

Einführung in die Bioinformatik

Kay Nieselt
 Integrative Transkriptomik
 Zentrum für Bioinformatik Tübingen
 Kay.Nieselt@uni-tuebingen.de
 SS 2013

5. Biologische Netzwerke - Gut vernetzt hält besser

Überblick

- Einleitung
- Hierarchie biologischer Netzwerke
- Netzwerkmaße
- Skalenfreie Netzwerke
- Kleine-Welt-Netzwerke

Von der Biologie zur Systembiologie

Im vergangenen Jahrhundert studierte die Biologie

- Struktur von Proteinen
- Struktur von DNA/RNA
- Replikation
- Transkription
- Translation
- Interaktion von Molekülen
- ...

Von der Biologie zur Systembiologie



Aufgrund neuer Technologien, die die Entschlüsselung von Genomen, Transkriptomen, Proteomen, usw. erlaubt, studiert die Biologie in diesem Jahrhundert systematisch

- Zellen
- Organe
- Organismen
- Zelluläre Prozesse: Kommunikation, ...

⇒ Systeme - Systembiologie

Biologische Systeme sind lebendige Netzwerke



- Netzwerk-Ansätze zur Beschreibung und Analyse komplexer vernetzter Systeme
- Netzwerke spielen eine entscheidende Rolle in der Systembiologie
- Daten von Genen, Proteinen und Metaboliten können direkt als Netzwerk interpretiert oder in Netzwerke eingebettet werden

Netzwerk-Biologie



NETWORK BIOLOGY: UNDERSTANDING THE CELL'S FUNCTIONAL ORGANIZATION

Albert-László Barabási* & Zoltán N. Oltvai[†]

A key aim of postgenomic biomedical research is to systematically catalogue all molecules and their interactions within a living cell. There is a clear need to understand how these molecules and the interactions between them determine the function of this enormously complex machinery, both in isolation and when surrounded by other cells. Rapid advances in network biology indicate that cellular networks are governed by universal laws and offer a new conceptual framework that could potentially revolutionize our view of biology and disease pathologies in the twenty-first century.

Barabási, Oltvai, Nature Rev. Genetics, 5:101 (2004)

Vielfalt biologischer Netzwerke



- Metabolische Netzwerke
- Regulatorische Netzwerke
- Interaktionsnetzwerke
- Phylogenetische Netzwerke
- Neuronale Netzwerke
- Soziale Netzwerke

Netzwerke sind Graphen

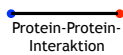


- Visualisierung / Modellierung von Netzwerken mittels Graphen
- Knoten sind Metabolite, Gene, Proteine, Neuronen, ...
- Kanten zwischen je zwei Knoten beschreiben Reaktionen, Interaktionen, neuronale Verbindungen, ...

Netzwerke sind Graphen

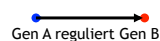


- Knoten: Gene/Proteine/Taxa ...
- Kanten: Interaktionen



- Kanten repräsentieren z.B.:
 - Protein-Protein-Interaktionen
 - Protein-DNA-Interaktionen (TF-Bindung)
 - Genetische Interaktionen
 - Regulatorische Interaktionen (positiv/negativ)

- Kanten können Richtungen haben



Eine Hierarchie biologischer Netzwerke

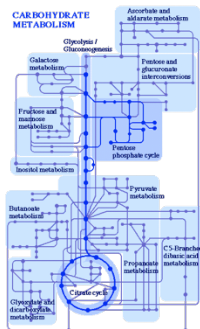


- Metabolische Netzwerke
- Interaktionsnetzwerke
- Regulatorische Netzwerke
- Signalnetzwerke
- Evolutionäre Netzwerke
- Soziale Netzwerke

Eine Hierarchie biologischer Netzwerke



- Metabolische Netzwerke
- Interaktionsnetzwerke
- Regulatorische Netzwerke
- Evolutionäre Netzwerke



Metabolische Netzwerke



- Für viele Zellprozesse wird Energie benötigt
- Energie wird durch **Metabolismus** erlangt:
 - (1) **katabolische Reaktion**: große Moleküle werden zu kleinen reduziert (Beispiel Glykolyse)
 - (2) **anabolische Reaktion**: komplexe Moleküle werden aus kleinen synthetisiert (Beispiel: Aminosäuresynthese)
- Metabolismus ist ein strukturierter Netzwerkprozess

