Harmonic Data Analysis for Shape and Centrality

Woong Kook
Department of Mathematical Sciences
Seoul National University

XAIENCE 2019
College of Natural Sciences
Seoul National University
November 08,2019

Presentation Overview

- Introduction: overview of Topological Data Analysis (TDA)
- Harmonic cycles and shape detection
- Effective conductance and network centrality
- Applications: mathematical summary of data

Topological Data Analysis (TDA): Main idea

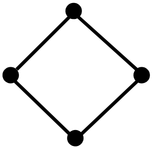
Topology: Mathematics of Shape Recognition

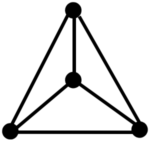
Topological Data Analysis:

Shape of Data

Point Cloud $\Rightarrow$ Topological Network $\Rightarrow$ Hidden Intelligence

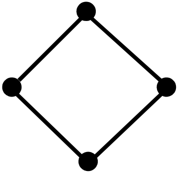
Simplicial Modeling

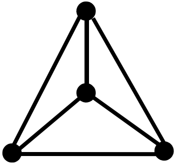




Euler characteristic $\chi(X)$

$$\chi(X) = 1 - |\mathit{vertices}| + |\mathit{edges}|$$



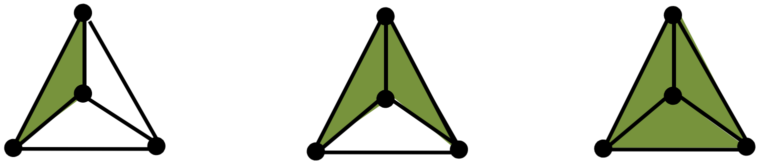


$\chi(R)$=1-1= **0** $\chi(D)$=1-4+4= **1** $\chi(P)$=1-4+6= **3**

$\chi(X)$ = # independent cycles (“holes” or “loops”)

Euler characteristic $\chi(X)$ for simplicial complexes

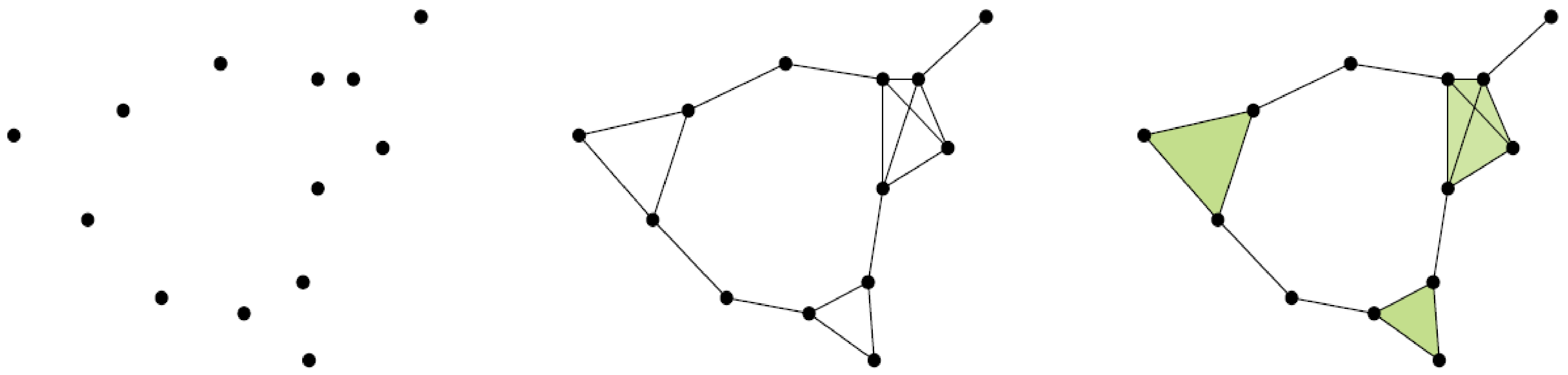
$$\chi(X) = 1 - |\mathit{vertices}| + |\mathit{edges}| - |\mathit{triangles}|$$



$\chi(P_1)$=1-4+6-1= **2** $\chi(P_2)$=1-4+6-2= **1** $\chi(P_3)$=1-4+6-3= **0**

$\chi(X)$ = # independent cycles ("holes" or "loops")

Persistent Homology

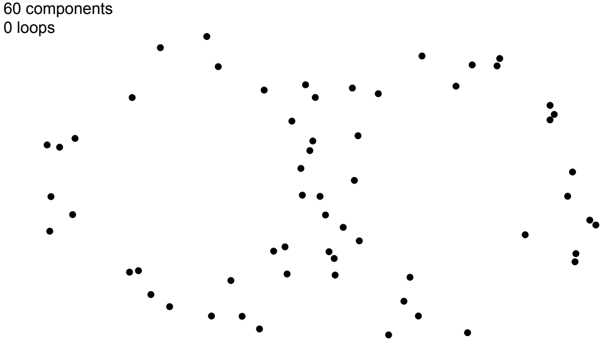


Vietoris-Rips Complex $VR(\varepsilon)$

β_0= # clusters

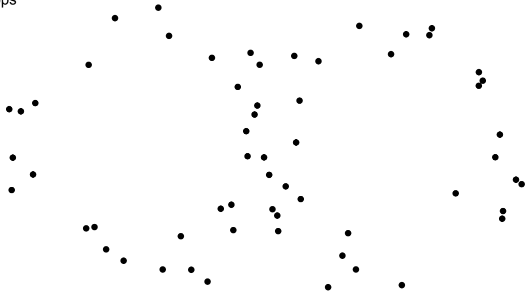
β_1= # cycles

A point cloud with two “loops”

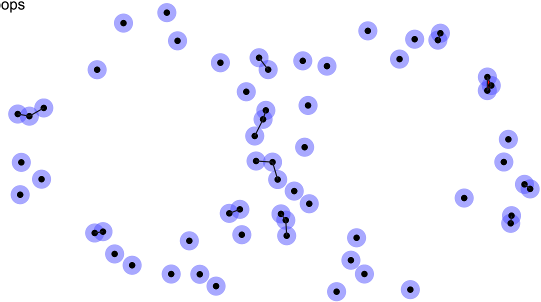


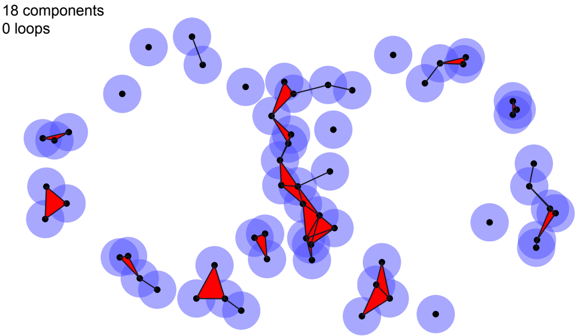
Question: How are the two loops recognized?

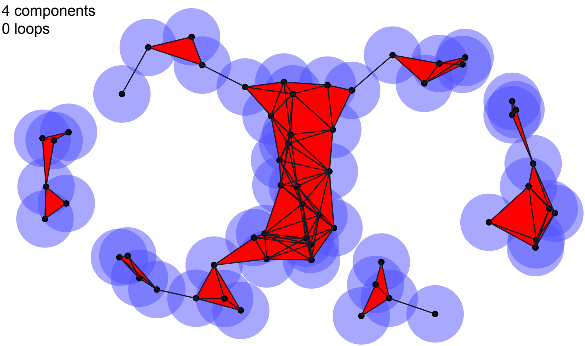
60 components
0 loops

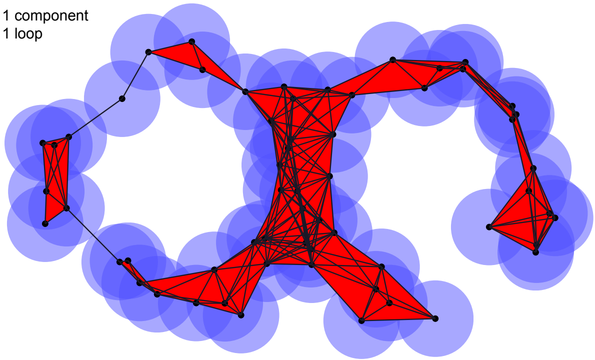


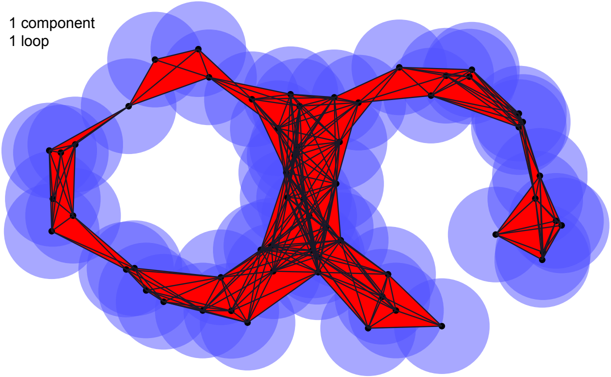
44 components
0 loops

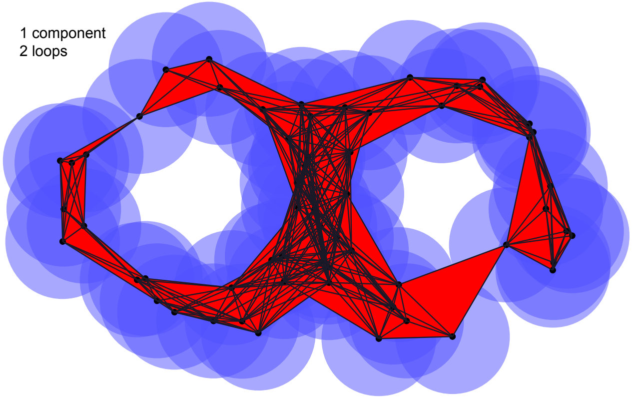


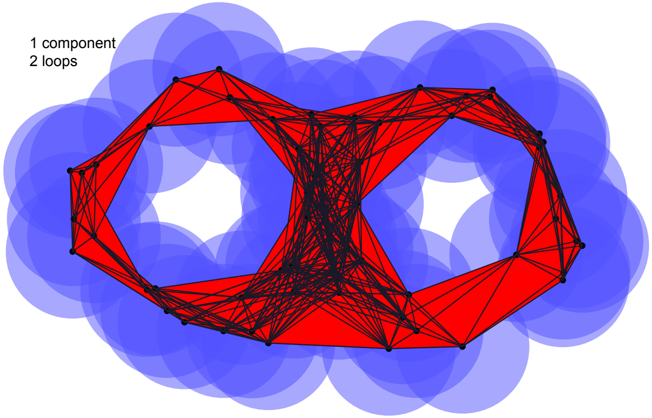

18 components
0 loops


4 components
0 loops

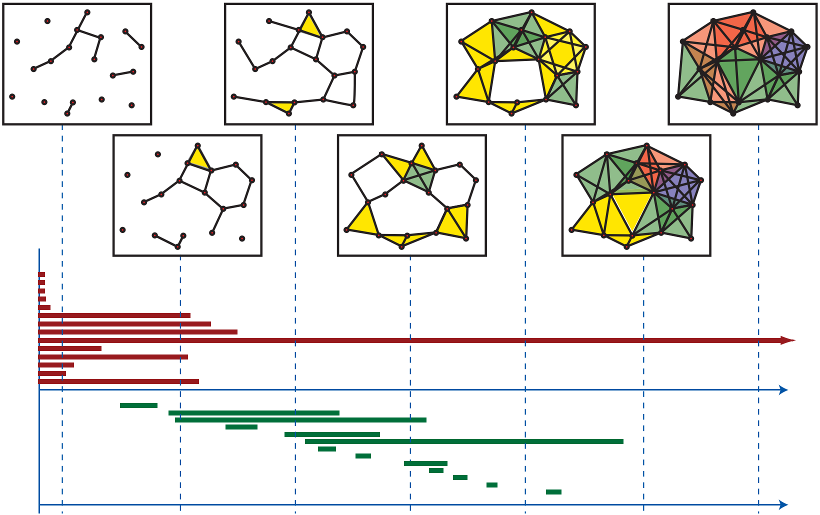

1 component
1 loop


1 component
1 loop

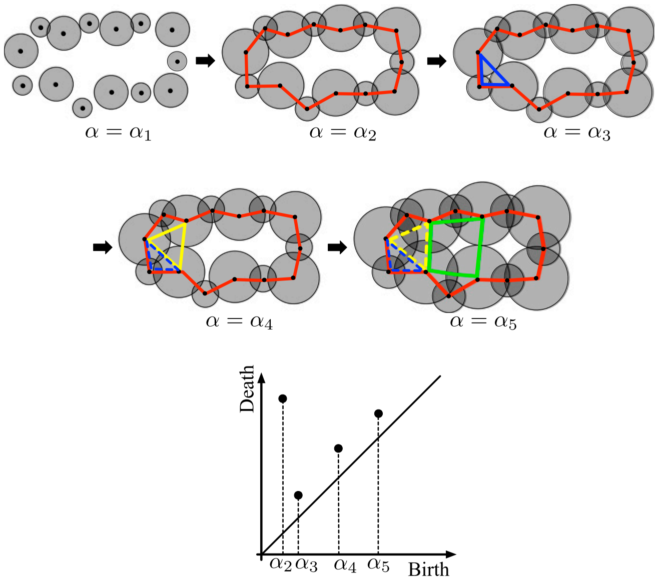

1 component
2 loops


1 component
2 loops

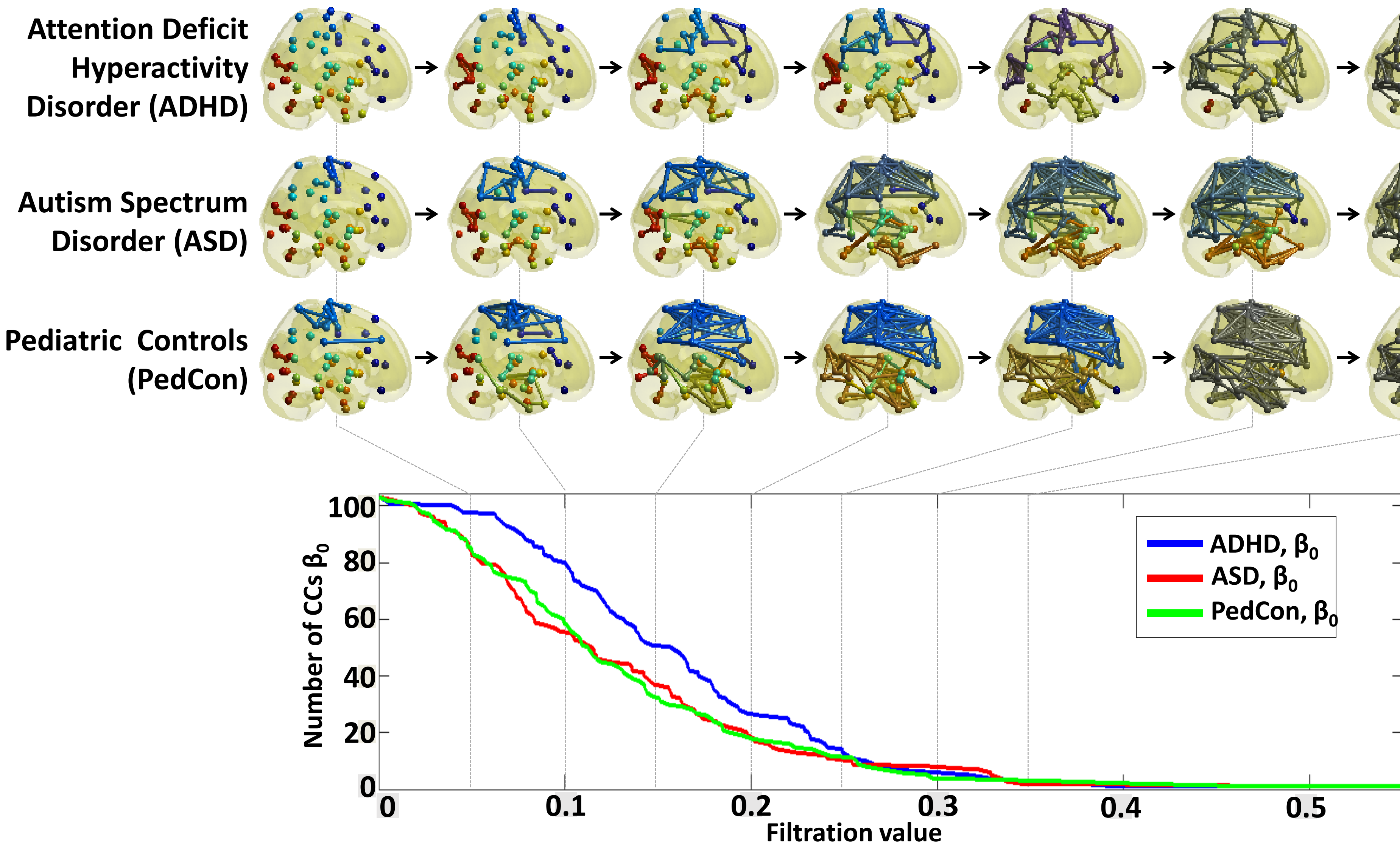
Barcode (Ghrist 2007)



Persistence Diagram (Hiraoka et al. 2016)

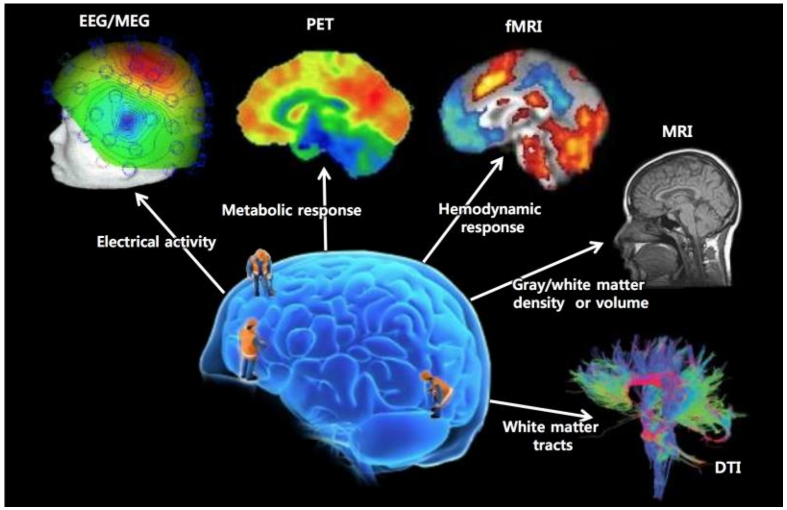


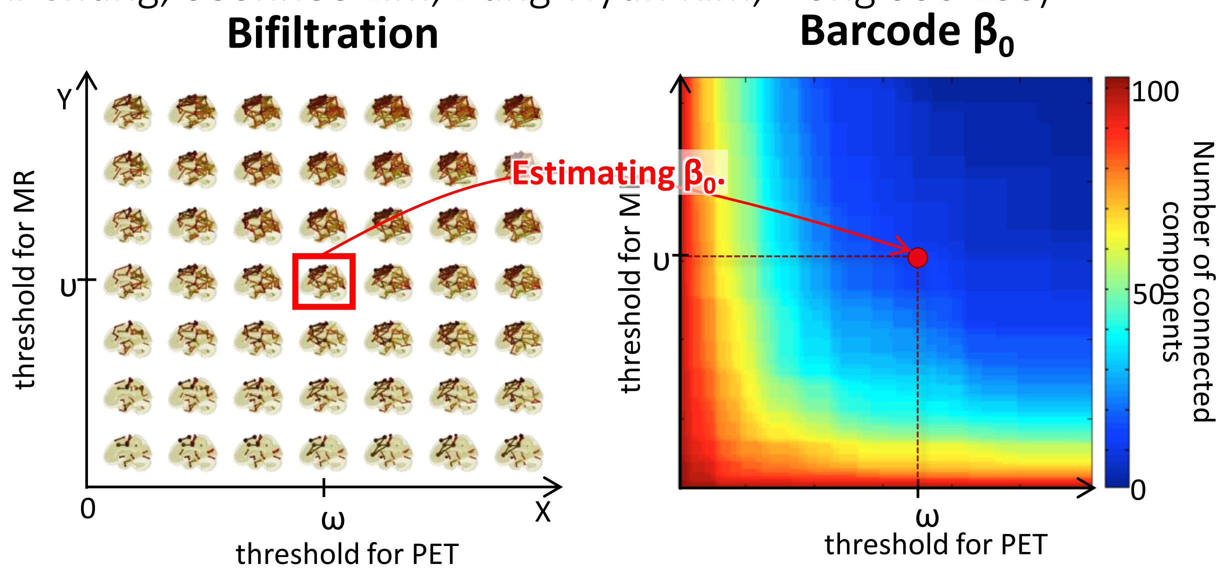
Brain network analysis based on persistent homology

