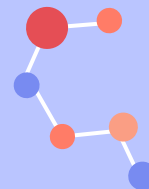


# Protein- Protein Interactions

Ashar Seif Al-Nasr  
Mariam Ashraf Mohamed  
Nourhan Yehia



# TABLE OF CONTENTS



## Data Description

Explore the data structure and content

01

03

## Shortest-Path Analysis

Analyzing the shortest acyclic paths between proteins in the network.

## Graph Construction

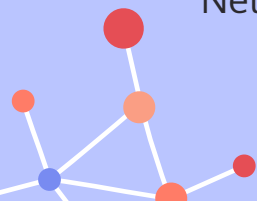
Construct the PPIs biological graph using NetworkX

02

04

## Connectivity Analysis

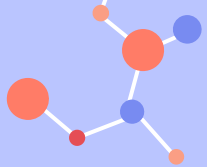
Analyzing the connectivity of the network and mapping protein UniProt IDs to gene names



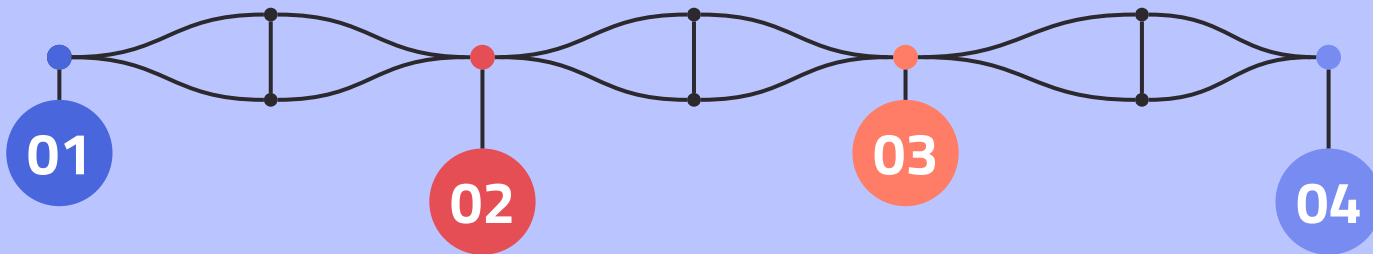


01

# Data Description



# Dataset Description



## Number of Samples

612515 PPI with interaction confidence

## Data Preprocessing

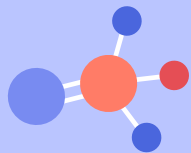
No Duplicates  
No Missing Values

## Tail And Head Nodes

Unique Tail Nodes :17081  
Unique Head Nodes :17095

## Bidirectional Interactors

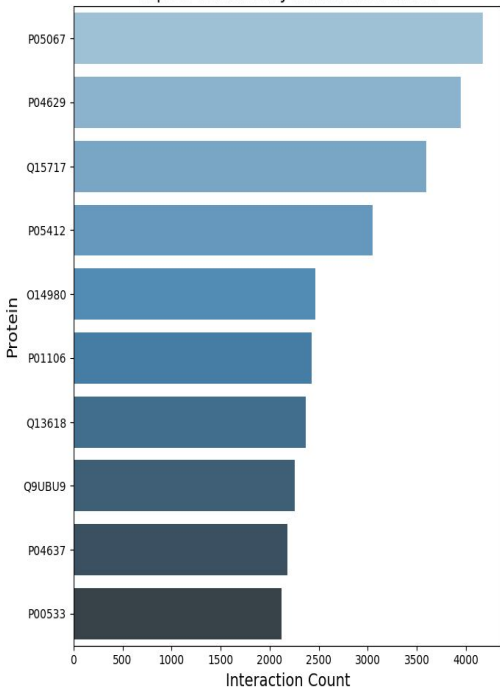
Number of Shared Nodes: 17008



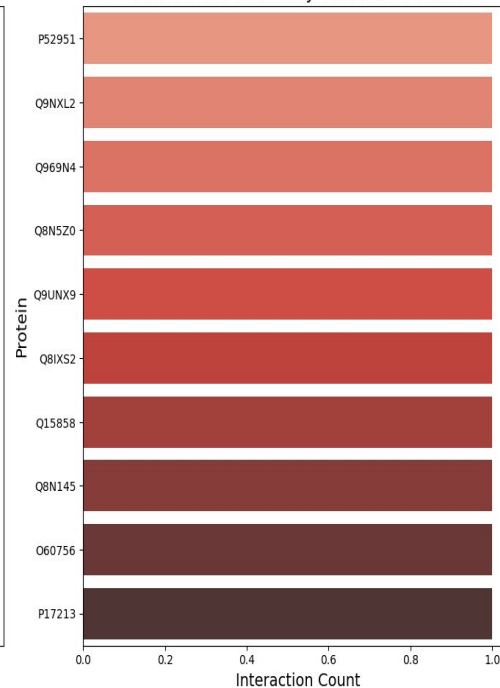
# Dataset Description



Top 10 Proteins by Interaction Count



Bottom 10 Proteins by Interaction Count



Tail Only

73

Example : P0CJ77

Functions as an **upstream signal initiator** by binding to interleukins. It is part of **receptor complexes** rather than a signaling intermediate or downstream protein

Head Only

87

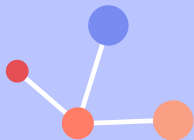
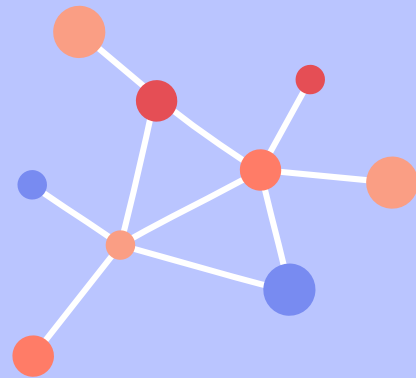
Example : Q8TBF5

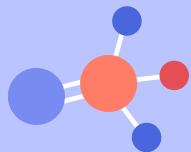
STAT6 is a **transcription factor** that plays a crucial role in mediating signals from cytokine receptors. STAT6 is a terminal effector in the **JAK-STAT signaling pathway** and does not initiate the signaling cascade



