All definitions and results

MATH 2740 – Mathematics of Data Science – Lecture 00

Julien Arino
julien.arino@umanitoba.ca

Department of Mathematics @ University of Manitoba

Fall 202X

The University of Manitoba campuses are located on original lands of Anishinaabeg, Ininew, Anisininew, Dakota and Dene peoples, and on the National Homeland of the Red River Métis. We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of Reconciliation and collaboration.

Definitions are colour coded

Memorising the definitions is part of the course. To help, definitions are colour coded

Definition 1 (Definitions)

These definitions are important, you need to know them

Definition 2 (Less important definitions)

These definitions are a little less important, you will not be asked to state them (although it is a good idea to know them anyway)

Results are colour coded

Memorising some of the results is part of the course. To help, results are colour coded

Theorem 3 (Theorems)

Theorems in blue boxes are worth knowing but you will not be asked to reproduce them

Theorem 4 (Important theorems)

Theorems in red boxes are important, you should know them and be able to reproduce them

You must know how to do some proofs

There are a few proofs (not many!) that I want you to know how to do

Such proofs appear on slides like the present one, with a red background

Outline

- Preliminary stuff
- Least squares
- The SVD
- Markov chains
- Graph theory
- Matrices associated to a graph/digraph
- Trees
- Why characterise graphs?

Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

Intersection and union of sets

Let X and Y be two sets

Definition 5 (Intersection)

The intersection of X and Y, $X \cap Y$, is the set of elements that belong to X **and** to Y,

$$X \cap Y = \{x : x \in X \textbf{ and } x \in Y\}$$

Definition 6 (Union)

The union of X and Y, $X \cup Y$, is the set of elements that belong to X **or** to Y,

$$X \cup Y = \{x : x \in X \textbf{ or } x \in Y\}$$

In mathematics, or=and/or in common parlance. We also have an **exclusive or** (xor)

Complex numbers

Definition 7 (Complex numbers)

A **complex number** is an ordered pair (a, b), where $a, b \in \mathbb{R}$. Usually written $a + ib$ or $a + bi$, where $i^2 = -1$ (i.e., $i = \sqrt{-1}$)
The set of all complex numbers is denoted $\mathbb{C}$,

$$\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}$$

Definition 8 (Addition and multiplication on $\mathbb{C}$)

Letting $a+ib$ and $c+id \in \mathbb{C}$, addition on $\mathbb{C}$ is defined by

$$(a+ib)+(c+id)=(a+c)+i(b+d)$$

and multiplication on $\mathbb{C}$ is defined by

$$(a+ib)(c+id)=(ac-bd)+i(ad+bc)$$

Latter is easy to obtain using regular multiplication and $i^2=-1$

Properties

$\forall \alpha, \beta, \gamma \in \mathbb{C}$,

$\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$ **[commutativity]**

$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ and $(\alpha\beta)\gamma = \alpha(\beta\gamma)$ **[associativity]**

$\gamma + 0 = \gamma$ and $\gamma 1 = \gamma$ **[identities]**

$\forall \alpha \in \mathbb{C}$, $\exists \beta \in \mathbb{C}$ unique s.t. $\alpha + \beta = 0$ **[additive inverse]**

$\forall \alpha \neq 0 \in \mathbb{C}$, $\exists \beta \in \mathbb{C}$ unique s.t. $\alpha\beta = 1$ **[multiplicative inverse]**

$\gamma(\alpha + \beta) = \gamma\alpha + \gamma\beta$ **[distributivity]**

Definition 9 (Real and imaginary parts)

Let $z = a + ib$. Then $\mathrm{Re}\, z = a$ is **real part** and $\mathrm{Im}\, z = b$ is **imaginary part** of z

If ambiguous, write $\mathrm{Re}\,(z)$ and $\mathrm{Im}\,(z)$

Definition 10 (Conjugate and Modulus)

Let $z = a + ib \in \mathbb{C}$. Then

- **Complex conjugate** of z is
$$\bar{z} = a - ib$$
- **Modulus** (or **absolute value**) of z is
$$|z| = \sqrt{a^2 + b^2} \geq 0$$

Properties of complex numbers

Let $w, z \in \mathbb{C}$, then

- $z + \bar{z} = 2\mathrm{Re}\ z$
- $z - \bar{z} = 2i\mathrm{Im}\ z$
- $z\bar{z} = |z|^2$
- $\overline{w+z} = \bar{w} + \bar{z}$ and $\overline{wz} = \bar{w}\bar{z}$
- $\bar{\bar{z}} = z$
- $|\mathrm{Re}\ z| \leq |z|$ and $|\mathrm{Im}\ z| \leq |z|$
- $|\bar{z}| = |z|$
- $|wz| = |w|\ |z|$
- $|w + z| \leq |w| + |z|$ **[triangle inequality]**

Vectors

A **vector** $\boldsymbol{v}$ is an ordered n-tuple of real or complex numbers

Denote $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ (real or complex numbers). For $v_1, \ldots, v_n \in \mathbb{F}$,

$$\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{F}^n$$

is a vector. $v_1, \ldots, v_n$ are the **components** of $\boldsymbol{v}$

If unambiguous, we write v. Otherwise, $\boldsymbol{v}$ or $\vec{v}$

Vector space

Definition 11 (Vector space)

A **vector space** over $\mathbb{F}$ is a set V together with two binary operations, **vector addition**, denoted $+$, and **scalar multiplication**, that satisfy the relations:

1. $\forall \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$
2. $\forall \boldsymbol{v}, \boldsymbol{w} \in V$, $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$
3. $\exists \boldsymbol{0} \in V$, the zero vector, such that $\boldsymbol{v} + \boldsymbol{0} = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$
4. $\forall \boldsymbol{v} \in V$, there exists an element $\boldsymbol{w} \in V$, the additive inverse of $\boldsymbol{v}$, such that $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{0}$
5. $\forall \alpha \in \mathbb{R}$ and $\forall \boldsymbol{v}, \boldsymbol{w} \in V$, $\alpha(\boldsymbol{v} + \boldsymbol{w}) = \alpha\boldsymbol{v} + \alpha\boldsymbol{w}$
6. $\forall \alpha, \beta \in \mathbb{R}$ and $\forall \boldsymbol{v} \in V$, $(\alpha + \beta)\boldsymbol{v} = \alpha\boldsymbol{v} + \beta\boldsymbol{v}$
7. $\forall \alpha, \beta \in \mathbb{R}$ and $\forall \boldsymbol{v} \in V$, $\alpha(\beta\boldsymbol{v}) = (\alpha\beta)\boldsymbol{v}$
8. $\forall \boldsymbol{v} \in V$, $1\boldsymbol{v} = \boldsymbol{v}$

Norms

Definition 12 (Norm)

Let V be a vector space over $\mathbb{F}$, and $\boldsymbol{v} \in V$ be a vector. The **norm** of $\boldsymbol{v}$, denoted $\|\boldsymbol{v}\|$, is a function from V to $\mathbb{R}_+$ that has the following properties:

1. For all $\boldsymbol{v} \in V$, $\|\boldsymbol{v}\| \geq 0$ with $\|\boldsymbol{v}\| = 0$ iff $\boldsymbol{v} = \mathbf{0}$
2. For all $\alpha \in \mathbb{F}$ and all $\boldsymbol{v} \in V$, $\|\alpha\boldsymbol{v}\| = |\alpha| \, \|\boldsymbol{v}\|$
3. For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$

Let V be a vector space (for example, $\mathbb{R}^2$ or $\mathbb{R}^3$)

The **zero element** (or **zero vector**) is the vector $\mathbf{0} = (0, \ldots, 0)$

The **additive inverse** of $\boldsymbol{v} = (v_1, \ldots, v_n)$ is $-\boldsymbol{v} = (-v_1, \ldots, -v_n)$

For $\boldsymbol{v} = (v_1, \ldots, v_n) \in V$, the length (or Euclidean norm) of $\boldsymbol{v}$ is the **scalar**

$$\|\boldsymbol{v}\| = \sqrt{v_1^2 + \cdots + v_n^2}$$

To **normalize** the vector $\boldsymbol{v}$ consists in considering $\tilde{\boldsymbol{v}} = \boldsymbol{v}/\|\boldsymbol{v}\|$, i.e., the vector in the same direction as $\boldsymbol{v}$ that has unit length

Dot product

Definition 13 (Dot product)

Let $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, $\boldsymbol{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$. The **dot product** of $\boldsymbol{a}$ and $\boldsymbol{b}$ is the **scalar**

$$\boldsymbol{a} \bullet \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + \cdots + a_n b_n$$

The dot product is a special case of **inner product**

Properties of the dot product

Theorem 14

For $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$ *and* $\alpha \in \mathbb{R}$,

- $\boldsymbol{a} \bullet \boldsymbol{a} = \|\boldsymbol{a}\|^2$ *(so* $\boldsymbol{a} \bullet \boldsymbol{a} \geq 0$*, with* $\boldsymbol{a} \bullet \boldsymbol{a} = 0$ *iff* $\boldsymbol{a} = \boldsymbol{0}$*)*
- $\boldsymbol{a} \bullet \boldsymbol{b} = \boldsymbol{b} \bullet \boldsymbol{a}$ *(*$\bullet$ *is commutative)*
- $\boldsymbol{a} \bullet (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \bullet \boldsymbol{b} + \boldsymbol{a} \bullet \boldsymbol{c}$ *(*$\bullet$ *distributive over* $+$*)*
- $(\alpha \boldsymbol{a}) \bullet \boldsymbol{b} = \alpha(\boldsymbol{a} \bullet \boldsymbol{b}) = \boldsymbol{a} \bullet (\alpha \boldsymbol{b})$
- $\boldsymbol{0} \bullet \boldsymbol{a} = 0$

Some results stemming from the dot product

Theorem 15

If θ is the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, then

$$\boldsymbol{a} \bullet \boldsymbol{b} = \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \, \cos\theta$$

Corollary 16 (Cauchy-Schwarz inequality)

For any two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we have

$$|\boldsymbol{a} \bullet \boldsymbol{b}| \leq \|\boldsymbol{a}\| \, \|\boldsymbol{b}\|$$

with equality if and only if $\boldsymbol{a}$ is a scalar multiple of $\boldsymbol{b}$, or one of them is $\mathbf{0}$.

Theorem 17

$\boldsymbol{a}$ and $\boldsymbol{b}$ are orthogonal if and only if $\boldsymbol{a} \bullet \boldsymbol{b} = 0$.

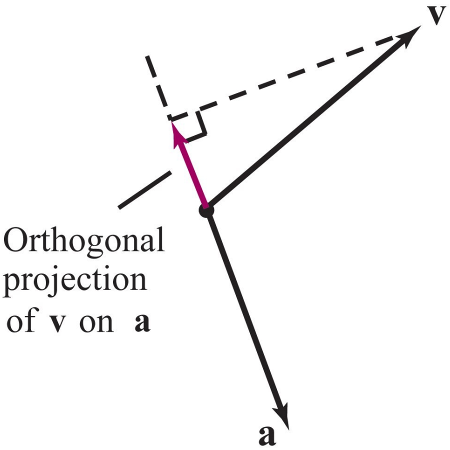
Scalar and vector projections

Scalar projection of $\boldsymbol{v}$ onto $\boldsymbol{a}$ (or component of $\boldsymbol{v}$ along $\boldsymbol{a}$):

$$\text{comp}_{\boldsymbol{a}}\boldsymbol{v} = \frac{\boldsymbol{a} \bullet \boldsymbol{v}}{\|\boldsymbol{a}\|}$$

Vector (or orthogonal) projection of $\boldsymbol{v}$ onto $\boldsymbol{a}$:

$$\text{proj}_{\boldsymbol{a}}\boldsymbol{v} = \left(\frac{\boldsymbol{a} \bullet \boldsymbol{v}}{\|\boldsymbol{a}\|}\right)\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} = \frac{\boldsymbol{a} \bullet \boldsymbol{v}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}$$



Linear systems

Definition 18 (Linear system)

A **linear system** of m equations in n unknowns takes the form

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_n \end{array} \tag{1}$$

The a_{ij}, x_j and b_j could be in $\mathbb{R}$ or $\mathbb{C}$, although here we typically assume they are in $\mathbb{R}$

The aim is to find $x_1, x_2, \ldots, x_n$ that satisfy all equations simultaneously

Theorem 19 (Nature of solutions to a linear system)

A linear system can have

- *no solution*
- *a unique solution*
- *infinitely many solutions*

Matrices and linear systems

Writing

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

where A is an $m \times n$ **matrix**, $\boldsymbol{x}$ and $\boldsymbol{b}$ are n (column) **vectors** (or $n \times 1$ matrices), then the linear system in the previous slide takes the form

$$A\boldsymbol{x} = \boldsymbol{b}$$

Consider the system

$$A\boldsymbol{x} = \boldsymbol{b}$$

If $\boldsymbol{b} = \mathbf{0}$, the system is **homogeneous** and always has the solution $\boldsymbol{x} = 0$ and so the “no solution” option in Theorem 19 goes away

Definition 20 (Matrix)

An m-by-n or $m \times n$ matrix is a rectangular array of elements of $\mathbb{R}$ or $\mathbb{C}$ with m rows and n columns,

$$A = [a_{ij}] = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

We always list indices as “row,column”

We denote $\mathcal{M}_{mn}(\mathbb{F})$ or $\mathbb{F}^{mn}$ the set of $m \times n$ matrices with entries in $\mathbb{F} = \{\mathbb{R}, \mathbb{C}\}$. Often, we omit $\mathbb{F}$ in $\mathcal{M}_{mn}$ if the nature of $\mathbb{F}$ is not important

When $m = n$, we usually write $\mathcal{M}_n$

Basic matrix arithmetic

Let $A \in \mathcal{M}_{mn}, B \in \mathcal{M}_{mn}$ be matrices (of the same size) and $c \in \mathbb{F} = \{\mathbb{R}, \mathbb{C}\}$ be a scalar

- **Scalar multiplication**
$$cA = [ca_{ij}]$$
- **Addition**
$$A + B = [a_{ij} + b_{ij}]$$
- **Subtraction** (addition of $-B = (-1)B$ to A)
$$A - B = A + (-1)B = [a_{ij} + (-1)b_{ij}] = [a_{ij} - b_{ij}]$$
- **Transposition** of A gives a matrix $A^T = \mathcal{M}_{nm}$ with
$$A^T = [a_{ji}], \quad j = 1, \ldots, n, \quad i = 1, \ldots, m$$

Matrix multiplication

The (matrix) **product** of A and B, AB, requires the "inner dimensions" to match, i.e., the number of columns in A must equal the number of rows in B

Suppose that is the case, i.e., let $A \in \mathcal{M}_{mn}$, $B \in \mathcal{M}_{np}$. Then the i, j entry in $C := AB$ takes the form

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Recall that the matrix product is not commutative, i.e., in general, $AB \neq BA$ (when both those products are defined, i.e., when $A, B \in \mathcal{M}_n$)

Special matrices

Definition 21 (Zero and identity matrices)

The **zero** matrix is the matrix 0_{mn} whose entries are all zero. The **identity** matrix is a square $n \times n$ matrix $\mathbb{I}_n$ with all entries on the main diagonal equal to one and all off diagonal entries equal to zero

Definition 22 (Symmetric matrix)

A square matrix $A \in \mathcal{M}_n$ is **symmetric** if $\forall i, j = 1, \ldots, n$, $a_{ij} = a_{ji}$. In other words, $A \in \mathcal{M}_n$ is symmetric if $A = A^T$

Making symmetric matrices

Theorem 23

1. *If $A \in \mathcal{M}_n$, then $A + A^T$ is symmetric*
2. *If $A \in \mathcal{M}_{mn}$, then $AA^T \in \mathcal{M}_m$ and $A^TA \in \mathcal{M}_n$ are symmetric*

Proof of Theorem 23

X symmetric $\iff X = X^T$, so use $X =$ the matrix whose symmetric property you want to check

1. True if $A + A^T = (A + A^T)^T$. We have

$$(A + A^T)^T = A^T + (A^T)^T = A^T + A = A + A^T$$

2. AA^T symmetric if $AA^T = (AA^T)^T$. We have

$$(AA^T)^T = (A^T)^T A^T = AA^T$$

$A^T A$ works similarly

Determinants

Definition 24 (Determinant)

Let $A \in \mathcal{M}_n$ with $n \geq 2$. The **determinant** of A is the *scalar*

$$\det(A) = |A| = \sum_{j=1}^{n} a_{ij} C_{ij}$$

where $C_{ij} = (-1)^{i+j} \det(A_{ij})$ is the (i,j)-**cofactor** of A and A_{ij} is the submatrix of A from which the ith row and jth column have been removed

This is a cofactor expansion along the ith row
This is a recursive formula: it gives result in terms of n $\mathcal{M}_{n-1}$ matrices, to which it must in turn be applied, all the way down to

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Two special matrices and their determinants

Definition 25

$A \in \mathcal{M}_n$ is **upper triangular** if $a_{ij} = 0$ when $i > j$, **lower triangular** if $a_{ij} = 0$ when $j > i$, **triangular** if it is *either* upper or lower triangular and **diagonal** if it is *both* upper and lower triangular

When A diagonal, we often write $A = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn})$

Theorem 26

Let $A \in \mathcal{M}_n$ be triangular or diagonal. Then

$$\det(A) = \prod_{i=1}^{n} a_{ii} = a_{11}a_{22}\cdots a_{nn}$$

Inversion/Singularity

Definition 27 (Matrix inverse)

$A \in \mathcal{M}_n$ is **invertible** (or **nonsingular**) if $\exists A^{-1} \in \mathcal{M}_n$ s.t.

$$AA^{-1} = A^{-1}A = \mathbb{I}$$

A^{-1} is the **inverse** of A. If A^{-1} does not exist, A is **singular**

Theorem 28

Let $A \in \mathcal{M}_n$, $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{F}^n$. Then

- *A invertible $\iff \det(A) \neq 0$*
- *If A invertible, A^{-1} is unique*
- *If A invertible, then $A\boldsymbol{x} = \boldsymbol{b}$ has the unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$*

Eigenvalues / Eigenvectors / Eigenpairs

Definition 29

Let $A \in \mathcal{M}_n$. A vector $\boldsymbol{x} \in \mathbb{F}^n$ such that $\boldsymbol{x} \neq \boldsymbol{0}$ is an **eigenvector** of A if $\exists \lambda \in \mathbb{F}$ called an **eigenvalue**, s.t.

$$A\boldsymbol{x} = \lambda \boldsymbol{x}$$

A couple $(\lambda, \boldsymbol{x})$ with $\boldsymbol{x} \neq \boldsymbol{0}$ s.t. $A\boldsymbol{x} = \lambda\boldsymbol{x}$ is an **eigenpair**

If $(\lambda, \boldsymbol{x})$ eigenpair, then for $c \neq 0$, $(\lambda, c\boldsymbol{x})$ also eigenpair since $A(c\boldsymbol{x}) = cA\boldsymbol{x} = c\lambda\boldsymbol{x}$ and dividing both sides by c..

Similarity

Definition 30 (Similarity)

$A, B \in \mathcal{M}_n$ are **similar** ($A \sim B$) if $\exists P \in \mathcal{M}_n$ invertible s.t.

$$P^{-1}AP = B$$

Theorem 31 ($\sim$ is an equivalence relation)

$A, B, C \in \mathcal{M}_n$, then

- $A \sim A$ *($\sim$ **reflexive**)*
- $A \sim B \implies B \sim A$ *($\sim$ **symmetric**)*
- $A \sim B$ *and* $B \sim C \implies A \sim C$ *($\sim$ **transitive**)*

Similarity (cont.)

Theorem 32

$A, B \in \mathcal{M}_n$ with $A \sim B$. Then

- $\det A = \det B$
- *A invertible $\iff$ B invertible*
- *A and B have the same eigenvalues*

Diagonalisation

Definition 33 (Diagonalisability)

$A \in \mathcal{M}_n$ is **diagonalisable** if $\exists D \in \mathcal{M}_n$ diagonal s.t. $A \sim D$

In other words, $A \in \mathcal{M}_n$ is diagonalisable if there exists a diagonal matrix $D \in \mathcal{M}_n$ and a nonsingular matrix $P \in \mathcal{M}_n$ s.t. $P^{-1}AP = D$

Could of course write $PAP^{-1} = D$ since P invertible, but $P^{-1}AP$ makes more sense for computations

Theorem 34

$A \in \mathcal{M}_n$ diagonalisable $\iff$ A has n linearly independent eigenvectors

Corollary 35 (Sufficient condition for diagonalisability)

$A \in \mathcal{M}_n$ has all its eigenvalues distinct $\implies$ A diagonalisable

For $P^{-1}AP = D$: in P, put the linearly independent eigenvectors as columns and in D, the corresponding eigenvalues

Linear combination and span

Definition 36 (Linear combination)

Let V be a vector space. A **linear combination** of a set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ of vectors in V is a *vector*

$$c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k$$

where $c_1, \ldots, c_k \in \mathbb{F}$

Definition 37 (Span)

The set of all linear combinations of a set of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ is the **span** of $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$,

$$\text{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = \{c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k : c_1, \ldots, c_k \in \mathbb{F}\}$$

Finite/infinite-dimensional vector spaces

Theorem 38

The span of a set of vectors in V is the smallest subspace of V containing all the vectors in the set

Definition 39 (Set of vectors spanning a space)

If $\text{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = V$, we say $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ **spans** V

Definition 40 (Dimension of a vector space)

A vector space V is **finite-dimensional** if some set of vectors in it spans V. A vector space V is **infinite-dimensional** if it is not finite-dimensional

Linear (in)dependence

Definition 41 (Linear independence/Linear dependence)

A set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ of vectors in a vector space V is **linearly independent** if

$$(c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k = 0) \Leftrightarrow (c_1 = \cdots = c_k = 0)\,,$$

where $c_1, \dots, c_k \in \mathbb{F}$. A set of vectors is **linearly dependent** if it is not linearly independent.

If linearly dependent, assume w.l.o.g. that $c_1 \neq 0$, then

$$\boldsymbol{v}_1 = -\frac{c_2}{c_1}\boldsymbol{v}_2 - \cdots - \frac{c_k}{c_1}\boldsymbol{v}_k$$

i.e., $\boldsymbol{v}_1$ is a linear combination of the other vectors in the set

Basis

Definition 42 (Basis)

Let V be a vector space. A **basis** of V is a set of vectors in V that is both linearly independent and spanning

Theorem 43 (Criterion for a basis)

A set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ of vectors in a vector space V is a basis of $V \iff \forall \boldsymbol{v} \in V$, $\boldsymbol{v}$ can be written uniquely in the form

$$\boldsymbol{v} = c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k,$$

where $c_1, \dots, c_k \in \mathbb{F}$

More on bases

Theorem 44

Any two bases of a finite-dimensional vector space have the same number of vectors

Definition 45 (Dimension)

The **dimension** $\dim V$ of a finite-dimensional vector space V is the number of vectors in any basis of the vector space

Constructing bases

Theorem 46

Let V be a finite-dimensional vector space. Then every linearly independent set of vectors in V with $\dim V$ elements is a basis of V

Theorem 47

Let V be a finite-dimensional vector space. Then every spanning set of vectors in V with $\dim V$ elements is a basis of V

Linear algebra in a nutshell

Theorem 48

Let $A \in \mathcal{M}_n$. The following statements are equivalent (TFAE)

1. *The matrix A is invertible*
2. $\forall \boldsymbol{b} \in \mathbb{F}^n$, $A\boldsymbol{x} = \boldsymbol{b}$ *has a unique solution* ($\boldsymbol{x} = A^{-1}\boldsymbol{b}$)
3. *The only solution to* $A\boldsymbol{x} = \mathbf{0}$ *is the trivial solution* $\boldsymbol{x} = \mathbf{0}$
4. $RREF(A) = \mathbb{I}_n$
5. *The matrix A is equal to a product of elementary matrices*
6. $\forall \boldsymbol{b} \in \mathbb{F}^n$, $A\boldsymbol{x} = \boldsymbol{b}$ *has a solution*
7. *There is a matrix $B \in \mathcal{M}_n$ such that $AB = \mathbb{I}_n$*
8. *There is an invertible matrix $B \in \mathcal{M}_n$ such that $AB = \mathbb{I}_n$*
9. $\det(A) \neq 0$
10. 0 *is not an eigenvalue of A*

Partial derivatives

How do we measure the “slope” on a 3D surface?

