Data Mining
CS-422
Homework - 4

Excercise 1

1.1 Leskovec, Ch. 3.

Ex. 3.1.1

Given, $\{1,2,3,4\}$, $\{2,3,5,7\}$, $\{2,4,6\}$.

Jaccard Similarity $= \cancel{\frac{A \cup B}{A \cap B}}$ $\frac{\{A \cap B\}}{\{A \cup B\}}$

$A = \{1, 2, 3, 4\}$
$B = \{2, 3, 5, 7\}$
$C = \{2, 4, 6\}$

$Sim(A,B) = \frac{2}{6} = 0.333$

$Sim(A,C) = \frac{2}{5} = 0.4$.

$Sim(B,C) = \frac{1}{6} = 0.167$.

Ex. 3.2.1

The first sentence of 3.2 is :-

"The most effective way to represent documents as sets, for the purpose of identifying lexically similar documents is to construct from the document the set of short strings that appear within it.

⇒ Set of first 10 3-shingles is ⟶

{"The most effective", "most effective way", "effective way to", "way to represent", "to represent documents", "represent documents as" "documents as sets", "as sets for", "sets for purpose", "for ~~the~~ purpose of"}.

Ex. 3.3.3

Element	S_1	S_2	S_3	S_4	$2x+1 \bmod 6$	$3x+2 \bmod 6$	$5x+2 \bmod 6$
0	0	1	0	1	1	2	2
1	0	1	0	0	3	5	1
2	1	0	0	1	5	2	0
3	0	0	1	0	1	5	5
4	0	0	1	1	3	2	4
5	1	0	0	0	5	5	3

a) Given,

$h_1(x) = 2x+1 \bmod 6$.

$h_2(x) = 3x+2 \bmod 6$.

$h_3(x) = 5x+2 \bmod 6$.

$h_1(0) = 2(0)+1 \bmod 6 = 1$

$h_2(0) = 3(0)+2 \bmod 6 = 2$

$h_3(0) = 5(0)+2 \bmod 6 = 2$.

$h_1(1) = 2(1)+1 \bmod 6 = 3$

$h_2(1) = 3(1)+2 \bmod 6 \Rightarrow 5$.

$h_3(1) = 5(1)+2 \bmod 6 = 1$

$h_1(2) = 2(2)+1 \bmod 6 = 5$.

$h_2(2) = 3(2)+2 \bmod 6 = 2$

$h_3(2) = 5(2)+2 \bmod 6 = 0$.

$h_1(3) = 2(3)+1 \bmod 6 = 1$

$h_2(3) = 3(3)+2 \bmod 6 = 5$

$h_3(3) = 5(3)+2 \bmod 6 = 5$.

Scanned with CamScanner

$h_1(4) = 2(4)+1 \bmod 6 = 3$
$h_2(4) = 3(4)+2 \bmod 6 = 2$
$h_3(4) = 5(4)+2 \bmod 6 = 4.$

$h_1(5) = 2(5)+1 \bmod 6 = 5$
$h_2(5) = 3(5)+2 \bmod 6 = 5$
$h_3(5) = 5(5)+2 \bmod 6 = 3.$

Final Minhash Signature Matrix:-

S_1	S_2	S_3	S_4
5	1	1	1
2	2	2	2
0	1	4	0.

Minhash Signature Matrix

	S_1	S_2	S_3	S_4
$h_1(0)$	/	1	/	1
$h_2(0)$	/	2	/	2
$h_3(0)$	/	2	/	2
$h_1(1)$	/	1	/	1
$h_2(1)$	/	2	/	2
$h_3(1)$	/	1	/	2
$h_1(2)$	5	1	/	1
$h_2(2)$	2	2	/	2
$h_3(2)$	0	1	/	0.
$h_1(3)$	5	1	1	1
$h_2(3)$	2	2	5	2
$h_3(3)$	0	1	5.	0.
$h_1(4)$	5	1	1	1
$h_2(4)$	2	2	2	2
$h_3(4)$	0	1	4.	0.
$h_1(5)$	5	1	1	1
$h_2(5)$	2	2	2	2
$h_3(5)$	0.	1	4	0

b) Only h_3 hash function is a true permutation. as it produces all different values.

