

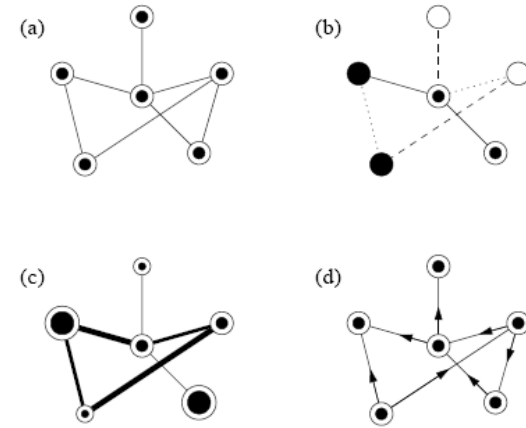
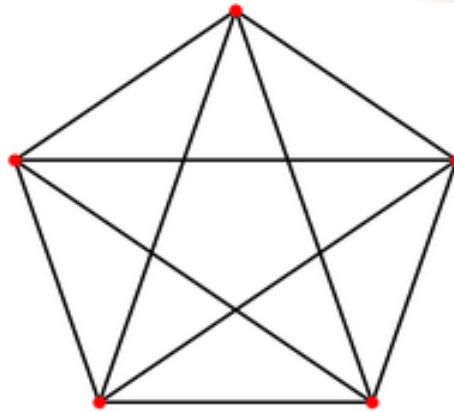
Aufbaukurs Bioinformatik



5. Tag: Interaktionsnetzwerke, Datenbanken
und Werkzeuge
+ Nachbesprechung

Ursula Kummer, Sven Sahle
mit Ulla Rost, Katja Wegner und Andreas Weidemann

Graphen



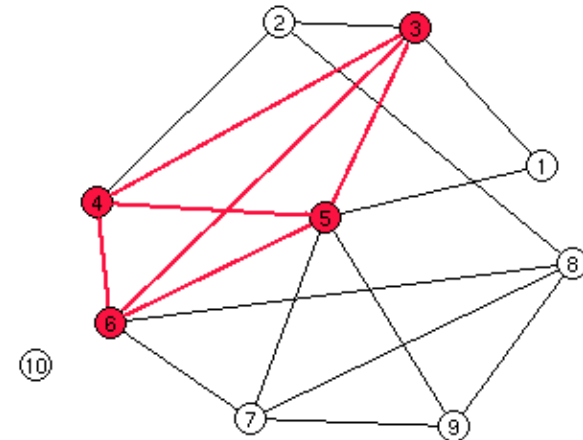
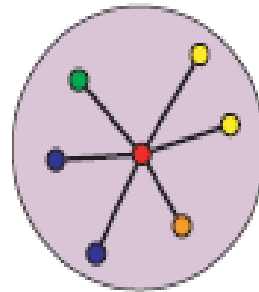
Begriffe:

Knoten

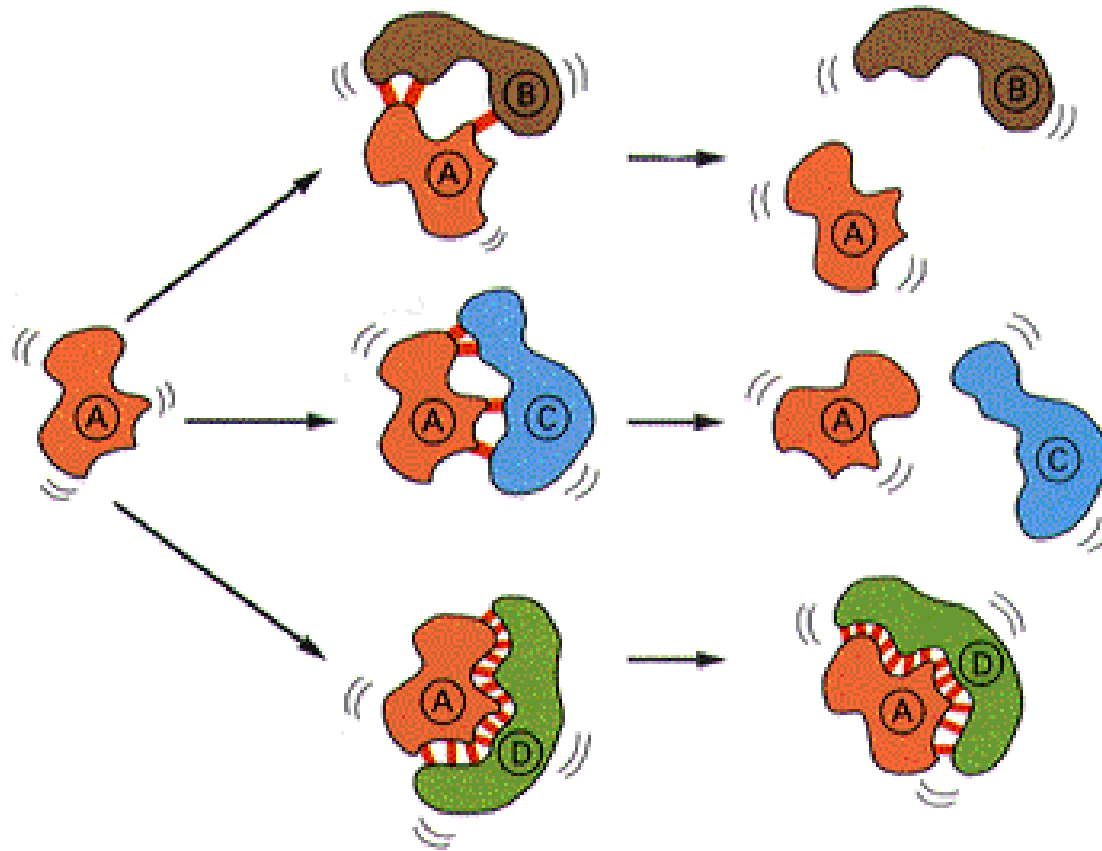
Kanten

Gerichtet/Ungerichtet

Hubs



Protein-Protein-Interaktions- Netzwerke



From The Art of MBoC³ © 1995 Garland Publishing, Inc.

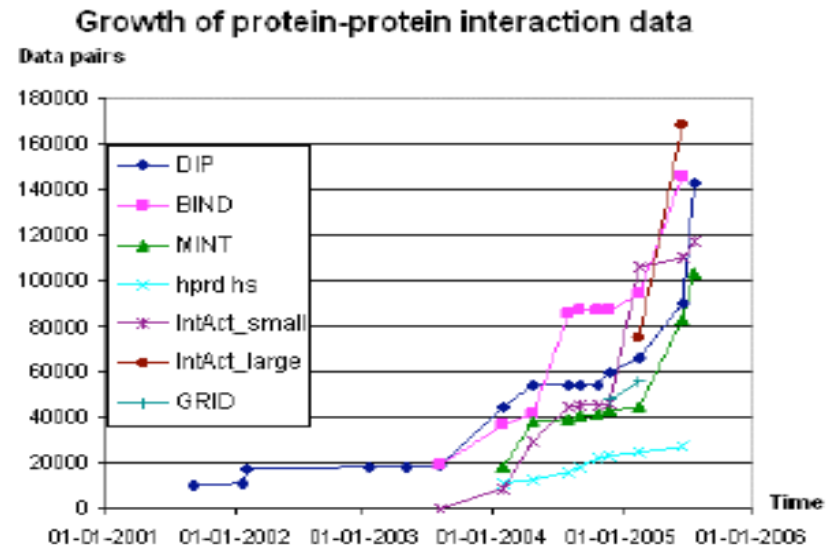
Woher kommen die Daten?