A **partial derivative** measures the slope in a direction parallel to one of the axes

- $\dfrac{\partial f}{\partial x}$ measures height change as we move only in the x direction. Treat y as a constant
- $\dfrac{\partial f}{\partial y}$ measures height change as we move only in the y direction. Treat x as a constant

The Lagrangian function

The condition $\nabla f = \lambda \nabla g$ is clever, but solving it can be messy

Instead, we combine our function and constraint into a single, new function called the **Lagrangian**

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - c]$$

- $f(x, y)$ the function we want to optimize
- $g(x, y) = c$ the constraint we must follow
- λ the Lagrange multiplier

Finding the unconstrained optimum of $\mathcal{L}$ solves the original constrained problem!

The method – step-by-step

To find the optimum of the Lagrangian $\mathcal{L}(x, y, \lambda)$, we find where its gradient is zero

We take the partial derivative with respect to *all* its variables (x, y, and λ) and set them to zero

1. $\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0$
2. $\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0$
3. $\frac{\partial \mathcal{L}}{\partial \lambda} = -(g(x, y) - c) = 0 \implies g(x, y) = c$

The first two equations rearrange to $\nabla f = \lambda \nabla g$ and the third equation is the original constraint

The gradient

$f : \mathbb{R}^n \to \mathbb{R}$ function of several variables, $\nabla = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right)$ the gradient operator

Then

$$\nabla f = \left(\frac{\partial}{\partial x_1} f, \ldots, \frac{\partial}{\partial x_n} f\right)$$

So ∇f is a *vector-valued* function, $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$; also written as

$$\nabla f = f_{x_1}(x_1, \ldots, x_n)\boldsymbol{e}_1 + \cdots f_{x_n}(x_1, \ldots, x_n)\boldsymbol{e}_n$$

where f_{x_i} is the partial derivative of f with respect to x_i and $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is the standard basis of $\mathbb{R}^n$

Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

The linear least squares problem

Given a collection of data points $(x_1, y_1), \ldots, (x_n, y_n)$, find the coefficients a, b of the line $y = a + bx$ such that

$$\|\mathbf{e}\| = \sqrt{\varepsilon_1^2 + \cdots + \varepsilon_n^2} = \sqrt{(y_1 - \tilde{y}_1)^2 + \cdots + (y_n - \tilde{y}_n)^2}$$

is minimal, where $\tilde{y}_i = a + bx_i$, for $i = 1, \ldots, n$

The least squares problem (simplest version)

Definition 49

Given a collection of points $(x_1, y_1), \ldots, (x_n, y_n)$, find the coefficients a, b of the line $y = a + bx$ such that

$$\|\mathbf{e}\| = \sqrt{\varepsilon_1^2 + \cdots + \varepsilon_n^2} = \sqrt{(y_1 - \tilde{y}_1)^2 + \cdots + (y_n - \tilde{y}_n)^2}$$

is minimal, where $\tilde{y}_i = a + bx_i$ for $i = 1, \ldots, n$

We just saw how to solve this by brute force using a genetic algorith to minimise $\|e\|$, let us now see how to solve this problem “properly”

For a data point $i = 1, \ldots, n$

$$\varepsilon_i = y_i - \tilde{y}_i = y_i - (a + bx_i)$$

So if we write this for all data points,

$$\begin{aligned} \varepsilon_1 &= y_1 - (a + bx_1) \\ &\vdots \\ \varepsilon_n &= y_n - (a + bx_n) \end{aligned}$$

In matrix form

$$\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$$

with

$$\boldsymbol{e} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, A = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} a \\ b \end{pmatrix} \text{ and } \boldsymbol{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

The least squares problem (reformulated)

Definition 50 (Least squares solutions)

Consider a collection of points $(x_1, y_1), \ldots, (x_n, y_n)$, a matrix $A \in \mathcal{M}_{mn}$, $\boldsymbol{b} \in \mathbb{R}^m$. A **least squares solution** of $A\boldsymbol{x} = \boldsymbol{b}$ is a vector $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ s.t.

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \quad \|\boldsymbol{b} - A\tilde{\boldsymbol{x}}\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\|$$

Needed to solve the problem

Definition 51 (Best approximation)

Let V be a vector space, $W \subset V$ and $\mathbf{v} \in V$. The **best approximation** to $\mathbf{v}$ in W is $\tilde{\mathbf{v}} \in W$ s.t.

$$\forall \mathbf{w} \in W, \mathbf{w} \neq \tilde{\mathbf{v}}, \quad \|\mathbf{v} - \tilde{\mathbf{v}}\| < \|\mathbf{v} - \mathbf{w}\|$$

Theorem 52 (Best approximation theorem)

Let V be a vector space with an inner product, $W \subset V$ and $\mathbf{v} \in V$. Then $\mathrm{proj}_W(\mathbf{v})$ *is the best approximation to* $\mathbf{v}$ *in W*

Putting things together

We just stated: The least squares solution of $\boldsymbol{Ax} = \boldsymbol{b}$ is a vector $\tilde{\boldsymbol{y}} \in \operatorname{col}(A)$ s.t.

$$\forall \boldsymbol{y} \in \operatorname{col}(A), \quad \|\boldsymbol{b} - \tilde{\boldsymbol{y}}\| \leq \|\boldsymbol{b} - \boldsymbol{y}\|$$

We know (reformulating a tad):

Theorem 53 (Best approximation theorem)

Let V be a vector space with an inner product, $W \subset V$ and $\mathbf{v} \in V$. Then $\operatorname{proj}_W(\mathbf{v}) \in W$ is the best approximation to $\mathbf{v}$ in W, i.e.,

$$\forall \mathbf{w} \in W, \mathbf{w} \neq \operatorname{proj}_W(\mathbf{v}), \quad \|\mathbf{v} - \operatorname{proj}_W(\mathbf{v})\| < \|\mathbf{v} - \mathbf{w}\|$$

$\implies W = \operatorname{col}(A)$, $\boldsymbol{v} = \boldsymbol{b}$ and $\tilde{\boldsymbol{y}} = \operatorname{proj}_{\operatorname{col}(A)}(\mathbf{b})$

Least squares theorem

Theorem 54 (Least squares theorem)

$A \in \mathcal{M}_{mn}$, $\boldsymbol{b} \in \mathbb{R}^m$. Then

1. *$A\boldsymbol{x} = \boldsymbol{b}$ always has at least one least squares solution $\tilde{\boldsymbol{x}}$*
2. *$\tilde{\boldsymbol{x}}$ least squares solution to $A\boldsymbol{x} = \boldsymbol{b} \iff \tilde{\boldsymbol{x}}$ is a solution to the normal equations $A^T A\tilde{\boldsymbol{x}} = A^T\boldsymbol{b}$*
3. *A has linearly independent columns $\iff A^T A$ invertible.*
   In this case, the least squares solution is unique and

$$\tilde{\boldsymbol{x}} = \left(A^T A\right)^{-1} A^T\boldsymbol{b}$$

We have seen 1 and 2, we will not show 3 (it is not hard)

Fitting the quadratic

We have the data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and want to fit

$$y = a_0 + a_1 x + a_2 x^2$$

At (x_1, y_1),

$$\tilde{y}_1 = a_0 + a_1 x_1 + a_2 x_1^2$$

$\vdots$

At (x_n, y_n),

$$\tilde{y}_n = a_0 + a_1 x_n + a_2 x_n^2$$

In terms of the error

$$\varepsilon_1 = y_1 - \tilde{y}_1 = y_1 - (a_0 + a_1 x_1 + a_2 x_1^2)$$
$$\vdots$$
$$\varepsilon_n = y_n - \tilde{y}_n = y_n - (a_0 + a_1 x_n + a_2 x_n^2)$$

i.e.,

$$\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$$

where

$$\boldsymbol{e} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, A = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \text{ and } \boldsymbol{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Theorem 54 applies, with here $A \in \mathcal{M}_{n3}$ and $\boldsymbol{b} \in \mathbb{R}^n$

Definition 55 (Orthogonal set of vectors)

The set of vectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\} \in \mathbb{R}^n$ is an **orthogonal set** if

$$\forall i, j = 1, \dots, k, \quad i \neq j \implies \boldsymbol{v}_i \bullet \boldsymbol{v}_j = 0$$

Theorem 56

$\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\} \in \mathbb{R}^n$ with $\forall i$, $\boldsymbol{v}_i \neq \mathbf{0}$, orthogonal set $\implies \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\} \in \mathbb{R}^n$ linearly independent

Definition 57 (Orthogonal basis)

Let S be a basis of the subspace $W \subset \mathbb{R}^n$ composed of an orthogonal set of vectors. We say S is an **orthogonal basis** of W

Orthonormal version of things

Definition 58 (Orthonormal set)

The set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \in \mathbb{R}^n$ is an **orthonormal set** if it is an orthogonal set and furthermore

$$\forall i = 1, \ldots, k, \quad \|\boldsymbol{v}_i\| = 1$$

Definition 59 (Orthonormal basis)

A basis of the subspace $W \subset \mathbb{R}^n$ is an **orthonormal basis** if the vectors composing it are an orthonormal set

$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \in \mathbb{R}^n$ is orthonormal if

$$\boldsymbol{v}_i \bullet \boldsymbol{v}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Projections

Definition 60 (Orthogonal projection onto a subspace)

$W \subset \mathbb{R}^n$ a subspace and $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ an orthogonal basis of W. $\forall \boldsymbol{v} \in \mathbb{R}^n$, the **orthogonal projection** of $\boldsymbol{v}$ **onto** W is

$$\mathrm{proj}_W(\boldsymbol{v}) = \frac{\boldsymbol{u}_1 \bullet \boldsymbol{v}}{\|\boldsymbol{u}_1\|^2}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{u}_k \bullet \boldsymbol{v}}{\|\boldsymbol{u}_k\|^2}\boldsymbol{u}_k$$

Definition 61 (Component orthogonal to a subspace)

$W \subset \mathbb{R}^n$ a subspace and $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ an orthogonal basis of W. $\forall \boldsymbol{v} \in \mathbb{R}^n$, the **component** of $\boldsymbol{v}$ **orthogonal to** W is

$$\mathrm{perp}_W(\boldsymbol{v}) = \boldsymbol{v} - \mathrm{proj}_W(\boldsymbol{v})$$

Gram-Schmidt process

Theorem 62

$W \subset \mathbb{R}^n$ a subset and $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ a basis of W. Let

$$
\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{x}_1 \\
\boldsymbol{v}_2 &= \boldsymbol{x}_2 - \frac{\boldsymbol{v}_1 \bullet \boldsymbol{x}_2}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 \\
\boldsymbol{v}_3 &= \boldsymbol{x}_3 - \frac{\boldsymbol{v}_1 \bullet \boldsymbol{x}_3}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 - \frac{\boldsymbol{v}_2 \bullet \boldsymbol{x}_3}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2 \\
&\vdots \\
\boldsymbol{v}_k &= \boldsymbol{x}_k - \frac{\boldsymbol{v}_1 \bullet \boldsymbol{x}_k}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 - \cdots - \frac{\boldsymbol{v}_{k-1} \bullet \boldsymbol{x}_k}{\|\boldsymbol{v}_{k-1}\|^2}\boldsymbol{v}_{k-1}
\end{aligned}
$$

and

$$W_1 = \text{span}(\boldsymbol{x}_1), W_2 = \text{span}(\boldsymbol{x}_1, \boldsymbol{x}_2), \ldots, W_k = \text{span}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$$

Then $\forall i = 1, \ldots, k$, $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i\}$ orthogonal basis for W_i

Theorem 63

Let $Q \in \mathcal{M}_{mn}$. The columns of Q form an orthonormal set if and only if

$$Q^T Q = \mathbb{I}_n$$

Definition 64 (Orthogonal matrix)

$Q \in \mathcal{M}_n$ is an **orthogonal matrix** if its columns form an orthonormal set

So $Q \in \mathcal{M}_n$ orthogonal if $Q^T Q = \mathbb{I}$, i.e., $Q^T = Q^{-1}$

Theorem 65 (NSC for orthogonality)

$Q \in \mathcal{M}_n$ *orthogonal* $\iff Q^{-1} = Q^T$

Theorem 66 (Orthogonal matrices “encode" isometries)

Let $Q \in \mathcal{M}_n$. TFAE

1. *Q orthogonal*
2. $\forall \boldsymbol{x} \in \mathbb{R}^n$, $\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$
3. $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $Q\boldsymbol{x} \bullet Q\boldsymbol{y} = \boldsymbol{x} \bullet \boldsymbol{y}$

Theorem 67

Let $Q \in \mathcal{M}_n$ be orthogonal. Then

1. *The rows of Q form an orthonormal set*
2. *Q^{-1} orthogonal*
3. $\det Q = \pm 1$
4. $\forall \lambda \in \sigma(Q)$, $|\lambda| = 1$
5. *If $Q_2 \in \mathcal{M}_n$ also orthogonal, then QQ_2 orthogonal*

The QR factorisation

Theorem 68

Let $A \in \mathcal{M}_{mn}$ with LI columns. Then A can be factored as

$$A = QR$$

where $Q \in \mathcal{M}_{mn}$ has orthonormal columns and $R \in \mathcal{M}_n$ is nonsingular upper triangular

Back to least squares

So what was the point of all that..?

Theorem 69 (Least squares with QR factorisation)

$A \in \mathcal{M}_{mn}$ with LI columns, $\boldsymbol{b} \in \mathbb{R}^m$. If $A = QR$ is a QR factorisation of A, then the unique least squares solution $\tilde{\boldsymbol{x}}$ of $A\boldsymbol{x} = \boldsymbol{b}$ is

$$\tilde{\boldsymbol{x}} = R^{-1}Q^T\boldsymbol{b}$$

Singular values

Definition 70 (Singular value)

Let $A \in \mathcal{M}_{mn}(\mathbb{R})$. The **singular values** of A are the real numbers

$$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n \geq 0$$

that are the square roots of the eigenvalues of $A^T A$

Singular values are real and nonnegative?

Recall that $\forall A \in \mathcal{M}_{mn}$, $A^T A$ is symmetric

Claim 1. Real symmetric matrices have real eigenvalues

Claim 2. For $A \in \mathcal{M}_{mn}(\mathbb{R})$, the eigenvalues of $A^T A$ are real and nonnegative

Proof. We know that for $A \in \mathcal{M}_{mn}$, $A^T A$ symmetric and from previous claim, if $A \in \mathcal{M}_{mn}(\mathbb{R})$, then $A^T A$ is symmetric and real and with real eigenvalues

Let $(\lambda, \boldsymbol{v})$ be an eigenpair of $A^T A$, with $\boldsymbol{v}$ chosen so that $\|\boldsymbol{v}\| = 1$

Norms are functions $V \to \mathbb{R}_+$, so $\|A\boldsymbol{v}\|$ and $\|A\boldsymbol{v}\|^2$ are ≥ 0 and thus

$$
\begin{aligned}
0 \leq \|A\boldsymbol{v}\|^2 &= (A\boldsymbol{v}) \bullet (A\boldsymbol{v}) = (A\boldsymbol{v})^T (A\boldsymbol{v}) \\
&= \boldsymbol{v}^T A^T A\boldsymbol{v} = \boldsymbol{v}^T (A^T A\boldsymbol{v}) = \boldsymbol{v}^T (\lambda\boldsymbol{v}) \\
&= \lambda(\boldsymbol{v}^T \boldsymbol{v}) = \lambda(\boldsymbol{v} \bullet \boldsymbol{v}) = \lambda\|\boldsymbol{v}\|^2 \\
&= \lambda
\end{aligned}
$$

Claim 3. For $A \in \mathcal{M}_{mn}(\mathbb{R})$, the nonzero eigenvalues of $A^T A$ and AA^T are the same

Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

The singular value decomposition (SVD)

Theorem 71 (SVD)

$A \in \mathcal{M}_{mn}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$

Then there exists $U \in \mathcal{M}_m$ orthogonal, $V \in \mathcal{M}_n$ orthogonal and a block matrix $\Sigma \in \mathcal{M}_{mn}$ taking the form

$$\Sigma = \begin{pmatrix} D & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{pmatrix}$$

where

$$D = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathcal{M}_r$$

such that

$$A = U\Sigma V^T$$

Definition 72

We call a factorisation as in Theorem 71 the **singular value decomposition** of A. The columns of U and V are, respectively, the **left** and **right singular vectors** of A

U and V^T are *rotation* or *reflection* matrices, Σ is a *scaling* matrix

$U \in \mathcal{M}_m$ orthogonal matrix with columns the eigenvectors of AA^T

$V \in \mathcal{M}_n$ orthogonal matrix with columns the eigenvectors of A^TA

Outer product form of the SVD

Theorem 73 (Outer product form of the SVD)

$A \in \mathcal{M}_{mn}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$, respectively, left and right singular vectors of A corresponding to these singular values

Then

$$A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^T \tag{2}$$

Computing the SVD (case of $\neq$ eigenvalues)

To compute the SVD, we use the following result

Theorem 74

Let $A \in \mathcal{M}_n$ symmetric, $(\lambda_1, \boldsymbol{u}_1)$ and $(\lambda_2, \boldsymbol{u}_2)$ be eigenpairs, $\lambda_1 \neq \lambda_2$. Then $\boldsymbol{u}_1 \bullet \boldsymbol{u}_2 = 0$

Proof of Theorem 74

$A \in \mathcal{M}_n$ symmetric, $(\lambda_1, \boldsymbol{u}_1)$ and $(\lambda_2, \boldsymbol{u}_2)$ eigenpairs with $\lambda_1 \neq \lambda_2$

$$\begin{aligned}
\lambda_1(\boldsymbol{v}_1 \bullet \boldsymbol{v}_2) &= (\lambda_1 \boldsymbol{v}_1) \bullet \boldsymbol{v}_2 \\
&= A\boldsymbol{v}_1 \bullet \boldsymbol{v}_2 \\
&= (A\boldsymbol{v}_1)^T \boldsymbol{v}_2 \\
&= \boldsymbol{v}_1^T A^T \boldsymbol{v}_2 \\
&= \boldsymbol{v}_1^T (A\boldsymbol{v}_2) \qquad [A \text{ symmetric so } A^T = A] \\
&= \boldsymbol{v}_1^T (\lambda_2 \boldsymbol{v}_2) \\
&= \lambda_2 (\boldsymbol{v}_1^T \boldsymbol{v}_2) \\
&= \lambda_2 (\boldsymbol{v}_1 \bullet \boldsymbol{v}_2)
\end{aligned}$$

So $(\lambda_1 - \lambda_2)(\boldsymbol{v}_1 \bullet \boldsymbol{v}_2) = 0$. But $\lambda_1 \neq \lambda_2$, so $\boldsymbol{v}_1 \bullet \boldsymbol{v}_2 = 0$ $\square$

Pseudoinverse of a matrix

Definition 75 (Pseudoinverse)

$A = U\Sigma V^T$ an SVD for $A \in \mathcal{M}_{mn}$, where

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \text{ with } D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$$

(D contains the nonzero singular values of A ordered as usual)

The **pseudoinverse** (or **Moore-Penrose inverse**) of A is $A^+ \in \mathcal{M}_{nm}$ given by

$$A^+ = V\Sigma^+ U^T$$

with

$$\Sigma^+ = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{nm}$$

Least squares revisited

Theorem 76

Let $A \in \mathcal{M}_{mn}$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^m$. The least squares problem $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least squares solution $\tilde{\boldsymbol{x}}$ of minimal length (closest to the origin) given by

$$\tilde{\boldsymbol{x}} = A^+\boldsymbol{b}$$

where A^+ is the pseudoinverse of A

The least squares problem

Problem Statement:

Given a system $A\boldsymbol{x} = \boldsymbol{b}$ where $A \in \mathcal{M}_{mn}$, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$ (typically $m > n$), find $\tilde{\boldsymbol{x}}$ that minimizes

$$\|\boldsymbol{b} - A\boldsymbol{x}\|^2 = \sum_{i=1}^{m}(b_i - \sum_{j=1}^{n} A_{ij}x_j)^2$$

Geometric interpretation: Find the vector $A\tilde{\boldsymbol{x}}$ in the column space of A that is closest to $\boldsymbol{b}$

Solution: $A\tilde{\boldsymbol{x}} = \text{proj}_{\text{col}(A)}(\boldsymbol{b})$

Method 1: Normal equations

The normal equations:

$$A^T A\tilde{\boldsymbol{x}} = A^T \boldsymbol{b}$$

When this works:

- Always has at least one solution
- Any solution $\tilde{\boldsymbol{x}}$ to the normal equations is a least squares solution

Computational issues:

- Forming $A^T A$ can be numerically unstable
- Condition number of $A^T A$ is the square of the condition number of A
- Still useful for theoretical analysis

Method 2: when A Has linearly independent columns

Condition: $A \in \mathcal{M}_{mn}$ has linearly independent columns

Then: $A^T A$ is invertible and the least squares solution is **unique**

$$\tilde{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$$

Properties:

- $A^T A \in \mathcal{M}_n$ is square, symmetric, and positive definite
- $(A^T A)^{-1} A^T$ is called the *left pseudoinverse* of A
- This gives the unique least squares solution

Drawback: Computing $(A^T A)^{-1}$ directly can be numerically unstable

Method 3: QR factorization

QR Factorization: If $A \in \mathcal{M}_{mn}$ has linearly independent columns, then

$$A = QR$$

where $Q \in \mathcal{M}_{mn}$ has orthonormal columns and $R \in \mathcal{M}_n$ is upper triangular and nonsingular

Least squares solution:

$$\tilde{\boldsymbol{x}} = R^{-1} Q^T \boldsymbol{b}$$

Advantages:

- More numerically stable than forming $A^T A$
- R is upper triangular $\Rightarrow$ solving $R\tilde{\boldsymbol{x}} = Q^T \boldsymbol{b}$ by back substitution
- Condition number of R equals condition number of A
- Gram-Schmidt or Householder reflections can compute QR factorization

Method 4: Singular Value Decomposition (SVD)

SVD: For any $A \in \mathcal{M}_{mn}$,

$$A = U\Sigma V^T$$

where $U \in \mathcal{M}_m$ orthogonal, $V \in \mathcal{M}_n$ orthogonal, $\Sigma \in \mathcal{M}_{mn}$ with $\Sigma_{ii} = \sigma_i \geq 0$ (singular values)

Pseudoinverse: $A^+ = V\Sigma^+U^T$ where Σ^+ has $(\Sigma^+)_{ii} = 1/\sigma_i$ if $\sigma_i > 0$, else 0

