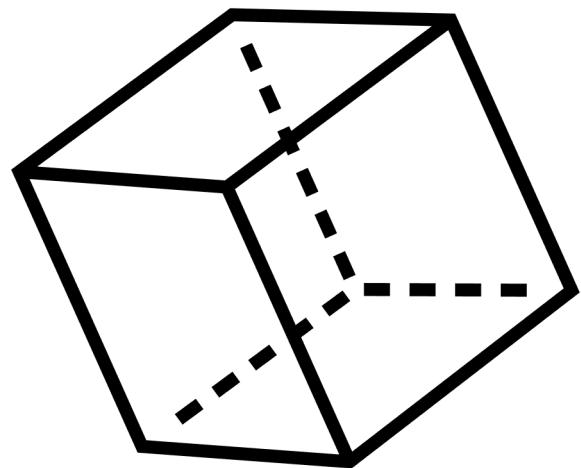


ULSAN NATIONAL INSTITUTE OF
SCIENCE AND TECHNOLOGY

Fifth Week

The Network Measurement



1. The Network Measurement

5.1 Centrality: Degree

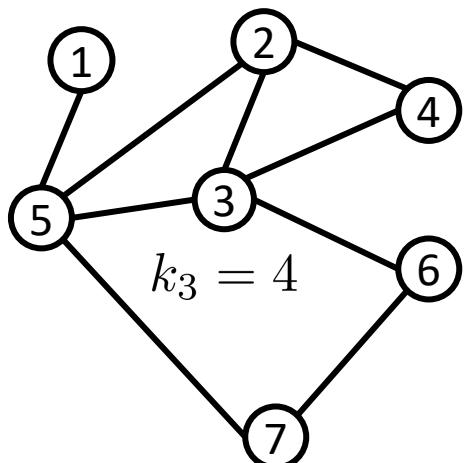
Definition 5.1.1 (Degree)

The **degree** of a vertex in a graph is the number of edges connected to it.

- ★ We will denote the degree of vertex i by k_i .
- ★ For an undirected graph of n vertices the degree can be written in terms of the adjacency matrix as

$$k_i = \sum_{j=1}^n A_{ij}.$$

Example (Degree)



$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$
$$k_3 = \sum_{j=1}^n A_{3j} = 4.$$

5.1 Centrality: Degree

Note

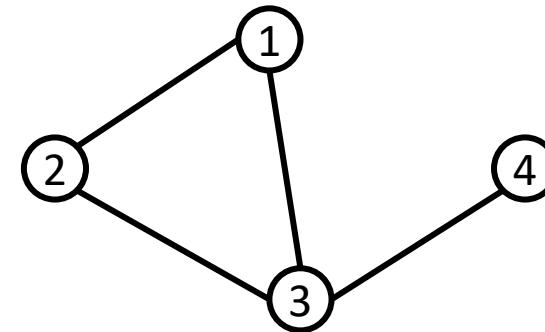
If there are m -edges in total then there are $2m$ ends of edges

$$\Rightarrow 2m = \sum_{i=1}^n k_i$$

$$\Rightarrow m = \frac{1}{2} \sum_{i=1}^n k_i$$

$$= \frac{1}{2} \sum_{i=1}^n \left(\sum_{j=1}^n A_{ij} \right)$$

$$= \frac{1}{2} \sum_{i,j=1}^n A_{ij}$$



edges = 4, $2m = 8$

$k_1 = 2, k_2 = 2, k_3 = 3, k_4 = 1$

5.1 Centrality: Degree

Note

The mean degree c of a vertex in an undirected graph is

$$\begin{aligned} c &= \frac{1}{n} \sum_{i=1}^n k_i, \\ \Rightarrow cn &= \sum_{i=1}^n k_i = 2m \\ \Rightarrow c &= \frac{2m}{n} \end{aligned}$$

Definition 5.1.2 (Density)

The maximum possible number of edges in a simple graph is $\binom{n}{2} = \frac{n(n-1)}{2}$.

The **connectance** or **density** ρ of a graph is the fraction of these edges that are actually present:

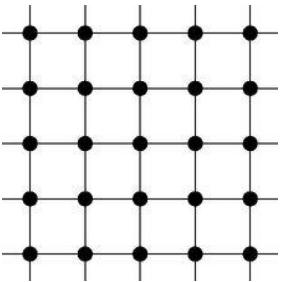
$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{2c}{n(n-1)} = \frac{c}{n-1}$$

5.1 Centrality: Degree

Note

$$0 < \rho < 1$$

- ① As $n \rightarrow \infty$, $\rho \rightarrow \text{constant} > 0 \Rightarrow$ A network is **dense**.
- ② As $n \rightarrow \infty$, $\rho \rightarrow 0 \Rightarrow$ A network is **sparse**.

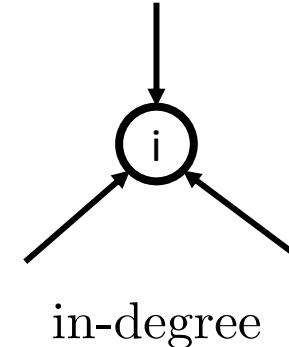
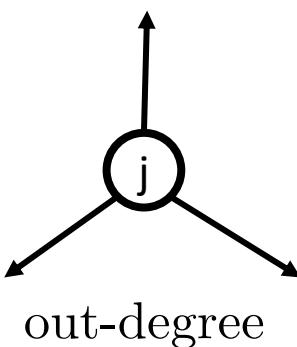


$k_i = 4 \rightarrow$ regular network (same degree)

5.1 Centrality: Degree

Definition 5.1.3 (in,out - degree)

In a directed network each vertex has two degrees. The **in-degree** is the number of ingoing edges connected to a vertex and the **out-degree** is the number of outgoing edges.