- **Phyikalische Interaktionen**

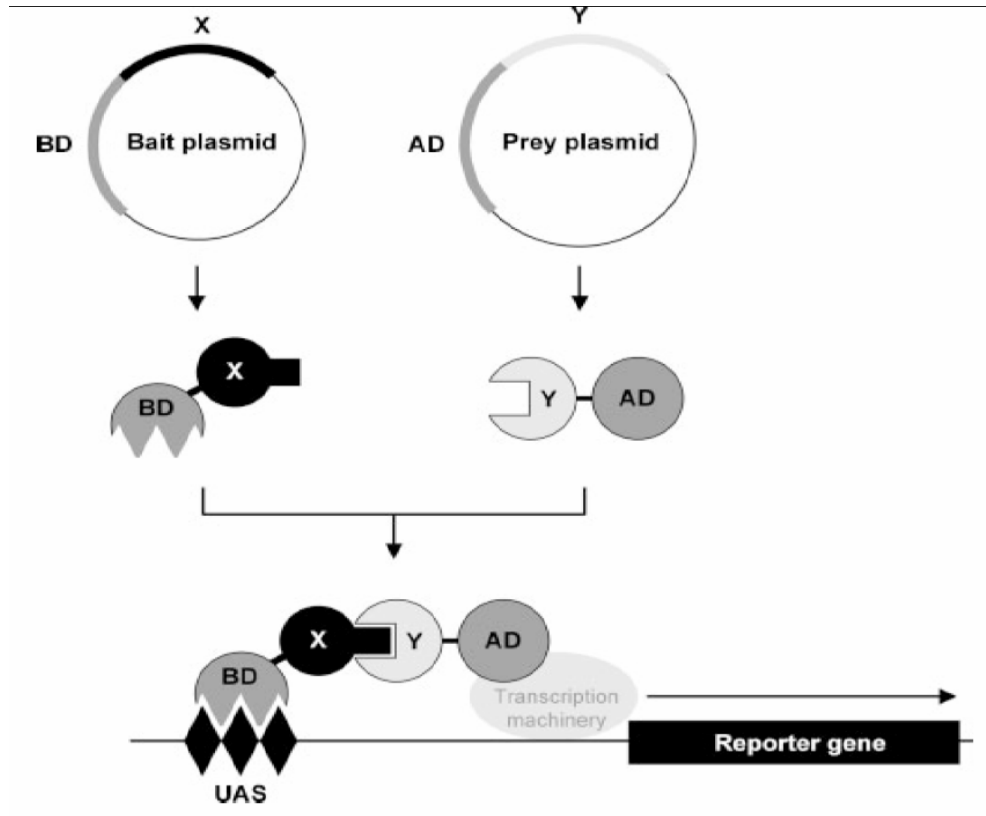
- z.B.
- Yeast two hybrid screens
- Affinitätsaufreinigung (MS)
-

- **Andere Interaktionen**

- Genetischer Kontext,
z.B. Correlationen bei
der Expression



Yeast two-hybrid

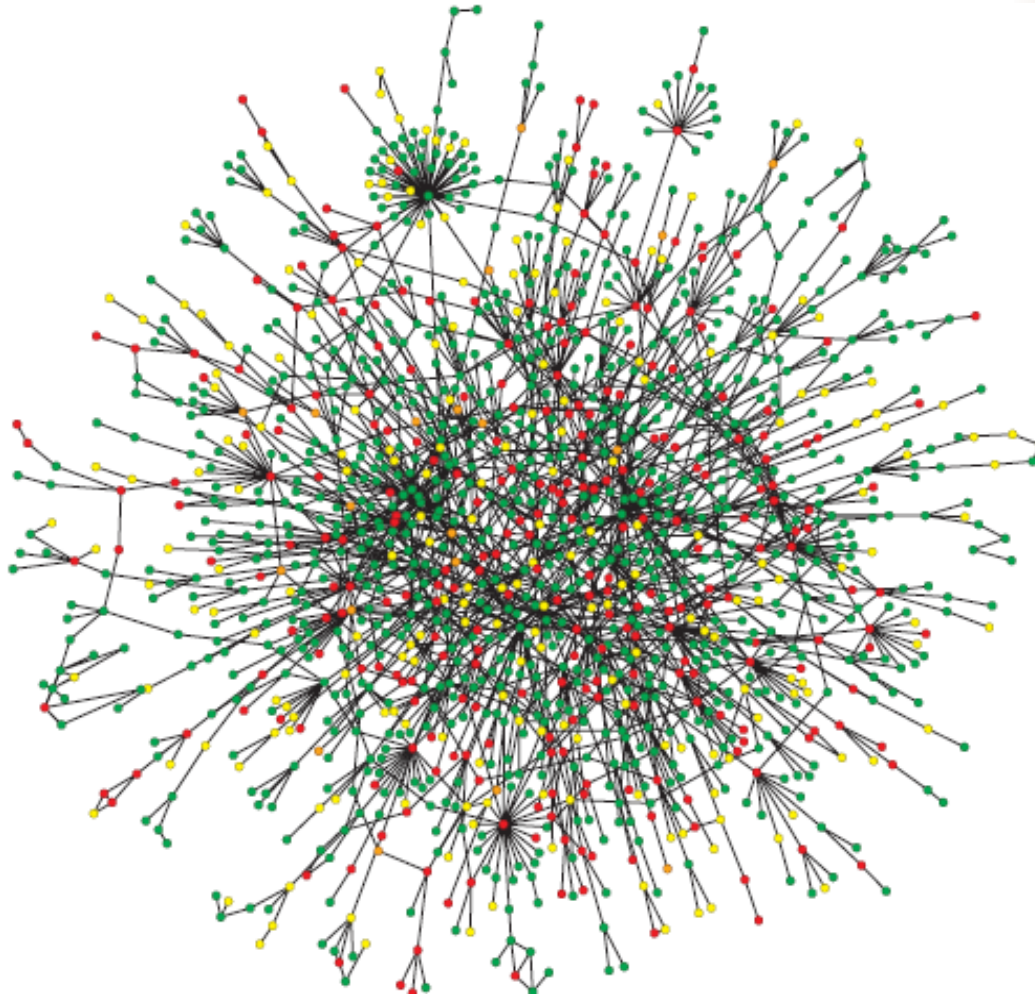


Y2H erkennt Interaktionen *in vivo*.

Ein Transkriptionsfaktor wird aus einer Aktivierungsdomäne (AD) und einer DNA-Bindungs-Domäne (DBD) aufgebaut.

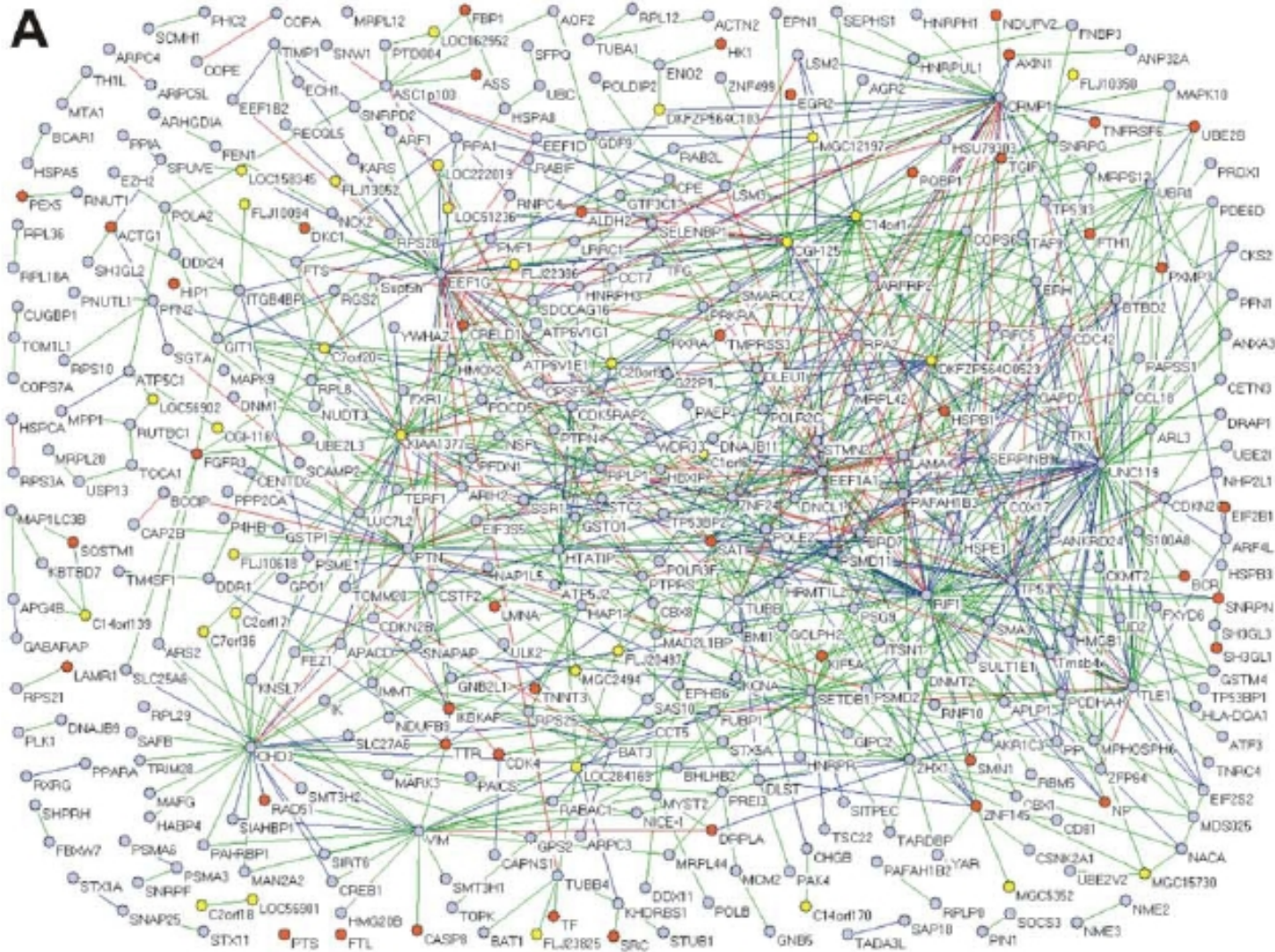
Ein Köderprotein ('bait') B wird an eine DBD angehängt und gegen eine Bibliothek von Beuteproteinen ("preys"), von denen jedes an einer AD hängt, gescreent.