c)

Similarities	S_1, S_2	$S_1 S_3$	$S_1 S_4$	$S_2 S_3$	$S_2 S_4$	$S_3 S_4$
Columns	$0/4 = 0.$	$0/4 = 0.$	$1/4 = 0.25$	$0/4 = 0.$	$1/4 = 0.25$	$1/4 = 0.25$
Sig matrix	$1/3 = 0.33$	$1/3 = 0.33$	$2/3 = 0.67$	$2/3 = 0.67$	$2/3 = 0.67$	$2/3 = 0.67$

The estimated Jaccard similarities are not close to the true ones at all.

Ex. 3.4.1.

S-curve $1-(1-s^r)^b$ for $s = 0.1, 0.2, \ldots .0.9$ for

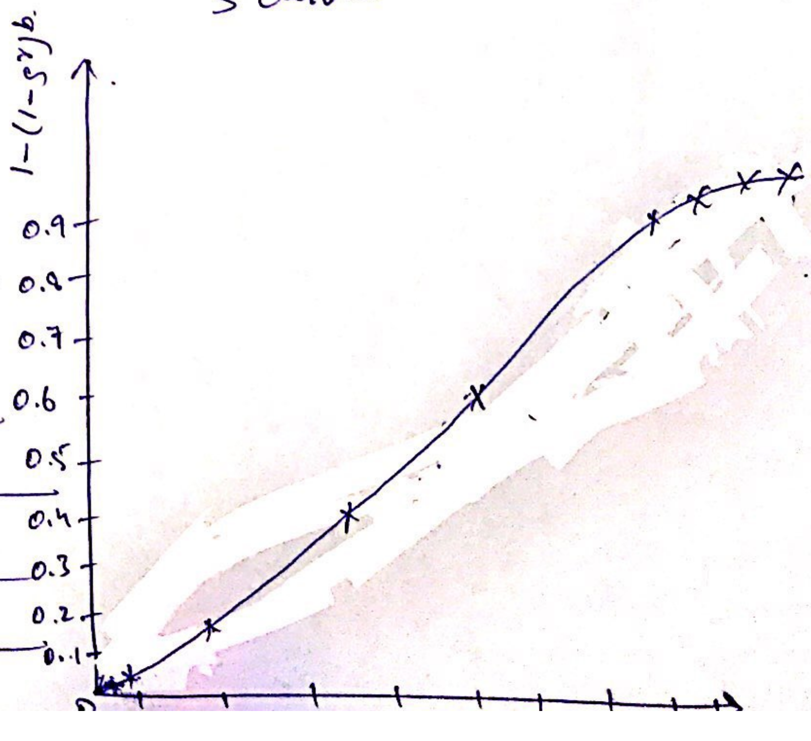
a) $r=3$, $b=10$

b) $r=6$, $b=20$.

a) $r=3$, $b=10$.

s	$1-(1-s^r)^b$
0.1	0.0100
0.2	0.0772
0.3	0.2394
0.4	0.4839
0.5	0.7369
0.6	0.9123
0.7	0.9850
0.8	0.9992

