Data Analysis with Information Centrality

Siwon Kim

Seoul National University

siwon.kim@snu.ac.kr

2026-06-26

About Me

GitHub

https://github.com/Siwon-S-Kim

Konkuk University

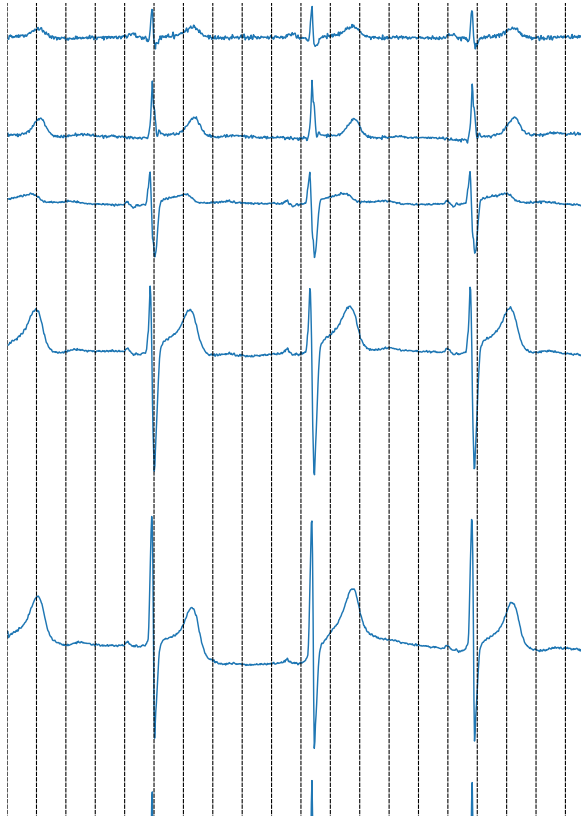
B.S. Double major in Physics and Mathematics

Seoul National University

Integrated Ph.D. in Mathematical Science

Data Scientist

ECG, Self-Supervised Learning



Advertisement

Advertisement

This slide is made in typst

using neovim 

experiment managed with uv 

Information Centrality

What, Why, How

1. What is information centrality?

What, Why, How

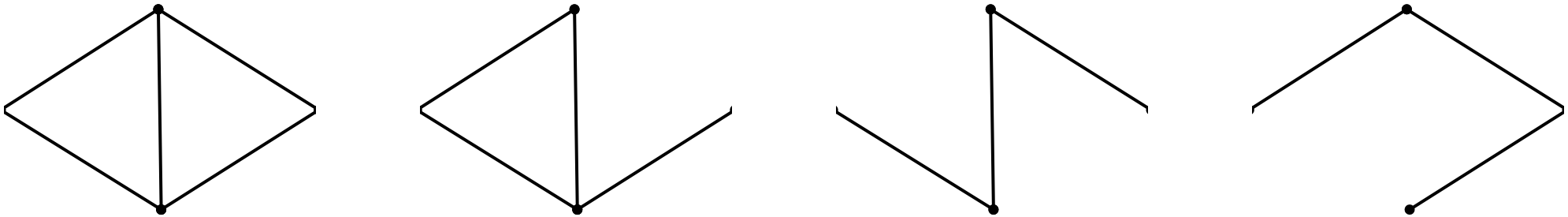
1. What is information centrality?

2. Why do we use information centrality?

What, Why, How

1. What is information centrality?
2. Why do we use information centrality?
3. How do we use information centrality?

Definition. *Graph* is a pair $G = (V, E)$ with finite set V and set of pairs E (of element in V)



Graph

Definition. *Graph* is a pair $G = (V, E)$ with finite set V and set of pairs E (of element in V)

1. Electric Circuit
2. Traffic Network
3. Social Network

Definition. A *weighted graph* $G = (V, E, w)$ is a graph with weight function

$$w : E \to \mathbb{R}^+$$

1. Electric Circuit
2. Traffic Network
3. Social Network

Definition. A *weighted graph* $G = (V, E, w)$ is a graph with weight function

$$w : E \to \mathbb{R}^+$$

1. Electric Circuit
2. Traffic Network
3. Social Network

w: Electric conductance of edge

Definition. A *weighted graph* $G = (V, E, w)$ is a graph with weight function

$$w : E \to \mathbb{R}^+$$

1. Electric Circuit — w: Electric conductance of edge
2. Traffic Network — w: Congestion of edge
3. Social Network

Definition. A *weighted graph* $G = (V, E, w)$ is a graph with weight function

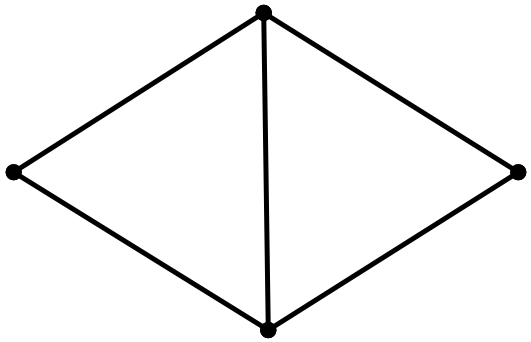
$$w : E \to \mathbb{R}^+$$

1. Electric Circuit — w: Electric conductance of edge
2. Traffic Network — w: Congestion of edge
3. Social Network — w: Relation intensity