Least squares solution:

$$\tilde{\boldsymbol{x}} = A^+\boldsymbol{b}$$

Key advantages:

- Works for *any* matrix A (even when columns are linearly dependent)
- Gives the solution of *minimal length* when multiple solutions exist
- Most numerically stable method
- Reveals the rank of A through the number of non-zero singular values

When to use which method

Method	**When to use**	**Advantages/Drawbacks**
Normal equations	Theory, small problems	Simple, but unstable
$(A^T A)^{-1} A^T$	A has LI columns	Explicit formula, unstable
QR Factorization	A has LI columns	Stable, efficient
SVD	Any A, rank-deficient	Most stable, handles all cases

Use QR for well-conditioned problems with LI columns, SVD for rank-deficient or ill-conditioned problems

Change of basis

Definition 77 (Change of basis matrix)

$\mathcal{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ and $\mathcal{C} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ bases of vector space V
The **change of basis matrix** $P_{\mathcal{C} \leftarrow \mathcal{B}} \in \mathcal{M}_n$,

$$P_{\mathcal{C} \leftarrow \mathcal{B}} = [[\boldsymbol{u}_1]_{\mathcal{C}} \cdots [\boldsymbol{u}_n]_{\mathcal{C}}]$$

has columns the coordinate vectors $[\boldsymbol{u}_1]_{\mathcal{C}}, \ldots, [\boldsymbol{u}_n]_{\mathcal{C}}$ of vectors in $\mathcal{B}$ with respect to $\mathcal{C}$

Theorem 78

$\mathcal{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ and $\mathcal{C} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ bases of vector space V and $P_{\mathcal{C} \leftarrow \mathcal{B}}$ a change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$

1. $\forall \boldsymbol{x} \in V$, $P_{\mathcal{C} \leftarrow \mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}} = [\boldsymbol{x}]_{\mathcal{C}}$
2. $P_{\mathcal{C} \leftarrow \mathcal{B}}$ *s.t.* $\forall \boldsymbol{x} \in V$, $P_{\mathcal{C} \leftarrow \mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}} = [\boldsymbol{x}]_{\mathcal{C}}$ *is* ***unique***
3. $P_{\mathcal{C} \leftarrow \mathcal{B}}$ *invertible and* $P_{\mathcal{C} \leftarrow \mathcal{B}}^{-1} = P_{\mathcal{B} \leftarrow \mathcal{C}}$

Row-reduction method for changing bases

Theorem 79

$\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ bases of vector space V. Let $\mathcal{E}$ be any basis for V,

$$B = [[\mathbf{u}_1]_{\mathcal{E}}, \ldots, [\mathbf{u}_n]_{\mathcal{E}}] \text{ and } C = [[\mathbf{v}_1]_{\mathcal{E}}, \ldots, [\mathbf{v}_n]_{\mathcal{E}}]$$

and let $[C|B]$ be the augmented matrix constructed using C and B. Then

$$RREF([C|B]) = [\mathbb{I}|P_{\mathcal{C}\leftarrow\mathcal{B}}]$$

If working in $\mathbb{R}^n$, this is quite useful with $\mathcal{E}$ the standard basis of $\mathbb{R}^n$ (it does not matter if $\mathcal{B} = \mathcal{E}$)

Definition 80 (Variance)

Let X be a random variable. The **variance** of X is given by

$$\operatorname{Var} X = E\left[(X - E(X))^2\right]$$

where E is the expected value

Definition 81 (Covariance)

Let X, Y be jointly distributed random variables. The **covariance** of X and Y is given by

$$\operatorname{cov}(X, Y) = E\left[(X - E(X))\,(Y - E(Y))\right]$$

Note that $\operatorname{cov}(X, X) = E\left[(X - E(X))^2\right] = \operatorname{Var} X$

Definition 82 (Unbiased estimators of the mean and variance)

Let $x_1, \ldots, x_n$ be data points (the *sample*) and

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

be the **mean** of the data. An unbiased estimator of the variance of the sample is

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Definition 83 (Unbiased estimator of the covariance)

Let $(x_1, y_1), \ldots, (x_n, y_n)$ be data points,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

be the means of the data. An estimator of the covariance of the sample is

$$\text{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

The covariance matrix (we usually have more than 2 variables)

Definition 84

Suppose p random variables $X_1, \ldots, X_p$. Then the covariance matrix is the symmetric matrix

$$\begin{pmatrix} \operatorname{cov}(X_1, X_1) & \operatorname{cov}(X_1, X_2) & \cdots & \operatorname{cov}(X_1, X_p) \\ \operatorname{cov}(X_2, X_1) & \operatorname{cov}(X_2, X_2) & \cdots & \operatorname{cov}(X_2, X_p) \\ \vdots & \vdots & & \vdots \\ \operatorname{cov}(X_p, X_1) & \operatorname{cov}(X_p, X_2) & \cdots & \operatorname{cov}(X_p, X_p) \end{pmatrix}$$

i.e., using the properties of covariance,

$$\begin{pmatrix} \operatorname{Var} X_1 & \operatorname{cov}(X_1, X_2) & \cdots & \operatorname{cov}(X_1, X_p) \\ \operatorname{cov}(X_1, X_2) & \operatorname{Var} X_2 & \cdots & \operatorname{cov}(X_2, X_p) \\ \vdots & \vdots & & \vdots \\ \operatorname{cov}(X_1, X_p) & \operatorname{cov}(X_2, X_p) & \cdots & \operatorname{Var} X_p \end{pmatrix}$$

Picking the right eigenvalue

$(\lambda, \boldsymbol{\alpha}_1)$ eigenpair of Σ, with $\boldsymbol{\alpha}_1$ having unit length

But which λ to choose?

Recall that we want Var $\boldsymbol{\alpha}_1^T \boldsymbol{x} = \boldsymbol{\alpha}_1^T \Sigma \boldsymbol{\alpha}_1$ maximal

We have

$$\text{Var } \boldsymbol{\alpha}_1^T \boldsymbol{x} = \boldsymbol{\alpha}_1^T \Sigma \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_1^T (\Sigma \boldsymbol{\alpha}_1) = \boldsymbol{\alpha}_1^T (\lambda \boldsymbol{\alpha}_1) = \lambda (\boldsymbol{\alpha}_1^T \boldsymbol{\alpha}_1) = \lambda$$

$\implies$ we pick $\lambda = \lambda_1$, the largest eigenvalue (covariance matrix symmetric so eigenvalues real)

What we have this far..

The first principal component is $\alpha_1^T \boldsymbol{x}$ and has variance λ_1, where λ_1 the largest eigenvalue of Σ and α_1 an associated eigenvector with $\|\alpha_1\| = 1$

We want the second principal component to be *uncorrelated* with $\alpha_1^T \boldsymbol{x}$ and to have maximum variance Var $\alpha_2^T \boldsymbol{x} = \alpha_2^T \Sigma \alpha_2$, under the constraint that $\|\alpha_2\| = 1$

$\alpha_2^T \boldsymbol{x}$ uncorrelated to $\alpha_1^T \boldsymbol{x}$ if $\mathrm{cov}(\alpha_1^T \boldsymbol{x}, \alpha_2^T \boldsymbol{x}) = 0$

We have

$$
\begin{aligned}
\mathrm{cov}(\boldsymbol{\alpha}_1^T \boldsymbol{x}, \boldsymbol{\alpha}_2^T \boldsymbol{x}) &= \boldsymbol{\alpha}_1^T \Sigma \boldsymbol{\alpha}_2 \\
&= \boldsymbol{\alpha}_2^T \Sigma^T \boldsymbol{\alpha}_1 \\
&= \boldsymbol{\alpha}_2^T \Sigma \boldsymbol{\alpha}_1 \quad [\Sigma \text{ symmetric}] \\
&= \boldsymbol{\alpha}_2^T (\lambda_1 \boldsymbol{\alpha}_1) \\
&= \lambda \boldsymbol{\alpha}_2^T \boldsymbol{\alpha}_1
\end{aligned}
$$

So $\boldsymbol{\alpha}_2^T \boldsymbol{x}$ uncorrelated to $\boldsymbol{\alpha}_1^T \boldsymbol{x}$ if $\boldsymbol{\alpha}_1 \perp \boldsymbol{\alpha}_2$

This is beginning to sound a lot like Gram-Schmidt, no?

In short

Take whatever covariance matrix is available to you (known Σ or sample S_X) – assume sample from now on for simplicity

For $i = 1, \ldots, p$, the ith principal component is

$$z_i = \boldsymbol{v}_i^T \boldsymbol{x}$$

where $\boldsymbol{v}_i$ eigenvector of S_X associated to the ith largest eigenvalue λ_i

If $\boldsymbol{v}_i$ is normalised, then $\lambda_i = \text{Var } z_k$

Covariance matrix

Σ the covariance matrix of the random variable, S_X the sample covariance matrix

$X \in \mathcal{M}_{mp}$ the data, then the (sample) covariance matrix S_X takes the form

$$S_X = \frac{1}{n-1} X^T X$$

where the data is centred!

Sometimes you will see $S_X = 1/(n-1)XX^T$. This is for matrices with observations in columns and variables in rows. Just remember that you want the covariance matrix to have size the number of variables, not observations, this will give you the order in which to take the product

Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

Definition 85 (Markov chain)

An experiment with finite number of possible outcomes $S_1, \ldots, S_n$ is repeated. The sequence of outcomes is a **Markov chain** if there is a set of n^2 numbers $\{p_{ij}\}$ such that the conditional probability of outcome S_i on any experiment given outcome S_j on the previous experiment is p_{ij}, i.e., for $1 \leq i, j \leq n$, $t = 1, \ldots$,

$$p_{ij} = \mathbb{P}(S_i \text{ on experiment } t+1 \mid S_j \text{ on experiment } t)$$

Outcomes $S_1, \ldots, S_n$ are **states** and p_{ij} are **transition probabilities**. $P = [p_{ij}]$ the **transition matrix**

In the following, we often write

$$\mathbb{P}(S_i \text{ on experiment } t+1 \mid S_j \text{ on experiment } t) \text{ as } \mathbb{P}(S_i(t+1) \mid S_j(t))$$

The matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

has

- entries that are probabilities, i.e., $0 \leq p_{ij} \leq 1$
- column sum 1, which we write

$$\sum_{i=1}^{n} p_{ij} = 1, \quad j = 1, \ldots, n$$

  or, using the notation $\mathbb{1}^T = (1, \ldots, 1)$,

$$\mathbb{1}^T P = \mathbb{1}^T$$

In matrix form

$$p(t+1) = Pp(t), \quad n = 1, 2, 3, \ldots$$

where $p(t) = (p_1(t), p_2(t), \ldots, p_n(t))^T$ is a probability vector and $P = (p_{ij})$ is an $n \times n$ *transition matrix*,

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

Stochastic matrices

Definition 86 (Stochastic matrix)

The nonnegative $n \times n$ matrix M is **row-stochastic** (resp. **column-stochastic**) if $\sum_{j=1}^{n} a_{ij} = 1$ for all $i = 1, \ldots, n$ (resp. $\sum_{i=1}^{n} a_{ij} = 1$ for all $j = 1, \ldots, n$)

We often say **stochastic** and let the context determine whether we mean row- or column-stochastic

If it is both row- and column-stochastic, the matrix is **doubly stochastic**

Theorem 87

Let $M \in \mathcal{M}_n$ be a stochastic matrix. Then all eigenvalues λ of M are such that $|\lambda| \leq 1$.

Theorem 88

Let $M \in \mathcal{M}_n$ be a stochastic matrix. $\lambda = 1$ is an eigenvalue of M. If M is row-stochastic, the eigenvalue 1 is associated to the column vector of ones (a right eigenvector of M); if M is column-stochastic, the eigenvalue 1 is associated to the row vector of ones (a left eigenvector of M)

Proof of Theorem 88

Suppose $M \in \mathcal{M}_n$ is row-stochastic. One way to write the requirement that each row sum equals 1 is as

$$M\mathbf{1} = \mathbf{1} \tag{3}$$

where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{C}^n$ is a column vector

If $M \in \mathcal{M}_n$, then the eigenpair equation takes the form

$$M\boldsymbol{v} = \lambda \boldsymbol{v}, \quad \boldsymbol{v} \neq \mathbf{0}$$

So, in (3), $\boldsymbol{v} = \mathbf{1}$ and $\lambda = 1$

This works the same way for a column-stochastic matrix, except that here the relation is $\mathbf{1}M = \mathbf{1}$ with $\mathbf{1}$ a row vector and the (left)eigenpair relation is $\boldsymbol{v}^T M = \lambda \boldsymbol{v}^T$ with $\boldsymbol{v}^T$ a row vector

Long time behaviour

Let $p(0)$ be the initial distribution vector. Then

$$\begin{aligned} p(1) &= Pp(0) \\ p(2) &= Pp(1) \\ &= P\left(Pp(0)\right) \\ &= P^2 p(0) \end{aligned}$$

Continuing, we get, for any t,

$$p(t) = P^t p(0)$$

Therefore,

$$\lim_{t \to +\infty} p(t) = \lim_{t \to +\infty} P^t p(0) = \left(\lim_{t \to +\infty} P^t \right) p(0)$$

if this limit exists

The matrix P^t

Theorem 89

If M, N are nonsingular stochastic matrices, then MN is a stochastic matrix

Corollary 90

If M is a nonsingular stochastic matrix, then for any $k \in \mathbb{N}$, M^k is a stochastic matrix

So P^t is stochastic

Regular Markov chains

Definition 91 (Regular Markov chain)

A **regular** Markov chain has P^k (entry-wise) positive for some integer $k > 0$, i.e., P^k has only positive entries

Definition 92 (Primitive matrix)

A nonnegative matrix M is **primitive** if, and only if, there is an integer $k > 0$ such that M^k is positive.

Theorem 93

Markov chain regular $\iff$ *transition matrix* P *primitive*

What is a directed graph?

Definition 94 (Digraph)

A **directed graph** (or **digraph**) G is a pair (V, A) where:

- V is a finite set of elements called **vertices** or **nodes**
- $A \subseteq V \times V$ is a set of ordered pairs of vertices called **arcs** or **directed edges**

Definition 95 (Arc)

An **arc** $a = (u, v) \in A$ represents a connection **from** vertex u **to** vertex v

- u is the **tail** of the arc
- v is the **head** of the arc

Definition 96 (Reducible/irrecible matrix)

A matrix $M \in \mathcal{M}_n$ is **reducible** if there exists a permutation matrix P such that

$$P^T M P = \begin{pmatrix} P & Q \\ \mathbf{0} & R \end{pmatrix},$$

i.e., M is similar to a block upper triangular matrix. The matrix M is **irreducible** if no such matrix exists

Definition 97 (Strongly connected digraph)

A digraph $\mathcal{G} = (V, A)$ is **strongly connected** if for any pair of vertices $u, v \in V$, there is a directed path from u to v

Theorem 98

$P \in \mathcal{M}_n$ irreducible $\iff$ $\mathcal{G}(P)$ strongly connected

A sufficient condition for primitivity

Theorem 99

Let $M \in \mathcal{M}_n$ be a nonnegative matrix. If $\mathcal{G}(M)$ is strongly connected and at least one of the diagonal entries m_{ii} of M is positive, then M is primitive

Behaviour of a regular MC

Theorem 100

If P is the transition matrix of a regular Markov chain, then

1. *the powers P^t approach a stochastic matrix W*
2. *each column of W is the same (column) vector $w = (w_1, \ldots, w_n)^T$*
3. *the components of w are positive*

So if the Markov chain is regular

$$\lim_{t \to +\infty} p(t) = \lim_{t \to +\infty} P^t p(0) = W p(0)$$

Computing W

Recall that since P is a stochastic matrix, 1 is an eigenvalue of P. As P is column stochastic, 1 is associated to the left (row) eigenvector $\mathbb{1}$

Now, if $\boldsymbol{p}(t)$ converges, then $\boldsymbol{p}(t+1) = P\boldsymbol{p}(t)$ at the limit, so $\boldsymbol{w} = \lim_{t\to\infty} \boldsymbol{p}(t)$ is a **fixed point** of the system. Replacing $\boldsymbol{p}$ with its limit, we have

$$\boldsymbol{w} = P\boldsymbol{w}$$

Solving for $\boldsymbol{w}$ thus amounts to finding $\boldsymbol{w}$ as a (right) eigenvector corresponding to the eigenvalue 1

Remember to normalise

$\boldsymbol{w}$ might have to be normalized since you want a probability vector

Check that the norm $\|\boldsymbol{w}\|_1$ defined by

$$\|\boldsymbol{w}\|_1 = |w_1| + \cdots + |w_n| = w_1 + \cdots + w_n$$

(since $\boldsymbol{w} \geq \boldsymbol{0}$) is equal to one

If not, use

$$\tilde{\boldsymbol{w}} = \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_1}$$

Absorbing Markov chains

Definition 101 (Absorbing state)

A state S_i in a Markov chain is **absorbing** if whenever it occurs on the t^{th} generation of the experiment, it then occurs on every subsequent step. In other words, S_i is absorbing if $p_{ii} = 1$ and $p_{ij} = 0$ for $i \neq j$

Definition 102 (Absorbing chain)

A Markov chain is **absorbing** if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state. In an absorbing Markov chain, a state that is not absorbing is called **transient**

Questions about absorbing chains

1. Does the process eventually reach an absorbing state?
2. What is the average number of steps spent in a transient state, if starting in a transient state?
3. What is the average number of steps before entering an absorbing state?
4. What is the probability of being absorbed by a given absorbing state, when there are more than one, when starting in a given transient state?

The answer to the first question ("Does the process eventually reach an absorbing state?") is given by the following result

Theorem 103

In an absorbing Markov chain, the probability of reaching an absorbing state is 1

To answer the other questions, write the transition matrix in **standard** form

For an absorbing chain with k absorbing states and $r-k$ transient states, write transition matrix as

$$P = \begin{pmatrix} \mathbb{I}_k & R \\ \mathbf{0} & Q \end{pmatrix}$$

with following meaning

	Absorbing states	Transient states
Absorbing states	$\mathbb{I}_k$	R
Transient states	$\mathbf{0}$	Q

with $\mathbb{I}_k$ the $k \times k$ identity matrix, $\mathbf{0}$ an $(r-k) \times k$ matrix of zeros, R an $k \times (r-k)$ matrix and Q an $(r-k) \times (r-k)$ matrix. The matrix $\mathbb{I}_{r-k} - Q$ is invertible. Let

- $N = (\mathbb{I}_{r-k} - Q)^{-1}$ the **fundamental matrix** of the MC
- T_i sum of the entries on column i of N
- $B = RN$

Answers to our remaining questions:

2. N_{ij} average number of times the process is in the ith transient state if it starts in the jth transient state

3. T_i average number of steps before the process enters an absorbing state if it starts in the ith transient state

4. B_{ij} probability of eventually entering the ith absorbing state if the process starts in the jth transient state

Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

Graphs vs digraphs vs multigraphs vs multidigraphs vs ...

Name-wise and notation-wise, this domain is a bit of a mess

- The vertex set V is essentially the only constant
- *Undirected graph* $G = (V, E)$, where E are the *edges*
- *Undirected multigraph* $G_M = (V, E)$
- *Directed graph* (or *digraph*) $G = (V, A)$, where A are the *arcs*
- *Directed multigraph* (or *multidigraph*) $G_M = (V, A)$
- Any of the above is called a *graph* and is denoted $G = (V, X)$, when we seek generality

And just to confuse the whole thing more: we often say *graph* for *unoriented graph*

Binary relation

Definition 104 (Binary relation)

- A **binary relation** is an arbitrary association of elements of one set with elements of another (maybe the same) set
- A binary relation over the sets X and Y is defined as a subset of the Cartesian product $X \times Y = \{(x, y) | x \in X, y \in Y\}$
- $(x, y) \in R$ is read "x is R-related to y" and is denoted xRy
- If $(x, y) \notin R$, we write "not xRy" or $x \not R y$

Definition 105 (Properties of binary relations)

A binary relation R over a set X is

- **Reflexive** if $\forall x \in X$, xRx
- **Irreflexive** if there does not exist $x \in X$ such that xRx
- **Symmetric** if $xRy \Rightarrow yRx$
- **Asymmetric** if $xRy \Rightarrow y\not{R}x$
- **Antisymmetric** if xRy and $yRx \Rightarrow x = y$
- **Transitive** if xRy and $yRz \Rightarrow xRz$
- **Total** (or **complete**) if $\forall x, y \in X$, xRy or yRx

Definition 106 (Equivalence relation)

A relation that is reflexive ($\forall x \in X$, xRx), symmetric ($xRy \Rightarrow yRx$) and transitive (xRy and $yRz \Rightarrow xRz$) is an **equivalence relation**

Definition 107 (Partial order)

*A relation that is reflexive ($\forall x \in X$, xRx), antisymmetric (xRy and $yRx \Rightarrow x = y$) and transitive (xRy and $yRz \Rightarrow xRz$) is a **partial order***

Definition 108 (Total order)

*A partial order that is total ($\forall x, y \in X$, xRy or yRx) is a **total order***

Graph

Intuitively: a graph is a set of points, and a set of relations between the points

The points are called the *vertices* of the graph and the relations are the *edges* of the graph

We can also think of the relations as being one directional, in which case the relations are the *arcs* of the digraph (a contraction of “directed graph”)

Graph, vertex and edge

Definition 109 (Graph)

An **undirected graph** is a pair $G = (V, E)$ of sets such that

- V is a set of points: $V = \{v_1, \ldots, v_p\}$
- E is a set of 2-element subsets of V: $E = \{\{v_i, v_j\}, \{v_i, v_k\}, \ldots, \{v_n, v_p\}\}$ or $E = \{v_i v_j, v_i v_k, \ldots, v_n v_p\}$

Definition 110 (Vertex)

The elements of V are the **vertices** (or nodes, or points) of the graph G. V (or $V(G)$) is the vertex set of the graph G

Definition 111 (Edge)

The elements of E are the **edges** (or lines) of the graph G. E (or $E(G)$) is the edge set of the graph G

Order and Size

Definition 112 (Order of a graph)

The number of vertices in G is the **order** of G. Using the notation $|V(G)|$ for the *cardinality* of $V(G)$,

$$|V(G)| = \text{order of G}$$

Definition 113 (Size of a graph)

The number of edges in G is the **size** of G,

$$|E(G)| = \text{size of G}$$

- A graph having order p and size q is called a $(p, q)-$graph
- A graph is finite if $|V(G)| < \infty$

Incident – Adjacent

Definition 114 (Incident)

- A vertex v is **incident** with an edge e if $v \in e$; then e is an edge at v
- If $e = uv \in E(G)$, then u and v are each incident with e
- The two vertices incident with an edge are its ends
- An edge $e = uv$ is incident with both vertices u and v

Definition 115 (Adjacent)

- Two vertices u and v are **adjacent** in a graph G if $uv \in E(G)$
- If uv and uw are distinct edges (i.e. $v \neq w$) of a graph G, then uv and uw are adjacent edges

Definition 116 (Multiple edge)

Multiple edges are two or more edges connecting the same two vertices within a multigraph

Definition 117 (Loop)

A **loop** is an edge with both the same ends; *e.g.* $\{u, u\}$ is a loop

Definition 118 (Simple graph)

A **simple graph** is a graph which contains no loops or multiple edges

Definition 119 (Multigraph)

A **multigraph** is a graph which can contain multiple edges or loops

Graph and binary relations

A simple graph G can be defined in term of a vertex set V and a binary relation over V that is

- irreflexive ($\forall u \in V, u \not R u$)
- symmetric ($\forall u, v \in V, uRv \implies vRu$)

The set of edges $E(G)$ is the set of symmetric pairs in R

If R is not irreflexive, the graph is not simple

Definition 120 (Degree of a vertex)

Let v be a vertex of $G = (V, E)$.

