

# Graphical models in Systems Biology

Francesc Rosselló (UIB)

ISBBC'13. Madrid, September 2013



# Networks

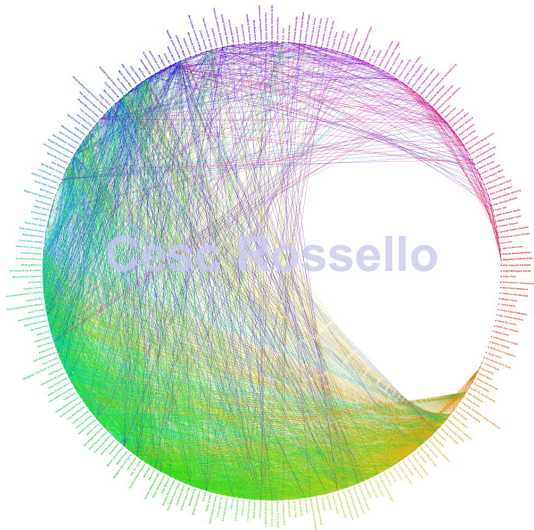
## Communication networks: Flight networks



Source: [http://content.united.com/ual/asset/UAL\\_NA\\_Map.pdf](http://content.united.com/ual/asset/UAL_NA_Map.pdf)

# Networks

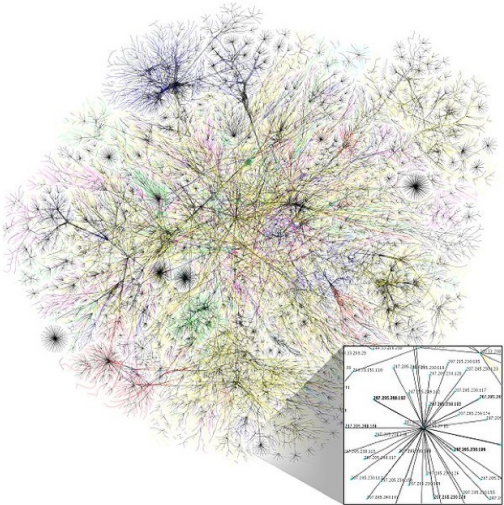
Social networks: Facebook



Source: generated with <http://friend-wheel.com>

# Networks

## Technological networks: Internet

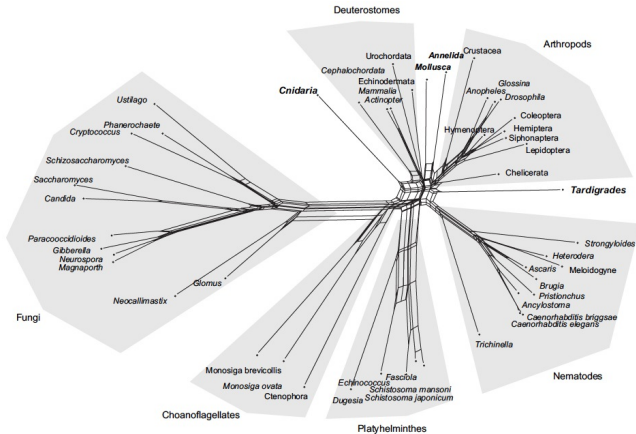


Source: Internet Mapping Project, <http://www.lumeta.com>



# Networks

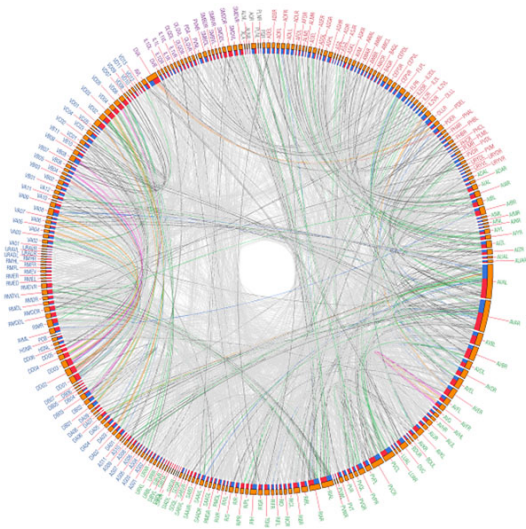
## Biological networks: Phylogenetic trees and networks



