

Daten

consensus_ppi_bioplex_biogrid_intact_huri_edgelist.tsv

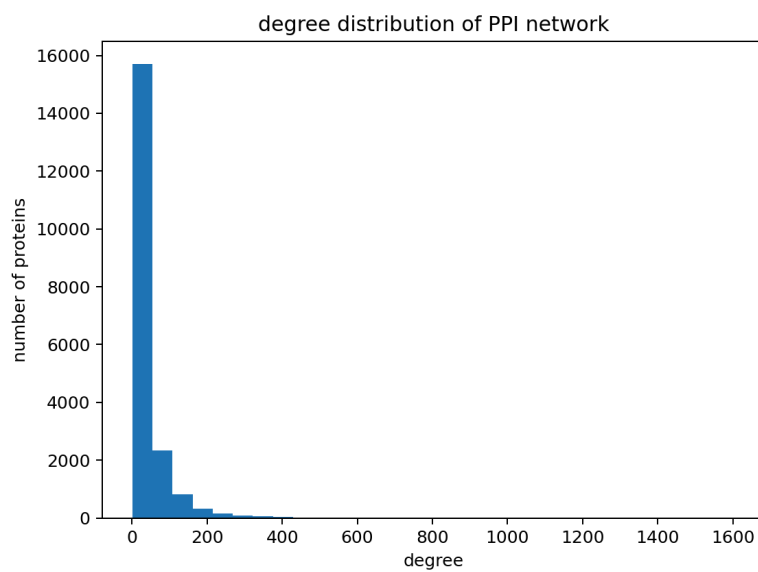
Nodes: 19582

Edges: 365500

avg. degree: 37.3302

median degree 18

max degree 1605



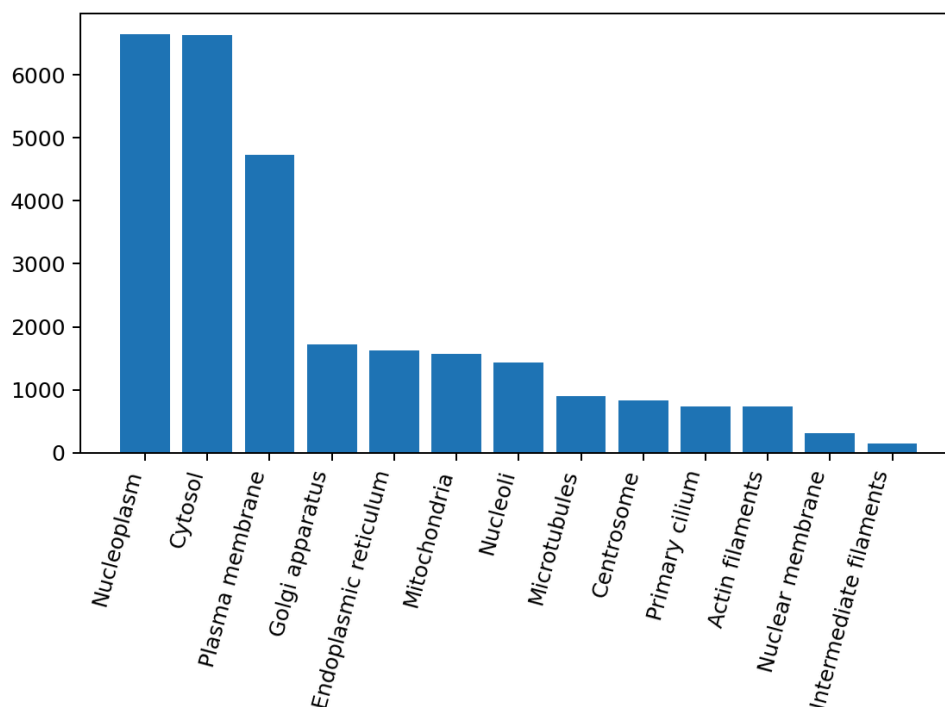
```
P04637,948
P42858,911
Q13618,859
P61981,814
P11142,763
Q9NRI5,739
```

O60260 (PRKN/Parkin), P01106 (MYC), P04637 (TP53), P42858 (HTT), Q13618 (CUL3)

Der fundamentale Unterschied zwischen Theorie und Praxis resultiert aus Study Bias: Die Forschungsintensität selbst folgt einer Power-Law-Verteilung, und aggregierte PPI-Netzwerke reflektieren wissenschaftliches Interesse mehr als biologische Realität. Mathematische Analysen demonstrierten, dass Power-Law-Topologien aus Study Bias emergieren können, selbst wenn das zugrunde liegende biologische Netzwerk einer vollständig anderen Distribution folgt

```
uniprot_A    uniprot_B
id_A    id_B line
```

protein_location_HPA_GO.tsv



```
cut -f2 protein_location_HPA_GO.tsv | sort -u
```

```
Actin filaments
Centrosome
Cytosol
```

Endoplasmic reticulum
Golgi apparatus
Intermediate filaments
Microtubules
Mitochondria
Nuclear membrane
Nucleoli
Nucleoplasm
Plasma membrane
Primary cilium

Multi-localized Proteins

```
cut -f1 protein_location_HPA_GO.tsv | sort | uniq -c | sort -k 1 | tail
```

Number of locations	Protein
7	STING1
8	CCDC88A
8	CSNK1D
8	DVL3
8	HSPA1A
8	HSPA1B
8	PSEN1
8	RPS3
9	MAPK1
9	MAPK3

-> MAPK1 and MAPK3 in most locations

project's architecture

Before writing any code, define the project's architecture, module boundaries, and data flow. Produce a flowchart and a short architecture document covering data verification, loading, processing, and how modules interact. The goal is a clear, modular design that we all agree on before implementation begins.

Set up the project's infrastructure. Based on research and flowchart-/ project's code design, start to setup the project files and structure accordingly.

Initial presentation

Introduction: Network Biology
Data (PPI, Localisation, Validation)

Flow Chart Challenges