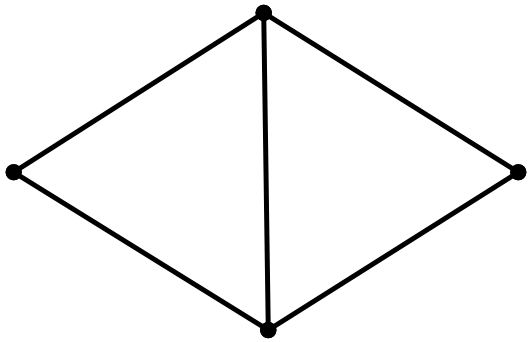
Required Properties

1. Kirchhoff's current law
2. Kirchhoff's voltage law
3. Ohm's law $V = IR$
4. Superposition
5. Equivalence theorems

Effective Resistance

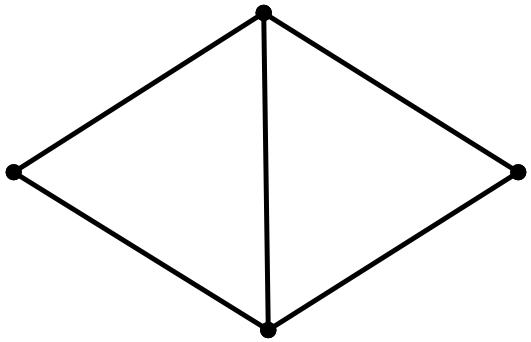


Effective Resistance



Definition. An *effective resistance* between v_i and v_j is "net opposition" of unit flow passing v_i,v_j

Effective Resistance



Definition. An *effective resistance* between v_i and v_j is "net opposition" of unit flow passing v_i,v_j

$$R_{ij} = \frac{V_i - V_j}{I}, \quad \text{when flow } I \text{ pass } v_i, v_j$$

Calculation

Ohm's law*

$$CV = \frac{V}{R} = I$$

Calculation

Ohm's law*

$$CV = \frac{V}{R} = I$$

$$\sum_{j \in N(v_i)} C_{ij}(V_i - V_j) = I_i \qquad V(G) = \{v_1, ..., v_n\}$$

Calculation

Ohm's law*

$$CV = \frac{V}{R} = I$$

$$\sum_{j \in N(v_i)} C_{ij}(V_i - V_j) = I_i \qquad V(G) = \{v_1, ..., v_n\}$$

$$\left(\sum_{j \in N(v_i)} C_{ij} \right) V_i - \sum_{j \in N(v_i)} C_{ij} V_j = I_i$$

Calculation

Ohm's law*

$$CV = \frac{V}{R} = I$$

$$\sum_{j \in N(v_i)} C_{ij}(V_i - V_j) = I_i \qquad V(G) = \{v_1, ..., v_n\}$$

$$\left(\sum_{j \in N(v_i)} C_{ij} \right) V_i - \sum_{j \in N(v_i)} C_{ij} V_j = I_i$$

$$A_{i1} V_1 + ... + A_{in} V_n = I_i, \qquad C_{ij} = 0 \text{ if not connected}$$

Calculation

Ohm's law*

$$CV = \frac{V}{R} = I$$

$$\sum_{j \in N(v_i)} C_{ij}(V_i - V_j) = I_i \qquad V(G) = \{v_1, \ldots, v_n\}$$

$$\left(\sum_{j \in N(v_i)} C_{ij} \right) V_i - \sum_{j \in N(v_i)} C_{ij} V_j = I_i$$

$$L(G)V = I \in \mathbb{R}^n$$

Calculation

Definition. *Discrete Laplace Equation*:

$$L(G)V = I \in \mathbb{R}^n$$

Calculation

$$L(G)V = I$$

“net opposition of unit flow passing the pair of vertices”

Calculation

$$L(G)V = I$$

"net opposition of unit flow passing the pair of vertices"

$$R_{ij} = V_i - V_j \quad \text{when } I : \text{unit flow between } v_i, v_j$$

Graph Laplacian

$$L(G) = \begin{pmatrix} \sum_j C_{1j} & -C_{12} & \cdots & -C_{1,n} \\ -C_{21} & \sum_j C_{2j} & \cdots & -C_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -C_{n1} & -C_{n,2} & \cdots & \sum_j C_{nj} \end{pmatrix}$$