Source: D. Huson, D. Bryant. *Mol. Biol. Evol.* 23 (2006), 254–267

# Networks

## Biological networks: Neuronal networks

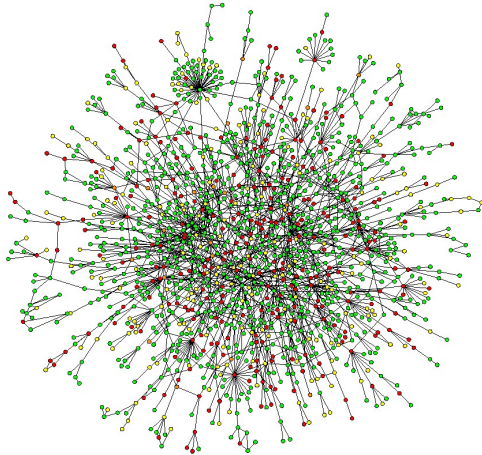


Source: Kim et al. *BMC Neuroscience* 13(Suppl 1):P110 (2012)



# Networks

Biological networks: Protein-Protein Interaction (PPI) networks



# Systems biology

- **Biomolecular interaction networks** are an integral part of Systems Biology
- Goal of **Systems Biology**: To understand
  - Organization of biological systems in terms of the interactions of cellular components
  - Dynamics of these interactions

# Biological networks

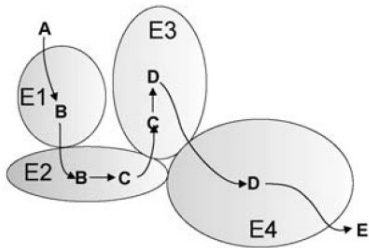
- Intra-cellular networks
  - PPI networks
  - Metabolic networks
  - Gene regulatory networks
  - Cell signaling networks
  - Gene expression networks
- Other biological networks
  - Neuronal synaptic connection networks
  - Protein structure networks
  - Brain functional networks
  - Ecological food webs
  - Phylogenetic networks
  - Disease–gene association networks

# PPI networks

**Proteins** are involved in most biological processes in the living cell

Usually, to be functionally active, proteins assembly into **protein complexes**

**Multienzyme complex:** A is converted to E through four sequentially acting enzymes E1–E4 assembled in a complex so that intermediate products do not diffuse

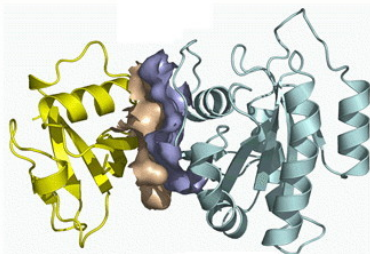


Source: B. Junker, F. Schreiber, *Analysis of biological networks*. Wiley (2008)

# PPI networks

- **Proteome:** set of proteins produced by a cell
- **Interactome:** set of all interactions between them

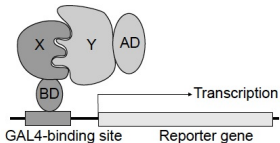
**Protein-protein interactions**  
(**PPI**) usually refer to physical interaction, i.e., binding



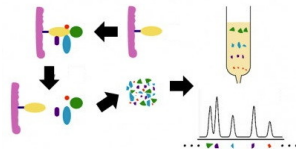
# PPI networks

PPI **detected** through high throughput screening methods,

- **Yeast 2-Hybrid System**



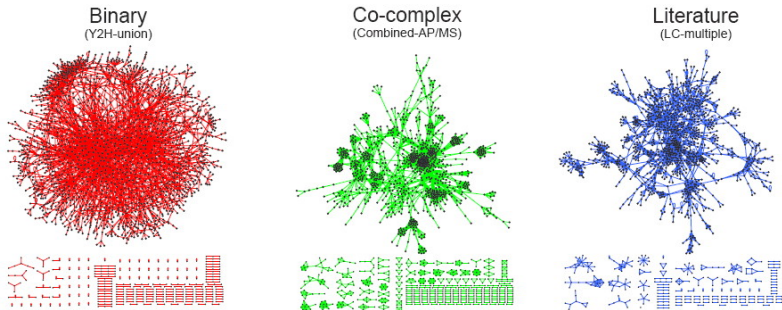
- **Affinity Capture of Protein Complexes**