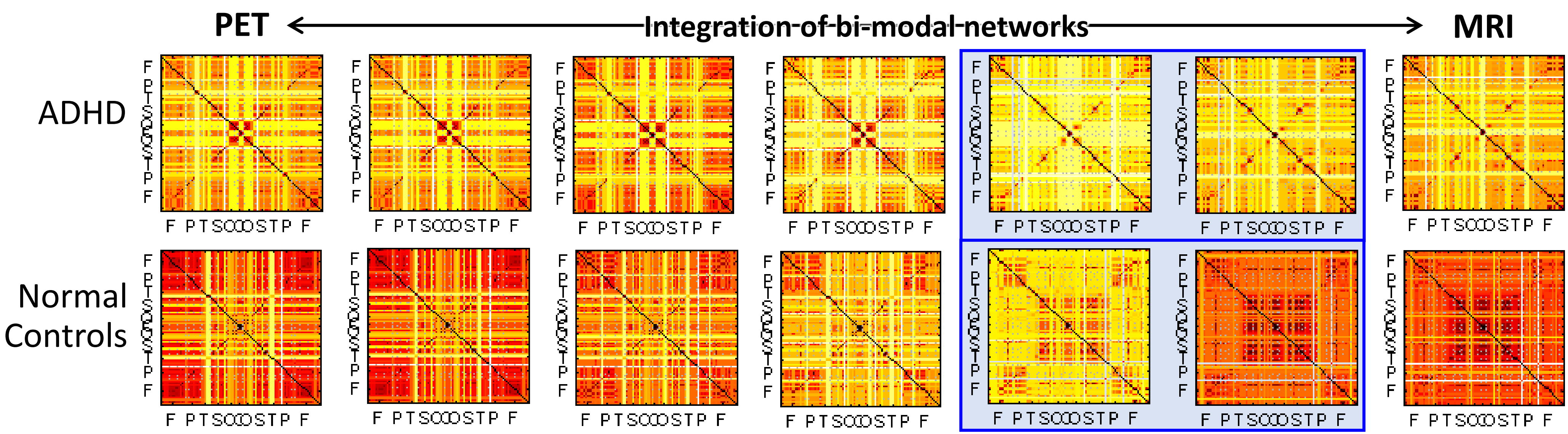


Multimodal brain network analysis based on multidimensional persistence

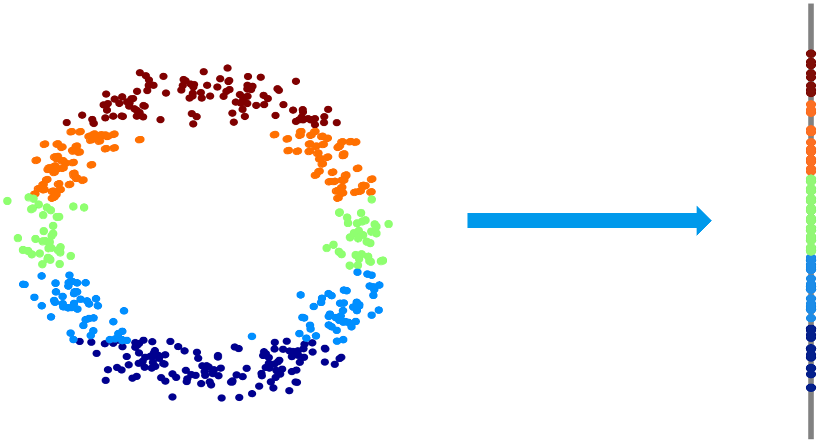
(Hyekyoung Lee, Hyejin Kang, Moo K. Chung, Seonhee Lim, Bung-Nyun Kim, Dong Soo Lee)





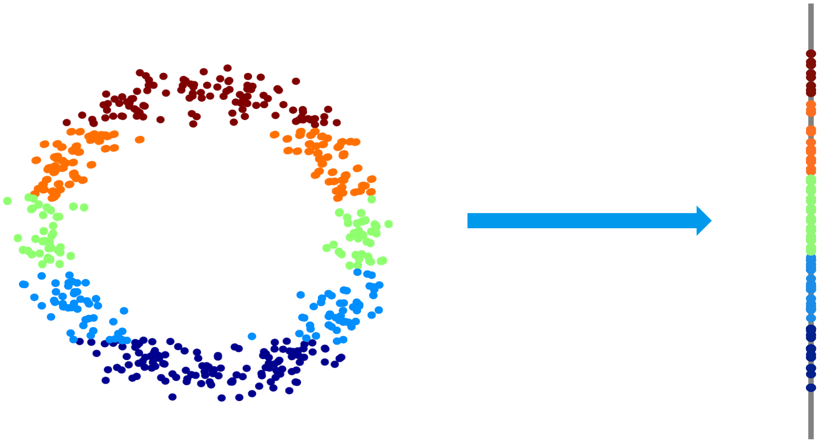


Mapper (Ayasdi 2015)



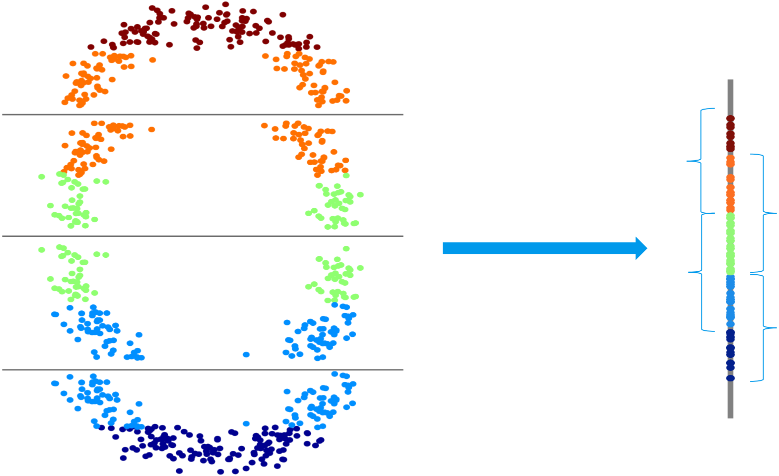
Filter Function : Point Cloud $\to \mathbb{R}$

Mapper (Ayasdi 2015)



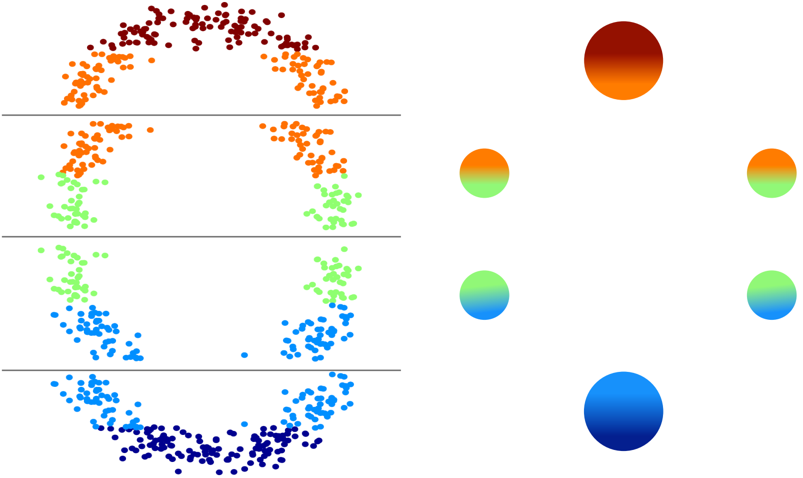
Filter Function : Point Cloud $\to \mathbb{R}$ (y-coordinates)

Mapper (Ayasdi 2015)



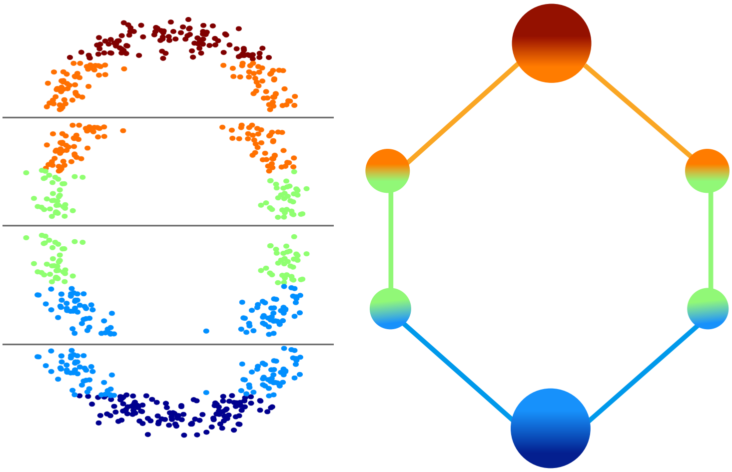
Dividing data points into *overlapping* level sets

Mapper (Ayasdi 2015)



Clustering within each level set

Mapper (Ayasdi 2015)



Topological Network

Examples of metrics and lenses (filter functions)

- Metric

  Cosine

  Correlation.

  Variance (or IQR) Normalized Euclidean

  Hamming

- Lense (filter function)

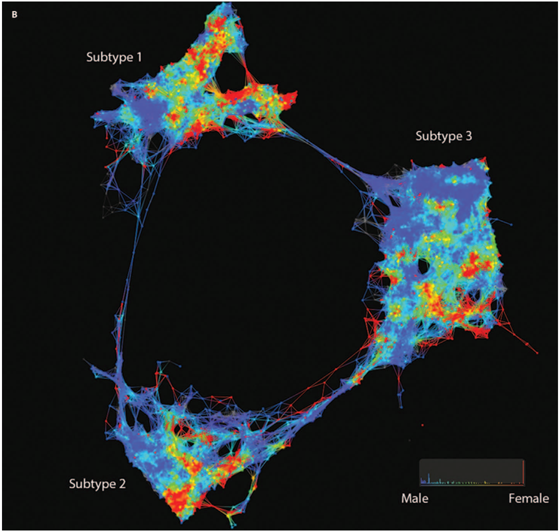
  L1, L-infinity centrality

  Gaussian density

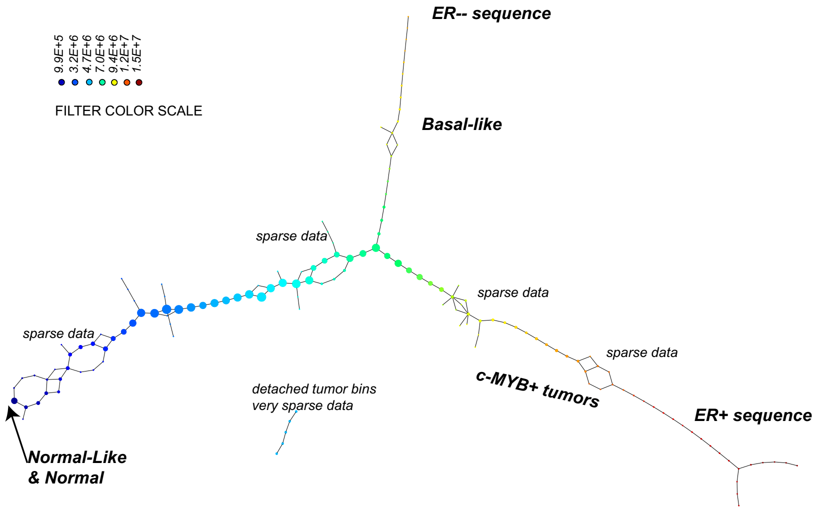
  k-nearest neighborhood (distance matrix via Dijkstra's algorithm)

  MDS coordinate (uses distance matrix)

Identification of type 2 diabetes subgroups through topological analysis of patient similarity (2015)

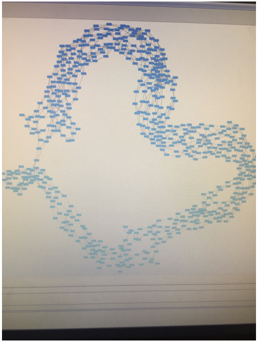


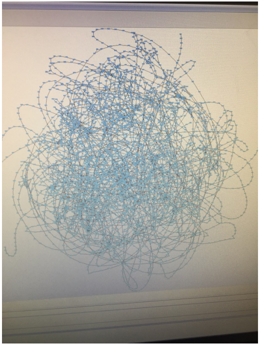
Nicolau, Levine, and Carlsson (2011)

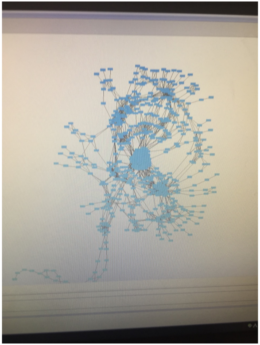


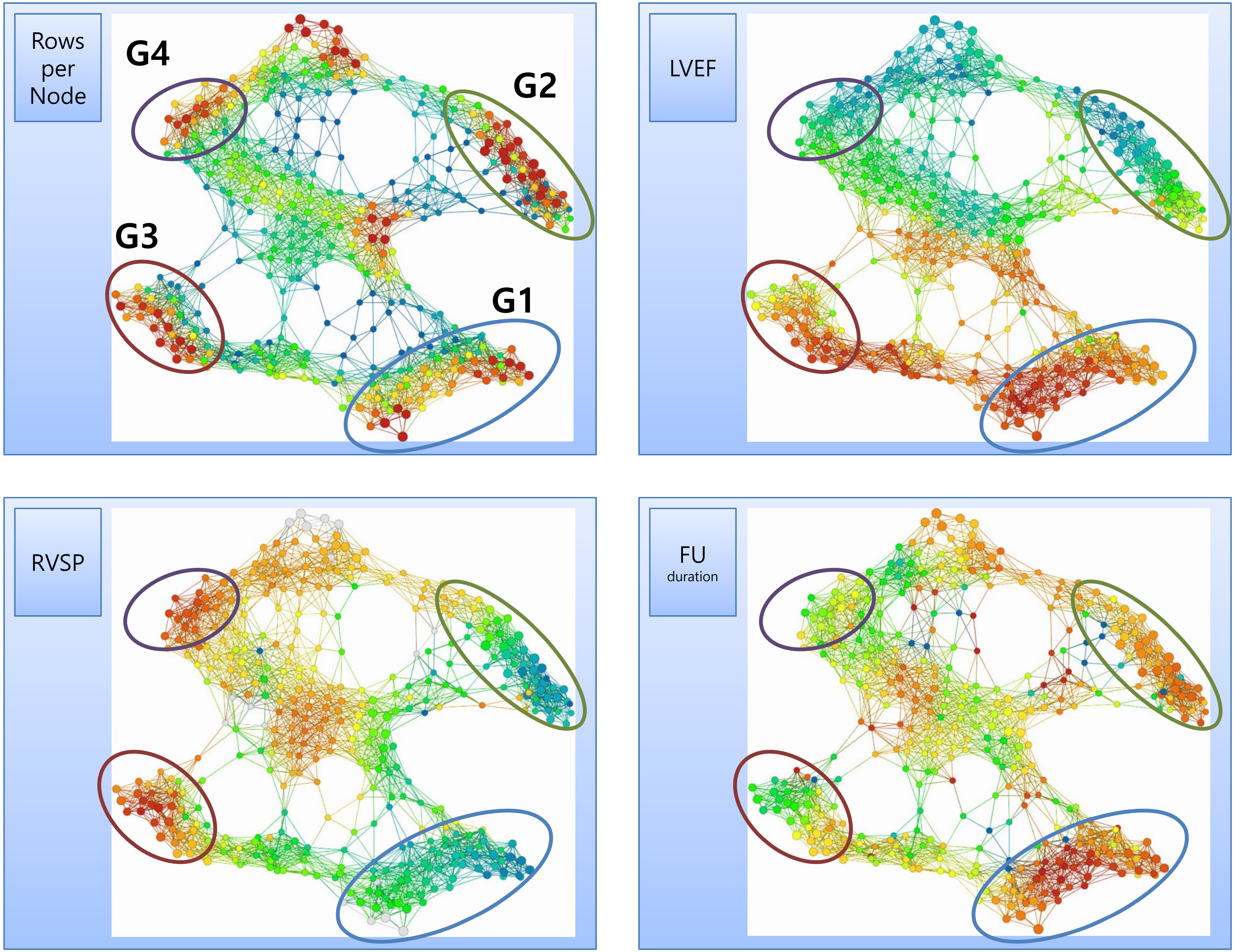
f: L^p norm on DSGA-transformed matrix d: correlation

Personal Account Data








Rows per Node
G4
G2
G3
G1
LVEF
RVSP
FU duration

		Standardization					Average					
		G_HE_LR	G_LE_LR	G_HE_HR	G_LE_HR	AbsMAX	G_HE_LR	G_LE_LR	G_HE_HR	G_LE_HR	Total Mean	Total STDEV
52	RVSP	-1.166	-1.208	1.786	1.433	1.786	26.36	25.73	70.69	65.40	43.87	15.02
156	LVEF_HF_Type	1.582	-0.756	1.582	-0.756	1.582	3.00	1.00	3.00	1.00	1.65	0.86
50	LVEF_combine_copy	1.480	-1.220	1.537	-1.221	1.537	60.85	18.61	61.73	18.60	37.69	15.64
49	EDD-ESD	1.147	-0.894	1.253	-0.955	1.253	17.86	7.83	18.37	7.53	12.22	4.91
48	LVESD	-1.188	0.832	-1.132	1.178	1.188	30.55	55.33	31.24	59.58	45.13	12.27
47	LVEDD	-0.896	0.577	-0.770	0.982	0.982	48.34	63.16	49.61	67.23	57.35	10.05
4	LV_dysf	0.979	-0.280	0.594	-0.309	0.979	1.41	1.02	1.29	1.01	1.11	0.31
54	Etiol_HF_val01	0.209	-0.238	0.797	-0.268	0.797	0.22	0.06	0.42	0.05	0.14	0.35
16	prevValveDis	-0.171	0.221	-0.727	0.129	0.727	1.80	1.93	1.60	1.90	1.86	0.35
64	Etiol_HF_others01	0.439	-0.193	0.679	-0.134	0.679	0.12	0.00	0.16	0.01	0.04	0.19
46	NP_decile	-0.675	0.029	-0.198	0.490	0.675	3.57	5.59	4.94	6.91	5.51	2.87
53	Etiol_HF_IHD01	-0.228	-0.024	-0.633	0.082	0.633	0.27	0.36	0.07	0.42	0.38	0.48
51	LA_dia	-0.226	-0.289	0.626	0.338	0.626	45.96	45.33	54.31	51.48	48.17	9.80
70	HF_Agg_PulmEmboli	0.074	-0.105	0.621	-0.052	0.621	0.02	0.00	0.08	0.01	0.01	0.10
82	Med_ACEI_durAdm	0.243	-0.559	0.397	-0.245	0.559	1.73	1.34	1.81	1.50	1.62	0.49
114	Med_IV_dobuta	-0.351	0.200	-0.102	0.526	0.526	0.08	0.31	0.18	0.45	0.23	0.42
15	DCMP	0.259	-0.269	0.220	-0.524	0.524	1.99	1.84	1.98	1.77	1.92	0.28
36	EKG_HR	-0.435	0.351	-0.504	0.029	0.504	83.82	105.61	81.91	96.66	95.87	27.72
1	SEX	-0.163	0.274	-0.498	0.358	0.498	0.45	0.67	0.28	0.71	0.53	0.50
83	Med_ACEI_dis	0.287	-0.487	0.321	-0.244	0.487	1.85	1.50	1.86	1.61	1.72	0.45
89	Med_BB_dis	0.035	-0.234	0.477	0.156	0.477	1.52	1.38	1.74	1.58	1.50	0.50
35	EKG_QTc	-0.458	0.208	-0.291	0.321	0.458	453.71	484.50	461.45	489.68	474.88	46.19
141	discharge_Hb_max	-0.073	0.457	-0.250	0.183	0.457	12.98	14.08	12.61	13.51	13.13	2.08
88	Med_BB_durAdm	0.057	-0.198	0.456	0.174	0.456	1.45	1.32	1.65	1.51	1.42	0.49
122	IVmed_inotropics	-0.259	0.086	-0.123	0.449	0.449	0.19	0.35	0.25	0.52	0.31	0.46
55	Etiol_HF_cong01	-0.047	-0.106	0.448	-0.053	0.448	0.01	0.00	0.06	0.01	0.01	0.10
65	HF_Agg_ACS	-0.008	0.005	-0.439	-0.162	0.439	0.26	0.26	0.07	0.19	0.26	0.44
56	Etiol_HF_CMP01	-0.326	0.425	-0.409	0.396	0.425	0.07	0.38	0.04	0.37	0.21	0.40
100	Med_Warfarin	0.095	-0.021	-0.422	0.143	0.422	2.48	2.38	2.02	2.52	2.40	0.90
163	FU duration	0.327	0.061	-0.144	-0.297	0.327	879.63	764.71	676.07	610.07	738.38	431.85
103	Med_Ivabradine	0.027	0.027	0.027	0.027	0.027	3.00	3.00	3.00	3.00	3.00	0.05