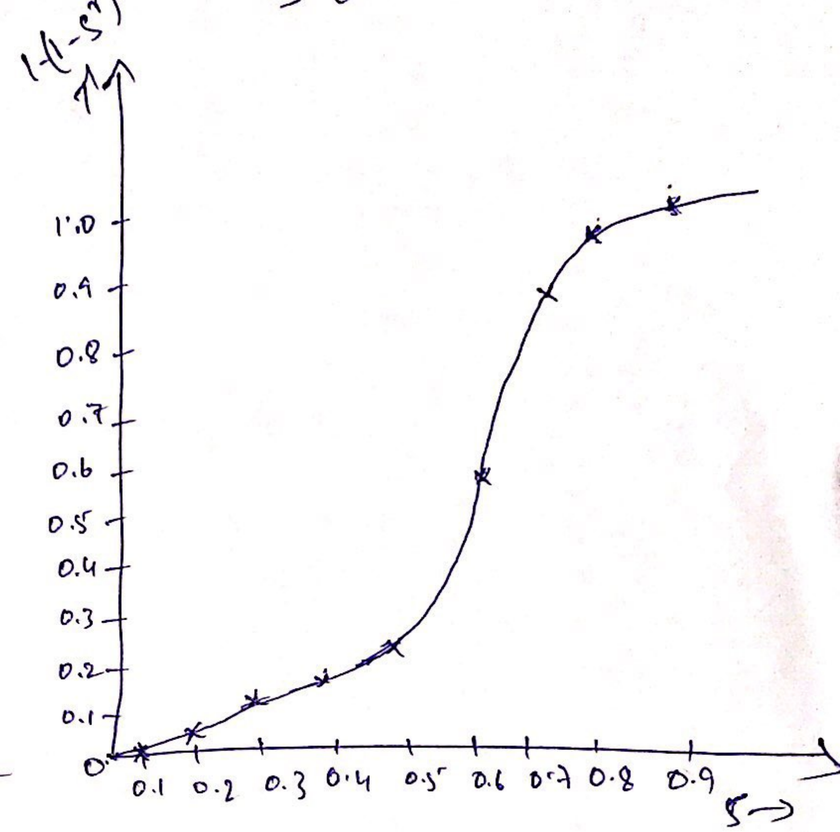
S-curve.



b) $r = 6$, $b = 20$.

s	$1-(1-s^r)^b$
0.1	0.000
0.2	0.0013
0.3	0.0145
0.4	0.0788
0.5	0.2702
0.6	0.6154
0.7	0.9152.
0.8	0.9977
0.9	1.000.

S-curve



Ex 3.5.4.

a) $\{1,2,3,4\}$ & $\{2,3,4,5\}$

Jaccard Distance $= \boxed{1 - \frac{A \cap B}{A \cup B}} = 1 - \frac{3}{5} \Rightarrow 0.4.$

b) $\{1,2,3\}$ & $\{4,5,6\}$

$JD = 1 - \frac{A \cap B}{A \cup B} \Rightarrow 1 - \frac{0}{6} = 1.$

Ex. 3.5.5 :

a) $\overset{A}{(3, -1, 2)}$ and $\overset{B.}{(-2, 3, 1)}$

$$\text{cosine}(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

$$A \cdot B \Rightarrow 3(-2) + (-1)3 + 2(1)$$

$$\Rightarrow -6 - 3 + 2.$$

$$\Rightarrow -7.$$

$$\|A\| = \sqrt{3^2 + (-1)^2 + 2^2} = \sqrt{14}.$$

$$\|B\| = \sqrt{(-2)^2 + 3^2 + 1^2} = \sqrt{14}.$$

$$\text{cosine}(\theta) = \frac{-7.}{\sqrt{14}\sqrt{14}} \Rightarrow \frac{-1}{2.}$$

$$\theta = \cos^{-1}\left(\frac{-1}{2}\right) = \cos^{-1}(-0.5) = 120^\circ.$$

---

b) $\overset{A}{(1, 2, 3)}$ and $\overset{B}{(2, 4, 6)}$.

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|}.$$

$$A \cdot B = 2(1) + 2(4) + 6(3)$$

$$\Rightarrow 2 + 8 + 18 \Rightarrow 28.$$

$$\|A\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{1+4+9} = \sqrt{14}.$$

$$\|B\| = \sqrt{2^2 + 4^2 + 6^2} = \sqrt{4+16+36} = \sqrt{56} = 2\sqrt{14}.$$

$$\cos(\theta) = \frac{28.}{\sqrt{14} \cdot 2\sqrt{14}.} \Rightarrow \frac{\cancel{28}^{14}}{\cancel{2}(14)} \Rightarrow 1$$

$$\cos^{-1}(1) = 0^\circ.$$

---

c) (5, 0, -4) and (-1, -6, 2).