Note

The adjacency matrix of a directed network has element $A_{ij} = 1$ if there is an edge from *j* to *i*, in- and out-degrees can be written

$$k_i^{in} = \sum_{j=1}^n A_{ij}, \quad k_j^{out} = \sum_{i=1}^n A_{ij}$$

(*n* is number of vertices)

5.1 Centrality: Degree

Note

The number of edges m in a directed network is equal to the total number of ingoing ends of edges at all vertices, or equivalently to the total number of outgoing ends of edges, so

$$m = \sum_{i=1}^n k_i^{in} = \sum_{j=1}^n k_j^{out} = \sum_{ij} A_{ij}$$

Note

Thus the mean in-degree c_{in} and the mean out-degree c_{out} of every directed network are equal:

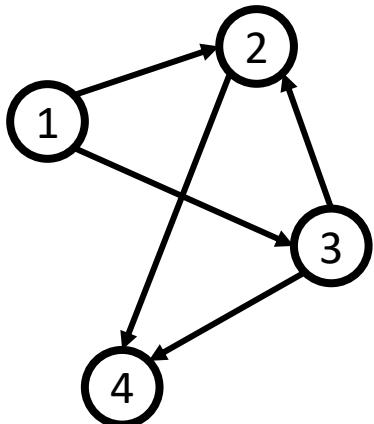
$$c_{in} = \frac{1}{n} \sum_{i=1}^n k_i^{in} = \frac{1}{n} \sum_{j=1}^n k_j^{out} = c_{out}$$

$$c = c_{in} = c_{out} = \frac{m}{n}$$

(n and m are number of vertices and edges in total)

5.1 Centrality: Degree

Example



$$n = 4, m = 5$$

$$k_1^{in} = 0, k_2^{in} = 2, k_3^{in} = 1, k_4^{in} = 2$$

$$k_1^{out} = 2, k_2^{out} = 1, k_3^{out} = 2, k_4^{out} = 0$$

$$c_{in} = \frac{0 + 2 + 1 + 2}{4} = \frac{5}{4} = \frac{2 + 1 + 2 + 0}{4} = c_{out}$$

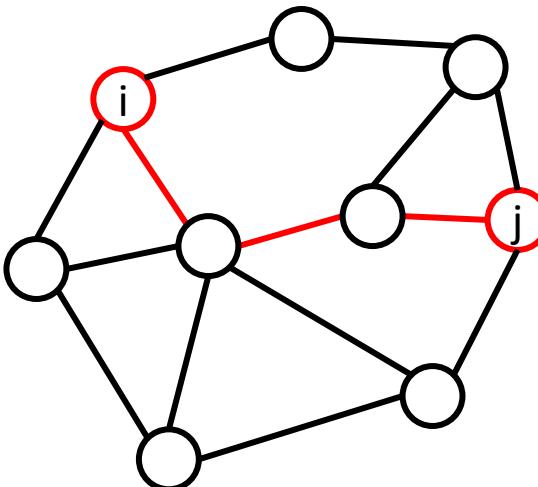
5.2 Centrality: Path

Definition 5.2.1 (Path)

The **path** in a network is any sequence of vertices such that every consecutive pair of vertices in the sequence is connected by an edge in the network.

The **length of a path** in a network is the number of edges traversed along the path (not the number of vertices).

Example (Path)

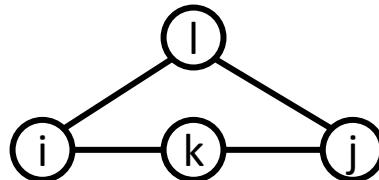


Red color vertices are path between i and j , and length of a path is 3.

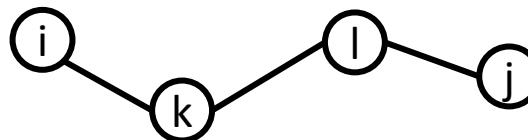
5.2 Centrality: Path

* $N_{ij}^{(n)}$: The total number of paths of length n from j to i .

i) $N_{ij}^{(2)} = \sum_{k=0}^n A_{ik}A_{kj} = [A^2]_{ij}$



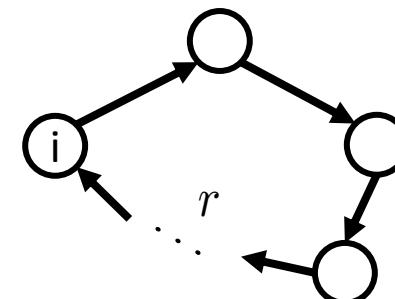
ii) $N_{ij}^3 = \sum_{k,l=1}^n A_{ik}A_{kl}A_{lj} = [A^3]_{ij}$



$$\therefore [N^{(r)}]_{ij} = [A^r]_{ij}$$

A special case of this result is that the number of paths of length r that start and end at the same vertex i is

$$[N^{(r)}]_{ii} = [A^r]_{ii}$$



5.2 Centrality: Path

Note

The total number L_r of loops of length r anywhere in a network is the sum of this quantity over all possible starting points i :

$$L_r = \sum_{i=1}^n [A^r]_{ii} = \text{Tr}(A^r) \quad - (*)$$

This expression counts separately loops consisting of the same vertices in the same order but with different starting points.



Equation (*) can also be expressed in terms of the eigenvalues of the adjacency matrix.

