Graph comparison applications in Systems Biology

Mercè Llabrés (UIB)

ISBBC 17 Valencia
June 28–30, 2017

Systems biology

- *Systems Biology*: computational and mathematical modeling of complex biological systems
- Main goal: improved understanding of biology and its mechanisms, in particular of the functions and interactions of key elements of living systems (DNA, RNA, proteins, cells)
  - Biological systems are modeled as networks: *Biological Networks*
  - A *biological network* is a set of entities connected by some kind of links

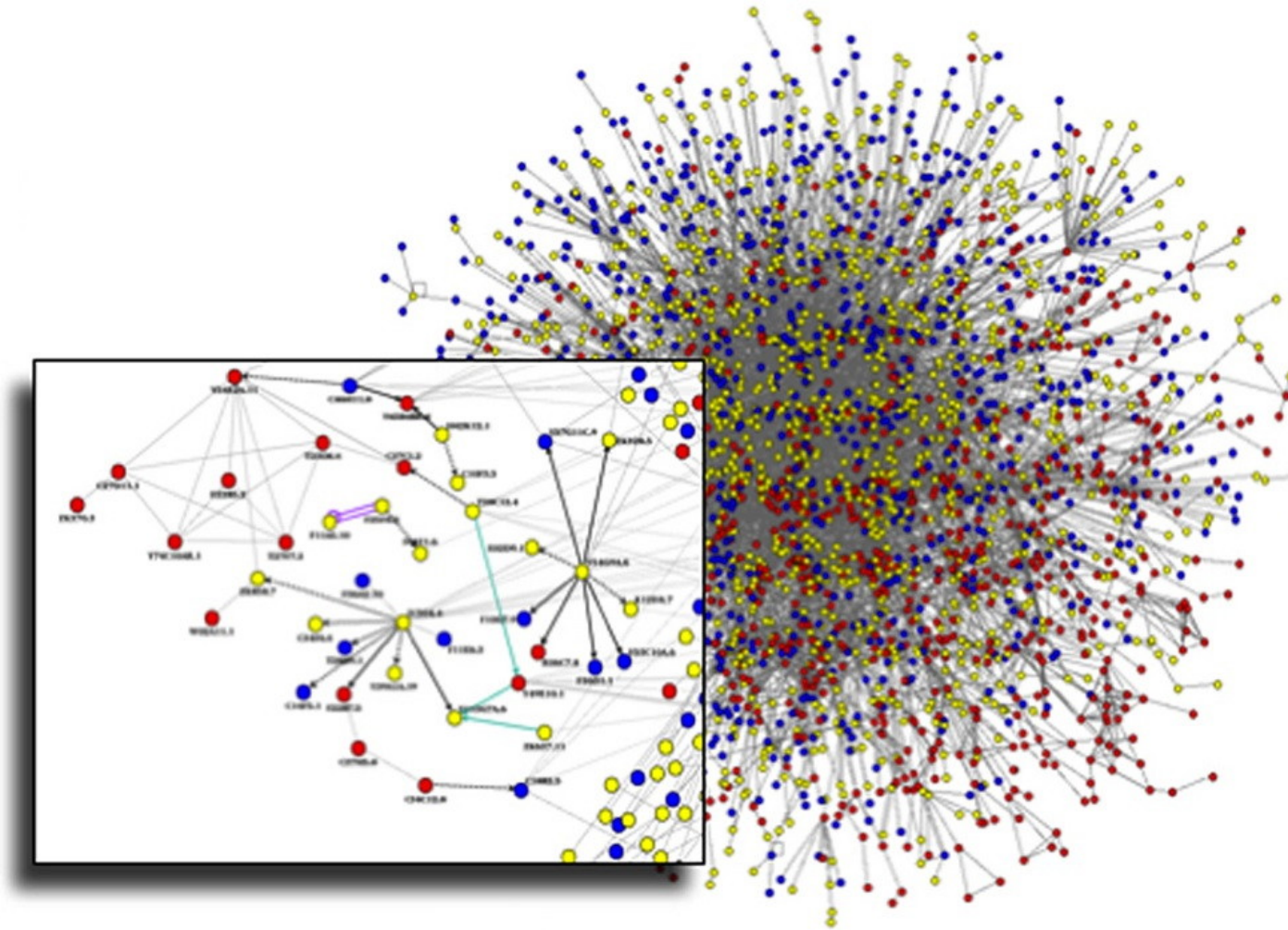
Biological networks

- Intra-cellular networks
  - *PPI networks*
  - *Metabolic networks*
  - Transcriptional regulation networks
  - Cell signaling networks
  - Gene expression networks
- Other biological networks
  - Neuronal synaptic connection networks
  - Protein structure networks
  - Brain functional networks
  - Phylogenetic networks
  - Disease–gene association networks
  - Trophic networks