- **Synthetic lethality**: When mutating 2 genes is lethal, corresponding proteins must be functionally related

# PPI networks

PPI networks of yeast obtained from different datasets



Source: [http://interactome.dfci.harvard.edu/S\\_cerevisiae/](http://interactome.dfci.harvard.edu/S_cerevisiae/)

# PPI networks

PPI **predicted** through computational methods

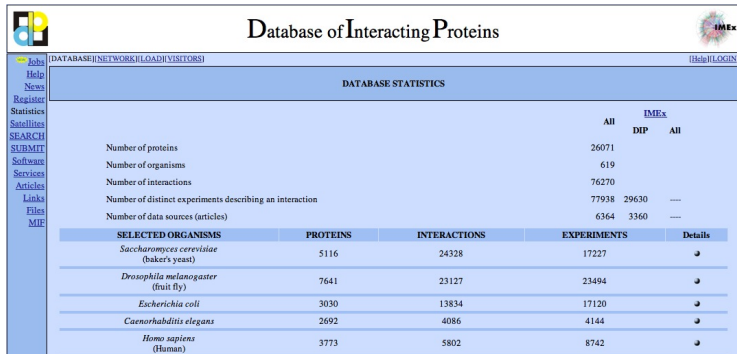
- Protein docking
- Simultaneous gene expression
- Comparing with other organisms
- Link prediction algorithms in complex networks  
(L. Lü, T. Zhou. *Physica A* 390 (2011), 1150-1170)



# PPI networks

Available on-line for a few organisms in different databases:

- DIP <http://dip.doe-mbi.ucla.edu>
- BioGRID <http://thebiogrid.org>
- HPRD (Human Protein Reference Database)  
<http://www.hprd.org>



Database of Interacting Proteins

Jobs Help News Register Statistics Satellites SEARCH SUBMIT Software Services Articles Links Files MIP

(DATABASE) (NETWORK) (LOAD) (VISITORS) (Help) (LOGIN)

### DATABASE STATISTICS

	IMEx		
	All	DIP	All
Number of proteins	26071		
Number of organisms	619		
Number of interactions	76270		
Number of distinct experiments describing an interaction	77938	29630	---
Number of data sources (articles)	6364	3360	---

SELECTED ORGANISMS	PROTEINS	INTERACTIONS	EXPERIMENTS	Details
<i>Saccharomyces cerevisiae</i> (baker's yeast)	5116	24328	17227	➤
<i>Drosophila melanogaster</i> (fruit fly)	7641	23127	23494	➤
<i>Escherichia coli</i>	3030	13834	17120	➤
<i>Caenorhabditis elegans</i>	2692	4086	4144	➤
<i>Homo sapiens</i> (Human)	3773	5802	8742	➤

# PPI networks

Issues with completeness:

- **HPRD** contains a PPI network of 41,000 interactions between 8,500 proteins. A human cell has 20,000–25,000 proteins, and 250,000 predicted interactions

Issues with bias:

- Identified interactions are not random, but opportunistic: highly expressed proteins, essential proteins, . . .

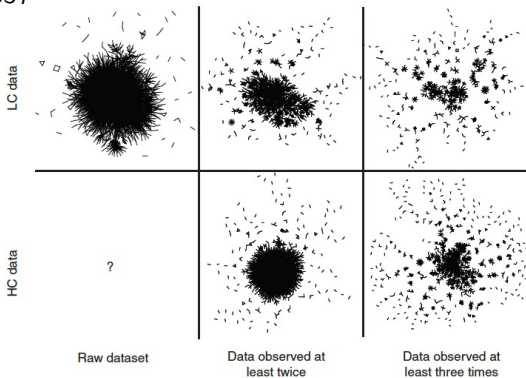
Issues with accuracy:

- The PPI network of yeast published by Ito *et al* in 2001 had a 80% of false positives

# PPI networks

Issues with accuracy:

- “About 80,000 interactions between yeast proteins are currently available [...] only 2,400 are supported by more than one method.” (C. von Mering *et al.* *Nature* 417 (2002), 399–403)



**LC data:** T. Reguly *et al.* *J. Biol.* 5(4):11 (2006); **HC data:** N. Batada *et al.* *PLoS Biol.* 4, e317 (2006)  
Source: L. Hakes *et al.* *Nature Biotech.* 26 (2008), 69–72.