$\cos(\theta) = \dfrac{A \cdot B}{\|A\| \|B\|}$ $\qquad A \cdot B = 5(-1) + 0(-6) + 2(-4)$

$\Rightarrow -5 + 0 - 8 \Rightarrow -13.$

$\|A\| = \sqrt{25 + 16} = \sqrt{41}$

$\|B\| = \sqrt{1 + 36 + 4} = \sqrt{41}$

$\theta = \cos^{-1}\left(\dfrac{-13}{41}\right) = \underline{108.48^\circ}$

---

d) (0, 1, 1, 0, 1, 1) and (0, 0, 1, 0, 0, 0)

$\cos\theta = \dfrac{A \cdot B}{\|A\| . \|B\|}$ $\qquad A \cdot B = 0 + 0 + 1 + 0 + 0 + 0 = 1.$

$\|A\| = \sqrt{1 + 1 + 1 + 1} = \sqrt{4} = 2.$

$\|B\| = \sqrt{1} = 1.$

$\cos\theta = \dfrac{1}{2}$ $\qquad \cos^{-1}\left(\dfrac{1}{2}\right) = \underline{60^\circ}.$

1.2 <u>Tan Chp 2.</u>

Q14) These attributes are all numerical, but can have widely varying ranges of values, depending on the scale used to measure them. The attributes are not asymmetric and the magnitude of an attribute matters. These latter 2 facts eliminate the cosine & correlation measure. Euclidean distance, applied after standardizing the attributes to have a mean of 0 and a std dev of 1, would be appropriate.

Q18)

a) $x = 0101010001$

$y = 0100011000$

Hamming distance = no. of different bits $\Rightarrow 3$.

Jaccard similarity $= \frac{2}{5} = \underline{\underline{0.4}}$.

---

b) The Hamming distance is similar to the Simple Matching Co-efficient (SMC).

$$SMC = \frac{\text{Hamming distance}}{\text{No. of bits}}$$

The Jaccard measure is similar to the cosine measure because both ignore 0-0 matches.

---

c) Jaccard is more appropriate for comparing the genetic makeup of two organisms. Since we want to see how many genes these 2 organisms share.

---

d) 2 Humans share > 99.9% of same genes. If we want to compare the genetic makeup of 2 human beings, we should focus on their differences. Thus, Hamming distance is more appropriate in this direction.

Q19)

(i) $x = (1, 1, 1, 1)$, $y = (2, 2, 2, 2)$

~~cosine = $\frac{8}{}$ ⇒~~

→ $\text{cosine}(x,y) = \frac{8}{\sqrt{1^2+1^2+1^2+1^2}\sqrt{2^2+2^2+2^2+2^2}} = \frac{8}{2\times 4} \Rightarrow \underline{\underline{1}}.$

→ corr(x,y) $\Rightarrow$ mean of $x = \frac{4}{4} = 1$

mean of $y = \frac{8}{4} = 2$.

$$\frac{(1-1)(2-2)+(1-1)(2-2)+(1-1)(2-2)+(1-1)(2-2)}{\sqrt{[(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2].[(2-2)^2+(2-2)^2+(2-2)^2+(2-2)^2]}}.$$

$= \underline{0}$ (undefined)

$$\text{corr}(x,y) = \frac{\sum_{i=1}^{n}(x_i-\hat{\mu}_x)(y_i-\hat{\mu}_y)}{\sqrt{\sum_{i=1}^{n}(x_i-\hat{\mu}_x)^2 \sum_{i=1}^{n}(y_i-\hat{\mu}_y)^2}}.$$

→ Euclidean $= \sqrt{(1-2)^2+(1-2)^2+(1-2)^2+(1-2)^2}$

$\Rightarrow \sqrt{1^2+1^2+1^2+1^2} = \underline{\underline{2}}.$

→

(ii) $x = (0,1,0,1)$ $y = (1,0,1,0)$
cosine, corr, Euclidean, Jaccard.

→ $\cos(x,y) = 0$

→ corr(x,y) = mean of $x = \frac{2}{4} = 0.5$

mean of $y = \frac{2}{4} = 0.5$

$$\text{corr}(x,y) = \frac{(0-0.5)(1-0.5) + (1-0.5)(0-0.5) + (0-0.5)(1-0.5) + (1-0.5)(0-0.5)}{\sqrt{[(0-0.5)^2 + (1-0.5)^2 + (0-0.5)^2 + (1-0.5)^2].[(1-0.5)^2 + (0-0.5)^2 + (1-0.5)^2 + (0-0.5)^2]}}$$

$\Rightarrow -1$.

