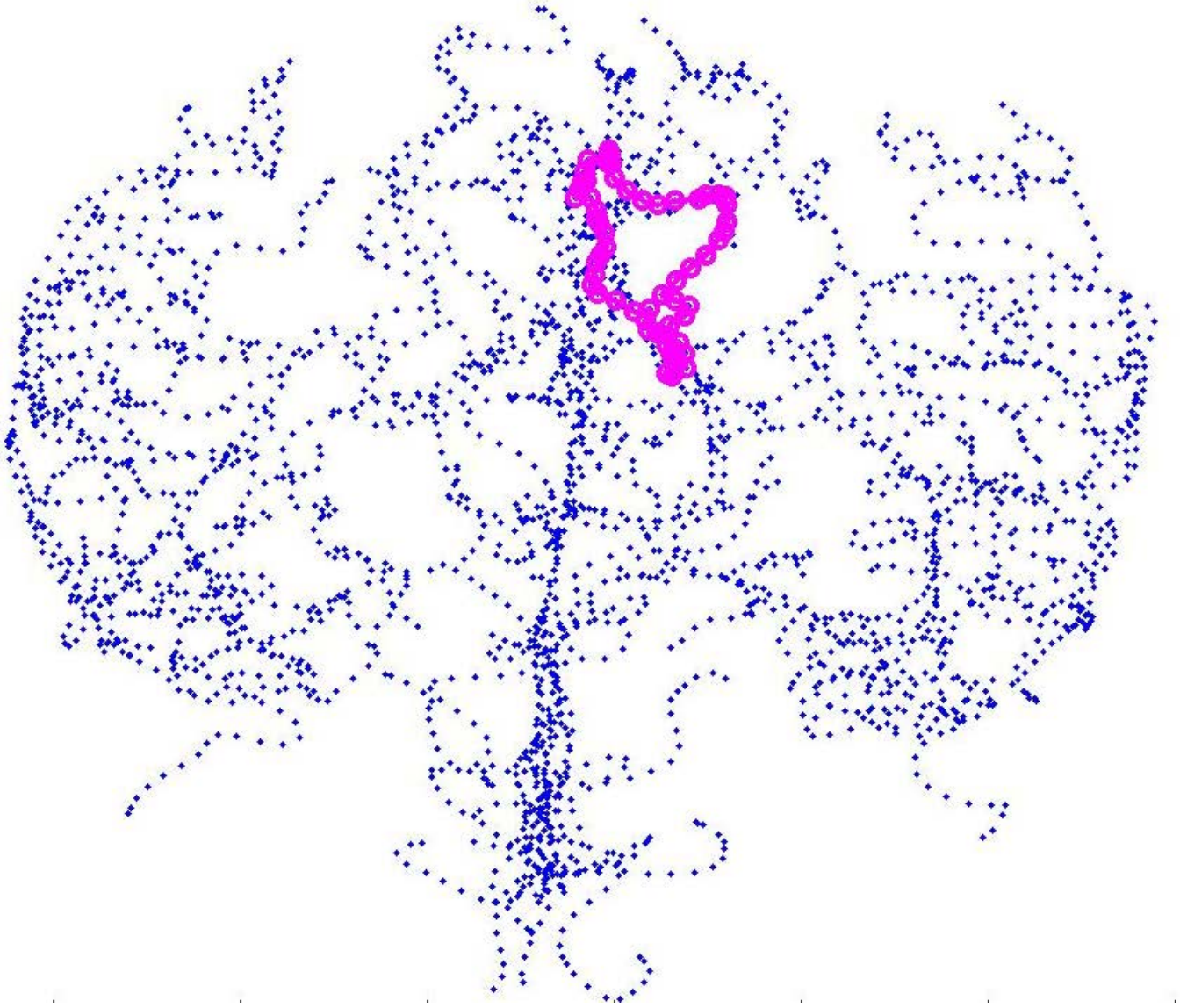


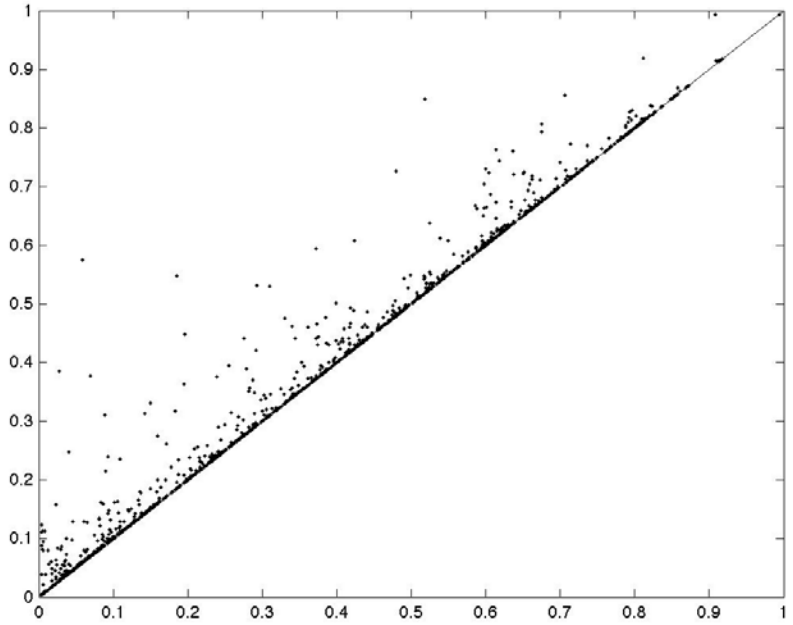

H
F
E
D
C
B
A

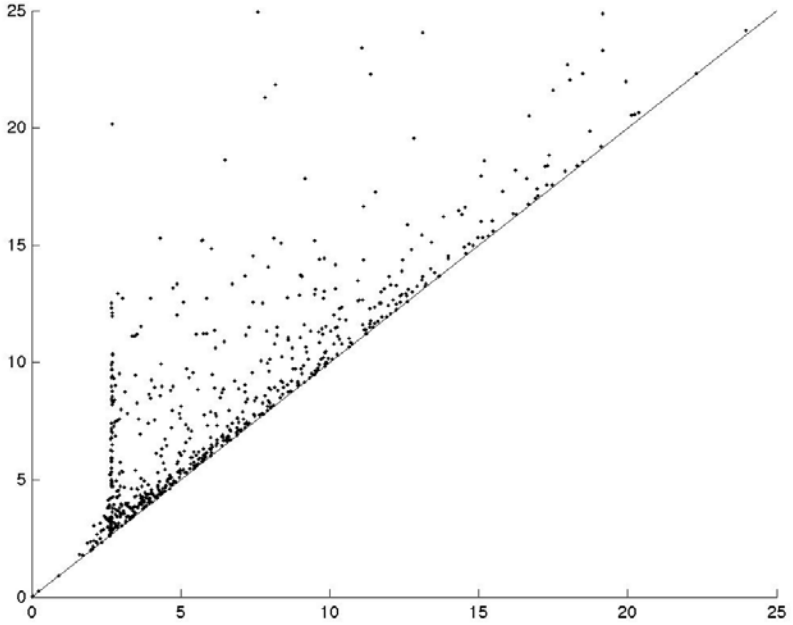

8.66912
200
180
160
140
120
100
80
60
40
40
60
80
100
120
140
160
180

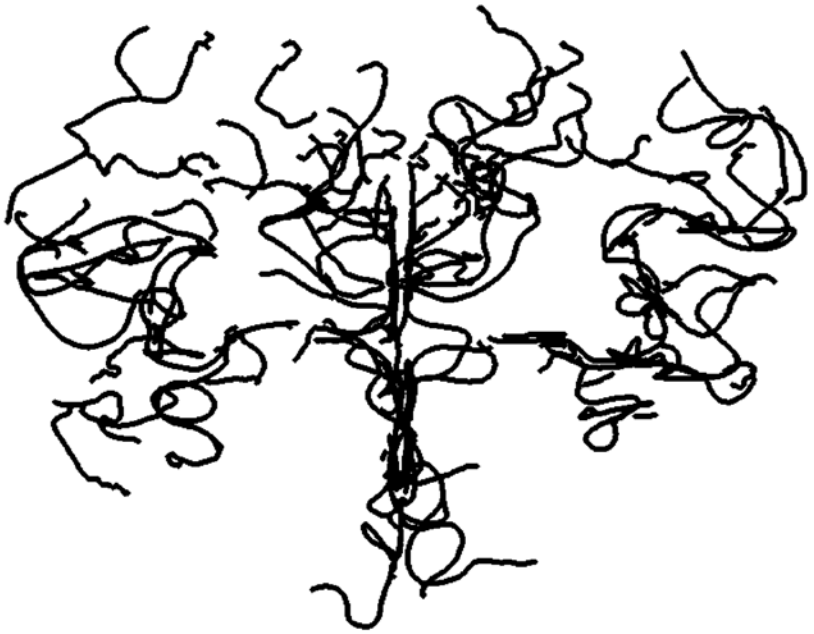

8.66912
200
180
160
140
120
100
80
60
40
40
60
80
100
120
140
160
180

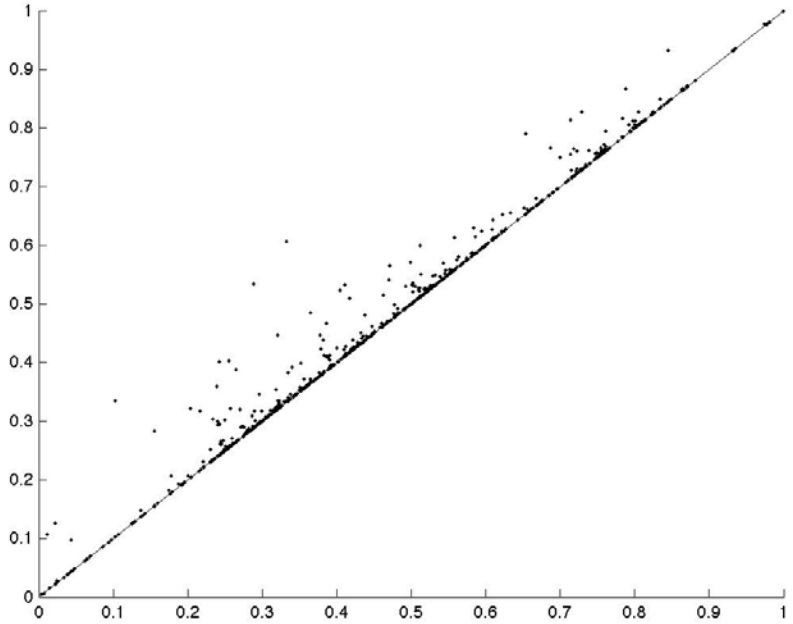


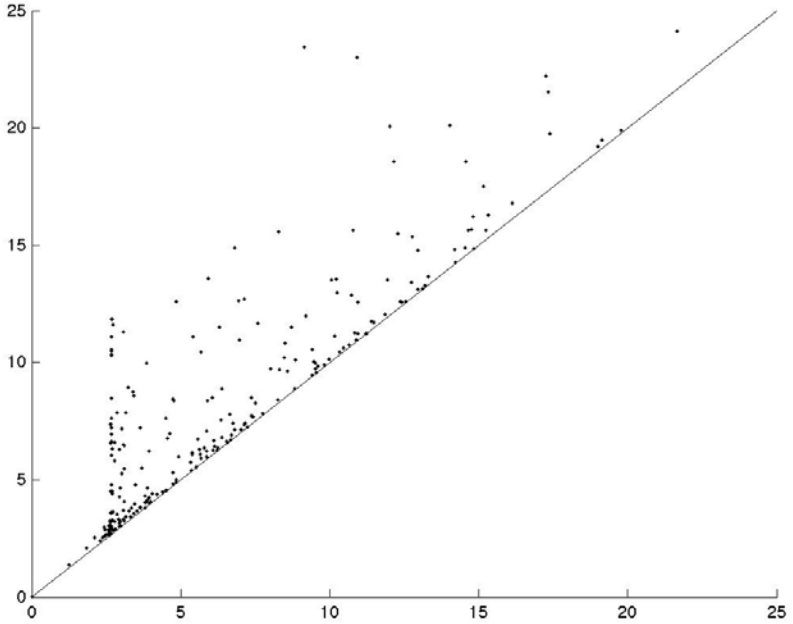


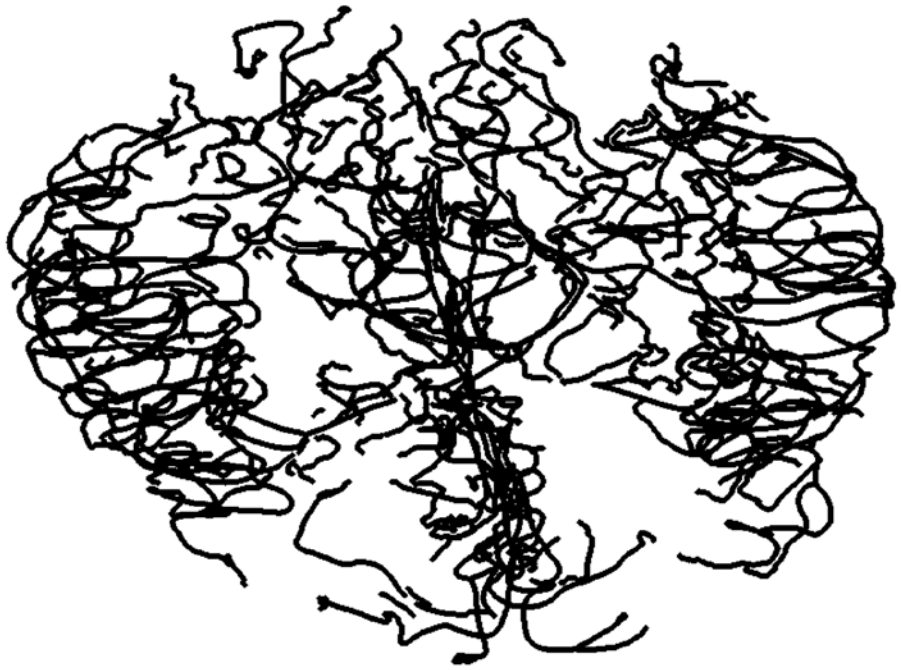

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

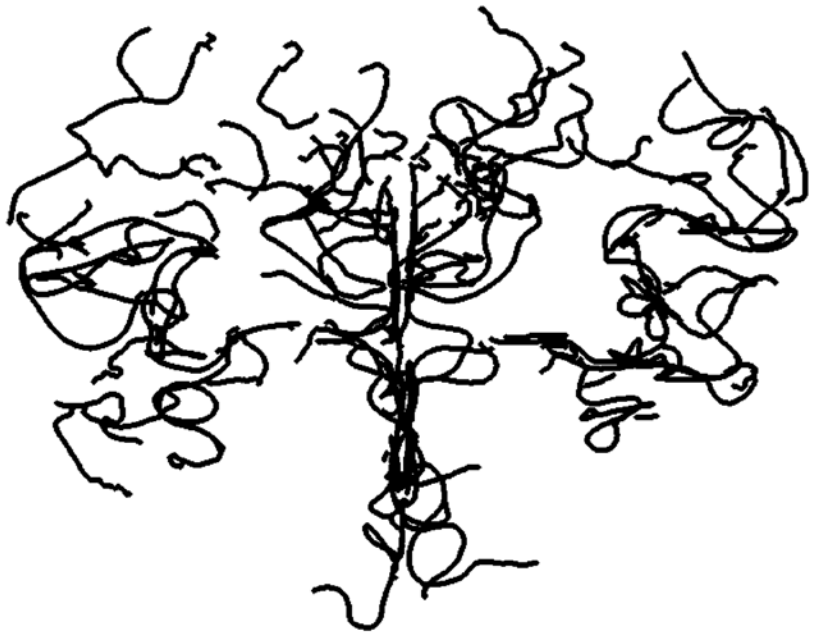

0
5
10
15
20
25

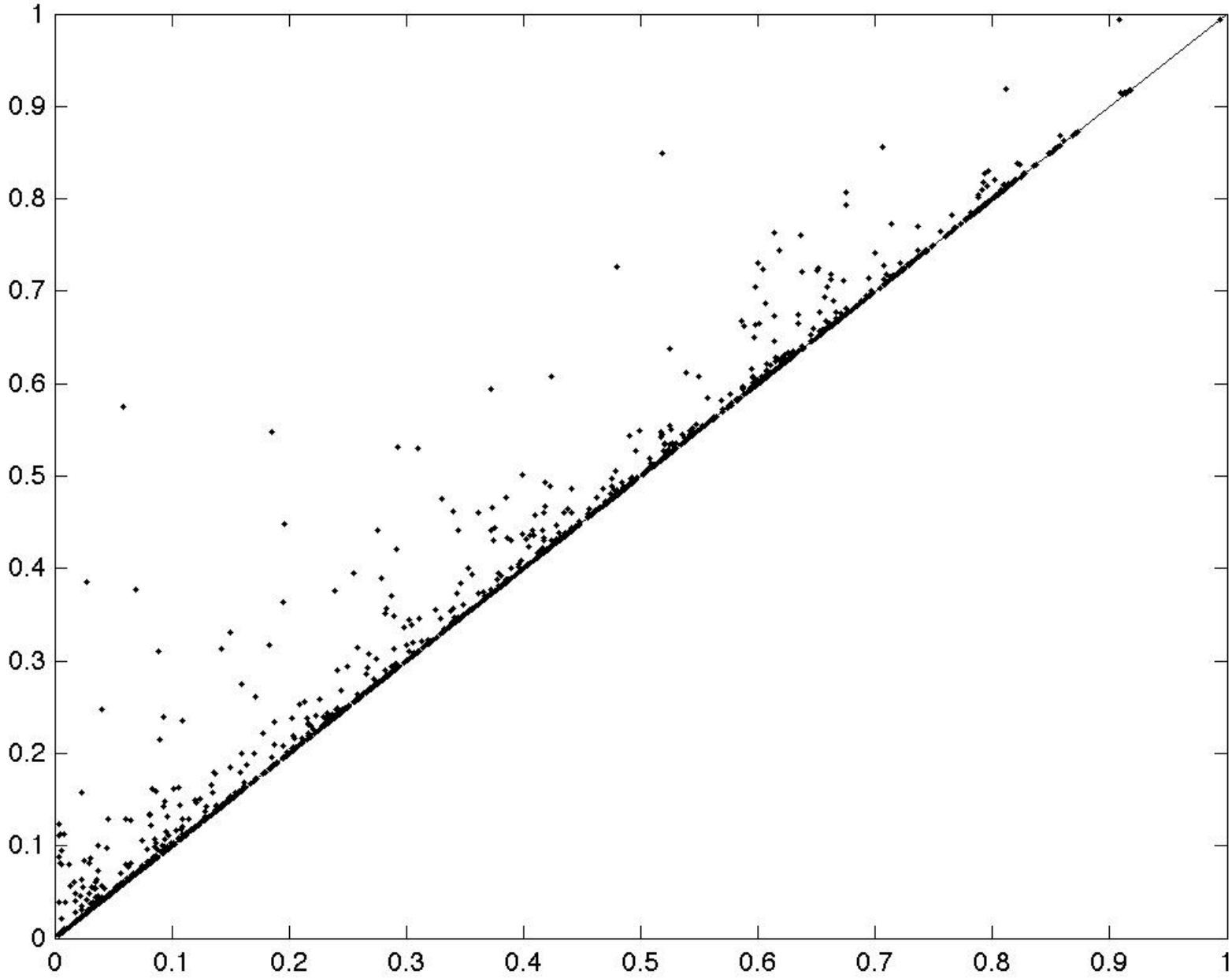


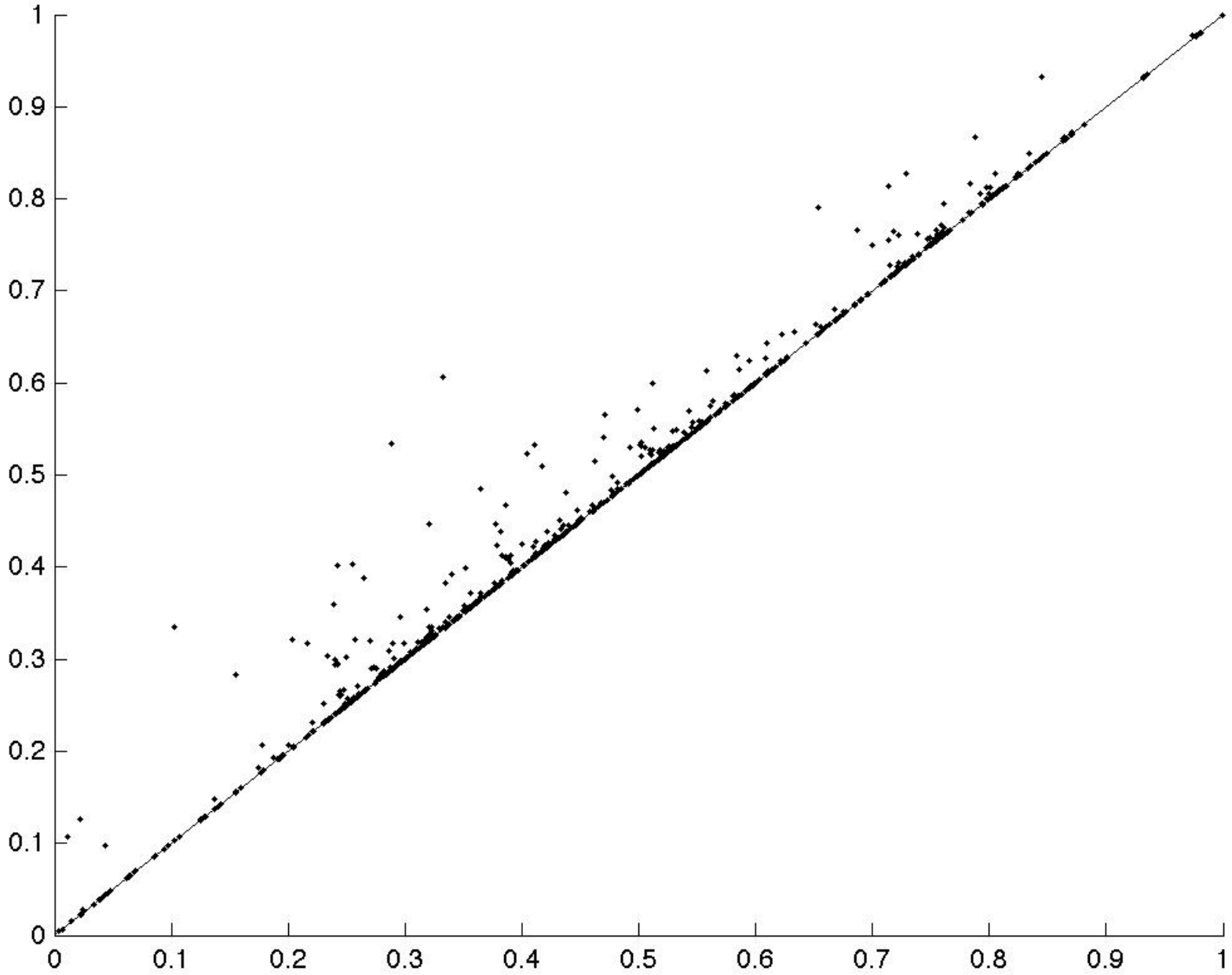

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

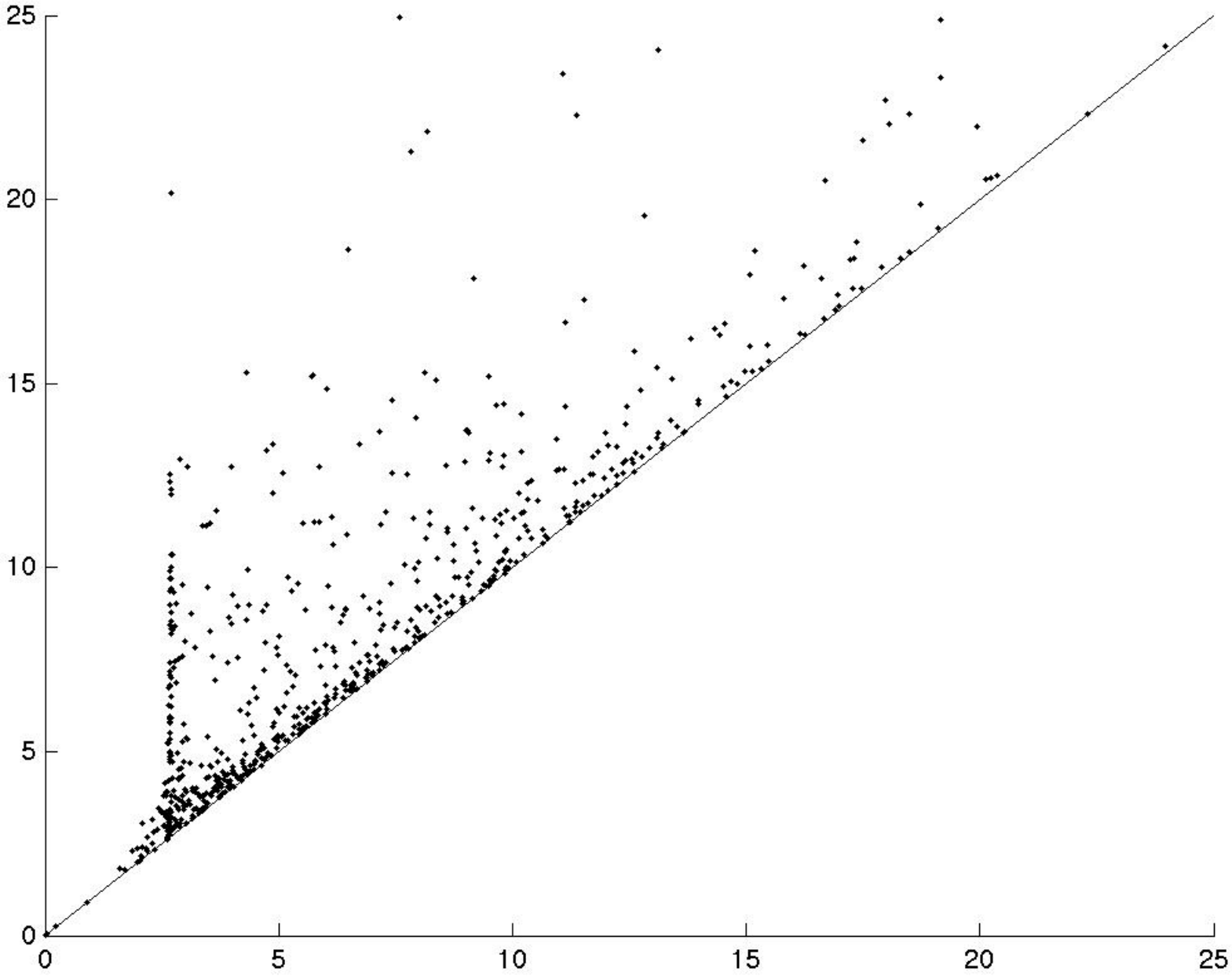

0
5
10
15
20
25

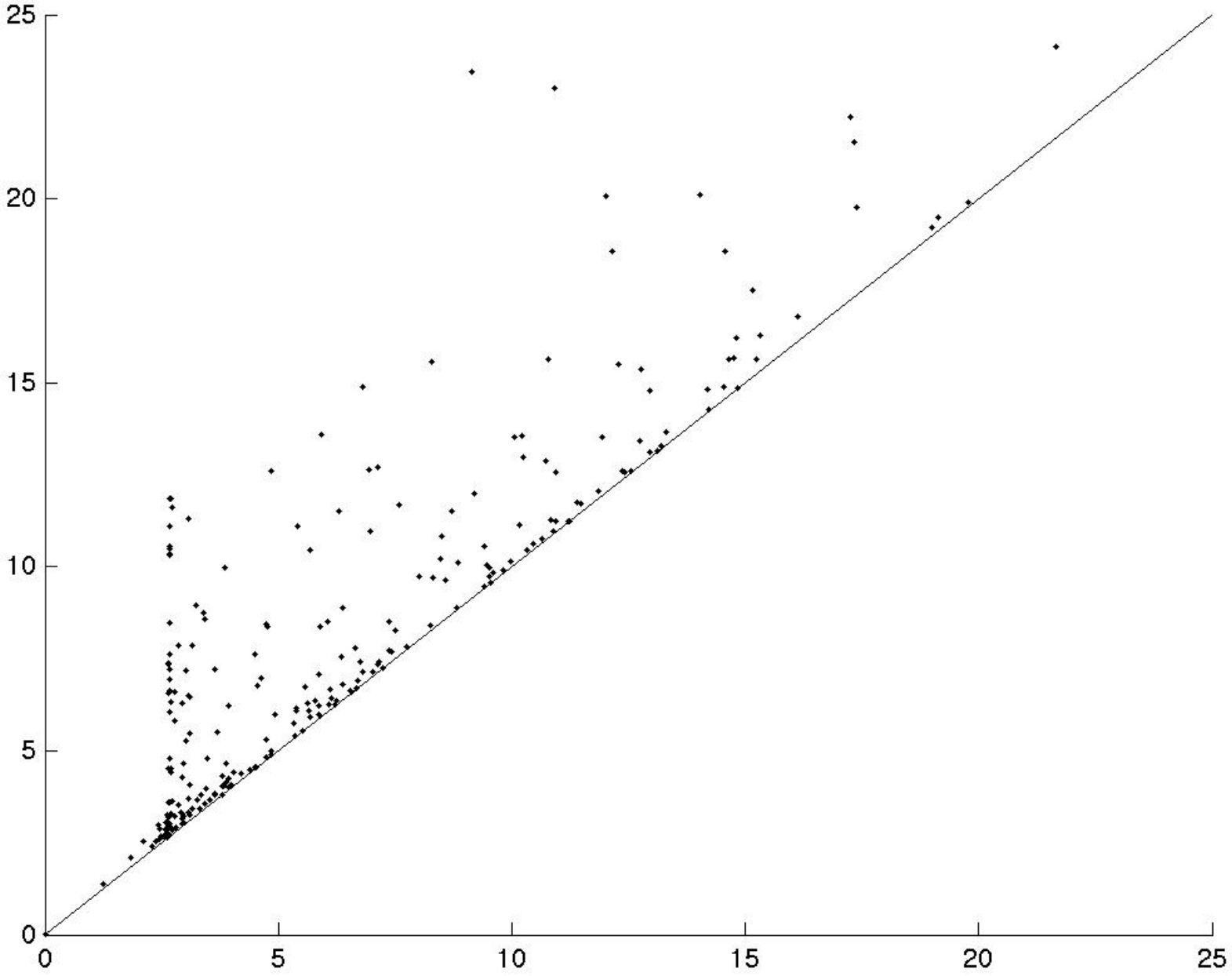





1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


25
20
15
10
5
0
0
5
10
15
20
25


25
20
15
10
5
0
0
5
10
15
20
25

The full data set consists of n = 98 (or 97) such trees from people whose ages range from 18 to 72 years old.

Each data point is a tree (representing arteries in human brains isolated via magnetic resonance imaging), embedded in 3-dimensional space, with additional attributes such as thickness (ignored).