PPI networks

Usual static representation as *undirected graphs*:

- Nodes: proteins
- Edges: (binary) interactions

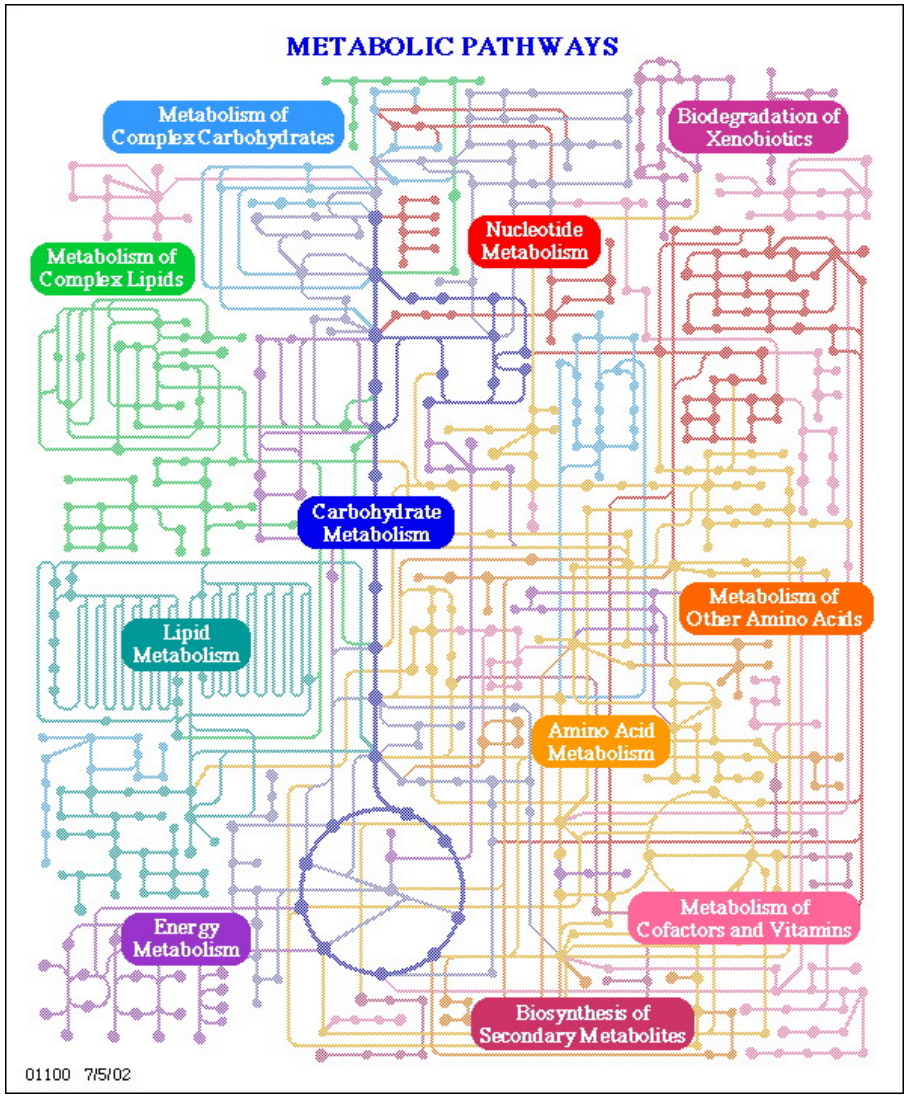


Metabolic networks

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

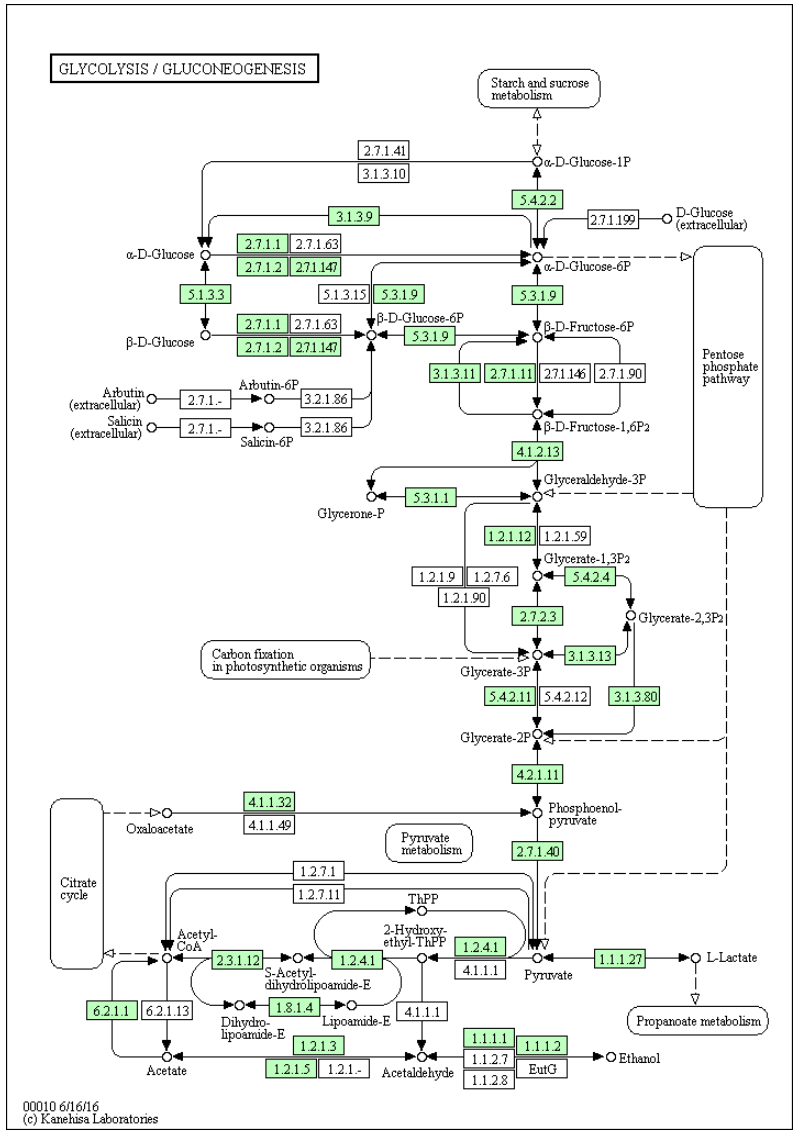