→ Euclidean $= \sqrt{(0-1)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2}$

$= 2$.

→ Jaccard $= 0$.

(iii) $x = (0,-1,0,1)$ $y = (1,0,-1,0)$
cosine, corr, Euclidean.

→ $\cos(x,y) = 0$

→ corr$(x,y) = 0$.

→ Euclidean $= \sqrt{(0-1)^2 + (-1-0)^2 + (0+1)^2 + (1-0)^2}$

$= 2$.

(iv) $x = (1,1,0,1,0,1)$ $y = (1,1,1,0,0,1)$
cosine, corr, Jaccard.

$\rightarrow \cos(x,y) = \dfrac{3}{\sqrt{1^2+1^2+1^2+1^2}\ \sqrt{1^2+1^2+1^2+1^2}} = \dfrac{3}{4} = \underline{\underline{0.75}}.$

$\rightarrow \text{corr}(x,y) =$ mean of $x \Rightarrow \dfrac{4}{6} = 0.67.$
mean of $y \Rightarrow \dfrac{4}{6} = 0.67.$

$$\Rightarrow \frac{(1-0.67)(1-0.67)+(1-0.67)(1-0.67)+(0-0.67)(1-0.67)+(1-0.67)(0-0.67)+(0-0.67)(0-0.67)+(1-0.67)(1-0.67)}{\sqrt{\left[(1-0.67)^2+(1-0.67)^2+(0-0.67)^2+(1-0.67)^2+(0-0.67)^2+(1-0.67)^2\right]\left[(1-0.67)^2+(1-0.67)^2+(1-0.67)^2+(0-0.67)^2+(0-0.67)^2+(1-0.67)^2\right]}}$$

$\Rightarrow \underline{\underline{0.25}}.$

$\rightarrow \text{Jaccard}(x,y) \Rightarrow \dfrac{3}{\cancel{6}5} \Rightarrow \cancel{0.5}\ \underline{\underline{0.6}}$

$\longrightarrow$

(v) $x = (2, -1, 0, 2, 0, -3)$ $y = (-1, 1, -1, 0, 0, -1)$

cosine, corr.

$\rightarrow \cos(x, y) = -2 - 1 + 0 + 0 + 0 + 3. = 0.$

$\rightarrow$ corr (x, y) mean of $x \Rightarrow 0.$

mean of $y \Rightarrow \frac{-2}{6} = -1/3$

$\Rightarrow (2-0)(-1+1/3) + (-1-0)(1+1/3) + 0 + (2-0)(0+1/3) + 0 + (-3-0)(-1+1/3)$

$\Rightarrow \underline{0.}$

Q20).

a) [-1, 1]. Many ~~item~~ times the data has only positive entries and in that case the range is [0, 1]

b) Not necessarily. The values of their attributes differ by a constant factor.

c) For 2 vectors, x & y that have a mean of 0, corr(x, y) = cos(x, y).

$\longrightarrow$

d) Since all 100,000 points fall on the curve, there is a functional relationship between Euclidean distance and cosine similarity for normalized data. More specifically, there is an inverse relationship between cosine similarity & Euclidean distance

e) All points (100,000) fall on the curve, there is a functional relationship b/w euclidean distance and correlation for normalized data. More specifically, there is an inverse relationship b/w correlation & Euclidean distance

f) Let x & y be 2 vectors where each vector has an L_2 & length of 1. For such vectors, the variance is just n times the sum of its squared attribute values and the correlation b/w the two vectors is their dot product divided by n.

$$d(x,y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}$$

$$= \sqrt{\sum_{k=1}^{n} x_k^2 - 2x_k y_k + y_k^2}$$

$$= \sqrt{1 - 2\cos(x,y) + 1}$$

$$= \sqrt{2(1-\cos(x,y))}$$

g) let x and y be 2 vectors where each vector has ~~an~~ mean of 0 and a std dev of 1. For such vectors, the variance (std dev squared) is just n times the sum of its squared attribute values & the correlation b/w the 2 vectors is their dot product divided by n.