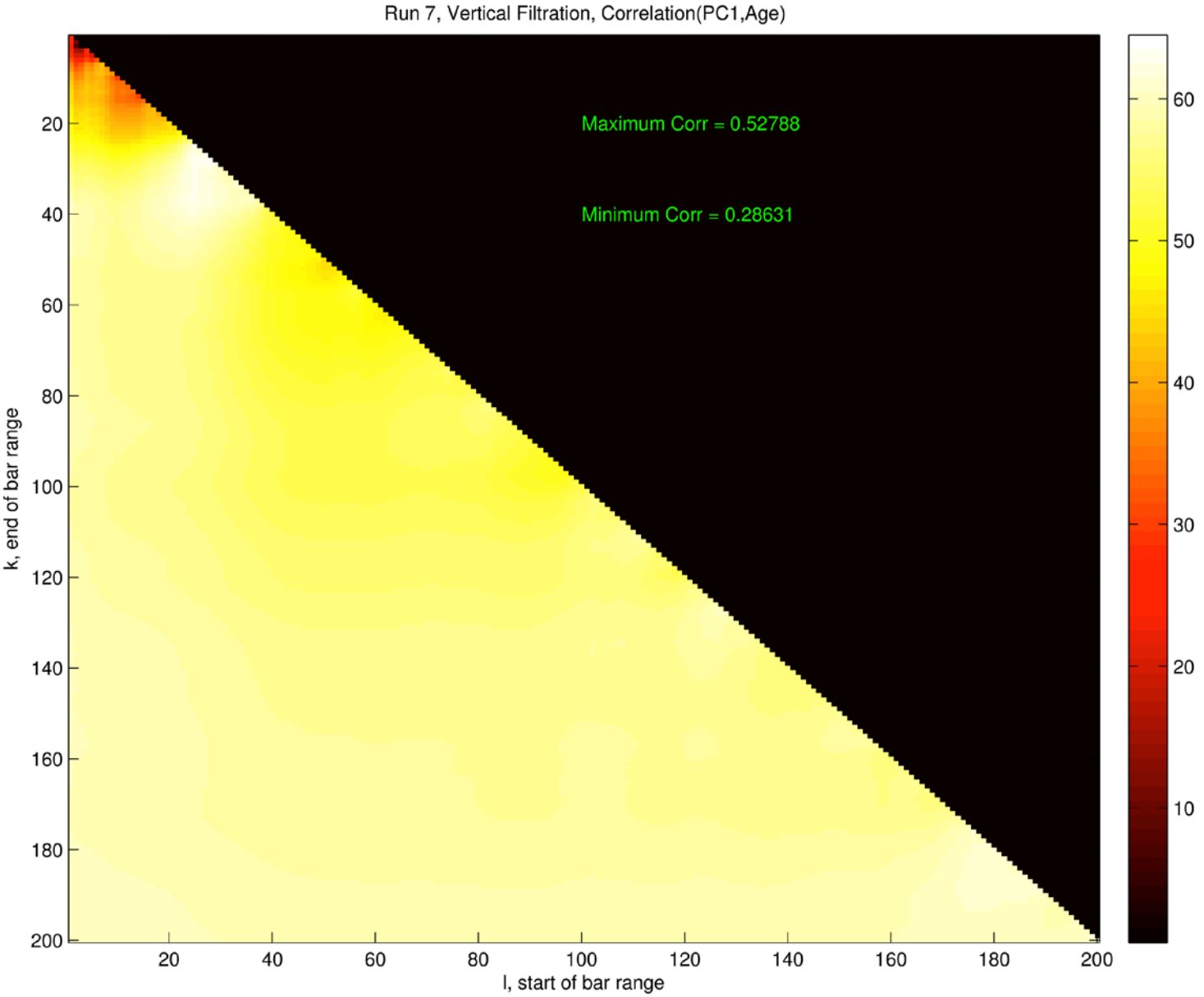
These diagrams are turned into feature vectors:

$(p_1, p_2, \ldots, p_{100})$ where p_i is the length of the i^{th} longest for for H_0.

$(q_1, q_2, \ldots, q_{100})$ where q_i is the length of the i^{th} longest for for H_1.

Age correlation heat map for features extracted from 0-dimensional persistent homology analysis. Color indicates the value of the function $\rho(n,N)$, which is the age correlation derived from the vectors $p_{n,N}$, with n: horizontal axis N: vertical axis.

p_1 = longest bar

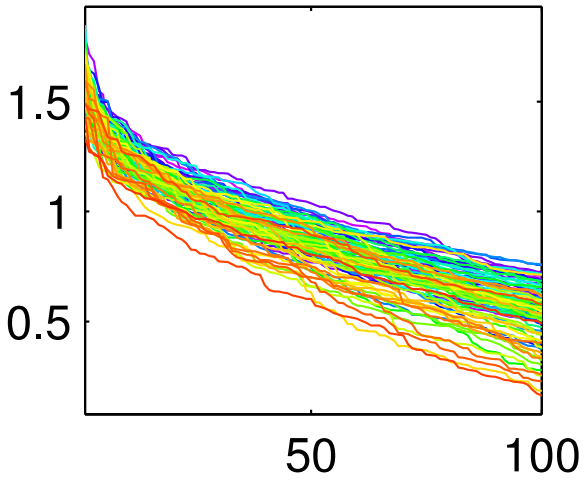


Color in the lower diagonal partof the plot codes correlation, ranging from very dark red (lowest)through hotter colors to white (highest correlation). The bottom of the color range is 0.29 and the top is 0.56, chosen to maximizeuse of the color scale.

TDA shows age to be correlated with certain measures of how brain arteries bend through space

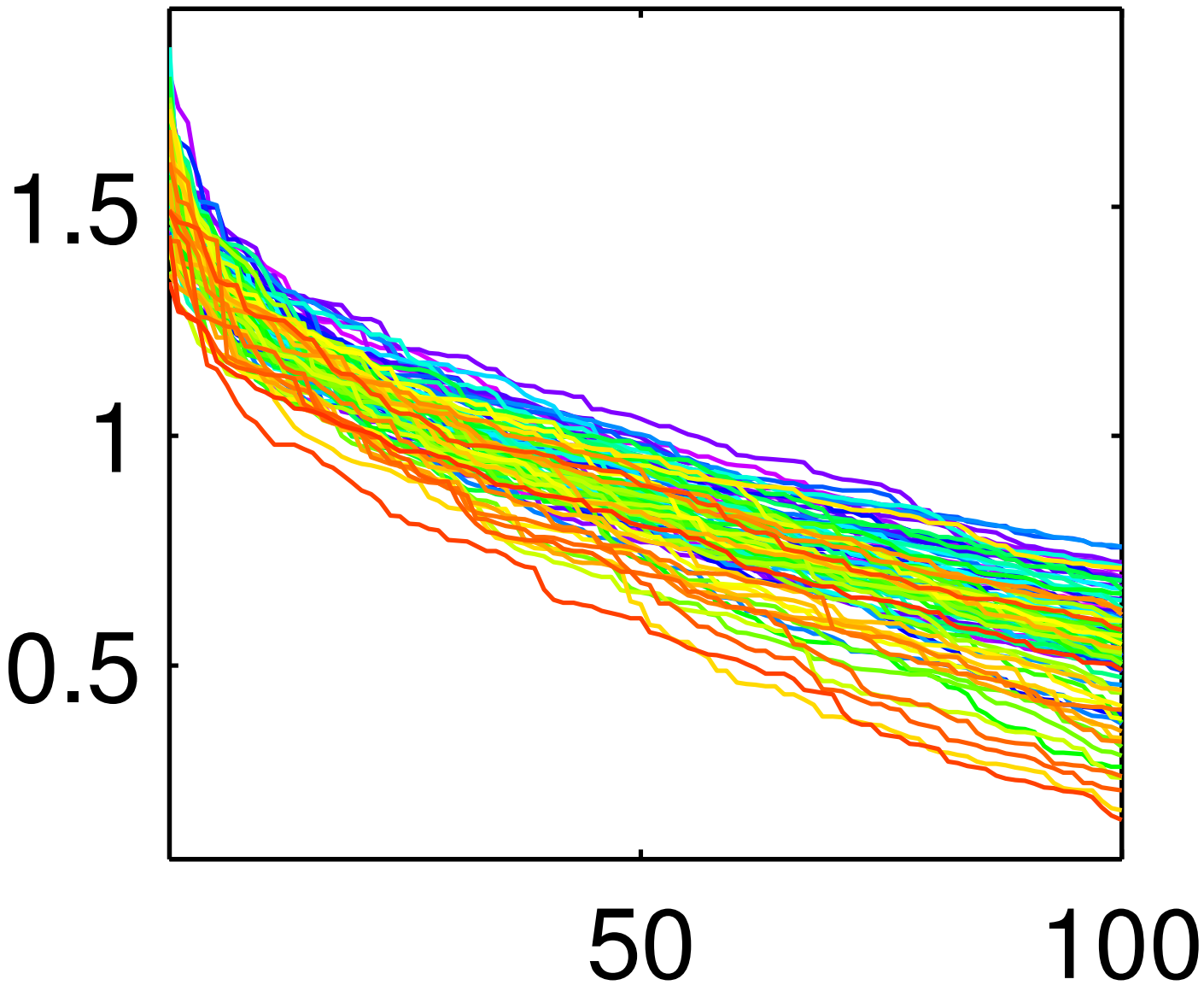
This contrasts with a previous study [Bullitt 2005] that correlates age with total artery length, and furthermore the TDA correlations are independent of that earlier one.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$



