# Lecture 2

September 2, 2025

# Matrices

$$\mathbf{A} = \begin{vmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{vmatrix}$$

- *m by n*
  - m is the number of row, n is the number of columns

# Why matrices?

- Represent systems of equations

$$X_1 + 3X_2 = 8$$
$$X_1 + 2X_2 = 6$$
$$X_2 = 2$$

# Why matrices?

- Represent systems of equations

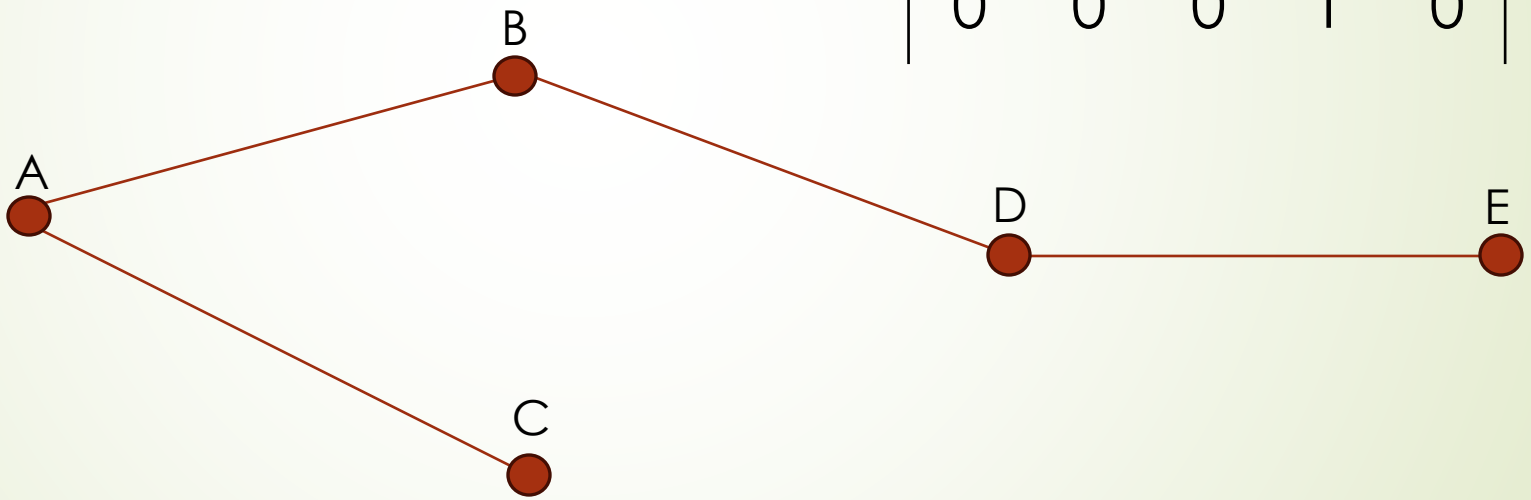
$$X_1 + 3X_2 = 8$$
$$X_1 + 2X_2 = 6$$
$$X_2 = 2$$

$$\left|\begin{array}{ccc} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{array}\right|$$

# Why matrices?

- Represent Graphs

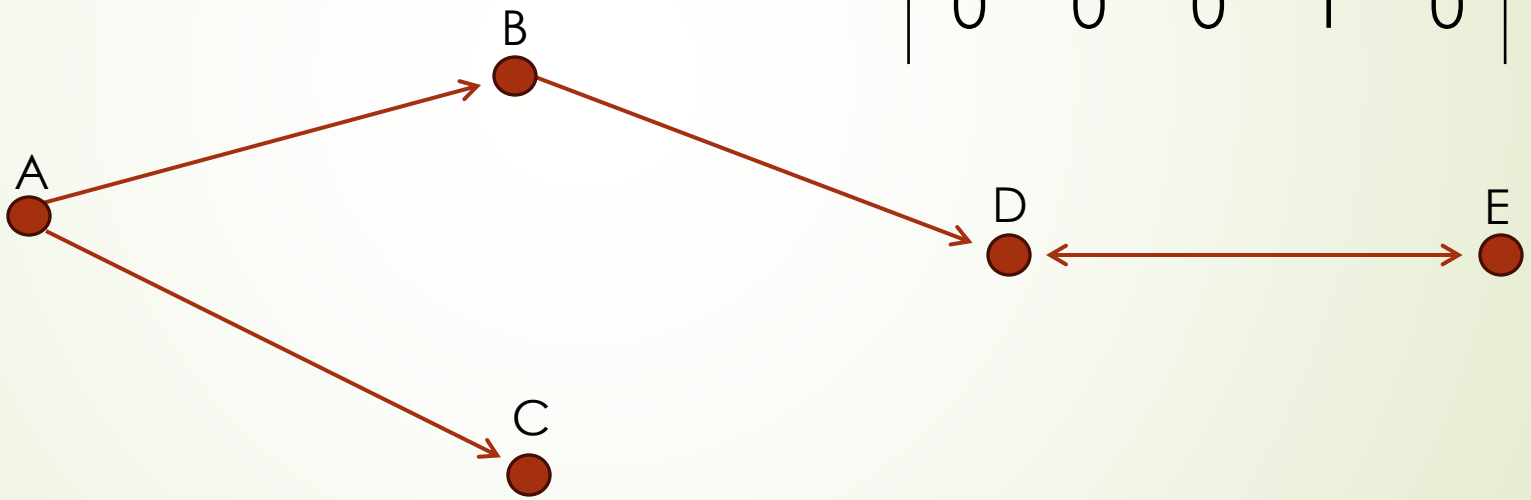
$$\begin{vmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{vmatrix}$$