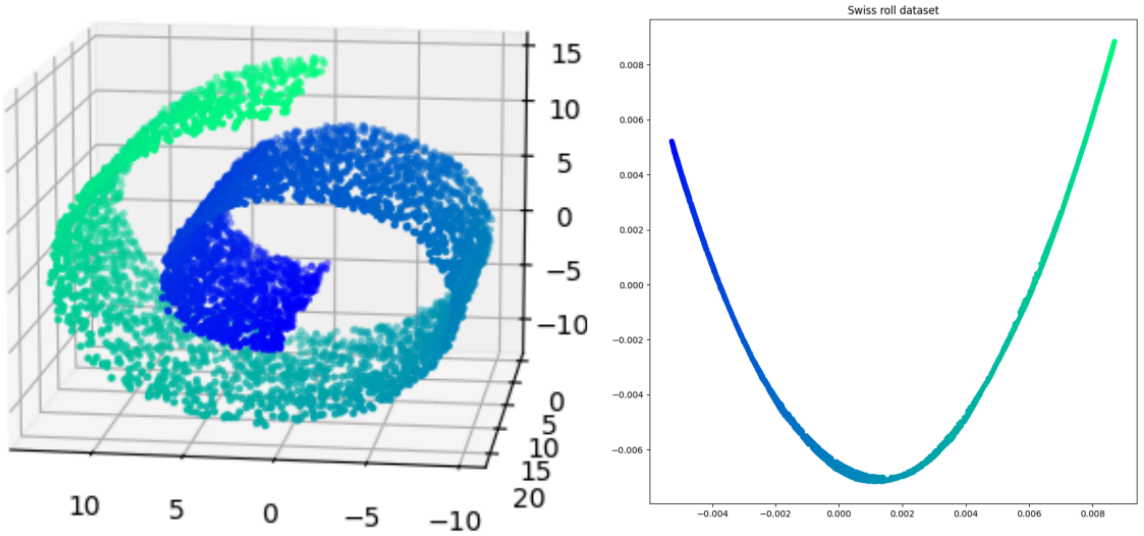
Graph Laplacian

$$L(G) = \begin{pmatrix} \sum_j C_{1j} & -C_{12} & \cdots & -C_{1,n} \\ -C_{21} & \sum_j C_{2j} & \cdots & -C_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -C_{n1} & -C_{n,2} & \cdots & \sum_j C_{nj} \end{pmatrix}$$

$$L(G)_{ii} = \sum_{j \neq i}^{n} C_{ij}, \quad L(G)_{ij} = -C_{ij} \quad i \neq j$$

Spectral Embedding

"Nonlinear dimension reduction technique"



- Implemented in UMAP
- Graph Laplacian

Formula

Definition. *Combinatorial Hodge Laplacian*

$$\Delta_G := L(G) + \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} \qquad \det(\Delta_G) \neq 0$$

Formula

Definition. *Combinatorial Hodge Laplacian*

$$\Delta_G := L(G) + \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} \qquad \det(\Delta_G) \neq 0$$

$$R_{ij} = g_{ii} + g_{jj} - g_{ij} - g_{ji}, \quad g_{ij} := \left(\Delta^{-1}\right)_{ij}$$

Definition of Information Centrality

Definition. An *effective conductance* is a reciprocal of effective resistance

Definition. An *information centrality* of vertex v_i is a harmonic mean of effective conductances between v_i and others

$$\mathrm{IC}_G(v_i) := \left(\frac{\sum_{j=1}^{n} R_{ij}}{n-1} \right)^{-1}, \quad (R_{ii} = 0)$$

Summary

Effective Resistance/Conductance

Connectivity measure

Information Centrality

Average of connectivity measure

Kernel Method

Positive Definite Kernel

"Extension of positive definite matrix"

Positive Definite Kernel

"Extension of positive definite matrix"

Definition. A symmetric function $k : \mathrm{X} \times \mathrm{X} \to \mathbb{R}$ is called *positive definite*, if for any finite subset $\{x_1, \ldots, x_m\} \subset \mathrm{X}$ associated Gram matrix K (or Kernel matrix) is positive semi definite.

Interpretation

Kernel method has ...