5.2 Centrality: Path

Note

For an undirected network, A is symmetric and all real eigenvalues.

$$A = UKU^\top$$

U : orthogonal matrix ($UU^\top = I$), $\text{Col}_i(U)$ is eigenvectors of A

K : a diagonal matrix whose element is eigenvalues of A

$$\underline{A^r} = (UKU^\top)^r = (UKU^\top)(UKU^\top) \cdots (UKU^\top) = \underline{UK^rU^\top}$$

$$\therefore \underline{L_r} = \text{Tr}(A^r) = \text{Tr}(UK^rU^\top)$$

Recall

$$\begin{aligned} \therefore \text{Tr}(PQ) &= \sum_{i=1}^n (PQ)_{ii} = \sum_{i=1}^n \left[\sum_{j=1}^n P_{ij} Q_{ji} \right] \\ \text{Tr}(PQ) &= \text{Tr}(QP) \\ &= \sum_{i=1}^n \left[\sum_{j=1}^n Q_{ji} P_{ij} \right] = \sum_{j=1}^n \left[\sum_{i=1}^n Q_{ji} P_{ij} \right] = \sum_{j=1}^n (QP)_{jj} = \text{Tr}(QP) \quad \square \end{aligned}$$

5.2 Centrality: Path

Note

$$\therefore L_r = \text{Tr}(UK^kU^\top) = \text{Tr}(K^rUU^\top) = \text{Tr}(K^r) = \sum_{i=1}^n \kappa_i^r$$

where κ_i is an eigenvalues of A , $K = \begin{bmatrix} k_1 & 0 & \cdots & 0 \\ 0 & k_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_n \end{bmatrix}$

Theorem 5.1.1

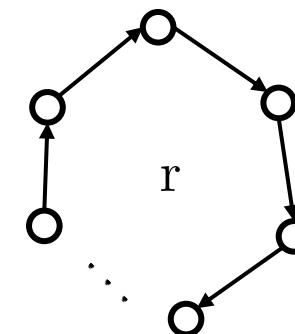
All of eigenvalues are zero \Leftrightarrow a network is acycle.

Proof) \Leftrightarrow acycle \Rightarrow a strictly upper triangular \Rightarrow zero eigenvalues

\Rightarrow Suppose a network is cycle, then $\exists r > 0$ such that $L_r > 0$

then, $L_r = \sum_{i=1}^n \kappa_i^r > 0 \Rightarrow \exists \kappa_j > 0$ for some j .

□



5.2 Centrality: Path

Note

For directed networks, the situation is more complicated. In general, asymmetric adjacency matrices and some asymmetric matrices cannot be diagonalized.

Every real matrix, whether diagonalizable or not, can be written in the form $A = QTQ^\top$, where Q is an orthogonal matrix ($QQ^\top = I$) and T is an upper triangular matrix. Q

Note (Schur decomposition)

Assume $Ax = \kappa x$

$$Ax = QTQ^\top x = \kappa x \Rightarrow TQ^\top x = \kappa Q^\top x$$

$\therefore \kappa$ is an eigenvalue of T .

5.2 Centrality: Path

Note

$$T = \begin{bmatrix} t_{11} & & & \mathbf{x} \\ & t_{22} & & \\ & & \ddots & \\ 0 & & & t_{nn} \end{bmatrix} \Rightarrow \text{eigenvalue of } T \text{ is diagonal component}$$

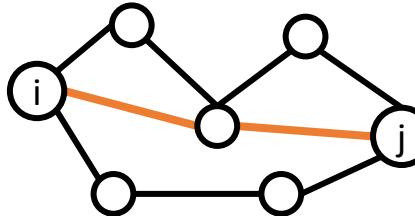
$$\Rightarrow T = \begin{bmatrix} \kappa_1 & & & \mathbf{x} \\ & \kappa_2 & & \\ & & \ddots & \\ 0 & & & \kappa_n \end{bmatrix} \Rightarrow T^r = \begin{bmatrix} \kappa_1^r & & & \mathbf{x} \\ & \kappa_2^r & & \\ & & \ddots & \\ 0 & & & \kappa_{nn}^r \end{bmatrix}$$

$$\begin{aligned} \therefore L_r &= \text{Tr}(A^r) = \text{Tr}(QTQ^\top)^r \\ &= \text{Tr}(QT^rQ^\top) = \text{Tr}(T^rQ^\top Q) \\ &= \text{Tr}(T^r) = \sum_{i=1}^n \kappa_i^r \end{aligned}$$

5.2 Centrality: Path

Definition 5.2.2 (Geodesic Paths)

A **geodesic path**, also called simply a shortest path, is a path between two vertices such that no shorter path exists:



Geodesic path between i and j is 2.

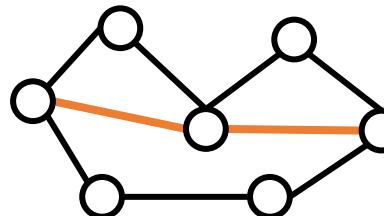
Definition 5.2.3 (Geodesic Distance)

The length of a geodesic path, often called the **geodesic distance** or **shortest distance**

The geodesic distance between vertices i and j is the smallest value of r such that $A_{ij}^r > 0$.

Definition 5.2.4 (Diameter)

The **diameter** of a graph is the length of the longest geodesic path between any pair of vertices in the network



Diameter is 4

5.2 Centrality: Path

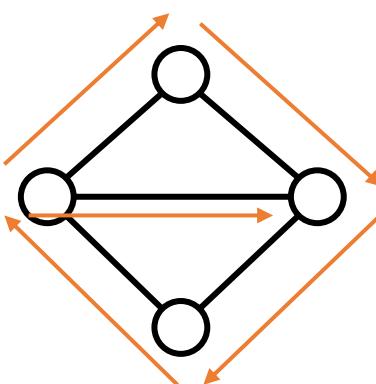
Definition 5.2.5 (Eulerian path)

A **Hamiltonian path** is a path that visits each vertex exactly once.

