School and Workshop on
Topological Data Analysis

Data Types

Methods for Reconstruction of Data Sets of Different Types

Hosein Masoomy

Computational Cosmology Group (CCG) [ccg.sbu.ac.ir]
&
Center for Complex Networks and Social Data Science (SSNSD) [ccnsd.ir]
@
Department of Physics, Shahid Beheshti University (SBU) [sbu.ac.ir]

August 24, 2022

Overview

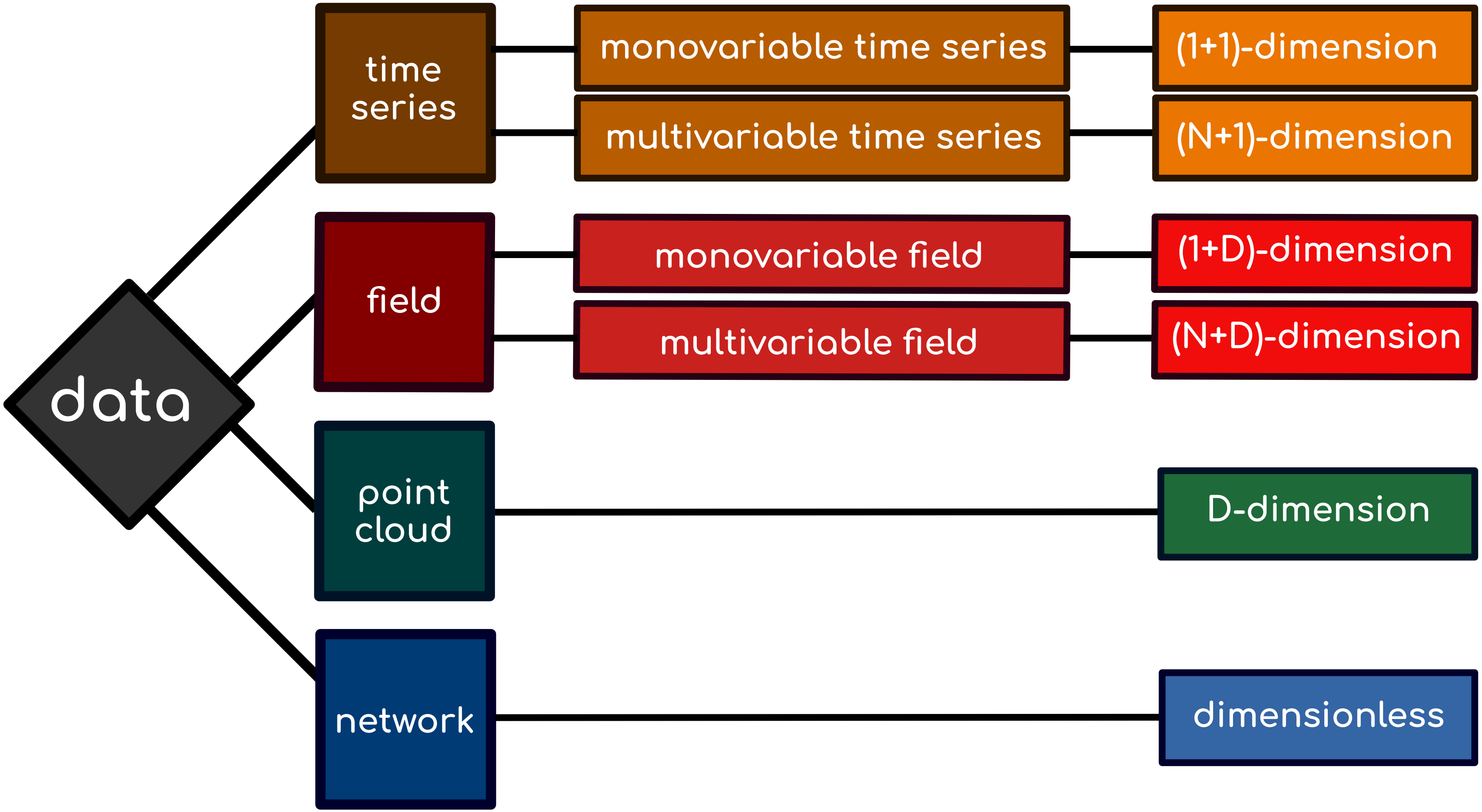
Data Types

Time Series
Field
Point Cloud
Network (Graph)

Methods for Reconstruction of Data Sets of Different Types

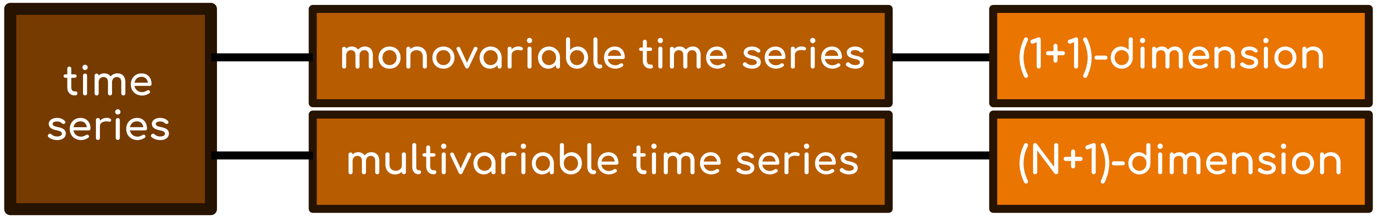
Time Delay Embedding (TDE)
Recurrent Plot (RP)
Visibility Graph (VG)
State Space (SS)
Correlation Network (CN)
Recurrent Network (RN)
Excursion Sets (ES)

Data Types



data
time series
monovariable time series
(1+1)-dimension
multivariable time series
(N+1)-dimension
field
monovariable field
(1+D)-dimension
multivariable field
(N+D)-dimension
point cloud
D-dimension
network
dimensionless

Data Types / Time Series



Data Types / Time Series

time series

monovariable time series

(1+1)-dimension

T = length of time series

$\mathcal{T} \subseteq \mathbb{R}^+ = [0, +\infty)$ domain

$X \subseteq \mathbb{R}$ range

$x : \mathcal{T} \rightarrow X$ map

$\vec{x} \equiv \left(x(t_i) \right)_{i=1}^{\mathcal{T}}$

H. Kantz, “Nonlinear time series analysis”, Vol. 7, Cambridge university press, (2004).

Data Types / Time Series

time series

- **monovariable time series** — **(1+1)-dimension**

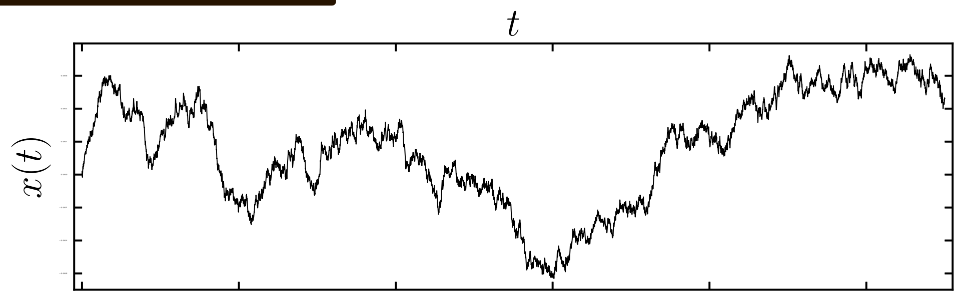
T = length of time series

$\mathcal{T} \subseteq \mathbb{R}^+ = [0, +\infty)$ domain

$X \subseteq \mathbb{R}$ range

$x : \mathcal{T} \to X$ map

$$\vec{x} = \left(x(t_i)\right)_{i=1}^{T}$$



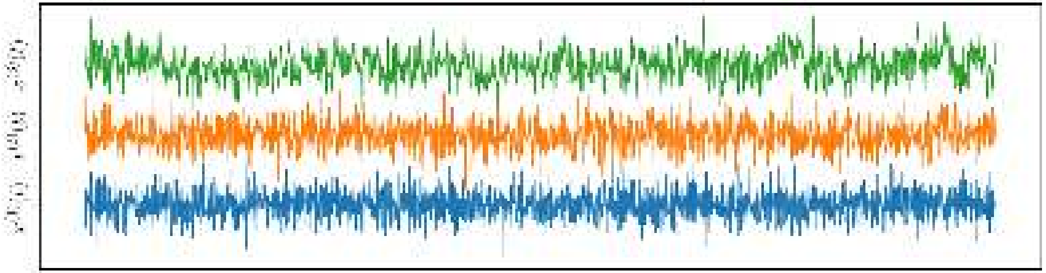
T=10000

- **multivariable time series** — **(N+1)-dimension**

N = number of degrees of freedom
(number of dependent variables)

To express the evolution of the system, we need N maps:

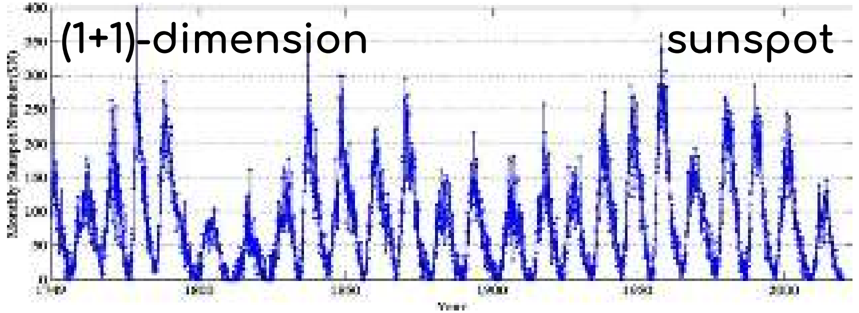
$$\mathcal{X} = \left\{ \vec{x}^{(j)} \;\middle|\; \vec{x}^{(j)} = \left(x^{(j)}(t_i)\right)_{i=1}^{T} \right\}_{j=1}^{N} = \left\{ x^{(j)}(t_i) \right\}_{i,j=1}^{T,N}$$