$(p_1, p_2, \ldots, p_{100})$

Color denotes age via a rainbow scheme starting with magenta for the youngest (19), ranging smoothly through blue,green, yellow to red for the oldest~(79)

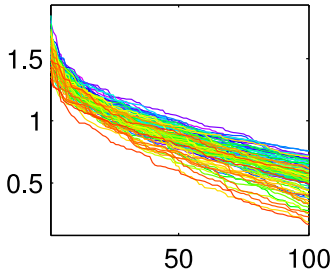


$(p_1, p_2, \ldots, p_{100})$

Color denotes age via a rainbow scheme starting with magenta for the youngest (19), ranging smoothly through blue,green, yellow to red for the oldest~(79)

$(p_1, p_2, \ldots, p_{100})$

Color denotes age via a rainbow scheme starting with magenta for the youngest (19), ranging smoothly through blue,green, yellow to red for the oldest~(79)

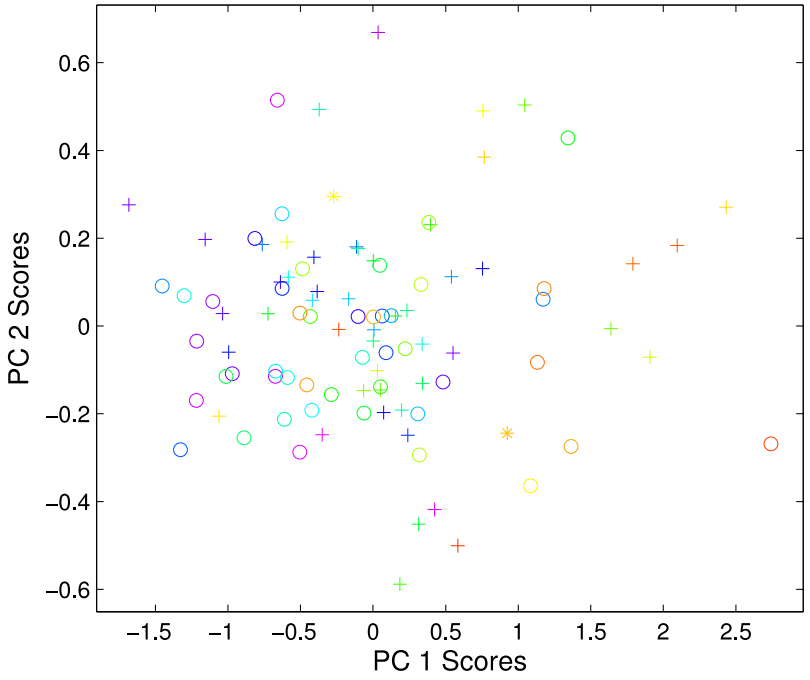


$(p_1, p_2, \ldots, p_{100})$

Color denotes age via a rainbow scheme starting with magenta for the youngest (19), ranging smoothly through blue,green, yellow to red for the oldest~(79)

s

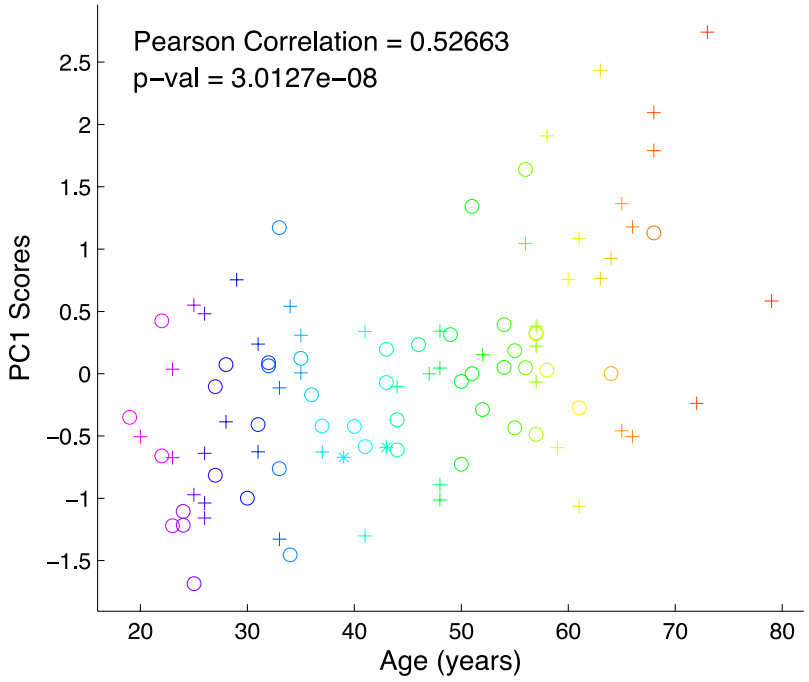
0-dim



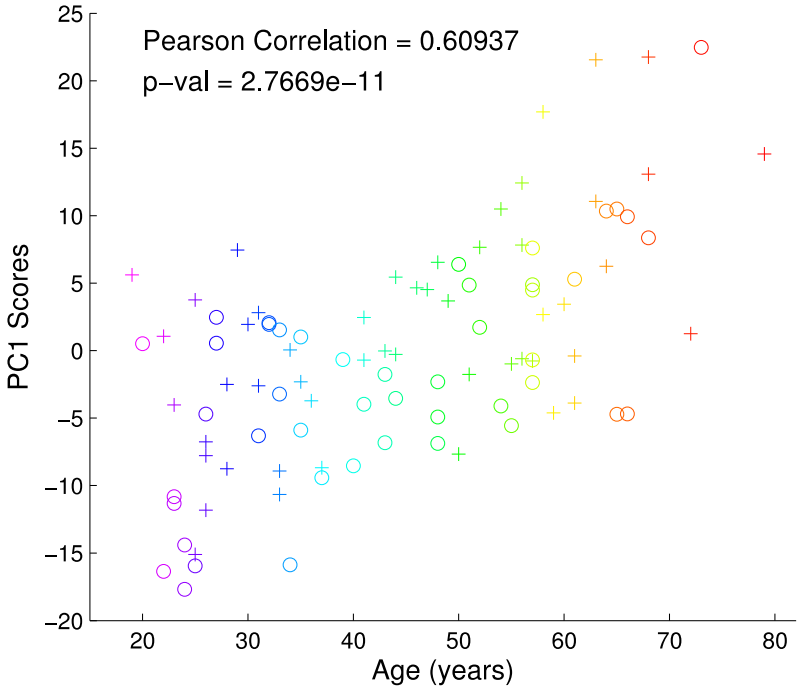
Females = red circles

Males = blue + signs

0-dim



1-dim

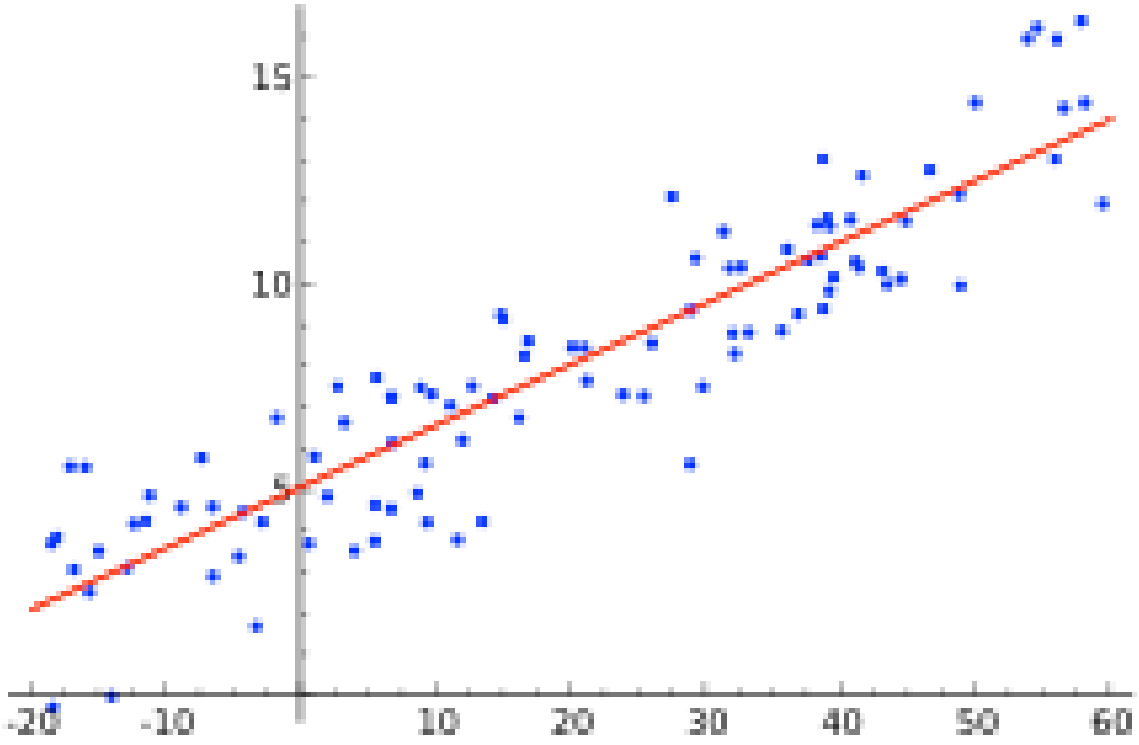


Linear regression:

For each I plot p_i versus L = total artery length

Residuals → Pearson correlation of $\rho = 0.52$

p-value = 10^{-8}



https://en.wikipedia.org/wiki/Linear_regression

TDA shows age to be correlated with certain measures of how brain arteries bend through space

This contrasts with a previous study [Bullitt 2005] that correlates age with total artery length, and furthermore the TDA correlations are independent of that earlier one.

Geometrically motivated methods to control for effects of total artery length yield similarly negligible increases or decreases in Pearson correlation and p -value. These methods simply divide the numbers p_i by (i) L or (ii) $L^{½}$ or (iii) $L^{2/3}$
before running the analysis in Section 3.2.

Exponents on L correspond to physical models where vessel length

(i) scales according to total linear skull size,

(ii) has constant flux (i.e. number of arteries passing) through each unit of cross-sectional area, or

(iii) remains constant per unit volume.

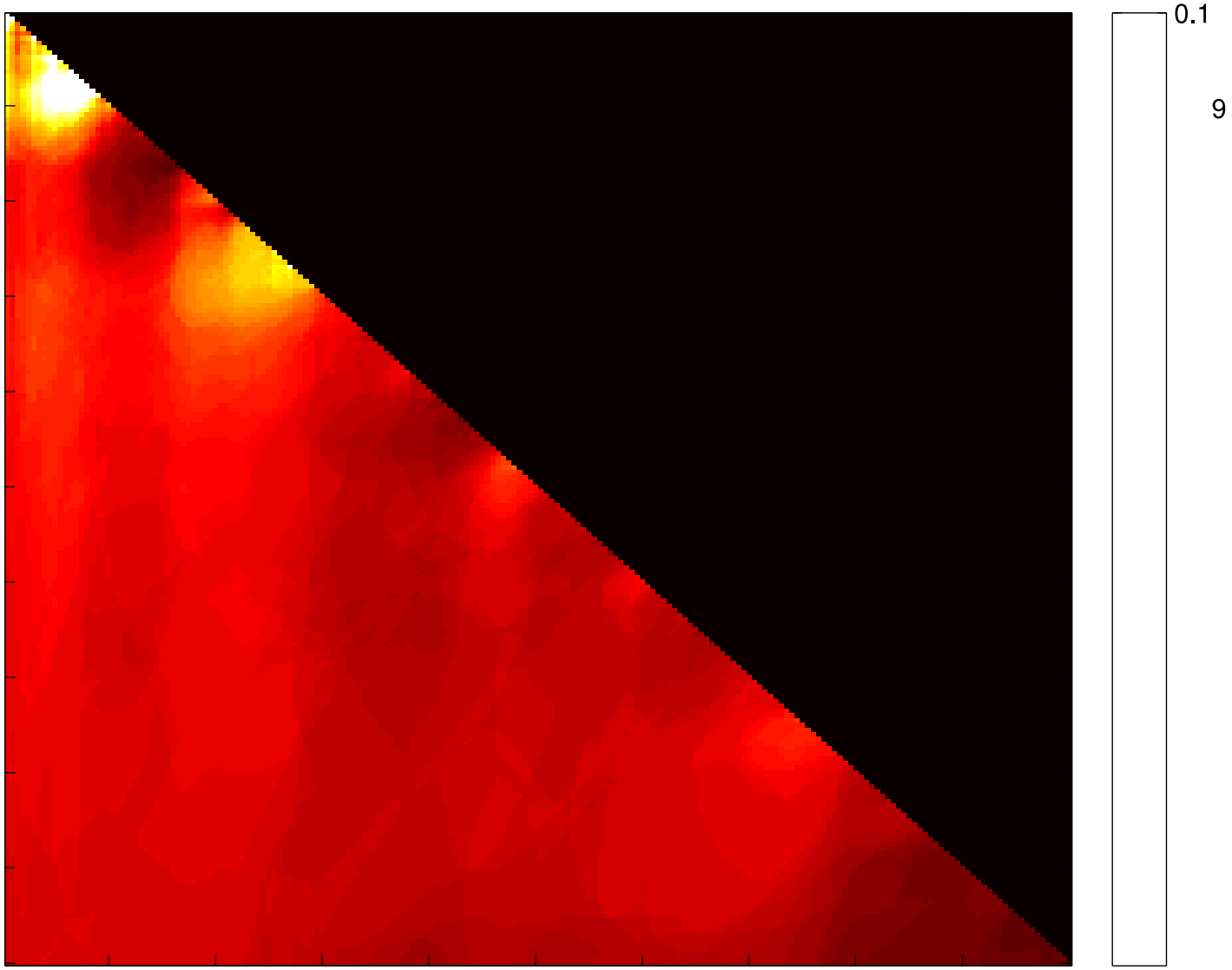
Controlling for total length in the one-dimensional persistence analysis from Section 3.2 yields decidedly weaker (but still non-negligible) age correlation: replacing the q_i features with their residuals, after running a linear regression between each q_i and L , results in Pearson correlation ρ = 0. 35.

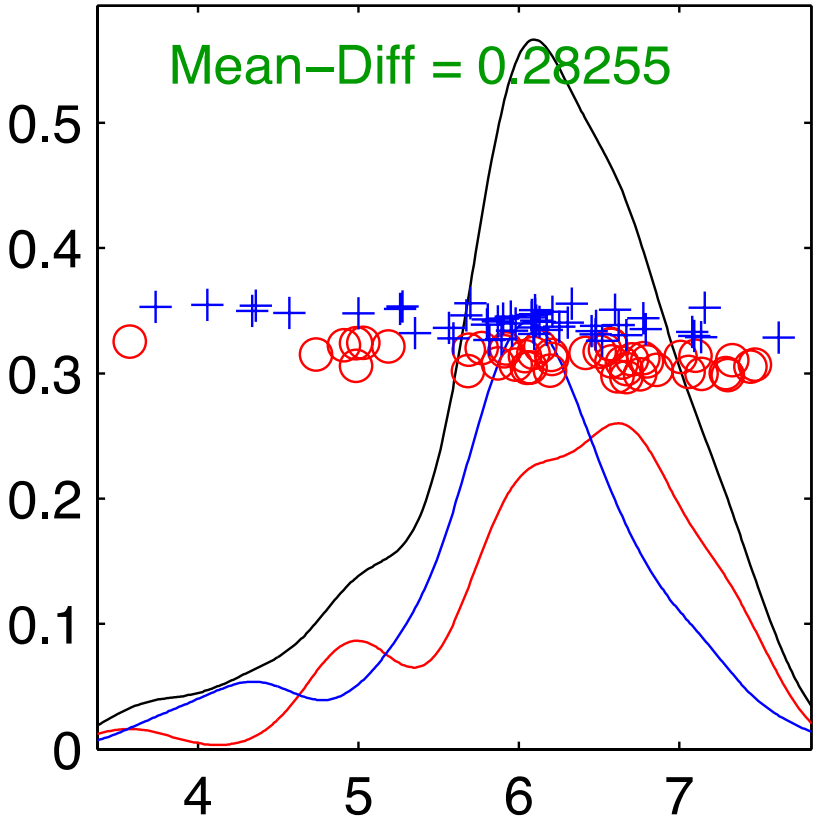
TDA shows age to be correlated with certain measures of how brain arteries bend through space

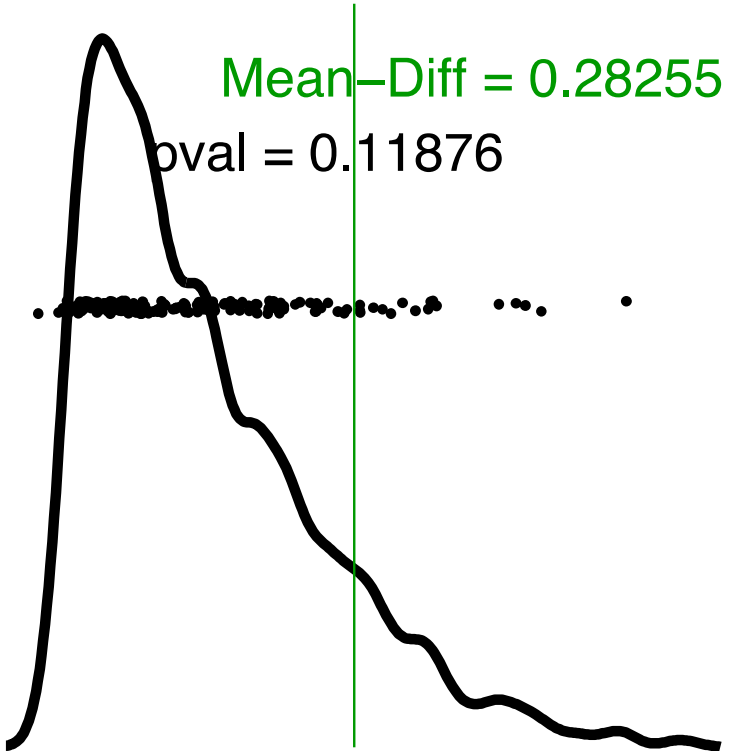
This contrasts with a previous study [Bullitt 2005] that correlates age with total artery length, and furthermore the TDA correlations are independent of that earlier one.

TDA in our context also finds stronger sex effects than the only other study [Shen 2013] to find any sex difference at all.

Sex difference significance heat map for features extracted from 1-dimensional persistent homology analysis..







Females = red circles

Males = blue + signs

Figure 7: Illustration of DiProPerm results on the zero-dimensional persistence features. Left: the result of projecting the data onto the direction vector determined by the means, suggesting some difference.

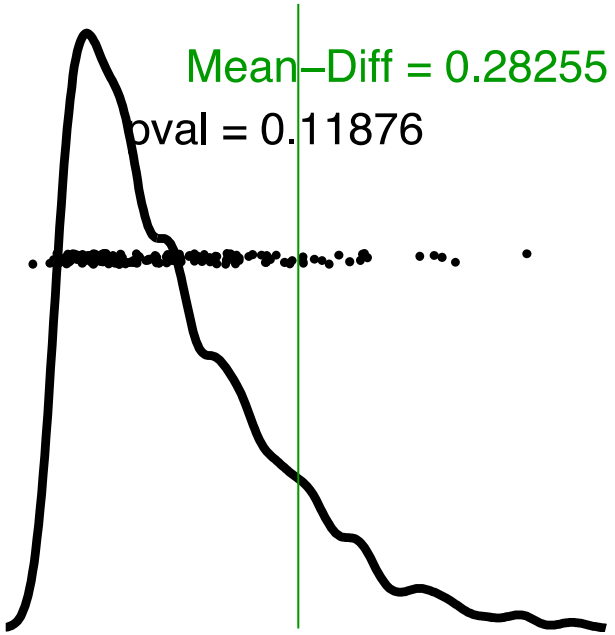
Right: The results of the permutation test, with the proportion of simulated differences that are bigger than that observed in the original data giving an empirical p-value.