HE : EF > 52 LE : EF < 25

HR : RVSP > 56 LR : RVSP < 32

Figure 1A

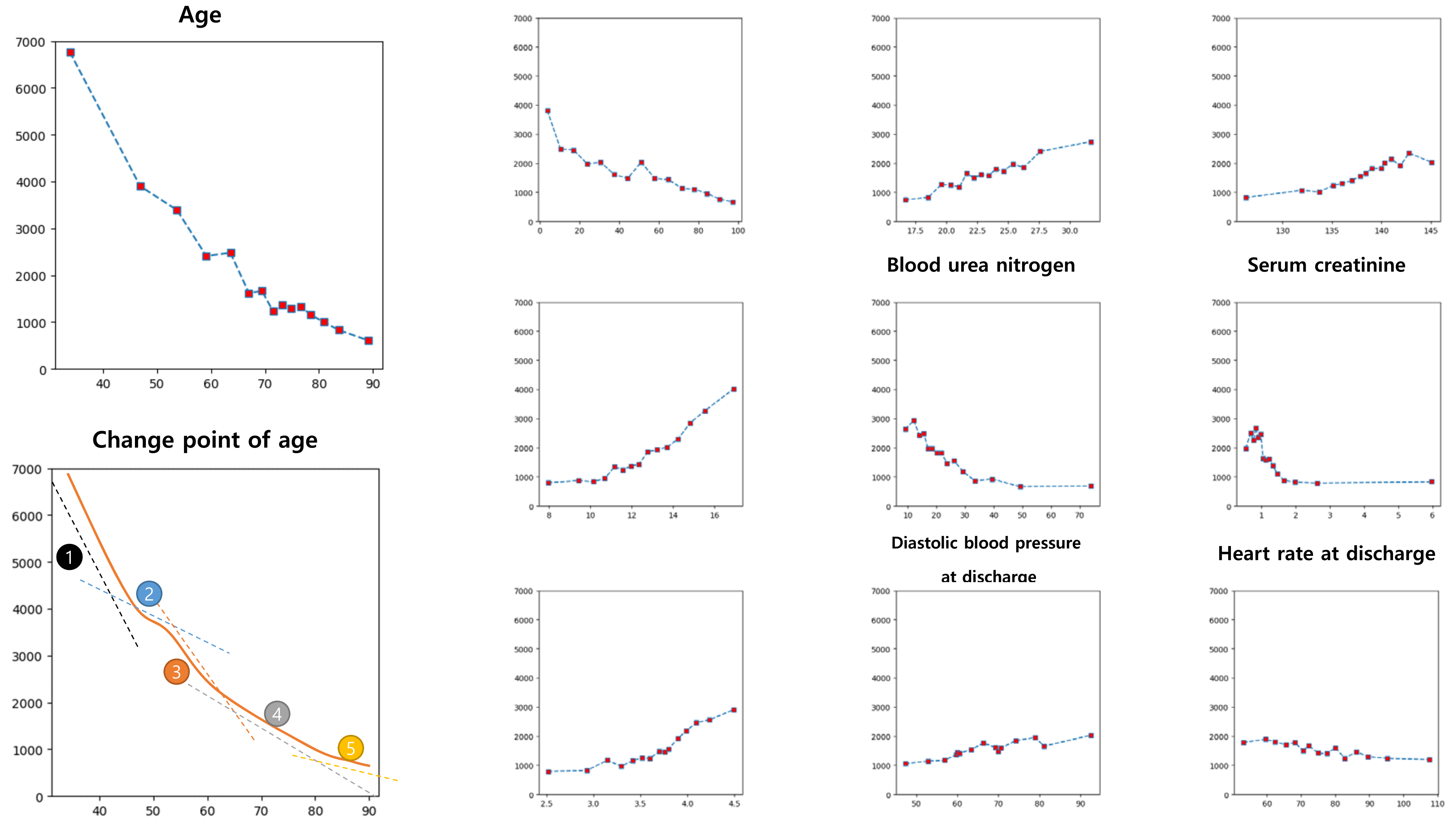
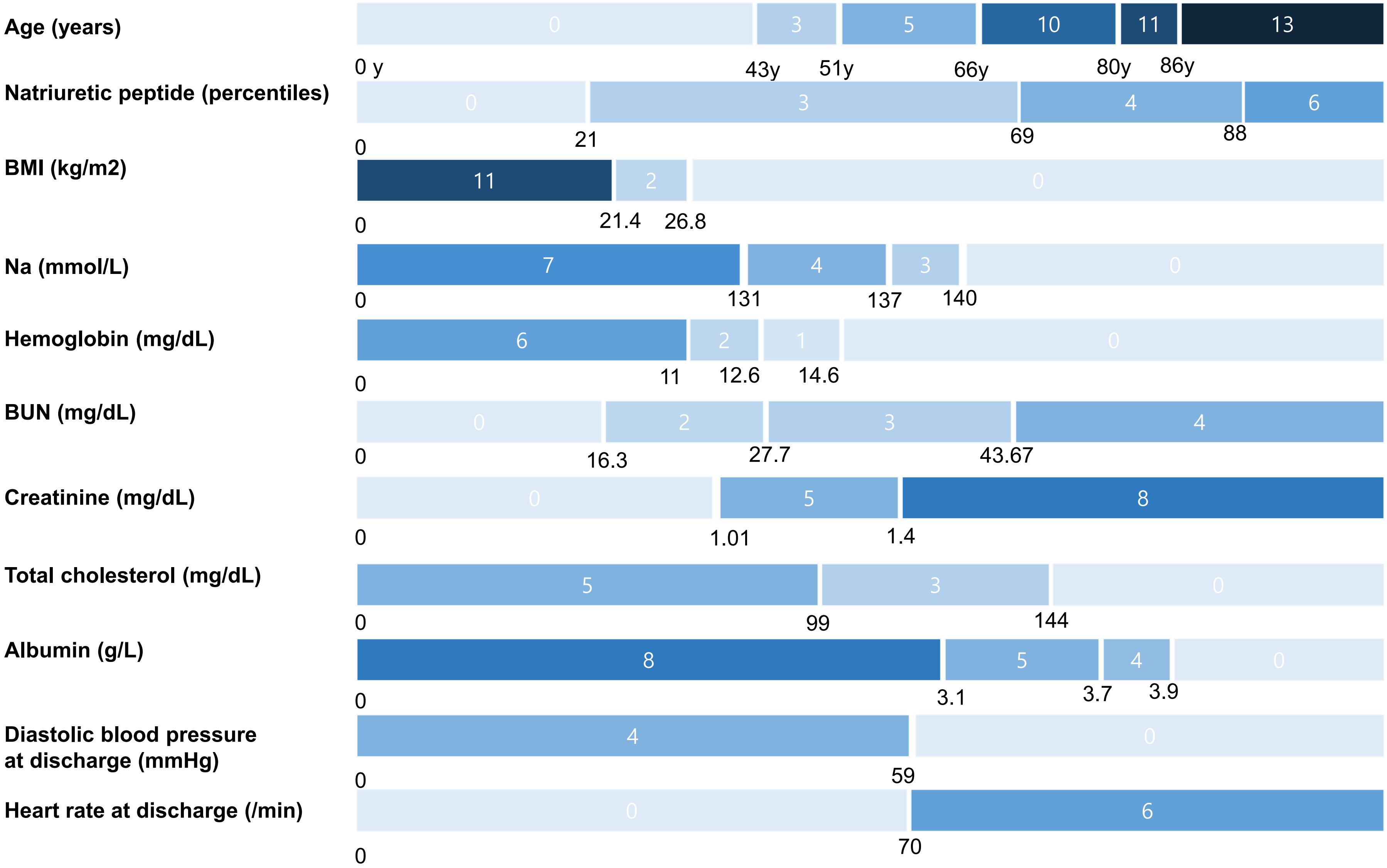

Age
Change point of age
1
2
3
4
5
Natriuretic peptide
BMI
Serum sodium
Blood urea nitrogen
Serum creatinine
Diastolic blood pressure
at discharge
Heart rate at discharge

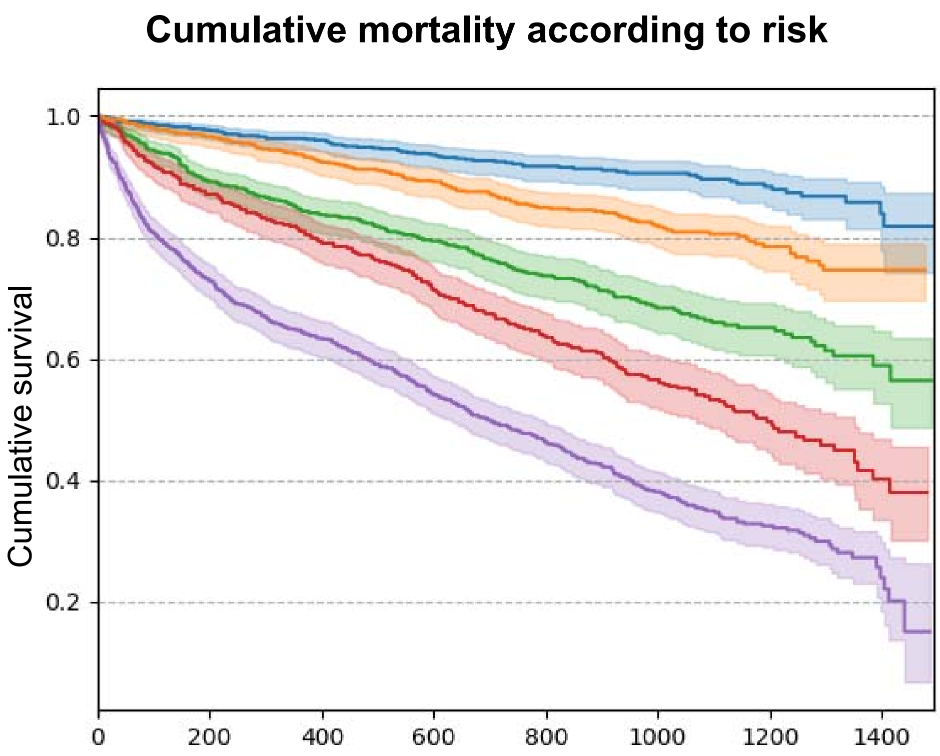
Figure 1B



Age (years)
0
3
5
10
11
13
0 y
43y
51y
66y
80y
86y
Natriuretic peptide (percentiles)
0
3
4
6
0
21
69
88
BMI (kg/m2)
11
2
0
0
21.4
26.8
Na (mmol/L)
7
4
3
0
0
131
137
140
Hemoglobin (mg/dL)
6
2
1
0
0
11
12.6
14.6
BUN (mg/dL)
0
2
3
4
0
16.3
27.7
43.67
Creatinine (mg/dL)
0
5
8
0
1.01
1.4
Total cholesterol (mg/dL)
5
3
0
0
99
144
Albumin (g/L)
8
5
4
0
0
3.1
3.7
3.9
Diastolic blood pressure at discharge (mmHg)
4
0
0
59
Heart rate at discharge (/min)
0
6
0
70

Figure 3

(A)


Cumulative mortality according to risk
Cumulative survival
1.0
0.8
0.6
0.4
0.2
0
200
400
600
800
1000
1200
1400

(B)

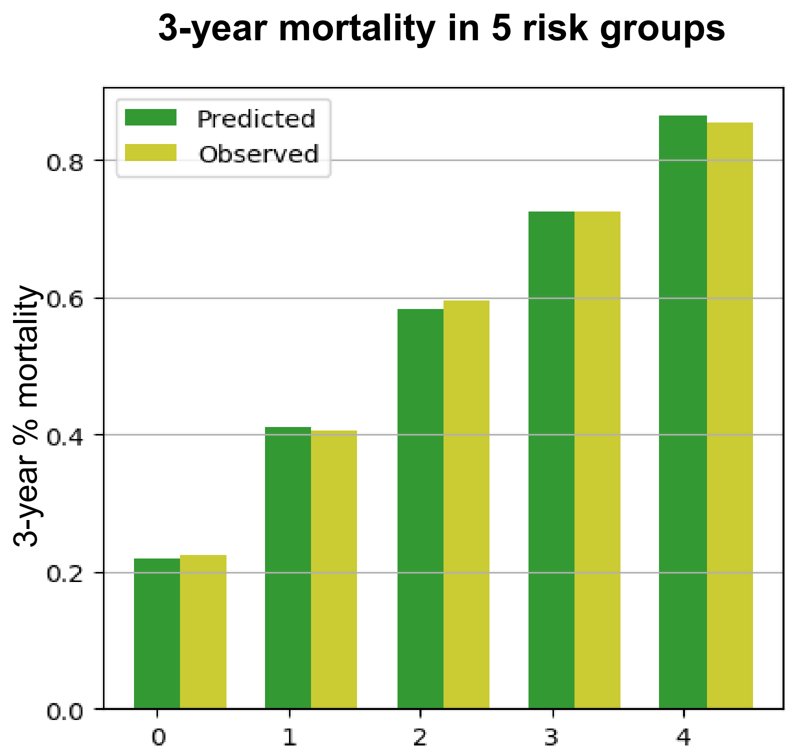
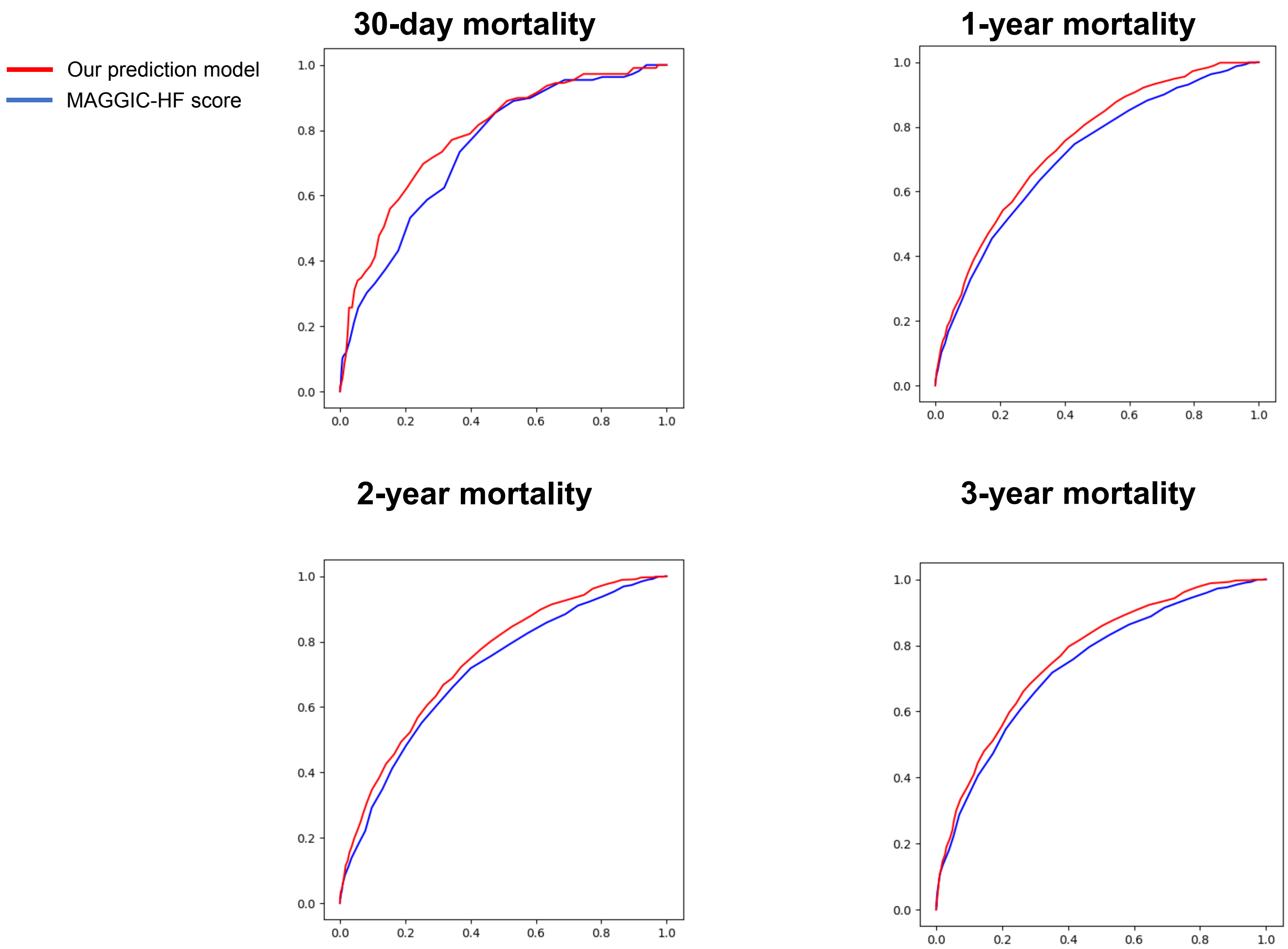
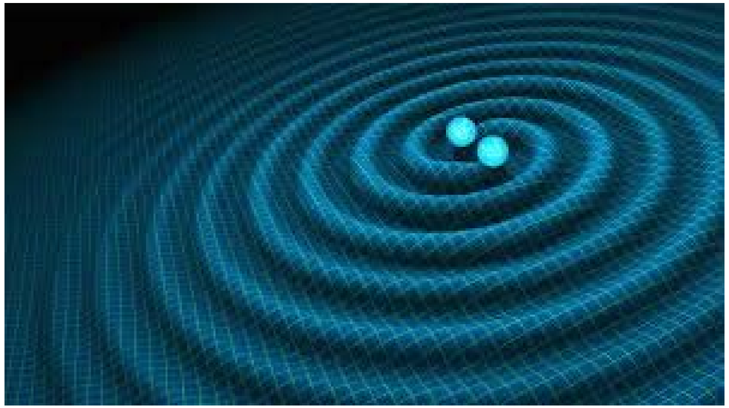

3-year mortality in 5 risk groups
Predicted
Observed
3-year % mortality
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4

Figure 4

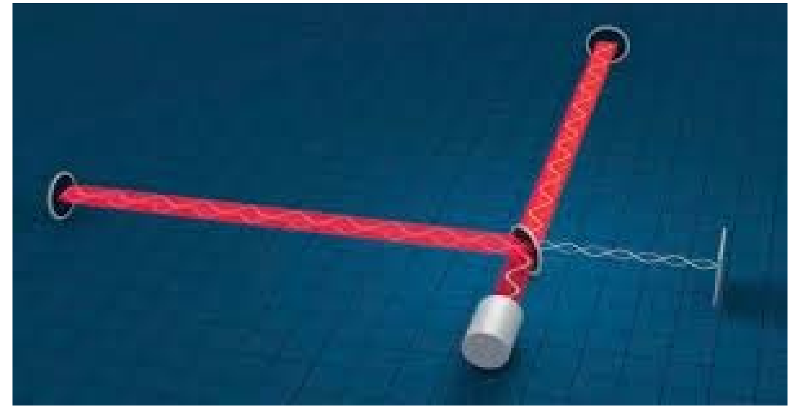

Our prediction model
MAGGIC-HF score
30-day mortality
1-year mortality
2-year mortality
3-year mortality
0.0
0.2
0.4
0.6
0.8
1.0

ICIAM 2019 & King of Spain

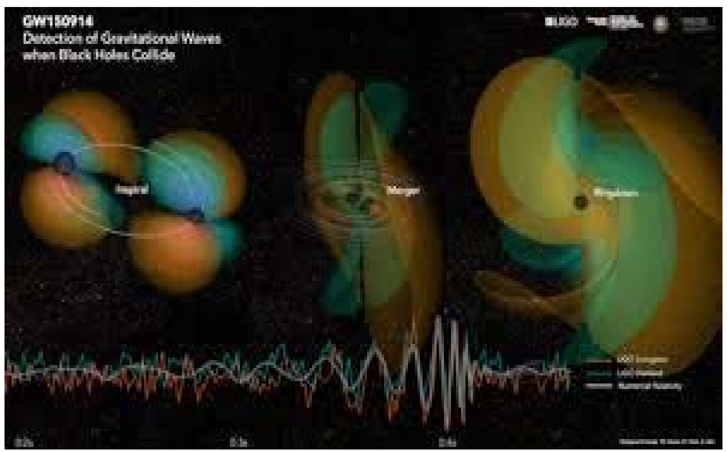
Applications of TDA (Gravitational Waves)



