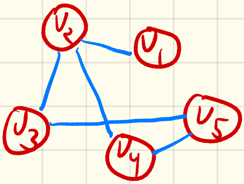
Representation

Let $G = (V, E)$ be a graph w/ n verts

We can write G as an $n \times n$ adjacency matrix

Example 1: Unweighted

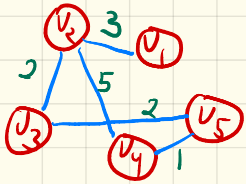


Graph G

$$
\begin{array}{c|ccccc}
 & v_1 & v_2 & v_3 & v_4 & v_5 \\
\hline
v_1 & 0 & 1 & 0 & 0 & 0 \\
v_2 & 1 & 0 & 1 & 1 & 0 \\
v_3 & 0 & 1 & 0 & 0 & 1 \\
v_4 & 0 & 1 & 0 & 0 & 1 \\
v_5 & 0 & 0 & 1 & 1 & 0
\end{array}
$$

Adjacency matrix of G

Example 2: Weighted



Graph G

$$
\begin{array}{c|ccccc}
 & v_1 & v_2 & v_3 & v_4 & v_5 \\
\hline
v_1 & 0 & 3 & 0 & 0 & 0 \\
v_2 & 3 & 0 & 2 & 5 & 0 \\
v_3 & 0 & 2 & 0 & 0 & 2 \\
v_4 & 0 & 5 & 0 & 0 & 1 \\
v_5 & 0 & 0 & 2 & 1 & 0
\end{array}
$$

Adjacency matrix of G

Graphs From Data

Graph $\xrightarrow{\text{sim}(x_i, x_j)}$ Similarity Matrix $\xrightarrow{\text{threshold}}$ Adj Matrix

$D =$

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

1. Pick similarity measure

example: $\|x_i - x_j\|_2$ ⟵ 2-norm "Euclidean dist"

$\|x_i - x_j\|_1$ ⟵ 1-norm "Manhattan dist"

$e^{-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}}$ ⟵ Gaussian ★ w/ $\sigma = 25$

$$\text{sim}_{25}(x_1, x_2) = e^{-\frac{\|x_1 - x_2\|_2^2}{2 \cdot 25^2}} = e^{-\frac{484.13}{2 * 25^2}} \approx .68$$

$D =$

	X_1	X_2	X_3
x_1	0.2	23	5.7
x_2	0.4	1	5.4
x_3	1.8	0.5	5.2
x_4	5.6	50	5.1
x_5	−0.5	34	5.3
x_6	0.4	19	5.4
x_7	1.1	11	5.5

$\xrightarrow{\text{sim}_{25}}$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1		0.68	0.67	0.55	0.91	0.99	0.89
x_2			1.0	0.14	0.42	0.77	0.92
x_3				0.14	0.41	0.76	0.92
x_4					0.79	0.45	0.29
x_5						0.83	0.65
x_6							0.95
x_7							

$$D = \begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ \hline x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

$$\xrightarrow{\text{sim}_{25}}$$

	x1	x2	x3	x4	x5	x6	x7
x1		0.68	0.67	0.55	0.91	0.99	0.89
x2			1.0	0.14	0.42	0.77	0.92
x3				0.14	0.41	0.76	0.92
x4					0.79	0.45	0.29
x5						0.83	0.65
x6							0.95
x7							

2. Pick a cutoff

$$A(i,j) = \begin{cases} 1 & \text{if } \text{sim}(x_i, x_j) \geq \tau \\ 0 & \text{otherwise} \end{cases}$$

Pick $\tau = .94$

$\text{sim}_{25}(x_1, x_2) = .68$: $\boxed{\cancel{.68 \overset{?}{\geq} .94}}$ $\Rightarrow A(i,j) = 0$

$\text{sim}_{25}(x_1, x_6) = .99$: $\boxed{.99 \overset{?}{\geq} .94}$ ✓ $\Rightarrow A(i,j) = 1$

	x1	x2	x3	x4	x5	x6	x7
x1		0.68	0.67	0.55	0.91	0.99	0.89
x2			1.0	0.14	0.42	0.77	0.92
x3				0.14	0.41	0.76	0.92
x4					0.79	0.45	0.29
x5						0.83	0.65
x6							0.95
x7							

$$\xrightarrow{\tau = .94}$$

$A =$

	x1	x2	x3	x4	x5	x6	x7
x1	0	0	0	0	1	1	1
x2	0	0	1	0	0	1	1
x3	0	1	0	0	0	1	0
x4	0	0	0	0	1	0	0
x5	1	0	0	1	0	1	0
x6	1	1	1	0	1	0	1
x7	1	1	0	0	0	1	0