Calculate the arithmetic mean of the vectors (p_1, …, p_{100}) corresponding to male subjects, and separately for the female subjects,

Compute the Euclidean distance between these means in R^{100} .

The size of this mean difference alone does not tell much as a raw number. Thus need

Permutation test: randomly reassign the 98 vectors into two groups of equal size,
compute the difference between the means of the two groups,
repeat this procedure 1000 times.



This method has been called DiProPerm.

In our test, 119 of the reassignments led to a larger mean-difference than the original male/female split, giving an estimated p-value of 0.1, which is not impressive.

However, we then repeated the entire procedure for the loop-vectors (q_1, …, q_{100}), and found a more compelling p-value of 0.03.

The p-value for the permutation test is the proportion of the *r* values that are larger than \ the Pearson correlation coefficient that was calculated from the original data.

http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-correctly-interpret-p-values

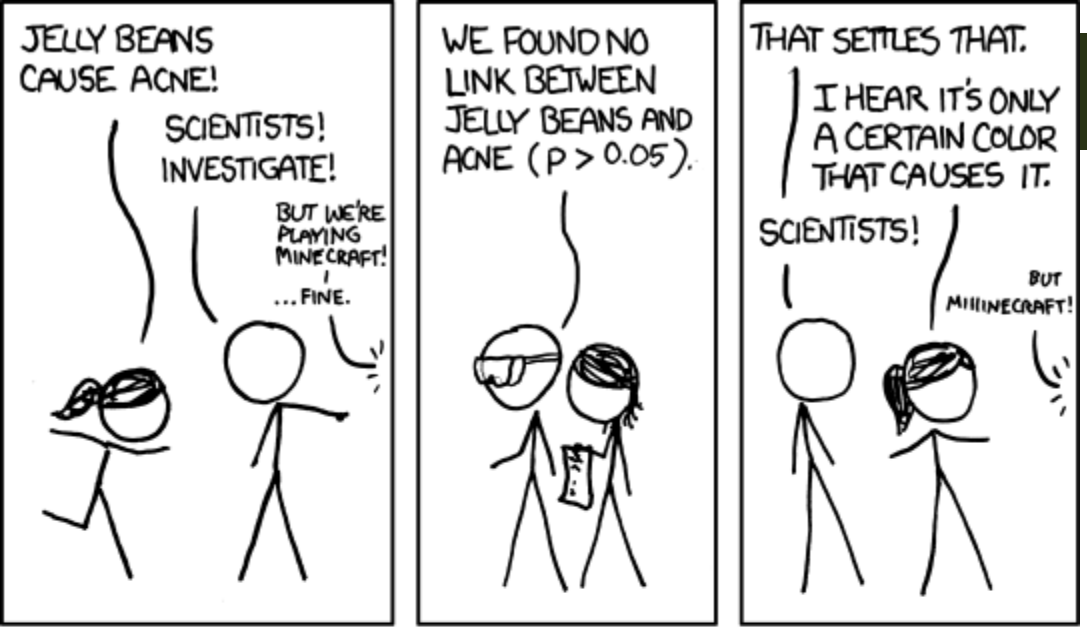
P value	Probability of incorrectly rejecting a true null hypothesis
0.05	At least 23% (and typically close to 50%)
0.01	At least 7% (and typically close to 15%)

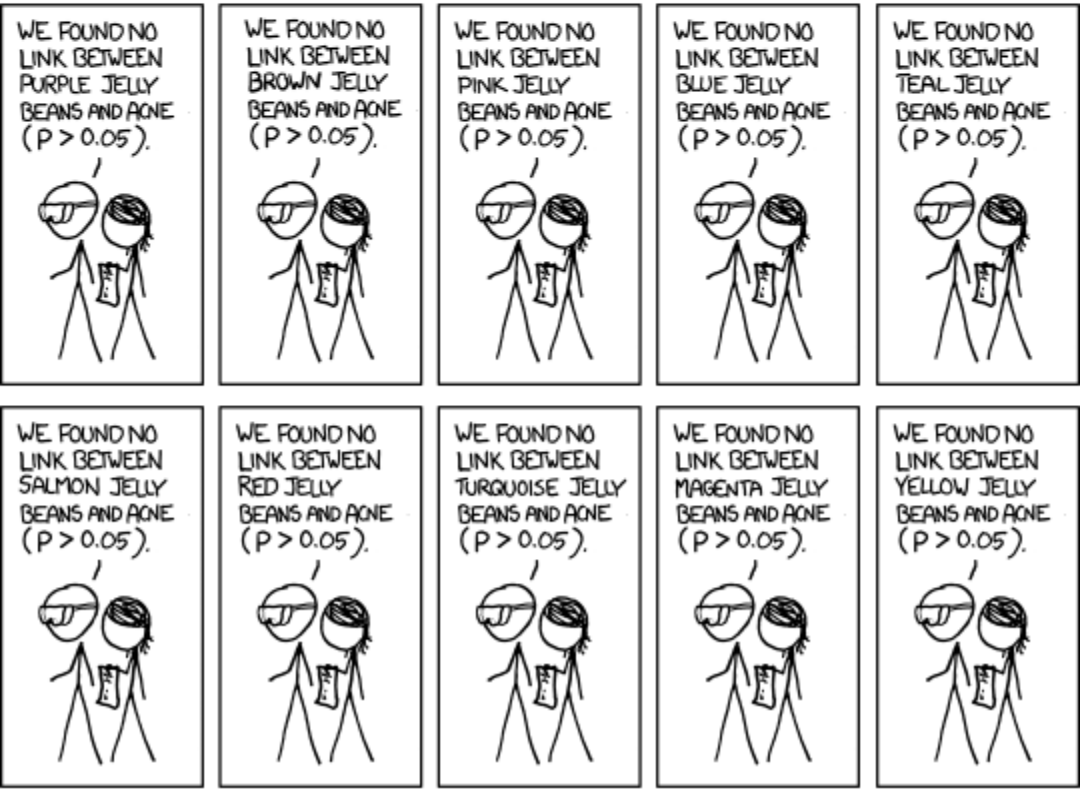
Example: vaccine study with P value of 0.04:

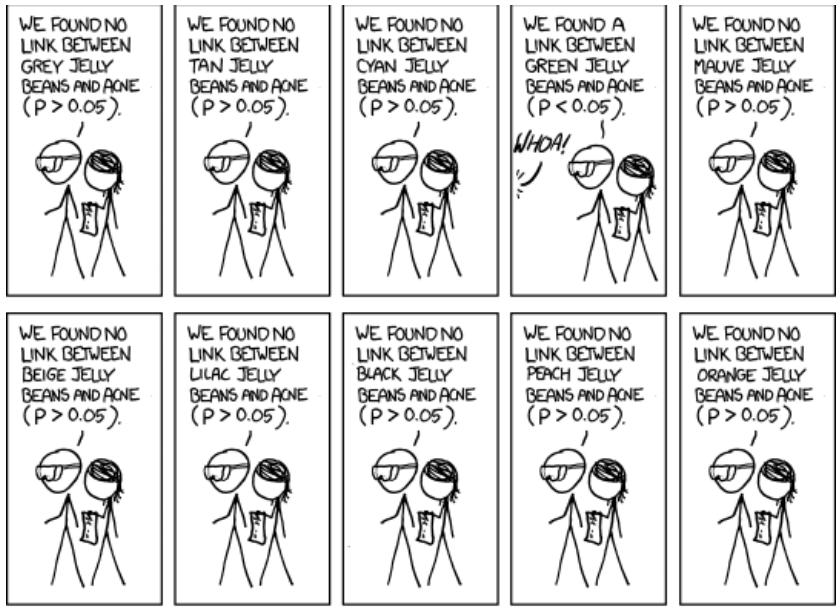
Correct: Assuming that the vaccine had no effect, you'd obtain the observed difference or more in 4% of studies due to random sampling error.

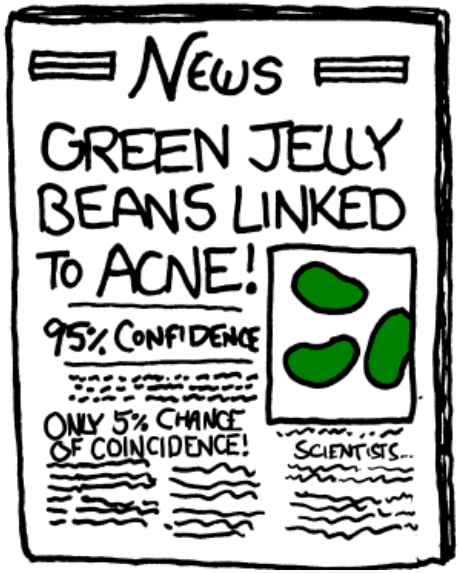
Incorrect: If you reject the null hypothesis, there's a 4% chance that you're making a mistake.

False Positives will occur









https://xkcd.com/882/

The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications

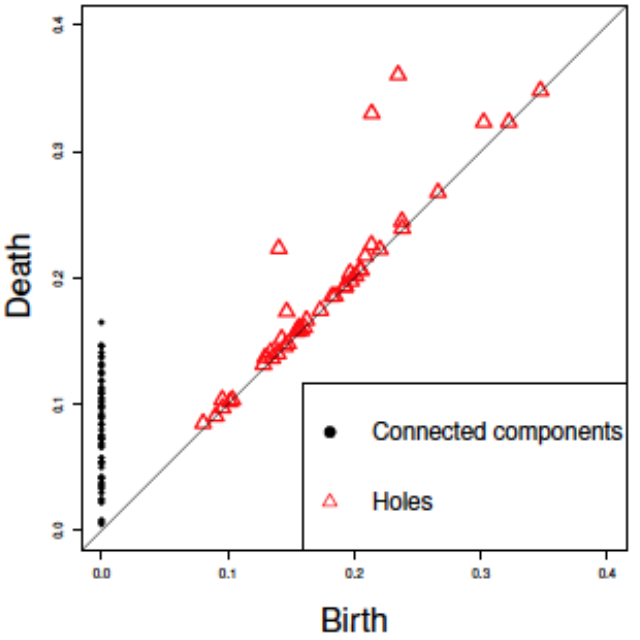
Christophe A.N. Biscio* and Jesper Møller*

November 3, 2016

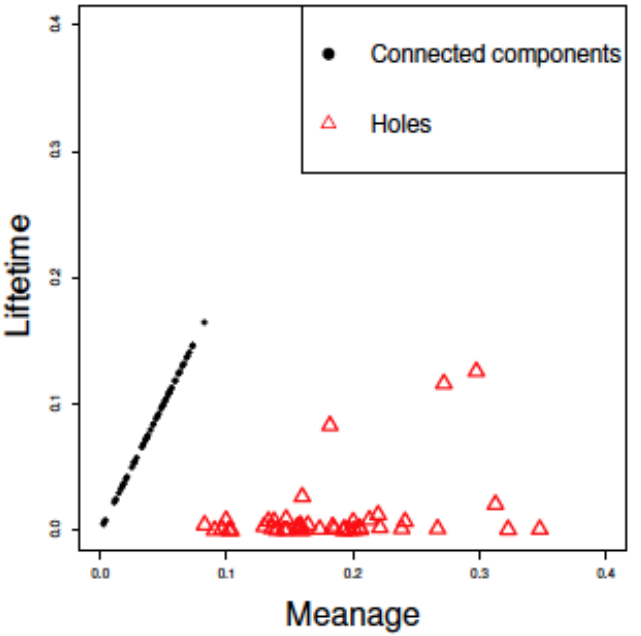
Persistent homology - Persistence diagram

- A persistent diagram consists of points (b_i, d_i) representing as r varies a connected components (holes) appearing at $r = b_i$ (birth) and disappearing at $r = d_i$ (death),
- possibly with multiplicity c_i.
- $(b_i, d_i) \leftrightarrow (m_i, l_i)$, where $m_i = \frac{b_i + d_i}{2}$ and $l_i = d_i - b_i$.

Persistence diagram



Rotated and rescaled persistence diagram



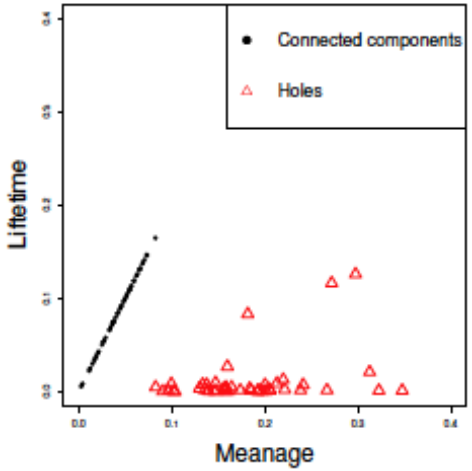
A new functional summary statistic

The **accumulated persistence function** (where $k = 0$ if connected components are considered, $k = 1$ if holes):

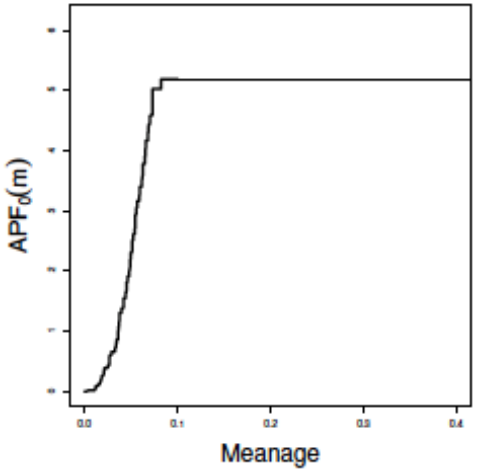
$$\mathrm{APF}_k(m) = \sum_i c_i l_i 1(m_i \leq m), \quad m \geq 0.$$

Example:

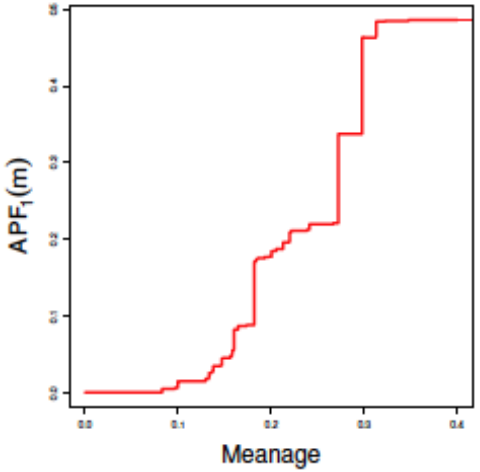
Rotated and rescaled persistent diagram



Connected components



Holes



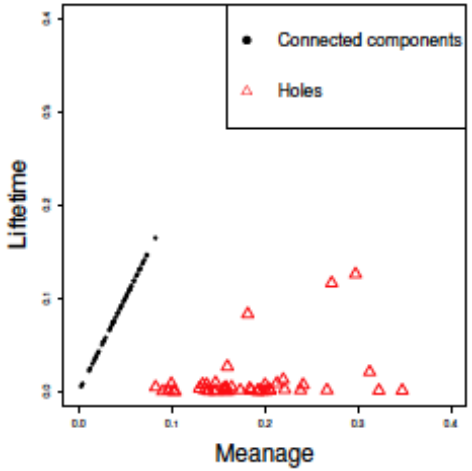
A new functional summary statistic

The **accumulated persistence function** (where $k = 0$ if connected components are considered, $k = 1$ if holes):

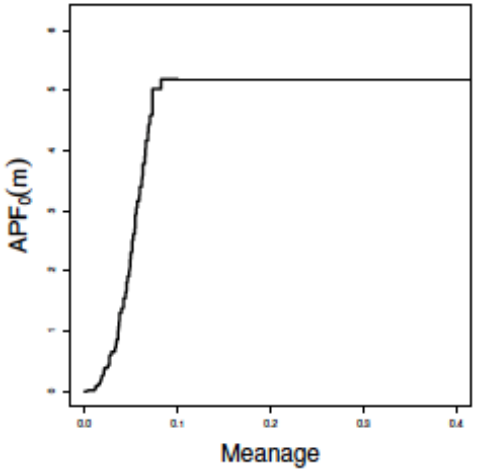
$$\mathrm{APF}_k(m) = \sum_i c_i l_i 1(m_i \leq m), \quad m \geq 0.$$

Example:

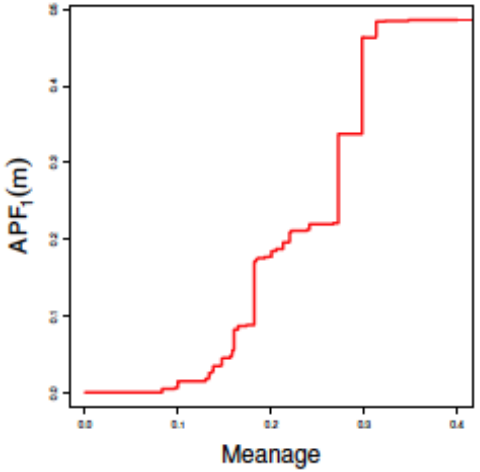
Rotated and rescaled persistent diagram



Connected components



Holes



If all $c_i = 1$ and all m_i are different, then barcode can be determined from APF.