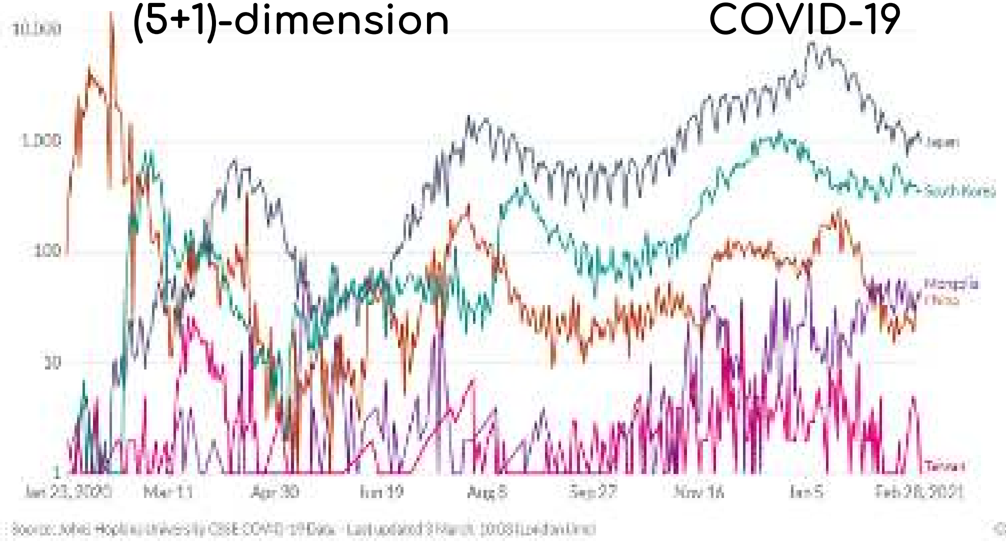
T=1000, N=3

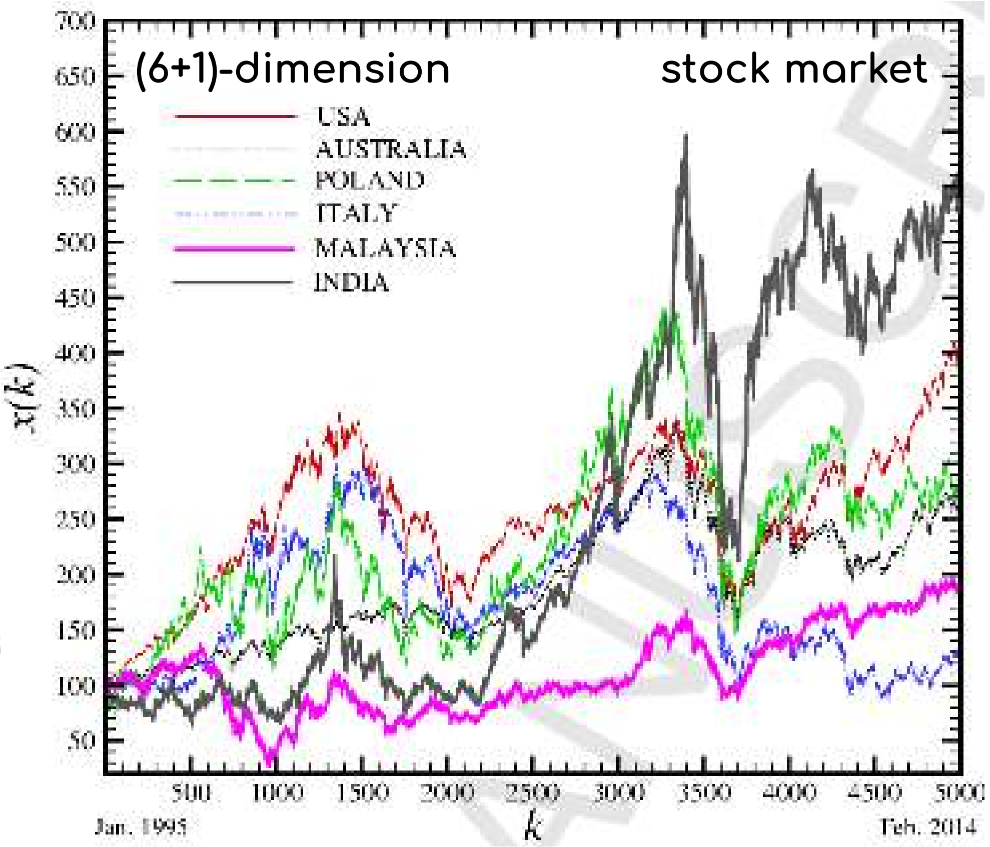
H. Kantz, “Nonlinear time series analysis”, Vol. 7, Cambridge university press, (2004).

Data Types / Time Series : real data



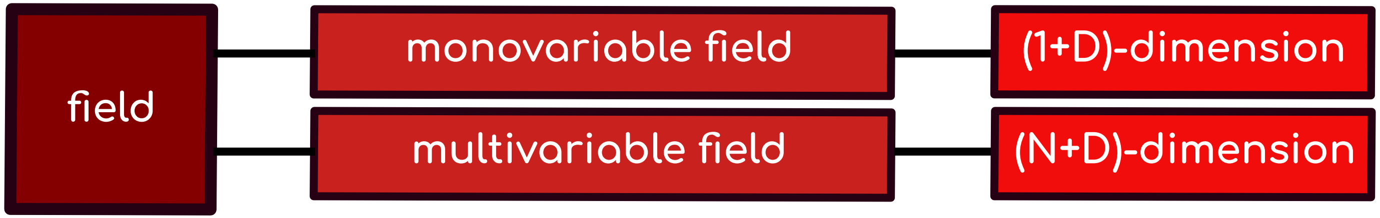
Panigrahi, Sibarama, Radha Mohan Pattanayak, Prabira Kumar Sethy, and Santi Kumari Behera. "Forecasting of sunspot time series using a hybridization of ARIMA, ETS and SVM methods." Solar Physics 296, no. 1 (2021): 1-19.





Ferreira, Paulo, Andreia Dionísio, and S. M. S. Movahed. "Assessment of 48 stock markets using adaptive multifractal approach." Physica A: Statistical Mechanics and its Applications 486 (2017): 730-750.

Data Types / Field



Data Types / Field

field

monovariable field

(1+D)-dimension

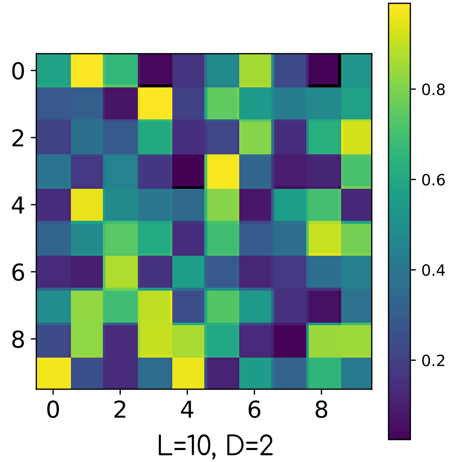
$$\mathcal{F} : \Pi \rightarrow \mathbb{R}$$

L = length of field

$$\Pi \subseteq \mathbb{R}^D$$

domain

D = dimension of domain
(number of independent variables)



L=10, D=2

R. J. Adler, “The geometry of random fields”, Society for Industrial and Applied Mathematics, (2010).

Data Types / Field

field

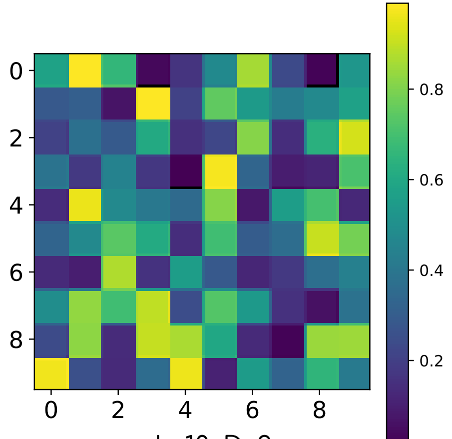
monovariable field — (1+D)-dimension

$$\mathcal{F} : \Pi \rightarrow \mathbb{R}$$

L = length of field

$$\Pi \subseteq \mathbb{R}^D$$ domain

D = dimension of domain (number of independent variables)