1. Robustness of Linear Method

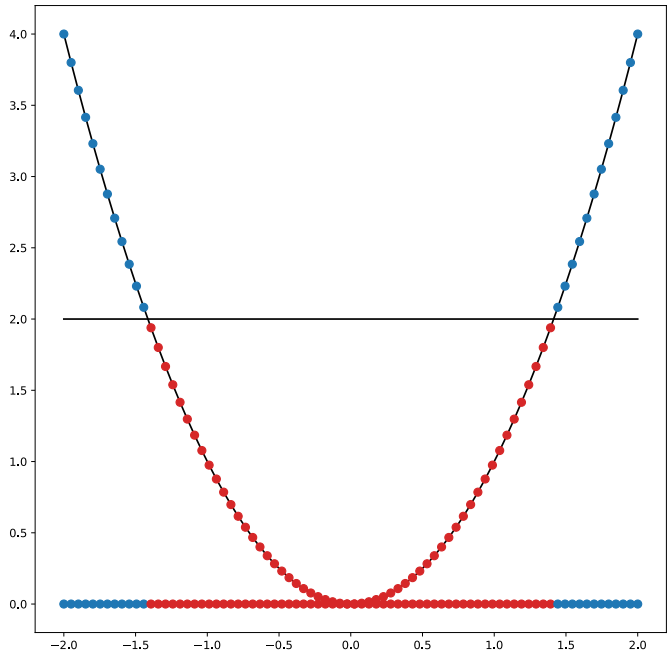
2. Power to Process Non-Linear Data

Interpretation

"Kernel method = Composition of Transform and Linear method"

Interpretation

"Kernel method = Composition of Transform and Linear method"



Interpretation

Theorem. *For each positive definite kernel* $k(x, y) : \mathrm{X} \times \mathrm{X} \to \mathbb{R}$*, there is map from data to Hilbert space* $\Phi : \mathrm{X} \to \mathbb{H}$*, that satisfies:*

$$k\big(x_i, x_j\big) = \langle \Phi(x_i), \Phi\big(x_j\big) \rangle.$$

Definition. We call Φ a *feature map* and the codomain Hilbert space a *feature space*

Kernel PCA

Kernel PCA

Lemma. *Spectral Embedding is composition of some feature map and Principal Component Analysis (PCA)*

with kernel matrix

$$K = \Delta^{-1}$$

$$k(x_i, x_j) = (\Delta^{-1})_{ij} = g_{ij}$$

Kernel PCA

$$\mathrm{d}\left(\Phi(x_i), \Phi\left(x_j\right)\right)^2$$

Kernel PCA

$$\mathrm{d}\left(\Phi(x_i), \Phi\left(x_j\right)\right)^2 = \langle \Phi(x_i) - \Phi\left(x_j\right), \Phi(x_i) - \Phi\left(x_j\right) \rangle$$

Kernel PCA

$$
\begin{aligned}
\mathrm{d}\left(\Phi(x_i), \Phi\left(x_j\right)\right)^2 &= \langle \Phi(x_i) - \Phi\left(x_j\right), \Phi(x_i) - \Phi\left(x_j\right) \rangle \\
&= \langle \Phi(x_i), \Phi(x_i) \rangle + \langle \Phi\left(x_j\right), \Phi\left(x_j\right) \rangle \\
&\quad - \langle \Phi(x_i), \Phi\left(x_j\right) \rangle - \langle \Phi\left(x_j\right), \Phi(x_i) \rangle
\end{aligned}
$$

Kernel PCA

$$
\begin{aligned}
\mathrm{d}\left(\Phi(x_i), \Phi\left(x_j\right)\right)^2 &= \langle \Phi(x_i) - \Phi\left(x_j\right), \Phi(x_i) - \Phi\left(x_j\right) \rangle \\
&= \langle \Phi(x_i), \Phi(x_i) \rangle + \langle \Phi\left(x_j\right), \Phi\left(x_j\right) \rangle \\
&\quad - \langle \Phi(x_i), \Phi\left(x_j\right) \rangle - \langle \Phi\left(x_j\right), \Phi(x_i) \rangle \\
&= K(x_i, x_i) + K\left(x_j, x_j\right) - K\left(x_i, x_j\right) - K\left(x_j, x_i\right)
\end{aligned}
$$

Kernel PCA

$$
\begin{aligned}
\mathrm{d}\left(\Phi(x_i), \Phi\left(x_j\right)\right)^2 &= \langle \Phi(x_i) - \Phi\left(x_j\right), \Phi(x_i) - \Phi\left(x_j\right) \rangle \\
&= \langle \Phi(x_i), \Phi(x_i) \rangle + \langle \Phi\left(x_j\right), \Phi\left(x_j\right) \rangle \\
&\quad - \langle \Phi(x_i), \Phi\left(x_j\right) \rangle - \langle \Phi\left(x_j\right), \Phi(x_i) \rangle \\
&= K(x_i, x_i) + K\left(x_j, x_j\right) - K\left(x_i, x_j\right) - K\left(x_j, x_i\right) \\
&= g_{ii} + g_{jj} - g_{ij} - g_{ji}
\end{aligned}
$$

