VI – Introduction

Mon, Oct 15, 2012 Bioinformatics 3 — Volkhard Helms

How Does a Cell Work?



Medalia et al, Science 298 (2002) 1209

A cell is a crowded environment => many different proteins, metabolites, compartments, ...

On a microscopic level => direct two-body interactions

At the macroscopic level => complex behavior

Can we understand the behavior from the interactions?

=> Connectivity

The view of traditional molecular biology

Molecular Biology: "One protein — one function"

mutation => phenotype

Linear one-way dependencies: regulation at the DNA level, proteins follow

DNA => RNA => protein => phenotype

Structural Biology: "Protein structure determines its function" biochemical conditions => phenotype

No feedback, just re-action:



The Network View of Biology

Molecular Systems Biology: "It's both + molecular interactions"



=> highly connected network of various interactions, dependencies

=> study networks

Example: Proteins in the Cell Cycle

From Lichtenberg et al, Science 307 (2005) 724:

color coded assignment of proteins in time-dependent complexes during the cell cycle

=> protein complexes are transient

=> describe with a time dependent network



Major Metabolic Pathways





http://www.mvv-muenchen.de/de/netz-bahnhoefe/netzplaene/index.html

Metabolism of E. coli



Euler @ Königsberg (1736)



Can one cross all seven bridges once in one continuous (closed) path???

The Königsberg Connections



Turn problem into a graph:i) each neighborhood is a nodeii) each bridge is a linkiii) straighten the layout



Continuous path $<=> \le 2$ nodes with odd degree

Closed continuous path <=> only nodes with even degree

see also: http://homepage.univie.ac.at/franz.embacher/Lehre/aussermathAnw/Spaziergaenge.html

Draw in one stroke:



http://homepage.univie.ac.at/franz.embacher/Lehre/aussermathAnw/Spaziergaenge.html

Quantify the "Hairy Monsters"



Network Measures:

- No. of edges, nodes
 => size of the network
- Average degree <k>
 => density of connections
- Degree distribution P(k)
 => structure of the network
- Cluster coefficient C(k)
 => local connectivity
- Connected components
- => subgraphs

Lecture – Overview

Protein complexes: spatial structure => experiments, spatial fitting, docking
Protein association: => interface properties, spatial simulations
Protein-Protein-Interaction Networks: pairwise connectivity => data from experiments, quality check
PPI: static network structure => network measures, clusters, modules,
Gene regulation: cause and response => Boolean networks
Metabolic networks: steady state of large networks => FBA, extreme pathways
Metabolic networks / signaling networks: dynamics => ODEs, modules, stochastic effects

Literature

Lecture **slides** — available before the lecture

Suggested **reading**

=> check our web page

http://gepard.bioinformatik.uni-saarland.de/teaching/...

Textbooks





AN INTRODUCTION TO SYSTEMS BIOLOGY

DESIGN PRINCIPLES OF BIOLOGICAL CIRCUITS







=> check computer science library

How to pass this course

Schein =	you need to pass 3 out of 4 short tests and	
you need to pass the final exam		

Short tests: 4 tests of 45 min each planned are: first half of lectures V6,V12,V18,V24 ⇒average grade is computed from 3 best tests
If you have passed 2 tests but failed 1-2 tests, you can select one failed test for an oral re-exam.

- Final exam: written test of 120 min about assignments requirements for participation:
 - 50% of the points from the assignments
 - one assignment task presented @ blackboard
 - 2 short tests passed

planned: Fri, Feb. 8, 2013???

Assignments

Tutors:Christian Spaniol, Ruslan Akulenko,Mohamed Hamed, Nadine Schaadt,

Tutorial: Wed, 12:00-14:00, E2 I, room 007

10 assignments with 100 points each

Assignments are part of the course material (not everything is covered in lecture)

=> one solution for two students (or one)

- => hand-written or one **printable** PDF/PS file per email
- => content: data analysis + interpretation think!
- => no 100% solutions required!!!
- => attach the source code of the programs for checking (no suppl. data)
- => present one task at the **blackboard**

Hand in at the following Fri electronically until 13:00 or

printed at the start of the lecture.

Some Graph Basics

Network <=> Graph

Formal **definition**:

A graph G is an ordered pair (V, E) of a set V of vertices and a set E of edges.

G = (V, E)



undirected graph

directed graph

If $E = V^{(2)} =$ fully connected graph

Graph Basics II

Subgraph:

$$G' = (V', E')$$
 is a subset of $G = (V, E)$

Weighted graph:

Weights assigned to the edges





Practical question: how to define useful subgraphs?

Note: no weights for vertices => why???

Walk the Graph

Path = sequence of connected vertices

start vertex => internal vertices => end vertex

Two paths are **independent** (internally vertex-disjoint), if they have no internal vertices in common.

Vertices u and v are **connected**, if there exists a path from u to v. otherwise: disconnected

Trail = path, in which all edges are distinct

Length of a path = number of vertices || sum of the edge weights



How many paths connect the green to the red vertex?