Metabolic networks are dissected into metabolic pathways



Metabolic pathways

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

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



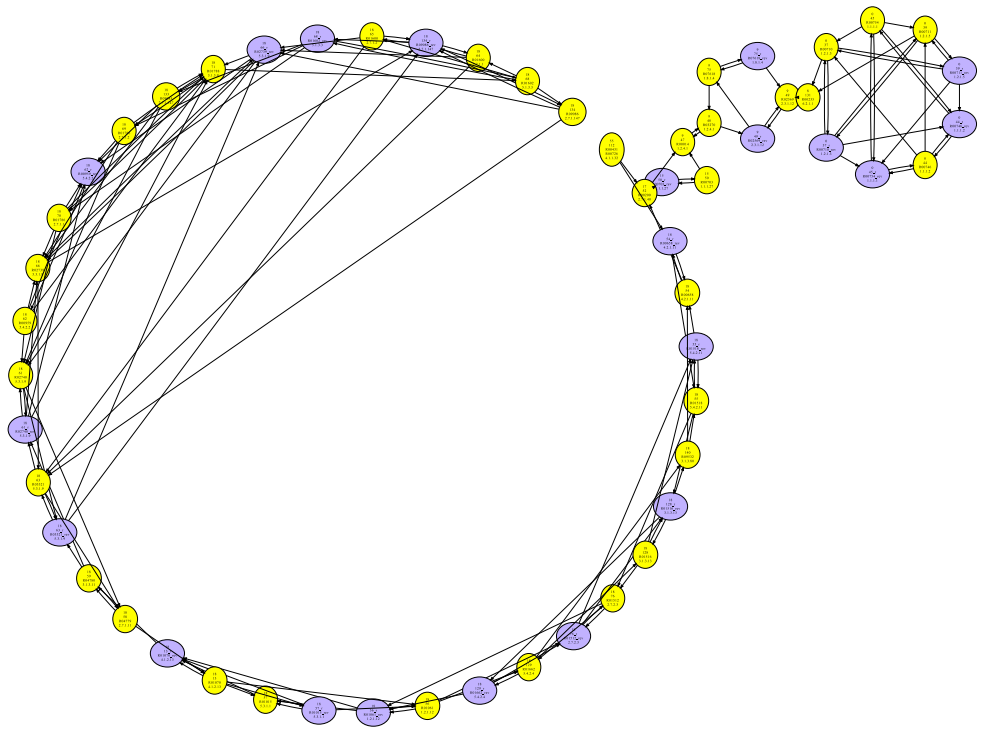
Human glycolysis pathway (Source: KEGG)