# PPI networks

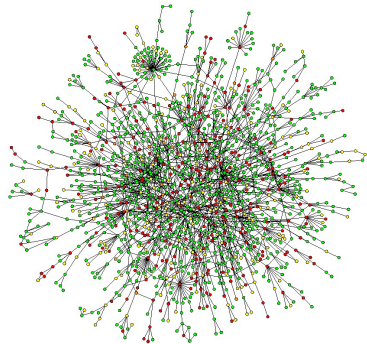
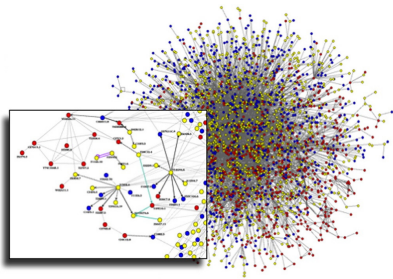
Usual static representation as **undirected graphs**:

- **Nodes**: proteins
- **Edges**: (binary) interactions

Multiple interactions are modeled by hyperedges, or inserted in the binary network as stars or cliques

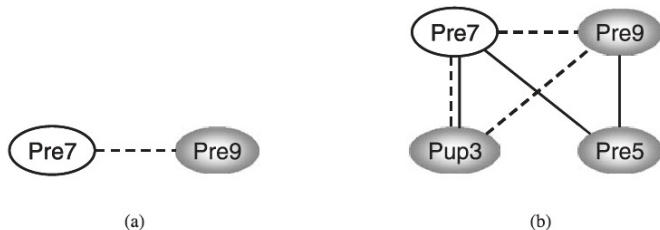
**Open problem**: Suitable models of:

- transient interactions
- spatial heterogeneity



# PPI networks

Modeling uncertainty through Markov relational networks

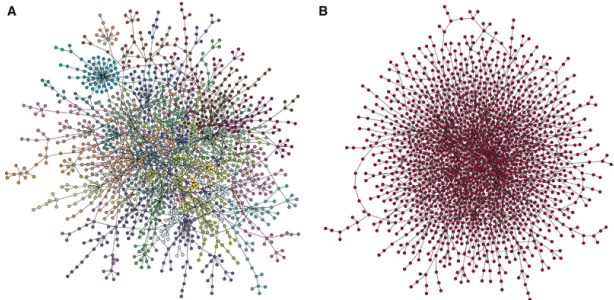


**FIG. 1.** Dependencies between interactions can be used to improve predictions. (a) A possible interaction of two proteins (Pre7 and Pre9). Pre9 is localized in the cytoplasm and in the nucleus (light gray) and Pre7 is not annotated to be in either one of those. This interaction was predicted by a computational assay (Sprinzak and Margalit, 2001) (dashed line). This evidence alone provides weak support for an interaction between the two proteins. (b) Two additional proteins Pre5 and Pup3. These were found to interact with Pre9 and Pre7 either by a computation assay (Sprinzak and Margalit, 2001) (dashed line) or experimental assays (Mewes *et al.*, 1998) (solid line). The combined evidence gives more support to the hypothesis that Pre7 and Pre9 interact.

Source: A. Jaimovich, PhD Thesis, Hebrew University (2010)

# PPI networks: Challenges

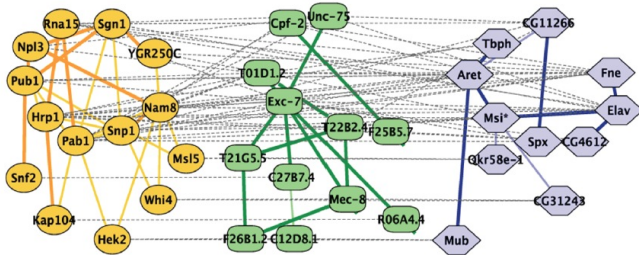
- Studying topological properties
  - Finding central proteins (**hubs**,...)
  - Finding functional modules (**motifs**,...)
  - Obtaining robust biological conclusions
- Modelling their topology and growth as a complex network