How long are the shortest paths?

Find the four trails from the green to the red vertex.

How many of them are independent?

Local Connectivity: Degree

Degree k of a vertex = number of edges at this vertex Directed graph => discern k_{in} and k_{out}

Degree distribution P(k) = fraction of nodes with k connections





Basic Types: Random Network

Generally: N vertices connected by L edges

More specific: **distribute** the edges **randomly** between the vertices

Maximal number of links between N vertices:

$$L_{max} = \frac{N(N-1)}{2}$$

=> **propability** *p* for an edge between two randomly selected nodes:

$$p = \frac{L}{L_{max}} = \frac{2L}{N(N-1)}$$

=> average degree λ

$$\lambda = \frac{2L}{N} = p(N-1)$$

=> path lengths grow with log(N) => small world

Random Network: P(k)

Network with *N* vertices, *L* edges => Probability for a random link:

$$p = \frac{2L}{N(N-1)}$$

Probability that random node has links to k other particular nodes:

$$W_k = p^k (1-p)^{N-k-1}$$

Probability that random node has links to any k other nodes:

$$P(k) = \binom{N-1}{k} W_k = \frac{(N-1)!}{(N-k-1)! \, k!} \, W_k$$

Limit of large graph: $N \rightarrow oo, p = \lambda / N$

$$\lim_{N \to \infty} P(k) = \lim_{N \to \infty} \frac{N!}{(N-k)! \, k!} \, p^k \, (1-p)^{N-k}$$

$$= \lim_{N \to \infty} \left(\frac{N(N-1) \dots (N-k+1)}{N^k} \right) \, \frac{\lambda^k}{k!} \, \left(1 - \frac{\lambda}{N} \right)^N \, \left(1 - \frac{\lambda}{N} \right)^{-k}$$

$$= 1 \qquad \qquad 1 \qquad \qquad \frac{\lambda^k}{k!} \quad e^{-\lambda} \qquad \qquad 1$$

$$= \frac{\lambda^k}{k!} \, e^{-\lambda}$$

Bioinformatics 3 – WS 12/13

Random Network: P(k)

Many independently placed edges => **Poisson statistics**

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$



=> Small probability for $k >> \lambda$

k	$P(k \mid \lambda = 2)$
0	0.135335283237
1	0.270670566473
2	0.270670566473
3	0.180447044315
4	0.0902235221577
5	0.0360894088631
6	0.0120298029544
7	0.00343708655839
8	0.000859271639598
9	0.000190949253244
10	3.81898506488e-05

Basic Types: Scale-Free

Growing network a la Barabasi and Albert (1999):

- start from a small "nucleus"
- add new node with *n* links
- connect new links to existing nodes with probability αk (preferential attachment; β (BA) = 1) (k_i)

$$p_i = \left(rac{k_i}{\sum k_i}
ight)^eta$$

=> "the rich get richer"

Properties:

• power-law degree distribution:

$$P(k) \, \propto \, k^{-\gamma}$$
 with $\, \gamma \,$ = 3 for the BA model

- self-similar structure with highly connected hubs (no intrinsic length scale)
 => path lengths grow with log(log(N))
 - => very small world

The Power-Law Signature

Power law (Potenzgesetz) $P(k) \propto k^{-\gamma}$

Take log on both sides:

$$\log(P(k)) = -\gamma \log(k)$$

Plot log(P) vs. log(k) => straight line



Note: for fitting γ against experimental data it is often better to use the integrated P(k) => integral smoothes the data

$$\int_{k_0}^k P(k)dk = \left[-rac{k^{-(\gamma-1)}}{\gamma}
ight]_{k_0}^k$$

Scale-Free: Examples

The World-Wide-Web:

=> growth via links to portal sites

Flight connections between airports

=> large international hubs, small local airports

Protein interaction networks => some central, ubiquitous proteins



http://a.parsons.edu/~limam240/blogimages/16_full.jpg

Saturation: Ageing + Costs

Example: network of movie actors (with how many other actors did an actor appear in a joint movie?)

Each actor makes new acquaintances for ~40 years before retirement => limits maximum number of links

Example: building up a physical computer network

It gets more and more expensive for a network hub to grow further => number of links saturates



Hierarchical, Regular, Clustered...

Tree-like network with similar degrees => like an organigram

=> hierarchic network

All nodes have the same degree and the same local neighborhood => regular network



P(k) for these example networks? (finite size!)

Note: most real-world networks are somewhere in between the basic types

Summary

What you learned **today**:

- => **networks** are everywhere
- => how to get the **"Schein**" for BI3

=> basic network **types** and **definitions**:

random, scale-free, degree distribution, Poisson distribution, ageing, ...

Next lecture:

- => clusters, percolation
- => algorithm on a graph: Dijkstra's shortest path algorithm
- => looking at graphs: graph layout

Further Reading: R.Albert and A–L Barabási, "Statistical mechanics of complex networks" *Rev. Mod. Phys.* **74** (2002) 47-97