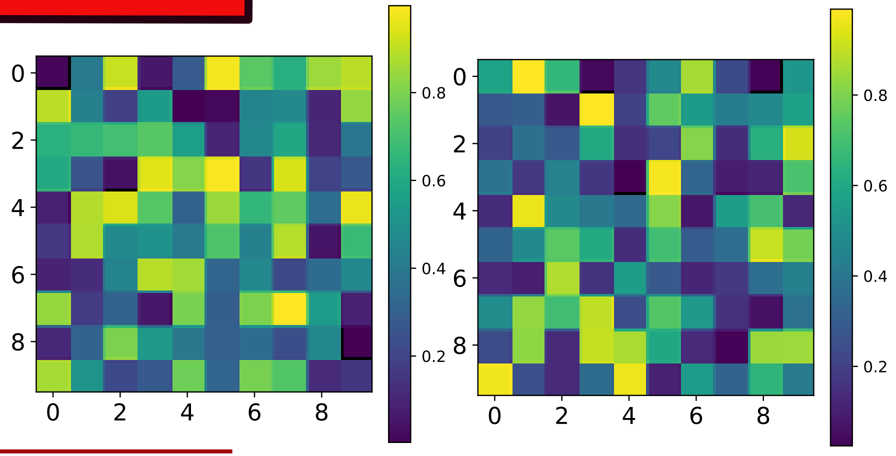
L=10, D=2

multivariable field — (N+D)-dimension

N = number of degrees of freedom (number of dependent variables)

The system is determined by N maps:

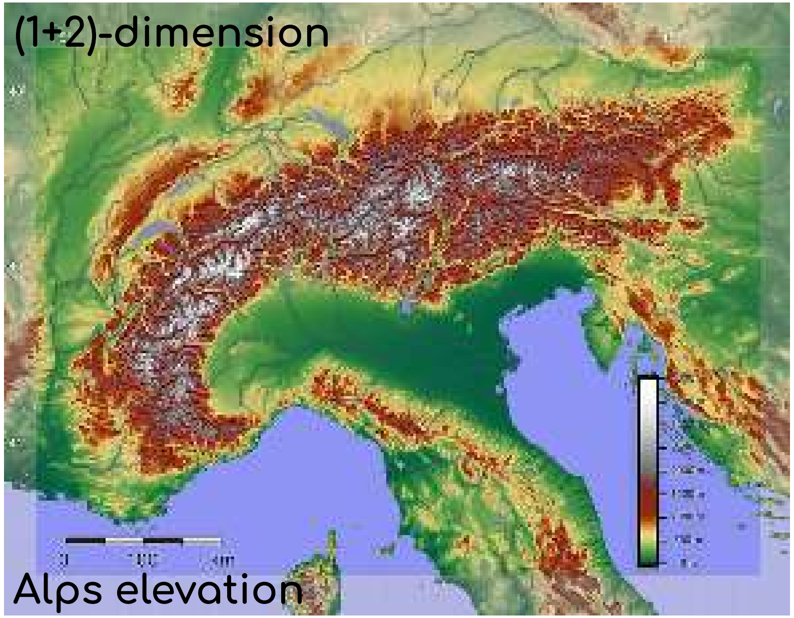
$$\mathfrak{F} = \left\{ \mathcal{F}_i \,\middle|\, \mathcal{F}_i : \Pi_i \rightarrow \mathbb{R} \,,\; \Pi_i \subseteq \mathbb{R}^D \right\}_{i=1}^{N}$$



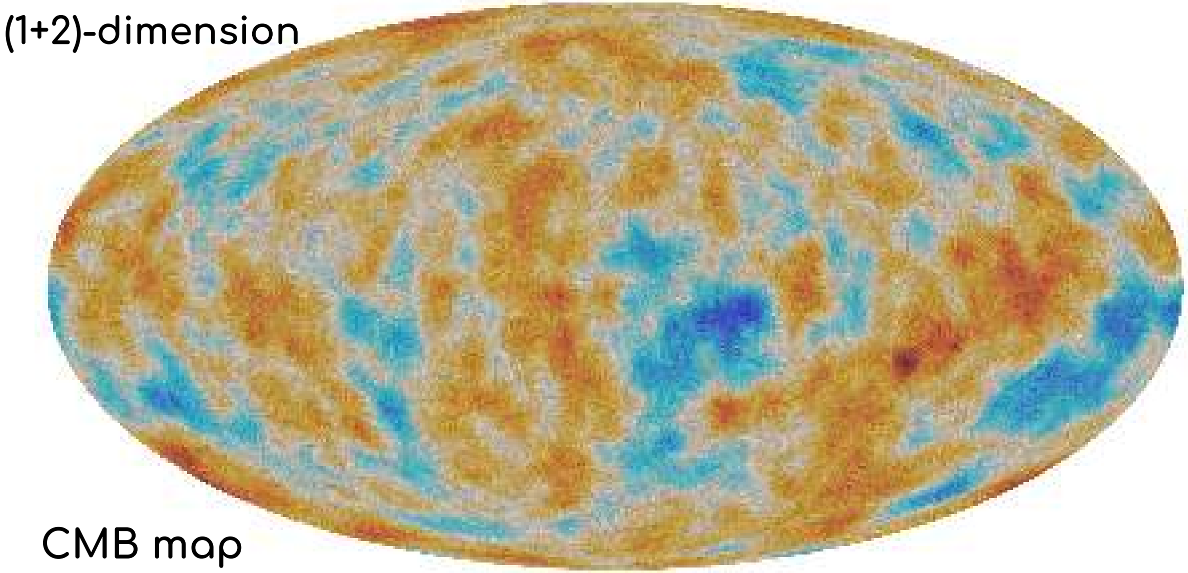
L=10, D=2, N=2

R. J. Adler, “The geometry of random fields”, Society for Industrial and Applied Mathematics, (2010).

Data Types / Field : real data

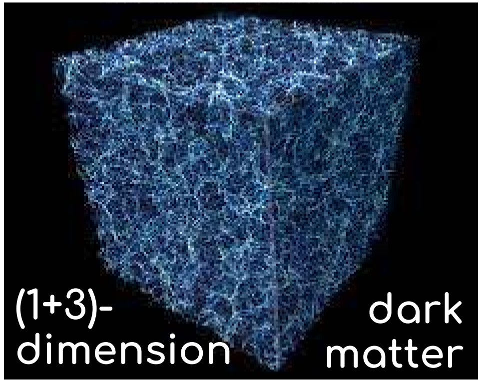


https://en.wikipedia.org/wiki/Alps



https://www.esa.int/ESA_Multimedia/Images/2015/02/Polarisation_of_the_Cosmic_Microwave_Background

https://en.wikipedia.org/wiki/The_Last_Supper_%28Leonardo%29



https://futurism.com/20-percent-universes-normal-matter-exists-dark-cosmic-voids

Data Types / Point Cloud

point cloud — D-dimension

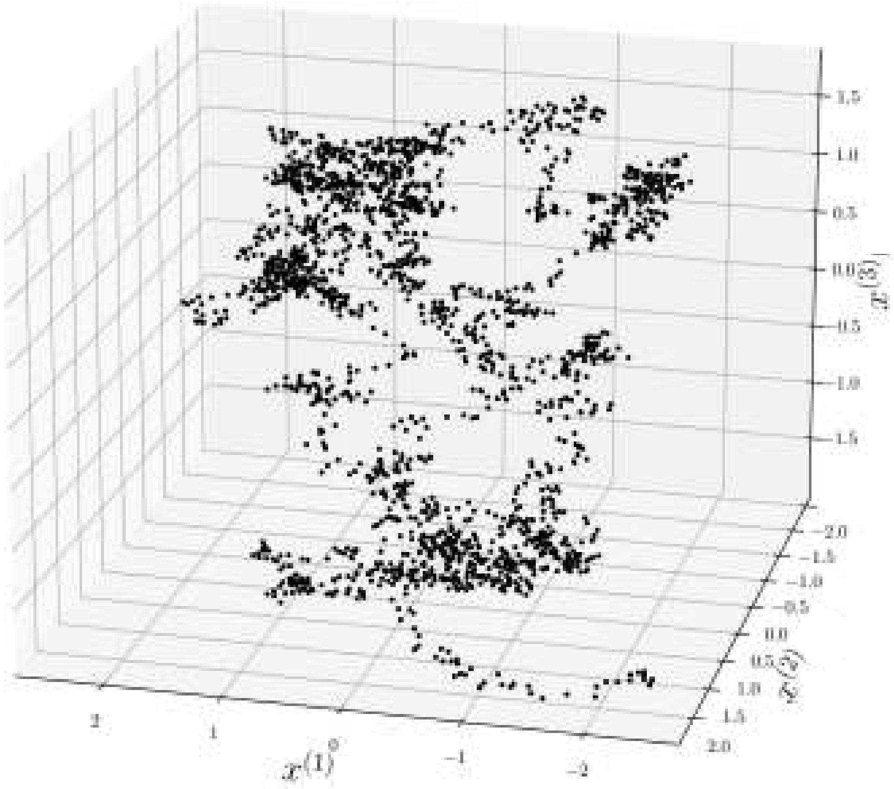
$$\mathbb{X} = \left\{ x_i \,\middle|\, x_i \equiv (x_i^{(d)})_{d=1}^{D} \,,\; x_i^{(d)} \in \mathbb{R} \right\}_{i=1}^{N \neq \infty}$$

x_i = ith point (ith element of point cloud)