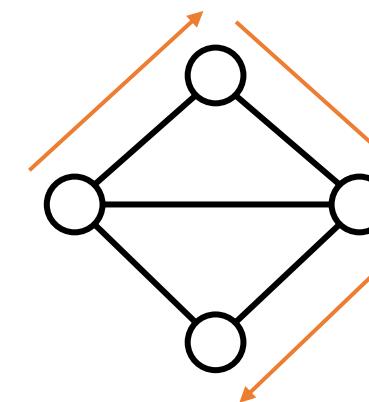
Definition 5.2.6 (Hamiltonian path)

An **Eulerian path** is a path that traverses each edge in a network exactly once.

Example



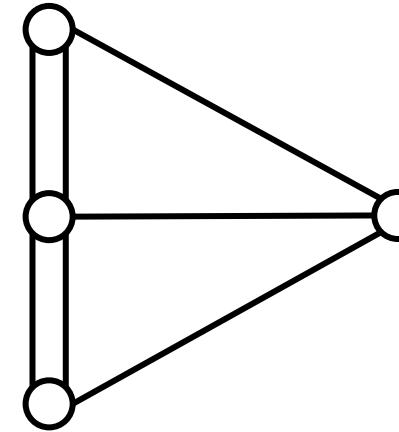
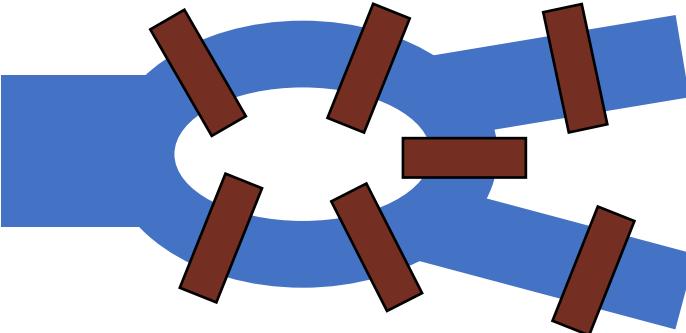
Eulerian path



Hamiltonian path

5.2 Centrality: Path

Example (The Königsberg bridges)



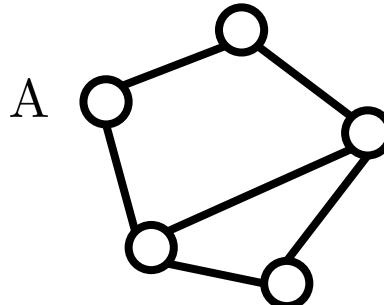
The Königsberg bridges.

Any Eulerian path must both enter and leave every vertex it passes through except the first and last, there can at most be two vertices in the network with odd degree if such a path is to exist. Since all four vertices in the Königsberg network have odd degree, the bridge problem necessarily has no solution.

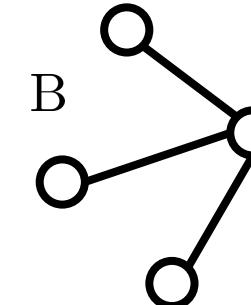
5.2 Path: Component

Definition 5.2.7 (Component)

The subgroups in a network like that of bellow figure are called **components**.



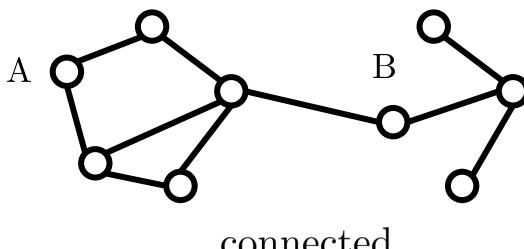
component 1



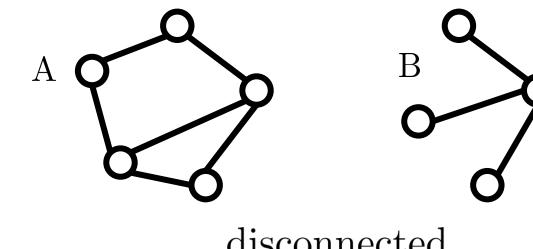
component 2

Definition 5.2.8 (disconnected, connected)

There is no path from the vertex labeled A to the vertex labeled B. A network of this kind is said to be **disconnected**. Conversely, if there is a path from every vertex in a network to every other vertex in the network is **connected**.