# Why matrices?

- Represent Graphs

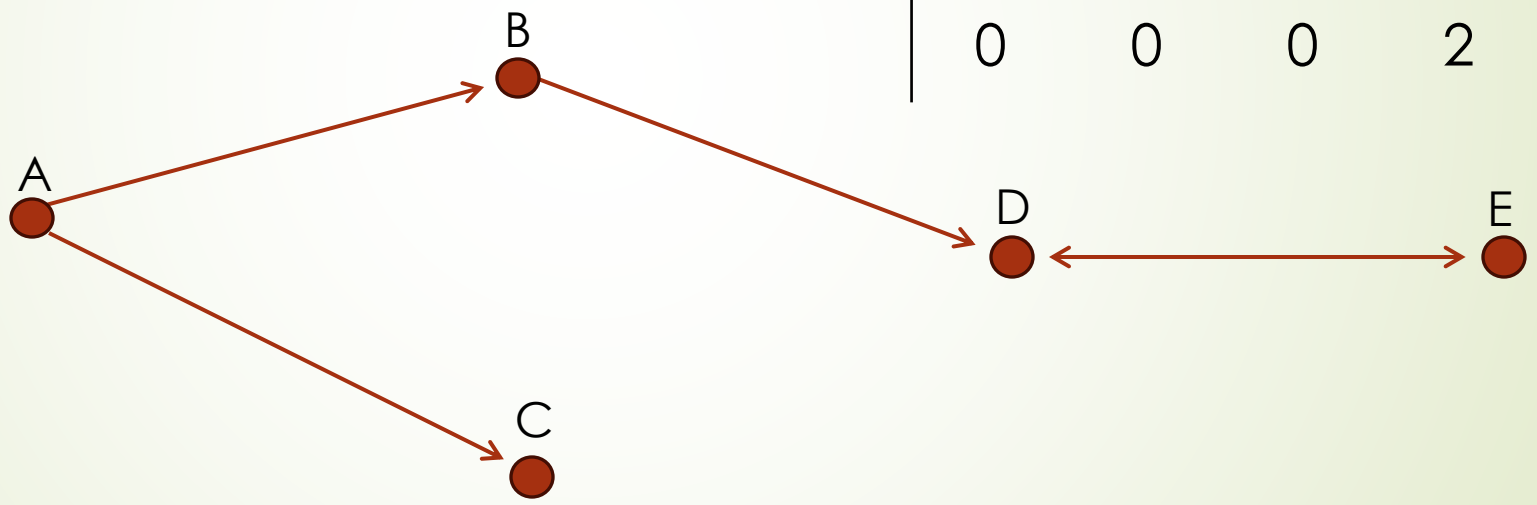
| | | | | |
|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 0 |



# Why matrices?

- Represent Graphs

| | | | | |
|---|---|---|---|---|
| 0 | 1.5 | 5 | 0 | 0 |
| 0 | 0 | 0 | 0.25 | 0 |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 2 |
| 0 | 0 | 0 | 2 | 0 |



# Use of matrices in biology

- Gene regulatory networks
- Metabolic networks
- Infectious disease models
- Survival Analysis

# Genomics

- Numerous genomes being sequenced
- Genomes filled junk DNA?
  - Genes
  - Regulatory factors
  - Repetitive elements
- Genes
  - Regulatory elements
  - Exons & Introns
  - Untranscribed elements

# Genomics

- Variations within the genome
  - Large deletions
  - Large duplications
  - Translocations
  - Single Nucleotide Variations
    - Insertions
    - Deletions
    - Polymorphisms

# What can we do with this information?

- Population/Ancestry structure
- GWAS studies
- Variant effects
  - Synonymous vs Non-Synonymous
  - Regulatory disruption

# Variant Call Format (VCF)

- File format containing sequence information
- Tab delimited file
- Header region
  - Version info
  - Chromosome/Scaffold/Contig info
  - Details on content

# Variant Call Format (VCF)

- Can contain info on one or more samples
- General position info
  - Chromosome, position, Reference, Alternative(s)
  - Gene info
  - Total counts – Individual allele and overall
- Sample specific info
  - Allele calls, site depth, number of counts per allele, quality
- Can contain all sites or those with just variants

# Group Project

- 3 Groups
- Form a testable hypothesis
  - Population analysis
  - Disease
- Compile Genetic Information
- Determine software necessary
  - Is it available or are you creating it?
- Preliminary presentation and report
- Run analysis on Biolinux machines
- Final Write-up and Presentation