$x_i^{(d)}$= dth element of ith point

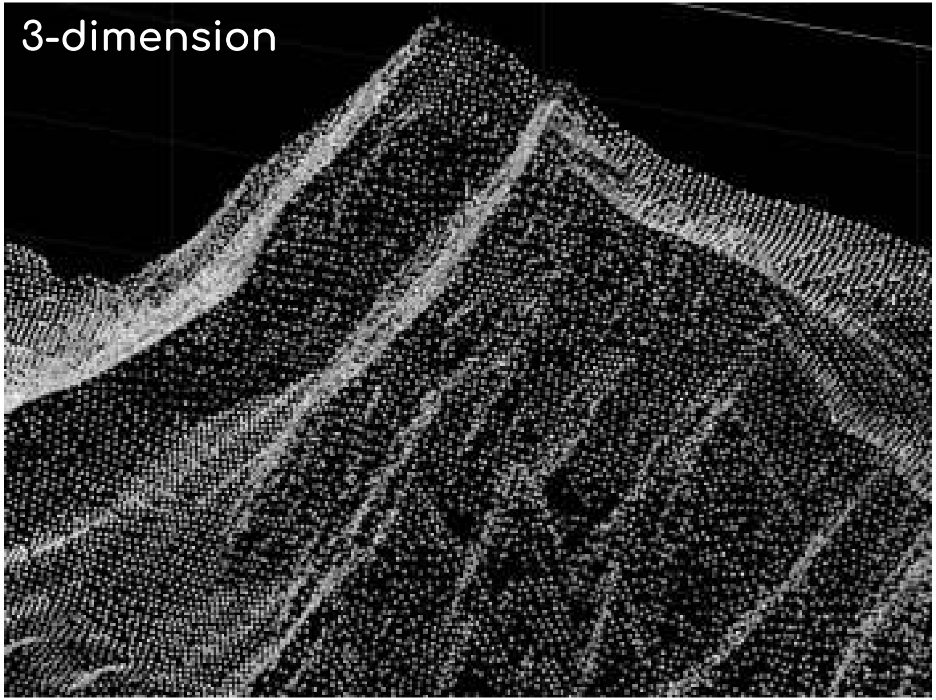
D = dimension of point cloud

N = size of point cloud (number of data points)



N=1000, D=3

Data Types / Point Cloud : real data



https://www.geo.tuwien.ac.at/downloads/pg/pctools/publish/pointCloudThinOut/html/pointCloudThinOut.html

https://www.wired.com/2014/09/shaun-kardinal-flying-formation/

Data Types / Network

network — dimensionless

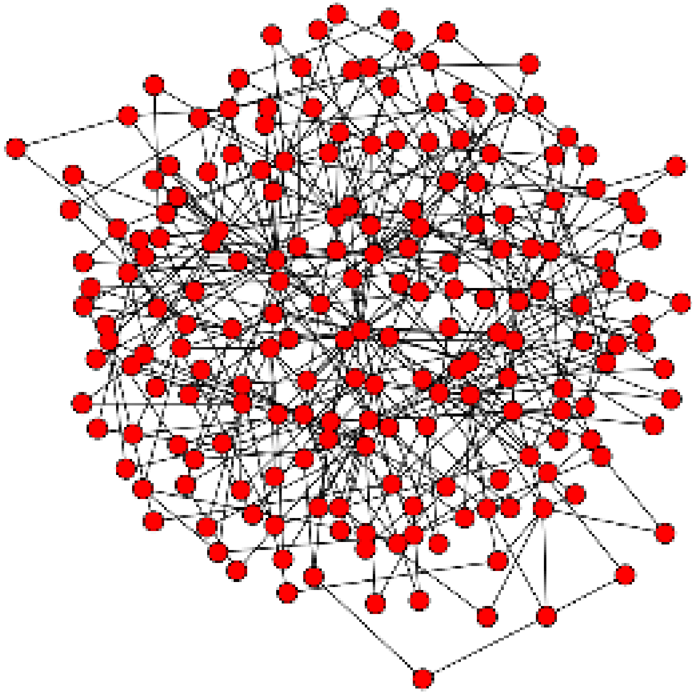
$G = (V, E, w)$ graph (network)

$V = \{v_i\}_{i=1}^{N}$ vertex (node) set
N = network size (number of nodes)

$E = V \times V$ edge (link) set (Cartesian product)
L = |E| = number of links

$w : E \rightarrow \mathbb{R}$ weight function (link map)

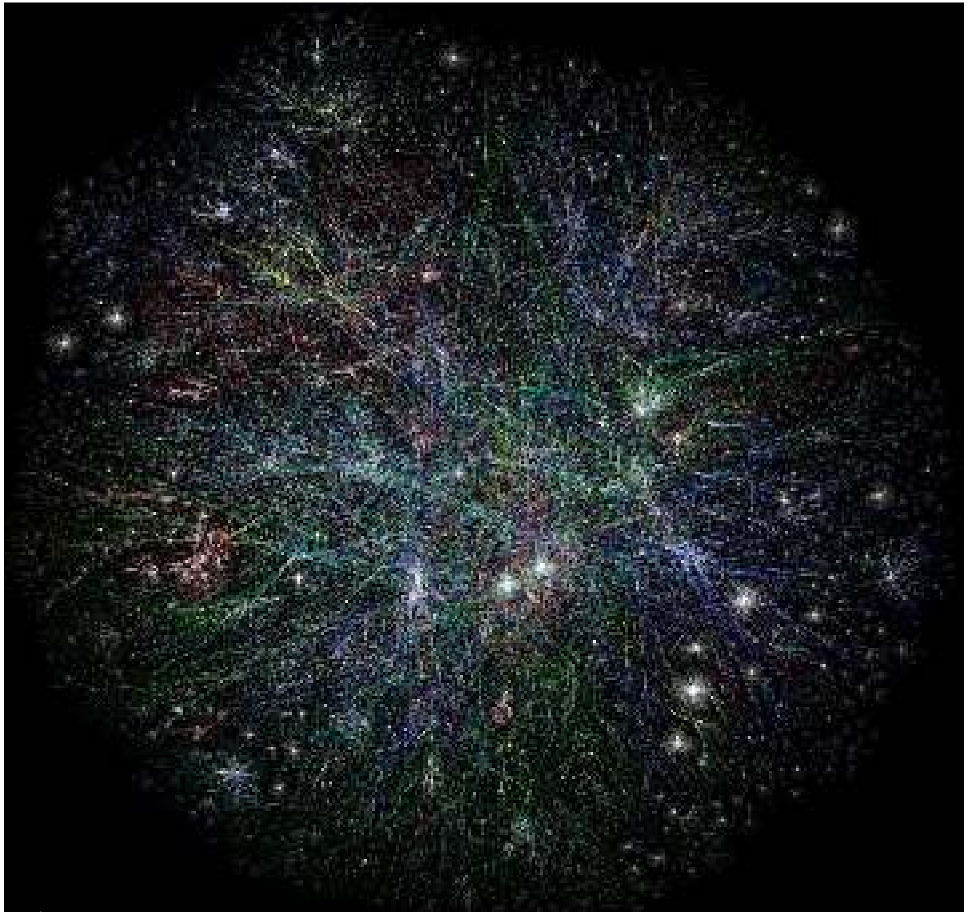
$$w(e_{ij}) \equiv w((v_i, v_j)) = w_{ij}$$



N=100, L=150

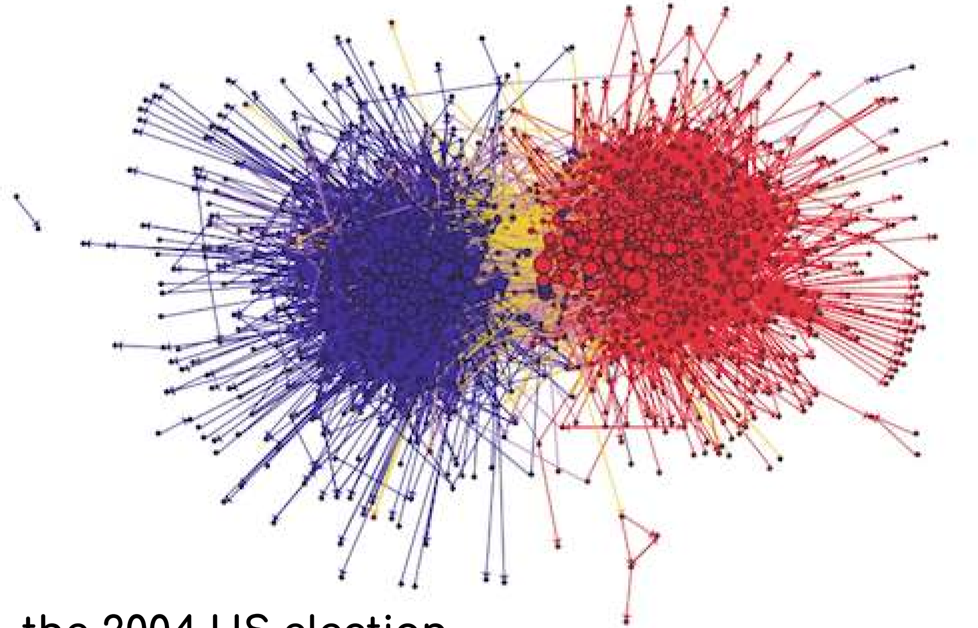
M. Newman, “Networks”, Oxford university press, (2018).