Beispiel



- Hefe PPI-Netzwerk
- Knoten – proteins
- Kanten – interactions

Beispiel: Mensch



Falsch Positive und falsch Negative

- Die Interaktome (z.B. Hefe) decken ca. 3-9% des kompletten Interaktoms ab
- > Falsch Negative durch limitierte Kenntnisse
- > Falsch Positive durch experimentelle Fehler und unphysiologisches Milieu

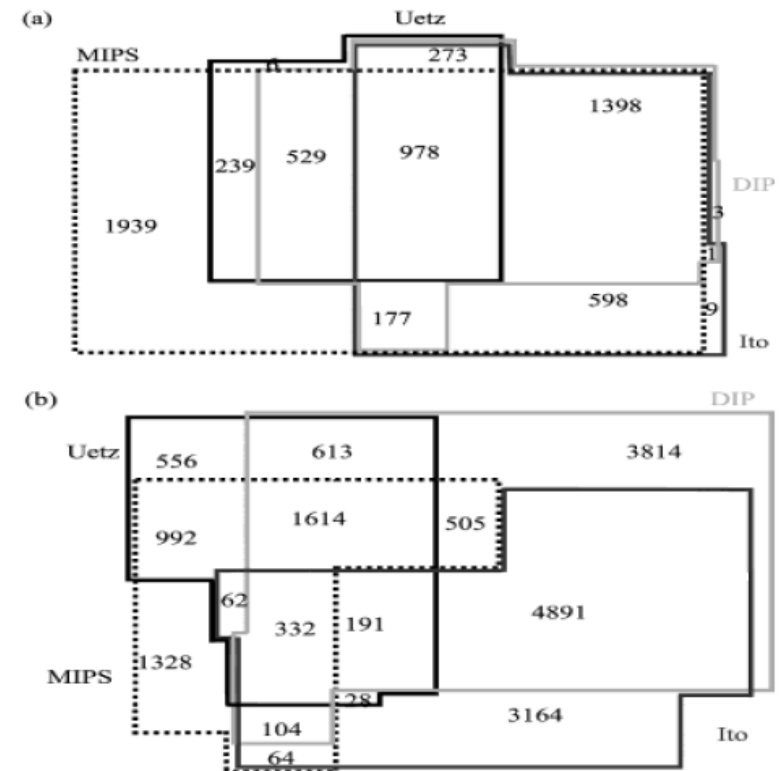
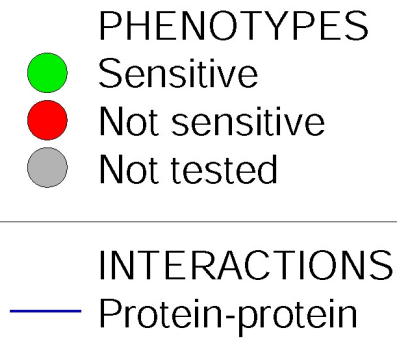
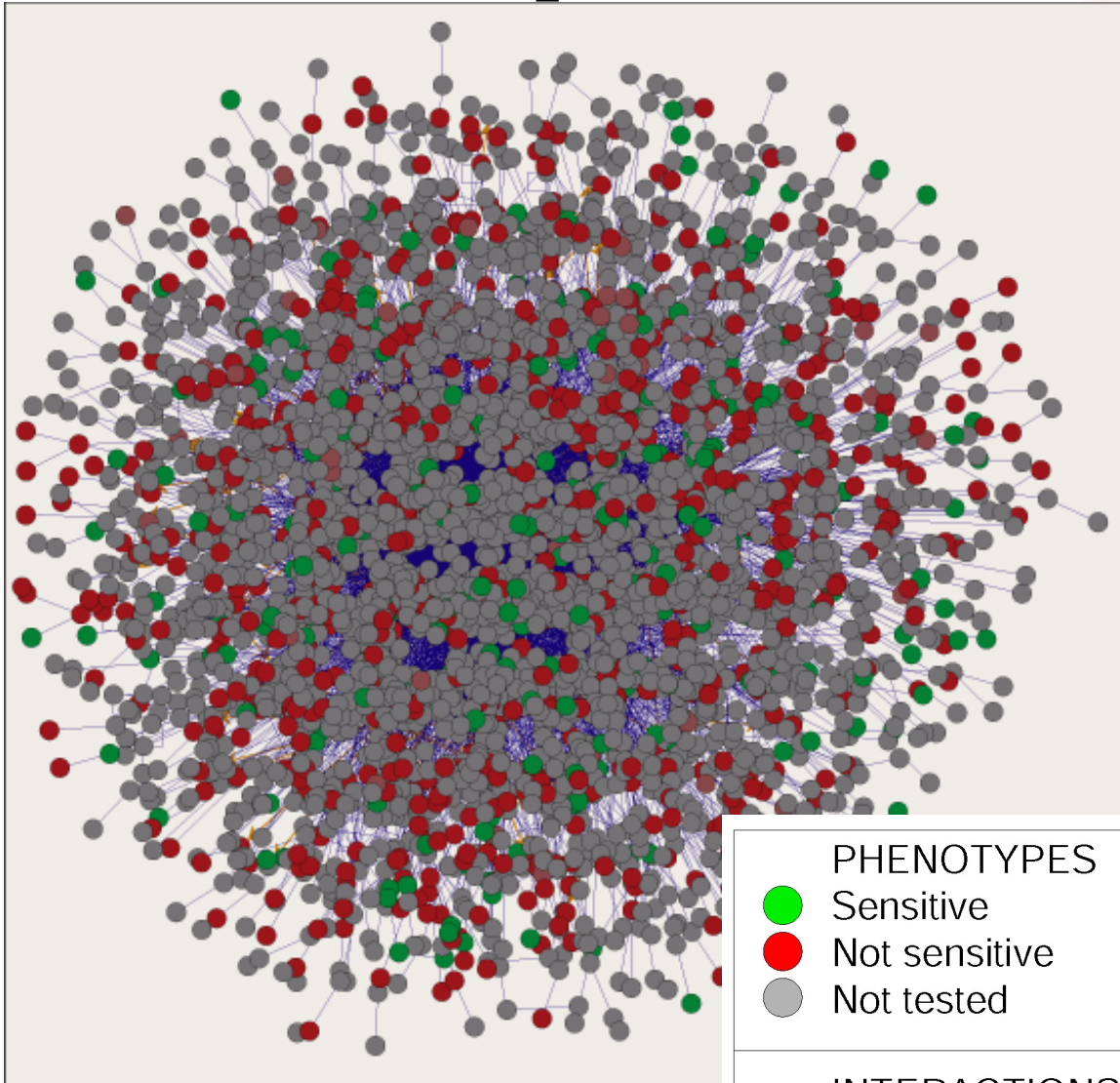


Figure 1. Relationship between the four studied databases (a) at the protein and at (b) the interaction level. The sum of the numbers within each color boundary denotes (a) the total number of proteins or (b) the total number of interactions found in the corresponding database. For example, (a) indicates that the MIPS database has altogether 6745 proteins, of which 1939 proteins with at least one interaction do not show interactions in any other databases. Note that while the overlap between the Uetz and Ito maps, (a) at the protein level is as high as 30%, (b) at the interaction level it is much smaller.