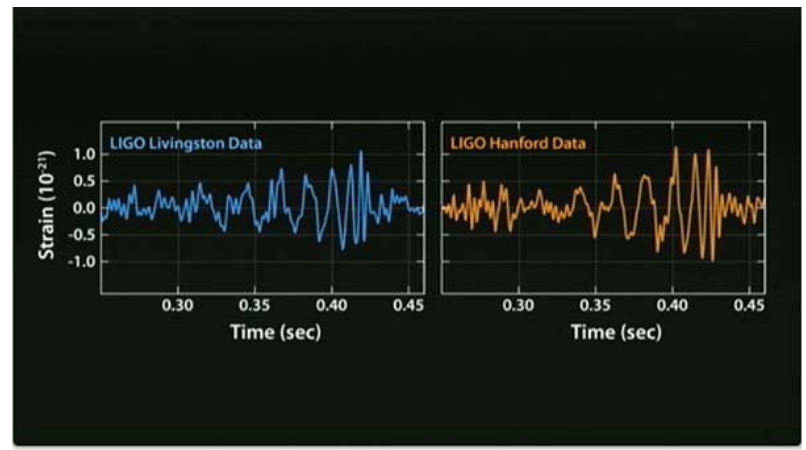
GW generated by binary neutron stars



GW observatory

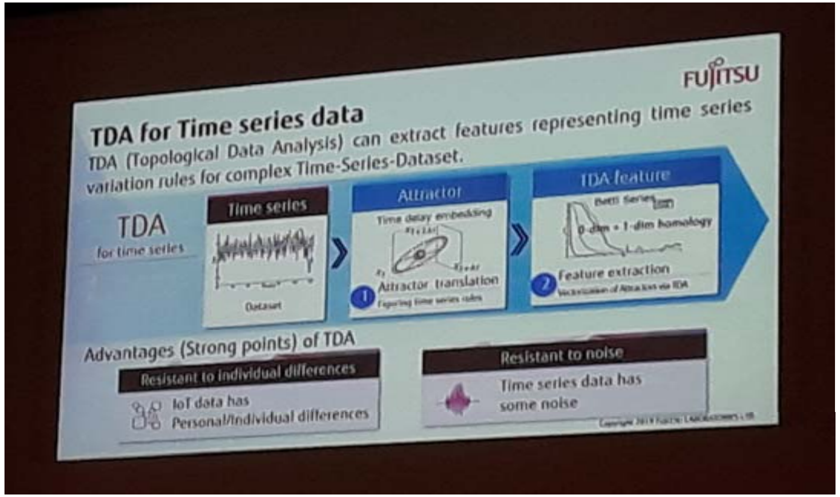


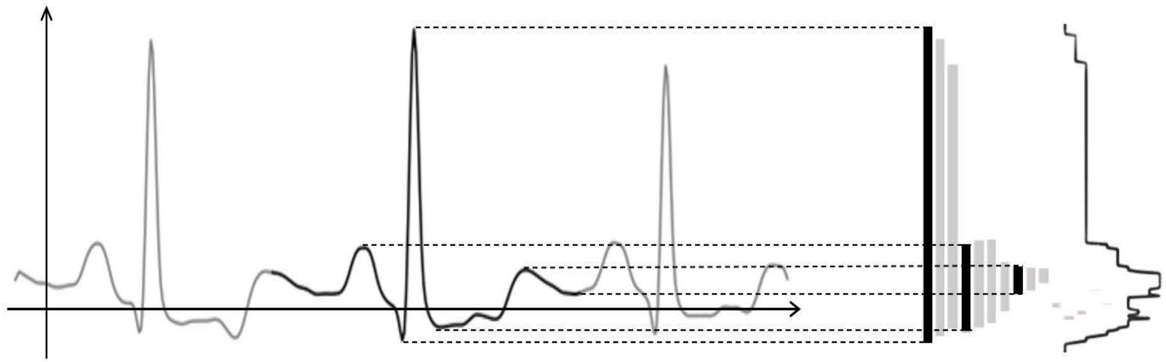
GW detection when black holes collide

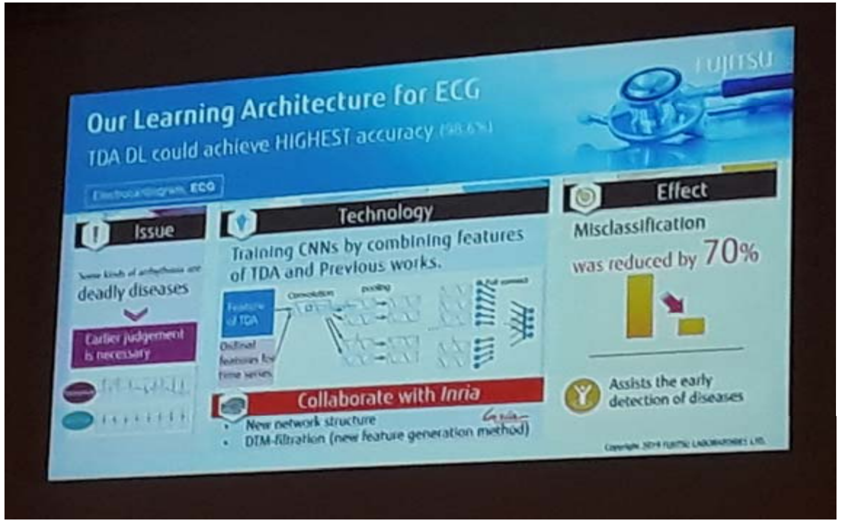


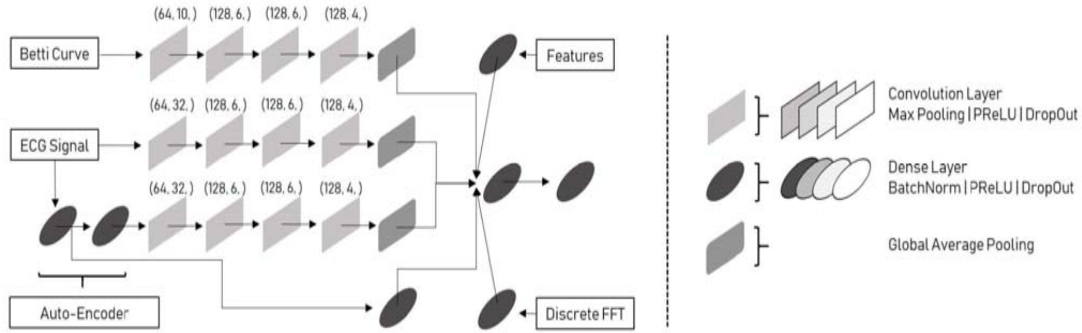
Time series images of GW

Applications of TDA (ECG Data)







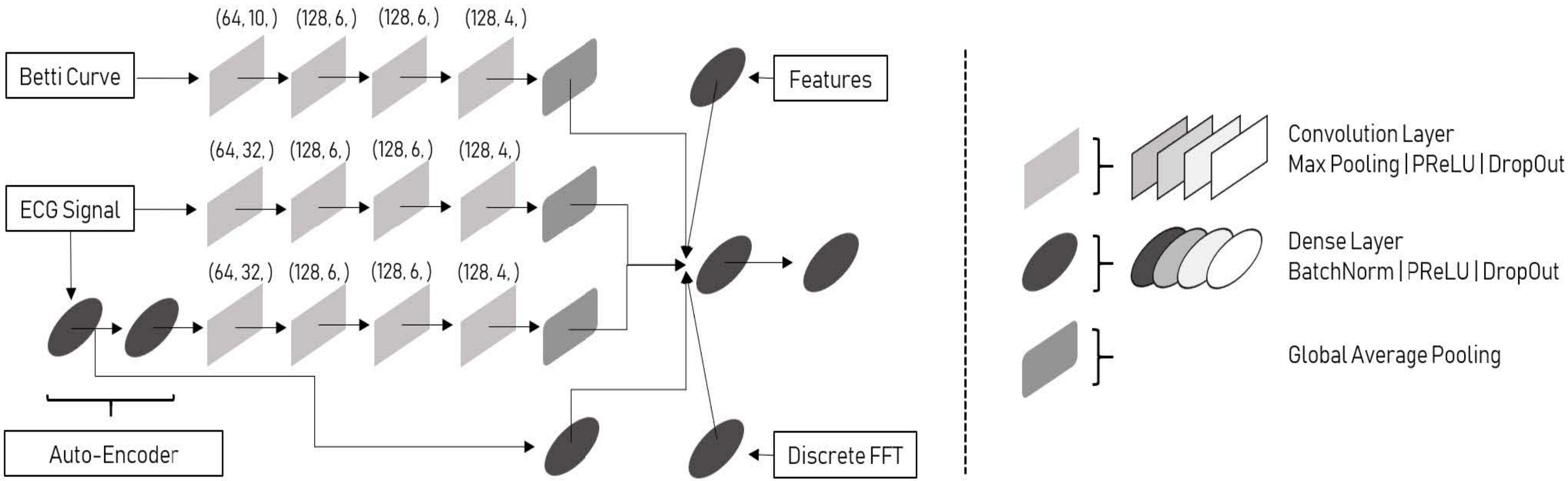


Source: TDA for Arrhythmia Detection through Modular Neural Networks, M. Dindin et al (2019)

N mapviews

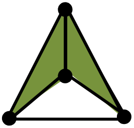
(N=# nodes)

Classification by Machine Learning



Source: TDA for Arrhythmia Detection, Dindin et al. 2019

Example: Boundary operators $\partial_i : C_i(X) \to C_{i-1}(X)$

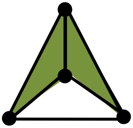


$$\partial_0 = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} : C_0(X) \to C_{-1}(X) = \mathbb{Z}$$

$$\partial_1 = \begin{pmatrix} -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} : C_1(X) \to C_0(X)$$

$$\partial_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} : C_2(X) \to C_1(X)$$

Example: Combinatorial Laplacians of a 2-dim simplicial complex



$$\Delta_i = \partial_{i+1}\partial_{i+1}^t + \partial_i^t\partial_i : C_i(X) \to C_i(X)$$

$$\Delta_{-1} = 4\, I_1 \qquad \Delta_0 = 4\, I_4$$

$$\Delta_1 = \begin{pmatrix} 2 & 1 & 1 & -1 & -1 & 0 \\ 1 & 3 & 0 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ -1 & 1 & 0 & 3 & 0 & 0 \\ -1 & 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix} \qquad \Delta_2 = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

Combinatorial Laplace Operator

Chain complex

$$\cdots \longrightarrow C_{d+1} \xrightarrow{\partial_{d+1}} C_d \xrightarrow{\partial_d} C_{d-1} \longrightarrow \cdots$$

$$\boxed{\Delta = \partial_d^t \partial_d + \partial_{d+1} \partial_{d+1}^t}$$

Ker Δ	⇒	Shape (TDA)
Det Δ	⇒	Connectivity
Inv Δ	⇒	Centrality

3) **K. Woong**, V. Reiner, and D. Stanton, Combinatorial Laplacians of matroid complexes, *Journal of the American Mathematical Society* (2000).

Harmonic space and combinatorial Hodge Theory

- Harmonic space $\mathscr{H}_i \subset C_i$

$$\mathscr{H}_i = \ker \Delta_i = \ker \partial_i \cap \ker \partial_{i+1}^t$$

- Energy minimizing property: for $h \in \mathscr{H}_i$ and $x = h + \partial_{i+1} y$

$$\langle h, h \rangle \ \leq \ \langle x, x \rangle$$

Proof: $\langle h, \partial_{i+1} y \rangle = \langle \partial_{i+1}^t h, y \rangle = 0 \quad \square$

- Hodge decomposition (J. Friedman, 1996)

$$C_i = \ker \Delta_i \oplus \operatorname{im} \Delta_i = \mathscr{H}_i \oplus \operatorname{im} \partial_{i+1} \oplus \operatorname{im} \partial_i^t$$

- Combinatorial Hodge Theory

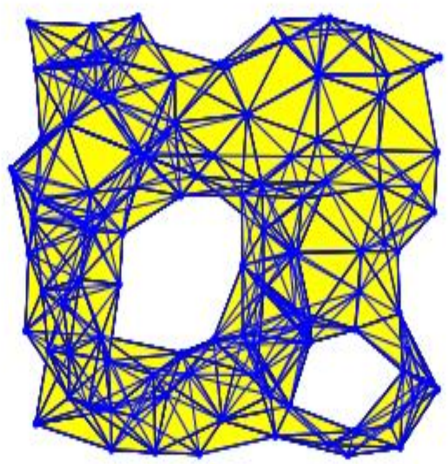
$$\mathscr{H}_i \cong \tilde{H}_i(C; \mathbb{R})$$

Proof: $\ker \partial_i \perp \operatorname{im} \partial_i^t \quad \Rightarrow \quad \ker \partial_i = \mathscr{H}_i \oplus \operatorname{im} \partial_{i+1} \quad \square$

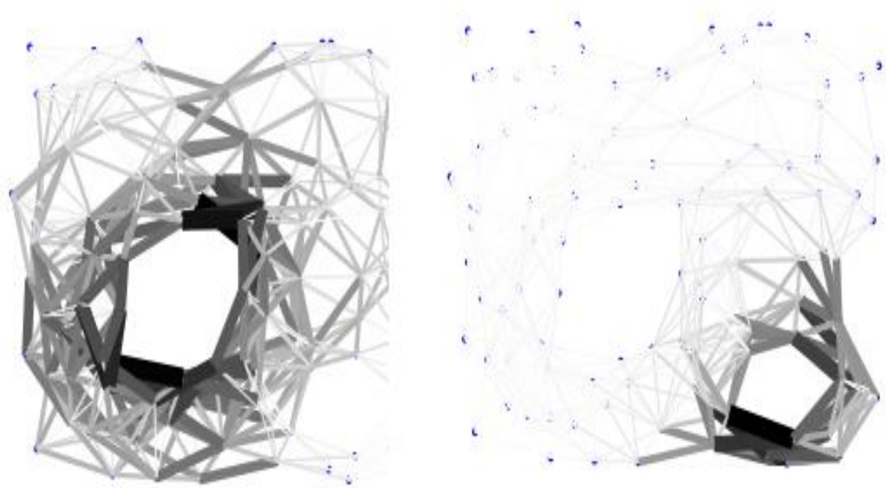
- Application to TDA

$$\beta_i = \dim \mathscr{H}_i$$

Hole Detection via Harmonic Cycles



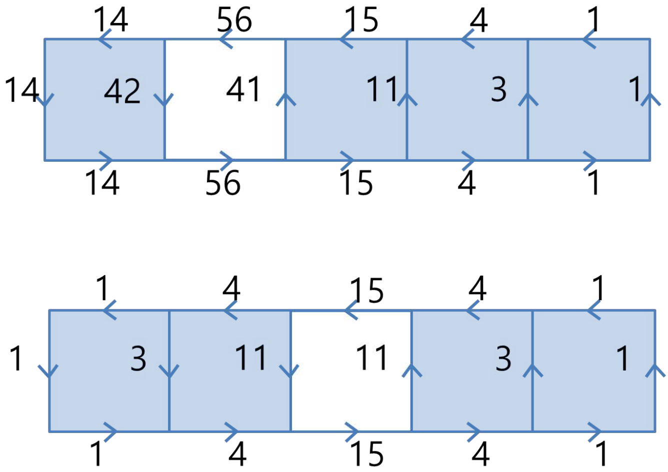
Shape of Data [4)]



Harmonic cycles (조화계)

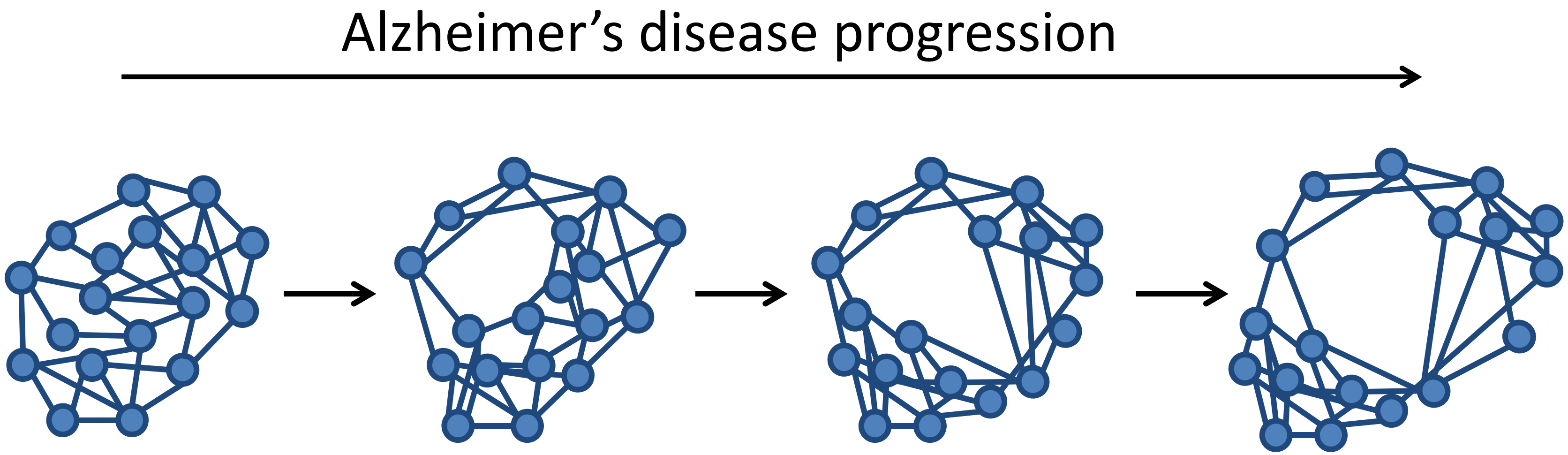
4) A. Muhammad, and M. Egerstedt, Control using higher order Laplacians in network topologies, In Proc. of 17th International Symposium on Mathematical Theory of Networks and Systems (2006).

Harmonic cycles as shape recognizer



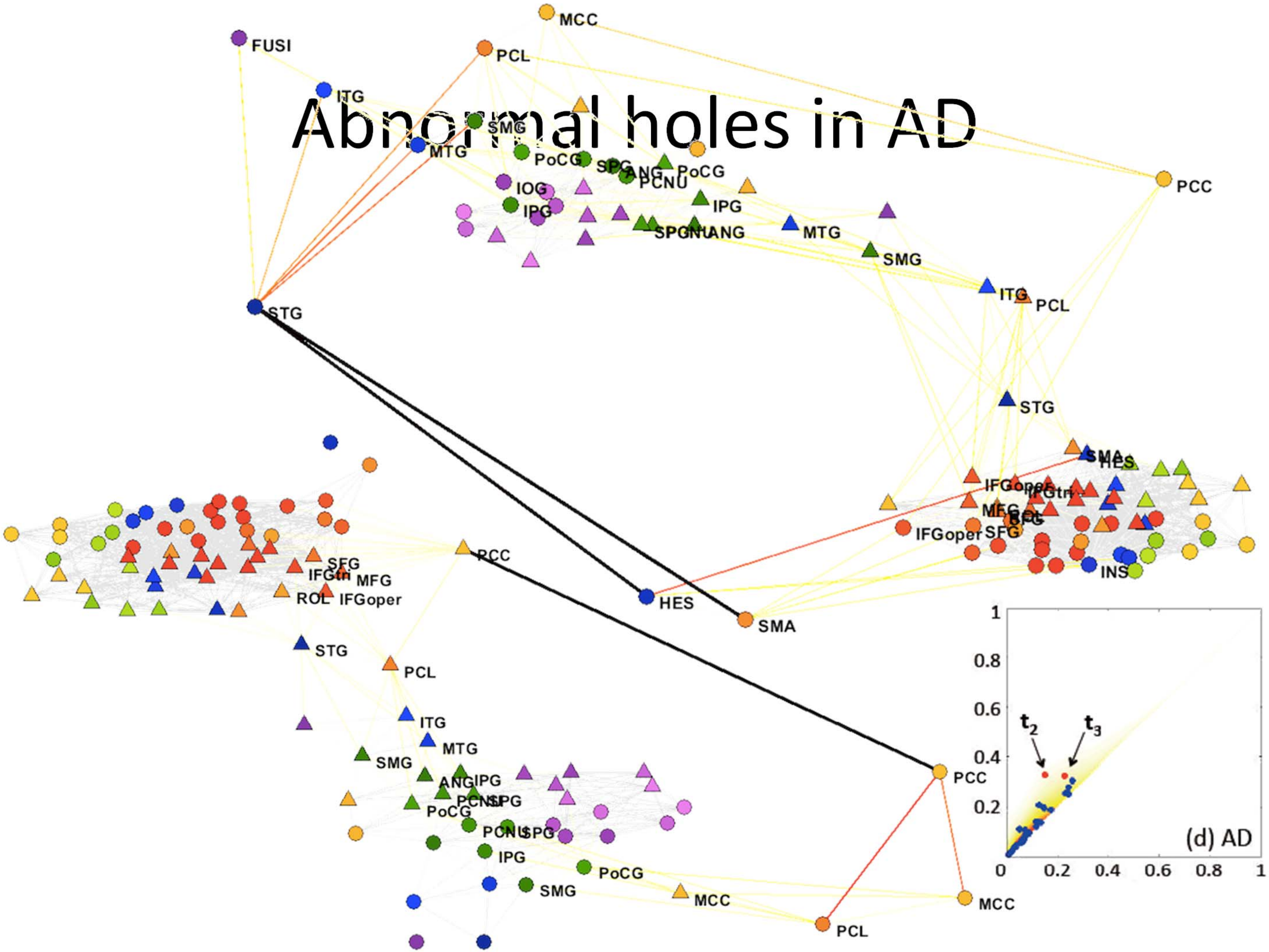
Both shapes have $\beta_1 = 1$ but different harmonic cycles.