Data Types / Network : real data



the internet

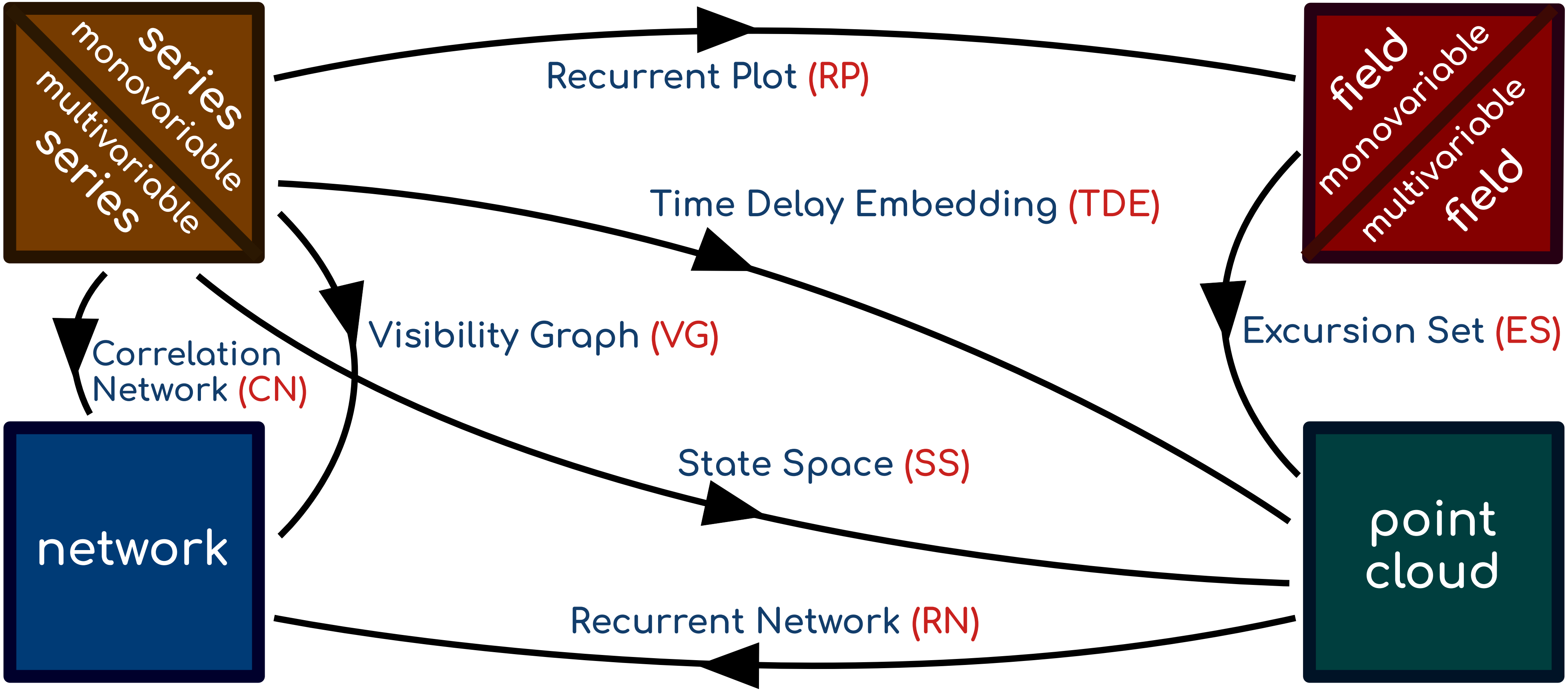
https://www.kaspersky.com/blog/amazing-internet-maps/10441/



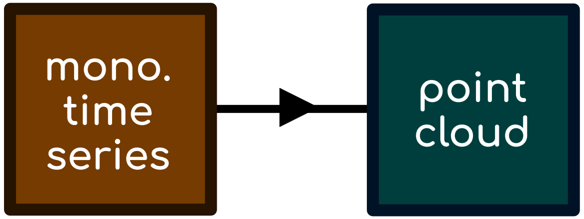
the 2004 US election

Adamic, Lada A., and Natalie Glance. "The political blogosphere and the 2004 US election: divided they blog." In Proceedings of the 3rd international workshop on Link discovery, pp. 36-43. 2005.

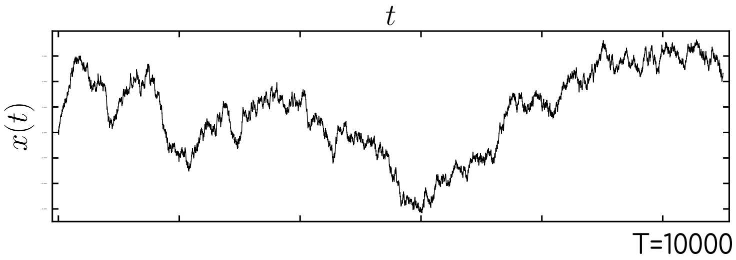
Methods for Reconstruction of Data Sets of Different Types



Methods ... / Time Delay Embedding (TDE)

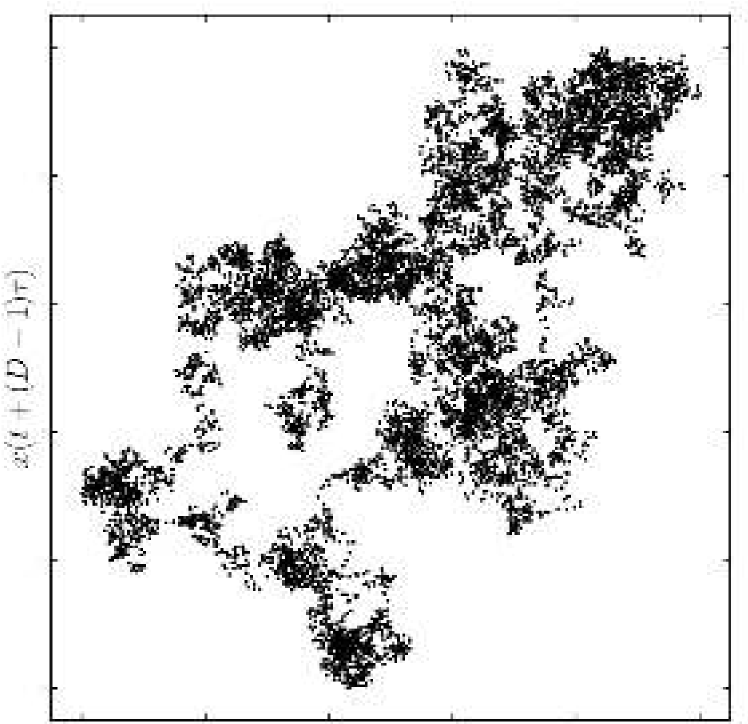


$$\vec{x} \equiv \left(x(t_i)\right)_{i=1}^{T}$$



D = embedding dimension
(dimension of reconstructed point cloud)

τ = time-delay



N=9000, D=2, \tau=1000

$$\mathbb{X}(\vec{x}) \equiv \left\{x_i \in \mathbb{R}^D \;\middle|\; x_i \equiv \left(x(t_i), x(t_i+\tau), x(t_i+2\tau), \dots, x(t_i+(D-1)\tau)\right)\right\}_{i=1}^{T-(D-1)\tau}$$

$|\mathbb{X}(\vec{x})| = |\vec{x}| - \left(dim(\mathbb{X}(\vec{x})) - 1\right)\tau$ $|\mathbb{X}(\vec{x})|$ = N $dim(\mathbb{X}(\vec{x}))$ = D $|\vec{x}|$ = T

F. Takens, "Detecting strange attractors in turbulence", In Dynamical systems and turbulence, Warwick 1980, pp. 366-381. Springer, Berlin, Heidelberg, (1981).

Methods … / Recurrent Plot (RP)

$$\vec{x} \equiv \left(x(t_i)\right)_{i=1}^{T}$$

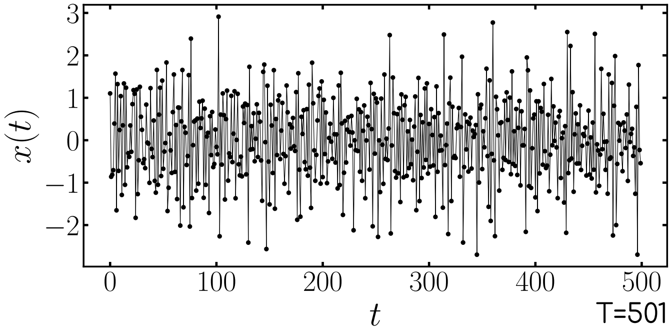
TDE : (D, τ)

$$\mathbb{X}(\vec{x}) \equiv \left\{x_i \in \mathbb{R}^D\right\}_{i=1}^{T-(D-1)\tau}$$