$$d(x,y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$$

$$= \sqrt{\sum_{k=1}^{n} x_k^2 - 2x_k y_k + y_k^2}$$

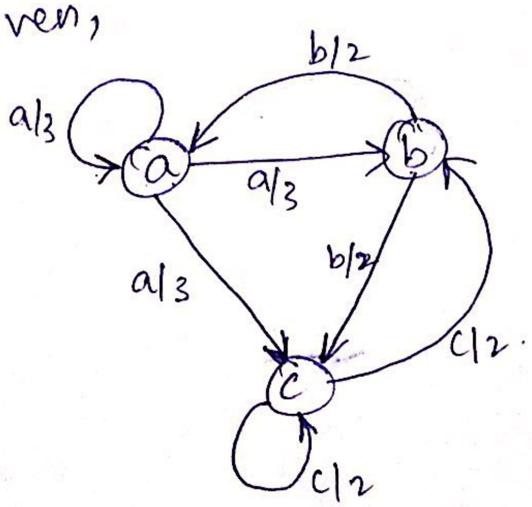
$$= \sqrt{n - 2n\,corr(x,y) + n}$$

$$= \sqrt{2n(1 - corr(x,y))}$$

1.3 Leskovec chp 5.

Ex 5.1.1

Given,



flow equations:-

$$r_a = r_a/3 + r_b/2$$

$$r_b = r_a/3 + r_c/2.$$

$$r_c = r_a/3 + r_b/2 + r_c/2.$$

Transition matrix

$$\begin{array}{c|ccc} & A & B & C \\ A & & & \\ B & & & \\ C & & & \end{array} \quad \begin{bmatrix} 1/3 & 1/2 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/2 \end{bmatrix}$$

choosing $x = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$ by iteration method using power theorem.

$$\begin{bmatrix} 1/3 & 1/2 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/2 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0.2778 \\ 0.2778 \\ 0.444 \end{bmatrix}$$

$$= \begin{bmatrix} 1/3 & 1/2 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0.2315 \\ 0.3148 \\ 0.4537 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/2 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0.2345 \\ 0.3040 \\ 0.4614 \end{bmatrix} \ldots\ldots$$

By, equation method, we get $(M\lambda = \lambda)$

$$\lambda = \left[\frac{3}{13}, \frac{4}{13}, \frac{6}{13}\right]^T.$$

Ex. 5.1.2

$\beta = 0.8 = 4/5$.

$\cancel{v' = \beta M}$ $v' = \beta M v + (1-\beta) e/n$.

$$\Rightarrow \begin{bmatrix} 4/15 & 2/5 & 0 \\ 4/15 & 0 & 2/5 \\ 4/15 & 2/5 & 2/5 \end{bmatrix} v + \begin{bmatrix} 1/15 \\ 1/15 \\ 1/15 \end{bmatrix}$$

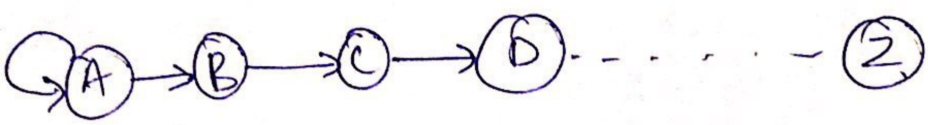
we get, $v = \left[\frac{7}{27}, \frac{25}{81}, \frac{35}{81} \right]^T$

By iteration method, we get the list :-

$$\begin{bmatrix} 0.3333 \\ 0.3333 \\ 0.3333 \end{bmatrix}, \begin{bmatrix} 0.2888 \\ 0.2888 \\ 0.4222 \end{bmatrix}, \begin{bmatrix} 0.2592 \\ 0.3125 \\ 0.4281 \end{bmatrix}, \begin{bmatrix} 0.2608 \\ 0.3071 \\ 0.4321 \end{bmatrix} \cdots\cdots$$

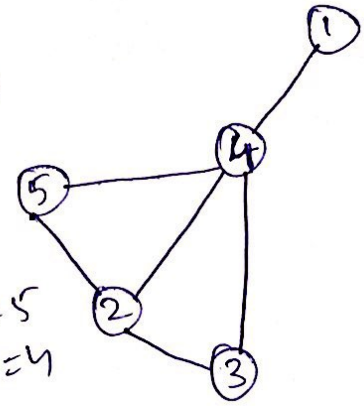
Ex 5.1.6.