- The number of edges of G incident with v is the **degree** of v in G
- The degree of v in G is noted $d_G(v)$ or $deg_G(v)$

Theorem 121

Let G be a $(p, q)-$graph with vertices $v_1, \ldots, v_p$, then

$$\sum_{i=1}^{p} d_G(v_i) = 2q$$

Definition 122 (Odd vertex)

A vertex is an **odd vertex** if its degree is odd

Definition 123 (Even vertex)

A vertex is called **even vertex** if its degree is even

Theorem 124

Every graph contains an even number of odd vertices

Isomorphic graphs

Definition 125 (Isomorphic graphs)

Let $G_1 = (V(G_1), E(G_1))$ and $G_2 = (V(G_2), E(G_2))$ be two graphs. G_1 and G_2 are ***isomorphic*** *if there exists an isomorphism ϕ from G_1 to G_2, that is defined as an injective mapping $\phi : V(G_1) \to V(G_2)$ such that two vertices u_1 and v_1 are adjacent in $G_1 \iff$ the vertices $\phi(u_1)$ and $\phi(v_1)$ are adjacent in G_2*

If ϕ is an isomorphism from G_1 to G_2, then the inverse mapping ϕ^{-1} from $V(G_2)$ to $V(G_1)$ also satisfies the definition of an isomorphism. As a consequence, if G_1 and G_2 are isomorphic graphs, then

- G_1 is isomorphic to G_2
- G_2 is isomorphic to G_1

Theorem 126

The relation "is isomorphic to" is an equivalence relation on the set of all graphs

Theorem 127

If G_1 and G_2 are isomorphic graphs, then the degrees of vertices of G_1 are exactly the degrees of vertices of G_2

Subgraph

Definition 128 (Subgraph)

Let $G = (V, E)$ be a graph. A graph $H = (V(H), E(H))$ is a **subgraph** of G if $V(H) \subseteq V$ and $E(H) \subseteq E$

Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs

Definition 129 (Union of G_1 and G_2)

$G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$

Definition 130 (Intersection of G_1 and G_2)

$G_1 \cap G_2 = (V_1 \cap V_2, E_1 \cap E_2)$

Definition 131 (Disjoint graphs)

If $G_1 \cap G_2 = (\emptyset, \emptyset) = \emptyset$ (empty graph) then G_1 and G_2 are **disjoint**

Definition 132 (Complement of G_1)

The **complement** $\bar{G}_1$ of G_1 is the graph on V_1, with the edge set
$E(\bar{G}_1) = [V_1]^2 \setminus E_1$ ($e \in E(\bar{G}_1) \iff e \notin E_1$)

Graph theory

Undirected graphs

Walks, trails, paths

Complete, bipartite and other notable graphs

Planar graphs

Graph colouring

Directed graphs

Directed graphs

Walks, paths, etc.

Connectivity in digraphs

Orientable graphs

Connected vertices and graph, components

Definition 133 (Connected vertices)

Two vertices u and v in a graph G are **connected** if $u = v$, or if $u \neq v$ and there exists a path in G that links u and v

(For *path*, see Definition 146 later)

Definition 134 (Connected graph)

A graph is **connected** if every two vertices of G are connected; otherwise, G is **disconnected**

A necessary condition for connectedness

Theorem 135

A connected graph on p vertices has at least $p - 1$ edges

In other words, a connected graph G of order p has $\text{size}(G) \geq p - 1$

Connectedness is an equivalence relation

Denote $x \equiv y$ the relation "$x = y$, or $x \neq y$ and there exists a path in G connecting x and y". $\equiv$ is an equivalence relation since

1. $x \equiv y$ [reflexivity]
2. $x \equiv y \implies y \equiv x$ [symmetry]
3. $x \equiv y, y \equiv z \implies x \equiv z$ [transitivity]

Definition 136 (Connected component of a graph)

The classes of the equivalence relation $\equiv$ partition V into connected sub-graphs of G called **connected components** (or **components** for short) of G

A connected subgraph H of a graph G is a component of G if H is not contained in any connected subgraph of G having more vertices or edges than H

Vertex deletion & cut vertices

Definition 137 (Vertex deletion)

If $v \in V(G)$ is a vertex of G, the graph $G - v$ is the graph formed from G by removing v and all edges incident with v

Definition 138 (Cut-vertices)

Let G be a connected graph. Then v is a **cut-vertex** G if $G - v$ is disconnected

Edge deletion & bridges

Definition 139 (Edge deletion)

If e is an edge of G, the graph $G - e$ is the graph formed from G by removing e from G

Definition 140 (Bridge)

An edge e in a connected graph G is a **bridge** if $G - e$ is disconnected

Theorem 141

Let G be a connected graph. An edge e of G is a bridge of G $\iff$ e does not lie on any cycle of G

(For *cycle*, see Definition 149 later)

Graph theory

- Undirected graphs
- **Walks, trails, paths**
- Complete, bipartite and other notable graphs
- Planar graphs
- Graph colouring
- Directed graphs
- Directed graphs
- Walks, paths, etc.
- Connectivity in digraphs
- Orientable graphs

Walk

Definition 142 (Walk)

A **walk** in a graph $G = (V, E)$ is a non-empty alternating sequence $v_0 e_0 v_1 e_1 v_2 \dots e_{k-1} v_k$ of vertices and edges in G such that $e_i = \{v_i, v_{i+1}\}$ for all $i < k$. This walk begins with v_0 and ends with v_k

Definition 143 (Length of a walk)

The **length** of a walk is equal to the number of edges in the walk

Definition 144 (Closed walk)

If $v_0 = v_k$, the walk is **closed**

Trail and path

Definition 145 (Trail)

If the edges in the walk are all distinct, it defines a **trail** in $G = (V, E)$

Definition 146 (Path)

If the vertices in the walk are all distinct, it defines a **path** in G

The sets of vertices and edges determined by a trail is a subgraph

Distance between two vertices

Definition 147 (Distance between two vertices)

The (**geodesic**) **distance** $d(u, v)$ in $G = (V, E)$ between two vertices u and v is the length of the shortest path linking u and v in G

If no such path exists, we assume $d(u, v) = \infty$

Circuit and cycle

Definition 148 (Circuit)

A trail linking u to v, containing at least 3 edges and in which $u = v$, is a **circuit**

Definition 149 (Cycle)

A circuit which does not repeat any vertices (except the first and the last) is a **cycle** (or **simple circuit**)

Definition 150 (Length of a cycle)

The **length of a cycle** is its number of edges

Eulerian trails and circuits

Definition 151 (Eulerian trail)

A walk in an undirected multigraph M that uses each edge **exactly once** is a **Eulerian trail** of M

Definition 152 (Traversable graph)

If a graph G has a Eulerian trail, then G is a **traversable graph**

Definition 153 (Eulerian circuit)

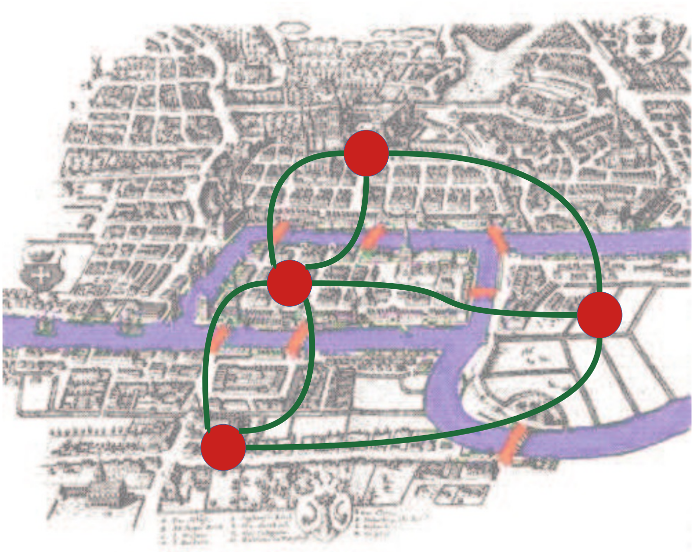
A circuit containing all the vertices and edges of a multigraph M is a **Eulerian circuit** of M

Definition 154 (Eulerian graph)

A graph (resp. multigraph) containing an Eulerian circuit is a **Eulerian graph** (resp. **Eulerian multigraph**)

Remember Euler's bridges of Königsberg?

Cross the 7 bridges in a single walk without recrossing any of them?



Theorem 155

A multigraph M is traversable $\iff$ M is connected and has exactly two odd vertices

Furthermore, any Eulerian trail of M begins at one of the odd vertices and ends at the other odd vertex

Theorem 156

A multigraph M is Eulerian $\iff$ M is connected and every vertex of M is even

Definition 157 (Hamiltonian path)

A path containing all vertices of a graph G is a **Hamiltonian path** of G

Definition 158 (Traceable graph)

If a graph G has an Hamiltonian path, then G is a **traceable graph**

Definition 159 (Hamiltonian cycle)

A cycle containing all vertices of a graph G is a **Hamiltonian cycle** of G

Definition 160 (Hamiltonian graph)

A graph containing a Hamiltonian cycle is a **Hamiltonian graph**

Theorem 161 (Dirac's theorem)

If G is a graph of order $p \geq 3$ such that $deg(v) \geq p/2$ for every vertex v of G, then G is Hamiltonian

Eulerian and Hamiltonian trails and circuits

Eulerian	Hamiltonian
A walk in an undirected multigraph M that uses each edge **exactly once** is a **Eulerian trail** of M	A path containing all vertices of a graph G is a **Hamiltonian path** of G
If a graph G has a Eulerian trail, then G is a **traversable graph**	If a graph G has an Hamiltonian path, then G is a **traceable graph**
A circuit containing all the vertices and edges of a multigraph M is a **Eulerian circuit** of M	A cycle containing all vertices of a graph G is a **Hamiltonian cycle** of G
A graph (resp. multigraph) containing an Eulerian circuit is a **Eulerian graph** (resp. **Eulerian multigraph**)	A graph containing a Hamiltonian cycle is a **Hamiltonian graph**

Graph theory

- Undirected graphs
- Walks, trails, paths
- **Complete, bipartite and other notable graphs**
- Planar graphs
- Graph colouring
- Directed graphs
- Directed graphs
- Walks, paths, etc.
- Connectivity in digraphs
- Orientable graphs

Definition 162 (Complete graph)

A graph is complete if every two of its vertices are adjacent

Definition 163 (n-clique)

A simple, complete graph on n vertices is called an n-**clique** and is often denoted K_n

Note that a complete graph of order p is $(p-1)$-regular

Bipartite graph

Definition 164 (Bipartite graph)

A graph is **bipartite** if its vertices can be partitioned into two sets V_1 and V_2, such that no two vertices in the same set are adjacent. This graph may be written $G = (V_1, V_2, E)$

Definition 165 (Complete bipartite graph)

A bipartite graph in which every two vertices from the 2 different partitions are adjacent is called a **complete bipartite graph**

We often denote $K_{p,q}$ a simple, complete bipartite graph with $|V_1| = p$ and $|V_2| = q$

Some specific graphs

Definition 166 (Tree)

Any connected graph that has no cycles is a **tree**

Definition 167 (Cycle C_n)

For $n \geq 3$, the **cycle** C_n is a connected graph of order n that is a cycle on n vertices

Definition 168 (Path P_n)

The **path** P_n is a connected graph that consists of $n \geq 2$ vertices and $n - 1$ edges. Two vertices of P_n have degree 1 and the rest are of degree 2

Definition 169 (Star S_n)

The **star** of order n is the complete bipartite graph $K_{1,n-1}$ (1 vertex of degree $n - 1$ and $n - 1$ vertices of degree 1)

Graph theory

- Undirected graphs
- Walks, trails, paths
- Complete, bipartite and other notable graphs
- **Planar graphs**
- Graph colouring
- Directed graphs
- Directed graphs
- Walks, paths, etc.
- Connectivity in digraphs
- Orientable graphs

Planar graph

Definition 170 (Planar graph)

A graph is **planar** if it *can be* drawn in the plane with no crossing edges (except at the vertices). Otherwise, it is **nonplanar**

Definition 171 (Plane graph)

A **plane graph** is a graph *that is drawn* in the plane with no crossing edges. (This is only possible if the graph is planar)

(To see the difference, have you ever played this game?)

Let G be a plane graph

- the connected parts of the plane are called **regions**
- vertices and edges that are incident with a region R make up a **boundary** of R

Theorem 172 (Euler's formula)

Let G be a connected plane graph with p vertices, q edges, and r regions, then

$$p - q + r = 2$$

Corollary 173

Let G be a plane graph with p vertices, q edges, r regions, and k connected components, then

$$p - q + r = k + 1$$

Two well-known non-planar graphs

$K_{3,3}$ and K_5 are nonplanar

Theorem 174 (Kuratowski Theorem)

A graph G is planar $\iff$ it contains no subgraph isomorphic to K_5 or $K_{3,3}$ or any subdivision of K_5 or $K_{3,3}$

Note: If a graph G is nonplanar and G is a subgraph of G', then G' is also nonplanar

Graph theory

- Undirected graphs
- Walks, trails, paths
- Complete, bipartite and other notable graphs
- Planar graphs
- **Graph colouring**
- Directed graphs
- Directed graphs
- Walks, paths, etc.
- Connectivity in digraphs
- Orientable graphs

Definition 175 (Colouring of a graph G)

A **colouring** of a graph G is an assignment of colours to the vertices of G such that adjacent vertices have different colours

Definition 176 (n-colouring of G)

A **n-colouring** is a colouring of G using n colours

Definition 177 (n-colourable)

G is **n-colourable** if there exists a colouring of G that uses n colours

Definition 178 (Chromatic number)

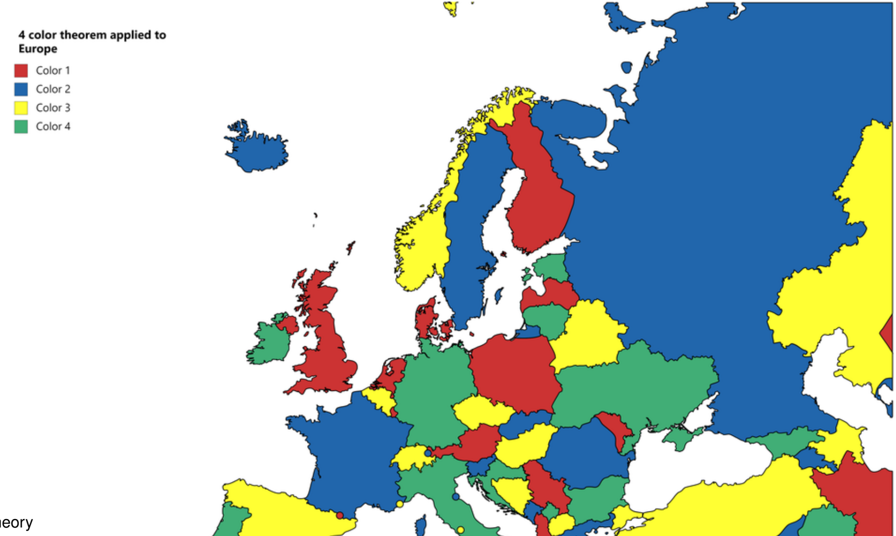
The **chromatic number** $\chi(G)$ of a graph G is the minimal value n for which an n-colouring of G exists

Property 179

- $\chi(G) = 1 \iff G$ *have no edges*
- *If* $G = K_{n,m}$, *then* $\chi(G) = 2$
- *If* $G = K_n$, *then* $\chi(G) = n$
- *For any graph G,*
$$\chi(G) \leq 1 + \Delta(G)$$
where $\Delta(G)$ *is the maximum degree of* G
- *If G is a planar graph, then* $\chi(G) \leq 4$

“Real life” problem

What is the minimal number of colours to colour all states in the map so that two adjacent states have different colours?



“Real life” problem

What is the minimal number of colours to colour all states in the map so that two adjacent states have different colours?

Mathematical representation:

- vertices correspond to the states
- vertices are adjacent $\iff$ the two states are adjacent (sharing an isolated point such as the “Four Corners” does not count)

Mathematical problem

What is the chromatic number of the graph associated to the map?

Welch-Powell algorithm for colouring a graph G

1. Order the vertices of G by decreasing degree. (Such an ordering may not be unique since some vertices may have the same degree)

2. Use one colour to paint the first vertex and to paint, in sequential order, each vertex on the list that is not adjacent to a vertex previously painted with this colour

3. Start again at the top of the list and repeat the process, painting previously unpainted vertices using a second colour

4. Repeat with additional colours until all vertices have been painted

Graph theory

Undirected graphs

Walks, trails, paths

Complete, bipartite and other notable graphs

Planar graphs

Graph colouring

Directed graphs

Directed graphs

Walks, paths, etc.

Connectivity in digraphs

Orientable graphs

Definitions

Definition 180 (Digraph)

A directed graph (or **digraph**) is a pair $G = (V, A)$ of sets such that

- V is a set of points: $V = \{v_1, v_2, v_3, .., v_p\}$
- A is a set of ordered pairs of V: $A = \{(v_i, v_j), (v_i, v_k), \ldots, (v_n, v_p)\}$ or $A = \{v_i v_j, v_i v_k, \ldots, v_n v_p\}$

Definition 181 (Vertex)

The elements of V are the vertices of the digraph G. V or $V(G)$ is the vertex set of the digraph G

Definition 182 (Arc)

The elements of A are the **arcs** (directed edges) of the digraph G. A or $A(G)$ is the arc set of the digraph G

Digraph and binary relation

A (simple) digraph D can be defined in term of a vertex set V and an irreflexive relation R over V

The defining relation R of the digraph G need not be symmetric

Directed network/weighted (di)graph

Definition 183 (Directed network)

A directed network is a digraph together with a function f,

$$f : A \to \mathbb{R},$$

which maps the arc set A into the set of real number. The value of the arc $uv \in A$ is $f(uv)$

Another name is **weighted** (di)graph

Loops & Multiple arcs

Definition 184 (Loop)

A **loop** is an arc with both the same ends; *e.g.* (u, u) is a loop

Definition 185 (Multiple arcs)

Multiple arcs (or multi-arcs) are two or more arcs connecting the same two vertices

Multidigraph/Digraph

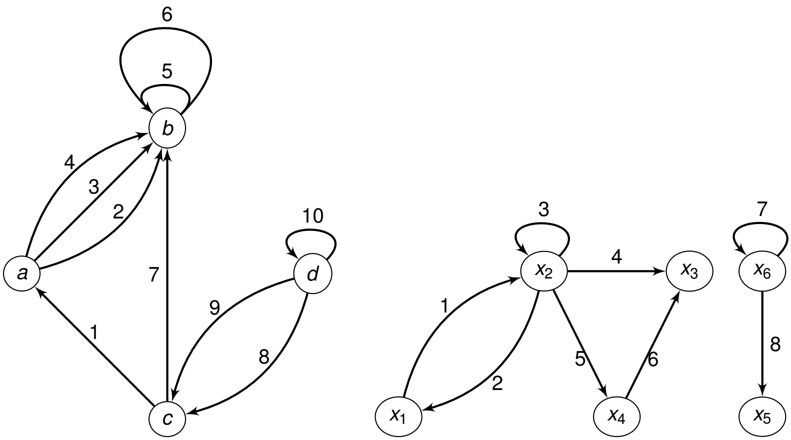
Definition 186 (Multidigraph)

A **multidigraph** is a digraph which allows repetition of arcs or loops

Definition 187 (Digraph)

In a digraph, no more than one arc can join any pair of vertices

Examples



Let $G = (V, A)$ be a digraph

Definition 188 (Arc endpoints)

For an arc $u = (x, y)$, vertex x is the **initial endpoint**, and vertex y is the **terminal endpoint**

Definition 189 (Predecessor - Successor)

If $(u, v) \in A(G)$ is an arc of G, then

- u is a **predecessor** of v
- v is a **successor** of u

Definition 190 (Neighbours of a vertex)

Let $x \in V$ be a vertex. The **neighbours** of x is the set $\Gamma(x) = \Gamma_G^+(x) \cup \Gamma_G^-(x)$, where $\Gamma_G^+(x)$ and $\Gamma_G^-(x)$ are, respectively, the set of successors and predecessors of v

Sources and sinks

Definition 191 (Directed away - Directed towards)

If $a = (u, v) \in A(G)$ is an arc of G, then

- *the arc a is said to be* ***directed away*** *from u*
- *the arc a is said to be* ***directed towards*** *v*

Definition 192 (Source - Sink)

- *Any vertex which has no arcs directed towards it is a* ***source***
- *Any vertex which has no arcs directed away from it is a* ***sink***

Adjacent arcs

Definition 193 (Adjacent arcs)

Two arcs are **adjacent** if they have at least one endpoint in common

Arcs incident to a subset of arcs

Definition 194 (Arc incident out of $X \subset A(G)$)

If the initial endpoint of an arc u belongs to $X \subset A(G)$ and if the terminal endpoint of arc u does not belong to X, then u is said to be **incident out of** X; we write $u \in \omega^+(X)$

Similarly, we define an **arc incident into** X and the set $\omega^-(X)$

Finally, the set of arcs **incident to** X is denoted

$$\omega(X) = \omega^+(X) \cup \omega^-(X)$$

Definition 195 (Subgraph of G generated by $A \subset V$)

The **subgraph** of G generated by A is the graph with A as its vertex set and with all the arcs in G that have both their endpoints in A. If $G = (V, \Gamma)$ is a 1-graph, then the subgraph generated by A is the 1-graph $G_A = (A, \Gamma_A)$ where

$$\Gamma_A(x) = \Gamma(x) \cap A \qquad (x \in A)$$

Definition 196 (Partial graph of G generated by $V \subset U$)

The graph (X, V) whose vertex set is X and whose arc set is V. In other words, it is graph G without the arcs $U - V$

Definition 197 (Partial subgraph of G)

A partial subgraph of G is the subgraph of a partial graph of G

Graph theory

- Undirected graphs
- Walks, trails, paths
- Complete, bipartite and other notable graphs
- Planar graphs
- Graph colouring
- Directed graphs
- **Directed graphs**
- Walks, paths, etc.
- Connectivity in digraphs
- Orientable graphs

Degree

Let v be a vertex of a digraph $G = (V, A)$

Definition 198 (Outdegree of a vertex)

The number of arcs directed away from a vertex v, in a digraph is called the **outdegree** of v and is written $d_G^+(v)$

Definition 199 (Indegree of a vertex)

The number of arcs directed towards a vertex v, in a digraph is called the **indegree** of v and is written $d_G^-(v)$

Definition 200 (Degree)

For any vertex v in a digraph, the **degree** of v is defined as

$$d_G(v) = d_G^+(v) + d_G^-(v)$$

```r
# Degrees in a digraph: out-, in-, and total degree
deg_out <- degree(G, mode = "out")
deg_in  <- degree(G, mode = "in")
deg_tot <- degree(G, mode = "all")
head(data.frame(v = as.integer(V(G)), out = deg_out, in_deg = deg_in, total
```

```
##   v out in_deg total
## 1 1   2      2     4
## 2 2   2      2     4
## 3 3   2      2     4
## 4 4   2      2     4
## 5 5   3      2     5
## 6 6   2      2     4
```

Theorem 201

For any (di)graph, the sum of the degrees of the vertices equals twice the number of edges (arcs)

Corollary 202

In any (di)graph, the sum of the degrees of the vertices is a nonnegative even integer

Theorem 203

If G is a digraph with vertex set $V(G) = \{v_1, \dots, v_p\}$ and q arcs, then

$$\sum_{i=1}^{p} d_G^+(v_i) = \sum_{i=1}^{p} d_G^-(v_i) = q$$

Definition 204 (Regular digraph)

A digraph G is r-regular if $d_G^+(v) = d_G^-(v) = r$ for all $v \in V(G)$

Symmetric/antisymmetric digraphs

Definition 205 (Symmetric digraph)

Let $G = (V, A)$ be a digraph with associated binary relation R. If R is *symmetric*, the digraph is symmetric

Definition 206 (Anti-symmetric digraph)

Let $G = (V, A)$ be a digraph with associated binary relation R. The digraph G is **anti-symmetric** if

$$xRy \implies y \not{R} x$$

Definition 207 (Symmetric multidigraph)

Let $G = (V, A)$ be a multidigraph. G is symmetric if $\forall x, y \in V(G)$, the number of arcs from x to y equals the number of arcs from y to x

```r
# A perfectly symmetric digraph would have every arc reciprocated
recip <- reciprocity(G)  # ratio of mutual connections
recip

## [1] 0.05128205

# Are all edges mutual? (strict symmetry)
all_mutual <- all(is.mutual(G, E(G)))
all_mutual

## [1] FALSE
```

Graph theory

- Undirected graphs
- Walks, trails, paths
- Complete, bipartite and other notable graphs
- Planar graphs
- Graph colouring
- Directed graphs
- Directed graphs
- **Walks, paths, etc.**
- Connectivity in digraphs
- Orientable graphs

Walks

Let $G = (V, A)$ be a digraph.