02

# Graph Construction

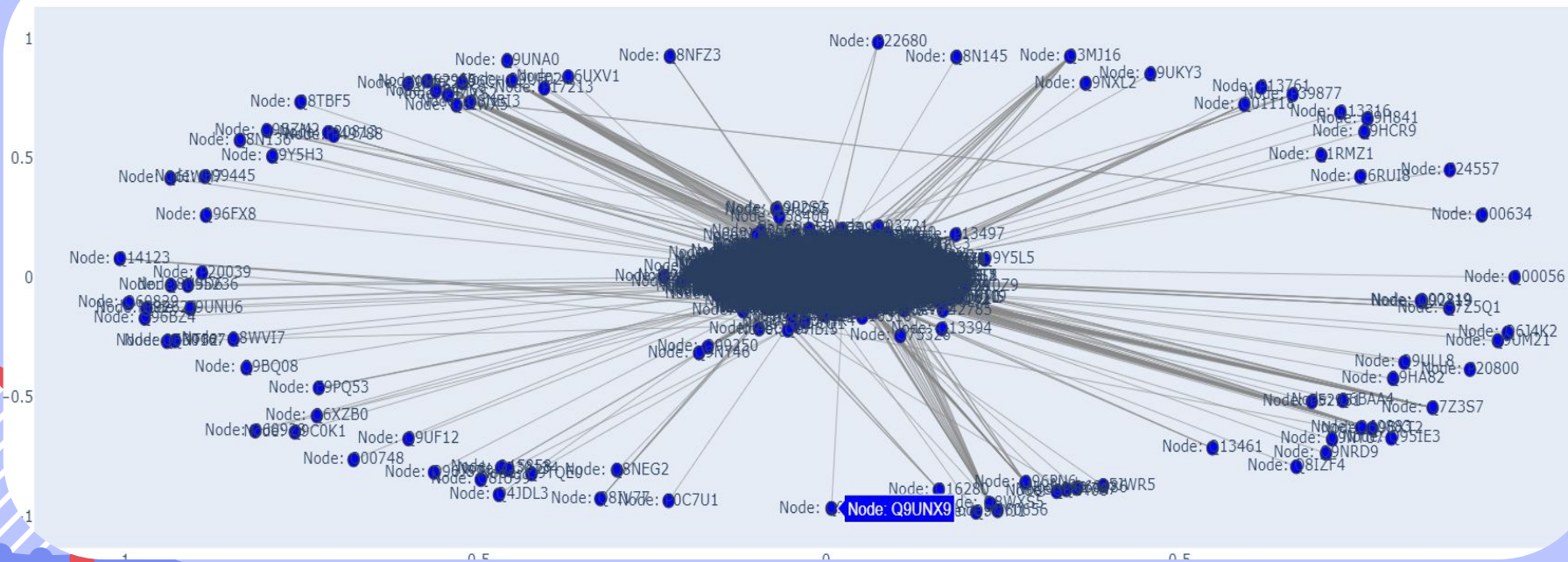


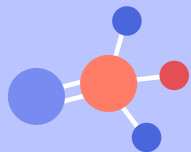


# Graph Construction



Protein-Protein Interaction Network





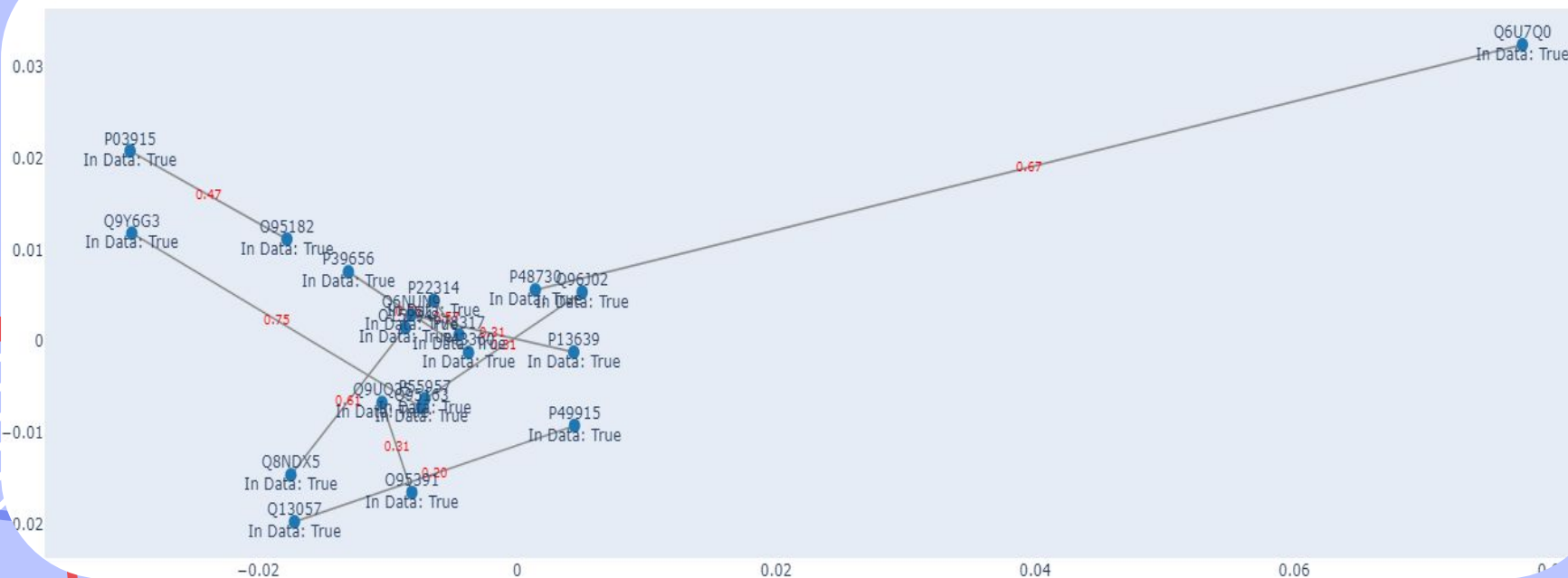
# Graph Construction



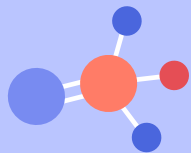
Number of Nodes in the Graph: **17168**

Number of Edges in the Graph: **612515**

Sample Edges Visualization with Edge Weights







# Graph Construction



Sampled Edges from Graph:

	Source	Target	Edge_Weight
0	P13639	Q6NUN9	0.311133
1	P48730	Q6U7Q0	0.674374
2	P78317	P22314	0.573519
3	Q9Y6G3	O95163	0.750000
4	Q9UQ35	O95391	0.311133
5	P49915	Q13057	0.201461
6	Q8NDX5	O15294	0.608042
7	P39656	P43360	0.750000
8	O95182	P03915	0.473922
9	P55957	Q96J02	0.311133