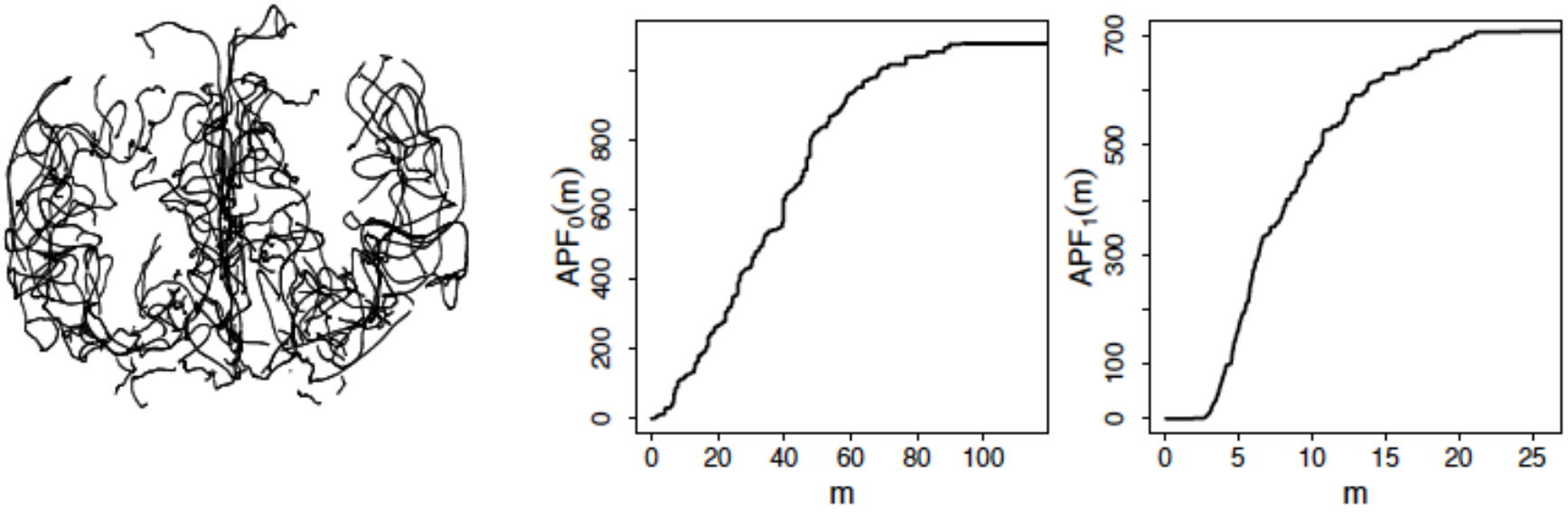
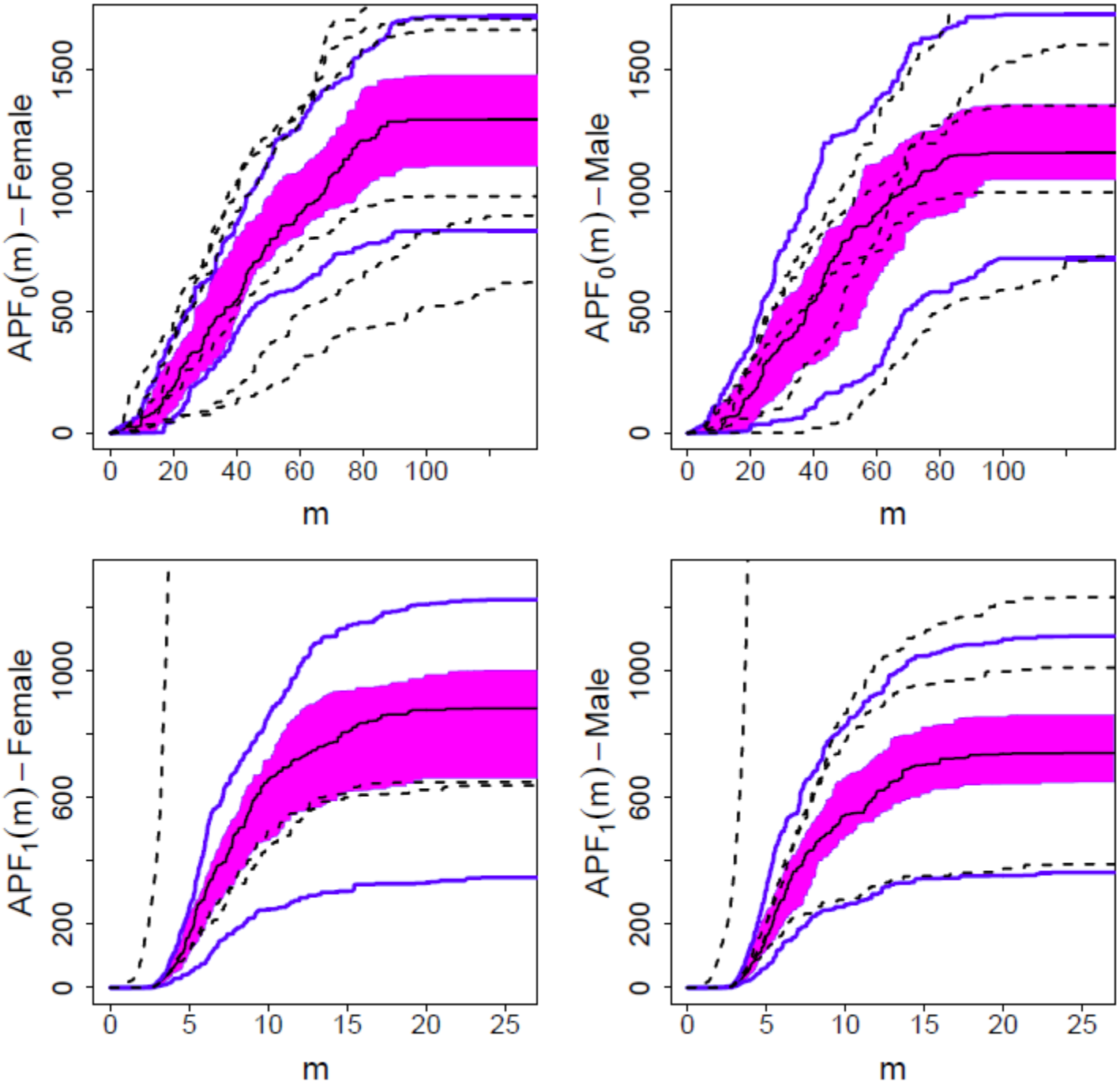


Figure 1: A brain artery tree (left panel), its corresponding APF_0 (middle panel) obtained from the sub-level set of the height function, and its corresponding APF_1 (right panel) obtained from the sub-level set of the distance function.

$$\mathrm{APF}_1(m) = \sum_{i=1}^{n} c_i l_i 1(m_i \leq m,\, m_i + l_i/2 \leq T), \quad m \geq 0.$$


APF0(m) – Female
APF0(m) – Male
APF1(m) – Female
APF1(m) – Male
m
0
500
1000
1500
200
600
0
20
40
60
80
100
5
10
15
20
25

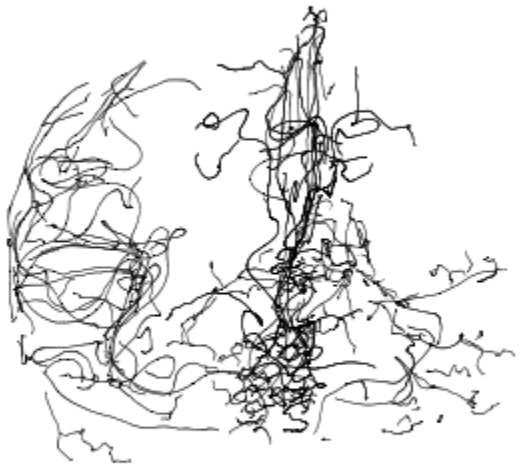
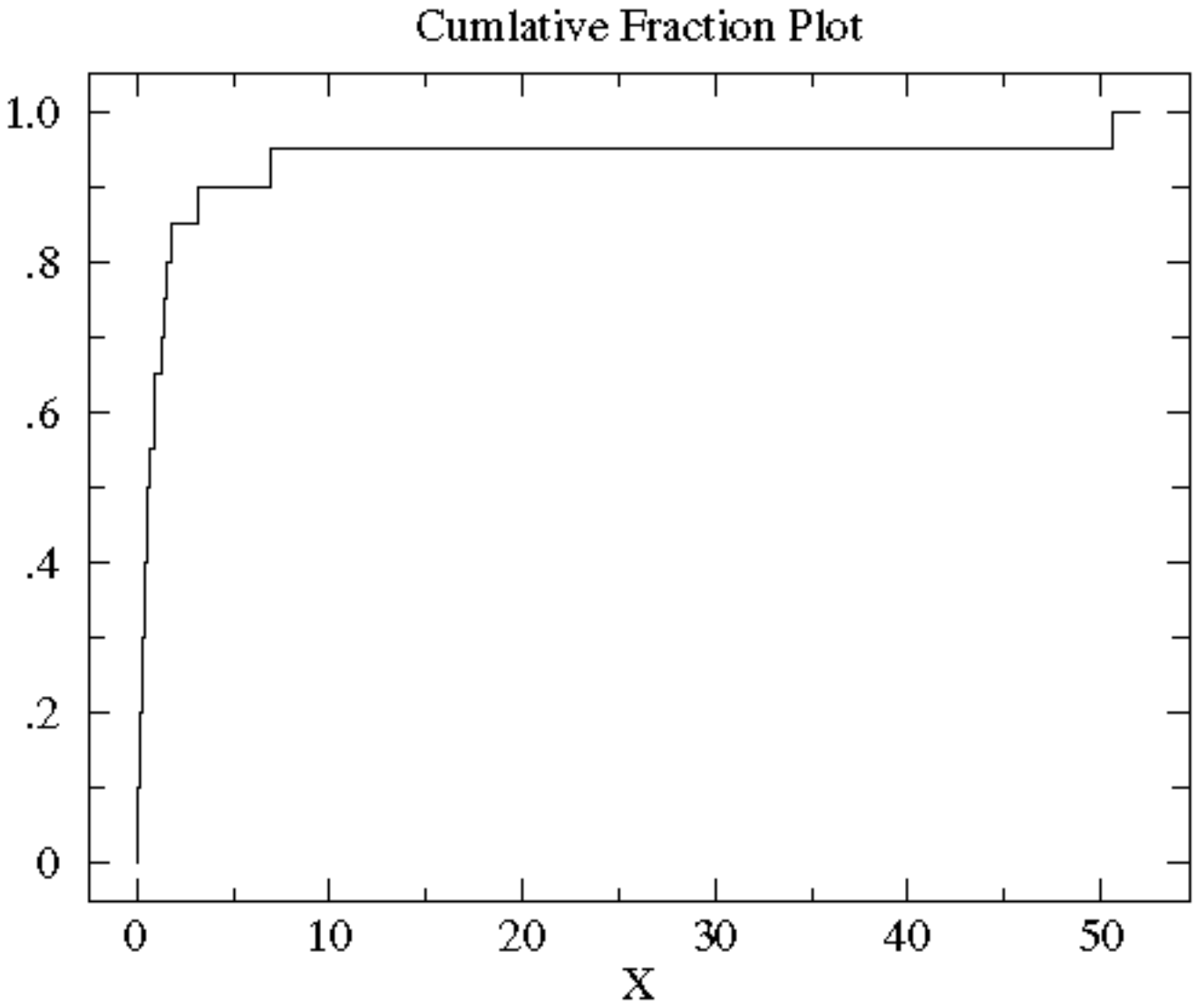


Figure 9: Brain artery tree of a male subject with APF_0^M and APF_1^M detected as outliers by the 1.5 criterion.

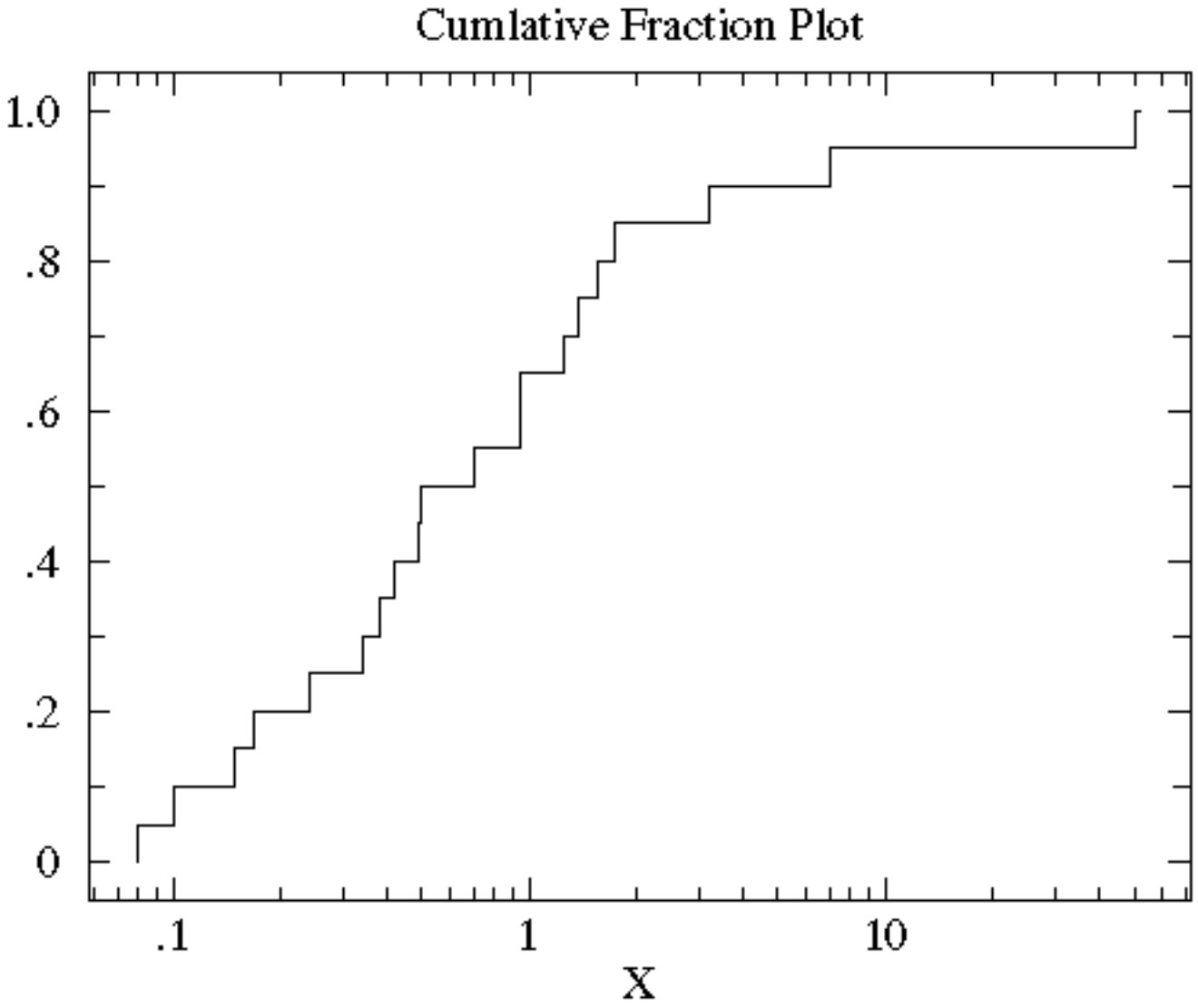
Kolmogorov-Smirnov Test

http://www.physics.csbsju.edu/stats/KS-test.html



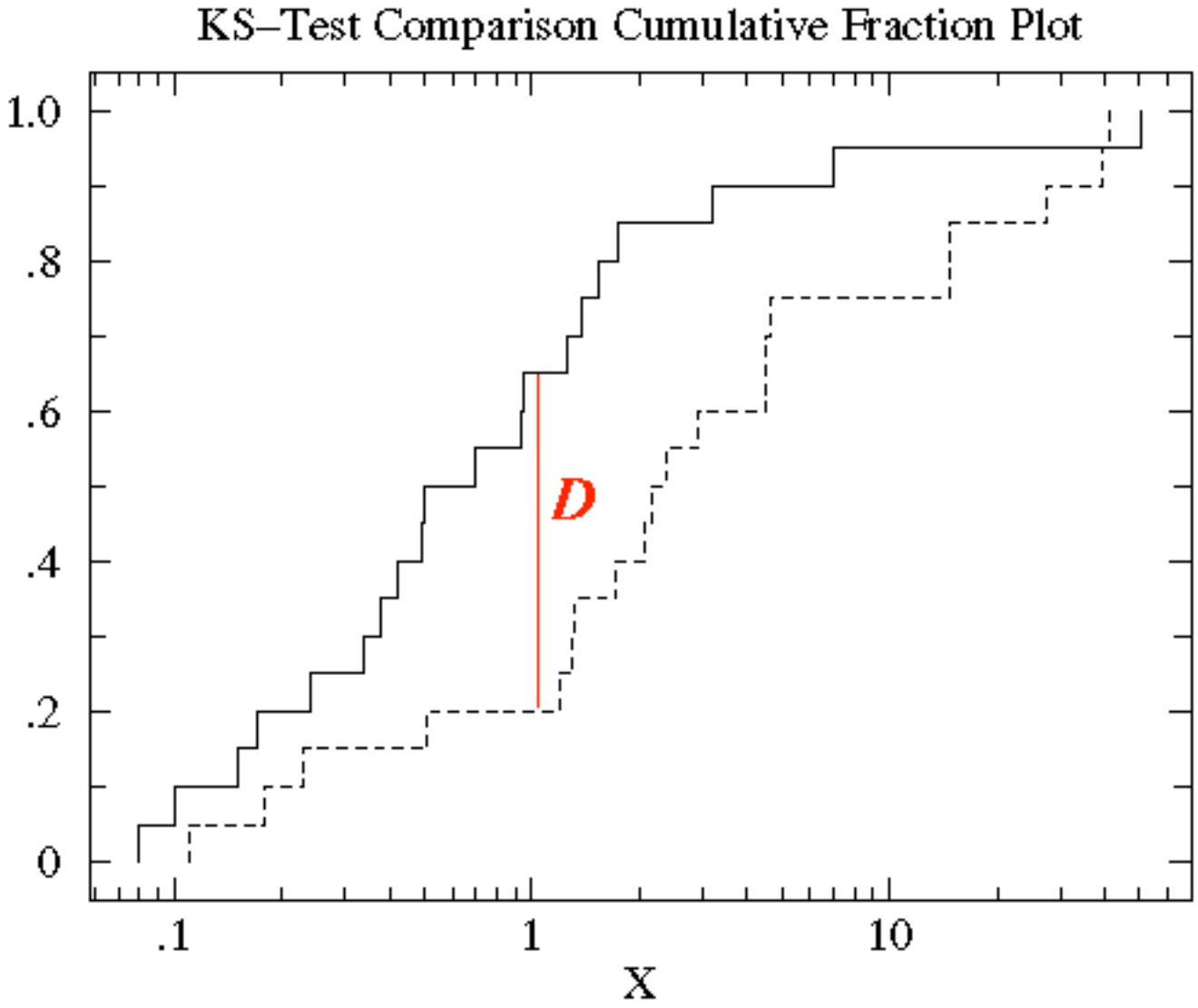
Sorted controlB={0.08, 0.10, 0.15, 0.17, 0.24, 0.34, 0.38, 0.42, 0.49, 0.50, 0.70, 0.94, 0.95, 1.26, 1.37, 1.55, 1.75, 3.20, 6.98, 50.57}

http://www.physics.csbsju.edu/stats/KS-test.html



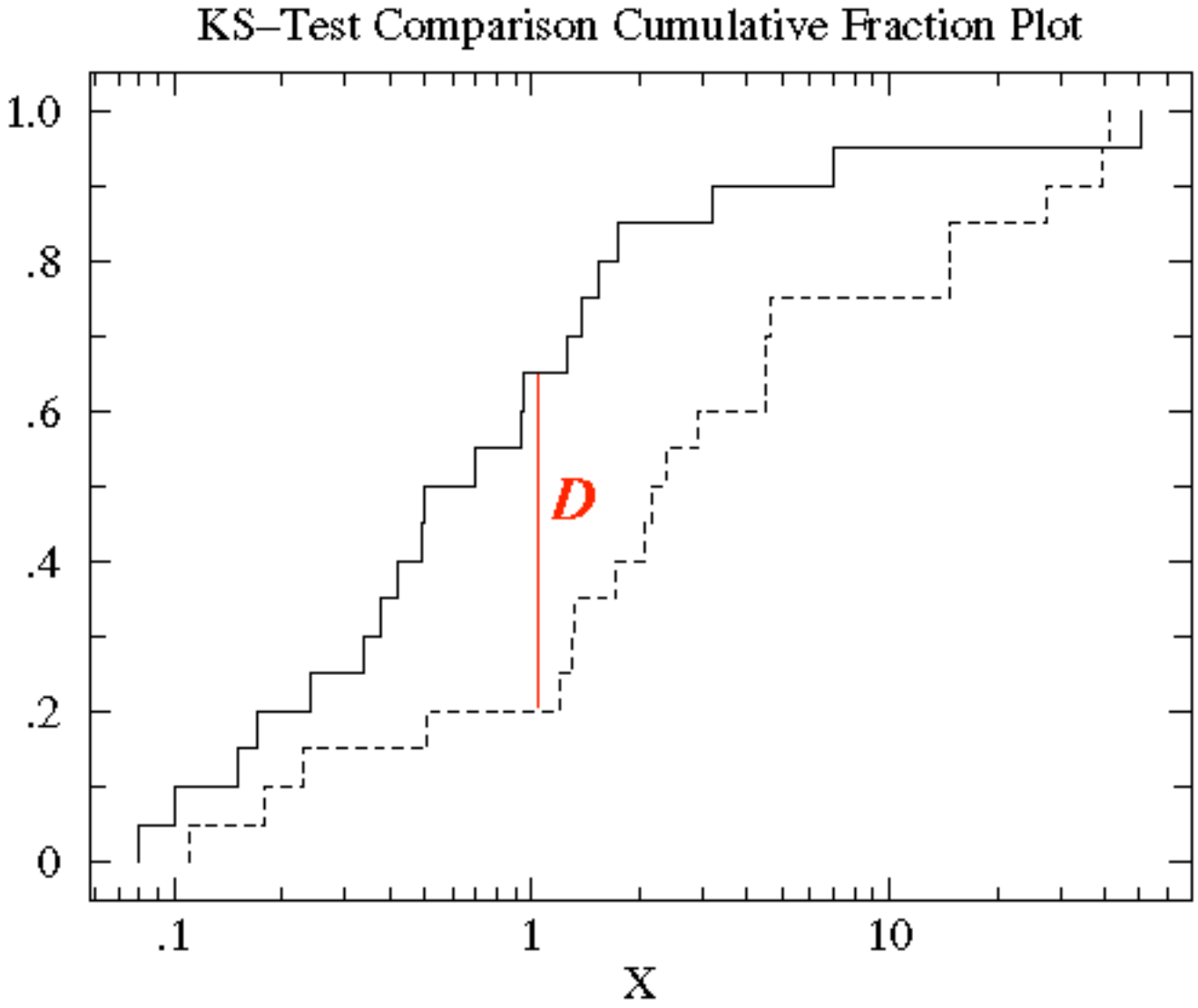
Sorted controlB={0.08, 0.10, 0.15, 0.17, 0.24, 0.34, 0.38, 0.42, 0.49, 0.50, 0.70, 0.94, 0.95, 1.26, 1.37, 1.55, 1.75, 3.20, 6.98, 50.57}