Definition 208 (Directed walk)

A **directed walk** in a digraph G is a non-empty alternating sequence $v_0 a_0 v_1 a_1 v_2 \ldots a_{k-1} v_k$ of vertices and arcs in G such that $a_i = (v_i, v_{i+1})$ for all $i < k$. This walk begins with v_0 and ends with v_k

Definition 209 (Length of a directed walk)

The length of a directed walk is equal to the number of arcs in the directed walk

Definition 210 (Closed walk)

If $v_0 = v_k$, the walk is closed

Trails

Let $G = (V, A)$ be a digraph.

Definition 211 (Directed trail)

A directed walk in G in which all arcs are distinct is a **directed trail** in G

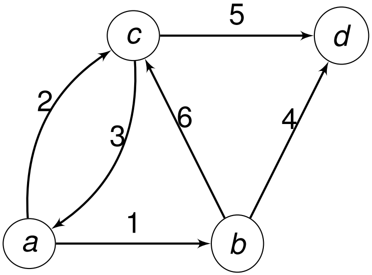
Definition 212 (Directed path)

A directed walk in G in which all vertices are distinct is a **directed path** in G

Definition 213 (Directed cycle)

A closed walk is a **directed cycle** if it contains at least three vertices and all its vertices are distinct except for $v_0 = v_k$

Examples of directed cycles



Cycles:

- $\mu^1 = (1, 6, 2) = [abca]$
- $\mu^2 = (1, 6, 3) = [abca]$
- $\mu^3 = (2, 3) = [aca]$
- $\mu^4 = (1, 4, 5, 2) = [abdca]$
- $\mu^5 = (6, 5, 4) = [acdb]$
- $\mu^6 = (1, 4, 5, 3) = [abdca]$

```r
# Shortest paths between two vertices
sp_1_to_15 <- shortest_paths(G, from = 1, to = 15, mode = "out")
sp_1_to_15$vpath

## [[1]]
## + 4/20 vertices, from 6555fe4:
## [1]  1  2  5 15

# All simple paths up to a given cutoff (to keep it small)
paths_1_to_6 <- all_simple_paths(G, from = 1, to = 6, mode = "out", cutoff =
length(paths_1_to_6)

## [1] 4

# Number of walks of length 2 via adjacency powers
Mdir <- as.matrix(as_adjacency_matrix(G, sparse = FALSE))
M2 <- Mdir %*% Mdir
M2[1, 6]

## [1] 0
```

Graph theory

- Undirected graphs
- Walks, trails, paths
- Complete, bipartite and other notable graphs
- Planar graphs
- Graph colouring
- Directed graphs
- Directed graphs
- Walks, paths, etc.
- **Connectivity in digraphs**
- Orientable graphs

Definitions

Definition 214 (Underlying graph)

Given a digraph, the undirected graph with each arc replaced by an edge is called the ***underlying graph***

Definition 215 (Weakly connected digraph)

If the underlying graph is a connected graph, then the digraph is **weakly connected**

Definition 216 (Strongly connected digraph)

A digraph G is **strongly connected** if for every two distinct vertices u and v of G, there exists a directed path from u to v

Definition 217 (Disconnected digraph)

A digraph is said to be **disconnected** if it is not weakly connected

```r
is_connected(G, mode = "weak")

## [1] TRUE

is_connected(G, mode = "strong")

## [1] FALSE

scc <- components(G, mode = "strong")
scc$no

## [1] 2

table(scc$membership)

##
##  1  2
## 12  8
```

Strong connectedness is an equivalence relation

Denote $x \equiv y$ the relation "$x = y$, or $x \neq y$ and there exists a directed path in G from x to y". $\equiv$ is an equivalence relation since

1. $x \equiv y$ [reflexivity]
2. $x \equiv y \implies y \equiv x$ [symmetry]
3. $x \equiv y, y \equiv z \implies x \equiv z$ [transitivity]

Definition 218 (Connected component of a graph)

Sets of the form

$$A(x_0) = \{x : x \in V, x \equiv x_0\}$$

are equivalence classes. They partition V into strongly connected sub-digraphs of G called **strongly connected components** (or **strong components**) of G

A strong component in G is a maximal strongly connected subdigraph of G

Theorem 219 (Properties)

Let $G = (V, A)$ be a digraph

- *If G is strongly connected, it has only one strongly connected component*
- *The strongly connected components partition the vertices $V(G)$, with every vertex in exactly one strongly connected component*

Algorithm for determining strongly connected components in $G = (V, A)$

- Determine the strongly connected component $C(v)$ containing the vertex v; if $V - C(v)$ is non-empty, re-do the same operation on the sub-digraph $G' = (V - C(v), A')$
- To determine $C(v)$, the strongly connected component containing v: let v be a vertex of a digraph , which is not already in any strongly connected component
    1. Mark the vertex v with $\pm$
    2. Mark with $+$ all successors (not already marked with $+$) of a vertex marked with $+$
    3. Mark with $-$ all predecessors (not already marked with $-$) of a vertex marked with $-$
    4. Repeat until no more possible marking with $+$ or $-$

  All vertices marked with $\pm$ belong to the same strongly connected component $C(v)$ containing the vertex v

Condensation of a digraph

Definition 220 (Condensation of a digraph)

The condensation G^* of a digraph G is a digraph having as vertices the strongly connected components (SCC) of G and such that there exists an arc in G^* from a SCC C_i to another SCC C_j if there is an arc in G from some vertex of S_i to a vertex of S_j

```r
# Condensation graph (SCC DAG) via contracting SCCs
scc <- components(G, mode = "strong")
G_cond <- contract(G, scc$membership, vertex.attr.comb = "ignore")
G_cond <- simplify(G_cond, remove.multiple = TRUE, remove.loops = TRUE)

vcount(G_cond); ecount(G_cond)

## [1] 2
## [1] 1

is_dag <- is.dag(G_cond)
is_dag

## [1] TRUE
```

Definition 221 (Articulation set)

For a connected graph, a set X of vertices is called an ***articulation set*** *(or a* ***cutset****) if the subgraph of G generated by $V - X$ is not connected*

Definition 222 (Stable set)

A set S of vertices is called a ***stable set*** *if no arc joins two distinct vertices in S*

Graph theory

Undirected graphs

Walks, trails, paths

Complete, bipartite and other notable graphs

Planar graphs

Graph colouring

Directed graphs

Directed graphs

Walks, paths, etc.

Connectivity in digraphs

Orientable graphs

Orientation

Definition 223 (Orienting a graph)

Given a connected *graph*, we describe the act of assigning a direction to each edge (edge $\rightarrow$ arc) as **orienting the graph**

Definition 224 (Strong orientation)

If the digraph resulting from orienting a graph is strongly connected, the orientation is a **strong orientation**

Orientable graph

Definition 225 (Orientable graph)

A connected graph G is **orientable** if it admits a strong orientation

Theorem 226

A connected graph $G = (V, E)$ is orientable $\iff$ G contains no bridges

(in other words, iff every edge is contained in a cycle)

```r
# Orientation and bridges (undirected case -> orientation)
Gu <- make_ring(8)
length(bridges(Gu))

## [1] 0

G_or <- as.directed(Gu, mode = "arbitrary")  # orient edges
is_connected(G_or, mode = "strong")

## [1] FALSE
```

Matrices associated to a graph/digraph

There are multiple matrices associated to a graph/digraph

The branch of graph theory that studies the properties of matrices derived from graphs and uses of these matrices in determining graph properties is *spectral graph theory*

Graphs greatly simplify some problems in linear algebra and vice versa

Matrices associated to a graph/digraph

- **Adjacency matrices**
- Other matrices associated to a graph/digraph
- Linking graphs and linear algebra

Adjacency matrix (undirected case)

Let $G = (V, E)$ be a graph of order p and size q, with vertices $v_1, \ldots, v_p$ and edges $e_1, \ldots, e_q$

Definition 227 (Adjacency matrix)

The **adjacency matrix** is

$$M_A = M_A(G) = [m_{ij}]$$

is a $p \times p$ matrix in which

$$m_{ij} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Theorem 228 (Adjacency matrix and degree)

The sum of the entries in row i of the adjacency matrix is the degree of v_i in the graph

We often write $A(G)$ and, reciprocally, if A is an adjacency matrix, $G(A)$ the corresponding graph

G undirected $\implies$ $A(G)$ symmetric

$A(G)$ has nonzero diagonal entries if G is not simple

Adjacency matrix (directed case)

Let $G = (V, A)$ be a digraph of order p with vertices $v_1, \ldots, v_p$

Definition 229 (Adjacency matrix)

The **adjacency matrix** $M = M(G) = [m_{ij}]$ is a $p \times p$ matrix in which

$$m_{ij} = \begin{cases} 1 & \text{if arc } v_i v_j \in A \\ 0 & \text{otherwise} \end{cases}$$

Theorem 230 (Properties of the adjacency matrix)

Let M be the adjacency matrix of a digraph G

- *M is not necessarily symmetric*
- *The sum of any column of M is equal to the number of arcs directed towards v_j*
- *The sum of the entries in row i is equal to the number of arcs directed away from vertex v_i*
- *The $(i, j)-$entry of M^n is equal to the number of walks of length n from vertex v_i to v_j*

```r
# Adjacency matrix (directed)
M <- as.matrix(as_adjacency_matrix(G, sparse = FALSE))

# Column sums = indegrees, row sums = outdegrees (directed case)
all.equal(rowSums(M), degree(G, mode = "out"))

## [1] TRUE

all.equal(colSums(M), degree(G, mode = "in"))

## [1] TRUE

# Walks of length 3 from 1 to 15 via M^3
M3 <- M %*% M %*% M
M3[1, 15]

## [1] 2
```

Definition 231 (Multiplicity of a pair)

The ***multiplicity*** *of a pair x, y is the number $m_G^+(x, y)$ of arcs with initial endpoint x and terminal endpoint y. Let*

$$m_G^-(x, y) = m_G^+(y, x)$$
$$m_G(x, y) = m_G^+(x, y) + m_G^-(x, y)$$

If $x \neq y$, then $m_G(x, y)$ is number of arcs with both x and y as endpoints. If $x = y$, then $m_G(x, y)$ equals twice the number of loops attached to vertex x. If $A, B \subset V$, $A \neq B$, let

$$m_G^+(A, B) = \{u : u \in U, u = (x, y), x \in A, y \in B\}$$
$$m_G(A, B) = m_G^+(A, B) + m_G^+(A, B)$$

Adjacency matrix (multigraph case)

Definition 232 (Adjacency matrix of a multigraph)

G an ℓ-graph, then the adjacency matrix $M_A = [m_{ij}]$ is defined as follows

$$m_{ij} = \begin{cases} k & \text{if arc there are } k \text{ arcs } (i,j) \in U \\ 0 & \text{otherwise} \end{cases}$$

with $k \leq \ell$

G undirected $\implies M_A(G)$ symmetric

$M_A(G)$ has nonzero diagonal entries if G is not simple.

Weighted adjacency matrices

Sometimes, adjacency matrices (typically for 1-graphs) have real entries, usually positive

This means that the arcs/edges have been given a weight

Theorem 233 (Number of walks of length n)

Let A be the adjacency matrix of a graph $G = (V(G), E(G))$, where $V(G) = \{v_1, v_2, \ldots, v_p\}$. Then the $(i,j)-$entry of A^n, $n \geq 1$, is the number of different walks linking v_i to v_j of length n in G.

(two walks of the same length are equal if their edges occur in exactly the same order)

Example: let A be the adjacency matrix of a graph $G = (V(G), E(G))$.

- the $(i,i)-$entry of A^2 is equal to the degree of v_i.
- the $(i,i)-$entry of A^3 is equal to twice the number of C_3 containing v_i.

Matrices associated to a graph/digraph

- Adjacency matrices
- **Other matrices associated to a graph/digraph**
- Linking graphs and linear algebra

Incidence matrix (undirected case)

Let $G = (V, E)$ be a graph of order p, and size q, with vertices $v_1, \ldots, v_p$, and edges $e_1, \ldots, e_q$

Definition 234 (Incidence matrix)

The incidence matrix is

$$B = B(G) = [b_{ij}]$$

is that $p \times q$ matrix in which

$$b_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is incident with } e_j \\ 0 & \text{otherwise} \end{cases}$$

Theorem 235 (Incidence matrix and degrees)

The sum of the entries in row i of the incidence matrix is the degree of v_i in the graph

Incidence matrix (directed case)

Let $G = (V, A)$ be a digraph of order p and size q, with vertices $v_1, \ldots, v_p$ and arcs $a_1, \ldots, a_q$

Definition 236 (Incidence matrix)

The **incidence matrix** $B = B(G) = [b_{ij}]$ is a $p \times q$ matrix in which

$$b_{ij} = \begin{cases} 1 & \text{if arc } a_j \text{ is directed away from a vertex } v_i \\ -1 & \text{if arc } a_j \text{ is directed towards a vertex } v_i \\ 0 & \text{otherwise} \end{cases}$$

Spectrum of a graph

We will come back to this later, but for now..

Definition 237 (Spectrum of a graph)

The **spectrum** of a graph G is the spectrum (set of eigenvalues) of its associated adjacency matrix $M(G)$

This is regardless of the type of adjacency matrix or graph

Degree matrix

Definition 238 (Degree matrix)

The **degree** matrix $D = [d_{ij}]$ for G is a $n \times n$ diagonal matrix defined as

$$d_{ij} = \begin{cases} d_G(v_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

In an undirected graph, this means that each loop increases the degree of a vertex by two

In a directed graph, the term “degree” may refer either to indegree (the number of incoming edges at each vertex) or outdegree (the number of outgoing edges at each vertex)

Laplacian matrix

Definition 239 (Laplacian matrix)

$G = (V, A)$ a simple graph with n vertices. The **Laplacian** matrix is

$$L = D(G) - M(G)$$

where $D(G)$ is the degree matrix and $M(G)$ is the adjacency matrix

Laplacian matrix (continued)

G simple graph $\implies$ $M(G)$ only contains 1 or 0 and its diagonal elements are all 0

For directed graphs, either the indegree or outdegree is used, depending on the application

Elements of L are given by

$$\ell_{ij} = \begin{cases} d_G(v_i) & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise} \end{cases}$$

```r
# Degree matrix and Laplacian
deg_vec <- degree(G, mode = "all")
Dmat <- diag(deg_vec)
L <- laplacian_matrix(G, normalized = FALSE, sparse = FALSE)

dim(Dmat); dim(L)

## [1] 20 20
## [1] 20 20

eigenvals <- eigen(as.matrix(L))$values
head(eigenvals)

## [1] 4.125310+0.0000000i 2.957548+1.4580892i 2.957548-1.4580892i
## [4] 2.598554+0.8441142i 2.598554-0.8441142i 2.172163+0.0000000i
```

Distance matrix

Let G be a graph of order p with vertices $v_1, \ldots, v_p$

Definition 240 (Distance matrix)

The distance matrix $\Delta(G) = [d_{ij}]$ is a $p \times p$ matrix in which

$$\delta_{ij} = d_G(v_i, v_j)$$

Note $\delta_{ii} = 0$ for $i = 1, \ldots, p$

```r
# Distance matrix (shortest path lengths)
D <- distances(G, mode = "out")
D[1, 15]

## [1] 3
```

Property 241

- *M is not necessarily symmetric*
- *The sum of any column of M is equal to the number of arcs directed towards v_j*
- *The sum of the entries in row i is equal to the number of arcs directed away from vertex v_i*
- *The $(i, j)-$entry of M^n is equal to the number of walks of length n from vertex v_i to v_j*

Matrices associated to a graph/digraph

- Adjacency matrices
- Other matrices associated to a graph/digraph
- **Linking graphs and linear algebra**

Counting paths

Theorem 242

G a digraph and $M_A(G)$ its adjacency matrix. Denote $P = [p_{ij}]$ the matrix $P = M_A^k$. Then p_{ij} is the number of distinct paths of length k from i to j in G

Definition 243 (Irreducible matrix)

A matrix $A \in \mathcal{M}_n$ is **reducible** if $\exists P \in \mathcal{M}_n$, permutation matrix, s.t. $P^T AP$ can be written in block triangular form. If no such P exists, A is **irreducible**

Theorem 244

A irreducible $\iff$ $G(A)$ strongly connected

Theorem 245

Let A be the adjacency matrix of a graph G on p vertices. A graph G on p vertices is connected $\iff$

$$I + A + A^2 + \cdots + A^{p-1} = C$$

has no zero entries

Theorem 246

Let M be the adjacency matrix of a digraph D on p vertices. A digraph D on p vertices is strongly connected $\iff$

$$I + M + M^2 + \cdots + M^{p-1} = C$$

has no zero entries

Nonnegative matrix

$A = [a_{ij}] \in \mathcal{M}_n(\mathbb{R})$ **nonnegative** if $a_{ij} \geq 0\ \forall i, j = 1, \ldots, n$; $\boldsymbol{v} \in \mathbb{R}^n$ nonnegative if $v_i \geq 0\ \forall i = 1, \ldots, n$. **Spectral radius** of A

$$\rho(A) = \max_{\lambda \in \mathrm{Sp}(A)} \{|\lambda|\}$$

$\mathrm{Sp}(A)$ the **spectrum** of A

Perron-Frobenius (PF) theorem

Theorem 247 (PF – Nonnegative case)

$0 \leq A \in \mathcal{M}_n(\mathbb{R})$. *Then* $\exists \mathbf{v} \geq \mathbf{0}$ *s.t.*

$$A\mathbf{v} = \rho(A)\mathbf{v}$$

Theorem 248 (PF – Irreducible case)

Let $0 \leq A \in \mathcal{M}_n(\mathbb{R})$ *irreducible. Then* $\exists \mathbf{v} > \mathbf{0}$ *s.t.*

$$A\mathbf{v} = \rho(A)\mathbf{v}$$

$\rho(A) > 0$ *and with algebraic multiplicity 1. No nonnegative eigenvector is associated to any other eigenvalue of* A

Primitive matrices

Definition 249

$0 \leq A \in \mathcal{M}_n(\mathbb{R})$ **primitive** (with **primitivity index** $k \in \mathbb{N}_+^*$) if $\exists k \in \mathbb{N}_+^*$ s.t.

$$A^k > 0,$$

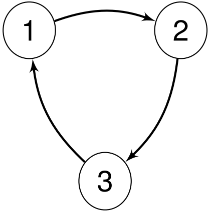
with k the smallest integer for which this is true. A **imprimitive** if it is not primitive

A primitive $\implies$ A irreducible; the converse is false

Theorem 250

$A \in \mathcal{M}_n(\mathbb{R})$ irreducible and $\exists i = 1, \ldots, n$ s.t. $a_{ii} > 0 \implies A$ primitive

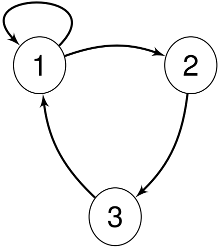
Here d is the index of imprimitivity (i.e., the number of eigenvalues that have the same modulus as $\lambda_p = \rho(A)$). If $d = 1$, then A is primitive. We have that $d =$ gcd of all the lengths of closed walks in $G(A)$



Adjacency matrix

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

Closed walks in $G(A)$ (lengths): $1 \to 1$ (3), $2 \to 2$ (3), $2 \to 2$ (3) $\implies \mathrm{gcd} = 3 \implies d = 3$ (here, all eigenvalues have modulus 1)



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

Closed walk $1 \to 1$ has length 1 $\implies$ gcd of lengths of closed walks is 1 $\implies$ A primitive

Let $\mathbf{0} \leq A \in \mathcal{M}_n$

Theorem 251

A primitive $\implies \exists\, 0 < k \leq (n-1)n^n$ *such that* $A^k > 0$

Theorem 252

If A *is primtive and the shortest simple directed cycle in* $G(A)$ *has length* s*, then the primitivity index is* $\leq n + s(n-1)$

Theorem 253

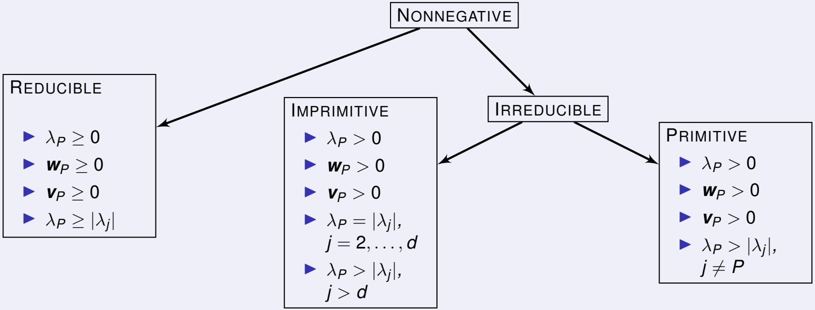
A primitive $\iff A^{n^2-2n+2} > \mathbf{0}$

Theorem 254

If A *is irreducible and has* d *positive entries on the diagonal, then the primitivity index* $\leq 2n - d - 1$

Theorem 255

$\mathbf{0} \leq A \in \mathcal{M}_n$, $\lambda_P = \rho(A)$ the Perron root of A, $\boldsymbol{v}_P$ and $\boldsymbol{w}_P$ the corresponding right and left Perron vectors of A, respectively, d the index of imprimitivity of A (with $d = 1$ when A is primitive) and $\lambda_j \in \sigma(A)$ the spectrum of A, with $j = 2, \ldots, n$ unless otherwise specified (assuming $\lambda_1 = \lambda_P$)



Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

Trees

Definition 256 (Forest, trees and branches)

- A connected graph with no cycle is a **tree**
- A tree is a connected acyclic graph, its edges are called **branches**
- A graph (connected or not) without any cycle is a **forest**. Each component is a tree

(A forest is a graph whose connected components are trees)

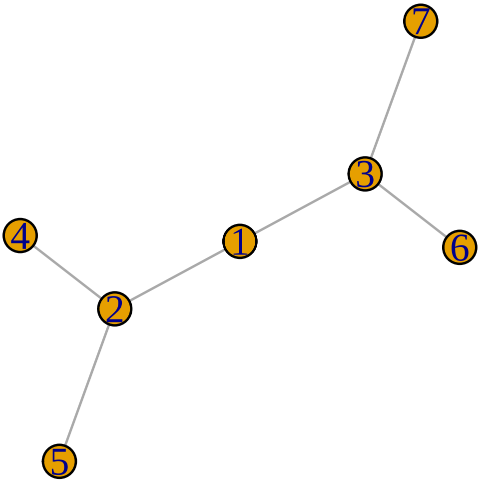
Is the “Karate graph” a tree?

```
is_acyclic(G_Z)

## [1] FALSE

is_tree(G_Z)

## [1] FALSE
```

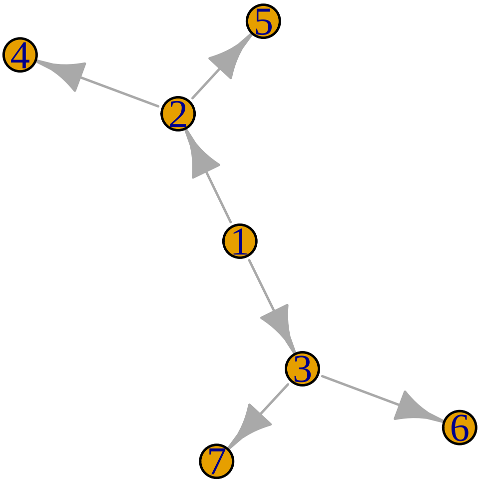
So we need a friend to play with!

```
G_tu <- make_tree(7, 2, mode = "undirected")
G_td <- make_tree(7, 2)
```

An undirected tree



An out directed tree



Property 257

- *Every edge of a tree is a bridge*
- *Given two vertices u and v of a tree, there is an unique path linking u to v*
- *A tree with p vertices and q edges satisfies $q = p - 1$. Thus, a tree is minimally connected*

(First property: the deletion of any edge of a tree diconnects it)

Every edge of a tree is a bridge

```
E(G_tu)

## + 6/6 edges from e1eaf96:
## [1] 1--2 1--3 2--4 2--5 3--6 3--7

bridges(G_tu)

## + 6/6 edges from e1eaf96:
## [1] 2--4 2--5 1--2 3--6 3--7 1--3

all(sort(E(G_tu)) == sort(bridges(G_tu)))

## [1] TRUE
```

Spanning tree

Definition 258 (Spanning tree)

A **spanning tree** of a connected graph G is a subgraph of G that contains all the vertices of G and is a tree.