In the case of chain of dead ends, headed by a node with a self loop there exists only one head node with self-loop and page-rank of 1. Page rank of the remaining nodes will be 1/2.



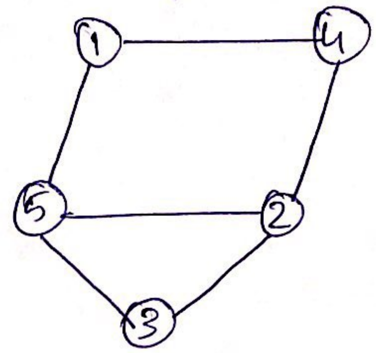
1.4. Centrality Measures

fig a)



n=5
n-1=4

fig b).



n=5
n-1=4

for fig(a).

a) Normalized degree centrality (NDC) for each node

Node.	$C_D(v)$/Link	NDC/($C_D^*(v)$).
1	1	1/4 = 0.25
2	3	3/4 = 0.75
3	2	1/2 = 0.5
4	4	1
5	2.	1/2 = 0.5

$$C_D^*(v) = \frac{1}{n-1} C_D(v)$$

$$C_D^*(1) = \frac{1}{4} \times 1 = 1/4.$$

$$C_D^*(2) = \frac{1}{4} \times 3 = 3/4$$

$$C_D^*(3) = \frac{1}{4} \times 2 = 1/2$$

$$C_D^*(4) = \frac{1}{4} \times 4 = 1$$

$$C_D^*(5) = \frac{1}{4} \times 2 = 1/2$$

b) Normalized closeness centrality of each node.

Node B

$C_c^*(v) = (n-1)\, C_c(v).$ | $C_c(v) = \frac{1}{\sum_j d(v,j)}.$

$C_c^*(1) = 4 \times \frac{1}{1+2+2+2}$

$= 4 \times \frac{1}{7} = 4/7.$

$C_c^*(2) = 4 \times \frac{1}{1+1+1+2} \Rightarrow 4/5.$

$C_c^*(3) = 4 \times \frac{1}{1+2+1+2} \Rightarrow 4/6 = 2/3$

$C_c^*(4) = 4 \times \frac{1}{1+1+1+1} = \frac{4}{4} = 1$

$C_c^*(5) = 4 \times \frac{1}{1+1+2+2} \Rightarrow \frac{4}{6} = 2/3$

Node	Score	Normalized Clossness Centrality
1	1/7.	4/7
2	1/5	4/5
3	1/6	2/3
4	1/4	1
5	1/6.	2/3

ⓒ Normalized b/wness centrality of each node.

$$C_B(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$$C^*_B(v) = \frac{C_B(v)}{2\binom{n-1}{2}}.$$

$C_B(1) \Rightarrow$

$v = 1$, ~~ss~~ $s = 2, 3, 4, 5$.

$2 \to 3. = 0/1 = 0$

$2 \to 4 = 0/1 = 0$

$2 \to 5 = 0/1 = 0$

$3 \to 4 = 0/1 = 0$

$3 \to 5 = 0/2 = 0$

$4 \to 5 = 0/1 = 0.$

$\boxed{C_B(1) = 0}$

$C_B(2) \Rightarrow$

$1 \to 4. = 0/1$

$1 \to 3 = 0/1$

$1 \to 5 = 0/1$

$3 \to 4 = 0/1$

$3 \to 5. = 1/2.$

$4 \to 5. \ 0/1$

$C_B(2) = \frac{1}{2} \times 2. = 1.$

& $C_B^*(2) = \frac{1}{12}.$

$C_B(3) =$

$1 \to 2 = 0/1$

$1 \to 4 = 0/1$ $\quad\quad C_B(3) = 0$

$1 \to 5 = 0/1$

$2 \to 4 = 0/1$

$2 \to 5 = 0/1$

$4 \to 5 = 0/1$

$C_B(4) =$

$1 \to 2 = 1/1$ $\quad\quad$ Total $= 7/2$.