Metabolic pathways

Static representation of metabolic pathways as directed graphs
$G = (V, E)$, $E \subseteq V \times V$

Reaction Graph

- Nodes: $R_i = (I_i, E_i, O_i)$ set of reactions
- Arcs: $(R_i, R_j) \in E$ if, and only if, there exists one metabolite c such that $c \in O_i \cap I_j$



Source: R. Alberich *et al.* *PLOS ONE*
https://doi.org/10.1371/journal.pone.0177031

Metabolic pathways

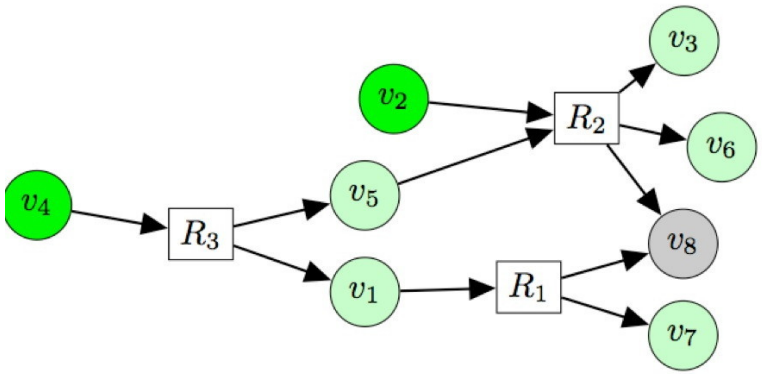
Static representation of metabolic pathways as hypergraphs

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

$$R_1\colon v_1 \to v_8 + v_7$$
$$R_2\colon v_2 + v_5 \to v_3 + v_6 + v_8$$
$$R_3\colon v_4 \to v_1 + v_5$$

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

Stoichiometric matrix



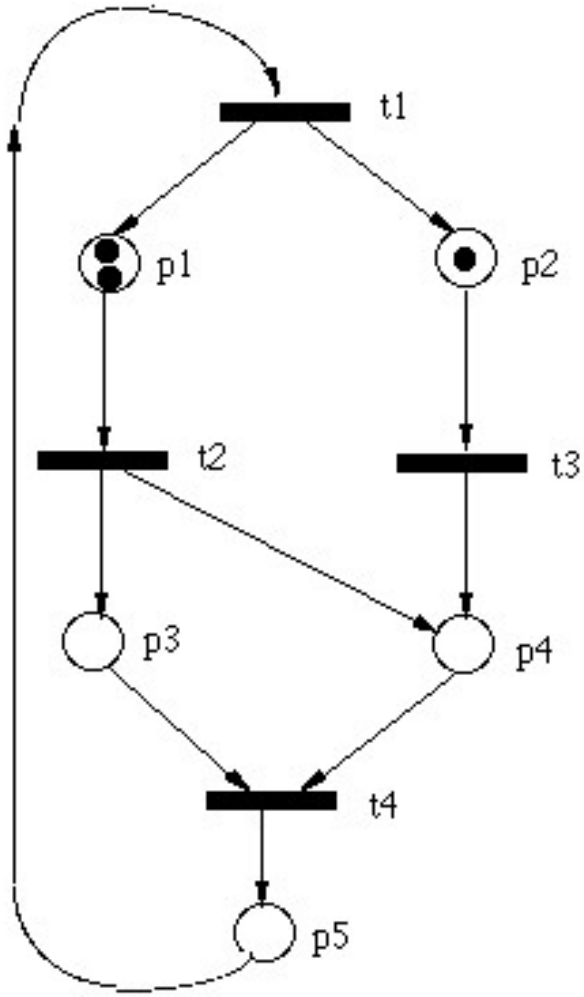
Source: P. Carbonell *et al.* *BMC Sys. Biol.* 6:10 (2012)

Metabolic pathways as Petri nets

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

Roughly, hypergraphs with *transitions* instead of hyperarcs

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



Biological networks as graphs

Graph topology	$\leftrightarrow$	Biological properties
Complete subgraphs in PPIN	$\leftrightarrow$	protein complexes
Cut nodes in Reaction Graphs	$\leftrightarrow$	essential reactions

Computational problem: networks are huge

- Modularization of the networks is needed.