Matched Edges in Dataset:

	Source	Target	Edge_Weight	Edge_Type
0	P48730	Q6U7Q0	0.674374	Phosphosite
1	P55957	Q96J02	0.311133	MI:0004 (affinity chromatography technology)
2	Q9Y6G3	O95163	0.750000	MI:0018 (two hybrid) MI:0398 (two hybrid pooli...
3	O95182	P03915	0.473922	SPIKE
4	Q8NDX5	O15294	0.608042	MI:0018 (two hybrid) MI:1356 (validated two hy...
5	P39656	P43360	0.750000	MI:0006 (anti bait coimmunoprecipitation) MI:0...
6	Q9UQ35	O95391	0.311133	MI:0004 (affinity chromatography technology)
7	P13639	Q6NUN9	0.311133	MI:0004 (affinity chromatography technology)
8	P49915	Q13057	0.201461	MI:0401 (biochemical)
9	P78317	P22314	0.573519	MI:0997 (ubiquitinase assay)

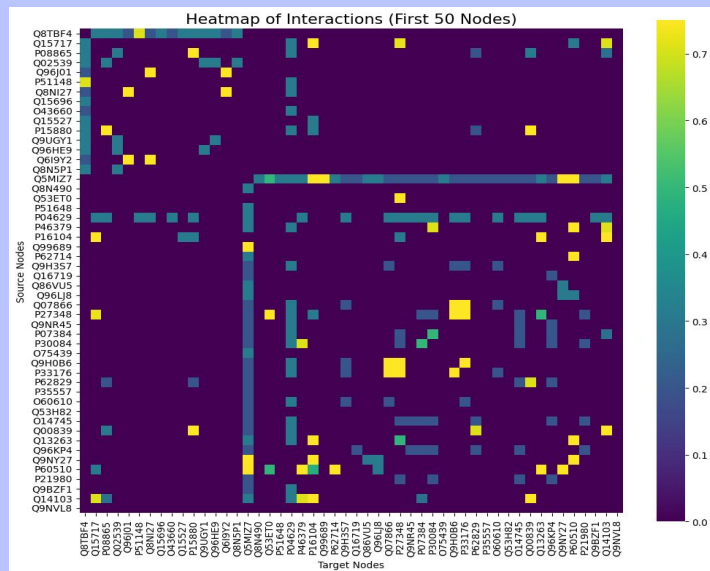
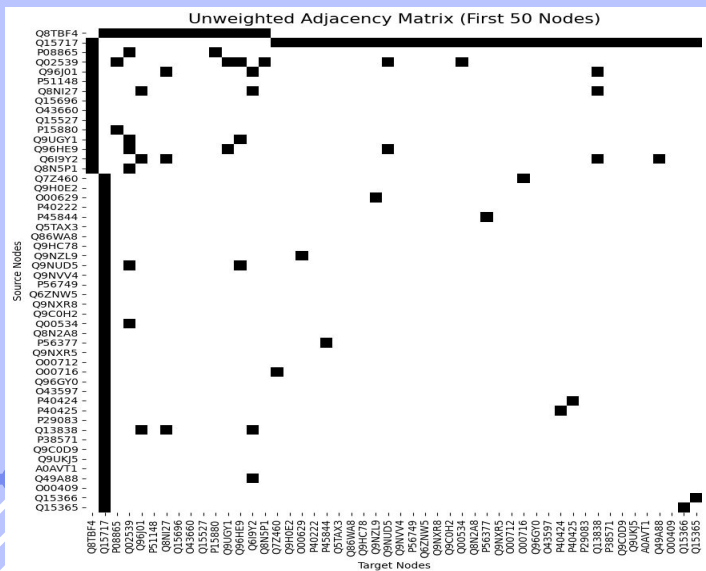
# Unweighted Graph And Adjacency Matrix

What is an Adjacency Matrix?

An **adjacency matrix** is a square matrix used to represent a graph. The rows and columns correspond to nodes, and the entries in the matrix indicate the presence (or absence) of edges between nodes:

$A[i][j]=1$  if there is an edge between node  $i$  and node  $j$ .

$A[i][j]=0$  otherwise.