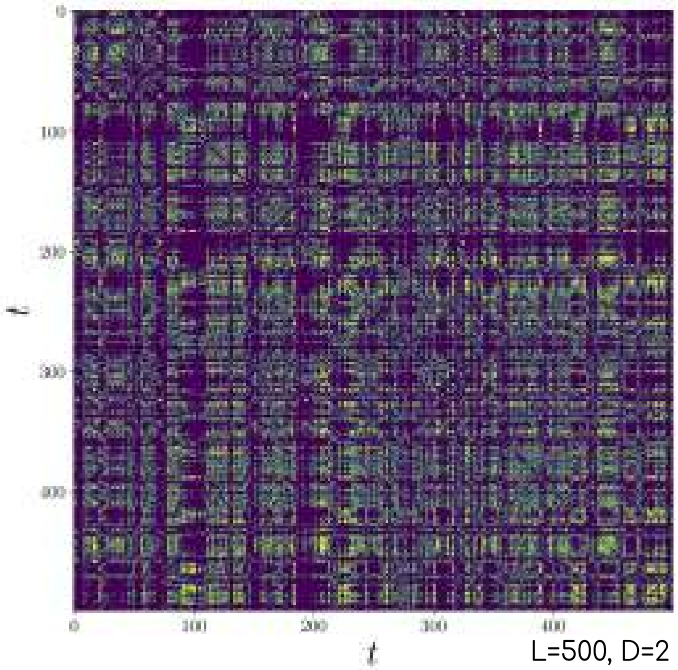
N = $T-(D-1)\tau$	size
D	dimension

$$\mathcal{F}_{ij}^{(\epsilon)} = \Theta(\epsilon - d(x_i, x_j))$$

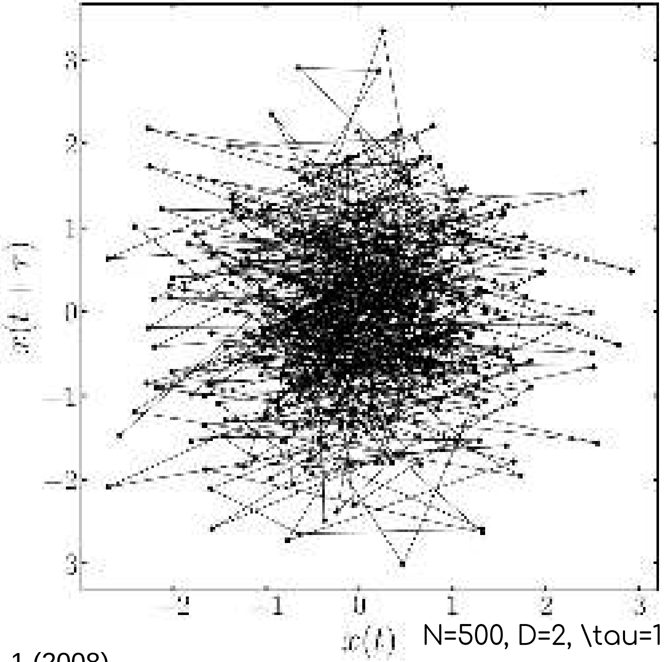
L = N	length
D	dimension



T=501



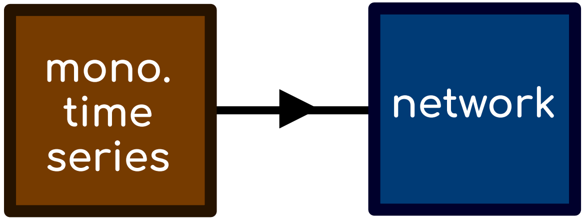
L=500, D=2

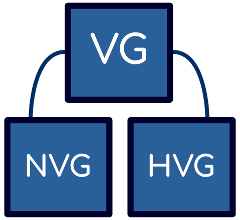


N=500, D=2, \tau=1

N. Marwan, "A historical review of recurrence plots", The European Physical Journal Special Topics 164, no. 1 (2008).

Methods ... / Visibility Graph (VG)



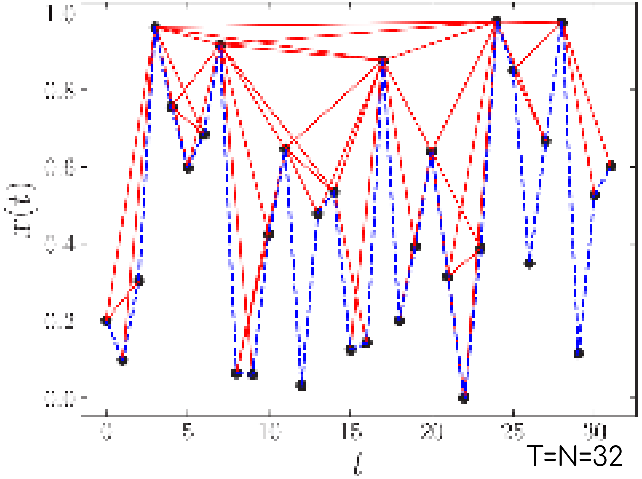


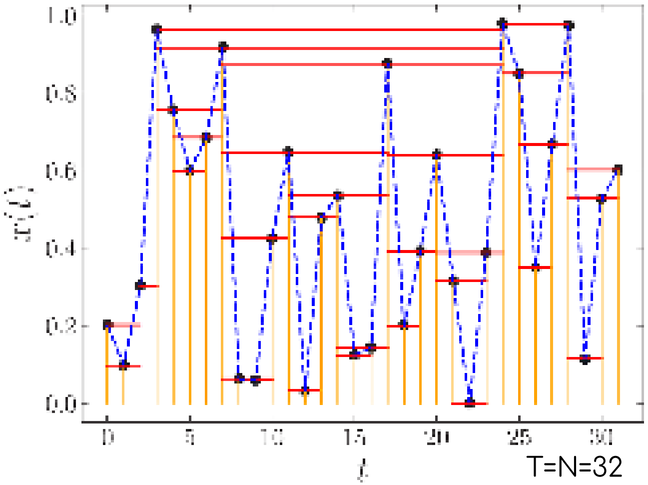
$$\vec{x} \equiv \left(x(t_i)\right)_{i=1}^{T}$$ time series $$G = (V, E, w)$$ network

$$f : V \equiv \{v_i\}_{i=1}^{N} \leftrightarrow \mathcal{T} \equiv (t_i)_{i=1}^{T}, \quad f(v_i) = t_i$$ bijection: N = T

$$w_{ij}^{(BN)} \equiv \begin{cases} 1, & |f(v_i) - f(v_j)| = 1 \\ \prod_{k=i+1}^{j-1} \Theta(s_{ij} - s_{ik}), & |f(v_i) - f(v_j)| > 1 \end{cases}$$ weight function

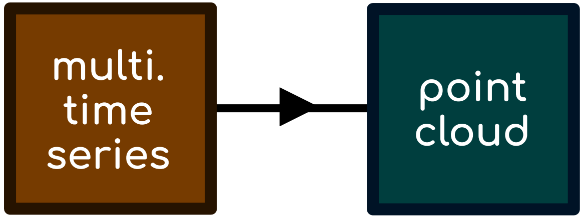
$$s_{ij} \equiv \frac{x(f(v_j)) - x(f(v_i))}{f(v_j) - f(v_i)}$$ slope of the visibility line between ith and jth data points





L. Lacasa, et al, "From time series to complex networks: The visibility graph", Proceedings of the National Academy of Sciences 105, no. 13 (2008).

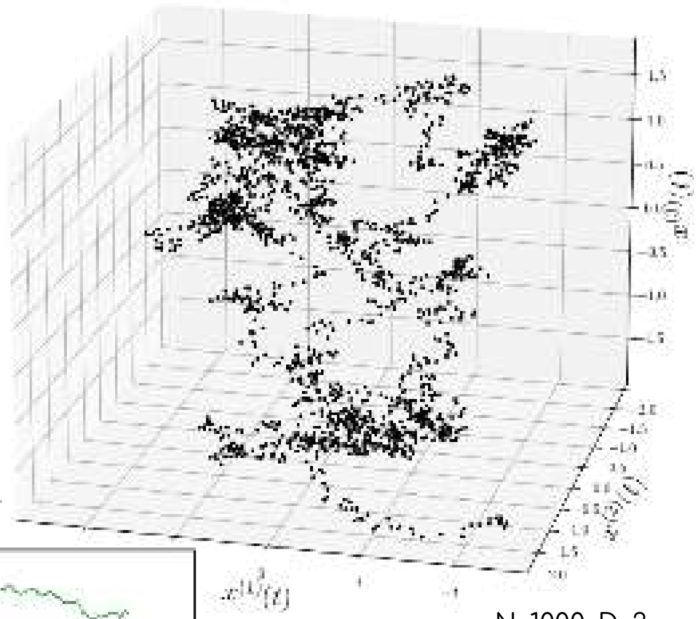
Methods ... / State Space (SS)



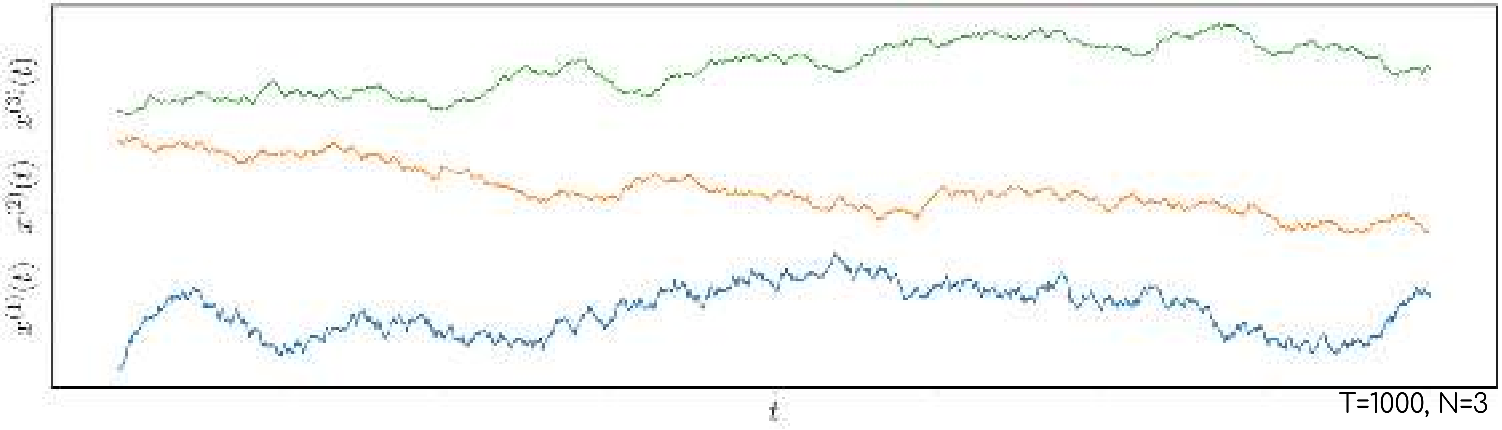
$$\mathcal{X} = \left\{ \vec{x}^{(j)} \,\middle|\, \vec{x}^{(j)} = \left(x^{(j)}(t_i) \right)_{i=1}^{T} \right\}_{j=1}^{N}$$