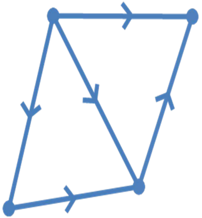
Change of hole structure



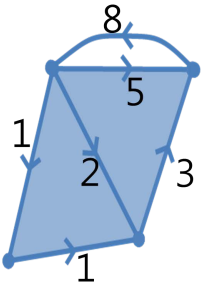
Source: *Abnormal Hole Detection in Brain Connectivity by Hodge Laplacian* by H. Lee

Abnormal holes in AD





G



$G^* \cup \omega \cong S^1$

$$I_\omega = I + 8\omega = \partial_1^t \phi + 8\omega = (1, -1, 3, -2, -5, 8)^t$$

$$I_\omega \in \ker \partial_1 \cap \ker \partial_2^t = \mathscr{H}_1 \cong \mathbb{R}$$

where $\mathscr{H}_1$ harmonic space for $G^* \cup \omega \cong S^1$.

Information Centrality

- Combinatorial Green's function $\Delta_0^{-1} = (g_{ij})$ (K. 2011)
- Information centrality I_{ab} (Stevenson and Zelen 1989)

$$I_{ab} = (g_{aa} + g_{bb} - 2g_{ab})^{-1} = C_{ab}$$

- Information centrality I_a of node a

$$I_a = |V(G)| \cdot \Big(\sum_{i \in V(G)} \frac{1}{I_{ai}} \Big)^{-1} \quad = \text{ harmonic mean of } I_{ai}$$

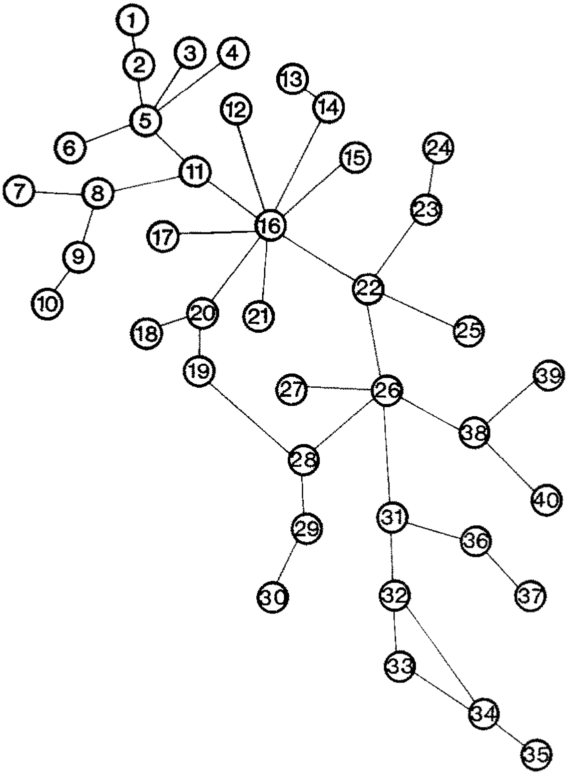
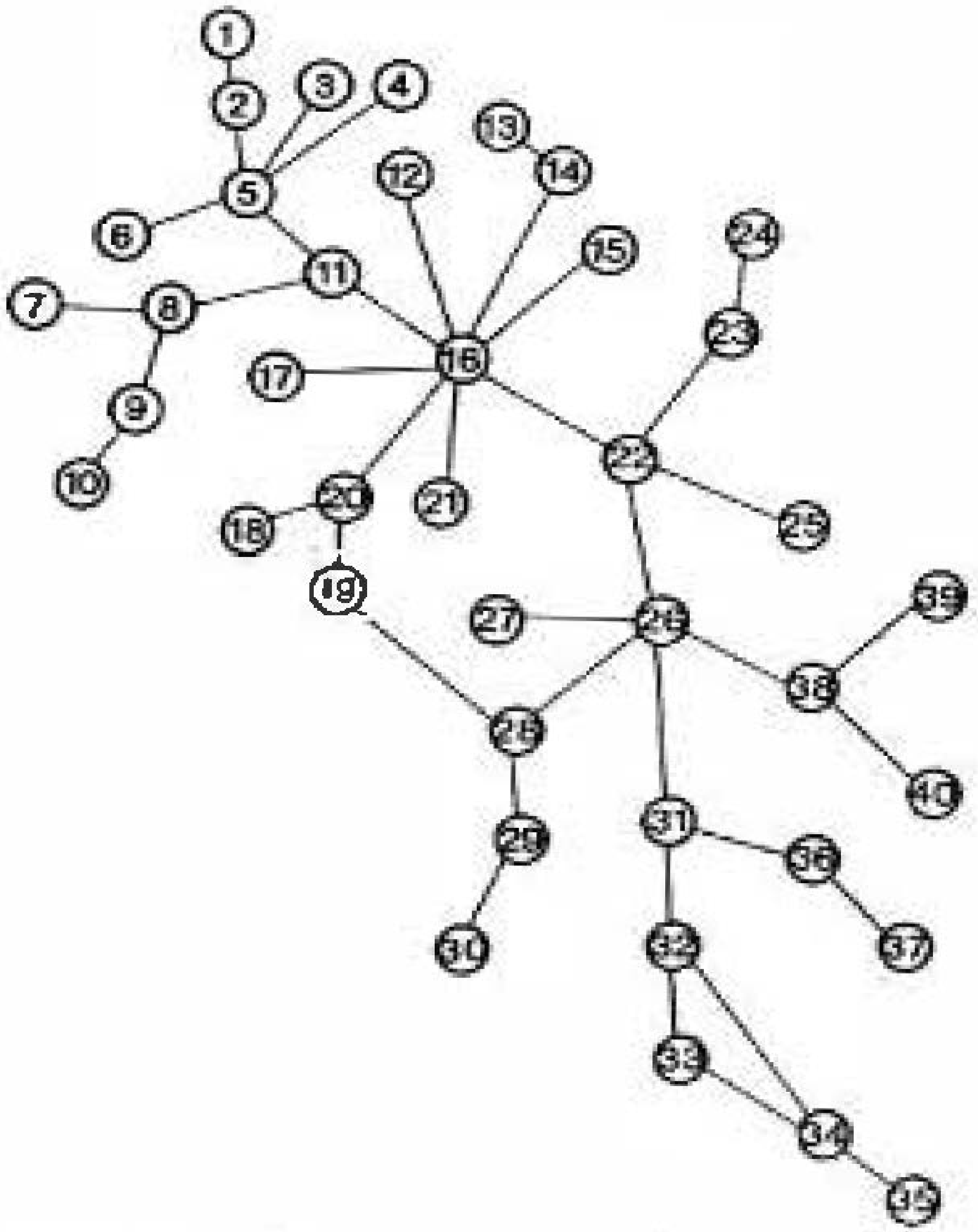
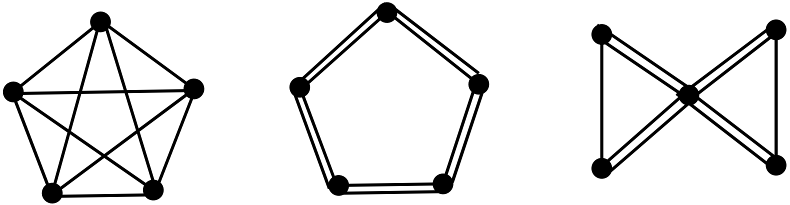


Fig. 2. Network of 40 homosexual men diagnosed with AIDS (from Figure 1 in Klovdahl 1985: 1204).



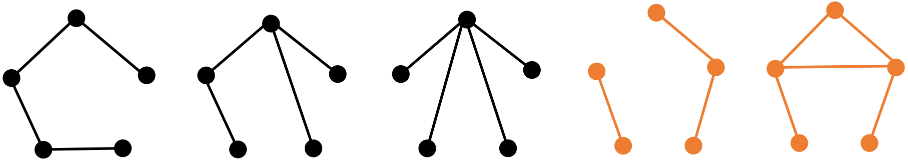
Comparison of centrality measures and rankings

Overall rank	Information	Betweenness	Distance	Degree
1	16 (0.417)	16 (17.7)	16 (0.351)	16 (0.205)
2	22 (0.392)	26 (14.0)	22 (0.345)	26 (0.128)
3	26 (0.388)	22 (13.5)	26 (0.322)	5 (0.128)
4	20 (0.357)	11 (11.6)	11 (0.302)	22 (0.103)
5	11 (0.351)	31 (7.6)	20 (0.281)	8 (0.077)
6	28 (0.348)	5 (6.6)	14 ([illegible]5)	11 (0.077)
7	19 (0.336)	8 (4.1)	19 (0.265)	20 (0.077)
8	31 (0.310)	32 (4.0)	31 (0.265)	28 (0.077)
9	14 (0.303)	28 (3.7)	12 (0.262)	31 (0.077)
10	15 ([illegible])	20 (3.1)	15 (0.262)	32 (0.077)



Question: Which network is most decentralized?

Spanning trees and non-spanning trees



spanning trees

non spanning trees

spanning tree = connected + acyclic

$k(G)$ = # spanning trees in G

Matrix-Tree Theorem

G connected graph with n vertices

- Matrix-Tree Theorem (Kirchhoff 1847)
  Let $L(G)_0 = \hat{\partial}_1 \hat{\partial}_1^t$ = $L(G)$ - (row i and column i) .

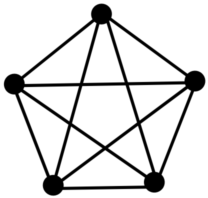
  $$\det L(G)_0 = k(G)$$

  Proof: By Cauchy-Binet theorem

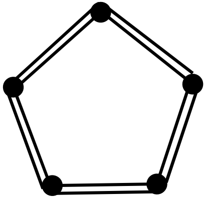
  $$\det \hat{\partial}_1 \hat{\partial}_1^t = \sum_B \det \hat{\partial}_B \det \hat{\partial}_B^t = \sum_{B \text{ sp. tree}} (\pm 1)^2 = k(G) \,.$$

- Every cofactor of $L(G) = k(G)$.

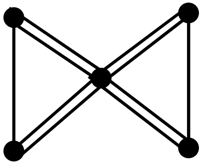
- Pseudo-determinant of $L(G)$ = $n \cdot k(G)$



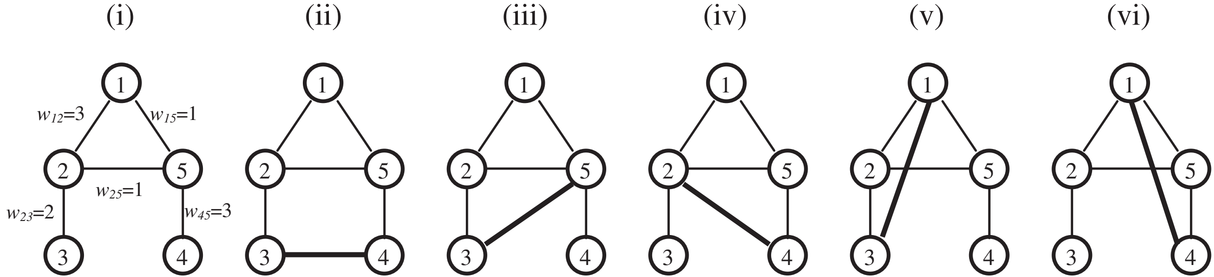
$k(G) = 125$



$k(G) = 80$



$k(G) = 64$



Information Flow	(i)	(ii)	(iii)	(iv)	(v)	(vi)
I_{12}	3.50	3.61	3.63	3.64	4.17	3.64
I_{13}	1.27	1.63	1.71	1.29	2.27	1.29
I_{14}	1.11	1.51	1.23	1.67	1.12	2.11
I_{15}	1.75	2.02	2.07	2.11	1.79	2.50
I_{23}	2.00	2.53	2.64	2.00	2.78	2.00
I_{24}	1.11	1.77	1.34	2.11	1.12	1.67
I_{25}	1.75	2.30	2.42	2.50	1.79	2.11
I_{34}	0.71	1.71	1.18	1.03	0.85	0.91
I_{35}	0.93	1.68	1.93	1.11	1.19	1.03
I_{45}	3.00	3.48	3.00	3.64	3.00	3.64
Mean(I_{ij}) (Std(I_{ij}))	1.71 (0.91)	2.22 (0.76)	2.11 (0.81)	2.11 (0.93)	2.01 (1.05)	2.09 (0.96)
Complexity $c(N)$	42	101	87	80	75	80

TDA and Topological Combinatorics

Topological Data Analysis(TDA)

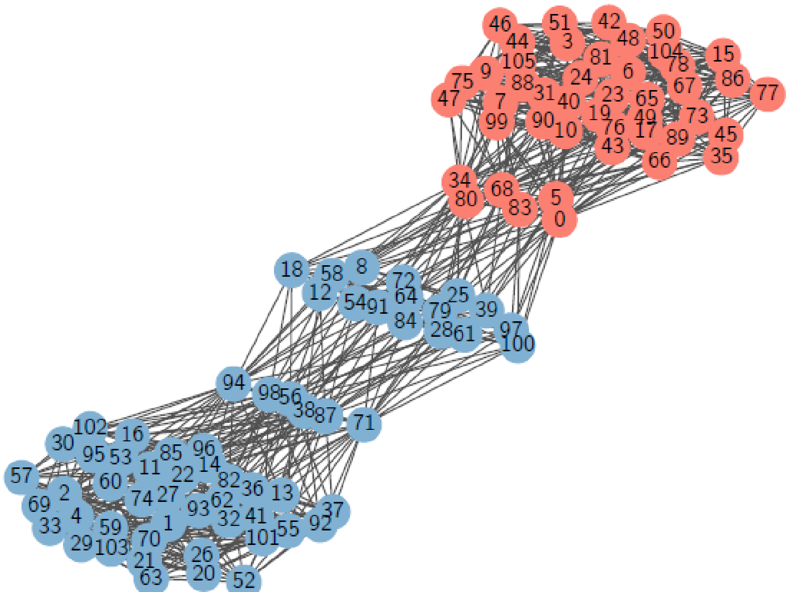
Point Cloud → Simplicial Complex → Topological Invariant

Topological Combinatorics

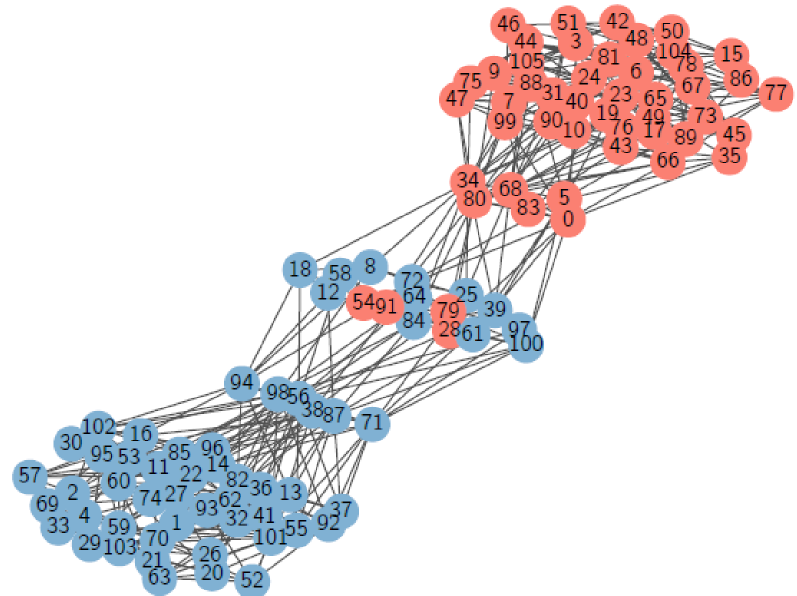
Discrete Objects → Simplicial Complex → Topological, Combinatorial Invariant

THANK YOU!

Spectral sparcification of simplicial complexes



Before



After

Keep n edges with the probability p_f proportional to $w(f)R(f,f)$ with the new accumulated weight $w(f)/\mathrm{n}p_f$ where $w(f)$ is a given weight and $R = \partial_1^t(\partial_1 W_1 \partial_1^t)^\dagger \partial_1$ is the effective resistance.

Source: Osting et al., *Towards Spectral Sparsification of Simplicial Complexes Based on Generalized Effective Resistance,* arXiv:1708.08436 (2017).

Definition of current and Thompson's Principle

For a current generator $\omega \in \binom{V(X)}{d+1}$, let

$$X_\omega := X \cup \omega \quad \text{and} \quad X^\omega := X^* \cup \omega \simeq S^d$$

$\mathscr{H}_d(X^\omega) \cong \tilde{H}_d(X^\omega; \mathbb{R}) \cong \mathbb{R}$ with $\partial_{X^\omega, d+1} = \partial_{X^*, d+1}$.

Definition: A current in X_ω is an element $I_\omega \in \mathscr{H}_d(X^\omega) \simeq \mathbb{R}$.

(TP) For a connected network (X, R), a cycle $y \in Z_1(X_\omega)$ is a current induced by a flow through ω if

$$y^t R y \leq z^t R z$$

for any cycle z with the same ω-component as I_ω.

When $R = \mathrm{id}$, I_ω is a current if $\langle y, y \rangle \leq \langle z, z \rangle$, i.e., if $y \in \mathscr{H}_1(X^\omega)$.

Comparison of centrality measures and rankings [a]

Overall rank	Information	Betweenness	Distance	Degree
1	16 (0.417)	16 (17.7)	16 (0.351)	16 (0.205)
2	22 (0.392)	26 (14.0)	22 (0.345)	26 (0.128)
3	26 (0.388)	22 (13.5)	26 (0.322)	5 (0.128)
4	20 (0.357)	11 (11.6)	11 (0.302)	22 (0.103)
5	11 (0.351)	31 (7.6)	20 (0.281)	8 (0.077)
6	28 (0.348)	5 (6.6)	14 (0.265)	11 (0.077)
7	19 (0.336)	8 (4.1)	19 (0.265)	20 (0.077)
8	31 (0.310)	32 (4.0)	31 (0.265)	28 (0.077)
9	14 (0.303)	28 (3.7)	12 (0.262)	31 (0.077)
10	15 (0.299)	20 (3.1)	15 (0.262)	32 (0.077)