Kernel PCA

$$
\begin{aligned}
\mathrm{d}\left(\Phi(x_i), \Phi\left(x_j\right)\right)^2 &= \langle \Phi(x_i) - \Phi\left(x_j\right), \Phi(x_i) - \Phi\left(x_j\right) \rangle \\
&= \langle \Phi(x_i), \Phi(x_i) \rangle + \langle \Phi\left(x_j\right), \Phi\left(x_j\right) \rangle \\
&\quad - \langle \Phi(x_i), \Phi\left(x_j\right) \rangle - \langle \Phi\left(x_j\right), \Phi(x_i) \rangle \\
&= K(x_i, x_i) + K\left(x_j, x_j\right) - K\left(x_i, x_j\right) - K\left(x_j, x_i\right) \\
&= g_{ii} + g_{jj} - g_{ij} - g_{ji} \\
&= R_{ij}
\end{aligned}
$$

Interpretation

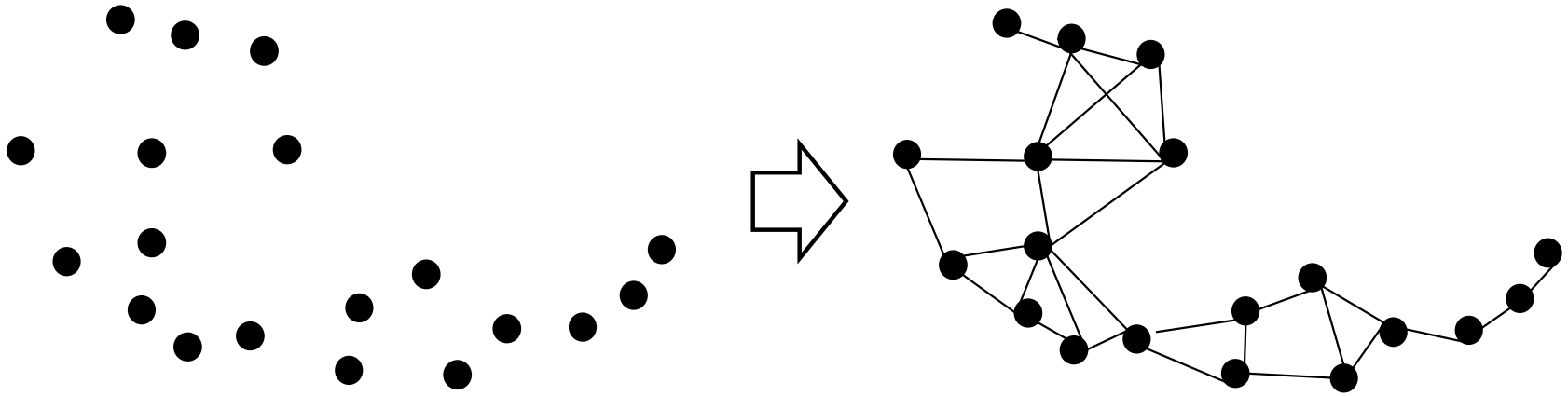
"Information centrality is a statistic of distances in the feature space"

High information centrality $\rightarrow$ Close to vertices $\rightarrow$ Affinity

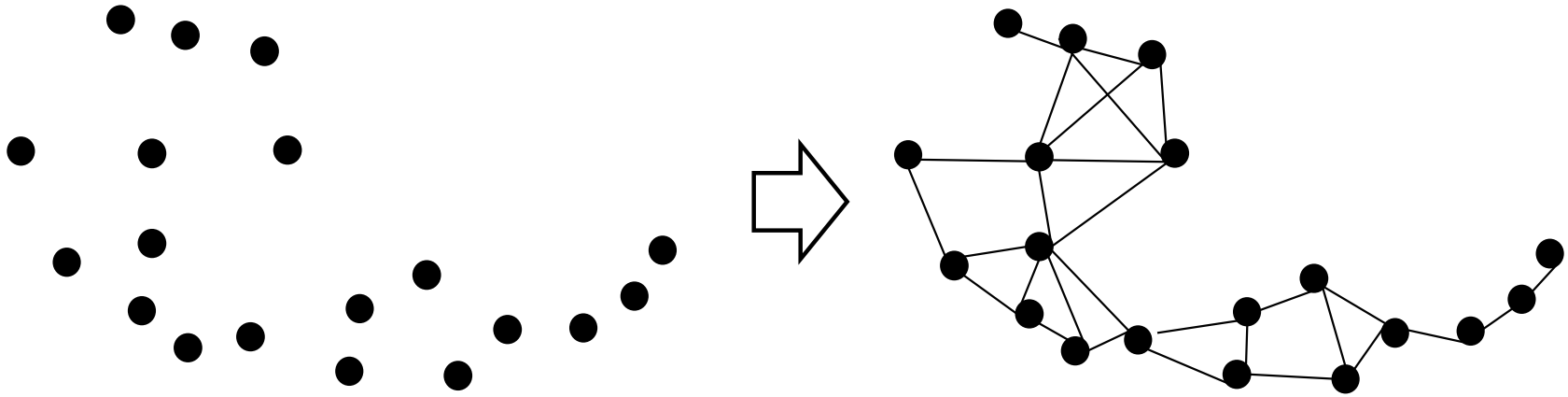
Low information centrality $\rightarrow$ Far from vertices $\rightarrow$ Deviation