03

# Shortest-Path Analysis



# Weight-Cost Transformation

Formula:

$$\text{cost} = -\log(\text{confidence} + \epsilon)$$

Path Selection Strategy:

- Apply shortest-path algorithms (e.g., Dijkstra's algorithm) to minimize the total sum of costs across the path.
- Since logarithms convert multiplication into addition, the total cost of a path is:

$$\sum_{i=1}^n -\log(\text{confidence}_i) = -\log \left( \prod_{i=1}^n \text{confidence}_i \right)$$

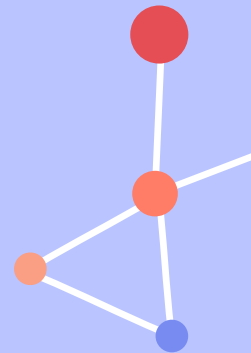
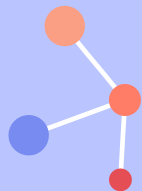
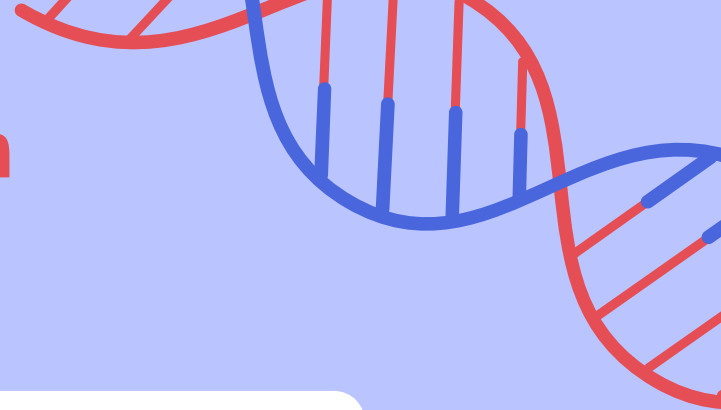
- Minimizing this sum is equivalent to **maximizing the product of confidences** across the path.

# Weight-Cost Transformation

$$\text{cost} = \frac{1}{\text{confidence} + \epsilon}$$

$$\text{cost} = 1 - \text{confidence}$$

$$\text{cost} = e^{-\alpha \cdot \text{confidence}}$$

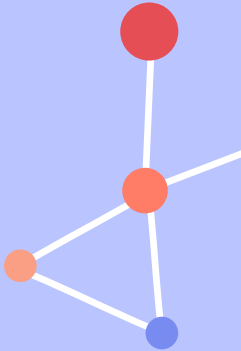
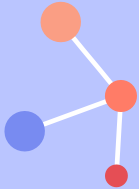


# Shortest-Path **Analysis**

**Method:** Dijkstra's algorithm for shortest paths

**Steps:**

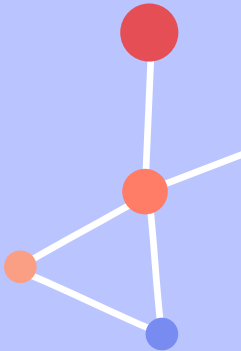
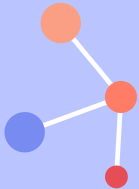
- ★ Convert interaction probabilities to cost values
- ★ Compute shortest paths
- ★ Extract path details including total score and costs