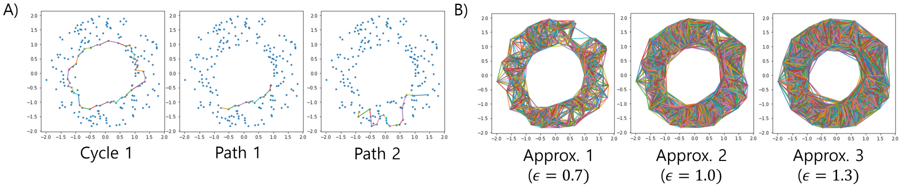
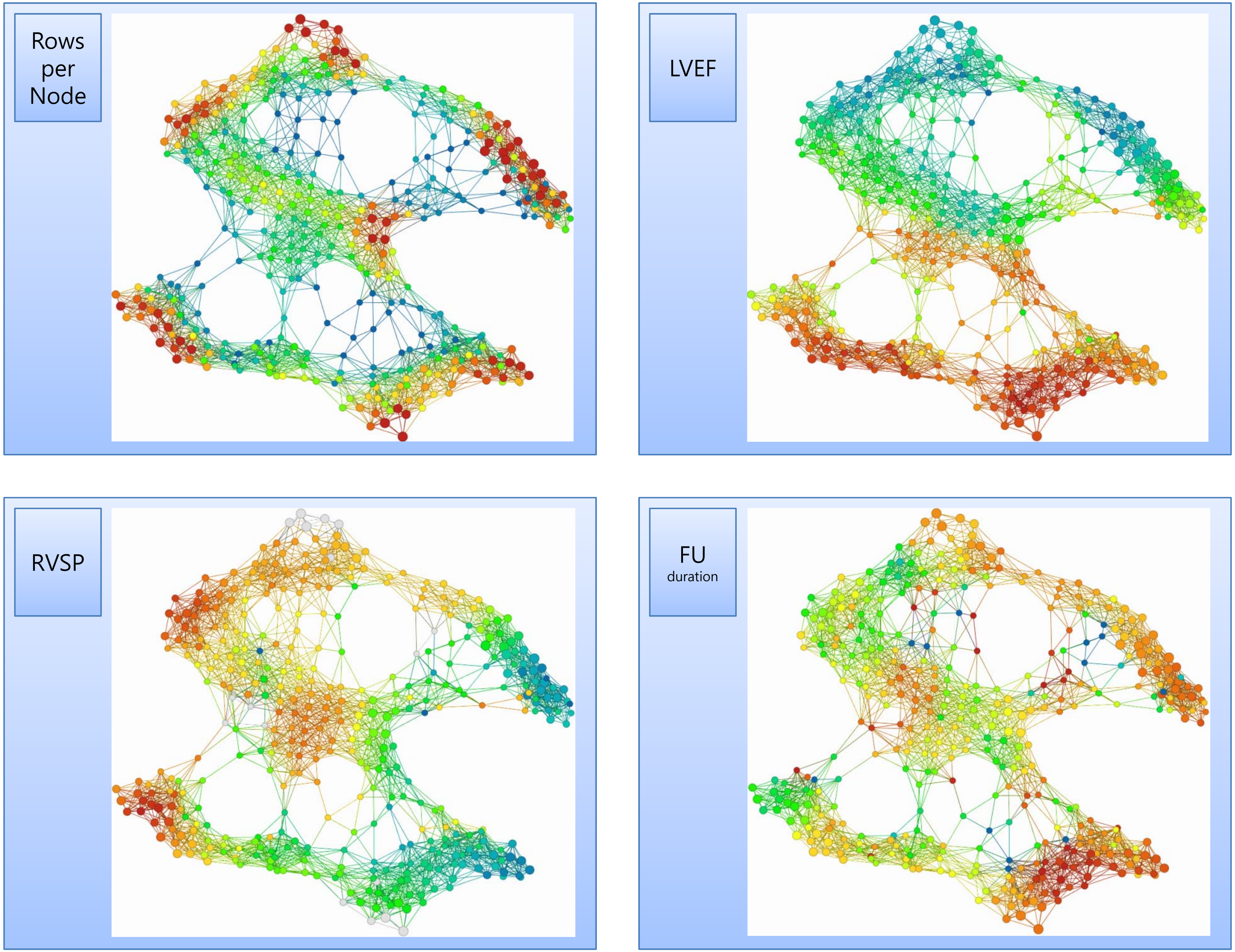
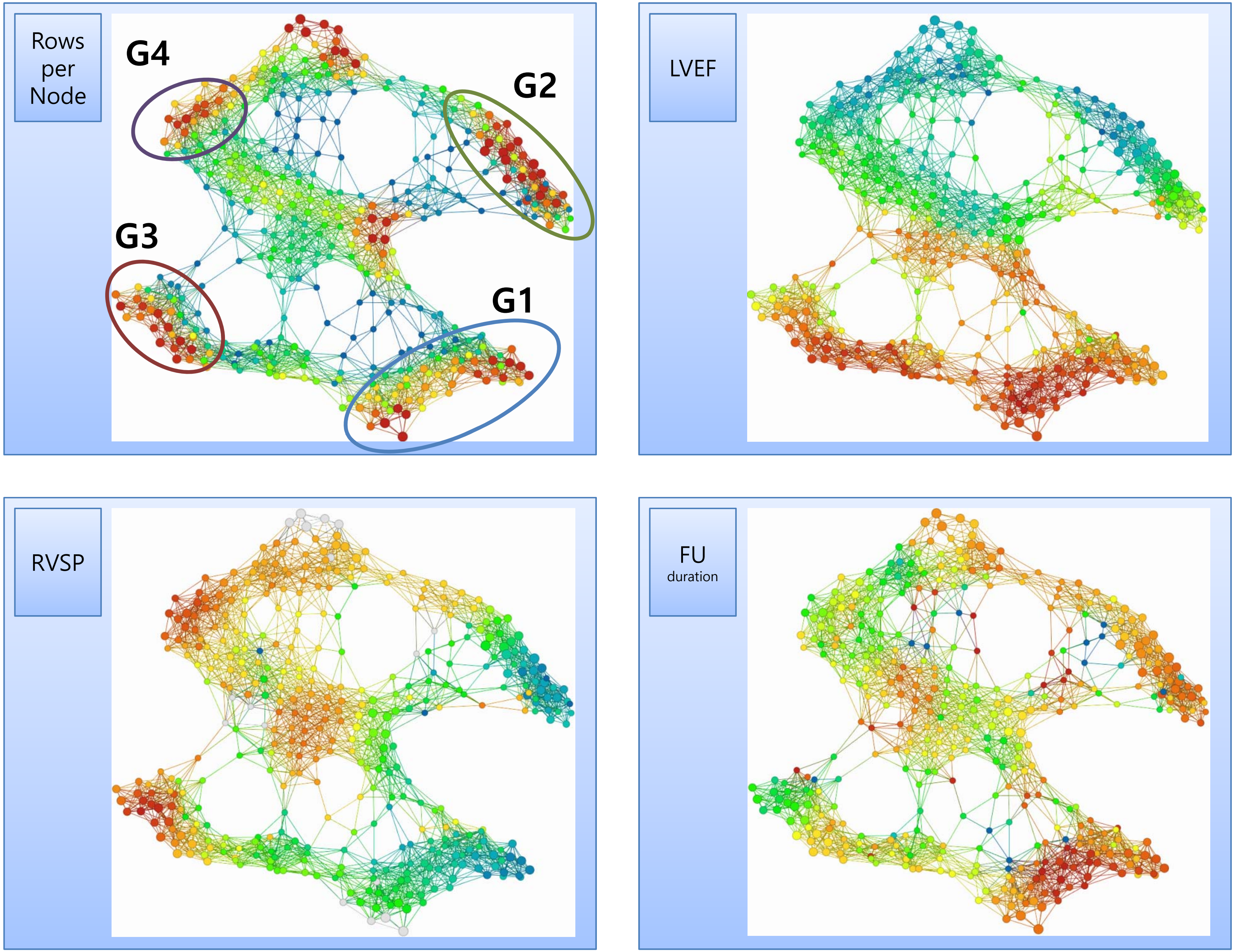


Fig. 3: In A, we have a cycle and two paths. In B, we have cell complexes approximating the given point cloud.

	$\epsilon = 0.7$	$\epsilon = 1.0$	$\epsilon = 1.3$
Cycle 1	1	1	1
Path 1	0.2307	0.2300	0.2537
Path 2	0.2095	0.2222	0.2351

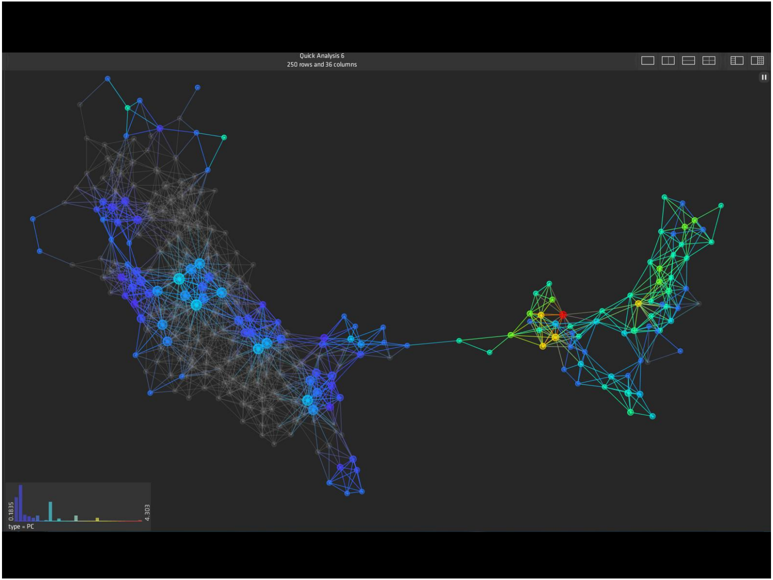
Table 1: The rational winding number of each case.


Rows per Node
LVEF
RVSP
FU duration

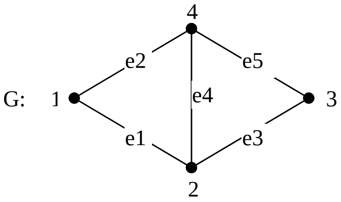


Rows per Node
G4
G2
G3
G1
LVEF
RVSP
FU duration

Pancreatic Cancer Data Analysis



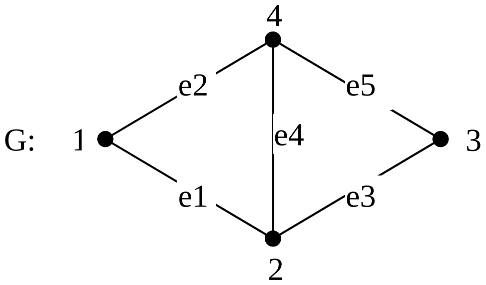
Example: Δ_0



$$\partial_1 = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \quad \partial_0 = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\Delta_0 = \partial_1 \partial_1^t + \partial_0^t \partial_0 = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Example: Δ_0



$$\Delta_0 = \partial_1 \partial_1^t + \partial_0^t \partial_0 = L + J = \begin{pmatrix} 3 & 0 & 1 & 0 \\ 0 & 4 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

Example: $\partial_{d+1}\partial_{d+1}^t$ vs Δ_d

$G = K_5$ complete graph on 5 vertices

$$\partial_1\partial_1^t = \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & -1 & 4 \end{pmatrix} \qquad \Delta_0 = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

always singular — non-singular for connected G

Combinatorial Laplacian

$C = \{C_i, \partial_i\}$ a finite chain complex over $\mathbb{Z}$ with $C_{-1} = \mathbb{Z}$

$$\partial_i : C_i \to C_{i-1} \qquad \partial_i^t : C_{i-1} \to C_i$$

$$\partial_{i+1}\partial_{i+1}^t : C_i \to C_i \qquad \partial_i^t \partial_i : C_i \to C_i$$

Definition (Combinatorial Laplacian $\Delta_i : C_i \to C_i$)

$$\Delta_i = \partial_{i+1}\partial_{i+1}^t + \partial_i^t \partial_i$$

Combinatorial Laplacians and harmonic space

X = finite simplicial (cell) complex

$C = \{C_i, \partial_i\}$ chain complex of X over $\mathbb{R}$

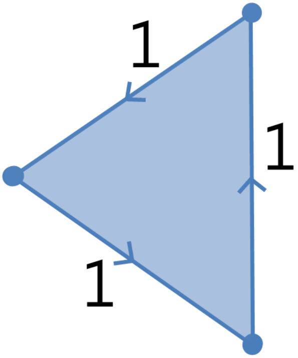
- Combinatorial Laplacian $\Delta_i : C_i \to C_i$ (Eckmann 1945)

$$\Delta_i = \partial_{i+1}\partial_{i+1}^t + \partial_i^t \partial_i$$

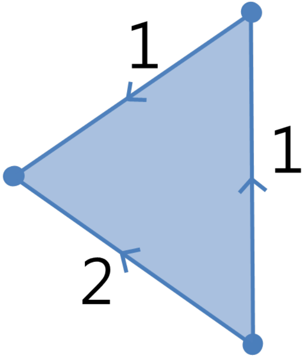
- Harmonic space $\mathscr{H}_i$

$$\mathscr{H}_i = \ker \Delta_i$$

Example: cycle and cocycle

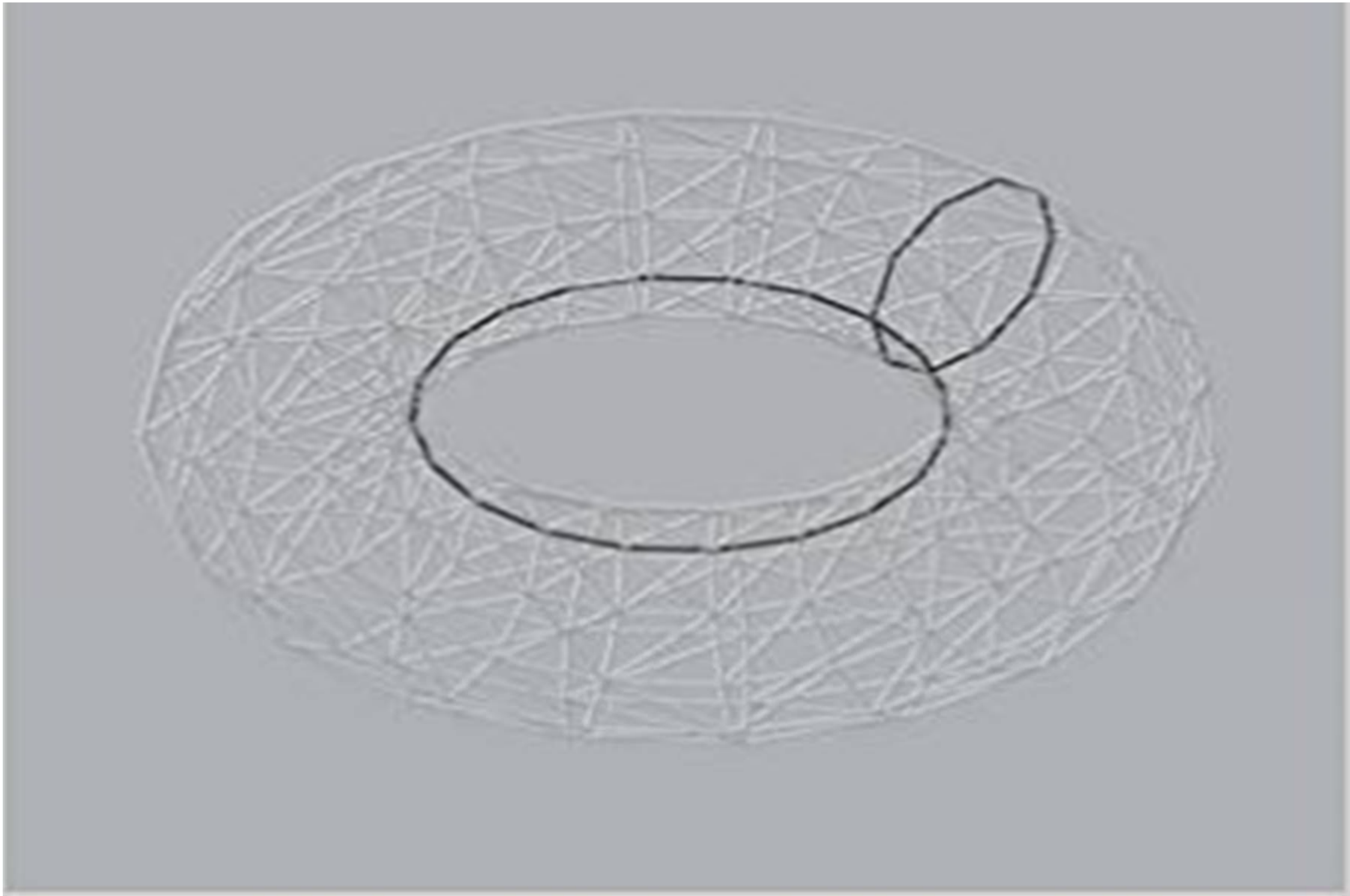


cycle : Yes
cocycle : No



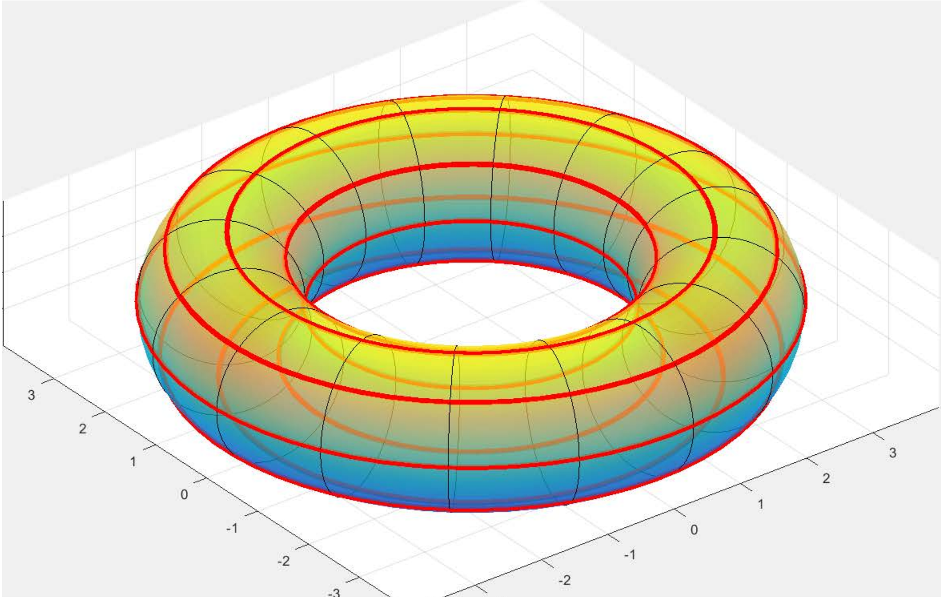
cycle : No
cocycle : Yes

Homology basis for torus

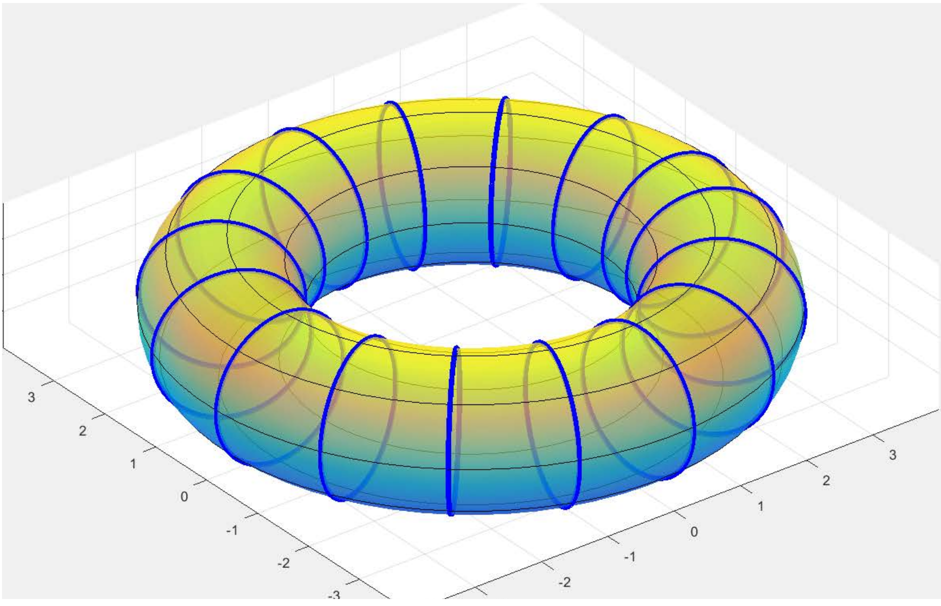


Source: Cover page of *Topology for Computing* by A. Zomorodian

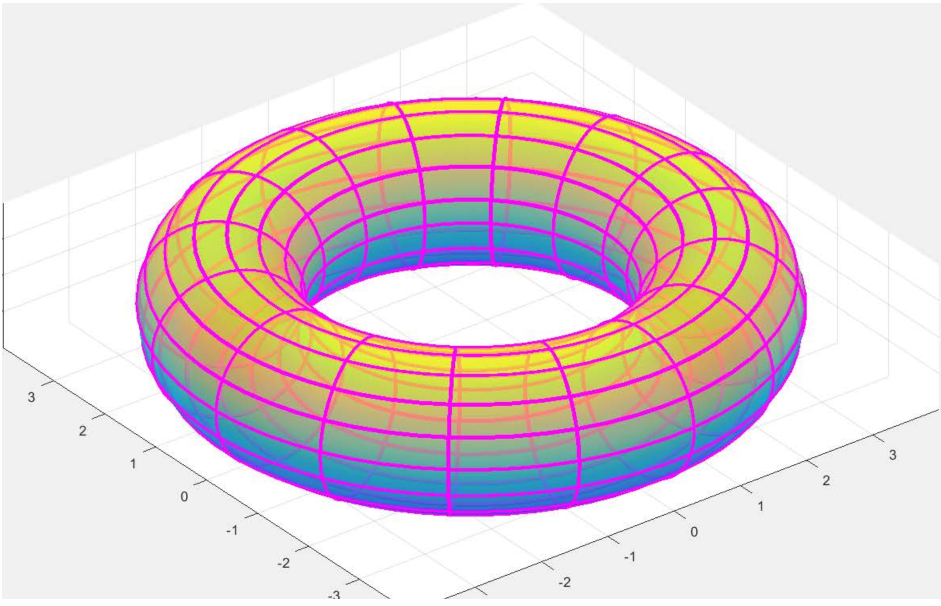
Example: harmonic cycles of a torus



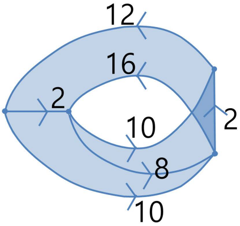
Example: harmonic cycles of a torus

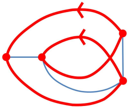


Example: harmonic cycles of a torus

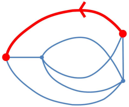


Winding number for paths (Kim and K. 2018)

$\lambda =$ 



$\circ\, \lambda \begin{aligned} &= 12 + 10 + 16 + 10 \\ &= \mathbf{2} \cdot k(G) \end{aligned}$



$\circ\, \lambda = 12 = \mathbf{\frac{1}{2}} \cdot k(G)$

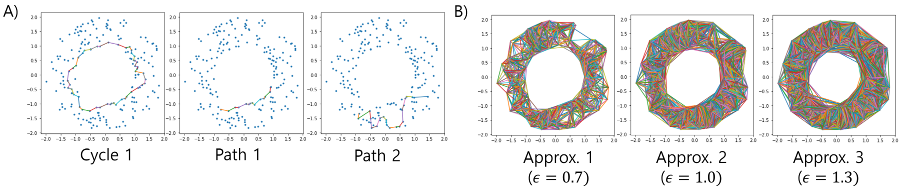
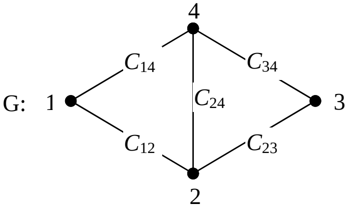


Fig. 3: In A, we have a cycle and two paths. In B, we have cell complexes approximating the given point cloud.