Metabolismus = Stoffwechsel

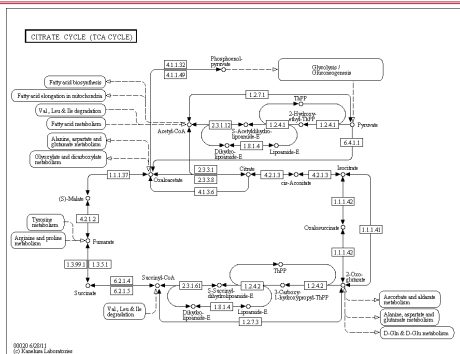
Metabolische Netzwerke



- **KEGG** (Kyoto Encyclopedia of Genes and Genomes)
- Große Datenbank mit Informationen über Gene, Biomoleküle, Stoffwechselwege, Reaktionsgleichungen etc.
- Sammlung von Netzwerkdiagrammen (manuelles Layout!)

<http://www.genome.jp/kegg/>

Netzwerk von KEGG



<http://www.genome.jp/kegg/>

Eine Hierarchie biologischer Netzwerke



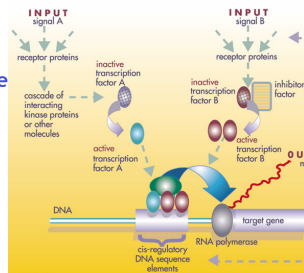
- Metabolische Netzwerke
- **Interaktionsnetzwerke**
- Regulatorische Netzwerke
- Evolutionäre Netzwerke



Barabási & Ottavai, Nat. Rev. Genetics (2004), 5:101

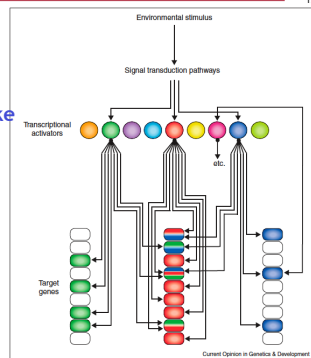
Eine Hierarchie biologischer Netzwerke

- Metabolische Netzwerke
- Interaktionsnetzwerke
- **Regulatorische Netzwerke**
- Evolutionäre Netzwerke



Eine Hierarchie biologischer Netzwerke

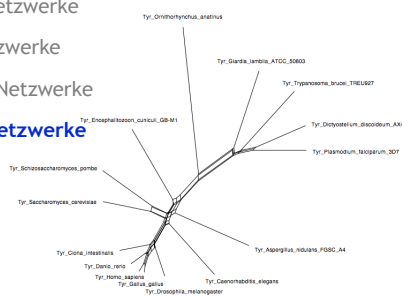
- Metabolische Netzwerke
- Interaktionsnetzwerke
- **Regulatorische Netzwerke**
- Evolutionäre Netzwerke



Wyrick and Young, Deciphering gene expression regulatory networks, Curr Opin Genet & Dev., 2002.

Eine Hierarchie biologischer Netzwerke

- Metabolische Netzwerke
- Interaktionsnetzwerke
- Regulatorische Netzwerke
- **Evolutionäre Netzwerke**



Graph generiert mit SplitsTree4, Huson et al.

Knoten/Kanten in biol. Netzwerken



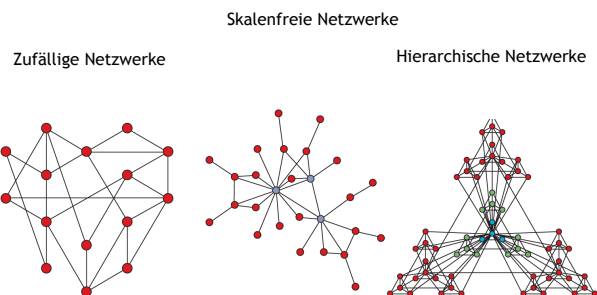
- Metabolische Netzwerke: Metabolite/Reaktionen
- Interaktionsnetzwerke: Proteine/Interaktionen
- Regulatorische Netzwerke: Proteine/Aktivierung, ...
- Evolutionäre Netzwerke: Spezies/Evolution

Analyse biologischer Netzwerke



- Untersuchung **struktureller Eigenschaften** von Netzwerken: besseres Verständnis für Aufbau, Robustheit gegenüber Störungen, Bedeutung zentraler Elemente
- Analyse und **Visualisierung** von Daten im Kontext zugrunde liegender biologischer Prozesse und Netzwerke
- Untersuchung des **dynamischen Verhaltens** der Prozesse mittels Simulation, um die Antwort eines biologischen Systems auf Änderungen vorherzusagen

