VECTOR SPACE ->
** There exists a zero vector 0 in V such that u + 0 = u for any u in Vector space.
** 1*v=v for any v in Vector space.
The null space of a matrix A is a vector space.

VECTOR SUBSPACE ->
A vector subspace W is a subset of a vector space V such that the zero vector 0 in V is also in W.
If we have a linear plane through the origin (0, 0, 0) in the 3-
dimensional space, then the set of all points on the linear plane becomes a
vector subspace.

*****Suppose we have a system of linear equations such that $A \cdot x = b$.
Then the system of linear equations has a unique solution if and only if Null(A) = {0}.

***Ax=b
if the right hand side vector b is in the column space, then the system of linear equations $A \cdot x = b$ is feasible,
i.e., there exists a solution.

*** Eigen values of a triangular matrix are the entries on the diagonal of the matrix.
** Suppose we have an n×n matrix A. A is invertible if and only if 0 is not an eigen value of A.

** Suppose v1, v2, . . . , vr are eigen vectors of a symmetric matrix A associated with the distinct eigen values λ1, λ2, . . . , λr of A. Then v1, v2, . . . , vr are linearly independent and they are orthogonal to each other.

DIAGONALIZATION ->
** Suppose A is an n×n matrix. Then A is diagonalizable if and only if A has n linearly independent eigen vectors. or A has n many distinct eigen values.

POLYHEDRON ->
A convex set is a subset S in R^d such that for any points x, y in S
$\alpha \cdot x + (1 - \alpha) \cdot y$ is also in S for all $0 \le \alpha \le 1$. A **convex hull** of a set {x1, x2, . . . , xk} is the smallest convex hull in R^d containing {x1, x2, . . . , xk}.

A polytope is a convex hull of a set of finitely many vertices in R^d.
A bounded polyhedron in R^d is a polytope.

If we define a polytope as a system of linear equations and inequalities, then we call this a **hyperplane representation of the polytope**. If we define a polytope by a convex hull of a set of finitely many vertices, then we call this a **vertex representation** of the polytope.

An optimal solution for a linear programming problem is always a vertex of the polyhedron.

###---Network
In this case, in order to visualise clearly, we take a subset of the collaboration network using the induced **subgraph()** function from the **igraph** package.

<u>Graph Theory</u>->
Suppose we have a finite set of vertices (or nodes) V = {1, 2, . . . , n}. Each vertex in V is labelled. If we want to pair vertices u, v in V , then we draw an edge between them. This is called an **undirected edge**. An undirected graph G is an object which consists of a **vertex set** V and a set of **undirected edges** E. We denote an undirected graph as G = (V, E).
Adjacency Matrix - symmetric

Suppose we have a finite set of vertices (or nodes) V = {1, 2, . . . , n}. Each vertex in V is labelled. If there is a relation from a vertex v in V to a vertex u in V , then we draw an arrow. This arrow is called a **directed edge**. A directed graph N is an object which consists of a vertex set V and a set of directed edges E. We denote a directed graph as N = (V, E). Often a **directed graph** is also called a **network**. In this chapter, We call a directed graph a network.
Adjacency Matrix - Not symmetric

The **degree** of a vertex v in V **for a graph or a network** is the number of edges adjacent to the vertex v. The **degree matrix** of a graph or a network where di is the degree of a vertex i in V with a vertex set V = {1, 2, . . . n} is an n × n diagonal matrix D.

The **Laplacian matrix** of a **graph or a network** with the vertex set $V = \{1, 2, . . . , n\}$ is an $n \times n$ matrix L such that **L = D − A**,
where **D is the degree matrix** of the graph or the network and **A is the adjacency** matrix of the graph or the network.
-> The **Laplacian matrix L of an **undirected graph G** is **symmetric**.
-> The **dimension of the null space of the Laplacian matrix L of an **undirected graph G** is the **number of components** in the graph G.

The **transition matrix** for an **undirected graph** G = (V, E)
with the vertex set $V = \{1, 2, . . . n\}$ is an $n \times n$ matrix such that where **pij = aij/di**
with aij is the (i, j)th element of the adjacency matrix of the graph G and di
is the degree of the vertex i in the graph G.

The **normalised Laplacian matrix** for an undirected graph G = (V, E) with the vertex set $V = \{1, 2, . . . n\}$ is an $n \times n$ matrix **Lp** such that **Identity− P** where **P is the transition matrix of the graph G.**

** Suppose λ0, . . . , λn−1 are distinct eigen values of the normalised Laplacian matrix LP of an **undirected graph** G = (V, E) with the vertex set $V = \{1, 2, . . . , n\}$ such that λ0 ≤ λ1 ≤ . . . ≤ λn−1. Then **sum of all these eigen values is <= n.**
and **the equality holds if and only if there is no isolated vertices.**

#####----Probability_Distributions

Given that the person has the disease, the conditional probability **P(+ | D)** that the diagnostic test is positive is called **sensitivity [positive predictive value]**. Given that the person does not have the disease, the conditional probability P(−| Dc) that the diagnostic test is negative is called the **specificity**.