connected



disconnected

5.3 Path: Independent path

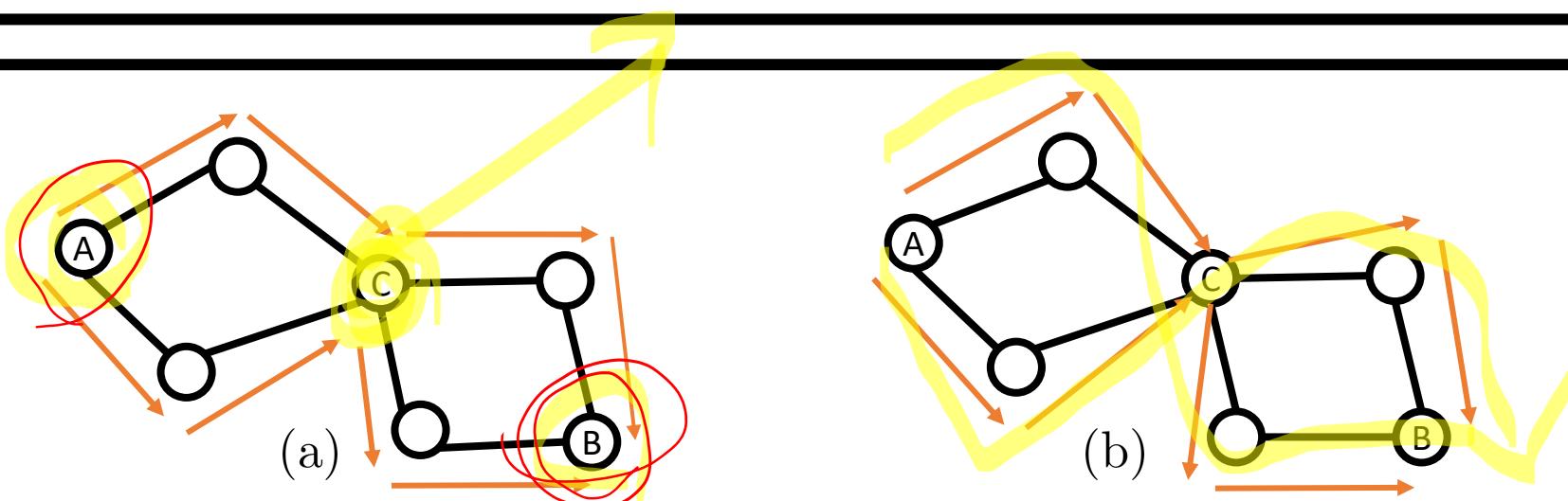
Definition 5.3.1 (Independent path)

Two paths connecting a given pair of vertices are **edge-independent** if they share no edges.

Two paths are **vertex-independent** (or **node-independent**) if they share no vertices other than the starting and ending vertices.

Path independent \Leftrightarrow Vertex independent

Example



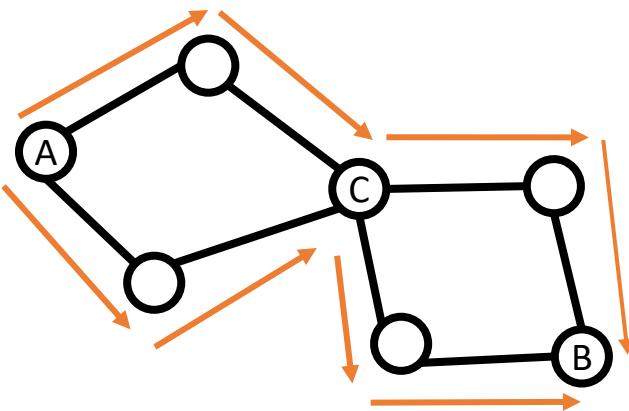
(a) There are two edge-independent paths from A to B in this figure, as denoted by the arrows, but there is only one vertex-independent path, because all paths must pass through the center vertex C. (b) The edge-independent paths are not unique; there are two different ways of choosing two independent paths from A to B in this case.

5.3 Path: Independent path

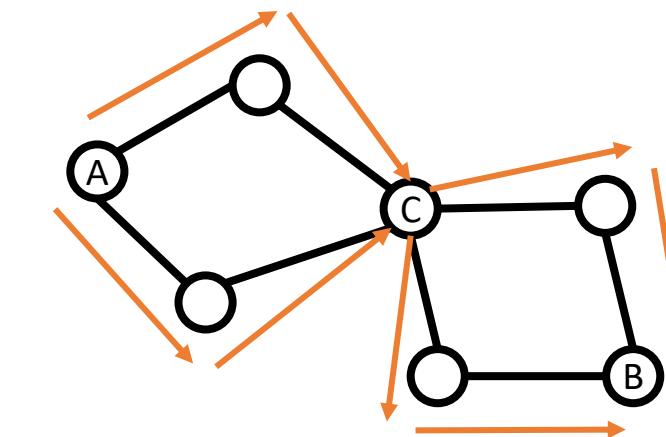
Definition 5.3.2 (Connectivity)

The number of independent paths between a pair of vertices is called the **connectivity** of the vertices. If we wish to be explicit about whether we are considering **edge-** or **vertex independence**, we refer to **edge** or **vertex connectivity**.

Example



Vertex-connectivity = 1



Edge-connectivity = 2

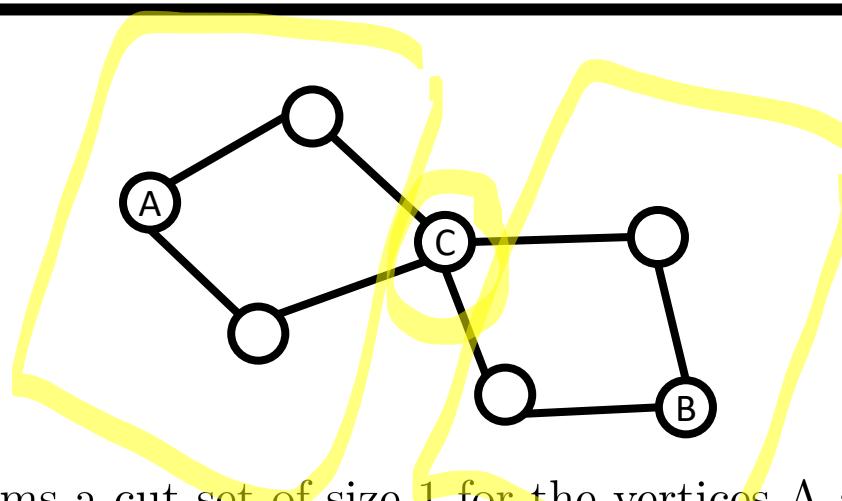
5.3 Path: Independent path

Definition 5.3.3 (Cut set)

A **cut set**, or more properly a **vertex cut set**, is a set of vertices whose removal will disconnect a specified pair of vertices.

An **edge cut set** is the equivalent construct for edges—it is a set of edges whose removal will disconnect a specified pair of vertices.