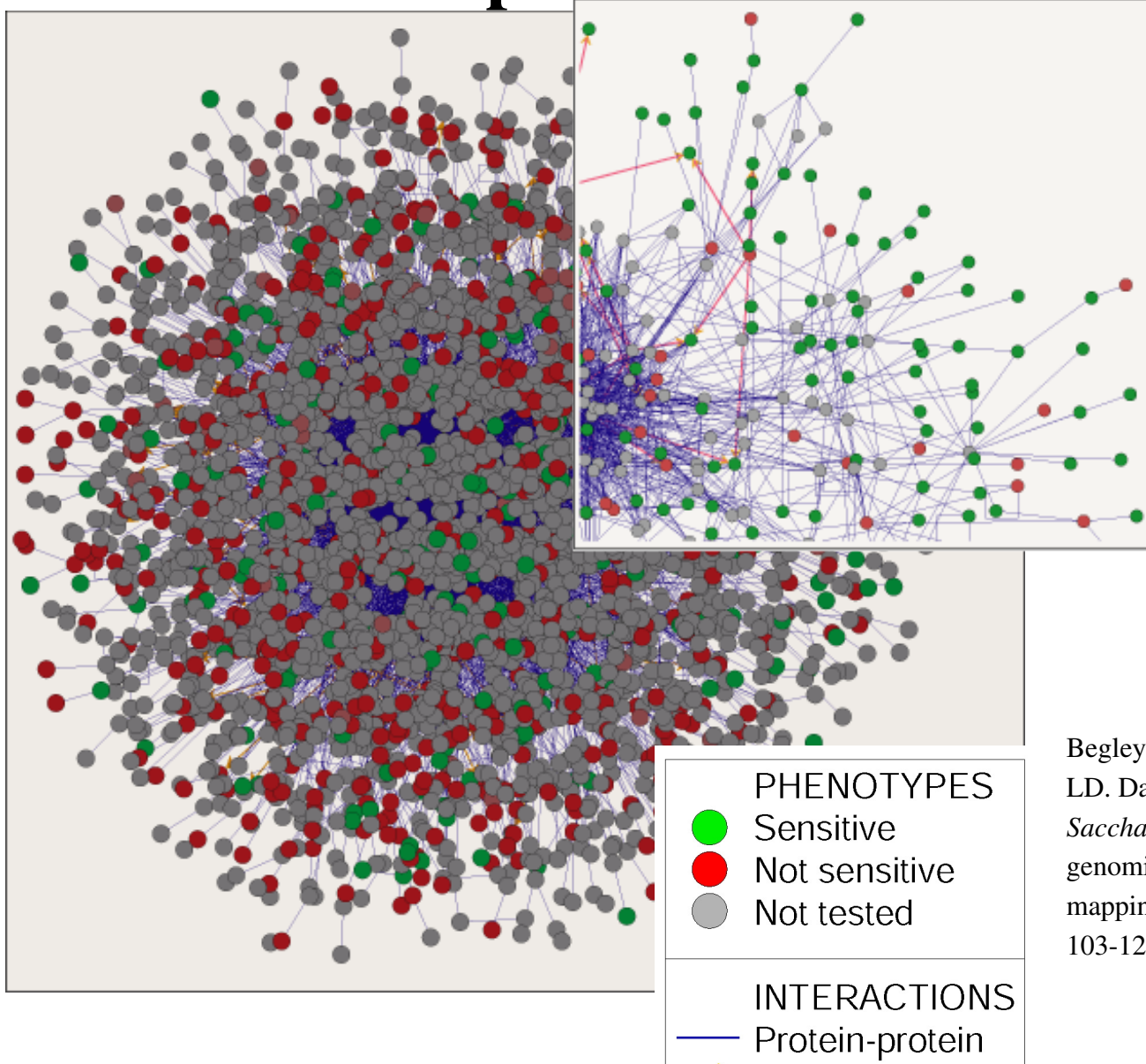
Beispiel: Hefe-Knockouts



- Systematische Phenotypisierung von 1615 Knockout-stämmen
- Evaluation des Wachstums in der Gegenwart von Methylmethansulfonat (MMS und anderen DNA schädigenden Substanzen)
- Vergleich mit einem PPI-Netzwerk

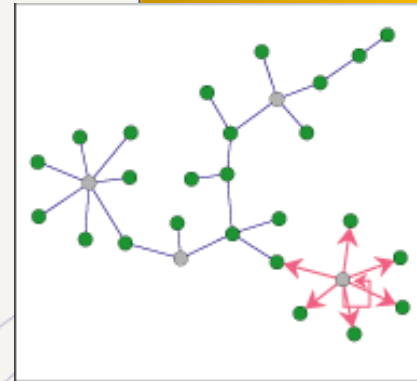
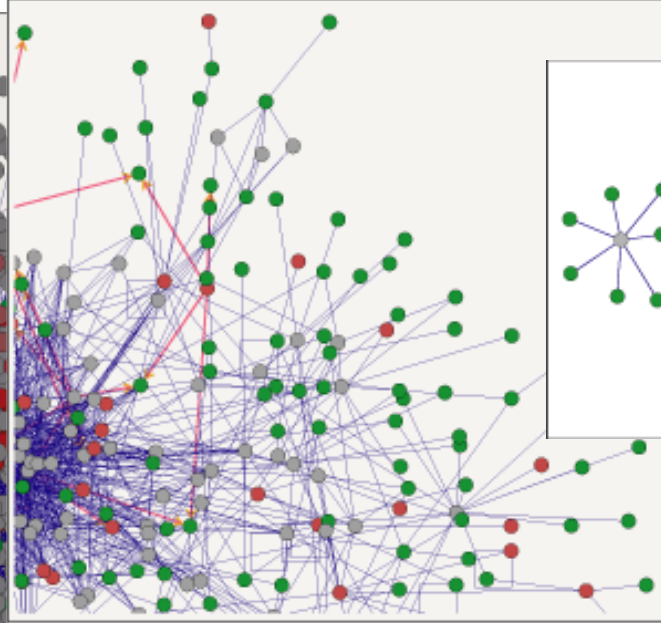
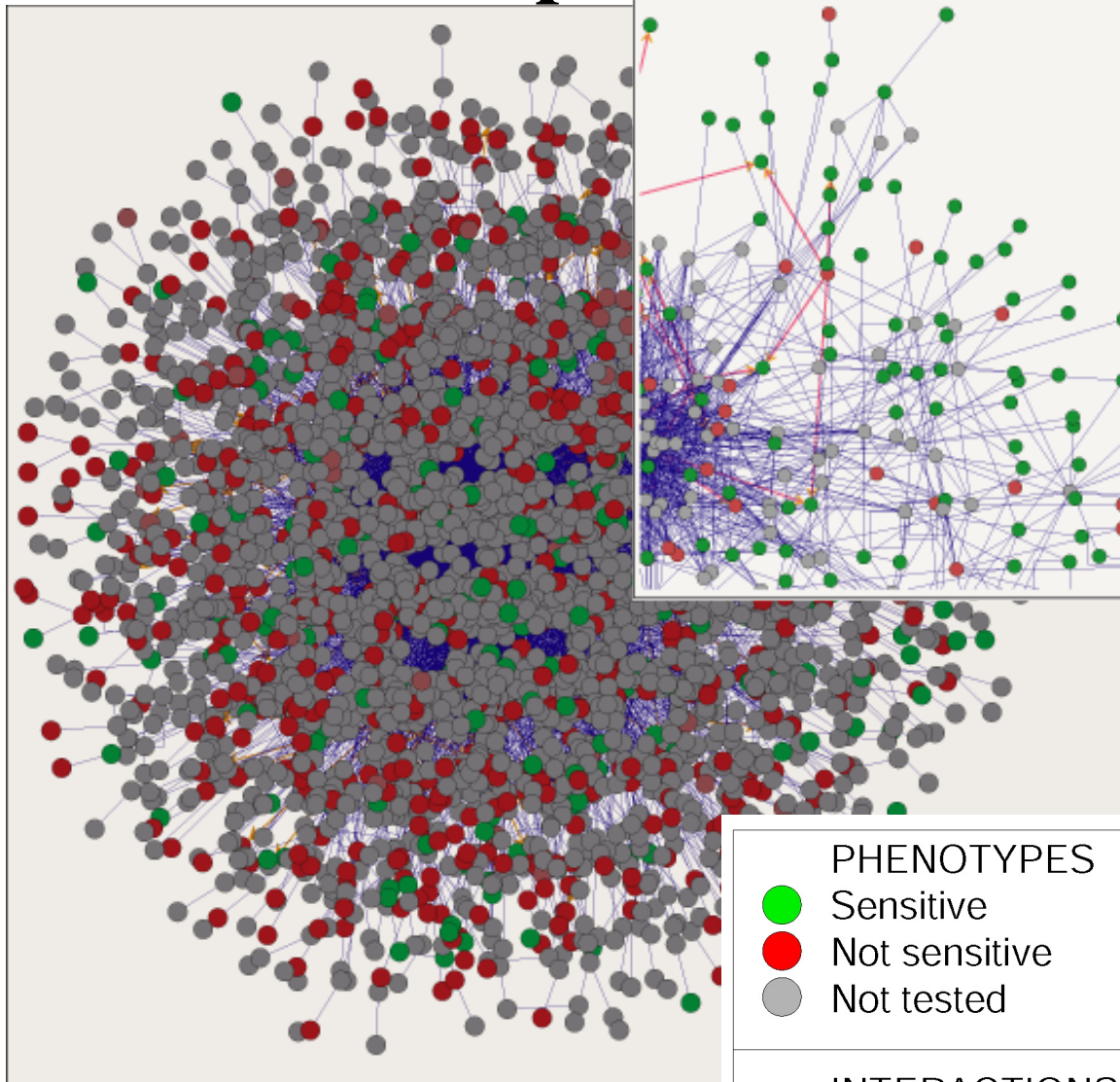
Begley TJ, Rosenbach AS, Ideker T, Samson LD. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol Cancer Res.* 2002 Dec;1(2): 103-12.