We now have a graph

$A =$

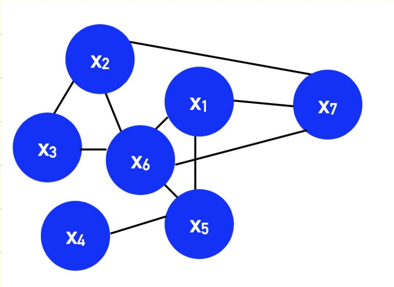
	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0	0	0	0	1	1	1
x_2	0	0	1	0	0	1	1
x_3	0	1	0	0	0	1	0
x_4	0	0	0	0	1	0	0
x_5	1	0	0	1	0	1	0
x_6	1	1	1	0	1	0	1
x_7	1	1	0	0	0	1	0

Visualized as

$$D = \begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ \hline x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

$\sigma = 25$
$\tau = .94$

$\longrightarrow$



Picking σ and τ:

Problem dependent ⍨

But we can understand what they mean!

Consider $sim_\sigma(x_i, x_j) = e^{-\frac{\boxed{\|x_i - x_j\|_2^2}}{2\sigma^2}}$ ← Squared Ecidean distance!

$$D = \begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ \hline x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

$\downarrow \|x_i - x_j\|_2$

	x1	x2	x3	x4	x5	x6	x7
x1		22.0	22.6	27.5	11.03	4.02	12.04
x2			1.5	49.28	33.01	18.0	10.03
x3				49.65	33.58	18.55	10.53
x4					17.12	31.43	39.26
x5						15.03	23.06
x6							8.03
x7							

Euclidean Dist Matrix

$\xrightarrow{sim_{10}}$

	x1	x2	x3	x4	x5	x6	x7
x1		0.09	0.08	0.02	0.54	0.92	0.48
x2			0.99	0.00	0.004	0.20	0.61
x3				0.00	0.004	0.18	0.57
x4					0.23	0.007	0.00
x5						0.32	0.07
x6							0.72
x7							

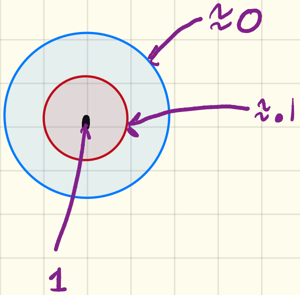
$\xrightarrow{Sim_{25}}$

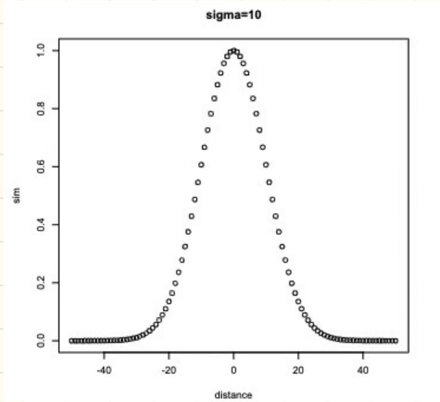
	x1	x2	x3	x4	x5	x6	x7
x1		0.68	0.67	0.55	0.91	0.99	0.89
x2			1.0	0.14	0.42	0.77	0.92
x3				0.14	0.41	0.76	0.92
x4					0.79	0.45	0.29
x5						0.83	0.65
x6							0.95
x7							

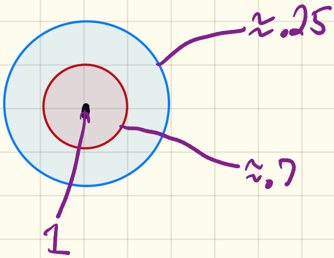
$\xrightarrow{sim_{50}}$

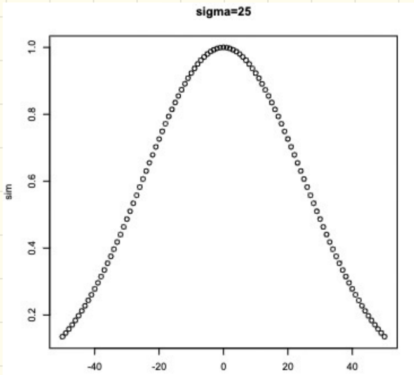
	x1	x2	x3	x4	x5	x6	x7
x1		0.91	0.90	0.86	0.98	1.0	0.97
x2			1.0	0.62	0.80	0.94	0.98
x3				0.61	0.80	0.93	0.98
x4					0.94	0.82	0.73
x5						0.96	0.90
x6							0.99
x7							