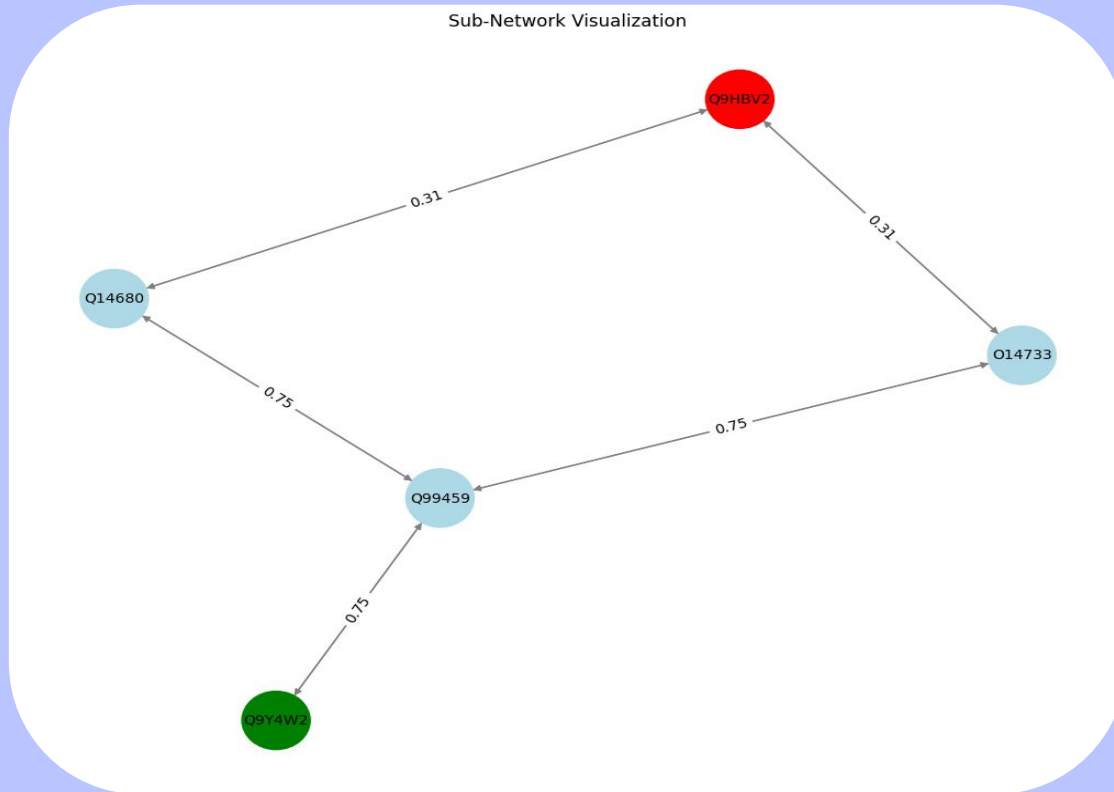
# Saving Path Data

```
Path: Q9HBV2 -> O14733 -> Q99459 -> Q9Y4W2  
Total Score: 0.1750123125  
Confidence: [0.311133, 0.75, 0.75]
```

```
Path: Q9HBV2 -> Q14680 -> Q99459 -> Q9Y4W2  
Total Score: 0.1750123125  
Confidence: [0.311133, 0.75, 0.75]
```



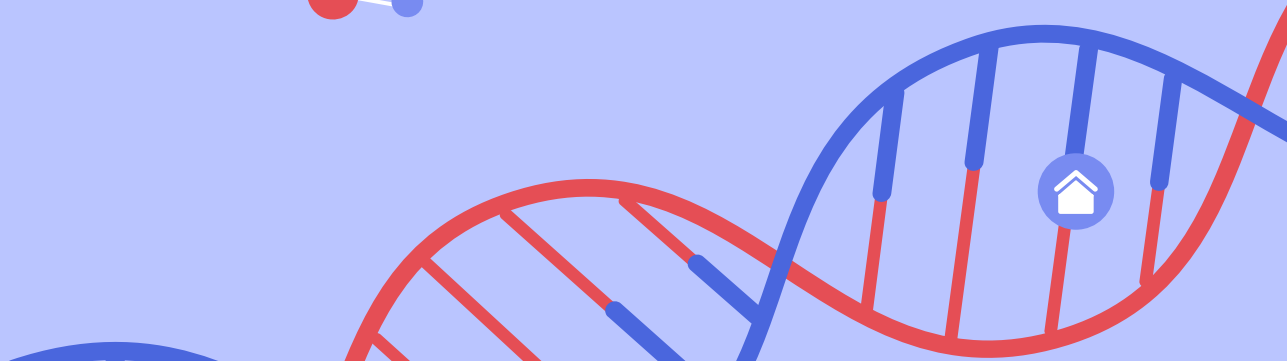
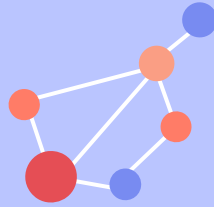
# Sub-Network Visualization