N = number of dependent variables
T = length of time series

$$\mathbb{X}(\mathcal{X}) = \left\{ \vec{x}(t_i) = \left(x^{(1)}(t_i), x^{(2)}(t_i), \ldots, x^{(N)}(t_i) \right) \in \mathbb{R}^N \,\middle|\, x^{(j)}(t_i) \in \mathbb{R} \right\}_{i=1}^{T}$$



N=1000, D=3

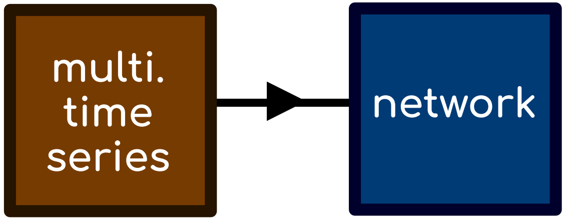


T=1000, N=3

$D(\mathbb{X}) = N(\mathcal{X})$

$N(\mathbb{X}) = T(\mathcal{X})$

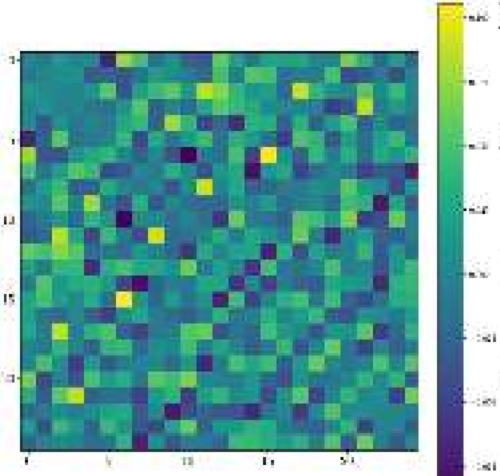
Methods ... / Correlation Network (CN)



$$\mathcal{X} = \left\{ \vec{x}^{(j)} \,\middle|\, \vec{x}^{(j)} = \left(x^{(j)}(t_i) \right)_{i=1}^{T} \right\}_{j=1}^{N}$$

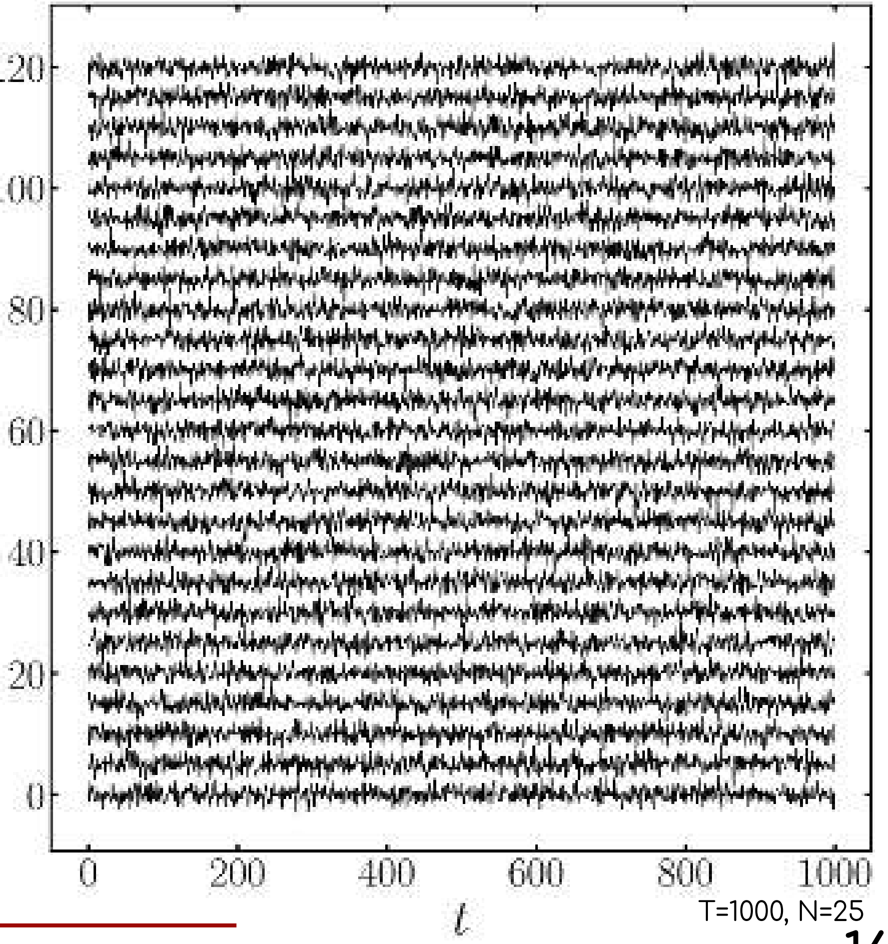
$$G = (V, E, w)$$

$$\mathcal{N} : \mathcal{X} \leftrightarrow V \qquad \text{N}(G) = \text{N}(\mathcal{X})$$



N=25

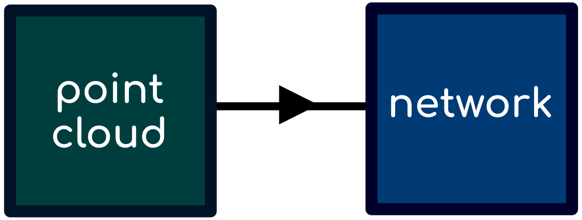
$$w_{jk} = \frac{\frac{1}{min\{T_j, T_k\}} \sum_{i=1}^{min\{T_j, T_k\}} \left(x^{(j)}(t_i) - \mu_{\vec{x}^{(j)}} \right)\left(x^{(k)}(t_i) - \mu_{\vec{x}^{(k)}} \right)}{\left[\frac{1}{T_j} \sum_{i=1}^{T_j} \left(x^{(j)}(t_i) - \mu_{\vec{x}^{(j)}} \right)^2 \quad \frac{1}{T_k} \sum_{i=1}^{T_k} \left(x^{(k)}(t_i) - \mu_{\vec{x}^{(k)}} \right)^2 \right]^{1/2}}$$



T=1000, N=25

Y. Zou, et al, "Complex network approaches to nonlinear time series analysis", Physics Reports 787 (2019).

Methods … / Recurrent Network (RN)

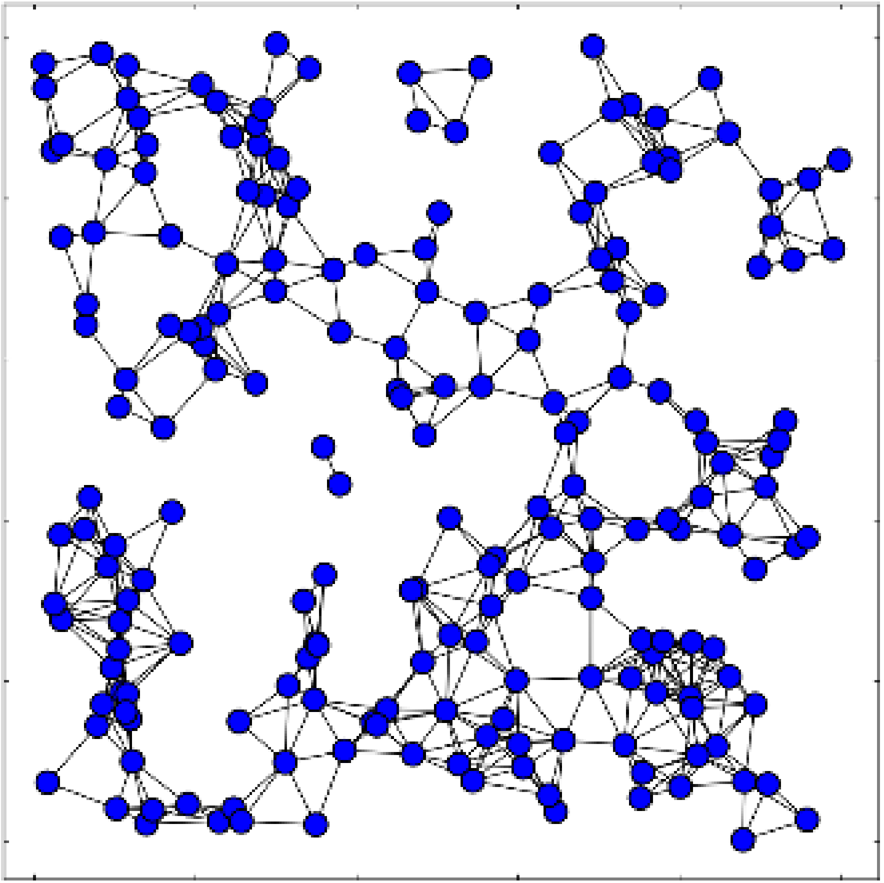


$$\mathbb{X} = \left\{ x_i \,\middle|\, x_i = (x_i^{(d)})_{d=1}^{D} \,,\, x_i^{(d)} \in \mathbb{R} \right\}_{i=1}^{N \not\to \infty}$$

$$G(\mathbb{X}, \epsilon) = (V, E, w^{(\epsilon)})$$

$f : \mathbb{X} \leftrightarrow V \quad ; \quad f(x_i) = v_i$ N(G) = N($\mathbb{X}$)