Netzwerktopologien



Barabási & Ottavai, Nat. Rev. Genetics (2004), 5:101

Netzwerkmaße

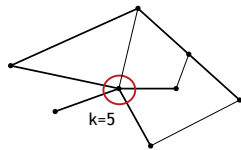


- Knotengrad, -verteilung
- Pfadlängen
- Durchmesser
- Zentralität
- Clusterkoeffizient

Gut vernetzt hält besser



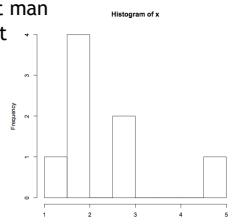
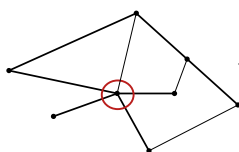
- Eines der einfachsten Maße ist der **Knotengrad** k für jeden Knoten in einem Graphen



Gut vernetzt hält besser



- Die **Gradverteilung** $P(k)$ entspricht der Wahrscheinlichkeit, dass ein Knoten den Grad k hat.
- Für ein geg. Netzwerk berechnet man $P(k)$, in dem man alle Knoten mit einem gegebenen Grad zählt.



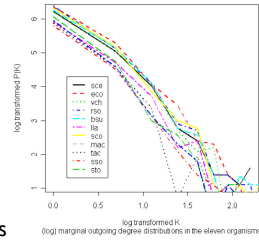
Skalenfreie Netzwerke

- In vielen biologischen Netzwerken folgt die Gradverteilung dem **Potenzgesetz**

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

γ ist der Gradexponent; γ liegt meistens zwischen 2 und 3

- Die meisten Knoten haben niedrigen Grad, aber es gibt **Hubs**, d.h. Knoten mit sehr vielen Verbindungen
- Wenn $\log(P(k))$ eines Netzwerkes als Funktion von k eine Gerade darstellt, ist das Hinweis auf Potenzgesetz $\Rightarrow$ **skalenfreies Netzwerk**



Zhu & Qin: Structural comparison of metabolic networks, *BMC Bioinformatics* 2005, 6:8

Hubs - Drehkreuze



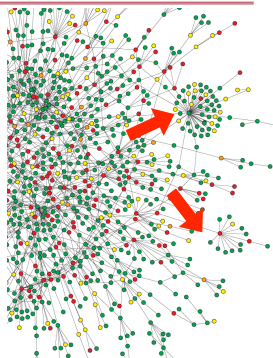
www.hemisphere magazine.com

Hubs

- Hubs finden sich oft in biologischen Netzwerken

Beispiel:

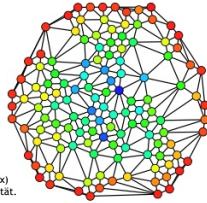
- Protein-Protein-Interaktionsnetzwerk in Hefe



Barabási & Ottavai, *Nat. Rev. Genetics* (2004), 5:101

Pfadlängen und Zentralität

- Die durchschnittliche kürzeste **Pfadlänge** zwischen je zwei Knoten ist ein Maß, wie schnell man von einem beliebigen Knoten zu einem anderen kommt.
- Weiterer wichtiger Parameter ist der **Graphendurchmesser**: die Länge des längsten kürzesten Pfades im Graphen
- Zentralität** eines Knoten: relative Anzahl der kürzesten Pfade, die den Knoten enthalten.



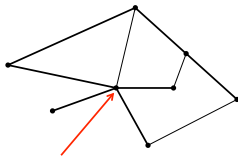
Farbton (Rot=0 nach Blau=max)
repräsentiert Knoten-Zentralität.

<http://en.wikipedia.org/wiki/Centrality>

Beispiel

Der folgende Netzwerk hat

- Ø Pfadlänge: 1,75
- Durchmesser: 3
- Zentralität für Knoten mit Grad 5: 1,3



Clusterkoeffizient

- Maß für **Transitivität** in einem Netzwerk (Ist Knoten A mit Knoten B verbunden und B mit Knoten C, ist dann auch A mit C verbunden?)
- Globaler Clusterkoeffizient:

$$C = \frac{3 \times \text{Anzahl der Dreiecke}}{\text{Anzahl der verbundenen Tripel}}$$

- Lokaler Clusterkoeffizient:

$$C_i = \frac{2n}{k_i(k_i - 1)}$$

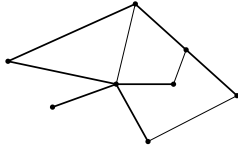
n = Anzahl der Kanten, die zwischen den k_i Nachbarn von Knoten i verlaufen

$k_i(k_i - 1)$ = Anzahl möglicher Kantenverbindungen zwischen allen Nachbarn von Knoten i .