http://www.physics.csbsju.edu/stats/KS-test.html

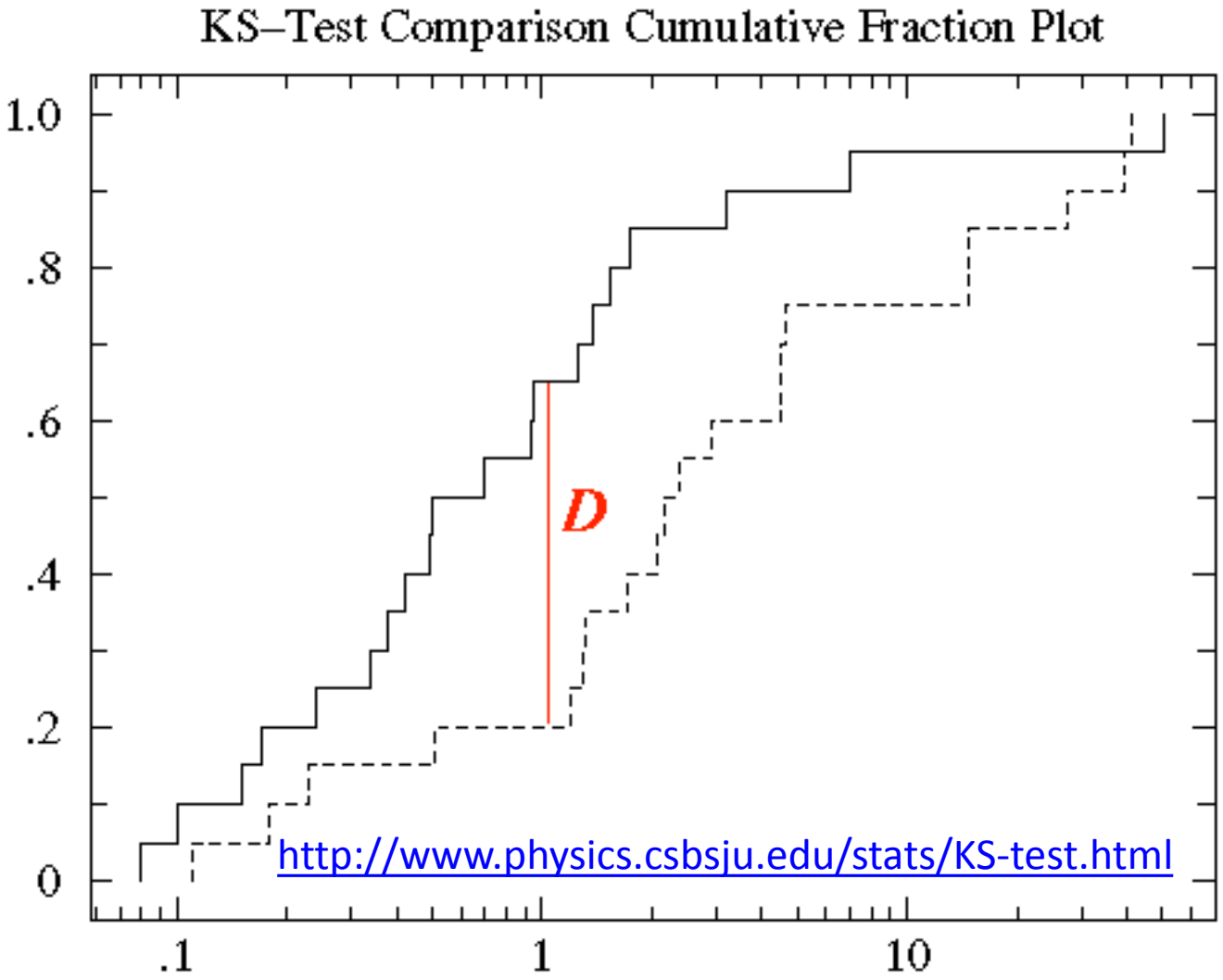


treatmentB= {2.37, 2.16, 14.82, 1.73, 41.04, 0.23, 1.32, 2.91, 39.41, 0.11, 27.44, 4.51, 0.51, 4.50, 0.18, 14.68, 4.66, 1.30, 2.06, 1.19}

http://www.physics.csbsju.edu/stats/KS-test.html



treatmentB= {0.11, 0.18, 0.23, 0.51, 1.19, 1.30, 1.32, 1.73, 2.06, 2.16, 2.37, 2.91, 4.50, 4.51, 4.66, 14.68, 14.82, 27.44, 39.41, 41.04}



The KS-test uses the maximum vertical deviation between the two curves as the statistic D. In this case the maximum deviation occurs near x=1 and has D=.45. (The fraction of the treatment group that is less then one is 0.2 (4 out of the 20 values); the fraction of the control group that is less than one is 0.65 (13 out of the 20 values). Thus the maximum difference in cumulative fraction is D=.45.)

(A) For $k = 0, 1$, we let PD_k' be the subset of PD_k corresponding to the 100 largest lifetimes. Then $D_1, \ldots, D_{46}$ and $E_1, \ldots, E_{49}$ are the RRPD_ks obtained from the PD_k's associated to female and male subjects, respectively. This is the setting used in Bendich *et al.* (2016).

(B) For $k = 0, 1$, we consider all lifetimes and let $D_1, \ldots, D_{46}$ and $E_1, \ldots, E_{49}$ be the RRPD_ks associated to female and male subjects, respectively.

(C) The samples are as in setting (B) except that we exclude the RRPD_ks where the corresponding APF_k was detected as an outlier in Example 4. Hence, $r_1 = 40$ and $r_2 = 43$ if $k = 0$, and $r_1 = 43$ and $r_2 = 45$ if $k = 1$.

	APF_0		APF_1	
	$I = [0, 137]$	$I = [0, 60]$	$I = [0, 25]$	$I = [15, 25]$
Setting (A)	5.26	3.26	3.18	2.72
Setting (B)	7.67	3.64	20.06	1.83
Setting (C)	4.55	2.61	0.92	0.85

Table 2: Estimated p-values of the two-sample test based on KS_{r_1,r_2} used with APF_0 and APF_1 on different intervals I to distinguish between male and female subjects under settings (A), (B), and (C) described in Example 8

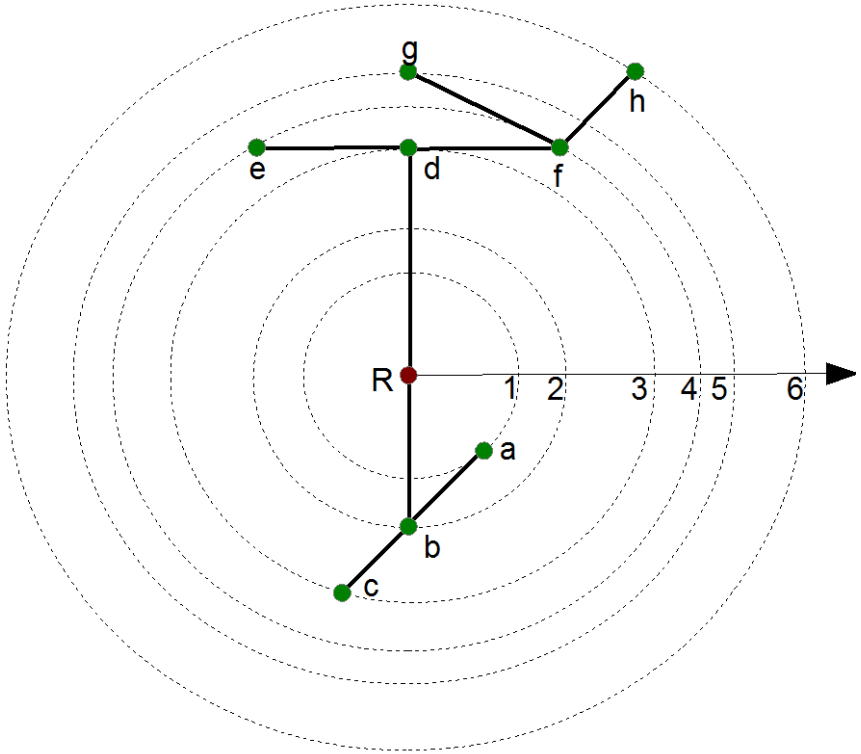
$$KS_{r_1,r_2} = \sqrt{\frac{r_1 r_2}{r}} \sup_{m \in I} \left| \overline{A_{r_1}}(m) - \overline{A_{r_2}}(m) \right|$$

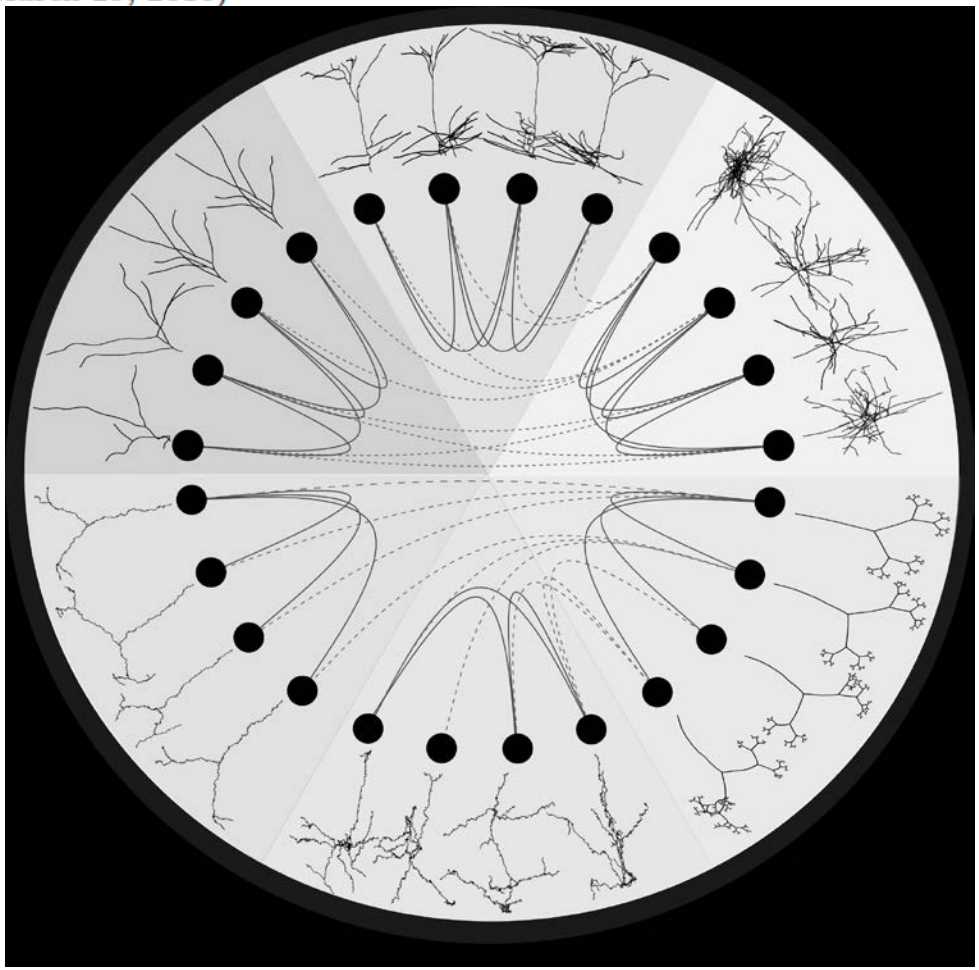
Quantifying topological invariants of neuronal morphologies

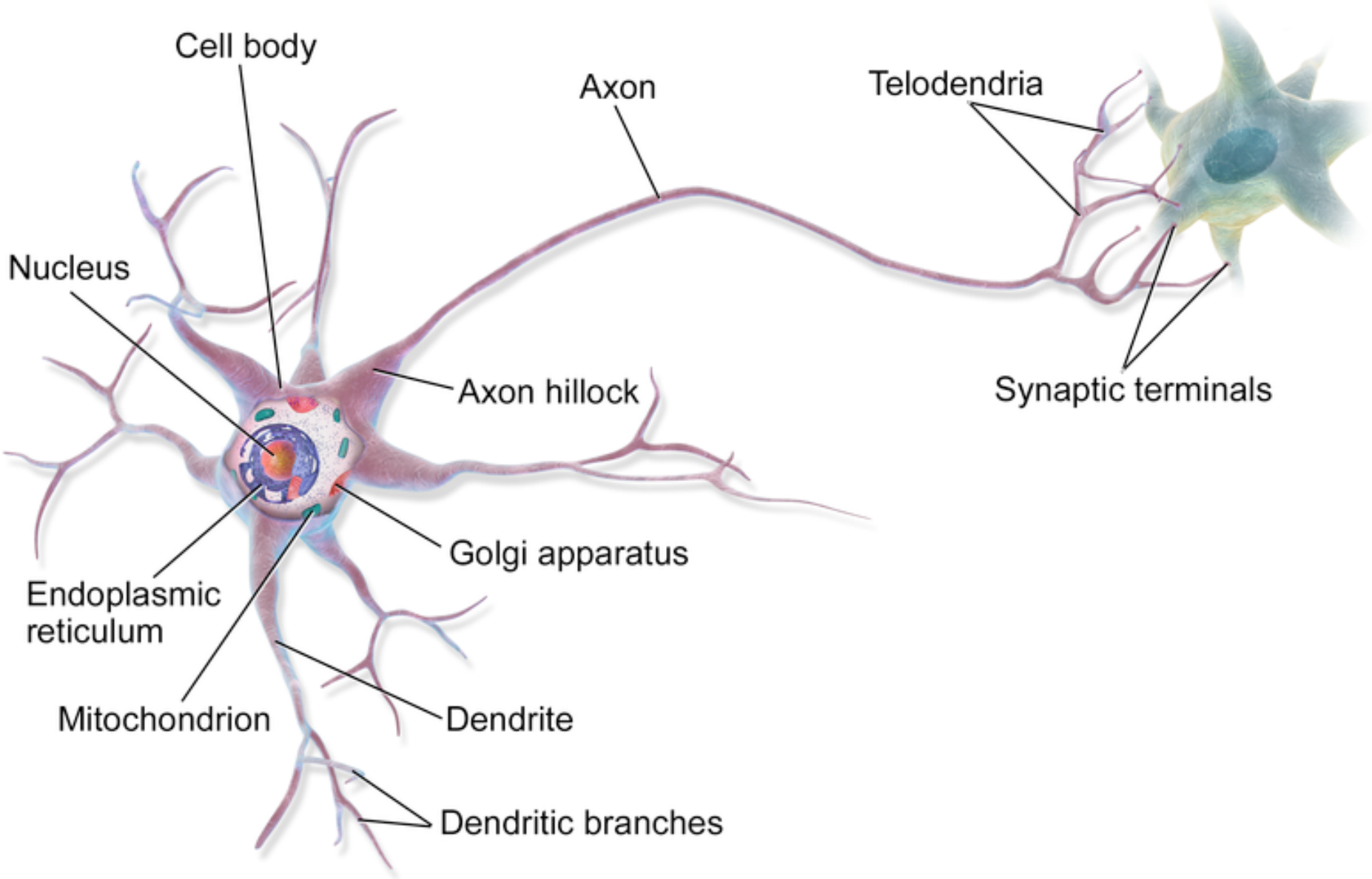
Lida Kanari,[1] Paweł Dłotko,[2] Martina Scolamiero,[3] Ran Levi,[4] Julian Shillcock,[1] Kathryn Hess,[3] and Henry Markram[1]

[1] *Blue Brain Project, EPFL*
[2] *DataShape, INRIA Saclay, Ile-de-France*
[3] *Laboratory for Topology and Neuroscience at the Brain Mind Institute, EPFL*
[4] *Institute of Mathematics, University of Aberdeen*

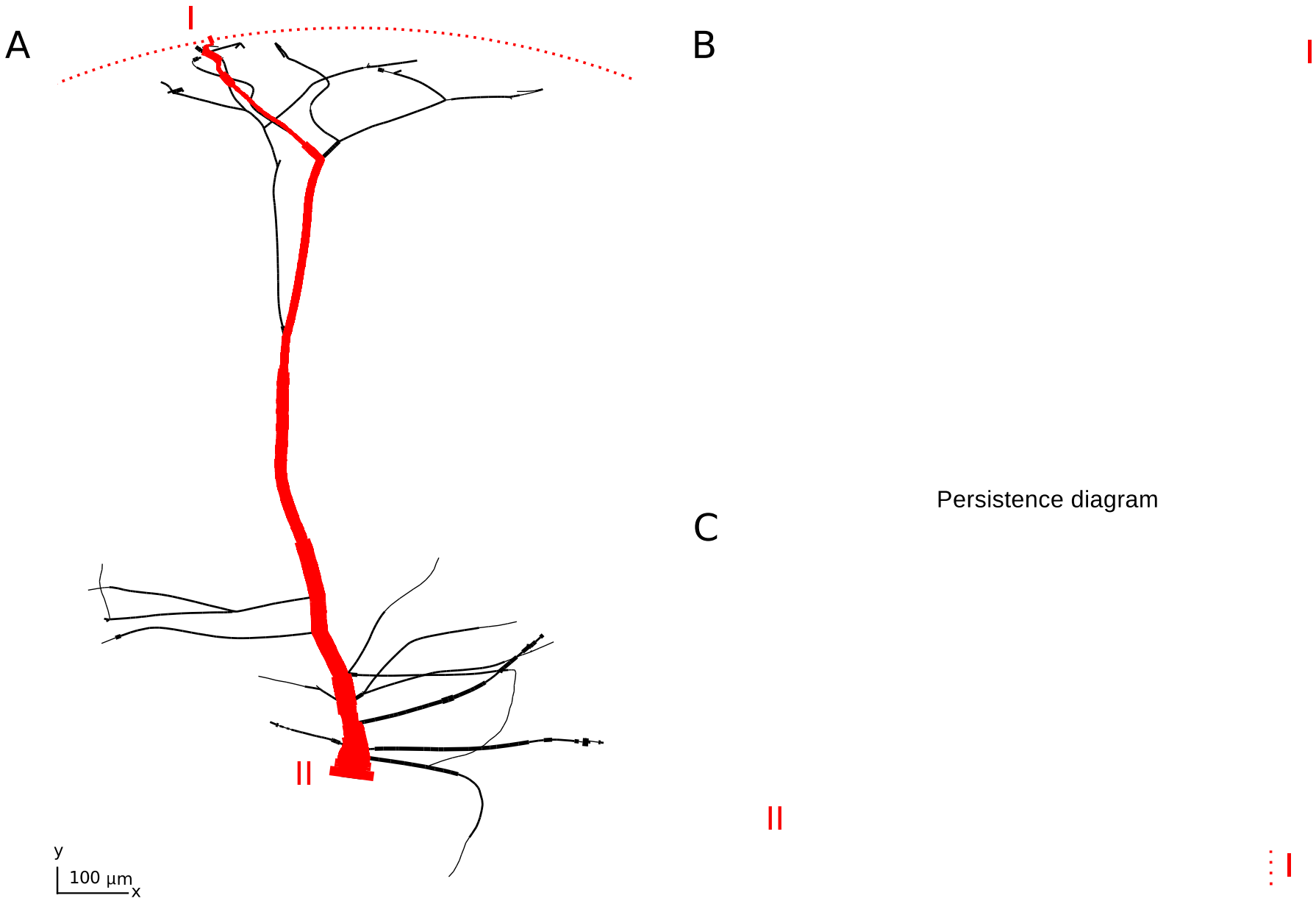
(Dated: March 29, 2016)







https://en.wikipedia.org/wiki/Neuron#/media/File:Blausen_0657_MultipolarNeuron.png