Beispiel: Hefe-Knockouts



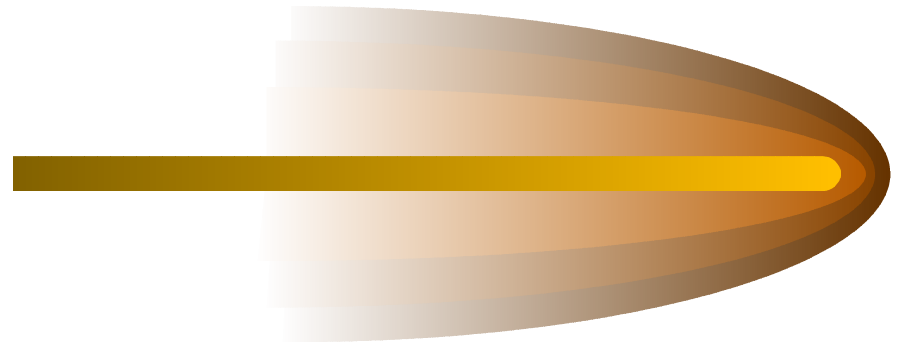
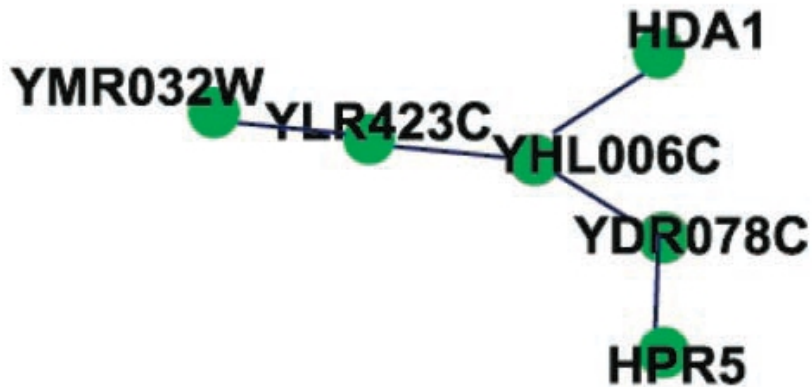
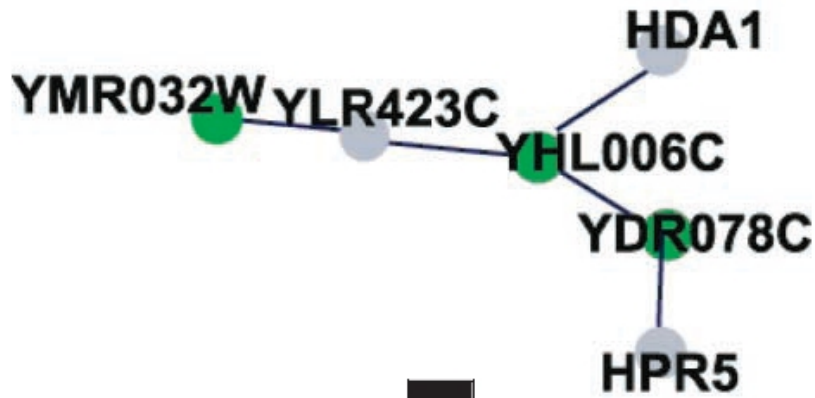
Begley TJ, Rosenbach AS, Ideker T, Samson LD. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol Cancer Res.* 2002 Dec;1(2): 103-12.

Beispiel: Hefe-Knockouts



- PHENOTYPES
 - Sensitive
 - Not sensitive
 - Not tested
- INTERACTIONS
 - Protein-protein

Begley TJ, Rosenbach AS, Ideker T, Samson LD. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol Cancer Res.* 2002 Dec;1(2): 103-12.



Grüne Knoten sind für MMS-Resistenz notwendig; graue Knoten wurden erst nicht getestet.

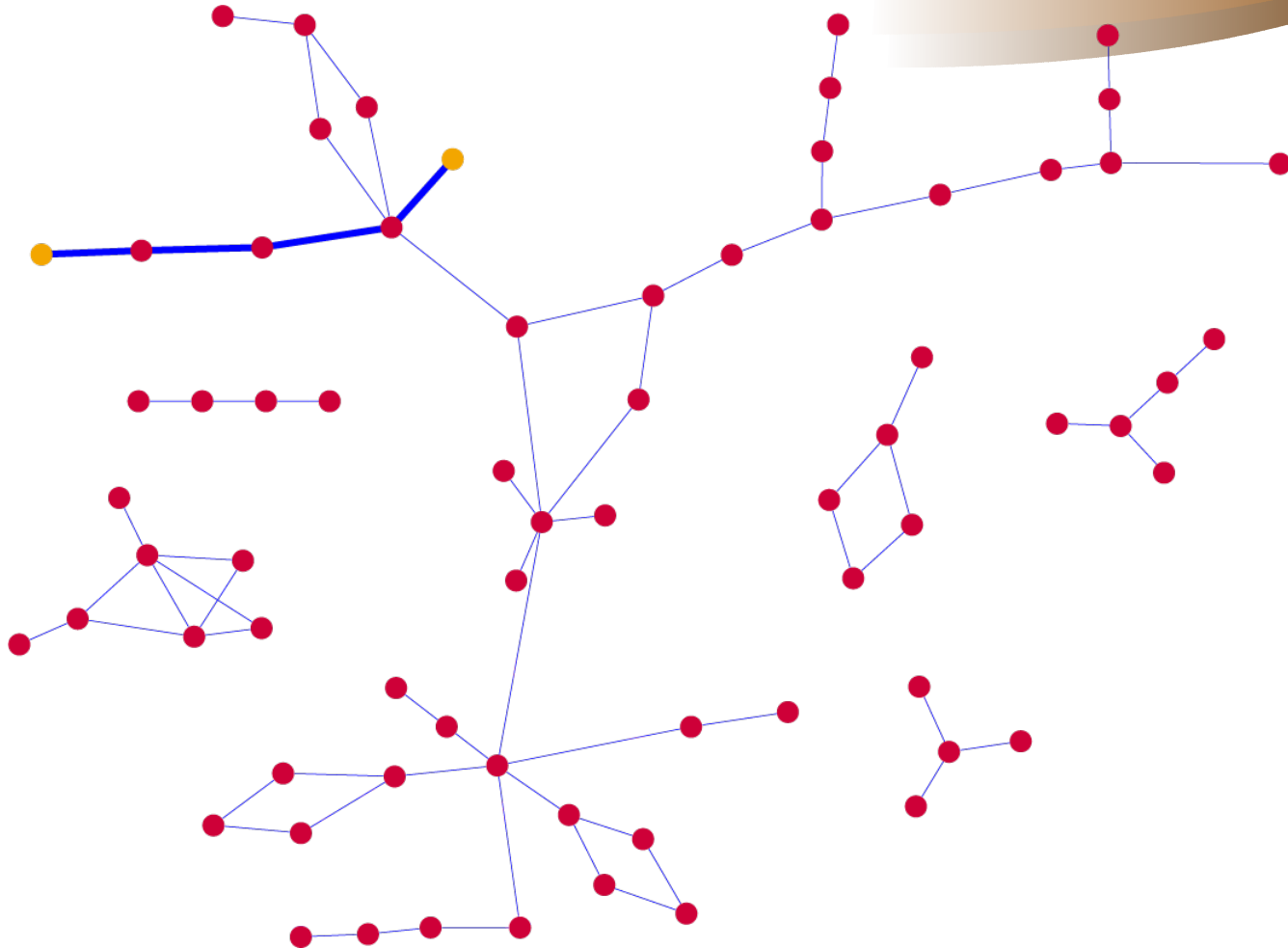
Später nachgetestet und alle sensitiv.

Begley TJ, Rosenbach AS, Ideker T, Samson LD. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol Cancer Res.* 2002 Dec;1(2):103-12.