Sources: R. Guimerà, M. Sales-Pardo, *Mol. Sys. Biol.* 2:42 (2006)  
V. Colliza et al. *Nat. Phys.* 2 (2006), 110–115

# PPI networks: Challenges

- Network alignment
- Prediction of interactions
- Prediction of protein function and structure
- Finding conserved functional modules across species



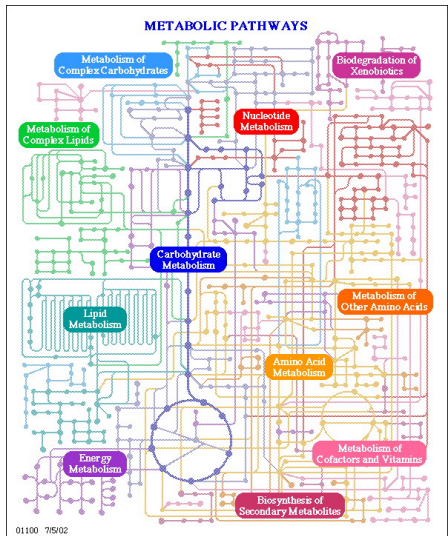
Alignment produced with Cytoscape <http://www.cytoscape.org>  
Source: R. Sharan *et al.* *PNAS* 102 (2004), 1974–1979

# Metabolic networks

**Metabolism:** the chemical system that generates the essential components for life

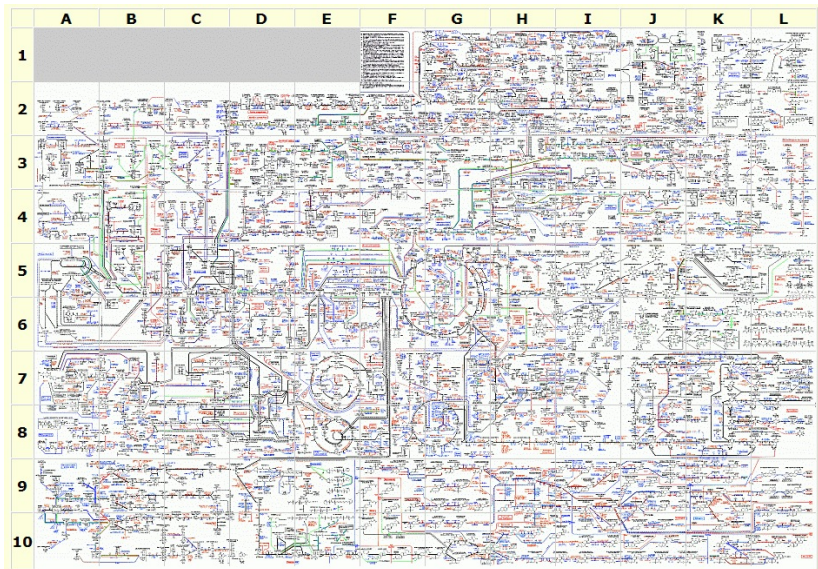
**Metabolic network:** the set of chemical reactions of metabolism and the regulatory interactions that guide them

Metabolic networks are dissected into **metabolic pathways**





# Metabolic networks

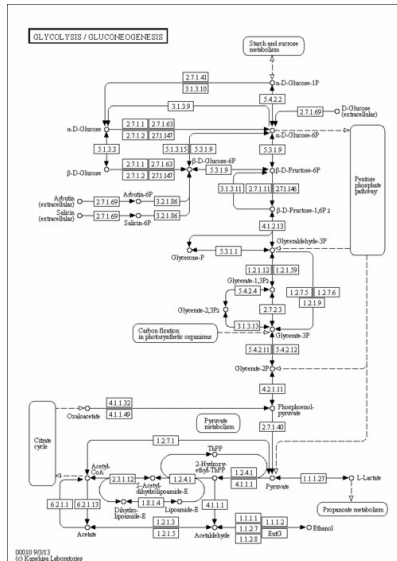


Source: Roche Applied Science Biochemical Pathways chart <http://web.expasy.org/pathways/>