	$\epsilon = 0.7$	$\epsilon = 1.0$	$\epsilon = 1.3$
Cycle 1	1	1	1
Path 1	0.2307	0.2300	0.2537
Path 2	0.2095	0.2222	0.2351

Table 1: The rational winding number of each case.

Laplacian for a Weighted Graph (Conductance Network)



$c_{ij} = c_{ji}$ = conductance between node i and node j ($i \neq j$)

$$L = \begin{pmatrix} c_{12}+c_{13}+c_{14} & -c_{12} & -c_{13} & -c_{14} \\ -c_{21} & c_{21}+c_{23}+c_{24} & -c_{23} & -c_{24} \\ -c_{13} & -c_{32} & c_{31}+c_{32}+c_{34} & -c_{34} \\ -c_{41} & -c_{42} & -c_{43} & c_{41}+c_{42}+c_{43} \end{pmatrix}$$

Discrete Laplace Equation $L\phi = \iota$

(G, R) = connected resistor network

G = connected finite graph with $\partial_1 : C_1(G) \to C_0(G)$

R = resistance (positive diagonal matrix) : $C_1(G) \to C_1(G)$

$C = R^{-1}$ = conductance (weights for edges)

Laplacian $L = \partial_1 C \partial_1^t \ : C_0(G) \to C_0(G)$

Net current flow $\iota = (\iota_v) \in C_0(G)$ satisfying $\sum \iota_v = 0$

$(\iota_a = \alpha, \quad \iota_b = -\alpha, \quad \iota_v = 0$ for $v \neq a, b)$

Potential $\phi = (\phi_v) \in C_0(G)$

Kirchhoff's Law: $\quad L\phi = \iota$

A solution to $L\phi = \iota$

$L = \partial_1 C \partial_1^t$ discrete Laplacian of a connected (G, R)
$J = \partial_0^t \partial_0$ all 1's matrix

$\Delta_0 = L + J$ nonsingular by combinatorial Hodge theory

For $\iota = (\iota_v)$ with $\sum \iota_v = 0$,

$$(L + J)\phi = \iota \quad \Rightarrow \quad L\phi = \iota\,.$$

A soluton to $L\phi = \iota$:

$$\phi = (L + J)^{-1}\iota$$

$\mathcal{G} = (L + J)^{-1}$ combinatorial Green's function (K. 2011)

Effective resistance between two nodes a and b

(G, R) = connected finite resistor network

Let a current α flow into a and out of b $\quad (a, b \in V(G))$.

Compute the potentials ϕ_a and ϕ_b : discrete Laplace equation

Effective resistance

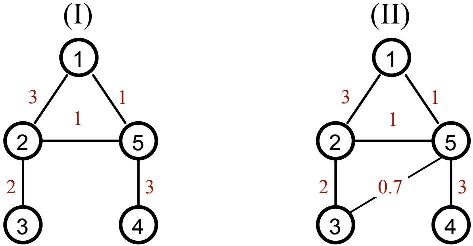
$$R_{ab} = \frac{\phi_a - \phi_b}{\alpha}$$

Effective conductance

$$C_{ab} = R_{ab}^{-1}$$

Information centrality I_a of node a

$$I_a = |V(G)| \cdot \Big(\sum_{i \in V(G)} \frac{1}{I_{ai}} \Big)^{-1}$$



Degree Centrality d_a	(I)	(II)	Information Centrality I_a	(I)	(II)
d_1	4	4	I_1	1.96	2.24
d_2	6	6	I_2	2.21	2.64
d_3	2	2.7	I_3	1.33	1.93
d_4	3	3	I_4	1.41	1.73
d_5	5	5.7	I_5	1.96	2.65

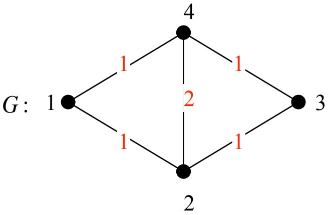
I_{ab} as a rate of change of $k(G)$ (Shin and K. 2014)

$$I_{ab} = \frac{k(G)}{k(G_{ab})} = \frac{k(G)}{\partial_x k(G)}$$

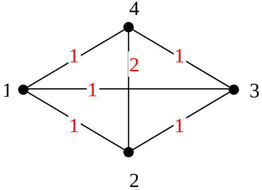
$\partial_x k(G)$ is inversely proportional to I_{ab}

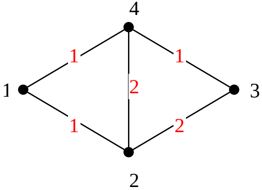
Hence, adding an edge *ab* where I_{ab} is smallest causes the largest increase in $k(G)$.

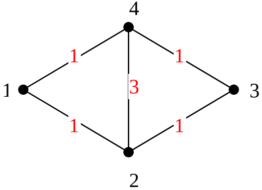
Decentralized network expansion algorithm ($k(G) = 12$)



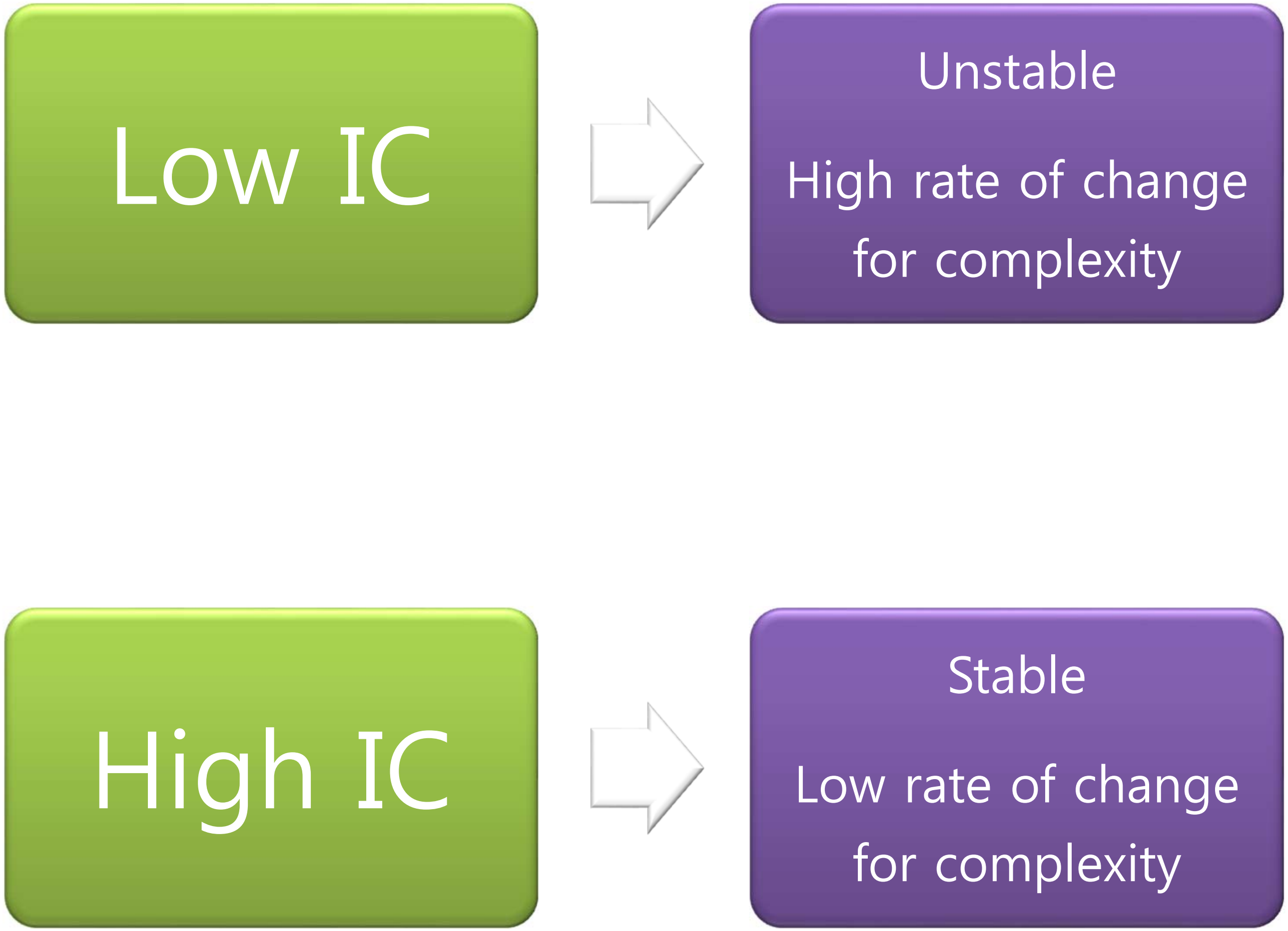
$l_{13} = 1$ $l_{23} = 12/7$ $l_{24} = 3$



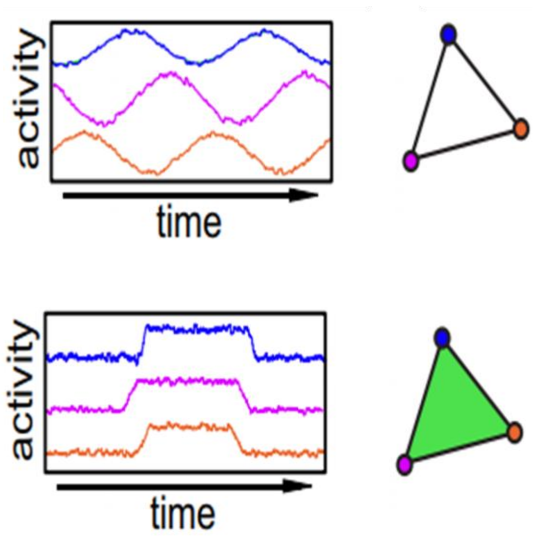


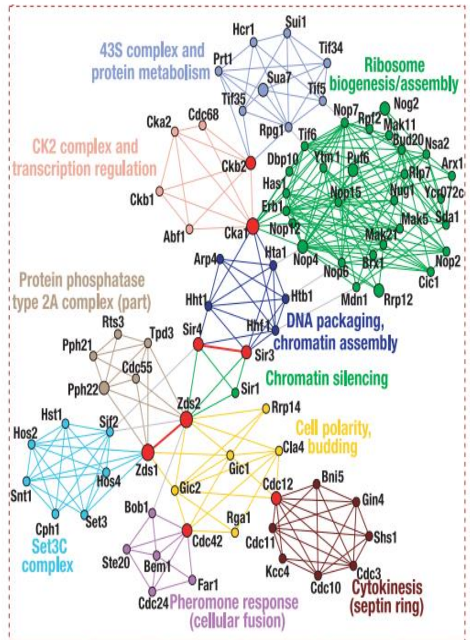


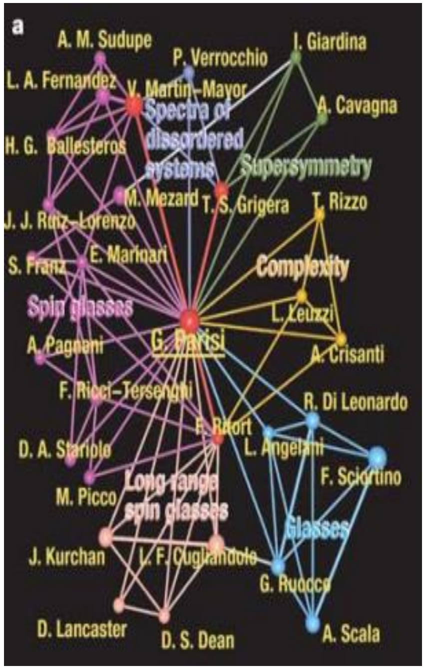
$k(G \cup 13) = 24$ $k(G \cup 23) = 19$ $k(G \cup 24) = 16$


Low IC
Unstable
High rate of change for complexity
High IC
Stable
Low rate of change for complexity

Emergence of Simplicial Networks





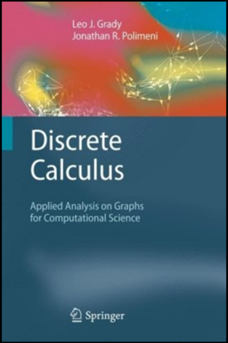


Brain network[6)]

Protein-Protein Interaction[7)]

Co-authorship Network[7)]

6) C. Giusti, R. Ghrist, D. S. Bassett, Two' s company, three (or more) is a simplex, *Journal of computational neuroscience* (2016).
7) Palla, I. Derényi, I. Farkas. T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, *Nature* (2015).
8) E. Estrada, G. J. Ross, Centralities in simplicial complexes. Applications to protein interaction networks, *Journal of theoretical biology* (2018).



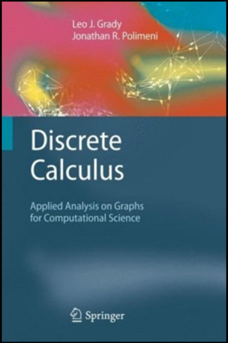
Author(s): Leo J. Grady, Jonathan R. Polimeni (auth.)

Publisher: Springer-Verlag London, Year: 2010

Although the authors are not aware of any attempts to use topological indices defined on the *edge Laplacian* to measure other aspects of the graph (or the chemical properties of the molecule), the measures above suggest an easy extension to global edge–face (edge–cycle) relationships. Specifically, a higher-order Kirchhoff index could be defined as

$$\text{Higher-OrderKI}(\mathcal{G}) = m\,\text{trace}\{\mathbf{L}_1^{\dagger}\}, \tag{8.9}$$

where $m = |\mathcal{E}|$ and $\mathbf{L}_1$ is the edge Laplacian (see Chap. 2). If the graph cycle set constitutes a basis (see Chap. 4), then the edge Laplacian matrix has full rank and the pseudoinverse is replaced by the true inverse of the edge Laplacian matrix. Recall from Chap. 4 that a set of $|\mathcal{E}| - |\mathcal{V}| + 1$ independent cycles will form a basis. This measure of the higher-order Kirchhoff index does not retain its original interpretation in terms of the effective resistance, since the effective resistance between two edges does not have a conventional definition. Despite losing this interpretation, this measure on the edge Laplacian matrix provides a value that indicates the relative "connectedness" of the edges via their incident nodes and cycles.



Author(s): Leo J. Grady, Jonathan R. Polimeni (auth.)

Publisher: Springer-Verlag London, Year: 2010

Although the authors are not aware of any attempts to use topological indices defined on the *edge Laplacian* to measure other aspects of the graph (or the chemical properties of the molecule), the measures above suggest an easy extension to global edge–face (edge–cycle) relationships. Specifically, a higher-order Kirchhoff index could be defined as

$$\text{Higher-OrderKI}(\mathcal{G}) = m\,\text{trace}\{\mathbf{L}_1^{\dagger}\}, \tag{8.9}$$

where $m = |\mathcal{E}|$ and $\mathbf{L}_1$ is the edge Laplacian (see Chap. 2). If the graph cycle set constitutes a basis (see Chap. 4), then the edge Laplacian matrix has full rank and the pseudoinverse is replaced by the true inverse of the edge Laplacian matrix. Recall from Chap. 4 that a set of $|\mathcal{E}| - |\mathcal{V}| + 1$ independent cycles will form a basis. This measure of the higher-order Kirchhoff index does not retain its original interpretation in terms of the effective resistance, since the effective resistance between two edges does not have a conventional definition. Despite losing this interpretation, this measure on the edge Laplacian matrix provides a value that indicates the relative "connectedness" of the edges via their incident nodes and cycles.

Choose a current $I_\omega = I + i_\omega[\omega]$ with $i_\omega \, (\neq 0)$ fixed

Define a voltage $V_\omega = V + v_\omega[\omega]$ by

$$V = RI \quad \text{and} \quad V_\omega \perp I_\omega$$

Then v_ω is given by

$$I^t RI + v_\omega i_\omega = 0$$

High-dimensional effective resistance for $\omega \in \binom{V(X)}{d+1}$

$$R_\omega = \left| \frac{v_\omega}{i_\omega} \right|$$

(If $|i_\omega| = 1$, then $R_\omega = I^t RI$.)

A formula for R_ω (K. and Lee 2018)

(X^d, R) with $\tilde{H}_{d-1}(X;\mathbb{R}) = 0$. (Assume R= id.)
$\Delta_{d-1} = \partial_{X,d}\partial^t_{X,d} + \partial^t_{X,d-1}\partial_{X,d-1} = L_{d-1} + J_{d-1}$
$\partial_\omega = \partial_{X_\omega,d}[\omega]$.

$$R_\omega = \partial^t_\omega \, \Delta^{-1}_{d-1} \, \partial_\omega$$

Proof: Let $i_\omega = -1$. Then $R_\omega = v_\omega$.

$V_\omega \in \ker \partial^t_{X^\omega,d+1} = \ker \partial^t_{X^*,d+1}$
$\Rightarrow \quad V_\omega = \partial^t_{X_\omega,d}\Phi$ for some $\Phi \in C_{d-1}(X) \quad \Rightarrow \quad v_\omega = \partial^t_\omega \Phi$

$I_\omega \in \ker \partial_{X^\omega,d}$ and $i_\omega = -1 \quad \Rightarrow \quad \partial_\omega = \partial_{X,d} I = \partial_{X,d} V = L_{d-1}\Phi$

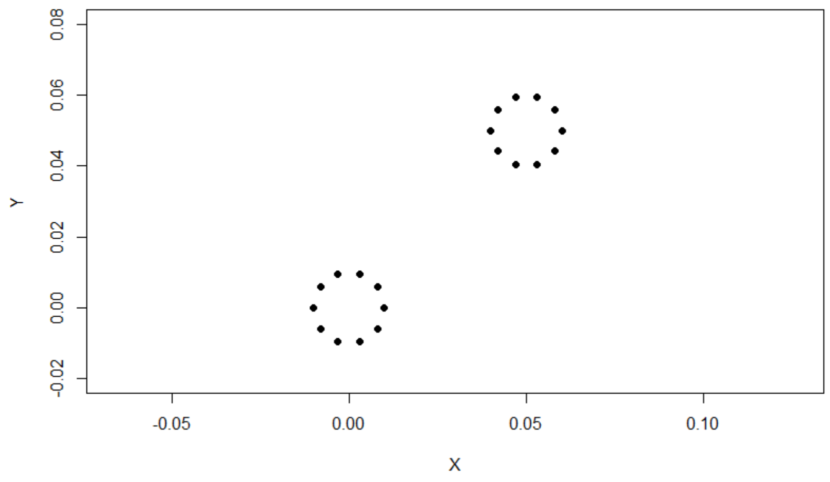
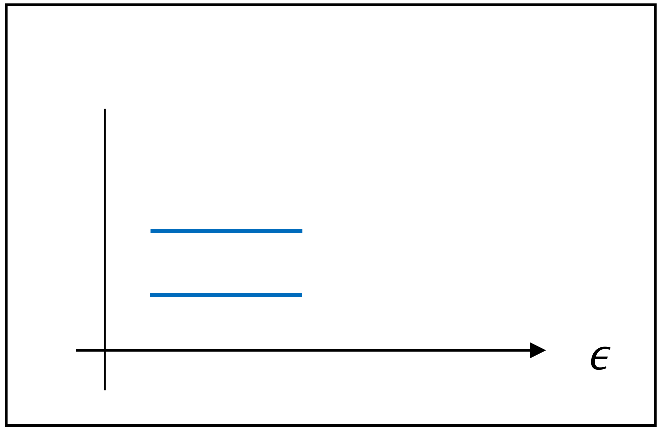
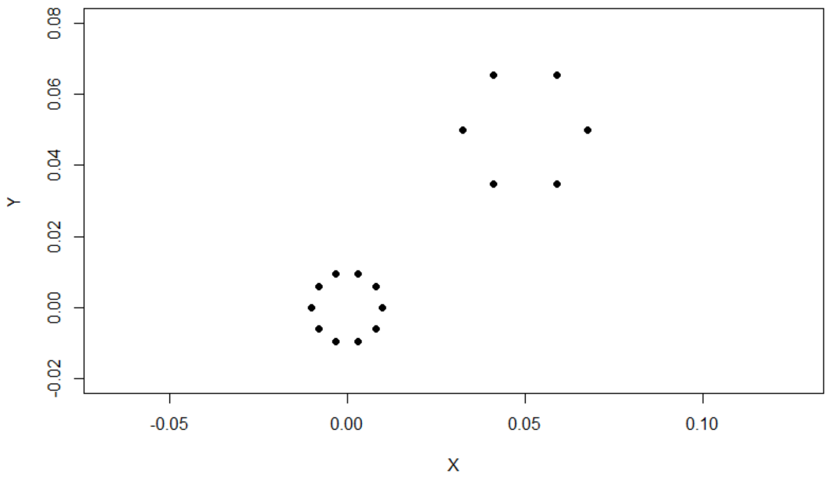
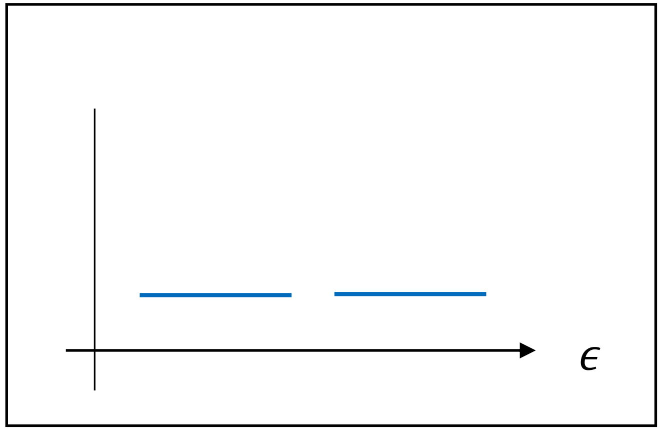
$$L_{d-1}\Phi = \partial_\omega \quad \Rightarrow \quad \Phi = (L_{d-1} + J_{d-1})^{-1}\partial_\omega \quad \square$$

High-dimensional information centrality

(X, R) simplicial network of dimension d (> 0)
Vertex set V with $|V| = n$, and $X_{d-1} = \binom{V}{d}$.
Effective conductance $C_\sigma = R_\sigma^{-1}$ for $\sigma \in \binom{V}{d+1}$.