A
I
II
y
100 µm
x
B
I
C
Persistence diagram
II
I

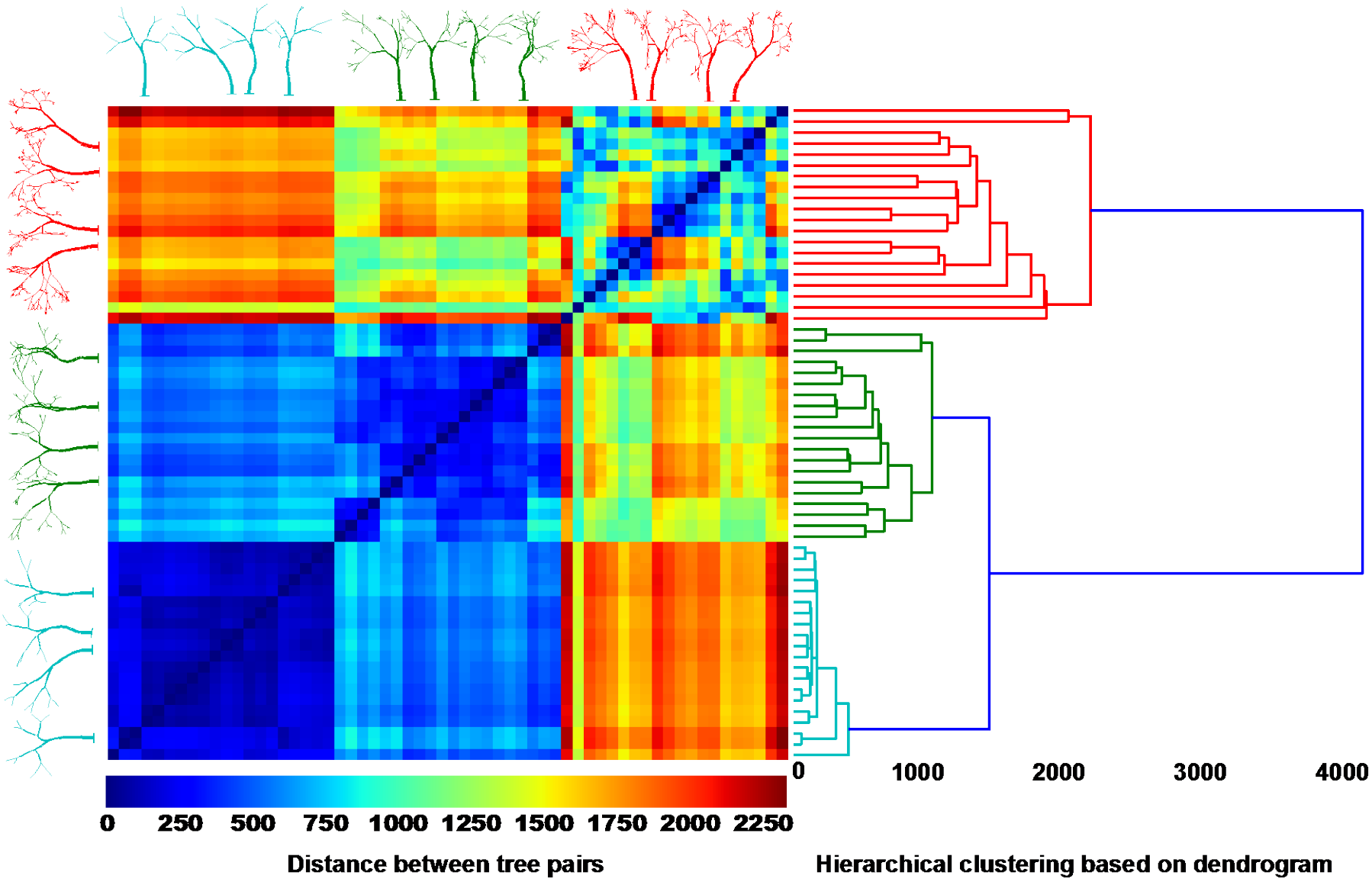
d_{Bar}:

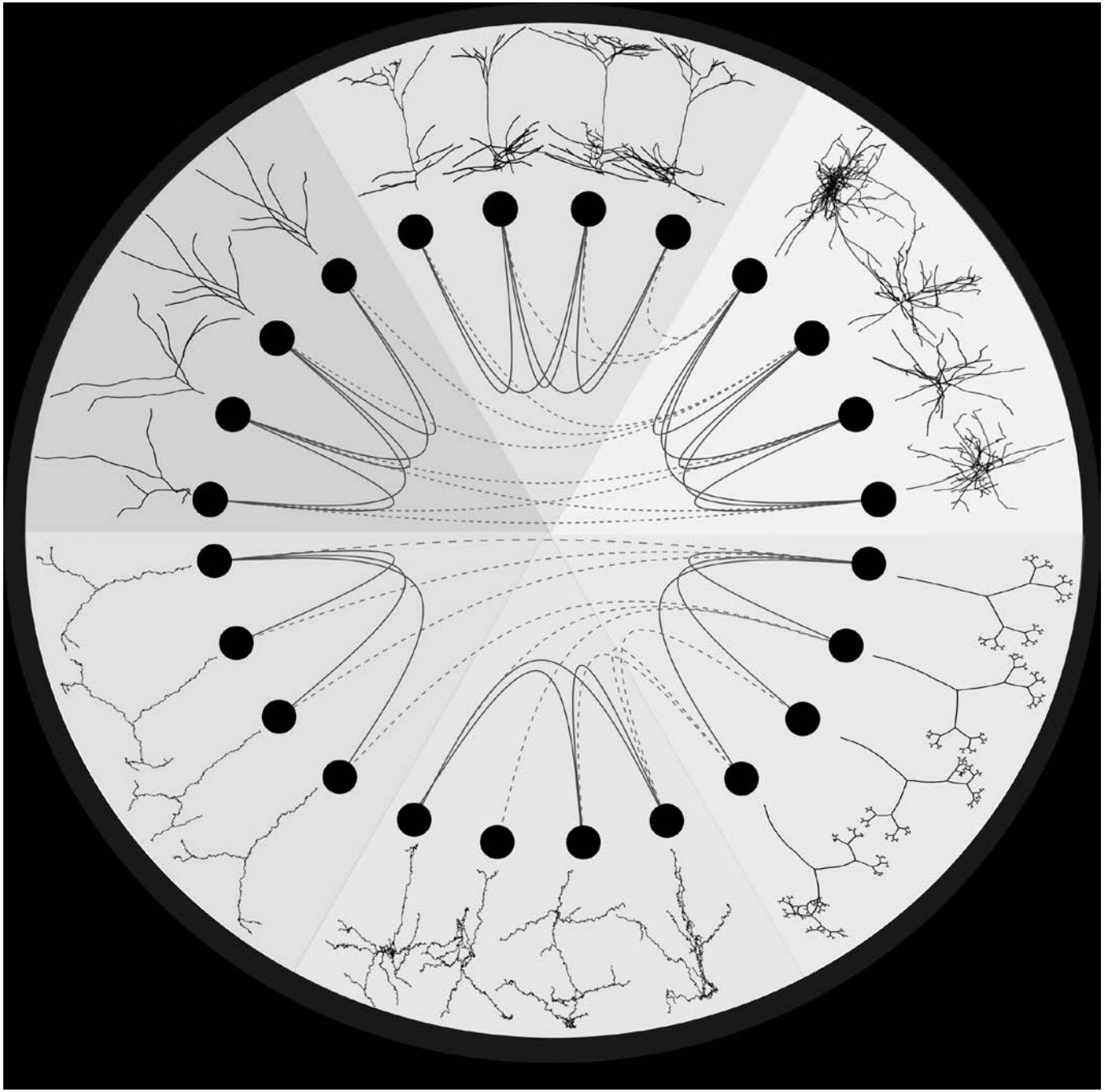
For each barcode we generate a density profile as follows:

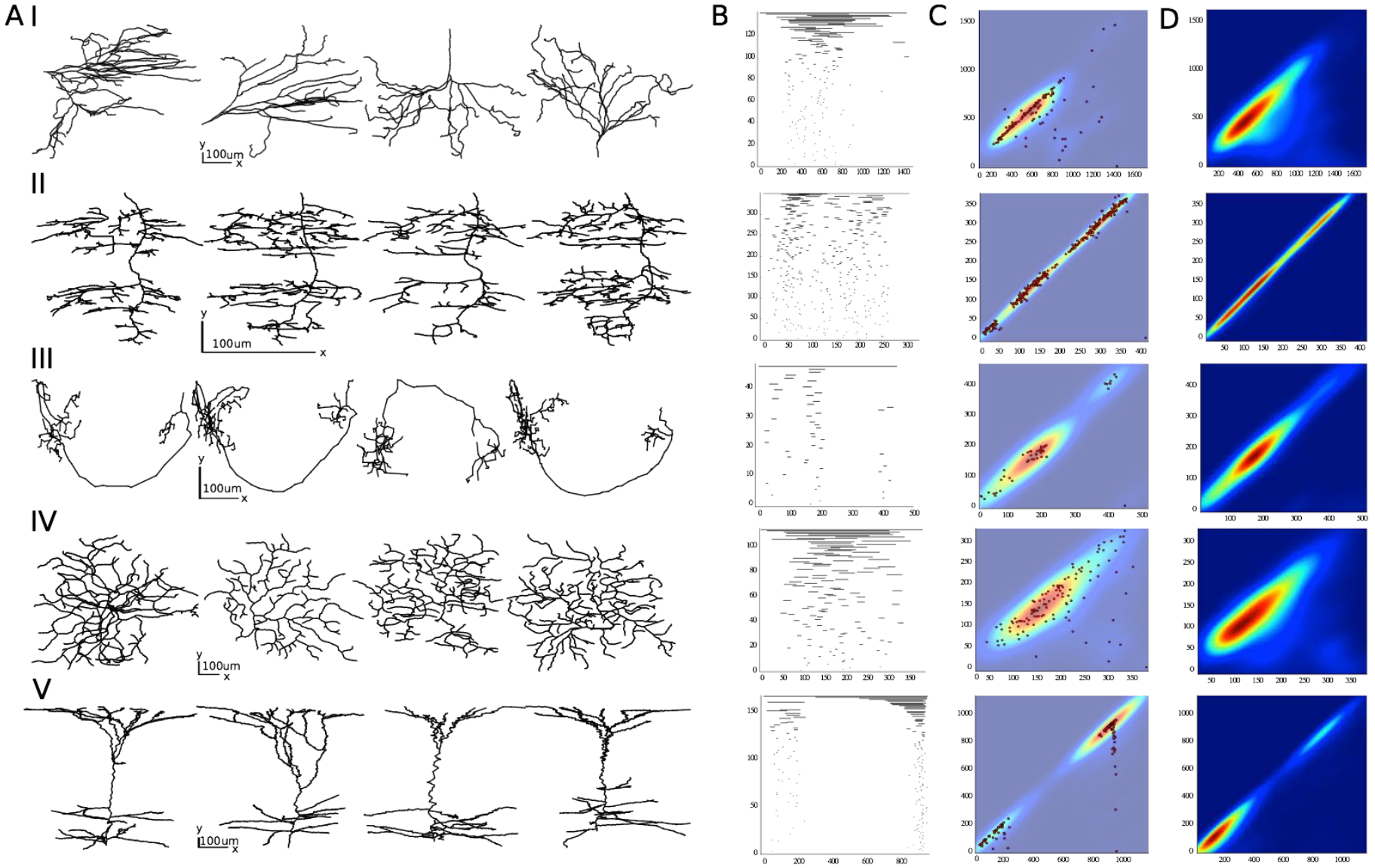
For all x in **R,** the value of the histogram is the number of intervals that contain x , i.e., the number of components alive at that point.

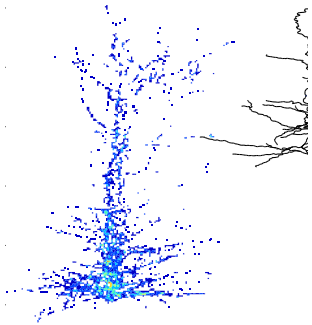
The distance between two barcodes D (T1) and D (T) is defined as the sum of the differences between the density profiles of the barcodes.

This distance is not stable with respect to Hausdorff distance, but it is the only distance we are aware of that succeeds in capturing the differences between distinct neuronal persistence barcodes.


0
250
500
750
1000
1250
1500
1750
2000
2250
Distance between tree pairs
0
1000
2000
3000
4000
Hierarchical clustering based on dendrogram




A I
II
III
IV
V
100um
B
C
D



B

C

Expert
Randomized

Accuracy (fraction)

https://arxiv.org/format/1608.03520

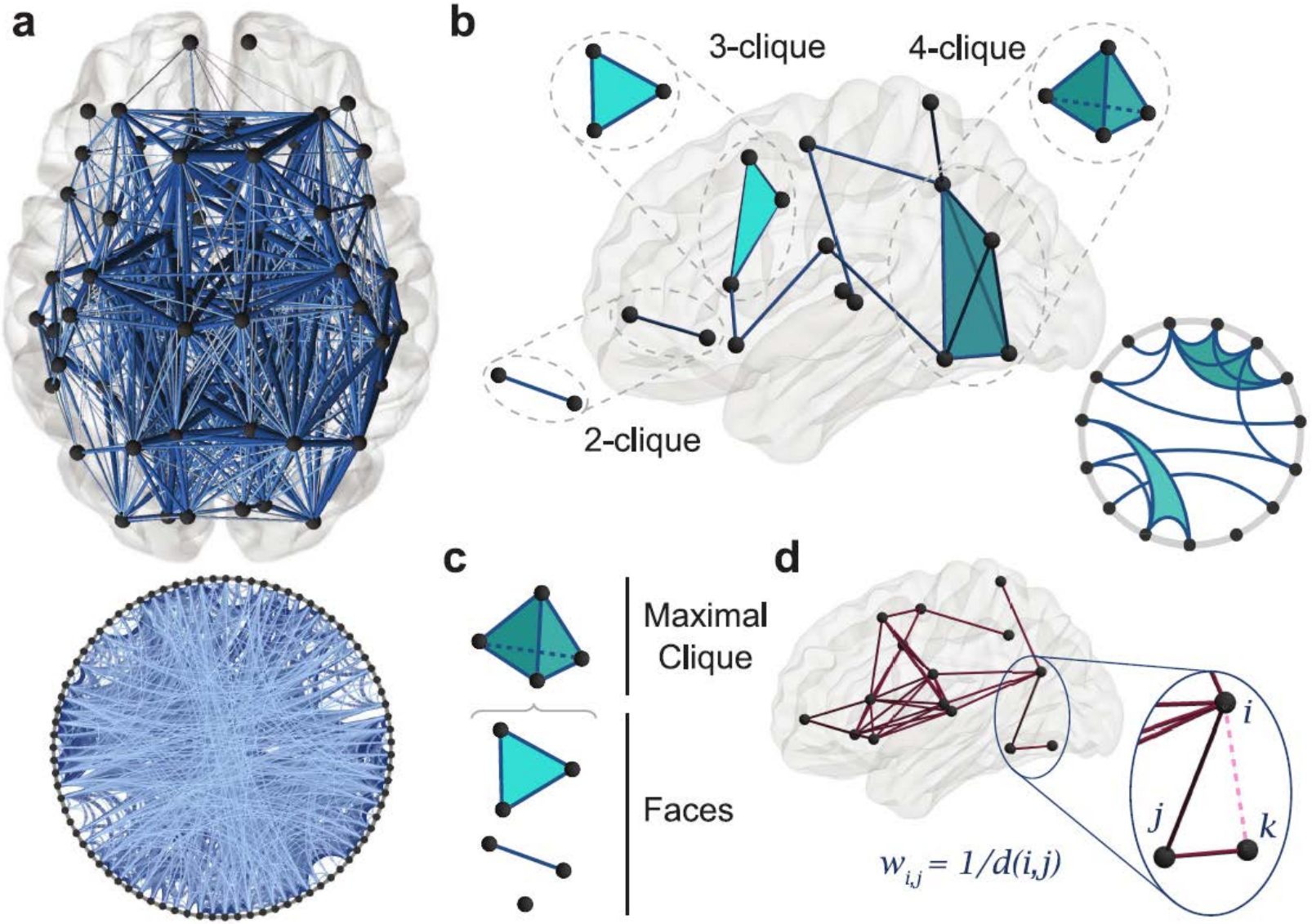
Cliques and Cavities in the Human Connectome

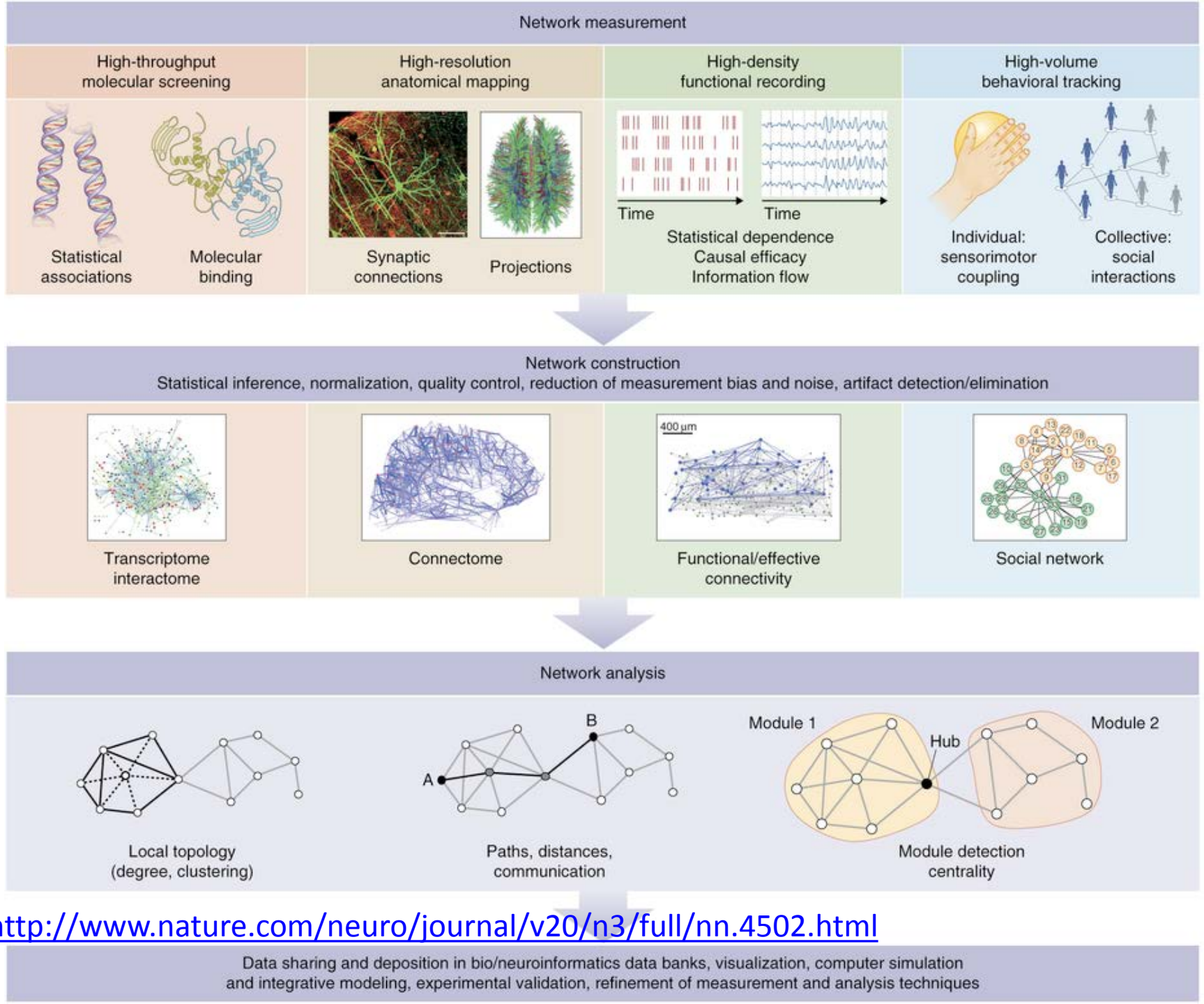
Ann Sizemore[1,2], Chad Giusti[1], Ari Kahn[1], Richard F. Betzel[1], and Danielle S. Bassett[1,3,*]