A graph may have many spanning trees

Minimal spanning tree

Definition 259 (Value of a spanning tree)

The **value of a spanning tree** T of order p is

$$\sum_{i=1}^{p-1} f(e_i)$$

where f is the function that maps the edge set into $\mathbb{R}$

Definition 260 (Minimal spanning tree)

Let G be an undirected network, and let T be a **minimal spanning tree** of G.
Then T is a spanning tree whose the value is minimum

Algorithm to find a minimal spanning tree

Let $G = (V(G), E(G))$ be an undirected network and T be a minimal spanning tree

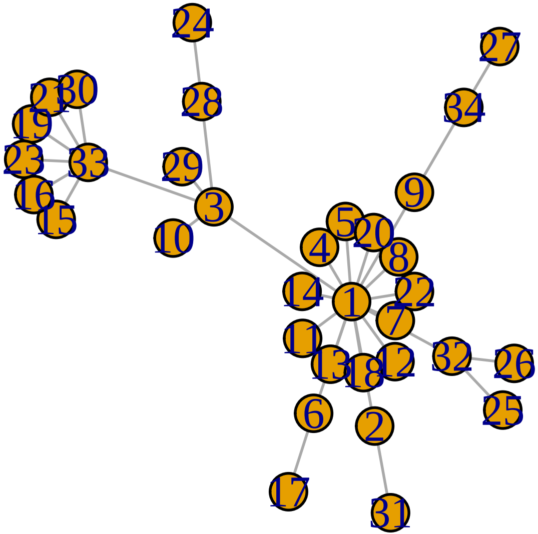
1. Sort the edges of G in increasing order by value
2. $T = (V(G), \emptyset)$
3. For each edge e in sorted order if the endpoints of e are disconnected in T add e to T

Finding a minimal spanning tree of the Karate graph

The function `mst` finds minimal spanning trees, using distances if no edge weights are provided

```
G_mst = mst(G_Z)
```

A minimal spanning tree of the Karate graph



Minimal connector problem

- Model: a graph G such that edges represent all possible connections, and each edge has a positive value which represents its cost; an undirected network G
- Solution: a minimal spanning tree T of G
  - a spanning tree of G is a subgraph of G that contains all the vertices of G and is a tree.
  - the cost of the spanning tree is the sum of values of the edges of T
  - a spanning tree such that no other spanning tree has a smaller cost is a minimmal spanning tree.

Theorem 261 (Characterisation of trees)

$H = (V, U)$ a graph of order $|V| = n > 2$. The following are equivalent and all characterise a tree :

1. *H connected and has no cycles*
2. *H has $n - 1$ arcs and no cycles*
3. *H connected and has exactly $n - 1$ arcs*
4. *H has no cycles, and if an arc is added to H, exactly one cycle is created*
5. *H connected, and if any arc is removed, the remaining graph is not connected*
6. *Every pair of vertices of H is connected by one and only one chain*

Definition 262 (Pendant vertex)

A vertex is **pendant** if it is adjacent to exactly one other vertex

Theorem 263

A tree of order $n \geq 2$ has at least two pendant vertices

Theorem 264

A graph $G = (V, U)$ has a partial graph that is a tree $\iff$ G connected

(A partial graph is a graph generated by a subset of the arcs)

Spanning tree

Can build a spanning tree as follows:

- Consider any arc u_0
- Find arc u_1 that does not form a cycle with u_0
- Find arc u_2 that does not form a cycle with $\{u_0, u_1\}$
- Continue
- When you cannot continue anymore, you have a spanning tree

Definition 265 (Minimally connected graph)

G is **minimally connected** if it is strongly connected and removal of any arc destroys strong-connectedness

A minimally connected graph is 1-graph without loops

Definition 266 (Contraction)

$G = (V, U)$. The **contraction** of the set $A \subset V$ of vertices consists in replacing A by a single vertex a and replacing each arc into (resp. out of) A by an arc with same index into (resp. out of) a

Theorem 267

G minimally connected, $A \subset V$ generating a strongly connected subgraph of G. Then the contraction of A gives a minimally connected graph

Arborescences

Definition 268 (Root)

Vertex $a \in V$ in $G = (V, U)$ is a **root** if all vertices of G can be reached by paths *starting* from a

Not all graphs have roots

Definition 269 (Quasi-strong connectedness)

G is **quasi-strongly connected** if $\forall x, y \in V$, exists $z \in V$ (denoted $z(x, y)$ to emphasize dependence on x, y) from which there is a path to x and a path to y

Strongly connected $\implies$ quasi-strongly connected (take $z(x, y) = x$); converse not true

Quasi-strongly connected $\implies$ connected

Arborescence

Definition 270 (Arborescence)

An **arborescence** is a tree that has a root

Lemma 271

$G = (V, U)$ has a root $\iff$ G quasi-strongly connected

Theorem 272

H graph of order $n > 1$. TFAE (and all characterise an arborescence)

1. *H quasi-strongly connected without cycles*
2. *H quasi-strongly connected with $n-1$ arcs*
3. *H tree having a root a*
4. *$\exists a \in V$ s.t. all other vertices are connected with a by 1 and only 1 path from a*
5. *H quasi-strongly connected and loses quasi-strong connectedness if any arc is removed*
6. *H quasi-strongly connected and $\exists a \in V$ s.t.*

$$d_H^-(a) = 0 \qquad \text{and} \qquad d_H^-(x) = 1, \quad \forall x \neq a$$

7. *H has no cycles and $\exists a \in V$ s.t.*

$$d_H^-(a) = 0 \qquad \text{and} \qquad d_H^-(x) = 1, \quad \forall x \neq a$$

Theorem 273

G has a partial graph that is an arborescence $\iff$ G quasi-strongly connected

Theorem 274

$G = (V, E)$ simple connected graph and $x_1 \in V$. It is possible to direct all edges of E so that the resulting graph $G_0 = (V, U)$ has a spanning tree H s.t.

1. *H is an arborescence with root x_1*
2. *The cycles associated with H are circuits*
3. *The only elementary circuits of G_0 are the cycles associated with H*

Counting trees

Proposition 275

X a set with n distinct objects, $n_1, \ldots, n_p$ nonnegative integers s.t. $n_1 + \cdots + n_p = n$. The number of ways to place the n objects into p boxes $X_1, \ldots, X_p$ containing $n_1, \ldots, n_p$ objects respectively is

$$\binom{n}{n_1, \ldots, n_p} = \frac{n!}{n_1! \cdots n_p!}$$

Proposition 276 (Multinomial formula)

Let $a_1, \ldots, a_p \in \mathbb{R}$ be p real numbers, then

$$(a_1 + \cdots + a_p)^n = \sum_{n_1, \ldots, n_p \geq 0} \binom{n}{n_1, \ldots, n_p} (a_1)^{n_1} \cdots (a_p)^{n_p}$$

Theorem 277

Denote $T(n; d_1, \ldots, d_n)$ the number of distinct trees H with vertices $x_1, \ldots, x_n$ and with degrees $d_H(x_1) = d_1, \ldots, d_H(x_n) = d_n$. Then

$$T(n; d_1, \ldots, d_n) = \binom{n-2}{d_1 - 1, \ldots, d_n - 1}$$

Theorem 278

The number of different trees with vertices $x_1, \ldots, x_n$ is n^{n-2}

There is a whole industry of similar results (as well as for arborescences), but we will stop here. The main point is that we are talking about a large number of possibilities..

Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

Geodesic distance

Definition 279 (Geodesic distance)

For $x, y \in V$, the **geodesic distance** $d(x, y)$ is the length of the shortest path from x to y, with $d(x, y) = \infty$ if no such path exists

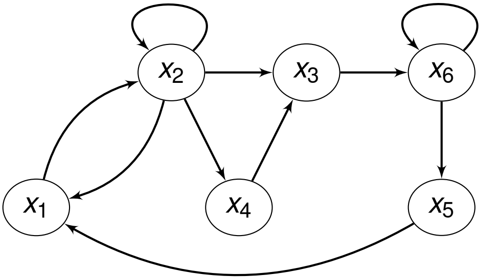
Eccentricity

Definition 280 (Vertex eccentricity)

The **eccentricity** $e(x)$ of vertex $x \in V$ is

$$e(x) = \max_{\substack{y \in V \\ y \neq x}} d(x, y)$$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & \mathbf{4} & 3 \\ 1 & 0 & 1 & 1 & \mathbf{3} & 2 \\ 3 & 4 & 0 & \mathbf{5} & 2 & 1 \\ 4 & \mathbf{5} & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & \mathbf{4} \\ 2 & 3 & \mathbf{4} & \mathbf{4} & 1 & 0 \end{pmatrix}$$



Central points, radius and centre

Definition 281 (Central point)

A **central point** of G is a vertex x_0 with smallest eccentricity

Definition 282 (Radius)

The **radius** of G is $\rho(G) = e(x_0)$, where x_0 is a centre of G In other words,

$$\rho(G) = \min_{x \in V} e(x)$$

Definition 283 (Centre)

The **centre** of G is the set of vertices that are central points of G, i.e.,

$$\{x \in V : e(x) = \rho(G)\}$$

Betweenness

Definition 284 (Betweenness)

$G = (V, A)$ a (di)graph. The **betweenness** of $v \in V$ is

$$b_{\mathcal{D}}(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where

- σ_{st} is number of shortest geodesic paths from s to t
- $\sigma_{st}(v)$ is number of shortest geodesic paths from s to t through v

In other words

- For each pair of vertices (s, t), compute the shortest paths between them
- For each pair of vertices (s, t), determine the fraction of shortest paths that pass through vertex v
- Sum this fraction over all pairs of vertices (s, t)

Normalising betweenness

Betweenness may be normalized by dividing through the number of pairs of vertices not including v:

- for directed graphs, $(n-1)(n-2)$
- for undirected graphs, $(n-1)(n-2)/2$

Number of shortest paths

Recall we found `distances(G, mode="out")`

$$\mathcal{D} = \begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ 3 & 4 & 0 & 5 & 2 & 1 \\ 4 & 5 & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & 4 \\ 2 & 3 & 4 & 4 & 1 & 0 \end{pmatrix}$$

To find the number of shortest paths between pairs of vertices, we can use powers of the adjacency matrix

Write $\mathcal{D} = [d_{ij}]$, for a given (i, j) $(i \neq j)$, if $d_{ij} = k$, then pick the (i, j) in A^k

We find

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

Recall that betweenness of v is

$$b_{\mathcal{D}}(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

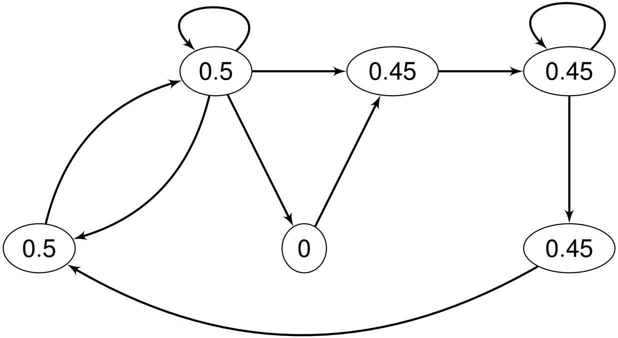
σ_{st} (# shortest paths from s to t) is found in the matrix above

What about $\sigma_{st}(v)$, # of those shortest paths that go through v?

We can use `all_shortest_paths(G, from = s, to = t, mode = "out")`

Example of betweenness

```
betweenness(G, directed = FALSE, normalized = TRUE)
```



Closeness

Definition 285

$G = (V, A)$. The **closeness** of $v \in V$ is

$$c_{\mathcal{D}}(v) = \frac{1}{n-1} \sum_{t \in V \setminus \{v\}} d_{\mathcal{D}}(v, t)$$

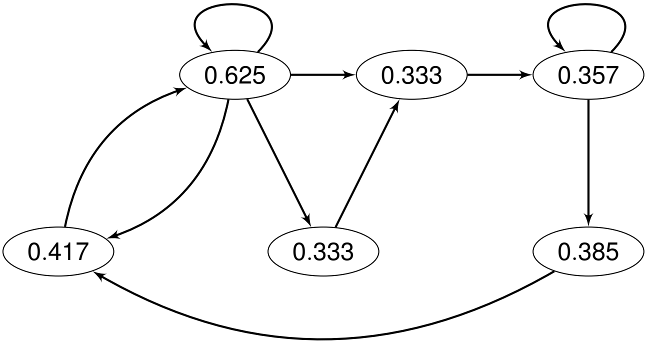
i.e., mean geodesic distance between a vertex v and all other vertices it has access to

Another definition is

$$c_{\mathcal{D}}(v) = \frac{1}{\sum_{t \in V \setminus \{v\}} d_{\mathcal{D}}(v, t)}$$

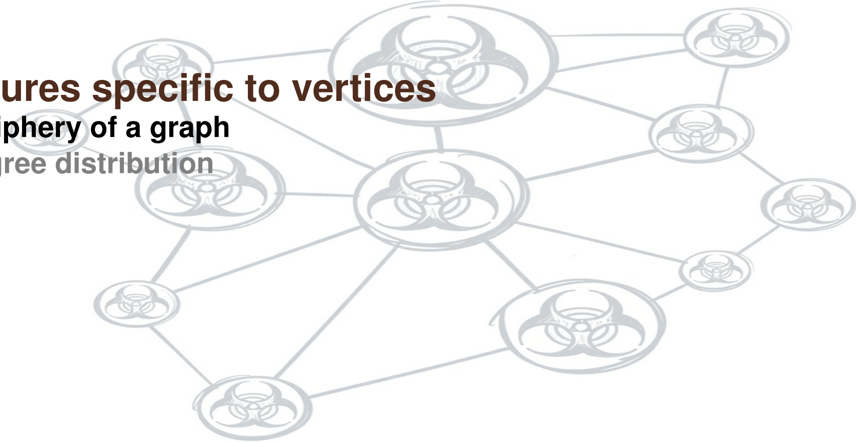
Example of (out) closeness

```
closeness(G, normalized = TRUE, mode=‘‘out’’)
```



Measures specific to vertices

- **Periphery of a graph**
- Degree distribution



Diametre and periphery of a graph

Definition 286 (Diametre of a graph)

The **diametre** of G is

$$\delta(G) = \max_{\substack{x,y \in V \\ x \neq y}} d(x, y) = \max_{x \in V} e(x)$$

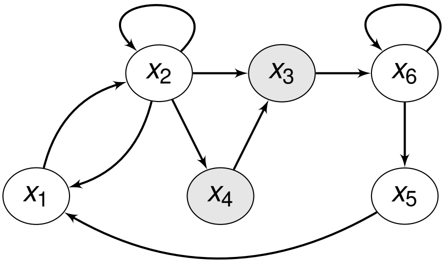
$\delta(G) < \infty \iff G$ strongly connected

Definition 287 (Periphery)

The **periphery** of a graph is the set of vertices whose eccentricity achieves the diametre, i.e.,

$$\{x \in V : e(x) = \delta(G)\}$$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & \mathbf{4} & 3 \\ 1 & 0 & 1 & 1 & \mathbf{3} & 2 \\ 3 & 4 & 0 & \mathbf{5} & 2 & 1 \\ 4 & \mathbf{5} & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & \mathbf{4} \\ 2 & 3 & \mathbf{4} & \mathbf{4} & 1 & 0 \end{pmatrix}$$



Diametre is $\delta(G) = 5$ and periphery is $\{x_3, x_4\}$

Definition 288 (Antipodal vertices)

Vertices $x, y \in V$ are **antipodal** if $d(x, y) = \delta(G)$

Measures specific to vertices

Periphery of a graph

Degree distribution

Degree distribution

Definition 289 (Arc incident to a vertex)

If a vertex x is the initial endpoint of an arc u, which is not a loop, the arc u is **incident out of vertex** x

The number of arcs incident out of x plus the number of loops attached to x is denoted $d_G^+(x)$ and is the **outer demi-degree** of x

An arc **incident into vertex** x and the **inner demi-degree** $d_G^-(x)$ are defined similarly

Definition 290 (Degree)

The **degree** of vertex x is the number of arcs with x as an endpoint, each loop being counted twice. The degree of x is denoted $d_G(x) = d_G^+(x) + d_G^-(x)$

If each vertex has the same degree, the graph is **regular**

Definition 291 (Isolated vertex)

A vertex of degree 0 is **isolated**.

Definition 292 (Average degree of G)

$d(G) = \frac{1}{|V|} \sum_{v \in V} deg_G(v)$.

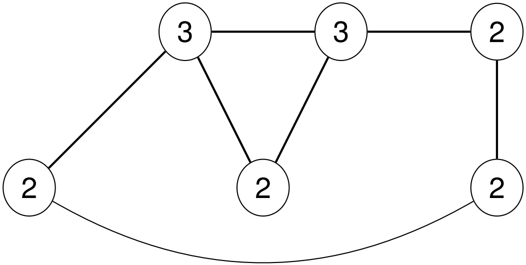
Definition 293 (Minimum degree of G)

$\delta(G) = \min\{deg_G(v) | v \in V\}$.

Definition 294 (Maximum degree of G)

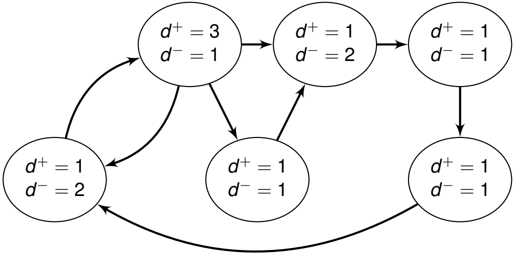
$\Delta(G) = \max\{deg_G(v) | v \in V\}$.

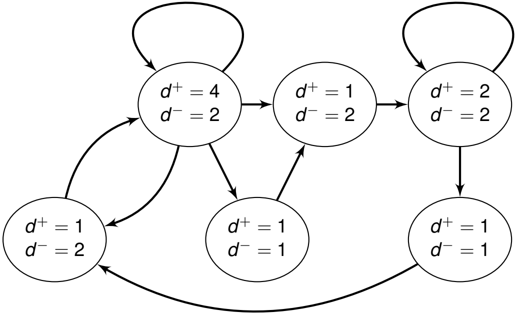
Degrees in an undirected graph



Here, vertices are labelled using the degree

Degrees in a directed graph





What to consider about degrees?

Degrees are often considered as a measure of popularity

Often write $k(i)$ (or k_i) for "degree of vertex i", $k^-(i)$ and $k^+(i)$ for in- and out-degree

- Minimum and maximum degree
- Minimum and maximum in/out-degree. E.g., if you consider the global air transportation network and the in/out-degree of airports, in-degree is a measure of a location's "popularity" as a travel destination
- Range of degrees in a graph: are there large discrepancies in connectivity between vertices in the graph?
- Average degree (often denoted $\langle k \rangle$ because of physicists)
- Average in/out-degree
- Variance of the degrees or in/out-degrees

- Average (nearest) neighbour degree, to encode for *preferential attachment* (one prefers to hang out with popular people)

$$k_i^{nn} = \frac{1}{k(i)} \sum_{j \in \mathcal{N}(i)} k(j)$$

or, in terms of the adjacency matrix $A = [a_{ij}]$,

$$k_i^{nn} = \frac{1}{k(i)} \sum_{j} a_{ij} k(j)$$

- *Excess degree*: take nearest neighbour degree but do not consider the edge/arc followed to get to the neighbour
- Degree, nearest neighbour and excess degree distributions

Degrees in `igraph`

- `degree` gives the degrees of the vertices
- `degree_distribution` gives numeric vector of the same length as the maximum degree plus one. The first element is the relative frequency zero degree vertices, the second vertices with degree one, etc.
- `knn` calculate the average nearest neighbor degree of the given vertices and the same quantity in the function of vertex degree
- `strength` sums up the edge weights of the adjacent edges for each vertex

Degree from adjacency matrix

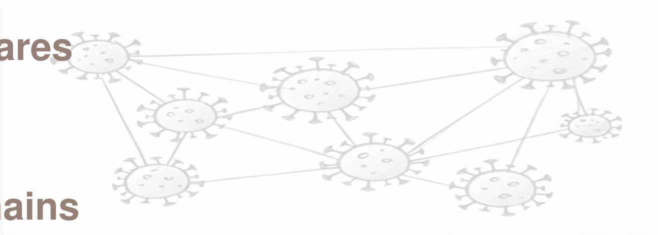
Suppose adjacency matrix take the form $A = [a_{ij}]$ with $a_{ij} = 1$ if there is an arc from the vertex indexed i to the vertex indexed j and 0 otherwise. (Could be the other way round, using A^T, just make sure)

Let $\boldsymbol{e} = (1, \ldots, 1)^T$ be the vector of all ones

$A\boldsymbol{e} = (d_G^+(1), \ldots, d_G^+(1))^T$ (out-degree)

$\boldsymbol{e}^T A = (d_G^-(1), \ldots, d_G^-(1))$ (in-degree)

Preliminary stuff

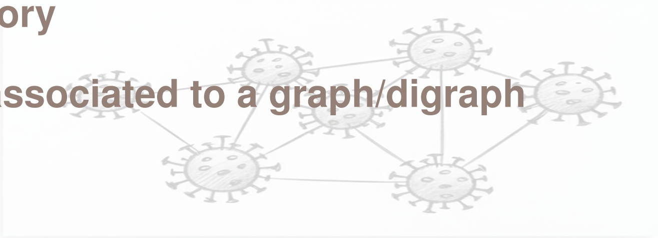


Least squares

The SVD

Markov chains

Graph theory



Matrices associated to a graph/digraph

Trees

Why characterise graphs?