E[(Y − μ)^3] describes **skewness**.
E[(Y − μ)^3]/(sigma^3) -> **skewness Coefficient.**
Jensen’s inequality states that for convex functions, **E[g(Y)] ≥ g[E(Y)].**

pbinom -> probability distribution function
dbinom -> Density
qbinom -> Quantile
rbinom -> random numbers

POISSON DISTRIBUTION→>

$$f(x) = \frac{\lambda^x}{x!}e^{-\lambda}$$

Mean = **μ.**
Variance = μ.
The Poisson distribution is **unimodal**. Its **mode** equals the **integer part of μ**.
SkewNess Coefficient = **1/(sqrt(μ)).**

When n is large, p is small **Binomial(n,p)** -> **Poisson(np).**

GAMMA DISTRIBUTION –>

Gamma distribution
The ***gamma distribution*** is characterized by the probability density function

$$f(y; k, \lambda) = \frac{\lambda^k}{\Gamma(k)} e^{-\lambda y} y^{k-1}, \ y \geq 0; \quad f(y; \lambda, k) = 0, \ y < 0, \tag{2.10}$$

for parameters $k > 0$ and $\lambda > 0$; k is a *shape parameter* and $1/\lambda$ is a *scale parameter*.

The gamma distribution has mean and standard deviation

$$\mu = k/\lambda, \quad \sigma = \sqrt{k}/\lambda, \tag{2.11}$$

Skewness parameter -> **2/sqrt(k)**.
K -> **shape parameter.**
1/λ -> **scale parameter.**

##########
The law of Iterated Expectation, E(Y) = E[E(Y|X)]

Covariance

The ***covariance*** between random variables X and Y having $E(X) = \mu_X$ and $E(Y) = \mu_Y$ is

$$\operatorname{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

Correlation

The ***correlation*** between a random variable X having $E(X) = \mu_X$ and $\operatorname{var}(X) = \sigma_X^2$ and a random variable Y having $E(Y) = \mu_Y$ and $\operatorname{var}(Y) = \sigma_Y^2$ is

$$\operatorname{corr}(X, Y) = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right] = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}. \tag{2.16}$$

$$E(Y \mid X = x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

Conditional distributions have higher moments as well as means. The variance of a conditional distribution is the expected squared distance from the conditional mean. The normal conditional distribution of Y given $X = x$ has variance

$$\begin{aligned}
\sigma_{Y|X}^2 &= E[Y - E(Y \mid X)]^2 = E\{Y - [\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)]\}^2 = E\{(Y - \mu_Y) - [\rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)]\}^2 \\
&= E(Y - \mu_Y)^2 + \left(\rho\frac{\sigma_Y}{\sigma_X}\right)^2 E(X - \mu_X)^2 - 2\left(\rho\frac{\sigma_Y}{\sigma_X}\right)E(X - \mu_X)(Y - \mu_Y) \\
&= \sigma_Y^2 + \left(\rho\frac{\sigma_Y}{\sigma_X}\right)^2 \sigma_X^2 - 2\left(\rho\frac{\sigma_Y}{\sigma_X}\right)\rho\sigma_X\sigma_Y = \sigma_Y^2(1 - \rho^2),
\end{aligned}$$

using that $E(X - \mu_X)(Y - \mu_Y) = \operatorname{cov}(X, Y) = \rho\sigma_X\sigma_Y$. The variance $\sigma_Y^2(1-\rho^2)$ of the conditional distribution of Y given X is smaller than the variance σ_Y^2 of the marginal distribution of Y. The larger the value of $|\rho|$, the smaller the conditional variance. As $|\rho|$ approaches 1, the conditional variance approaches 0 and the joint bivariate normal distribution falls more tightly along a straight line.

The proportion of observations that fall at least k standard deviations from the mean can be no greater than 1/(k^2). The result is called **Chebyshev's inequality**.
For example, no more than 4% of the observations can fall at least five standard deviations from the mean.

1. If $\{Y_i\}$ are independent normal random variables, with $Y_i \sim N(\mu_i, \sigma_i^2)$, then $\sum_{i=1}^n Y_i \sim N(\sum_{i=1}^n \mu_i,\ \sum_{i=1}^n \sigma_i^2)$.

   The sum is also normal, with means and variances also adding.

2. If $\{Y_i\}$ are independent binomial random variables, with $Y_i \sim \text{binom}(n_i, \pi)$, then $\sum_{i=1}^n Y_i \sim \text{binom}(\sum_{i=1}^n n_i,\ \pi)$.

   The sum is also binomial, with number of trials adding likewise, if the success probability π is the same for each binomial.

3. If $\{Y_i\}$ are independent Poisson random variables, with $Y_i \sim \text{Pois}(\mu_i)$, then $\sum_{i=1}^n Y_i \sim \text{Pois}(\sum_{i=1}^n \mu_i)$.

   The sum is also Poisson, with means also adding.