Beispiel



- Lokaler Clusterkoeffizient für Knoten mit Grad 5: $(2 \cdot 1) / 20 = 0,1$



Small-World Netzwerke

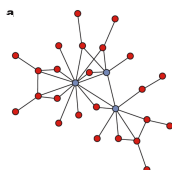


- Nachbarn eines Knotens stärker vernetzt als in zufällig organisierten Netzwerken
- Durchschnittliche Pfadlänge zwischen zwei Knoten des Netzwerks sehr kurz
- Skalenfreie Netzwerke haben diese Eigenschaft, da 'Hubs' das schnelle Durchlaufen ermöglichen
- Beispiel: Metabolische Netzwerke sind **Kleine-Welt-Netzwerke**, die meisten Paare von Metaboliten sind mittels 3-4 Reaktionen verbunden.

Stabilität



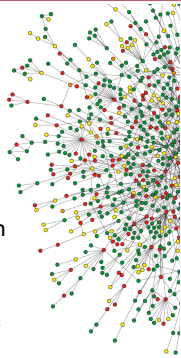
- Biologische Netzwerke sind meist robust gegenüber Störungen (Perturbationen)
- Skalenfreie Netzwerke sind sehr robust gegenüber zufälligen Deletionen: die Mehrheit der Knoten haben geringen Grad
- Hubs, auf der anderen Seite, sind sehr gefährdet: ihre Deletion könnte zum Zerfall des Netzwerkes in Teilnetzwerke führen



Stabilität

Beispiel des Protein-Protein-Interaktionsnetzwerkes in Hefe:

- Knoten sind farbkodiert hinsichtlich Effekts einer Knock-out-Mutante
 - Rot: letal
 - Grün: nicht letal
 - Orange: langsames Wachstum
 - Gelb: unbekannt
- Hubs sind oft rot!

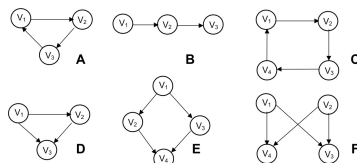


Aufgaben der Bioinformatik

- Repräsentation der Netzwerke und zugehöriger Daten
 - Datenstrukturen
 - Datenbanken
 - Datenintegration
- Analyse und Simulation der Netzwerke und der durch sie repräsentierten Prozesse
 - Algorithmen
 - Simulation
 - Statistik
 - Maschinelles Lernen
- Darstellung und Exploration
 - Visualisierung

Analyse biologischer Netzwerke

- Biologische Netzwerke sind oft modular und organisiert
- Untersuchung von kleinen, vielfach vorkommenden Teilgraphen -> Netzwerk-Motive!
- Netzwerk-Motive haben besondere Bedeutung für Signaltransduktions- und Genregulatorische Netzwerke

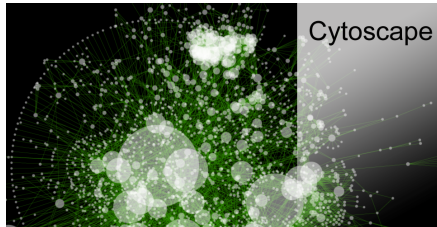


Periopoulos GA, BioData Mining 2011

Cytoscape



- open source Software-Plattform
- Visualisierung komplexer Netzwerke
- Integration beliebiger Attributdaten
- > 100 Plugins
- Anwendungsdomänen: Bioinformatik, soziale Netzwerke, semantisches Web



http://www.cytoscape.org/what_is_cytoscape.html

Zusammenfassung



- Netzwerke finden sich in allen Bereichen der Biologie
- Viele dieser sind skalenfreie-Netzwerke und weisen Eigenschaften der “Kleinen Welt” auf.
- Die Bioinformatik und die Systembiologie tragen in vielen Bereichen wie Analyse, Visualisierung, Interpretation zum Verständnis bei

Links



- <http://www.genome.jp/kegg>
- <http://string.embl.de/>
- <http://www.cytoscape.org/>