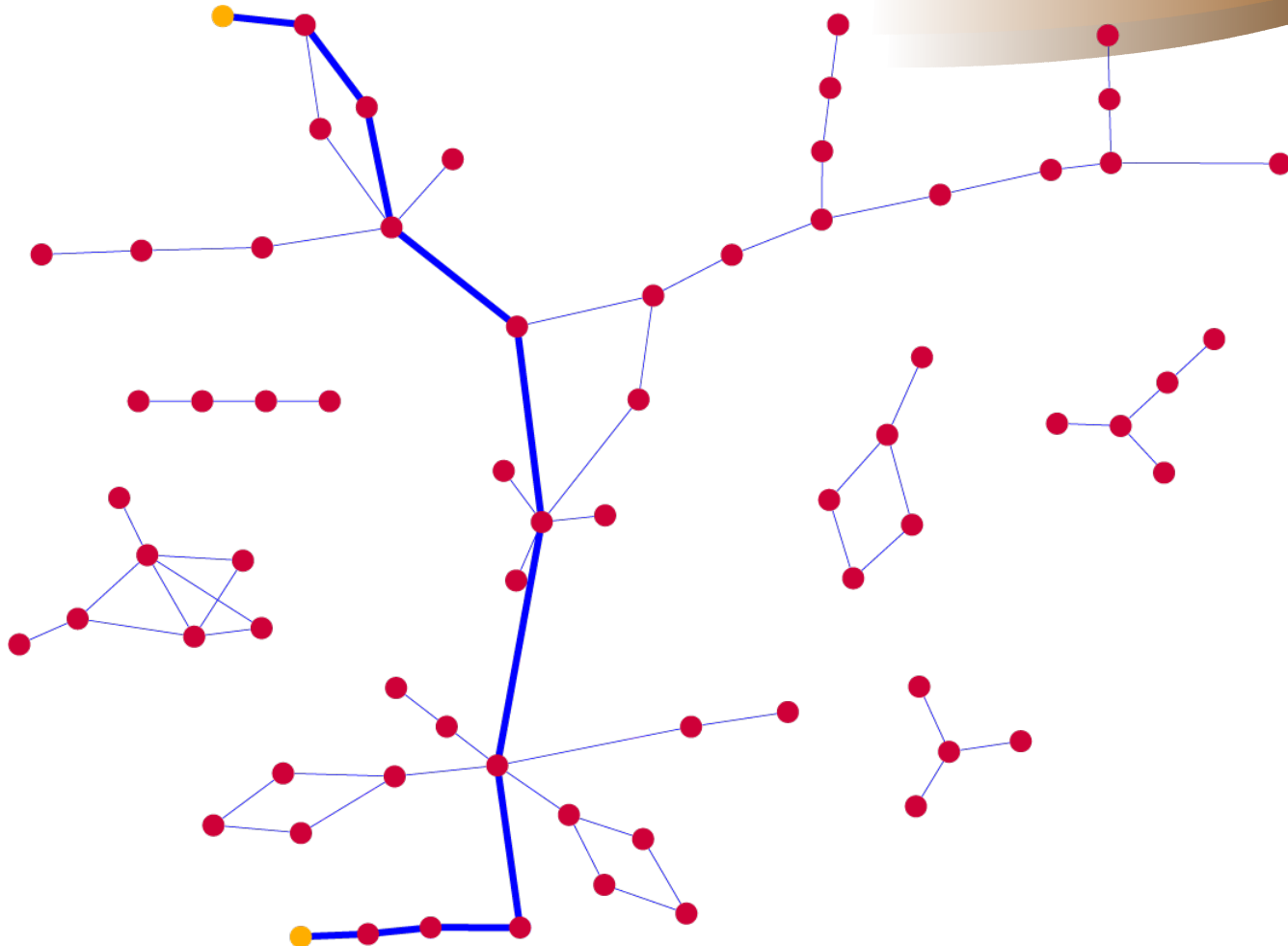
Netzwerkparameter

- Grad (Degree) k_i
Die Zahl der Kanten, an denen Knoten i beteiligt ist
- Gradverteilung? -> Degree distribution $P(k)$
Die Häufigkeit (Frequenz) von Knoten des Grades k
- Mittlere Pfadlänge
Die durchschnittlichen kürzesten Pfade zwischen allen Knotenpaaren
- Clusterkoeffizient? -> Clustering Coefficient
Mass für die Dichte der Nachbarschaft

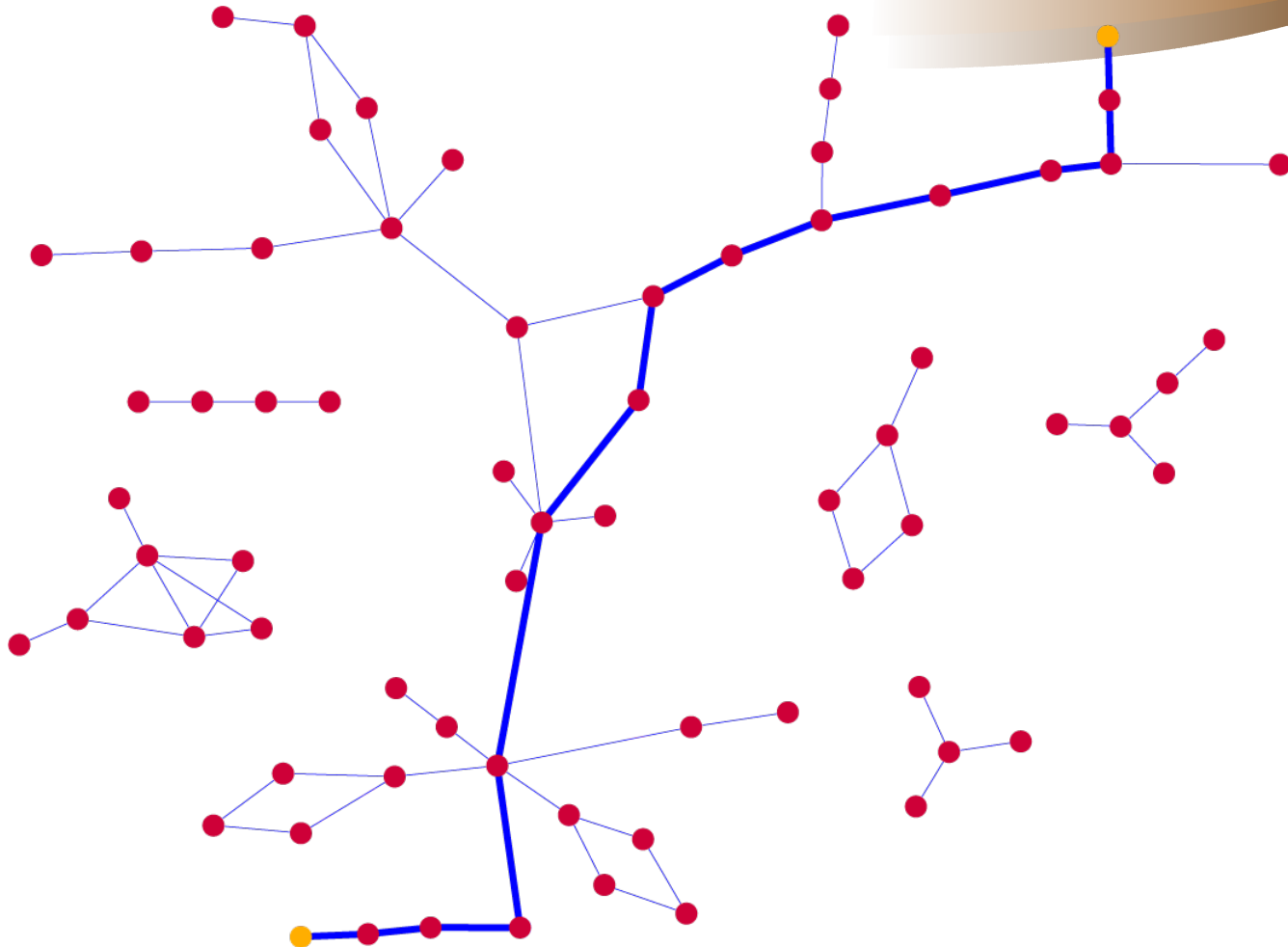
Kürzeste Pfade



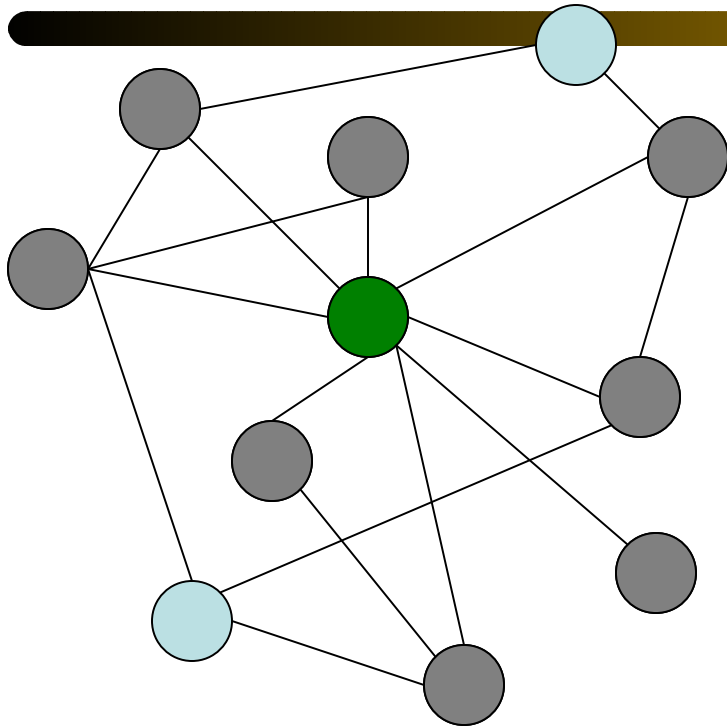
Kürzeste Pfade



Der längste kürzeste Pfad



Clustering Coefficient



Maß für die Dichte des Netzwerks in der Umgebung des Knoten I

$$C_I = \frac{2n_I}{k \cdot (k-1)}$$

k : Nachbarn von I

n_I : Kanten zwischen den Nachbarn

Der grüne Knoten hat 8 Nachbarn

Es gibt 4 Kanten zwischen den Nachbarn

$$C = 2 \cdot 4 / (8 \cdot (8-1)) = 8/56 = 1/7$$

Small-world-Netze

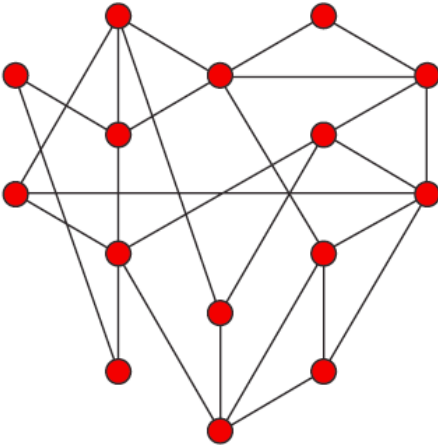


- Jeder Knoten kann von jedem anderen durch wenige Schritte erreicht werden
- Grosser Clustering Coefficient und kurze durchschnittliche kürzeste Pfade
- Soziale Netze, das Internet und biologische Netze haben diese Eigenschaften