Measures at the graph level

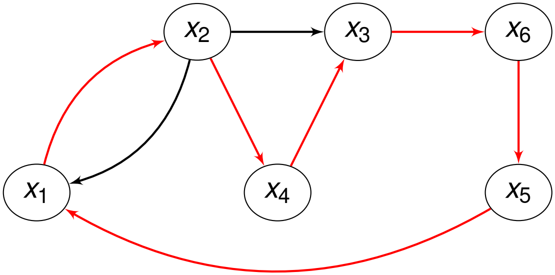
- **Circumference & Girth**
- Graph density
- Graph connectivity
- Cliques
- *k*-cores

Circumference

Definition 295 (Circumference)

In an undirected (resp. directed) graph, the total number of edges (resp. arcs) in the longest cycle of graph G is the **circumference** of G

Circumference is 6.

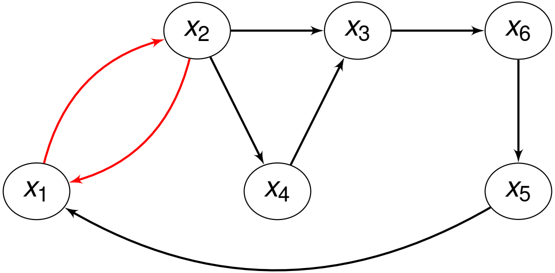


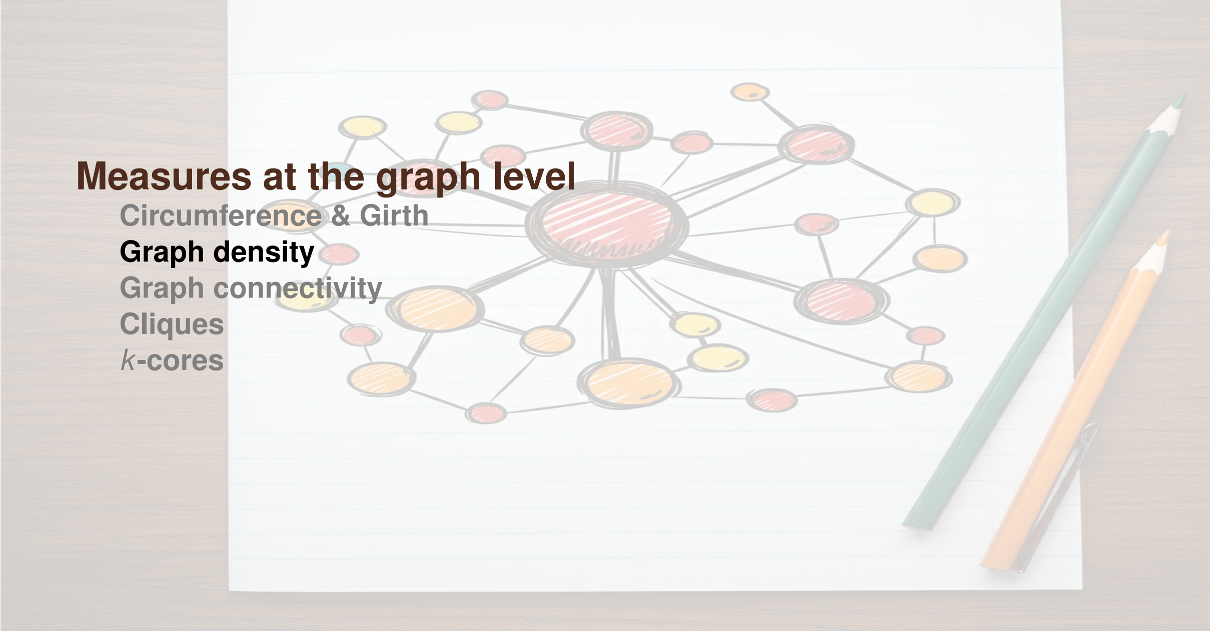
Girth

Definition 296 (Girth)

The total number of edges in the shortest cycle of graph G is the **girth** $g(G)$

Girth is 2.





Measures at the graph level

- Circumference & Girth
- **Graph density**
- Graph connectivity
- Cliques
- *k*-cores

Completeness

Definition 297 (Complete undirected graph)

An undirected graph is complete if every two of its vertices are adjacent.

Definition 298 (Complete digraph)

A digraph $D(V, A)$ is complete if $\forall u, v \in V$, $uv \in A$.

In case of simple graphs, completeness effectively means that "information" can be transmitted from every vertex to every other vertex quickly (1 step)

It can be useful to know how far away we are from being complete

Number of edges/arcs in a complete graph

$G = (V, E)$ undirected and simple of order n has at most

$$\frac{n(n-1)}{2}$$

edges, while $G = (V, A)$ directed and simple of order n has at most

$$n(n-1)$$

arcs

Density of a graph

Definition 299 (Density)

The fraction of maximum number of edges or arcs present in the graph is the **density** of the graph.

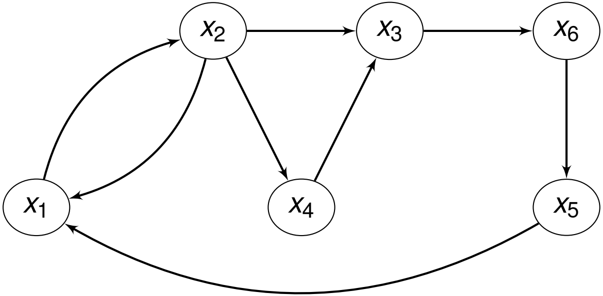
If the graph has p edges or arcs, then its density is, respectively,

$$\frac{2p}{n(n-1)}$$

or

$$\frac{p}{n(n-1)}$$

Example of density



Graph has order 6 and thus a max of 30 arcs. Here, 8 arcs $\Longrightarrow$ density 0.267 (26.7% of arcs are present)

Measures at the graph level

- Circumference & Girth
- Graph density
- **Graph connectivity**
- Cliques
- *k*-cores

Connectedness

We have already seen connectedness (quasi- or strong in the oriented case)

Connectedness is important in terms of characteristing graph properties, as it shows the capacity of the graph to convey information to all the members of the graph (the vertices)

Definition 300 (Connected graph)

A **connected graph** is a graph that contains a chain $\mu[x, y]$ for each pair x, y of distinct vertices

Denote $x \equiv y$ the relation "$x = y$, or $x \neq y$ and there exists a chain in G connecting x and y". $\equiv$ is an equivalence relation since

1. $x \equiv y$ [reflexivity]
2. $x \equiv y \implies y \equiv x$ [symmetry]
3. $x \equiv y, y \equiv z \implies x \equiv z$ [transitivity]

Definition 301 (Connected component of a graph)

The classes of the equivalence relation $\equiv$ partition V into connected sub-graphs of G called **connected components**

Articulation set

Definition 302 (Articulation set)

For a connected graph, a set A of vertices is called an **articulation set** (or a **cutset**) if the subgraph of G generated by $V - A$ is not connected

`articulation_points(G)` in `igraph` (assumes the graph is undirected, makes it so if not)

Strongly connected graphs

$G = (V, U)$ connected. A **path of length 0** is any sequence $\{x\}$ consisting of a single vertex $x \in V$

For $x, y \in V$, let $x \equiv y$ be the relation "there is a path $\mu_1[x, y]$ from x to y as well as a path $\mu_2[y, x]$ from y to x". This is an equivalence relation (it is reflexive, symmetric and transitive)

Definition 303 (Strong components)

Sets of the form

$$A(x_0) = \{x : x \in V, x \equiv x_0\}$$

are equivalence classes; they partition V and are the **strongly connected components** of G

Definition 304 (Strongly connected graph)

G **strongly connected** if it has a single strong component

Definition 305 (Minimally connected graph)

G is **minimally connected** if it is strongly connected and removal of any arc destroys strong-connectedness

Definition 306 (Contraction)

$G = (V, U)$. The **contraction** of the set $A \subset V$ of vertices consists in replacing A by a single vertex a and replacing each arc into (resp. out of) A by an arc with same index into (resp. out of) a

Quasi-strong connectedness

Definition 307 (Quasi-strong connectedness)

G **quasi-strongly connected** if $\forall x, y \in V$, exists $z \in V$ (denoted $z(x, y)$ to emphasize dependence on x, y) from which there is a path to x and a path to y

Strongly connected $\implies$ quasi-strongly connected (take $z(x, y) = x$); converse not true

Quasi-strongly connected $\implies$ connected

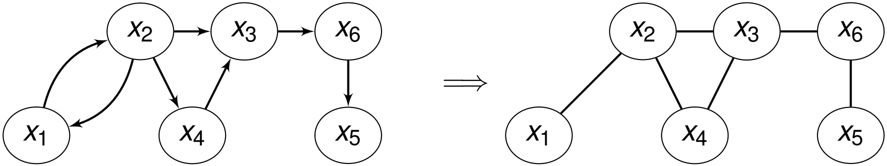
Lemma 308

$G = (V, U)$ has a root $\iff$ G quasi-strongly connected

Weak-connectedness

Definition 309 (Weakly connected graph)

$G = (V, U)$ **weakly connected** if $G = (V, E)$ connected, where E is obtained from U by ignoring the direction of arcs



Weak components

Define for $x, y \in V$ the relation $x \equiv y$ as "$x = y$ or $x \neq y$ and there is a chain in G connecting x and y" [like for components in an undirected graph, except the graph is directed here]

This defines an equivalence relation

Definition 310 (Weak components)

Sets of the form

$$A(x_0) = \{x : x \in V, x \equiv x_0\}$$

are equivalence classes partitioning V into the **weakly connected components** of G

$G = (V, U)$ is weakly connected if there is a single weak component

Components in igraph

- is_connected decides whether the graph is weakly or strongly connected
- components finds the maximal (weakly or strongly) connected components of a graph
- count_components does almost the same as components but returns only the number of clusters found instead of returning the actual clusters
- component_distribution creates a histogram for the maximal connected component sizes
- decompose creates a separate graph for each component of a graph
- subcomponent finds all vertices reachable from a given vertex, or the opposite: all vertices from which a given vertex is reachable via a directed path

Measures at the graph level

- Circumference & Girth
- Graph density
- Graph connectivity
- **Cliques**
- *k*-cores

Cliques

Definition 311 (Clique in undirected graphs)

$G = (V, E)$ a simple undirected graph. A **clique** is a subgraph G' of G such that all vertices in G' are adjacent

Definition 312 (n-clique)

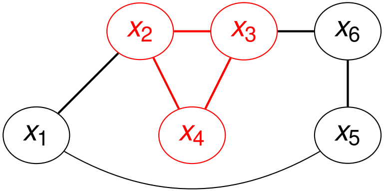
A simple, complete graph on n vertices is called an n-**clique** and is often denoted K_n

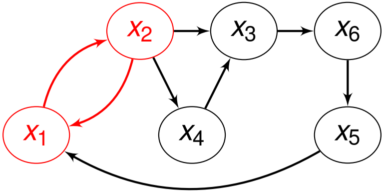
Definition 313 (Clique in directed graphs)

$G = (V, U)$ a simple directed graph. A **clique** is a subgraph G' of G such that all vertices in G' are mutually adjacent

Definition 314 (Maximal clique)

A **maximal clique** is a clique that cannot be extended by adding another adjacent vertex


x_2
x_3
x_6
x_1
x_4
x_5


x_2
x_3
x_6
x_1
x_4
x_5

Cliques in `igraph`

- `cliques` find all complete subgraphs in the input graph, obeying the size limitations given in the min and max arguments
- `largest_cliques` finds all largest cliques in the input graph
- `max_cliques` finds all maximal cliques in the input graph (The largest cliques are always maximal, but a maximal clique is not necessarily the largest)
- `count_max_cliques` counts the maximal cliques
- `clique_num` calculates the size of the largest clique(s)

Measures at the graph level

- Circumference & Girth
- Graph density
- Graph connectivity
- Cliques
- ***k*-cores**

k-core

Definition 315 (k-core of a graph)

$G = (V, U)$ a graph. The **k-core** of G is a maximal subgraph in which each vertex has degree at least k

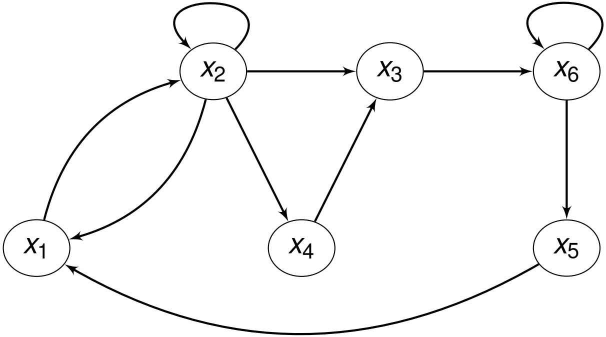
Definition 316 (Coreness of a vertex)

$G = (V, U)$ a graph, $x \in V$. The **coreness** of x is k if x belongs to the k-core of G but not to the $k + 1$ core of G

For directed graphs, in-cores or out-cores depending on whether in-degree or out-degree is used

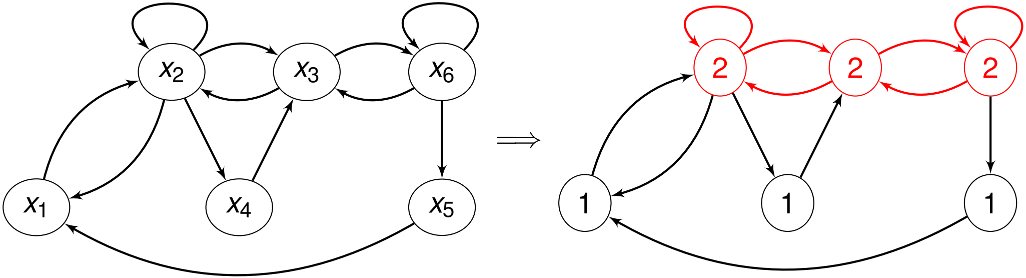
In `igraph`: `coreness`

Coreness in the directed case

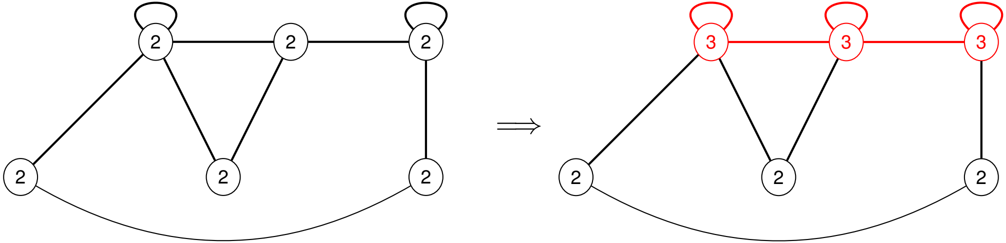


G has only a 1-in-core and 1-out-core: there is no (maximal) subgraph in which the in- or out-degree is larger than 1

In-coreness in the directed case



Coreness in the undirected case



Preliminary stuff

Least squares

The SVD

Markov chains

Graph theory

Matrices associated to a graph/digraph

Trees

Why characterise graphs?

Why characterise a graph

Graphs are everywhere!

To compare graphs, understand their properties, we need ways to describe their shape and characteristics

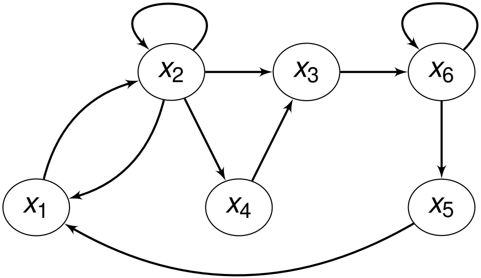
Geodesic distance

Definition 317 (Geodesic distance)

For $x, y \in V$, the **geodesic distance** $d(x, y)$ is the length of the shortest path from x to y, with $d(x, y) = \infty$ if no such path exists

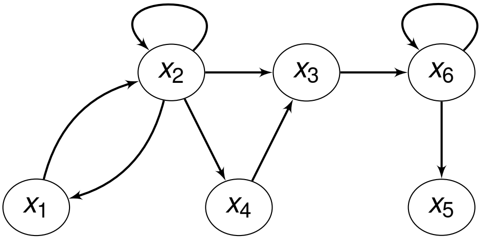
- $d(x_1, x_2) = 1$
- $d(x_1, x_3) = 2$
- $\dots$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ 3 & 4 & 0 & 5 & 2 & 1 \\ 4 & 5 & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & 4 \\ 2 & 3 & 4 & 4 & 1 & 0 \end{pmatrix}$$



- $d(x_5, x_1) = \infty$
- $d(x_3, x_1) = \infty$
- $\dots$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ \infty & \infty & 0 & \infty & 2 & 1 \\ \infty & \infty & 1 & 0 & 3 & 2 \\ \infty & \infty & \infty & \infty & 0 & \infty \\ \infty & \infty & \infty & \infty & 1 & 0 \end{pmatrix}$$



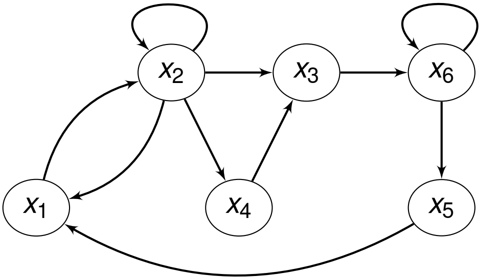
Eccentricity

Definition 318 (Vertex eccentricity)

The **eccentricity** $e(x)$ of vertex $x \in V$ is

$$e(x) = \max_{\substack{y \in V \\ y \neq x}} d(x, y)$$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & \mathbf{4} & 3 \\ 1 & 0 & 1 & 1 & \mathbf{3} & 2 \\ 3 & 4 & 0 & \mathbf{5} & 2 & 1 \\ 4 & \mathbf{5} & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & \mathbf{4} \\ 2 & 3 & \mathbf{4} & \mathbf{4} & 1 & 0 \end{pmatrix}$$



Central points, radius and centre

Definition 319 (Central point)

A **central point** of G is a vertex x_0 with smallest eccentricity

Definition 320 (Radius)

The **radius** of G is $\rho(G) = e(x_0)$, where x_0 is a centre of G In other words,

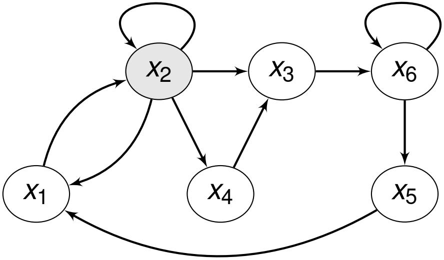
$$\rho(G) = \min_{x \in V} e(x)$$

Definition 321 (Centre)

The **centre** of G is the set of vertices that are central points of G, i.e.,

$$\{x \in V : e(x) = \rho(G)\}$$

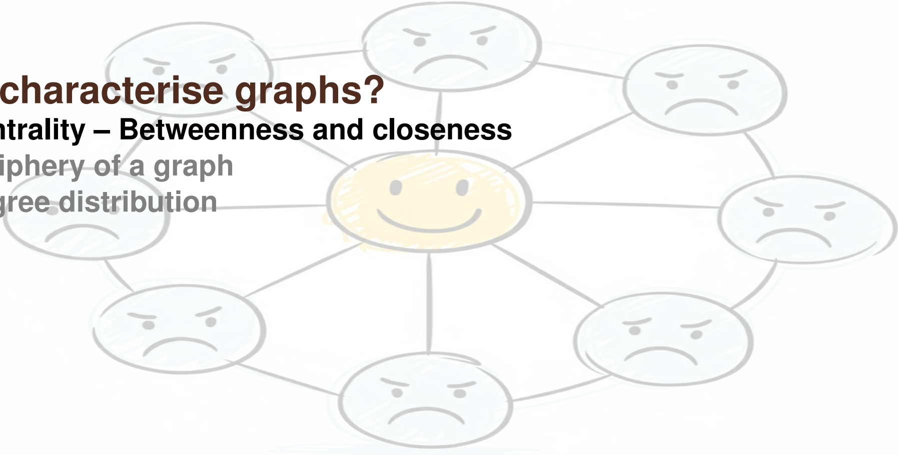
$$
\begin{pmatrix}
0 & 1 & 2 & 2 & 4 & 3 \\
1 & 0 & 1 & 1 & 3 & 2 \\
3 & 4 & 0 & 5 & 2 & 1 \\
4 & 5 & 1 & 0 & 3 & 2 \\
1 & 2 & 3 & 3 & 0 & 4 \\
2 & 3 & 4 & 4 & 1 & 0
\end{pmatrix}
$$



Radius is 3, x_2 is a central point (the only one) and the centre is $\{x_2\}$

Why characterise graphs?

- **Centrality – Betweenness and closeness**
- Periphery of a graph
- Degree distribution



How *central* is a vertex?

Centrality tries to answer the question: what are the most influent vertices?

We have seen central vertices and vertices on the periphery, let us consider two other measures of centrality

- Betweenness centrality
- Closeness centrality

Many other forms (we will come back to this, e.g., degree centrality)

Betweenness

Definition 322 (Betweenness)

$G = (V, A)$ a (di)graph. The **betweenness** of $v \in V$ is

$$b_{\mathcal{D}}(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where

- σ_{st} is number of shortest geodesic paths from s to t
- $\sigma_{st}(v)$ is number of shortest geodesic paths from s to t through v

In other words

- For each pair of vertices (s, t), compute the shortest paths between them
- For each pair of vertices (s, t), determine the fraction of shortest paths that pass through vertex v
- Sum this fraction over all pairs of vertices (s, t)

Normalising betweenness

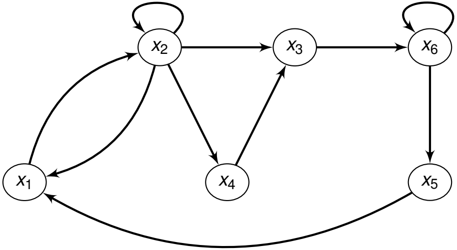
Betweenness may be normalized by dividing through the number of pairs of vertices not including v:

- for directed graphs, $(n-1)(n-2)$
- for undirected graphs, $(n-1)(n-2)/2$

Example of betweenness

```
distances(G, mode="out")
```

$$\begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ 3 & 4 & 0 & 5 & 2 & 1 \\ 4 & 5 & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & 4 \\ 2 & 3 & 4 & 4 & 1 & 0 \end{pmatrix}$$



Number of shortest paths

Recall we found `distances(G, mode="out")`

$$\mathcal{D} = \begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ 3 & 4 & 0 & 5 & 2 & 1 \\ 4 & 5 & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & 4 \\ 2 & 3 & 4 & 4 & 1 & 0 \end{pmatrix}$$

To find the number of shortest paths between pairs of vertices, we can use powers of the adjacency matrix

Write $\mathcal{D} = [d_{ij}]$, for a given (i, j) $(i \neq j)$, if $d_{ij} = k$, then pick the (i, j) in A^k

We find

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

Recall that betweenness of v is

$$b_{\mathcal{D}}(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

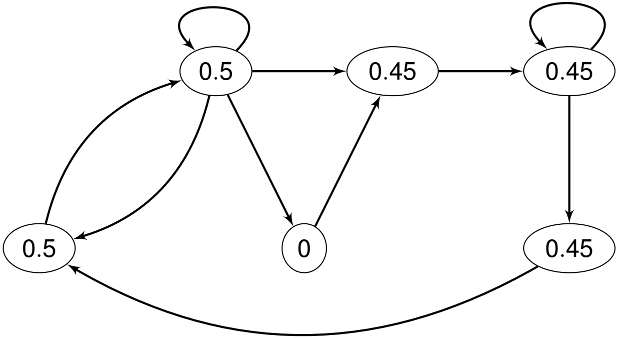
σ_{st} (# shortest paths from s to t) is found in the matrix above

What about $\sigma_{st}(v)$, # of those shortest paths that go through v?