Example



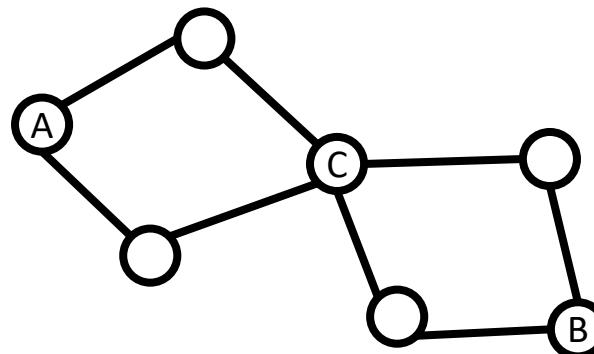
The central vertex C forms a cut set of size 1 for the vertices A and B. There are also other cut sets for A and B in this network, although all the others are larger than size 1.

5.3 Path: Independent path

Definition 5.3.4 (Minimum Cut set)

A **minimum cut set** is the smallest cut set that will disconnect a specified pair of vertices.

Example

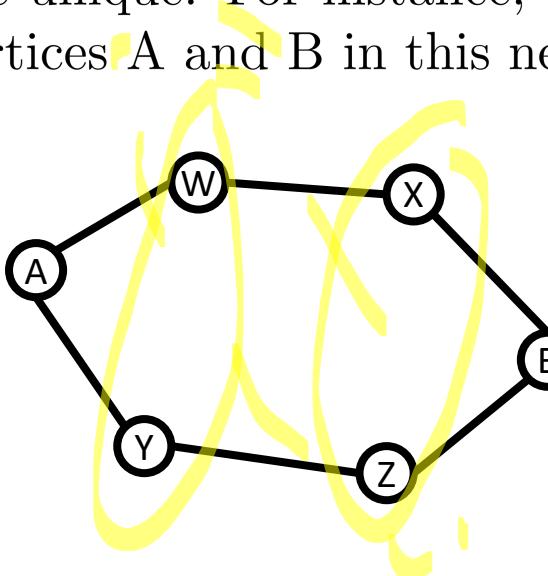


The single vertex C is a minimum vertex cut set for vertices A and B.

5.3 Path: Independent path

Note

A minimum cut set need not be unique. For instance, there is a variety of minimum vertex cut sets of size two between the vertices A and B in this network:



$\{W,Y\}$, $\{W,Z\}$, $\{X,Y\}$, and $\{X,Z\}$ are all minimum cut sets for this network.

Theorem 5.3.1 (Menger's theorem)

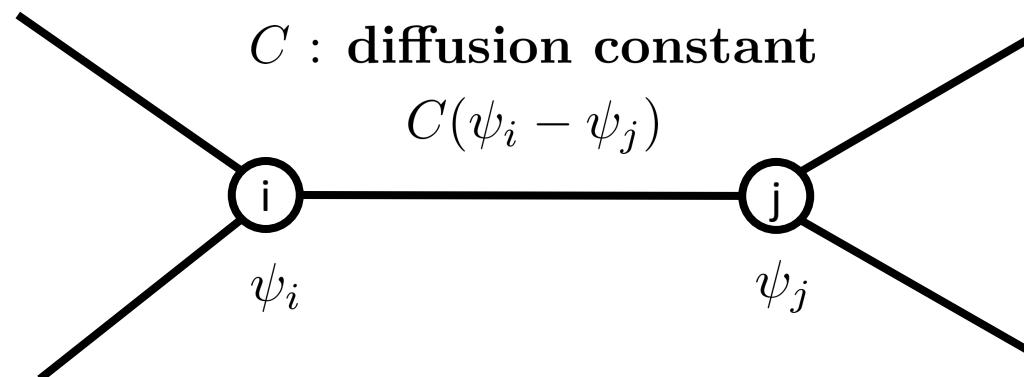
If there is no cut set of size less than n between a given pair of vertices, then there are at least n independent paths between the same vertices.

5.4 Graph Laplacian

Definition 5.4.1 (Diffusion)

Diffusion is, among other things, the process by which gas moves from regions of high density to regions of low, driven by the relative pressure (or partial pressure) of the different regions.

One can also consider diffusion processes on networks, and such processes are sometimes used as a simple model of spread across a network, such as the spread of an idea or the spread of a disease.



That is, in a small interval of time the amount of fluid flowing from j to i is $C(\psi_j - \psi_i)dt$.

$$\Rightarrow \frac{d\psi_i}{dt} = C \sum_j A_{ij} (\psi_j - \psi_i)$$

5.4 Graph Laplacian

$$\frac{d\psi_i}{dt} = C \sum_j A_{ij} \psi_j - C \psi_i \sum_j A_{ij} = C \sum_j A_{ij} \psi_j - C \psi_i k_i = C \sum_j (A_{ij} - \delta_{ij} k_i) \psi_j$$

where k_i is the degree of vertex i as usual and we have made use of the result $k_i = \sum_j A_{ij}$, and δ_{ij} is Kronecker delta, which is 1 if $i = j$ and 0 otherwise.

$$\therefore \frac{d\psi}{dt} = C(A - D)\psi$$

where $D = \begin{bmatrix} k_1 & & & \mathbf{0} \\ & k_2 & & \\ & & \ddots & \\ \mathbf{0} & & & k_n \end{bmatrix}$, $\psi = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_n \end{bmatrix}$

Let $L := D - A$: Laplacian matrix

$$\Rightarrow \frac{d\psi}{dt} + CL\psi = 0 \quad - (*)$$

$\left(\frac{d\psi}{dt} + C\Delta\psi = 0, \Delta \text{ is called Laplacian, and defined by } \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \Delta \right)$