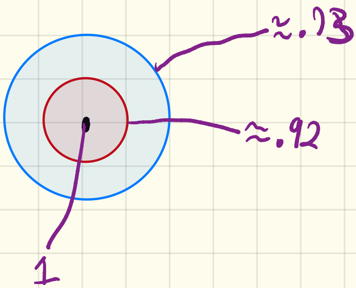
dist is 20
dist is 40

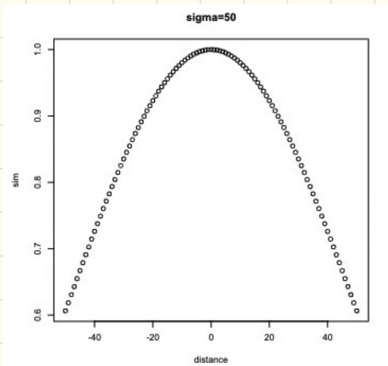








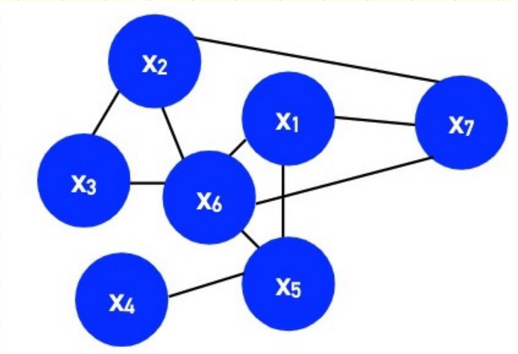




Measures of Centrality

Q: How to measure the importance of a node?

Graph



Adj Matrix

$A =$

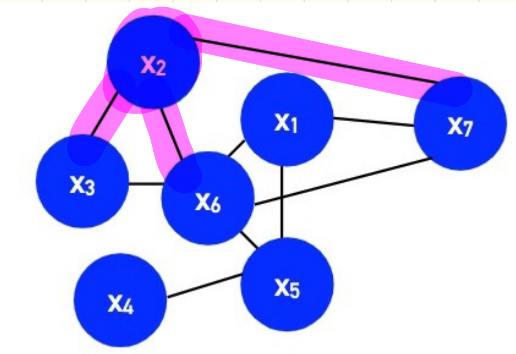
	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0	0	0	0	1	1	1
x_2	0	0	1	0	0	1	1
x_3	0	1	0	0	0	1	0
x_4	0	0	0	0	1	0	0
x_5	1	0	0	1	0	1	0
x_6	1	1	1	0	1	0	1
x_7	1	1	0	0	0	1	0

Degree Centrality - nodes connected to other nodes are central

$$d(x_i) = \sum_{j=1}^{n} A(i,j)$$

Example

$$d(x_2) = \sum_{j=1}^{7} A(2,j) = 0+0+1+0+0+1+1 = 3$$

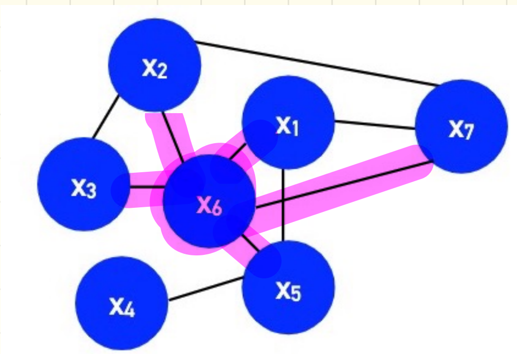


$A =$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0	0	0	0	1	1	1
x_2	0	0	1	0	0	1	1
x_3	0	1	0	0	0	1	0
x_4	0	0	0	0	1	0	0
x_5	1	0	0	1	0	1	0
x_6	1	1	1	0	1	0	1
x_7	1	1	0	0	0	1	0

$$d(x_6) = \sum_{j=1}^{7} A(6,j) = 5$$

Graph



Adj Matrix

$A =$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0	0	0	0	1	1	1
x_2	0	0	1	0	0	1	1
x_3	0	1	0	0	0	1	0
x_4	0	0	0	0	1	0	0
x_5	1	0	0	1	0	1	0
x_6	1	1	1	0	1	0	1
x_7	1	1	0	0	0	1	0

Closeness Centrality – nodes close to lots of nodes are central

$$cc(x_i) = \frac{1}{\sum_{j=1}^{n} \boxed{d(x_i, x_j)}}$$

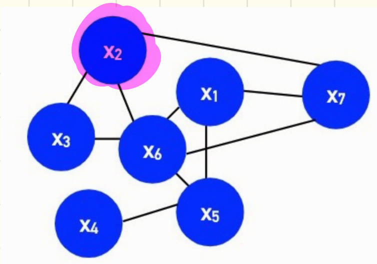
shortest path between x_i and x_j

Example:

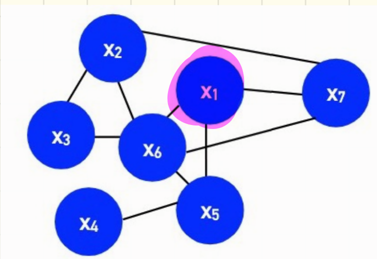
$$cc(x_2) = \frac{1}{\underset{2}{d(x_2,x_1)} + \underset{0}{d(x_2,x_2)} + \cdots + d(x_2,x_7)}$$

x_3 x_4 x_5 x_6 x_7

$$= \frac{1}{2+0+1+3+2+1+1} = \frac{1}{10}$$



$$cc(x_1) = \frac{1}{0+2+2+2+1+1+1} = \frac{1}{9}$$



Betweenness Centrality - nodes on many shortest paths

Let η_{jk} be # of shortest paths between x_j and x_k

$\eta_{jk}(x_i)$ # of shortest paths between x_j and x_k that pass through x_i

Betweeness Centrality $$bc(x_i) = \sum_{j \neq i} \sum_{k \neq j} \frac{\eta_{jk}(x_i)}{\eta_{jk}}$$

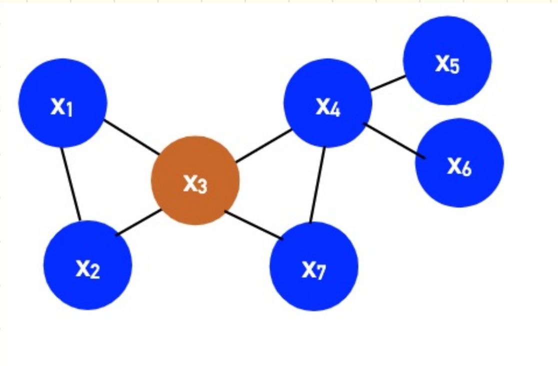
Example: $bc(x_3)$

$\eta_{12} = 1$ $\quad \eta_{12}(x_3) = 0$

~~$\eta_{13} =$~~

$\eta_{14} = 1$ $\quad \eta(x_3) = 1$

$\vdots$



	Shortest paths	Shortest paths passing through x_3 ($\eta_{jk}(x_3)$)
η_{12}	1	0
η_{14}	1	1
η_{15}	1	1
η_{16}	1	1
η_{17}	1	1
η_{24}	1	1
η_{25}	1	1
η_{26}	1	1
η_{27}	1	1
η_{45}	1	0
η_{46}	1	0
η_{47}	1	0
η_{56}	1	0
η_{57}	1	0
η_{67}	1	0

$$bc(x_3) = \sum_{j \neq i} \sum_{k \neq i} \frac{\eta_{jk}(x_3)}{\eta_{jk}}$$

$$= \frac{\eta_{11}(x_3)}{\eta_{11}} + \frac{\eta_{12}(x_3)}{\eta_{12}} + \frac{\eta_{14}(x_3)}{\eta_{14}} + \cdots = 8$$

Eccentricity – the less eccentric a node is the more central it is

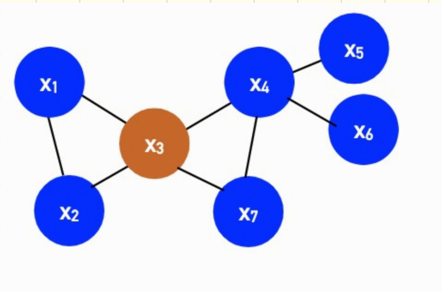
Let $d(x_i, x_j)$ be the length of the shortest path between x_i and x_j

Eccentricity of x_i is

$$e(x_i) = \max_j \{ d(x_i, x_j) \}$$

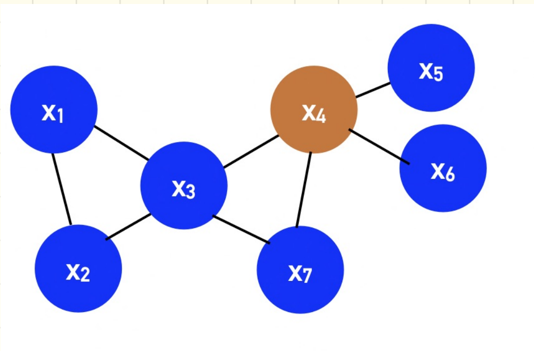
Example:

$e(x_3) = 2$



x_i	$d(x_3, x_i)$
x_1	1
x_2	1
x_3	0
x_4	1
x_5	2
x_6	2
x_7	1

$e(x_4) = 2$



x_i	$d(x_4, x_i)$
x_1	2
x_2	2
x_3	1
x_4	0
x_5	1
x_6	1
x_7	1

Prestige Centrality - a node is prestigious if prestigious nodes point to it

Let $N_i = \{x_j \in V : d(x_i, x_j) = 1\}$ ← all verts 1 hop from x_i

The prestige of x_i is

$$p(x_i) = \sum_{x_j \in N_i} p(x_j)$$

Example:

$$p(x_3) = p(x_1) + p(x_5) + p(x_4)$$