# Metabolic networks

**Metabolic pathway:** a subsystem of a metabolic network dealing with some specific process:

- a network of **chemical reactions**, linked to each other, catalyzed by **enzymes** where **substrates** are transformed into **products**
- the network **kinetics** is represented by the rate equation associated with each reaction

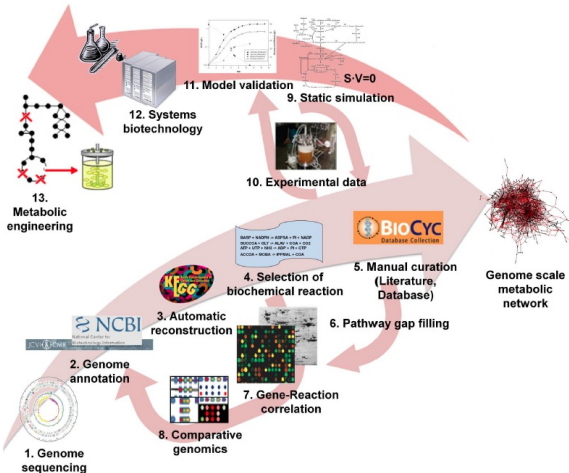


Human glycolysis pathway (Source: KEGG)

# Metabolic networks

Constructed:

- Partially experimentally
- Partially from genome sequencing and annotation



# Metabolic pathways

Available on-line for many organisms in different databases:

- KEGG (Kyoto Encyclopedia of Genes and Genomes)  
(<http://www.genome.jp/kegg/pathway.html>)
- BioModels (<http://www.ebi.ac.uk/biomodels-main>)
- MetaCyc (Encyclopedia of Metabolic Pathways)  
(<http://metacyc.org>)
- ...

No database of complete metabolic networks of complex organisms, only for specific organisms

- Human <http://humanmetabolism.org>
- Arabidopsis  
[http://www.metnetdb.org/MetNet\\_db.htm](http://www.metnetdb.org/MetNet_db.htm)
- Yeast <http://www.comp-sys-bio.org/yeastnet/>
- ...

# KEGG

- At present it contains around 95,000 pathways
- Pathways are represented by maps with additional information
- Models are coded in KGML (KEGG Markup Language)



## KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE MODULE DISEASE DRUG KO GENOME GENES LIGAND DBGET

Select prefix:   Enter keywords:   [Help](#)

### Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps and update history) representing our knowledge on the molecular interaction and reaction networks for:

1. **Metabolism**  
Global map Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
2. **Genetic Information Processing**
3. **Environmental Information Processing**
4. **Cellular Processes**
5. **Organismal Systems**
6. **Human Diseases**

and also on the structure relationships (KEGG drug structure maps) in:

7. **Drug Development**

### Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretation of higher-level systemic functions.

- Search Pathway - basic pathway mapping tool
- Search&Color Pathway - advanced pathway mapping tool
- Color Pathway - selected pathway map coloring tool

### 1. Metabolism

#### 1.0 Global and overview maps

Metabolic pathways	[KEGG Atlas]	KEGG modules
Biosynthesis of secondary metabolites	[KEGG Atlas]	KEGG reaction modules
Microbial metabolism in diverse environments	[KEGG Atlas]	
Carbon metabolism <i>New!</i>	[KEGG Atlas]	
2-Oxocarboxylic acid metabolism	[KEGG Atlas]	
Biosynthesis of amino acids	[KEGG Atlas]	
Degradation of aromatic compounds	[KEGG Atlas]	

#### 1.1 Carbohydrate metabolism

Glycolysis / Gluconeogenesis		Enzymes
Citrate cycle (TCA cycle)		Compounds with biological roles

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



# Metabolic networks

Simple, static representation as directed graphs

## Metabolite graphs

- **Nodes:** metabolites
- **Arcs:** Existence of a reaction

## Reaction graphs

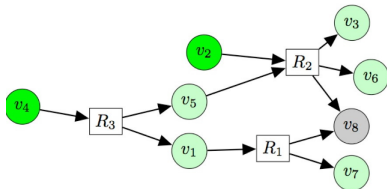
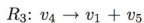
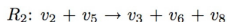
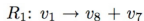
- **Nodes:** reactions
- **Arcs:** A product of the source is a substrate of the end

