Data Mining (CSE542)

Homework 03

ID: ____ Name: ___조원석_____ Date:__2023-04-03__

Task-1

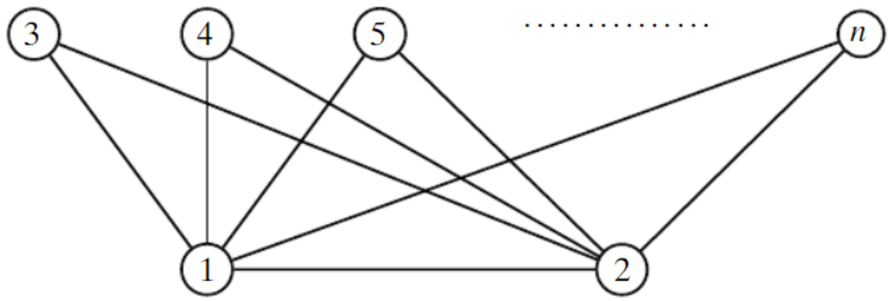


Figure 4.17.

Consider the double star graph given in Figure 4.17 with n nodes, where only nodes 1 and 2 are connected to all other vertices, and there are no other links. Answer the following questions (treating n as a variable).

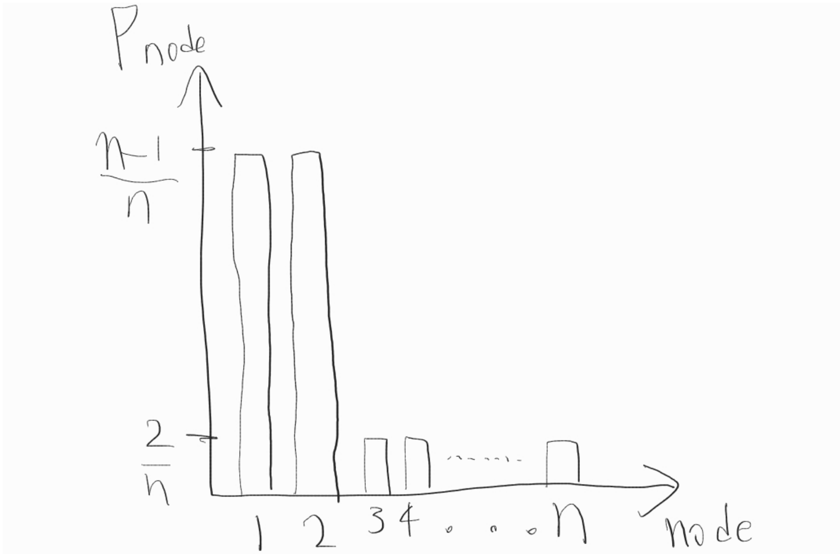
(a) What is the degree distribution for this graph?

(b) What is the mean degree?

(c) What is the clustering coefficient for vertex 1 and vertex 3?

(d) What is the clustering coefficient $C(G)$ for the entire graph? What happens to the clustering coefficient as $n \to \infty$?

(a) degree distribution



P	Pnode
2	$\frac{\mathbf{n-2}}{\mathbf{n}}$
n-1	$\frac{\mathbf{2}}{\mathbf{n}}$

(b) mean degree = $\frac{1}{n} * \sum_{node=1}^{n} P_{node} = 4 - \frac{6}{n}$

(c) clustering coefficient for

- vertex 1 : C(1) = $\frac{n-2}{\binom{n-1}{2}} = 2 * \frac{(n-2)}{(n-1)*(n-2)} = \frac{2}{n-1}$

- vertex 3 : C(3) = 1

(d)

- clustrering coefficient C(G) for the entire graph

C(G) = $((n-2)*1 + 2*2/(n-1))/n = \frac{n^2-3n+6}{n^2-n}$

- clustering coefficient as n -> ∞

C(G) -> $\lim_{n \to \infty} C(G) = \lim_{n \to \infty} \frac{n^2-3n+6}{n^2-n} = 1$

Task-2

Consider the data shown in Table 5.1. Assume the following kernel function: $K(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|^2$. Compute the kernel matrix $\mathbf{K}$.

Table 5.1.