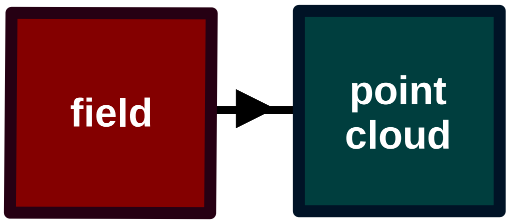
$$w^{(\epsilon)} : E \to \{0, 1\} \quad ; \quad w_{ij}^{(\epsilon)} = \Theta(\epsilon - d(x_i, x_j))$$



D=2, N=100, \epsilon=0.1, L=150

C. Chen, et al, "Recurrence network modeling and analysis of spatial data", Chaos: An Interdisciplinary Journal of Nonlinear Science 28, no. 8 (2018).

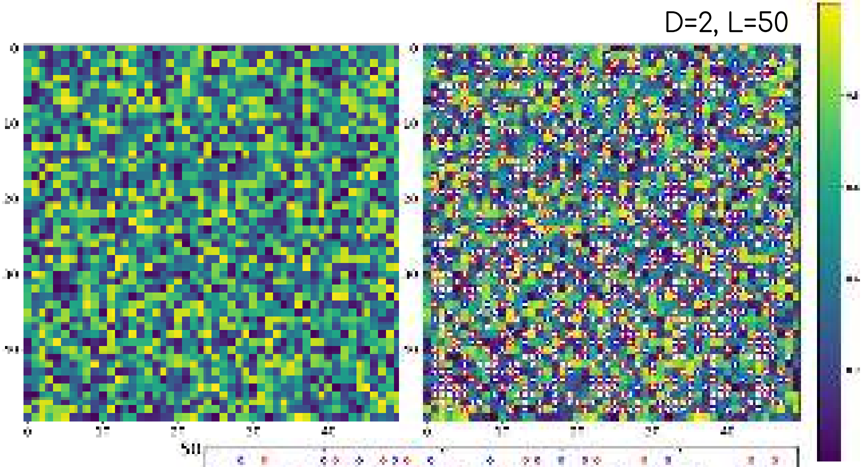
Methods … / Excursion Set (ES)



$\mathcal{F} : \Pi \rightarrow \mathbb{R}$ (D,L)

$$\mathbb{X}(\mathcal{F}) = \left\{ \pi \in \Pi \,\middle|\, \pi \in \mathcal{E}(\mathcal{F}) \right\}$$

for D=3: [local maxima and minima]

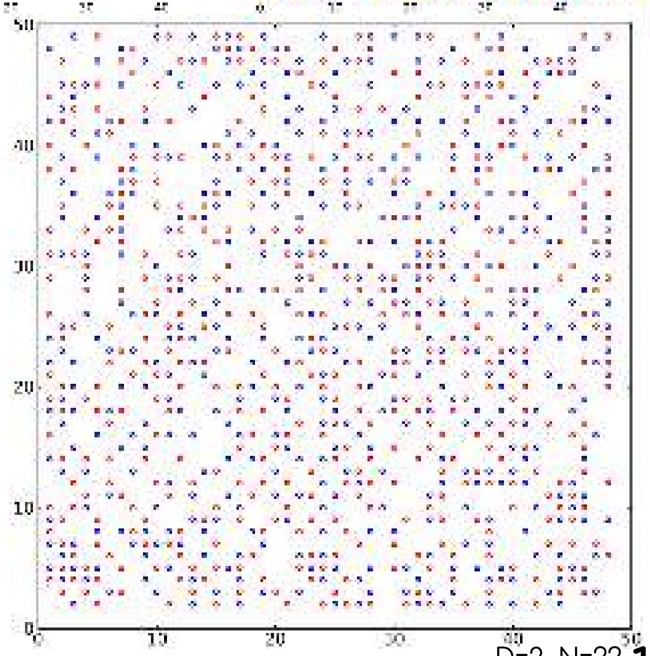


- $\mathcal{E}_{max}(\Pi) = \max(\Pi) = \left\{ (i,j,k) \in \Pi \,\middle|\, \mathcal{F}(i,j,k) > \max\{\mathcal{F}(\mathcal{N}(i,j,k))\} \right\}_{i,j,k=2}^{L-1}$
- $\mathcal{E}_{min}(\Pi) = \min(\Pi) = \left\{ (i,j,k) \in \Pi \,\middle|\, \mathcal{F}(i,j,k) < \min\{\mathcal{F}(\mathcal{N}(i,j,k))\} \right\}_{i,j,k=2}^{L-1}$

$$\begin{aligned}\mathcal{N}(i,j,k) = \{ & (i-1,j,k),(i+1,j,k),(i,j-1,k),(i,j+1,k),(i,j,k-1),(i,j,k+1),\\ & (i-1,j-1,k),(i-1,j+1,k),(i-1,j,k-1),(i-1,j,k+1),(i,j-1,k-1),\\ & (i,j-1,k+1),(i,j+1,k-1),(i,j-1,k+1),(i+1,j-1,k),(i+1,j+1,k),\\ & (i+1,j,k-1),(i+1,j,k+1),(i-1,j-1,k-1),(i-1,j+1,k-1),\\ & (i-1,j-1,k+1),(i-1,j+1,k+1),(i+1,j-1,k-1),(i+1,j+1,k-1),\\ & (i+1,j-1,k+1),(i+1,j+1,k+1)\} \subset \Pi(\mathcal{F})\end{aligned}$$

first neighbors

$D(\mathcal{F}) = D(\mathbb{X})$



References

[1] H. Kantz, “Nonlinear time series analysis”, Vol. 7, Cambridge university press, (2004).

[2] R. J. Adler, “The geometry of random fields”, Society for Industrial and Applied Mathematics, (2010).

[3] M. Newman, “Networks”, Oxford university press, (2018).

[4] F. Takens, "Detecting strange attractors in turbulence", In Dynamical systems and turbulence, Warwick 1980, pp. 366-381. Springer, Berlin, Heidelberg, (1981).

[5] N. Marwan, "A historical review of recurrence plots", The European Physical Journal Special Topics 164, no. 1 (2008).

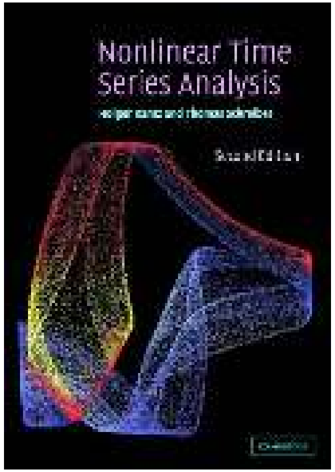
[6] L. Lacasa, et al, "From time series to complex networks: The visibility graph", Proceedings of the National Academy of Sciences 105, no. 13 (2008).

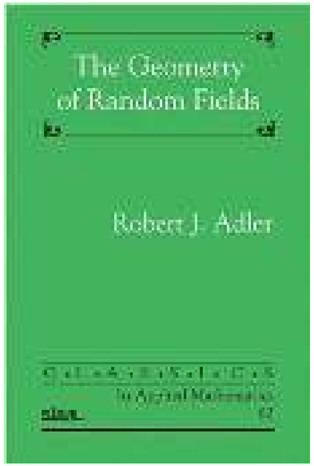
[7] Y. Zou, et al, "Complex network approaches to nonlinear time series analysis", Physics Reports 787 (2019).

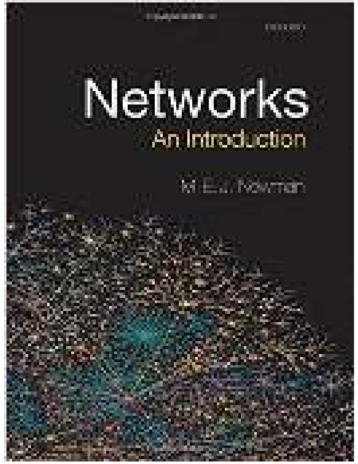
[8] C. Chen, et al, "Recurrence network modeling and analysis of spatial data", Chaos: An Interdisciplinary Journal of Nonlinear Science 28, no. 8 (2018).

[9] H. Masoomy, B. Askari, M. N. Najafi, and S. M. S. Movahed, "Persistent homology of fractional Gaussian noise", Physical Review E 104, no. 3, (2021).

[10] H. Masoomy, and M. N. Najafi, "The Visibility Graphs of Correlated Time Series Violate the Barthelemy's Conjecture for Degree and Betweenness Centralities", arXiv preprint arXiv:2112.07698, (2021).







Thanks :)

hoseingmasoomy@gmail.com

Designed by Hana Masoomy