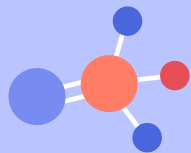




# Connectivity Analysis

04

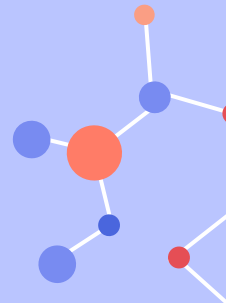
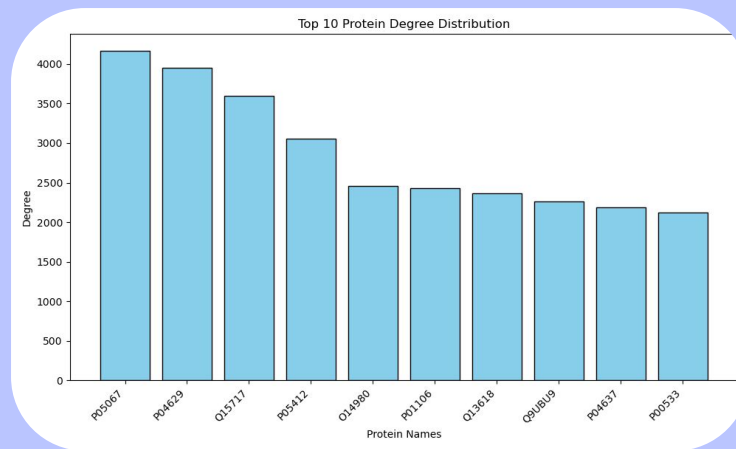
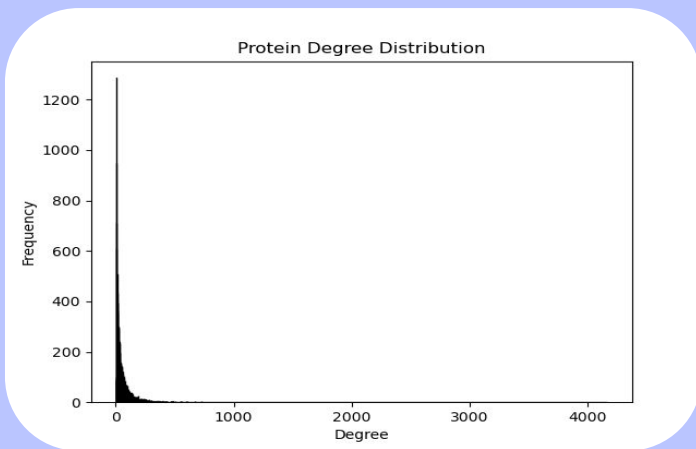


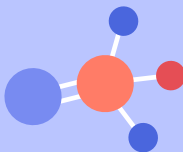


# Proteins Degree

What is the proteins degree?

The **degree** of a protein in a PPI network represents the number of other proteins it directly interacts with (direct relationships)



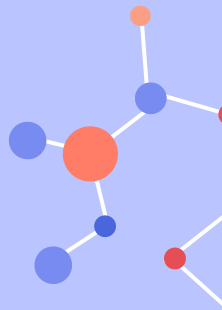


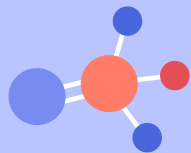
# Degree Ranking & Directly Connected Proteins



P05067	4170
P04629	3950
Q15717	3597
P05412	3052
O14980	2462
P01106	2429
Q13618	2367
Q9UBU9	2259
P04637	2183
P00533	2123
Q9HCE1	1981
P62993	1916
P49736	1817
P07900	1692
P03372	1663
Q92905	1621
P02751	1515
Q99496	1462
P24941	1441

Protein: P04629	
Degree: 3950	
Connected Proteins:	
Q5MIZ7	Weight: 0.311133
Q7Z392	Weight: 0.311133
Q96N67	Weight: 0.311133
P40227	Weight: 0.311133
P40222	Weight: 0.311133
P20936	Weight: 0.311133
P33527	Weight: 0.311133
A3KN83	Weight: 0.311133
Q9NQT8	Weight: 0.311133
Q9NZL4	Weight: 0.311133
P16435	Weight: 0.311133
Q13724	Weight: 0.311133
Q9H4M9	Weight: 0.311133
Q7Z460	Weight: 0.311133
Q01974	Weight: 0.311133
Q9BZQ6	Weight: 0.311133
Q7Z4K8	Weight: 0.311133
Q00535	Weight: 0.311133
Q00534	Weight: 0.311133