$1 \to 3 = 1/1$

$1 \to 5 = 1/1$ $\quad\quad C_B(4) = \frac{7}{2} \times 2 = 7$

$2 \to 3 = 0/1$ $\quad\quad C_B^*(4) = \frac{7}{12} = 0.5833$

$2 \to 5 = 0/1$

$3 \to 5 = 1/2$.

$\underline{C_B(5) =}$

$1 \to 2 = 0/1$

$1 \to 3 = 0/1$ $\quad\quad$ Total $= 0$

$1 \to 4 = 0/1$ $\quad\quad C_B(5) = 0$.

$2 \to 3 = 0/1$

$2 \to 4 = 0/1$

$3 \to 4 = 0/1$

for fig(b)

a)

Node	Link	NDC
1	2	$1/2 = 0.5$
2	3	$3/4 = 0.75$
3	2.	$1/2 = 0.5$
4	2	$1/2 = 0.5$
5	3	$3/4 = 0.75$

$$C_D^*(1) = \frac{1}{4} * 2 \Rightarrow 2/4 = 1/2$$

$$C_D^*(2) = \frac{1}{4} * 3 \Rightarrow 3/4$$

$$C_D^*(3) = \frac{1}{4} * 2 = \frac{2}{4} = 1/2$$

$$C_D^*(4) = \frac{1}{4} * 2 = \frac{2}{4} = 1/2$$

$$C_D^*(5) = \frac{1}{4} * 3 = 3/4$$

$\longrightarrow$

b)

$$C_c^*(1) = 4 \times \frac{1}{1+1+2+2} \Rightarrow \frac{4}{6} = 2/3$$

$$C_c^*(2) = 4 \times \frac{1}{1+1+1+2} \Rightarrow \frac{4}{5}$$

$$C_c^*(3) = 4 \times \frac{1}{1+1+2+2} = \frac{4}{6} = 2/3$$

$$C_c^*(4) = 4 \times \frac{1}{1+1+2+2} = \frac{4}{6} = 2/3.$$

$$C_c^*(5) = 4 \times \frac{1}{1+1+1+2} = \frac{4}{5}.$$

Node	Scores	Normalized closeness centrality
1	1/6	2/3 = 0.67
2	1/5	4/5 = 0.8
3	1/6	2/3 = 0.67
4	1/6	2/3 = 0.67.
5.	1/5.	4/5 = 0.8

c)

$C_B(1) =$

$2 \to 3 = 0/1$

$2 \to 4 = 0/1$

$2 \to 5 = 0/1$

$3 \to 4 = 0/1$

$3 \to 5 = 0/1$

$4 \to 5 = 1/2.$

total = 0.5.

$C_B(1) = \frac{1}{12}.$

$C_B(2) =$

$1 \to 3 = 0/1$

$1 \to 4 = 0/1$

$1 \to 5 = 0/1$

$3 \to 4 = 1/1$

$3 \to 5 = 0/1$

$4 \to 5 = 1/2$

total = 0.5

$C_B(2) = \frac{1}{12}.$

$C_B(3)$

$1 \to 2 = 0/2$

$1 \to 4 = 0/1$

$1 \to 5 = 0/1$

$2 \to 4 = 0/1$

$2 \to 5 = 0/1$

$4 \to 5 = 0/2$

total = 0

$C_B(3) = 0$

$C_B(4)$

$1 \to 2 = 1/2$

$1 \to 3 = 0/1$

$1 \to 5 = 0/1$

$2 \to 3 = 0/1$

$2 \to 5 = 0/1$

$3 \to 5 = 0/1$

total = 0.5

$C_B(4) = \frac{1}{12}.$

$C_B(5)$

$1 \to 2 = 1/2$

$1 \to 3 = 1/1$

$1 \to 4 = 0/1$

$2 \to 3 = 0/1$

$2 \to 4 = 0/1$

$3 \to 4 = 0/1$

total $\frac{1}{2} + 1 = 3/2$

$C_B(5) = 2 \times \frac{3}{2} = 3.$