	X_1	X_2
$\mathbf{x}_1^T$	4	2.9
$\mathbf{x}_2^T$	2.5	1
$\mathbf{x}_3^T$	3.5	4
$\mathbf{x}_4^T$	2	2.1

$$\mathrm{K}(X_1, X_2) = \left|\left|X_1 - X_2\right|\right|^2 = 1.5^2 + 1.9^2 = 5.86$$

$$\mathrm{K}(X_1, X_3) = \left|\left|X_1 - X_3\right|\right|^2 = 0.5^2 + 1.1^2 = 1.46$$

$$\mathrm{K}(X_1, X_4) = \left|\left|X_1 - X_4\right|\right|^2 = 2.0^2 + 0.8^2 = 4.64$$

$$\mathrm{K}(X_2, X_3) = \left|\left|X_2 - X_3\right|\right|^2 = 1.0^2 + 3.0^2 = 10.0$$

$$\mathrm{K}(X_2, X_4) = \left|\left|X_2 - X_4\right|\right|^2 = 0.5^2 + 1.1^2 = 1.46$$

$$\mathrm{K}(X_3, X_4) = \left|\left|X_3 - X_4\right|\right|^2 = 1.5^2 + 1.9^2 = 5.86$$

- Kernel matrix K

$$\mathrm{K} = \begin{matrix} 0 & 5.86 & 1.46 & 4.64 \\ 5.86 & 0 & 10 & 1.46 \\ 1.46 & 10 & 0 & 5.86 \\ 4.64 & 1.46 & 5.86 & 0 \end{matrix}$$

Task-3:

Consider the following data matrix **D**:

X_1	X_2
8	−20
0	−1
10	−19
10	−20
2	0

(a) Compute the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ for **D**.
(b) Compute the eigenvalues of $\boldsymbol{\Sigma}$.
(c) What is the "intrinsic" dimensionality of this dataset (discounting some small amount of variance)?
(d) Compute the first principal component.

(a) Compute the mean μ and convariance matrix ∑ for D

$\mu = [\frac{8+0+10+10+2}{5}, \frac{-20,-1,-19,-20,0}{5}]$ **= [6.0, -12.0]**

X1	X2	X1^2	X2^2	X1 × X2
8- (6.0)=2	-20 - (-12.0)=-8	4	64	-16
0- (6.0)=-6	-1- (-12.0)=11	36	121	-66
10- (6.0)=4	-19- (-12.0)=-7	16	49	-28
10- (6.0)=4	-20- (-12.0)=-8	16	64	-32
2- (6.0)=-4	0- (-12.0)=12	16	144	-48
SUM		88	442	-190
SUM / 5		17.6	88.4	-38

$$\sum \text{ for D} = \frac{1}{5}\begin{pmatrix} \boldsymbol{X1^2} & \boldsymbol{X1\,X2} \\ \boldsymbol{X2\,X1} & \boldsymbol{X2^2} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{17.6} & \mathbf{-38} \\ \mathbf{-38} & \mathbf{88.4} \end{pmatrix}$$

(b) Compute the eigenvalues of ∑

solving eigenvalues by det(∑ - λI),

$$(\mathbf{17.6} - \boldsymbol{\lambda})(\mathbf{88.4} - \boldsymbol{\lambda}) - \mathbf{38}^2$$

$$= \boldsymbol{\lambda}^2 - \mathbf{106}\boldsymbol{\lambda} + \mathbf{1555.84} - \mathbf{1444}$$

$$= \boldsymbol{\lambda}^2 - \mathbf{106}\boldsymbol{\lambda} + \mathbf{111.84} = \mathbf{0}$$

$$\boldsymbol{\lambda} = \frac{106 \pm \sqrt{106^2 - 4 * 111.84}}{2} = \frac{106 \pm \sqrt{10788.64}}{2}$$

$$\boldsymbol{\lambda_1} = \frac{106 + \sqrt{10788.64}}{2} \fallingdotseq 104.9342 \ \boldsymbol{\lambda_2} = \frac{106 - \sqrt{10788.64}}{2} \fallingdotseq 1.0658$$

(c) What is the "intrinsic" dimensionality of this dataset(discounting some small amount of variance)?

$$\frac{\mathbf{104.9342}}{\mathbf{104.9342} + \mathbf{1.0658}} \fallingdotseq 0.9899$$

=> first principal component is 105.

(d) Compute the first principal component.

=> ∑ ui=λiui, $\boldsymbol{\lambda_1} \fallingdotseq 104.9342$

Solving eigenvector)

$$\begin{pmatrix} \mathbf{17.6} - \mathbf{104.9342} & -\mathbf{38} \\ -\mathbf{38} & \mathbf{88.4} - \mathbf{104.9342} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

$$\begin{pmatrix} -\mathbf{87.3342} & -\mathbf{38} \\ -\mathbf{38} & -\mathbf{16.5342} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

$$-\mathbf{87.3342x} - \mathbf{38y} = \mathbf{0}$$

$$\mathbf{x} = -\frac{\mathbf{38}}{\mathbf{87.3342}}\boldsymbol{y} \fallingdotseq -\mathbf{0.4351}\boldsymbol{y},$$

If y = 1, x≒0.4351

$$\frac{\mathbf{1}}{\mathbf{1.09}} \begin{pmatrix} -\mathbf{0.4351} \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} -\mathbf{0.3992} \\ \mathbf{0.9173} \end{pmatrix}$$