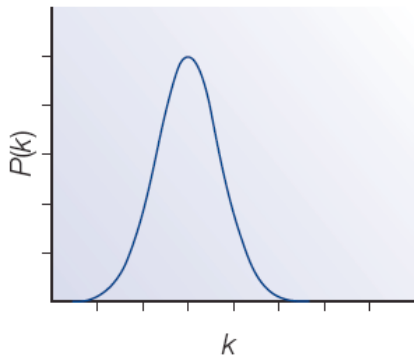
Degree Distribution

A Random network

Aa

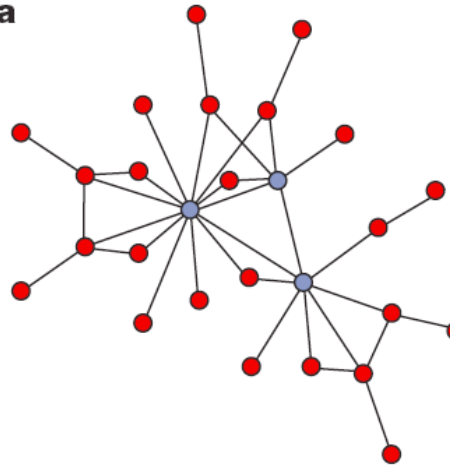


Ab

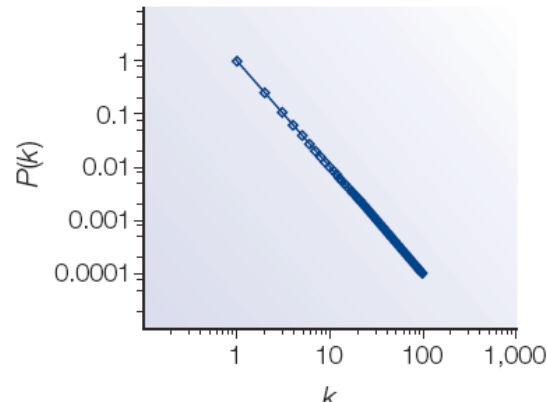


B Scale-free network

Ba



Bb



Für zufällige Netzwerke, ist $P(k)$ normalverteilt

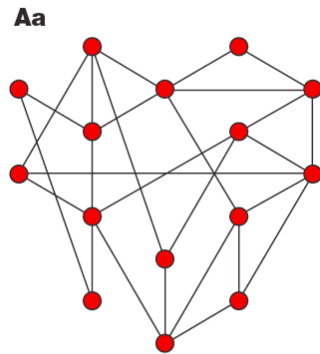
Für reale Netzwerke sind die Verteilungen oft einem Potenzgesetz folgend:

$$P(k) \sim k^{-\gamma}$$

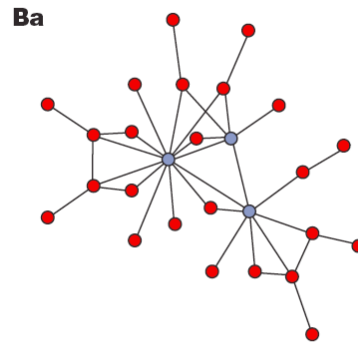
-> scale-free

Analyse der Topologie

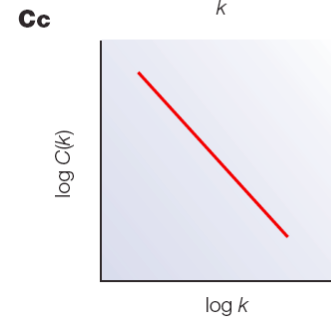
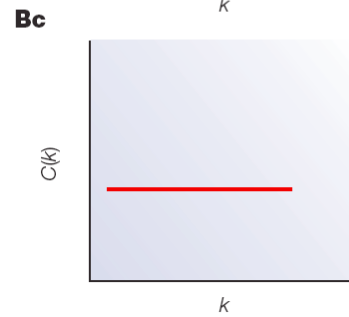
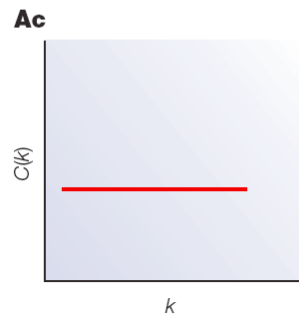
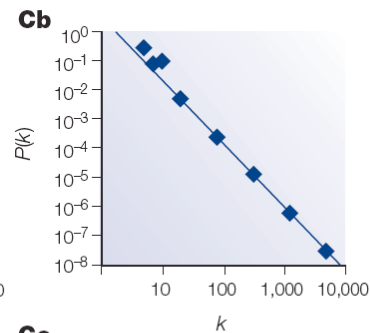
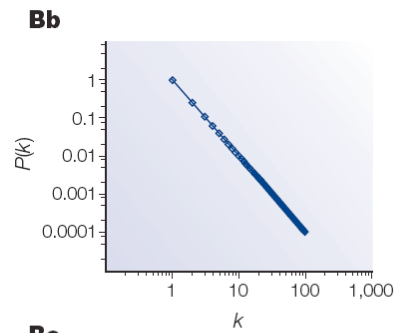
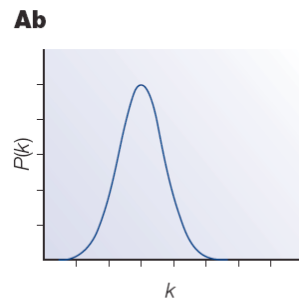
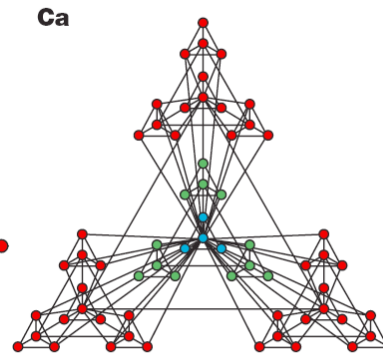
A Random network



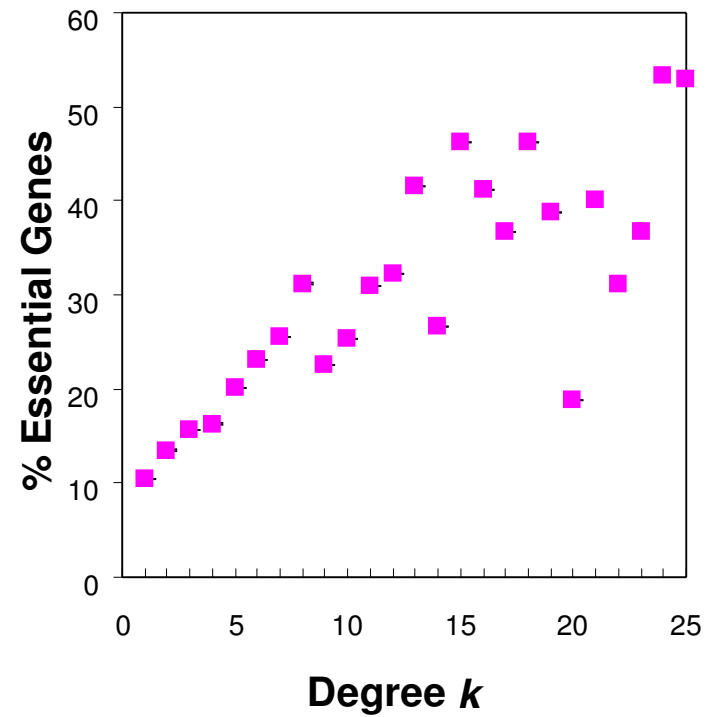
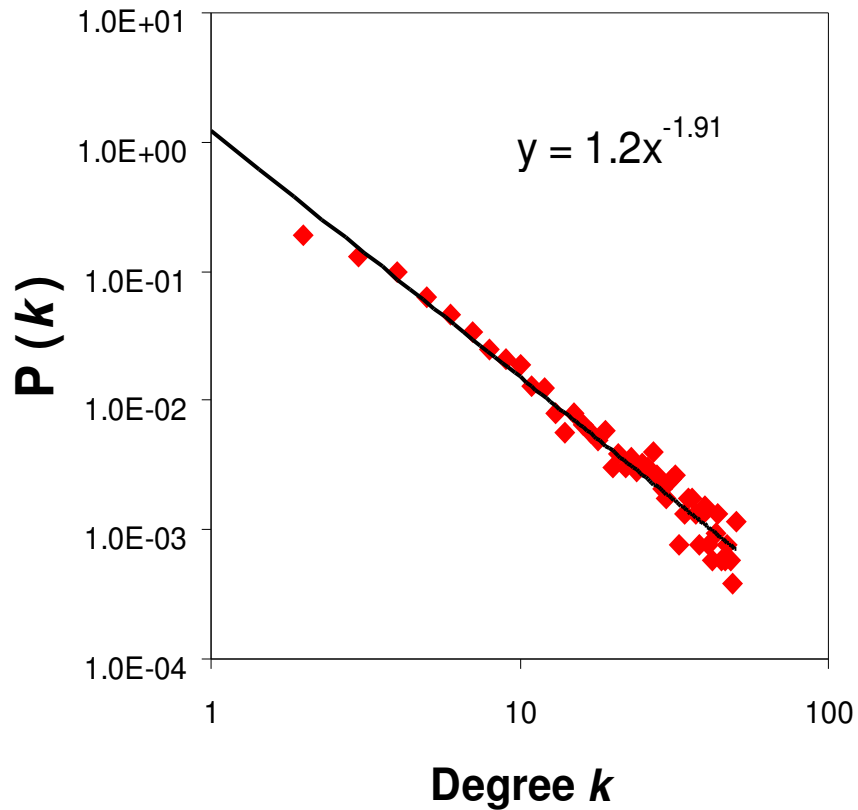
B Scale-free network