[1] *Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19041 USA*
[2] *Broad Institute, Harvard University and the Massachusetts Institute of Technology, Cambridge, MA 02142 USA*
[3] *Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA 19041 USA*
[*] *To whom correspondence should be addressed: dsb@seas.upenn.edu*

In this network, nodes correspond to 83 brain regions defined by the Lausanne parcellation [26] and

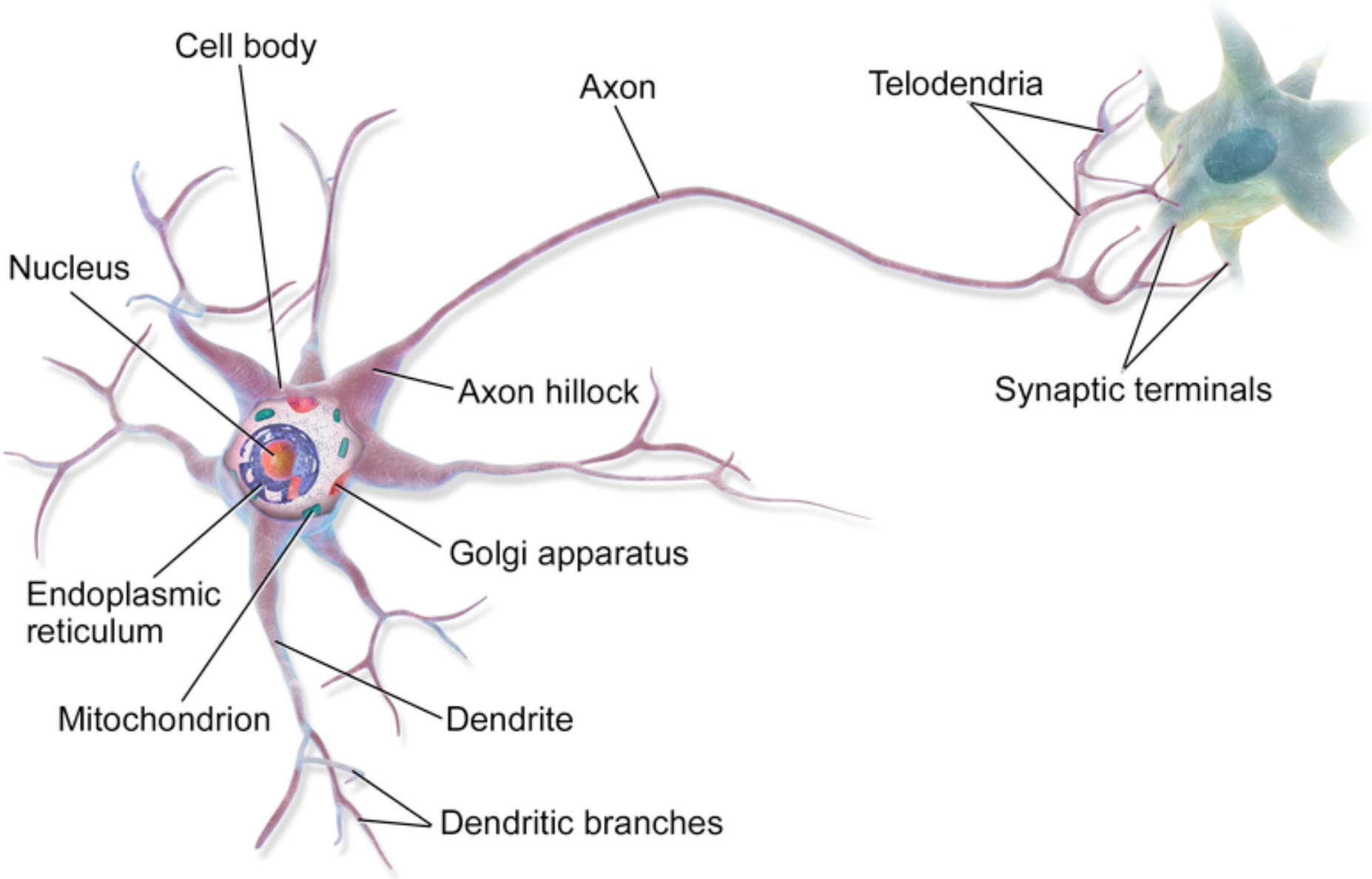
edges correspond to the density of white matter tracts between node pairs


a
b
3-clique
4-clique
2-clique
c
Maximal Clique
Faces
d
i
j
k
$w_{i,j} = 1/d(i,j)$

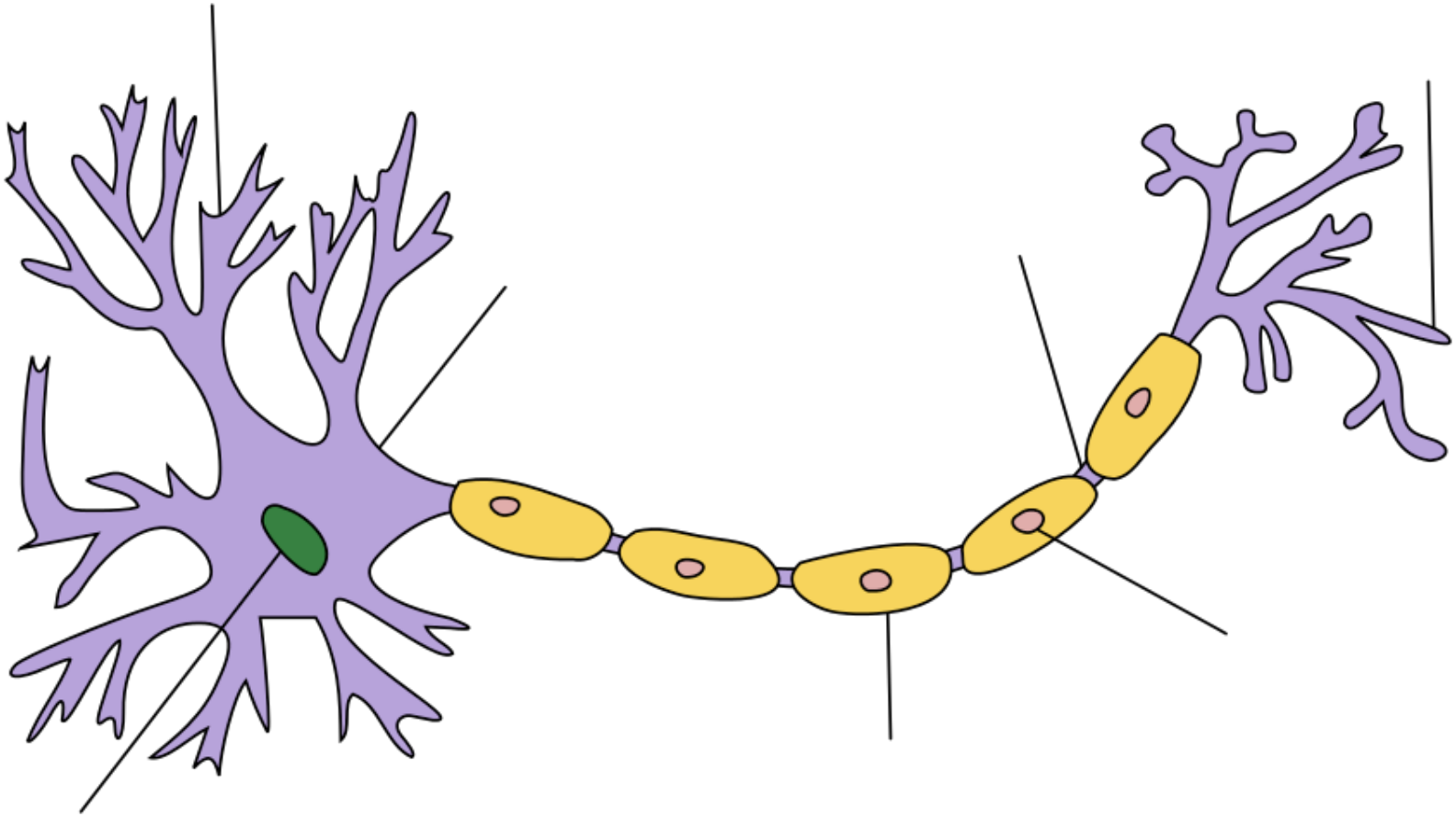


http://www.nature.com/neuro/journal/v20/n3/full/nn.4502.html

Debbie Maizels/Springer Nature

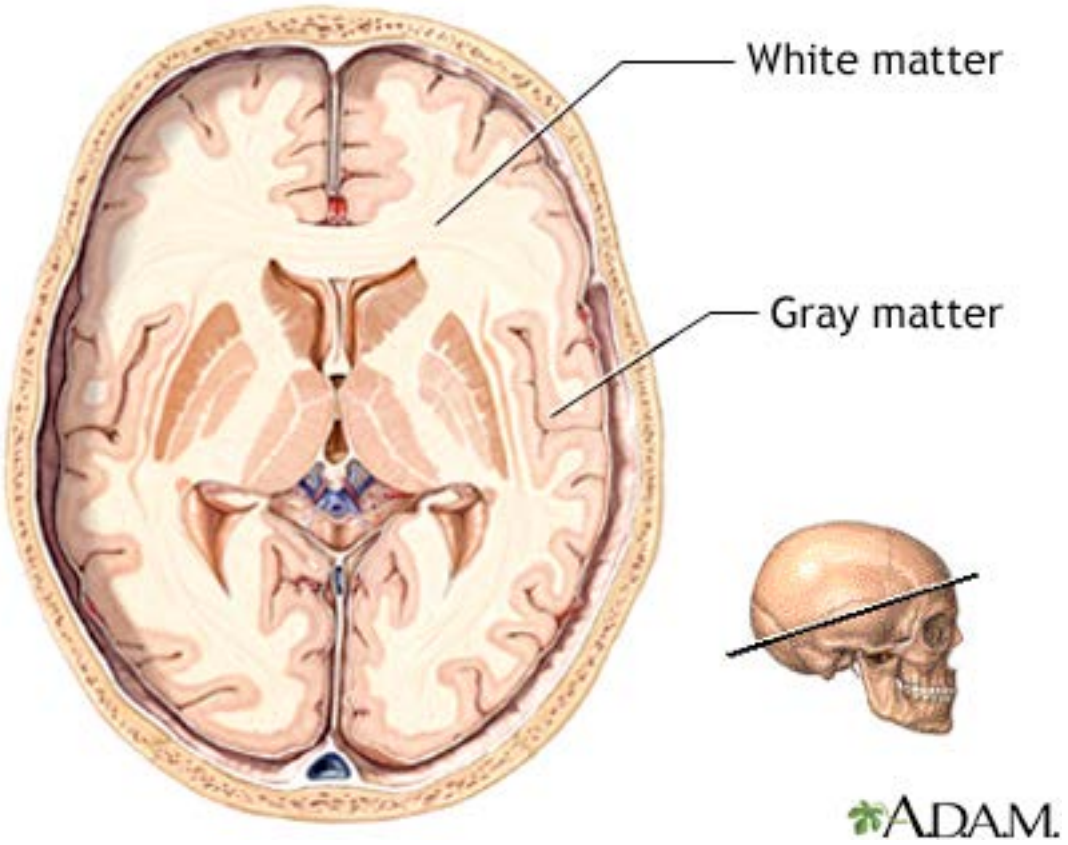


https://en.wikipedia.org/wiki/Neuron#/media/File:Blausen_0657_MultipolarNeuron.png



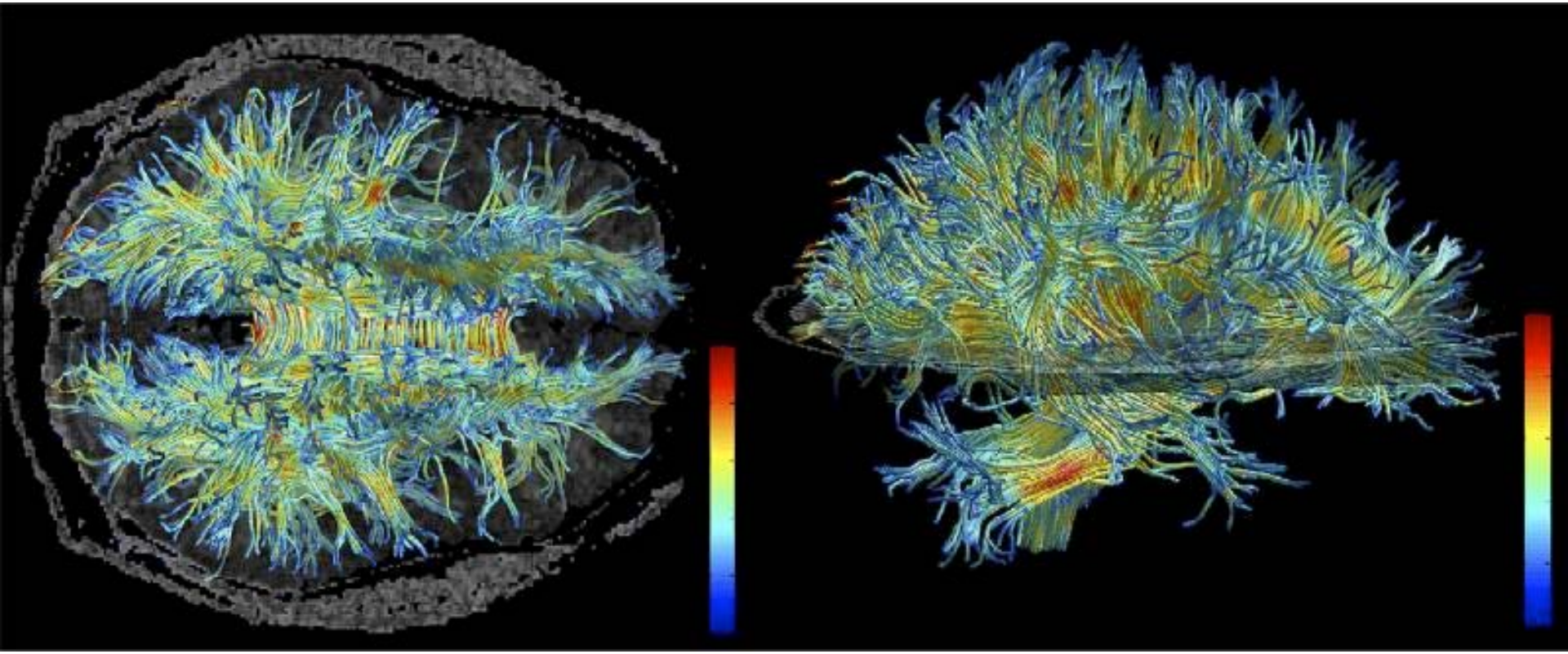
https://en.wikipedia.org/wiki/Axon#/media/File:Neuron_Hand-tuned.svg

https://medlineplus.gov/ency/imagepages/18117.htm



The tissue called "gray matter" in the brain and spinal cord is made up of cell bodies.

"White matter" is composed of nerve fibers (axons).



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2768134/

https://arxiv.org/format/1608.03520

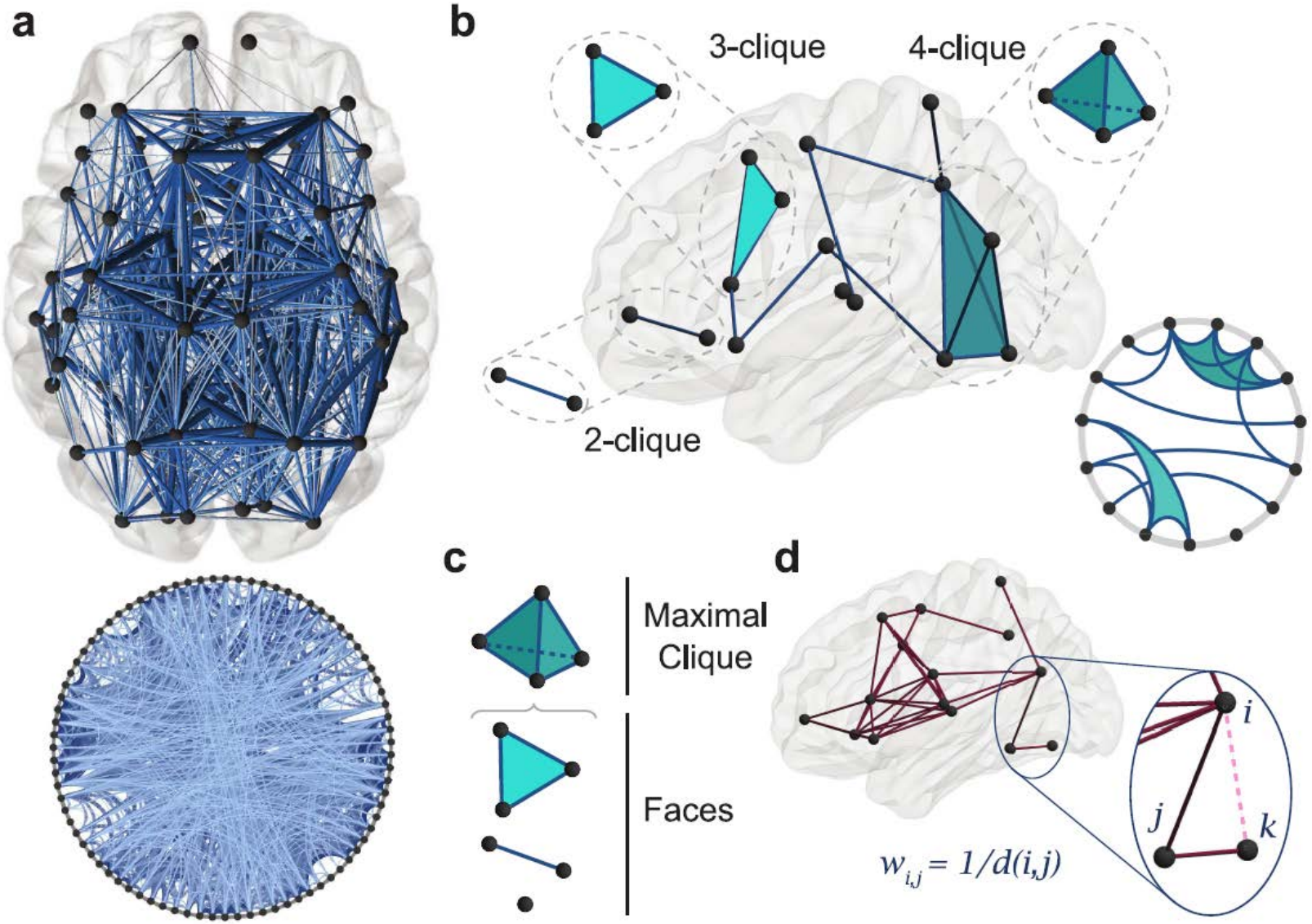
Cliques and Cavities in the Human Connectome

Ann Sizemore[1,2], Chad Giusti[1], Ari Kahn[1], Richard F. Betzel[1], and Danielle S. Bassett[1,3,*]

[1] *Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19041 USA*
[2] *Broad Institute, Harvard University and the Massachusetts Institute of Technology, Cambridge, MA 02142 USA*
[3] *Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA 19041 USA*
[*] *To whom correspondence should be addressed: dsb@seas.upenn.edu*

In this network, nodes correspond to 83 brain regions defined by the Lausanne parcellation [26] and

edges correspond to the density of white matter tracts between node pairs


a
b
3-clique
4-clique
2-clique
c
Maximal Clique
Faces
d
i
j
k
$w_{i,j} = 1/d(i,j)$