For nonempty $A \subset V$, let $X_A = \{\sigma \in X_d \mid A \subset \sigma\}$
$\mathcal{I}_A$ = high-dimensional information centrality of A is the harmonic mean of C_σ for $\sigma \in X_A$:

$$\mathcal{I}_A = \binom{n-1}{d+1-|A|} \left(\sum_{\sigma \in X_A} \frac{1}{C_\sigma} \right)^{-1} .$$

Point Cloud	Barcode	L^1- persistence Landscape
		8.244295e-05
		≀≀
		8.272145e-05

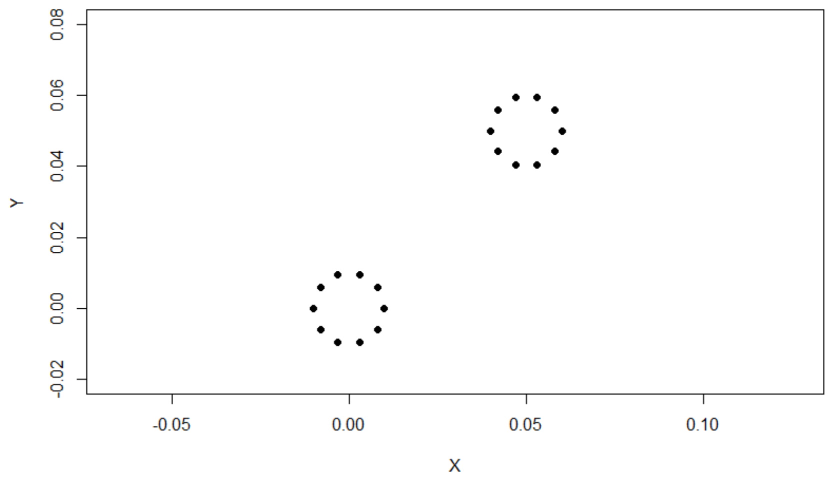
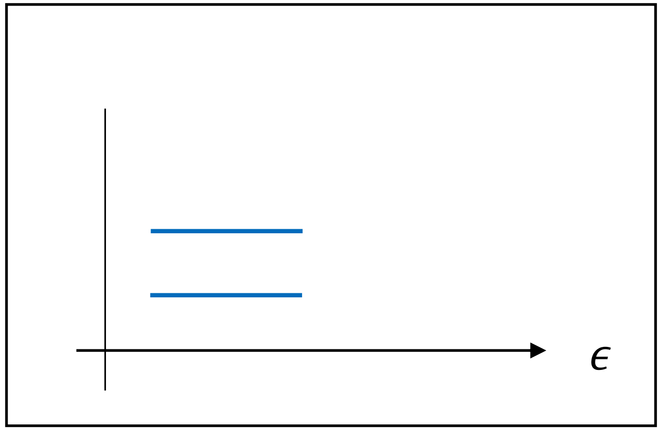
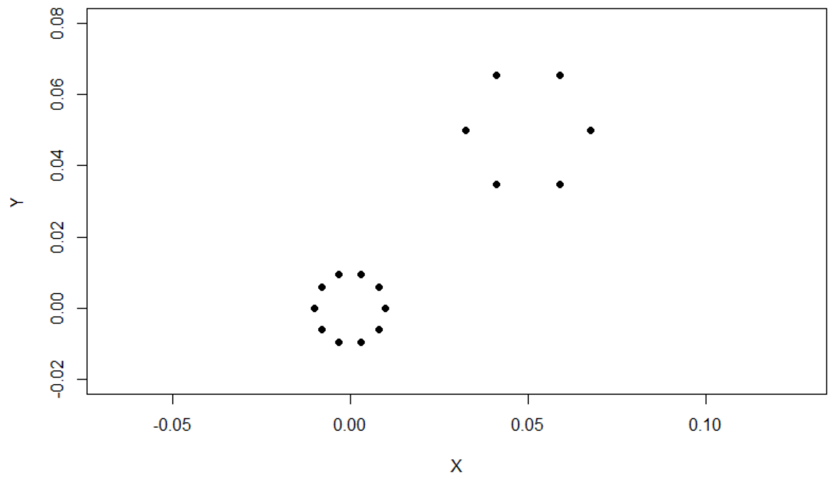
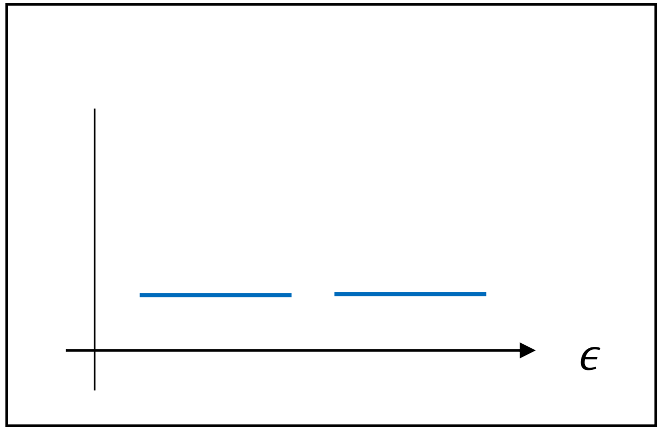
Modified Persistence Landscape

$$f_{[a,b]}(t) \coloneqq \begin{cases} 0 & if\ t \notin [a,b] \\ t-a & if\ a \leq t \leq \frac{a+b}{2} \\ b\ -t & if\ \frac{a+b}{2} \leq t\ \ \leq b \end{cases} \quad \Rightarrow \quad f_{[a,b]}(t) \coloneqq \begin{cases} 0 & if\ t \notin [a,b] \\ \frac{a+b}{b-a}t - a\frac{a+b}{b-a} & if\ a \leq t \leq \frac{a+b}{2} \\ b\frac{a+b}{b-a} - \frac{a+b}{b-a}t & if\ \frac{a+b}{2} \leq t\ \ \leq b \end{cases}$$

For a collection $[a_i, b_i]$ of intervals in a barcode B, we get a sequence λ_k of functions, for $k \in \mathbb{N}$:

$$\lambda_k(t) \coloneqq k-th\ largest\ \{f_{[a_i,b_i]}(t)\},$$

($\lambda_k(t)$ = 0 if $k > number\ of\ intervals$)

Point Cloud	Barcode	L^1-modified persistence Landscape
		0.0001618034
		$\neq$
		0.0002357817

Various network centralities

- Betweenness centrality (Importance as a **bridge**)

$$B(v) = \sum_{\substack{s,t \in V \\ s \neq v \neq t}} \frac{\#\ of\ shortest\ paths\ s \rightarrow t\ passing\ v}{\#\ of\ shortest\ paths\ s \rightarrow t}$$

- Closeness centrality (How close the node is to others)

$$\mathrm{C}(v) = \frac{1}{\sum_u d(u,v)}$$

- Eigenvector centrality (Links to "important" nodes are "important")

$$x(v) = \frac{1}{\lambda} \sum_{v \sim t} x(t) \quad \longrightarrow \quad (Matrix\ form)\ Ax = \lambda x$$

- PageRank (named after **Larry Page** in **Google™**)

$$PR(v) = k \sum_{u \rightarrow v} \frac{PR(u)}{outdeg(v)} \quad \longrightarrow \quad (Matrix\ form)\ R = kMR$$

M. Alstyne and E. Brynjolfsson: Cyberbalkanization

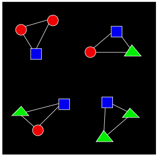


Figure 1.A

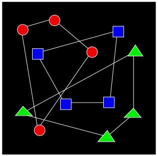


Figure 1.B

IT can reconstitute geographic communities (locations in 1.A) by research discipline (colors in 1.B).

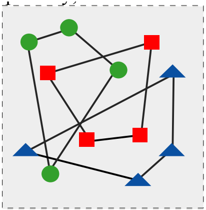


Figure 4.A

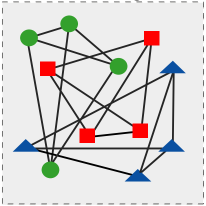


Figure 4.B

Increasing channel capacity does not necessarily increase the measure of integrated affiliations.

Sources

Widening Access and Narrowing Focus: Could the Internet Balkanize Science?, Science (1996)

Global Village or Cyber Balkans? Modeling and Measuring the Integration of Electronic Communities, Management Sci. (2005)

4차 산업혁명 미래 신기술 : 새로운 신용 정보 창출

위상수학 Topology

조합론 Combinatorics

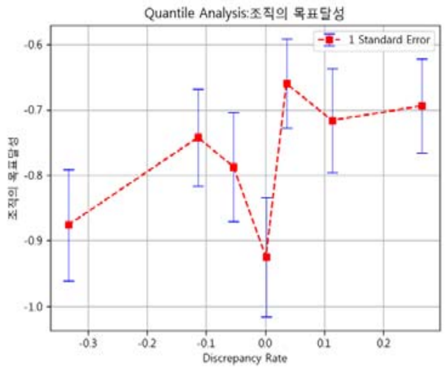
+

=

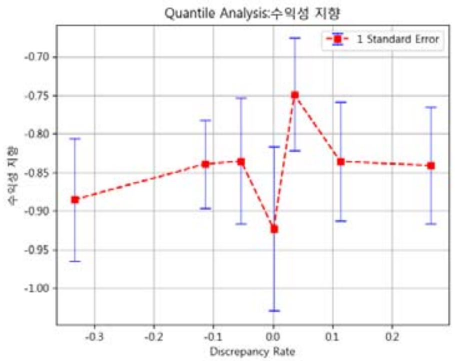
씨앗 네트워크 프로젝트

구성원들 사이의 씨앗(**선물)**을 통하여
새로운 금융·신용 지표를 창출하는 프로젝트

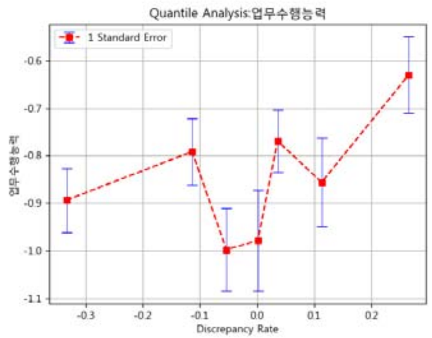
1. 블록체인을 이용한 신용정보의 축적
2. 금융 거래 외에서 창출되는 신용 지표
3. 스스로 신용정보를 관리하는 구조



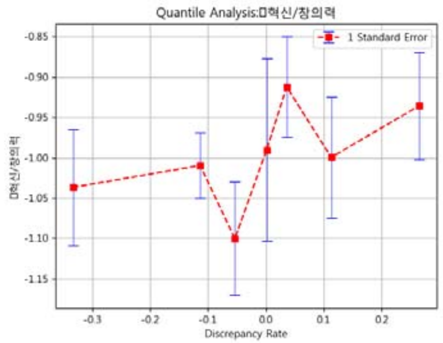
Statistics	Value
Correlation	15%
(P-value)	9%
2 sample t-test (P-value)	12%



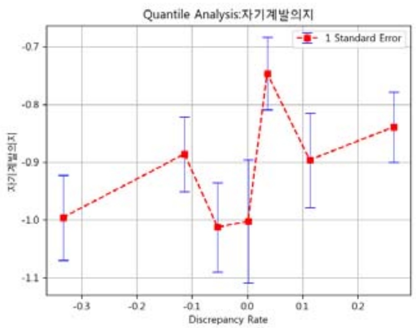
Statistics	Value
Correlation	4%
(P-value)	62%
2 sample t-test (P-value)	70%



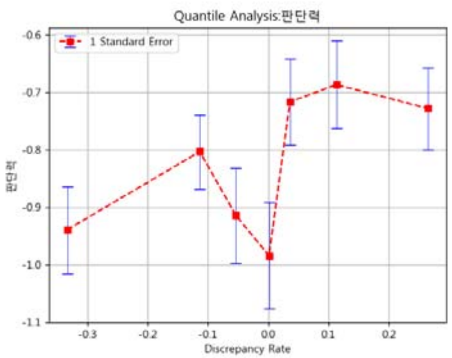
Statistics	Value
Correlation	**17%**
(P-value)	**4%**
2 sample t-test (P-value)	**2%**



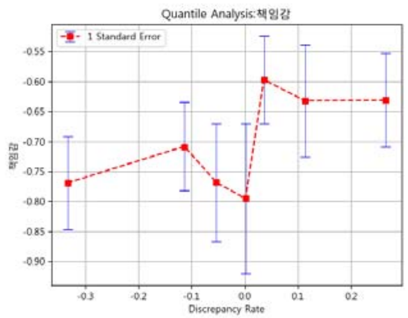
Statistics	Value
Correlation	13%
(P-value)	11%
2 sample t-test (P-value)	32%



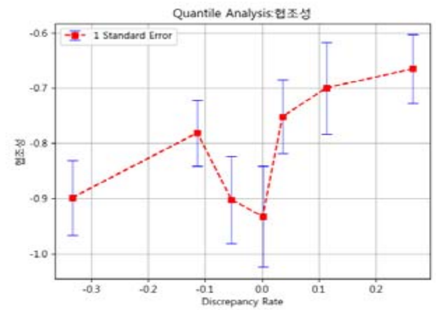
Statistics	Value
Correlation	16%
(P-value)	6%
2 sample t-test (P-value)	12%



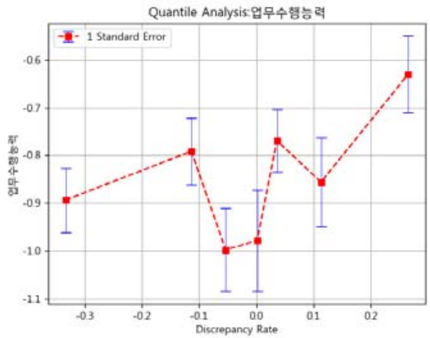
Statistics	Value
Correlation	**21%**
(P-value)	**1%**
2 sample t-test (P-value)	**5%**



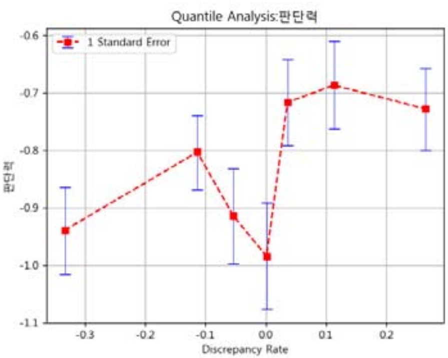
Statistics	Value
Correlation	13%
(P-value)	11%
2 sample t-test (P-value)	23%



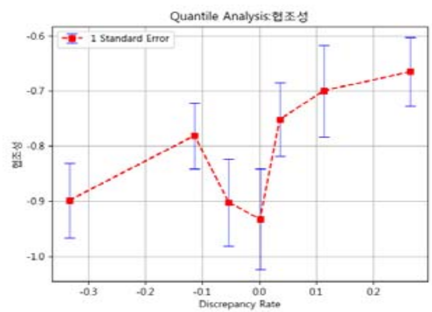
Statistics	Value
Correlation	21%
(P-value)	1%
2 sample t-test (P-value)	1%



Statistics	Value
Correlation	**17%**
(P-value)	**4%**
2 sample t-test (P-value)	**2%**



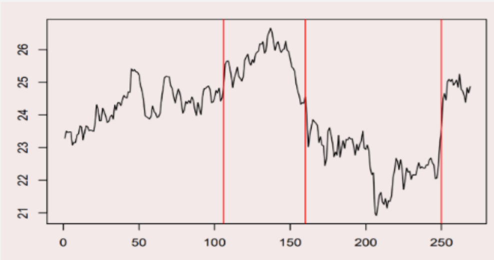
Statistics	Value
Correlation	**21%**
(P-value)	**1%**
2 sample t-test (P-value)	**5%**



Statistics	Value
Correlation	21%
(P-value)	1%
2 sample t-test (P-value)	1%

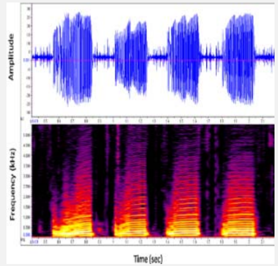
1. 영역 분할

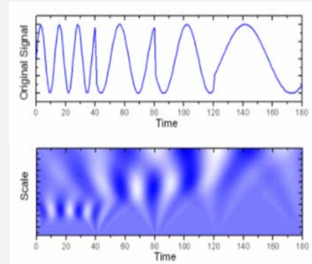
- 기본적으로 R peak을 이용 분할 (CPD 알고리즘 적용)

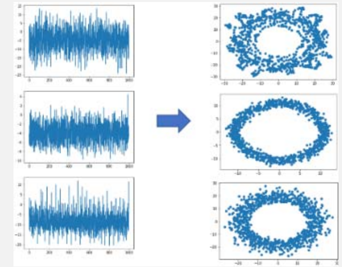


2. 데이터 특징 추출

- Fourier, wavelet 변형 및 Taken's embedding 사용
- Local high order statistics, Hermite basis decomposition 등 적용

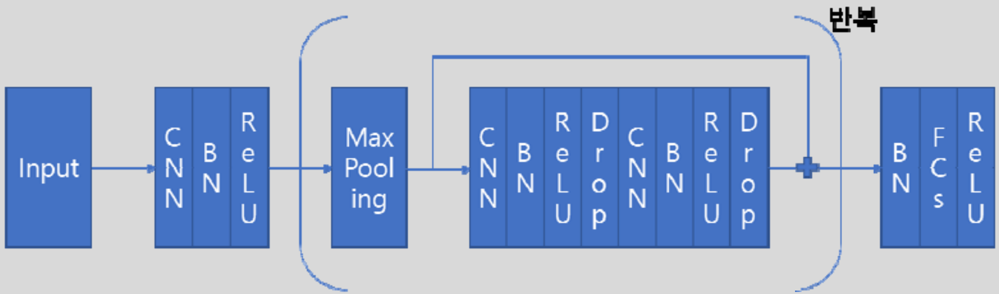






3. DNN

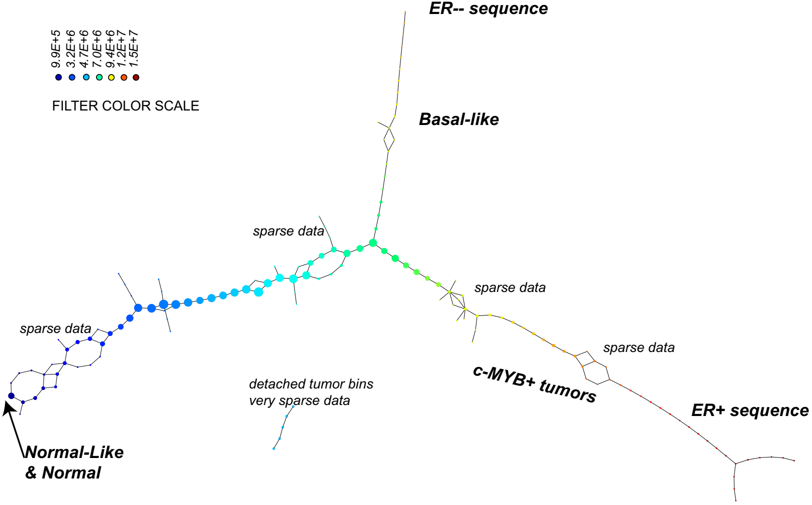
- DNN을 통해 P, Q, R, S, T Peak의 정보 추출
- DNN 결과에서 local extreme point를 찾는 알고리즘으로 최종 peak 결정



4. 부정맥 판독 알고리즘 구현

- P, Q, R, S, T peak 정보를 기반으로 시계열 데이터를 표현
- peak 정보를 바탕으로 부정맥 세부질환 진단

Nicolau, Levine, and Carlsson (2011)



f: L^p norm on DSGA-transformed matrix $\qquad d$: correlation

Identification of type 2 diabetes subgroups through topological analysis of patient similarity (2015)