# Proteins Uniprot ID



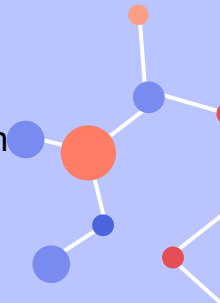
## Universal Protein Resource Identifier

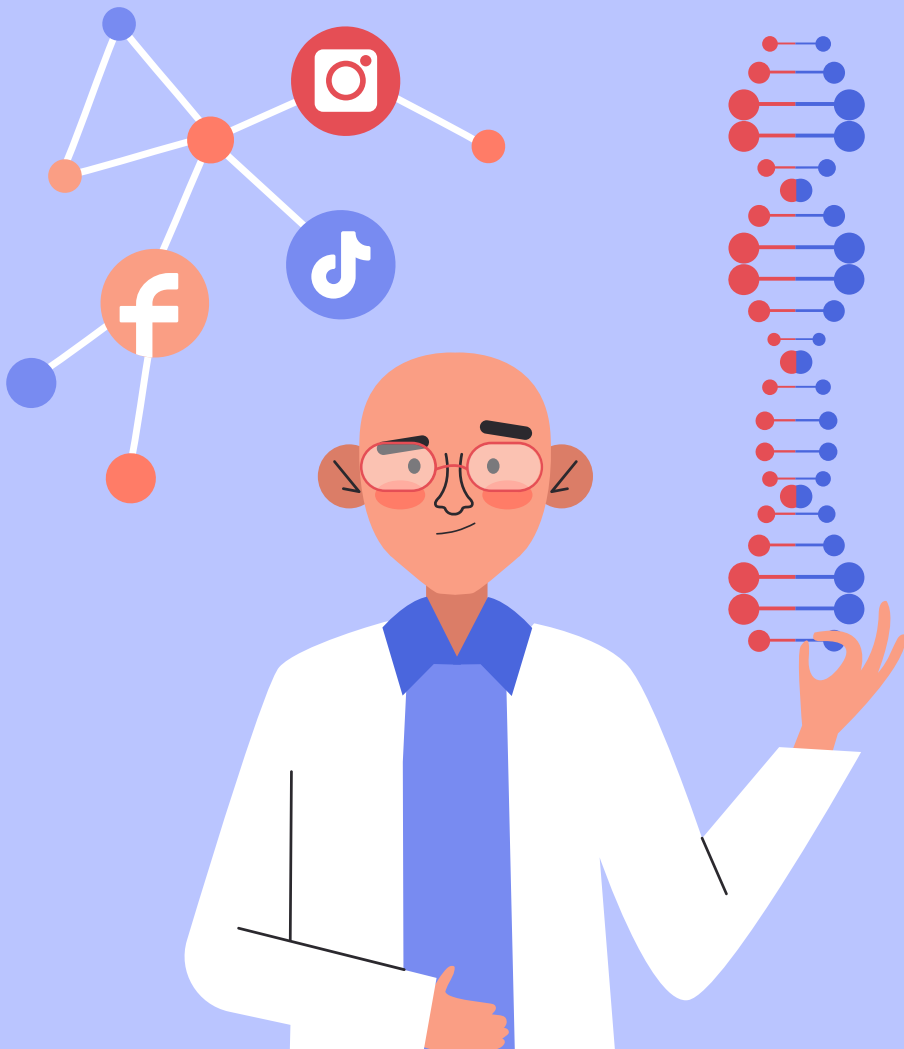
UniProt database provides a standardized way to refer to proteins across different species and studies.

Q8NGE1	OR6C4
Q14943	KIR3DS1
Q8N6M6	AOPEP
Q9Y5P0	OR51B4
Q6IN84	MRM1
Q13607	OR2F1
P30556	AGTR1
Q9UNP9	PP1E
P28749	RBL1
Q9BRQ0	PYGO2
P48730	CSNK1D
Q9UQC9	CLCA2
Q96CQ1	SLC25A36
P23945	FSHR
Q5T7M9	DIPK1A
Q9BV23	ABHD6
Q9HCR9	PDE11A
P08949	NMB
Q8IYI6	EXOC8
P14735	IDE

**Format:** A UniProt ID typically consists of 6-10 alphanumeric characters, such as **P04629** (for TP53 in humans).

**UniProt Entry Name:** Proteins also have descriptive entry names like **TP53\_HUMAN**, which include the protein name and organism.





# THANKS!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**