5.4 Graph Laplacian

$$L_{ij} = \begin{cases} k_i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and there is an edge,} \\ 0 & \text{otherwise} \end{cases}$$
$$(L_{ij} = \delta_{ij}k_i - A_{ij})$$

Note

We can solve the diffusion equation (*) by writing the vector ψ as a linear combination of the eigenvectors v_i of the Laplacian thus:

$$\psi(t) = \sum_i a_i(t)v_i$$

where $a_i(t)$ varying over time and v_i is an eigenvector of A .

$$\therefore \frac{d\psi}{dt} + CL\psi = 0, \Rightarrow \sum_i \frac{da_i(t)}{dt}v_i + CL \sum_i a_i(t)v_i \frac{da_i(t)}{dt}v_i + C \sum_i a_i(t)Lv_i = 0$$

5.4 Graph Laplacian

Let $Lv_i = \lambda_i v_i$, where λ_i is the eigenvalue corresponding to the eigenvector v_i ,

$$\Rightarrow \sum_i \frac{da_i(t)}{dt} v_i + C \sum_i a_i(t) \lambda_i v_i = 0$$

$$\Rightarrow \sum_i \left(\frac{da_i(t)}{dt} + C \lambda_i a_i(t) \right) v_i = 0$$

↑
basis

$$\Rightarrow \frac{da_i(t)}{dt} + C \lambda_i a_i(t) = 0, \forall i.$$

$$\therefore a_i(t) = e^{-C\lambda_i t} a_i(0), a_i(0) : \text{initial condition.}$$

$$\text{Therefore } \psi(t) = \sum_i a_i(0) e^{-C\lambda_i t} v_i$$

5.4 Graph Laplacian

Note

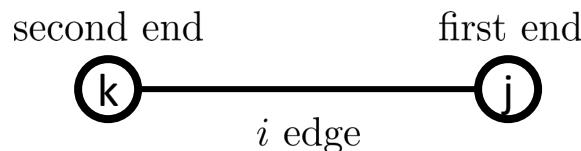
Recall. at undirected network $L = D - A \Rightarrow L^\top = L$: symmetric

\Rightarrow All eigenvalues are real

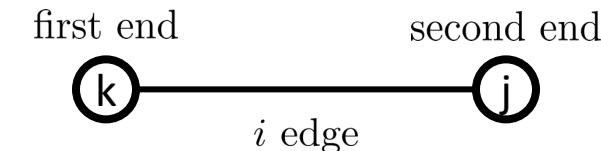
Claim : All eigenvalues are non-negative

Introducing $m \times n$ incidence matrix B , consider an undirected network with m vertices and m edges,

$$B_{ij} = \begin{cases} +1 & \text{if end 1 of edge } i \text{ is attached to vertex } j, \\ -1 & \text{if end 2 of edge } i \text{ is attached to vertex } j, \\ 0 & \text{otherwise} \end{cases}$$



$$B_{ij} = 1$$



$$B_{ij} = -1$$

5.4 Graph Laplacian

Note

$$\sum_k B_{ki} B_{kj} = \sum_k B_{ik}^\top B_{kj} = \begin{cases} -1 & \text{if } i \neq j \text{ and there is an edge from } i \text{ to } j \\ k_i & \text{if } i = j, (\sum_k B_{ki}^2 = 1) \\ 0 & \text{otherwise} \end{cases}$$
$$= L_{ij} \quad (L = D - A)$$
$$\therefore L = B^\top B$$

Note

Suppose $Lv_i = \lambda_i v_i$ with $\|v_i\| = 1$

$$\begin{aligned} \|Bv_i\|^2 &= (Bv_i) \cdot (Bv_i) = (Bv_i)^\top (Bv_i) = v_i^\top B^\top B v_i \\ &= v_i^\top \perp v_i = v_i^\top \lambda_i v_i = \lambda_i v_i^\top v_i \\ \therefore \lambda_i &= \|Bv_i\|^2 \geq 0, \quad (\because \|v_i\|^2 = 1) \\ \Rightarrow \psi(t) &\longrightarrow 0 \text{ as } t \longrightarrow \infty \end{aligned}$$

5.4 Graph Laplacian

Note

While the eigenvalues of the Laplacian cannot be negative, they can be zero, and in fact the Laplacian always has at least one zero eigenvalue.

consider $\mathbf{1} = (1, 1, \dots, 1)^\top$

$$\sum_j L_{ij} \times 1 = \sum_j (\delta_{ij} k_i - A_{ij}) = k_i - \sum_j A_{ij} = k_i - k_i = 0$$

$$\left(\because \sum_j A_{ij} = k_i \right)$$

$$L_{i,:} \cdot \mathbf{1} = [L_{i1} \quad L_{i2} \quad \cdots \quad L_{in}] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Rightarrow L \cdot \mathbf{1} = 0 = 0 \cdot \mathbf{1}$$