Allow the study of their topology as complex networks

# Metabolic pathways

Static representation of metabolic pathways as **hypergraphs**

- Nodes: metabolites
- Hyperarcs: reactions
- Hyperarcs' labels: enzymes



$$\mathbf{S} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

Stoichiometric matrix

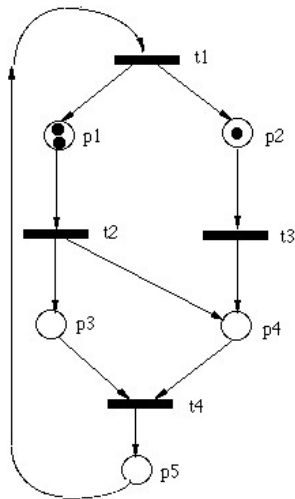


# Metabolic pathways as Petri nets

Petri nets allow a natural representation of metabolic pathways and their dynamics.

Roughly, enriched bipartite graph version of hypergraph

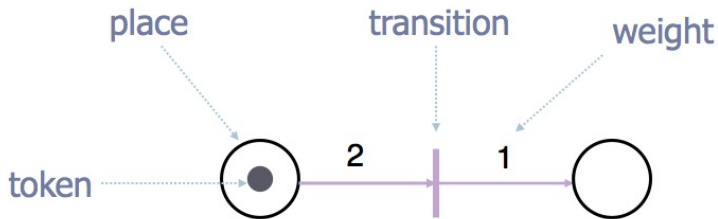
There is a clear correspondence between Petri net concepts and metabolic pathways concepts



# Metabolic pathways as Petri nets

A **Petri net** is a directed, bipartite graph

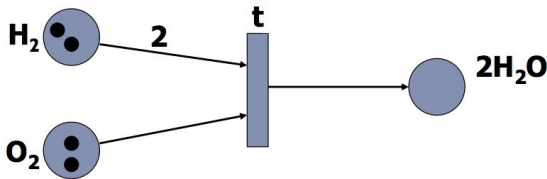
- Arcs connecting **places** (metabolites) and **transitions** (reactions)
- Arcs weighted in  $\mathbb{N}^+$  (**multiplicities**) and places weighted in  $\mathbb{N}$  (**tokens**)



# Metabolic pathways as Petri nets

A transition may fire when each of its input places contain at least as many tokens as the weight of the corresponding input arc

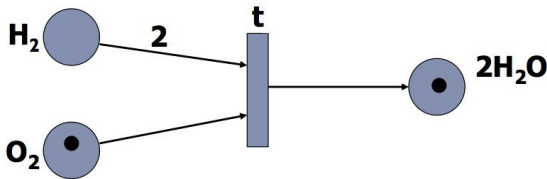
The firing of a transition changes the inputs and output places's weights by moving tokens according to the arcs' weights



# Metabolic pathways as Petri nets

A transition may fire when each of its input places contain at least as many tokens as the weight of the corresponding input arc

The firing of a transition changes the inputs and output places's weights by moving tokens according to the arcs' weights



# Metabolic pathways as Petri nets

Many useful extensions:

- **Inhibitor** input arcs (with them, Petri nets  $\equiv$  Turing machines)
- **Functional** Petri nets have arc weights defined as functions of input tokens
- **Continuous** Petri nets may have real-valued numbers of tokens (**marks**: concentrations)
- **Timed** Petri nets assign deterministic time frames to transitions
- **Stochastic** Petri nets assign delays to transitions with a probability distribution depending on **transition rates**

# Metabolic pathways as Petri nets

Correspondence between metabolic pathways and Petri nets

Pathway elements	Petri net elements
Metabolites, enzymes, compounds	Places
Reactions, interactions	Transitions
Substrates, reagents	Input places
Reaction products	Output places
Stoichiometric coefficients	Arc weights
Metabolites, enzymes, compounds quantities	Number of tokens on places
Kinetic laws of reactions	Transition rates