Application

Graph formulation

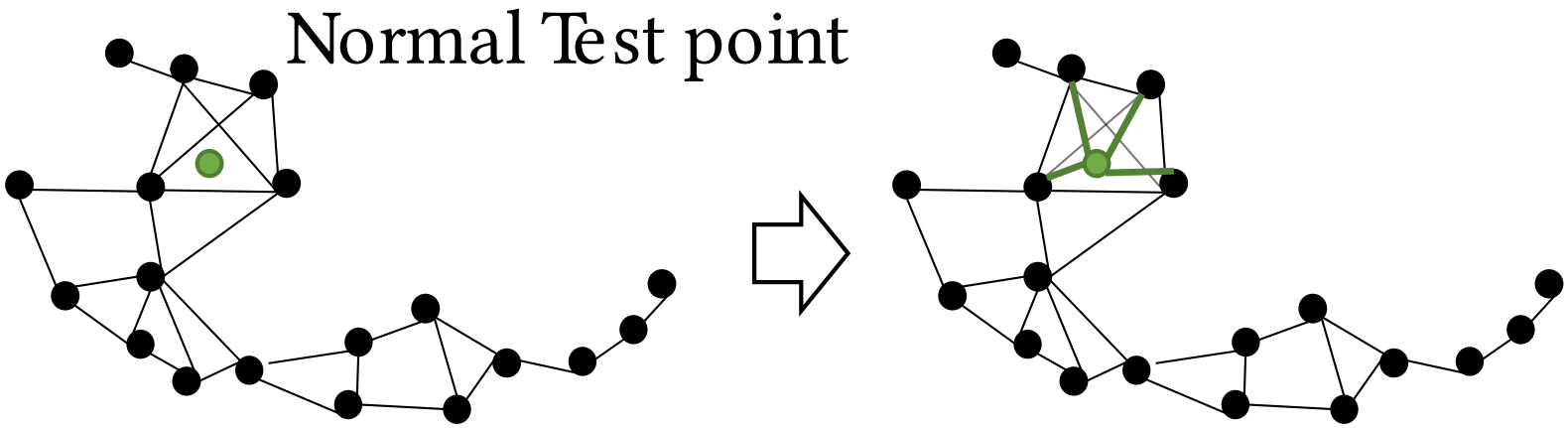


Graph formulation

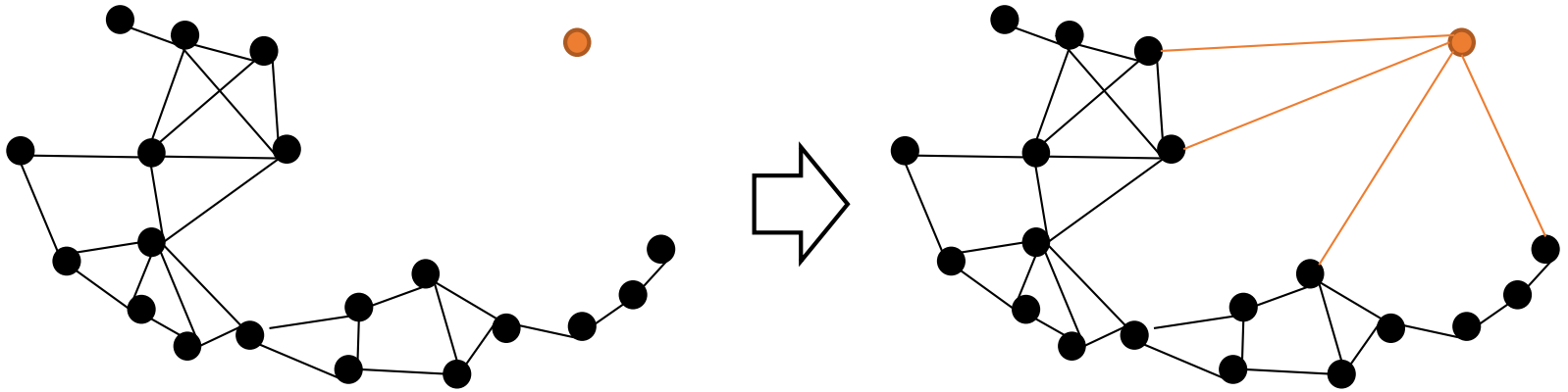


k-Nearest Neighbourhood

$$w(v_i, v_j) = e^{-k\,\mathrm{d}(v_i, v_j)}$$


Normal Test point

Outlier Test point



Pseudo Code

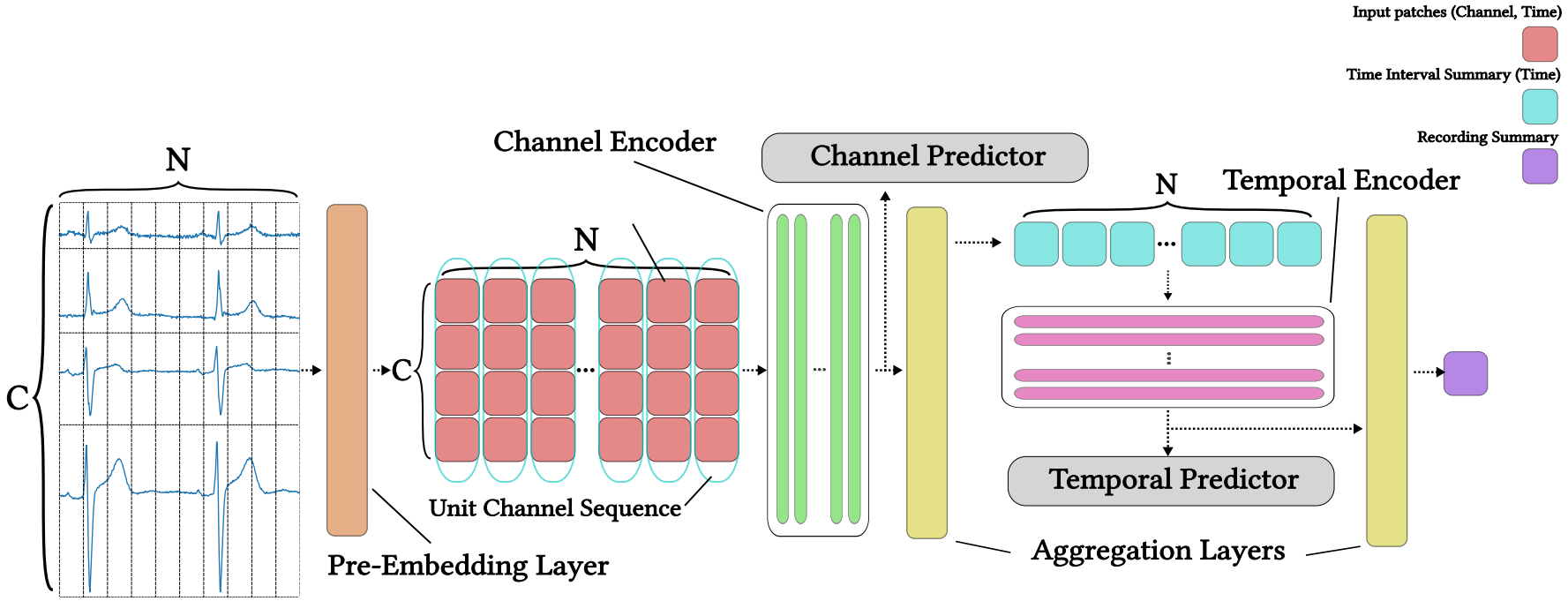
Test Information Centrality

Normal points $\{X_i\}$, Input Points $\{Z_j\}$, Nbhd Scheme S
▷ Fit normal points
$G \leftarrow$ graph formed by S with $\{X_i\}$
$\mathrm{w}(\cdot) \leftarrow$ Assign weight on $E(G)$
for $z \in \{Z_j\}$ **do**
    $N_z \leftarrow S$-Nbhd of z in $\{X_i\}$
    $\mathrm{w}(\cdot) \leftarrow$ Extend weight on $\{(z, x) |\ x \in N_z\}$
    $G' \leftarrow$ Subgraph sampled from path component of N_z
    ▷ Find subgraph to reduce computation
    $H \leftarrow G' \cup z$
    $I_j \leftarrow \mathrm{IC}_{H(z)}$
end
return $\{I_j\}$