↑ ↑
eignevalue non-zero eigenvector

5.5 Experiments

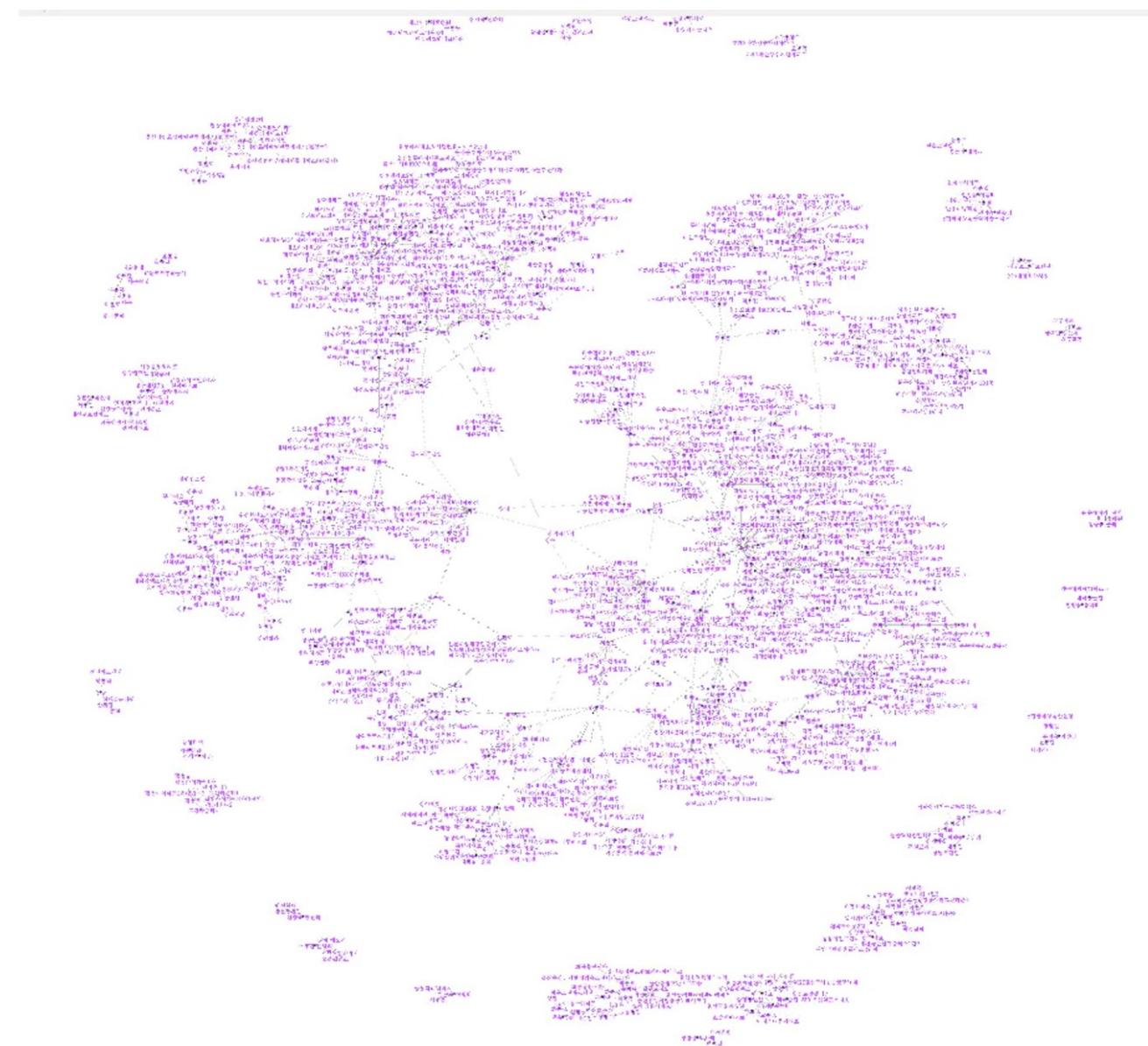
- [데이터셋] 울산광역시 2022년 아파트 매매 및 전월세 실거래 자료

울산광역시 2022.01~2022.07 아파트 매매 및 전월세 실거래 자료 | 코...
국토교통부_아파트매매 실거래 상세 자료 * 부동산 거래신고에 관한 법률에 따라 신고
된 주택의 실거래 자료를 제공 * <https://www.data.go.kr/dataset/414Ns9kb>



5.5 Experiments

- [노트북] 울산 지역 부동산 데이터를 이용한 도로명-아파트 네트워크 구축 및 중심성 분석



5.5 Experiments

- [노트북] 울산 지역 부동산 데이터를 이용한 **군, 구 아파트** 네트워크 구축 및 중심성 분석

울산 부동산 데이터를 이용한 네트워크 분석 | 코어닷투데이

울산 부동산 데이터를 군, 구로 나누어서 분석해보았습니다.

 <https://v5.core.today/notebook/s4363jV4z#Untitled0.ipynb>

	도로명	법정동	거래금액	법정동읍면동코드	도로명코드	process
0	옥교9길	학성동	29002	10100	4307360	(4307360, 10100)
1	학성4길	학성동	19100	10100	4307437	(4307437, 10100)
2	옥교15길	학성동	35900	10100	4307351	(4307351, 10100)
3	구역전1길	학성동	31000	10100	4307060	(4307060, 10100)
4	학성로	학성동	36000	10100	3169023	(3169023, 10100)
...						
6079	삼동로	웅촌면 대복리	8450	34029	3000110	(3000110, 34029)
6080	삼동로	웅촌면 대복리	7100	34029	3000110	(3000110, 34029)
6081	삼동로	웅촌면 대복리	7200	34029	3000110	(3000110, 34029)
6082	삼동로	웅촌면 대복리	6250	34029	3000110	(3000110, 34029)
6083	남천로	상북면 천전리	7900	38028	3173018	(3173018, 38028)

5.5 Experiments

교통데이터거래소 - 울산정보산업진흥원 '마이비 카드 이력 정보'

울산광역시에서 운행되는 버스에서 사용되는 마이비 카드 이력 정보 (수집일자, 노선명, 환승시간, 정류장명, 승객수 등)

최다 환승 정류장은 **공업탑**으로 일 평균 환승 인원이 **1,326명**으로 다른 정류장에 비해 압도적으로 환승객이 많다.

아래의 지도는 환승 기준 상위 50개 정류장을 히트맵으로 시각화한 결과입니다.