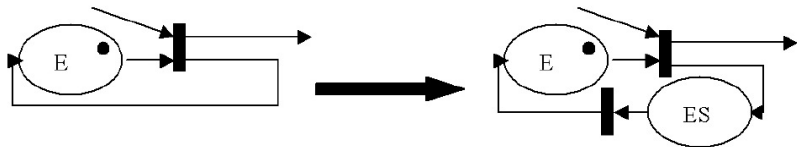
  

Stoichiometric matrix	Incidence matrix
Biochemical systems	Petri nets
Conservation relations	P-invariants
Semi-positive (non-negative) conservation relations	Semi-positive P-invariants
Steady-state flux distributions	T-invariants
Elementary flux modes	Minimal T-invariants

# Metabolic pathways as Petri nets

Some decisions:

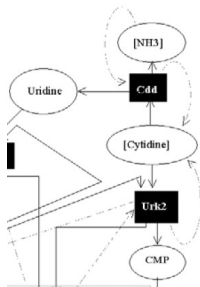
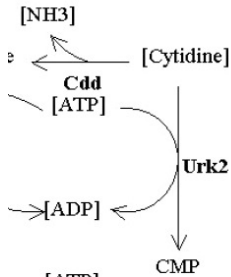
- What to do with enzymes? (they would create self-loops)
  - Labels of transitions
  - Break loops by adding new places enzyme-substrate in loops



# Metabolic pathways as Petri nets

Some decisions:

- What to do with external (source or sink) metabolites?
  - Allow “infinite” tokens in them
  - Allow rules that do not “consume ” or do not “produce” tokens
- Allow self-loops for external metabolites: arcs back to transition from sinks and from transitions to sources





# Metabolic pathways as Petri nets

Some decisions:

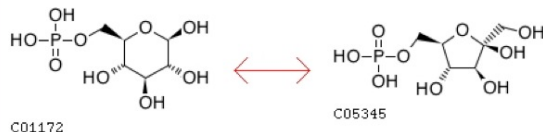
- Omit or maintain ubiquitous molecules (water, ADP, ATP, etc.)?
- Represent reversible reactions as two reactions
- Use colors to model spatial information on substances

[Survey](#): P. Baldan *et al*, *Nat. Comput.* 9 (2010), 955–989



# Metabolic pathways as Petri nets

**MPath2PN**: translates metabolic pathways (in KEGG format) into Petri nets in PNML (Petri Net Markup Language)

<b>Entry</b>	R03321	Reaction
<b>Name</b>	beta-D-Glucose 6-phosphate ketol-isomerase	
<b>Definition</b>	beta-D-Glucose 6-phosphate $\rightleftharpoons$ beta-D-Fructose 6-phosphate	
<b>Equation</b>	C01172 $\rightleftharpoons$ C05345	
	 <p>C01172 <span style="margin-left: 300px;">C05345</span></p>	
<b>RPair</b>	RP02940	C01172_C05345 main
<b>Enzyme</b>	5.3.1.9	
<b>Pathway</b>	rn00010 Glycolysis / Gluconeogenesis rn01100 Metabolic pathways rn01110 Biosynthesis of secondary metabolites rn01120 Microbial metabolism in diverse environments	
<b>Orthology</b>	K01810 glucose-6-phosphate isomerase [EC:5.3.1.9] K06859 glucose-6-phosphate isomerase, archaeal [EC:5.3.1.9] K13810 transaldolase / glucose-6-phosphate isomerase [EC:2.2.1.2 5.3.1.9]	



# Metabolic pathways as Petri nets

**MPath2PN**: translates metabolic pathways (in KEGG format) into Petri nets in PNML (Petri Net Markup Language)

Available at <http://www.dsi.unive.it/~biolab>

# Metabolic pathways as Petri nets

Modelling metabolic pathways as Petri nets allows to use Petri nets tools and techniques to

- Study the pathway characteristics and behavior
- Simulate the pathway on an initial configuration
- Compare metabolic pathways taking into account structural and behavioral aspects

# Metabolic pathways as Petri nets

Modelling metabolic pathways as Petri nets allows to use Petri nets tools and techniques to

- Study the pathway characteristics and behavior
- Simulate the pathway on an initial configuration
- Compare metabolic pathways taking into account structural and behavioral aspects

P-systems are related to Petri nets (P. Frisco, 2006) and they are better at modelling metabolic pathways, but they lack these techniques for the first and third problems

# Basic bibliography