We can use `all_shortest_paths(G, from = s, to = t, mode = "out")`

Example of betweenness

`betweenness(G, directed = FALSE, normalized = TRUE)`



Closeness

Definition 323

$G = (V, A)$. The **closeness** of $v \in V$ is

$$c_{\mathcal{D}}(v) = \frac{1}{n-1} \sum_{t \in V \setminus \{v\}} d_{\mathcal{D}}(v, t)$$

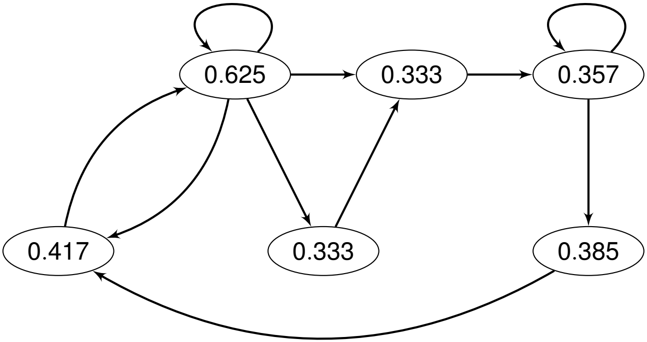
i.e., mean geodesic distance between a vertex v and all other vertices it has access to

Another definition is

$$c_{\mathcal{D}}(v) = \frac{1}{\sum_{t \in V \setminus \{v\}} d_{\mathcal{D}}(v, t)}$$

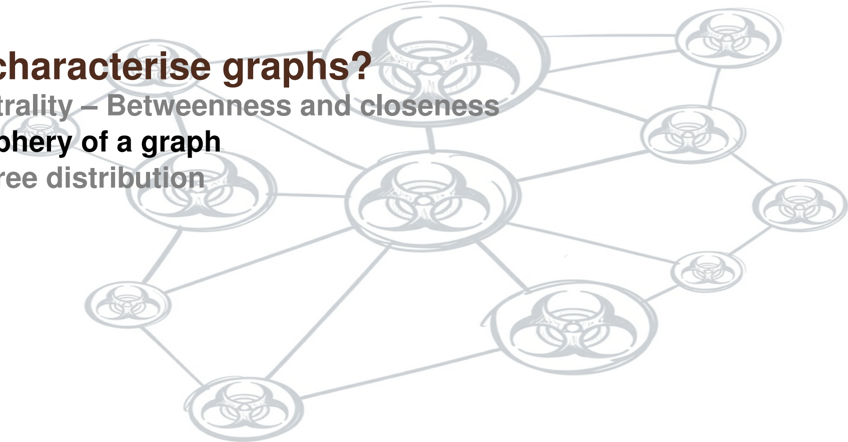
Example of (out) closeness

```
closeness(G, normalized = TRUE, mode=''out'')
```



Why characterise graphs?

- Centrality – Betweenness and closeness
- **Periphery of a graph**
- Degree distribution



Diametre and periphery of a graph

Definition 324 (Diametre of a graph)

The **diametre** of G is

$$\delta(G) = \max_{\substack{x,y \in V \\ x \neq y}} d(x,y) = \max_{x \in V} e(x)$$

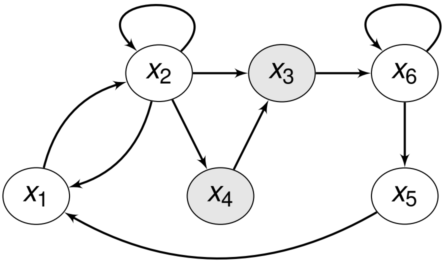
$\delta(G) < \infty \iff G$ strongly connected

Definition 325 (Periphery)

The **periphery** of a graph is the set of vertices whose eccentricity achieves the diametre, i.e.,

$$\{x \in V : e(x) = \delta(G)\}$$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ 3 & 4 & 0 & 5 & 2 & 1 \\ 4 & 5 & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & 4 \\ 2 & 3 & 4 & 4 & 1 & 0 \end{pmatrix}$$



Diametre is $\delta(G) = 5$ and periphery is $\{x_3, x_4\}$

Definition 326 (Antipodal vertices)

Vertices $x, y \in V$ are **antipodal** if $d(x, y) = \delta(G)$

Why characterise graphs?

- Centrality – Betweenness and closeness
- Periphery of a graph
- **Degree distribution**

Degree distribution

Definition 327 (Arc incident to a vertex)

If a vertex x is the initial endpoint of an arc u, which is not a loop, the arc u is **incident out of vertex** x

The number of arcs incident out of x plus the number of loops attached to x is denoted $d_G^+(x)$ and is the **outer demi-degree** of x

An arc **incident into vertex** x and the **inner demi-degree** $d_G^-(x)$ are defined similarly

Definition 328 (Degree)

The **degree** of vertex x is the number of arcs with x as an endpoint, each loop being counted twice. The degree of x is denoted $d_G(x) = d_G^+(x) + d_G^-(x)$

If each vertex has the same degree, the graph is **regular**

Definition 329 (Isolated vertex)

A vertex of degree 0 is **isolated**.

Definition 330 (Average degree of G)

$d(G) = \frac{1}{|V|} \sum_{v \in V} deg_G(v)$.

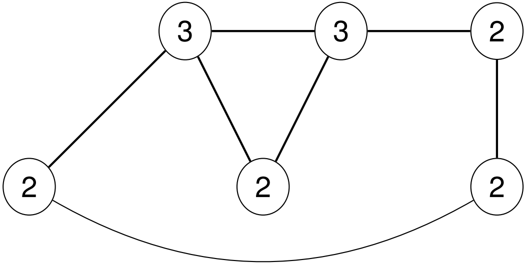
Definition 331 (Minimum degree of G)

$\delta(G) = \min\{deg_G(v) | v \in V\}$.

Definition 332 (Maximum degree of G)

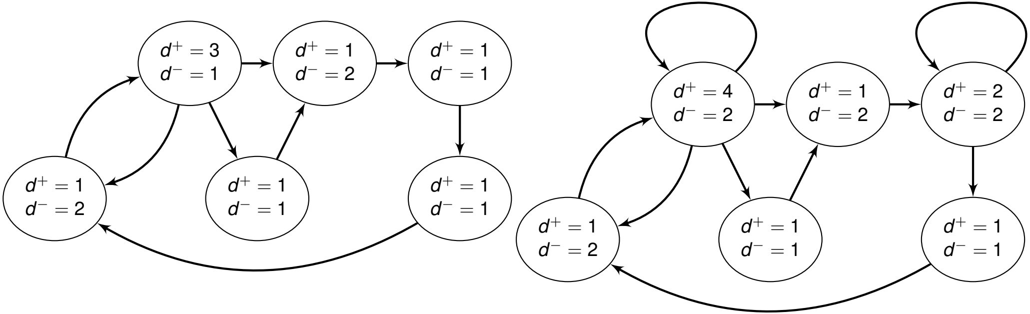
$\Delta(G) = \max\{deg_G(v) | v \in V\}$.

Degrees in an undirected graph



Here, vertices are labelled using the degree

Degrees in a directed graph



What to consider about degrees?

Degrees are often considered as a measure of popularity

Often write $k(i)$ (or k_i) for "degree of vertex i", $k^-(i)$ and $k^+(i)$ for in- and out-degree

- Minimum and maximum degree
- Minimum and maximum in/out-degree. E.g., if you consider the global air transportation network and the in/out-degree of airports, in-degree is a measure of a location's "popularity" as a travel destination
- Range of degrees in a graph: are there large discrepancies in connectivity between vertices in the graph?
- Average degree (often denoted $\langle k \rangle$ because of physicists)
- Average in/out-degree
- Variance of the degrees or in/out-degrees

- Average (nearest) neighbour degree, to encode for *preferential attachment* (one prefers to hang out with popular people)

$$k_i^{nn} = \frac{1}{k(i)} \sum_{j \in \mathcal{N}(i)} k(j)$$

or, in terms of the adjacency matrix $A = [a_{ij}]$,

$$k_i^{nn} = \frac{1}{k(i)} \sum_{j} a_{ij} k(j)$$

- *Excess degree*: take nearest neighbour degree but do not consider the edge/arc followed to get to the neighbour
- Degree, nearest neighbour and excess degree distributions

Degrees in `igraph`

- `degree` gives the degrees of the vertices
- `degree_distribution` gives numeric vector of the same length as the maximum degree plus one. The first element is the relative frequency zero degree vertices, the second vertices with degree one, etc.
- `knn` calculate the average nearest neighbor degree of the given vertices and the same quantity in the function of vertex degree
- `strength` sums up the edge weights of the adjacent edges for each vertex

Degree from adjacency matrix

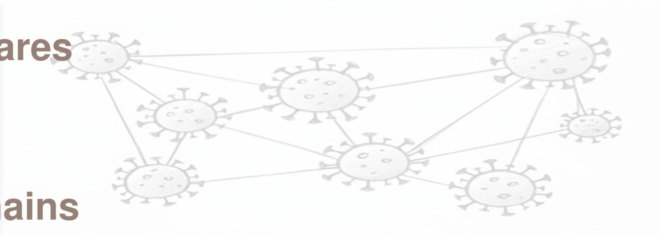
Suppose adjacency matrix take the form $A = [a_{ij}]$ with $a_{ij} = 1$ if there is an arc from the vertex indexed i to the vertex indexed j and 0 otherwise. (Could be the other way round, using A^T, just make sure)

Let $\boldsymbol{e} = (1, \ldots, 1)^T$ be the vector of all ones

$A\boldsymbol{e} = (d_G^+(1), \ldots, d_G^+(1))^T$ (out-degree)

$\boldsymbol{e}^T A = (d_G^-(1), \ldots, d_G^-(1))$ (in-degree)

Preliminary stuff

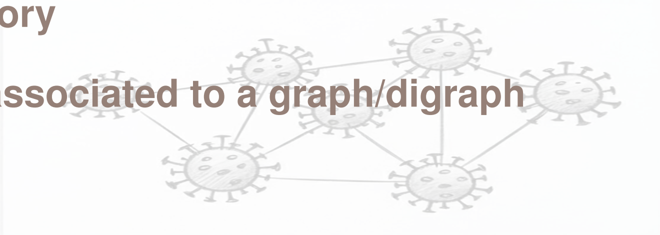


Least squares

The SVD

Markov chains

Graph theory



Matrices associated to a graph/digraph

Trees

Why characterise graphs?

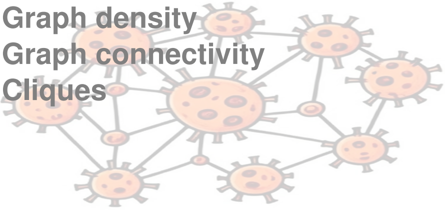
Measures at the graph level

Circumference & Girth

Graph density

Graph connectivity

Cliques

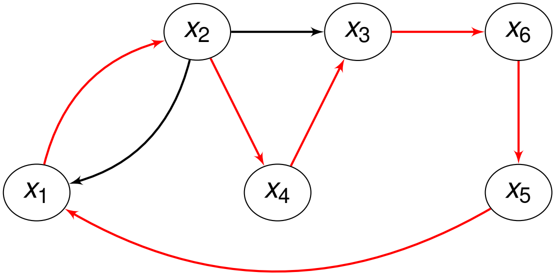


Circumference

Definition 333 (Circumference)

In an undirected (resp. directed) graph, the total number of edges (resp. arcs) in the longest cycle of graph G is the **circumference** of G

Circumference is 6.

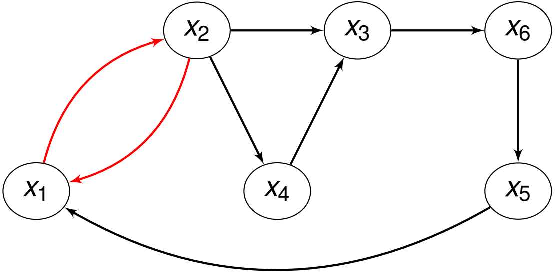


Girth

Definition 334 (Girth)

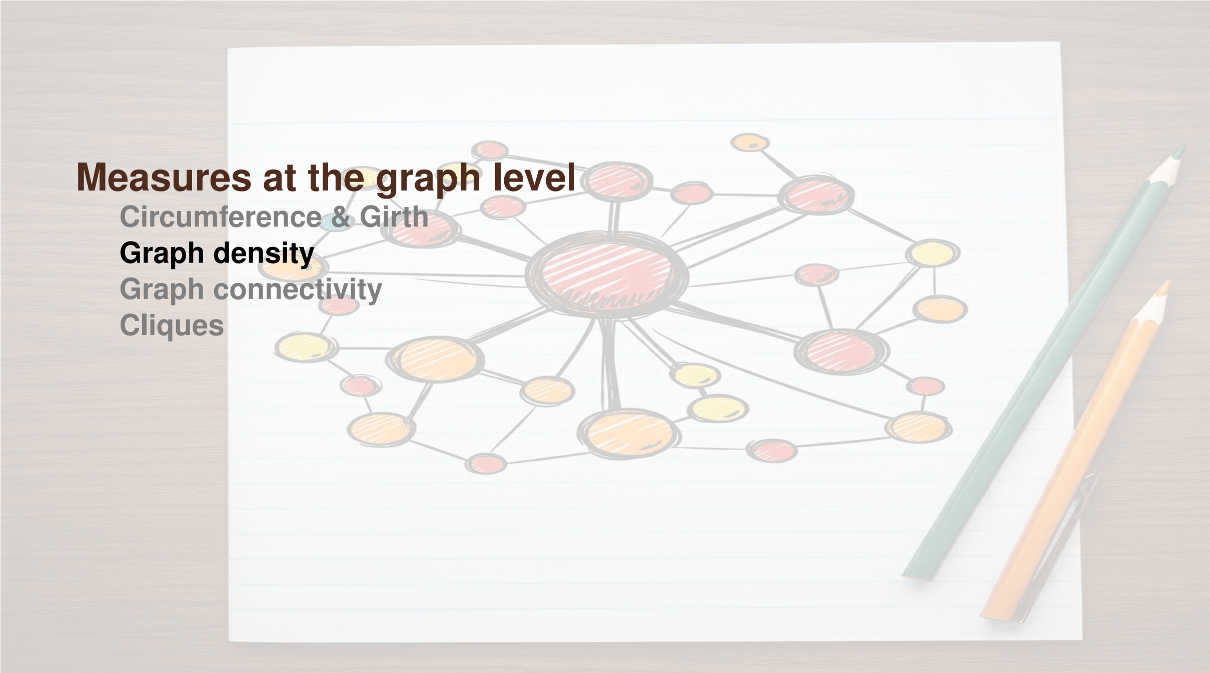
The total number of edges in the shortest cycle of graph G is the **girth** $g(G)$

Girth is 2.



Measures at the graph level

- Circumference & Girth
- **Graph density**
- Graph connectivity
- Cliques



Completeness

Definition 335 (Complete undirected graph)

An undirected graph is complete if every two of its vertices are adjacent.

Definition 336 (Complete digraph)

A digraph $D(V, A)$ is complete if $\forall u, v \in V$, $uv \in A$.

In case of simple graphs, completeness effectively means that “information” can be transmitted from every vertex to every other vertex quickly (1 step)

It can be useful to know how far away we are from being complete

Number of edges/arcs in a complete graph

$G = (V, E)$ undirected and simple of order n has at most

$$\frac{n(n-1)}{2}$$

edges, while $G = (V, A)$ directed and simple of order n has at most

$$n(n-1)$$

arcs

Density of a graph

Definition 337 (Density)

The fraction of maximum number of edges or arcs present in the graph is the **density** of the graph.

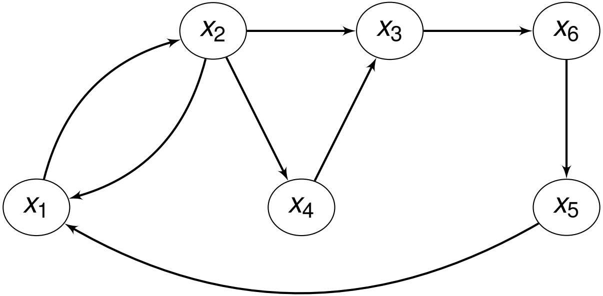
If the graph has p edges or arcs, then its density is, respectively,

$$\frac{2p}{n(n-1)}$$

or

$$\frac{p}{n(n-1)}$$

Example of density



Graph has order 6 and thus a max of 30 arcs. Here, 8 arcs $\implies$ density 0.267 (26.7% of arcs are present)

Measures at the graph level

- Circumference & Girth
- Graph density
- **Graph connectivity**
- Cliques

Connectedness

We have already seen connectedness (quasi- or strong in the oriented case)

Connectedness is important in terms of characteristing graph properties, as it shows the capacity of the graph to convey information to all the members of the graph (the vertices)

Definition 338 (Connected graph)

A **connected graph** is a graph that contains a chain $\mu[x, y]$ for each pair x, y of distinct vertices

Denote $x \equiv y$ the relation "$x = y$, or $x \neq y$ and there exists a chain in G connecting x and y". $\equiv$ is an equivalence relation since

1. $x \equiv y$ [reflexivity]
2. $x \equiv y \implies y \equiv x$ [symmetry]
3. $x \equiv y, y \equiv z \implies x \equiv z$ [transitivity]

Definition 339 (Connected component of a graph)

The classes of the equivalence relation $\equiv$ partition V into connected sub-graphs of G called **connected components**

Articulation set

Definition 340 (Articulation set)

For a connected graph, a set A of vertices is called an **articulation set** (or a **cutset**) if the subgraph of G generated by $V - A$ is not connected

`articulation_points(G)` in `igraph` (assumes the graph is undirected, makes it so if not)

Strongly connected graphs

$G = (V, U)$ connected. A **path of length 0** is any sequence $\{x\}$ consisting of a single vertex $x \in V$

For $x, y \in V$, let $x \equiv y$ be the relation "there is a path $\mu_1[x, y]$ from x to y as well as a path $\mu_2[y, x]$ from y to x". This is an equivalence relation (it is reflexive, symmetric and transitive)

Definition 341 (Strong components)

Sets of the form

$$A(x_0) = \{x : x \in V, x \equiv x_0\}$$

are equivalence classes; they partition V and are the **strongly connected components** of G

Definition 342 (Strongly connected graph)

G **strongly connected** if it has a single strong component

Definition 343 (Minimally connected graph)

G is **minimally connected** if it is strongly connected and removal of any arc destroys strong-connectedness

Definition 344 (Contraction)

$G = (V, U)$. The **contraction** of the set $A \subset V$ of vertices consists in replacing A by a single vertex a and replacing each arc into (resp. out of) A by an arc with same index into (resp. out of) a

Quasi-strong connectedness

Definition 345 (Quasi-strong connectedness)

G **quasi-strongly connected** if $\forall x, y \in V$, exists $z \in V$ (denoted $z(x, y)$ to emphasize dependence on x, y) from which there is a path to x and a path to y

Strongly connected $\implies$ quasi-strongly connected (take $z(x, y) = x$); converse not true

Quasi-strongly connected $\implies$ connected

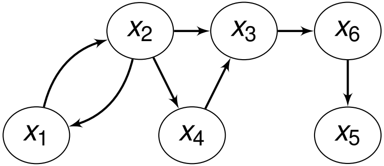
Lemma 346

$G = (V, U)$ has a root $\iff$ G quasi-strongly connected

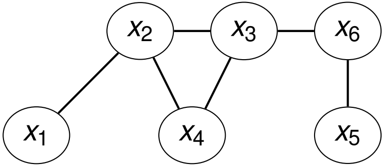
Weak-connectedness

Definition 347 (Weakly connected graph)

$G = (V, U)$ **weakly connected** if $G = (V, E)$ connected, where E is obtained from U by ignoring the direction of arcs



$\Longrightarrow$



Weak components

Define for $x, y \in V$ the relation $x \equiv y$ as "$x = y$ or $x \neq y$ and there is a chain in G connecting x and y" [like for components in an undirected graph, except the graph is directed here]

This defines an equivalence relation

Definition 348 (Weak components)

Sets of the form

$$A(x_0) = \{x : x \in V, x \equiv x_0\}$$

are equivalence classes partitioning V into the **weakly connected components** of G

$G = (V, U)$ is weakly connected if there is a single weak component

Components in `igraph`

- `is_connected` decides whether the graph is weakly or strongly connected
- `components` finds the maximal (weakly or strongly) connected components of a graph
- `count_components` does almost the same as components but returns only the number of clusters found instead of returning the actual clusters
- `component_distribution` creates a histogram for the maximal connected component sizes
- `decompose` creates a separate graph for each component of a graph
- `subcomponent` finds all vertices reachable from a given vertex, or the opposite: all vertices from which a given vertex is reachable via a directed path

Measures at the graph level

- Circumference & Girth
- Graph density
- Graph connectivity
- **Cliques**

Cliques

Definition 349 (Clique in undirected graphs)

$G = (V, E)$ a simple undirected graph. A **clique** is a subgraph G' of G such that all vertices in G' are adjacent

Definition 350 (n-clique)

A simple, complete graph on n vertices is called an n-**clique** and is often denoted K_n

Definition 351 (Clique in directed graphs)

$G = (V, U)$ a simple directed graph. A **clique** is a subgraph G' of G such that all vertices in G' are mutually adjacent

Definition 352 (Maximal clique)

A **maximal clique** is a clique that cannot be extended by adding another adjacent vertex

k-core

Definition 353 (k-core of a graph)

$G = (V, U)$ a graph. The **k-core** of G is a maximal subgraph in which each vertex has degree at least k

Definition 354 (Coreness of a vertex)

$G = (V, U)$ a graph, $x \in V$. The **coreness** of x is k if x belongs to the k-core of G but not to the $k + 1$ core of G

For directed graphs, in-cores or out-cores depending on whether in-degree or out-degree is used

Linearly separable points

Let X_1 and X_2 be two sets of points in $\mathbb{R}^p$

Then X_1 and X_2 are **linearly separable** if there exist $w_1, w_2, .., w_p, k \in \mathbb{R}$ such that

- every point $x \in X_1$ satisfies $\sum_{i=1}^{p} w_i x_i > k$
- every point $x \in X_2$ satisfies $\sum_{i=1}^{p} w_i x_i < k$

where x_i is the ith component of x