Outlier Detection on ECG

Input Data

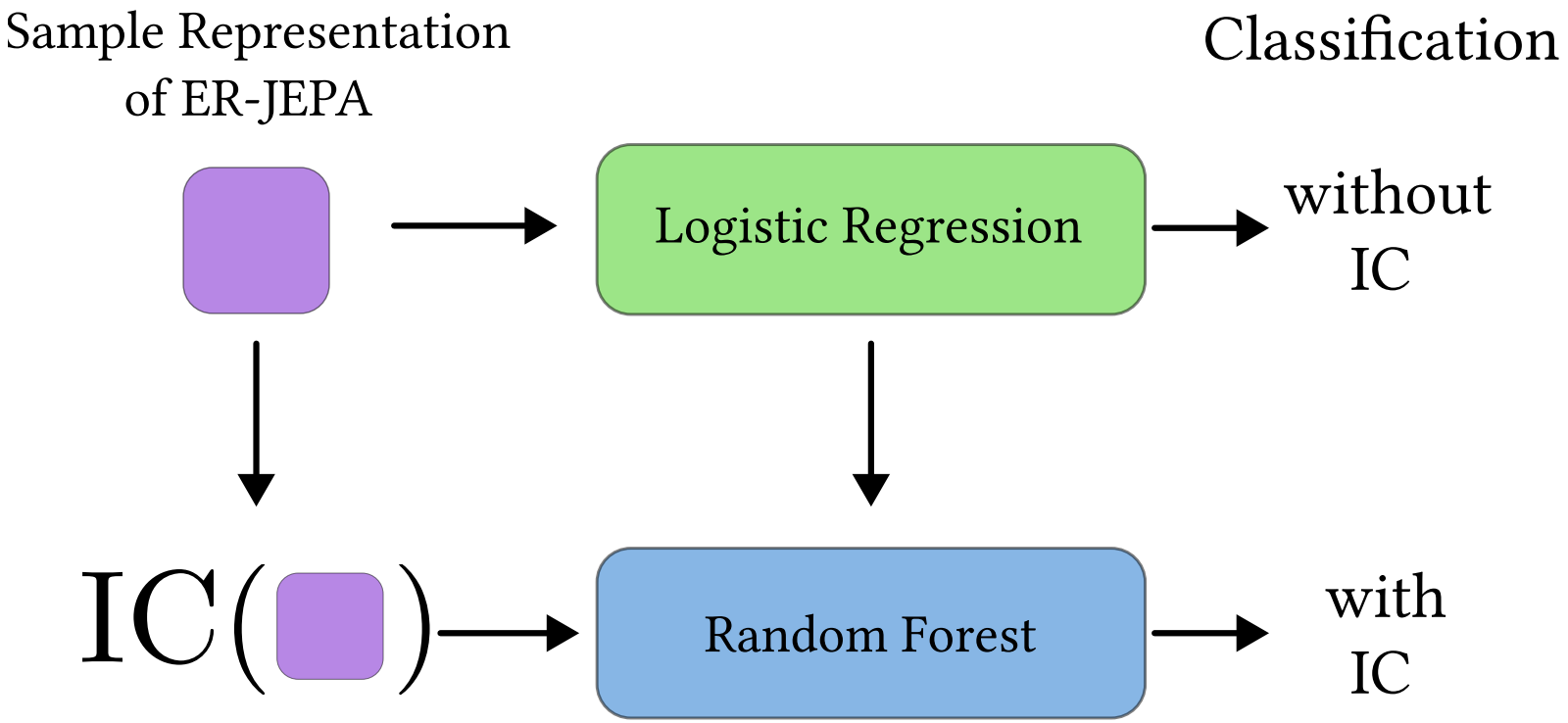
Input Data


Input patches (Channel, Time)
Time Interval Summary (Time)
Recording Summary
N
C
Pre-Embedding Layer
Channel Encoder
N
C
Unit Channel Sequence
Channel Predictor
N
Temporal Encoder
Temporal Predictor
Aggregation Layers

Lightweight SSL Framework for Multivariate Time Series with Hierarchical-JEPA on Electrocardiogram Data

https://github.com/Siwon-S-Kim/er-jepa-ecg (not available right now)

Schematic


Sample Representation
of ER-JEPA
Classification
Logistic Regression
without
IC
IC( )
Random Forest
with
IC

Result

Category	Precision	Recall	F1-score	Support
NSR	0.941 -> 0.945	0.905 -> 0.915	0.923 -> 0.930	284
Not NSR	0.986 -> 0.987	0.991 -> 0.992	0.989 -> 0.990	1869
Accuracy			0.980 -> 0.982	2153
Macro avg	0.964 -> 0.966	0.948 -> 0.954	0.956 -> 0.960	2153
Weighted avg	0.980 -> 0.982	0.980 -> 0.982	0.980 -> 0.982	2153

$$\text{Average of } \Delta\,\text{Acc} = +0.12\% \quad \text{Std. Dev. of } \Delta\,\text{Acc} = 0.05\%$$

$$\text{Average of } \Delta\,\text{F1} = +0.29\% \quad \text{Std. Dev. of } \Delta\,\text{F1} = 0.11\%$$

Theory

Continuity Result

Theorem. *Let $k \in \mathbb{N}, \varepsilon > 0$ and G connected graph constructed from data points $\mathrm{X} = \{x_1, \ldots, x_n\} \subset \mathbb{R}^N$ with appropriate weight function w. Then test effective conductance between the test point $x \in \mathbb{R}^N$ and any vertex is locally Lipschitz continuous almost everywhere, with upper bound of local Lipschitz constant*

$$\propto \frac{\kappa\left(\tilde{G}/e\right)\kappa\left(\tilde{G}\right)}{\kappa(G)^2}$$

Next Step

Stronger stability theorem

Differential Geometry, Random Matrix Theory, Optimal Transport