C Hierarchical network



Letalität von Knock-outs

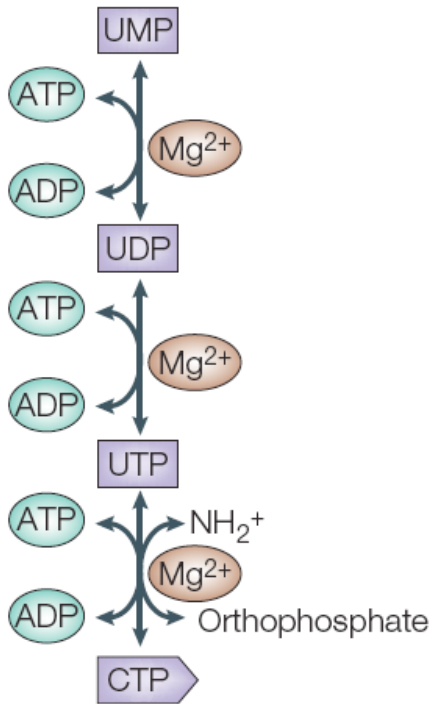


Scale-Free-Netze sind robust

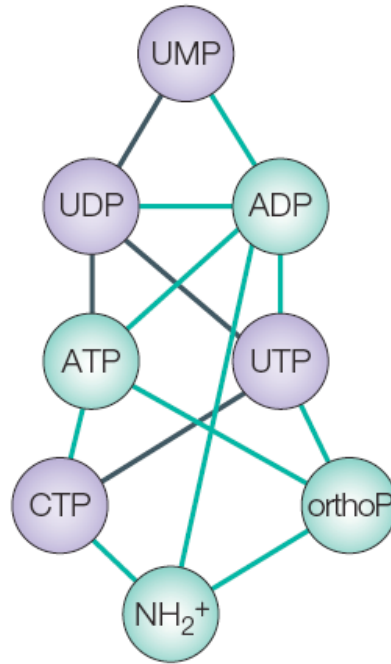
- Komplexe Systeme sind robust gegenüber dem Ausfall einzelner Komponenten
- Die Topologie spielt dabei eine wichtige Rolle
 - Sogar wenn ~80% der Knoten ausfallen, können die übrigen ~20% verlinkt bleiben
- In Hefe sind nur ca. ~20% der Proteinausfälle letal und es ist 5mal wahrscheinlicher, dass sie einen Grad $k > 15$ als $k < 5$ haben.

Metabolische Netze

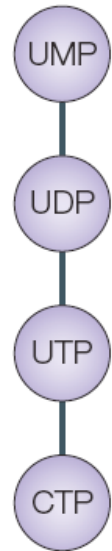
a



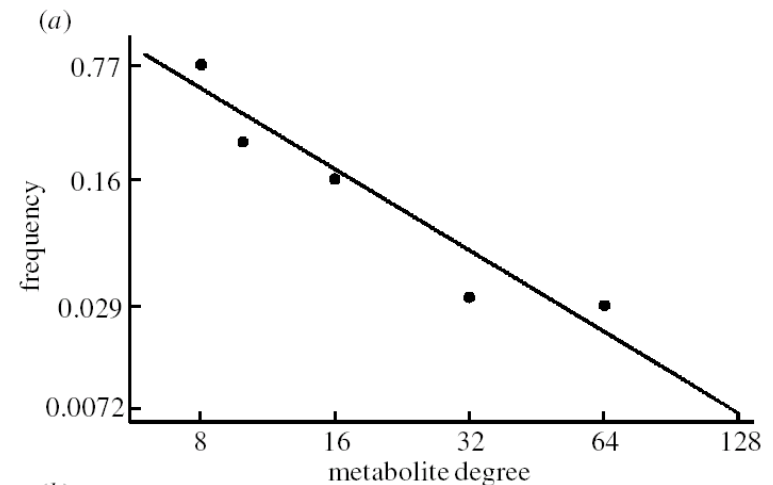
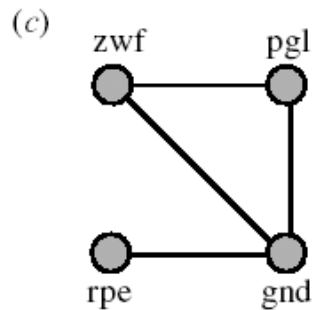
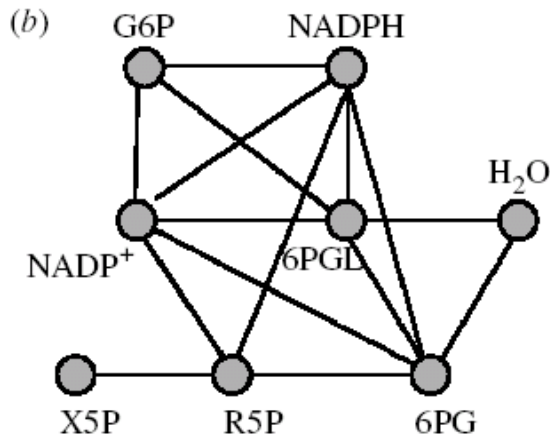
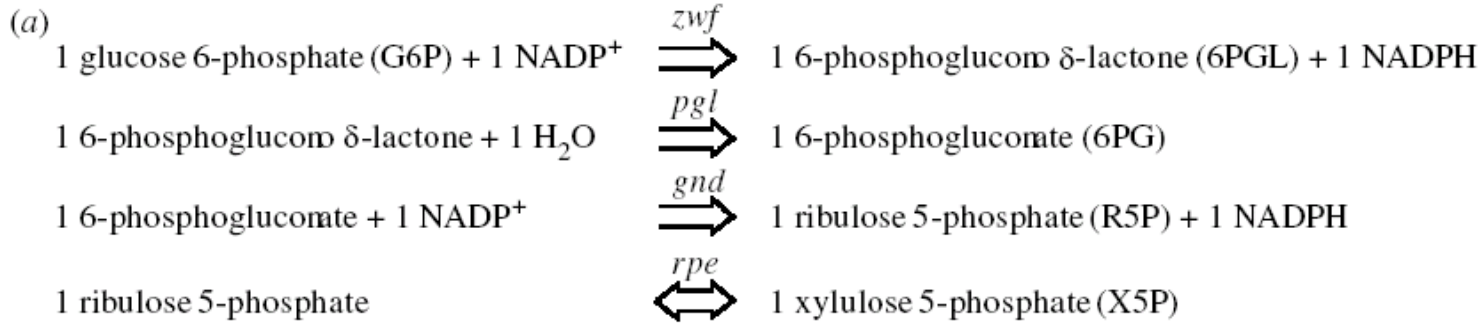
b



c



Metabolische Netze



Wagner, A, Fell DA

Proc Biol Sci. 2001 Sep 7;268(1478):1803-10.

PPI – Datenbanken + Tools

- München (Mammalian PPI Database)
<http://mips.gsf.de/proj/ppi/>
- BIOGRID – genetische und PPI
<http://biodata.mshri.on.ca/grid/>
- <http://128.100.65.8/ophidv2.201/> - PPI (exp. und IT) ursprünglich nur Mensch
- Intact (EBI) - PPI
<http://www.ebi.ac.uk/intact/site/index.jsf>
- Visant – Visualization von Interaktionen – z.Zt. sehr schlecht
<http://visant.bu.edu/>
- Apid – greift auf mehrere DB zu – auch Statistik
<http://bioinfow.dep.usal.es/apid/index.htm>