Biological networks size

Pathways (hsa)	Nodes	Edges
Purine metabolism pathway	141	527
Glycolisis/Glucogenesis	35	157
Entire metabolic network	1812	7074

Metabolic pathways from the KEGG database.

Biological networks size

- PPI networks are huge: $\sim$ 6000 nodes, $\sim$ 20000 edges.

Network	Nodes	Edges
C. Elegans	2,805	4,495
D. Melanogaster	7,518	25,635
H. Sapiens	9,633	34,327
S. Cerevisiae	5,499	31,261

PPI networks from the IsoBase database.

- We need computer tools to analyze and compare PPI networks.

Why do we need to compare biological networks?

PPI networks comparison

Why do we need to compare PPI networks?

- *Protein networks in disease, like for instance cancer, increase considerably their number of edges. That is, the amount of proteins that interact is much bigger.*

  Ideker T, Sharan R *Protein networks in disease*. Genome Res. 18:644–652 (2008)

- Detection of proteins that causes the disease

Metabolic networks comparison

Why do we need to compare metabolic networks?

- Networks similarities as well as networks dynamics are a key point to identify the causes of certain diseases.

Braun P, Rietman E, Vidal M *Networking metabolites and diseases*. PNAS 2008 105 (29) 9849-9850; doi:10.1073/pnas.0805644105

Toledo, Jon B. et al. *Metabolic network failures in Alzheimer's disease – A biochemical road map*. Alzheimer's & Dementia. The Journal of the Alzheimer's Association.

Metabolic networks comparison

Why do we need to compare metabolic networks?

- Similarity parts between pathways provide insights for drug target identification.

Sridhar P, Kahveci T, Ranka S *An iterative algorithm for metabolic network-based drug target identification*. Pac. Symp. Biocomput. 2007, 12:88–99.

Watanabe N, Cherney M, van Belkum et al M *Crystal structure of LL-diaminopimelate aminotransferase from Arabidopsis thaliana: a recently discovered enzyme in the biosysthesis of L.lysine by plants and Chlamydia.*. J. Mol. Biol. 2007, 371:685–702.

Biological Networks comparison by

- Definition of similarity functions (nodes degree, cut nodes, bridges edges, complete subgraphs, etc..)
- Network Alignment (injective application that assigns the nodes from one graph to another)

Main Goal: to infer useful biological information from one network to another (biological function, evolutionary relationships,..)

Biological networks aligners

Local alignments

- Aimed at finding local regions with the same network structure: low coverage.
- Local alignments are mutually inconsistent.

Global alignment

- Aimed at finding the best overall alignment between protein interaction networks.
- Designed to either obtain a high number of conserved edges (relations) or a functional consistence between aligned nodes.

Biological networks aligners

Pairwise alignments

- Two biological networks are aligned.

Multiple alignments

- More than two biological networks are aligned.

PPI networks...

PPI networks aligners

- The right balance between network topology and biological information, is one of the most difficult and key points in every protein interaction network alignment algorithm.

The global pairwise alignment is usually obtained by defining

- node score similarity based on network topology and sequence similarity
- final node assignment maximizing an overall score

PPI Network Aligners (pairwise global)

- PINALOG Node similarity is sequence similarity. Alignment obtained by extending seeds.

  Phan HT, Sternberg MJ. *PINALOG: a novel approach to align protein interaction networks–implications for complex detection and function prediction.* Bioinformatics. 2012 May 1;28(9):1239-45. doi: 10.1093/bioinformatics/bts119.

- SPINAL Node similarity is sequence similarity. Alignment obtained by solving weighted bipartite matching problem.

  Aladag AE, Erten C. *SPINAL: scalable protein interaction network alignment.* Systems biology Vol. 29 no. 7 2013, pages 917âĂŞ924 doi:10.1093/bioinformatics/btt071.

- AligNet Node similarity is sequence similarity and node degree. Alignment obtained by first constructing an overlapping clustering, the clusters alignment and a final extension.

  Alcalá et al. *AligNet:Alignment of Protein-Protein Interaction Networks*. Submited to Algorithms for Molecular Biology

PPI Network Aligners (pairwise local)

Based at finding conserved substructures: graph isomorphism

- NetAlign

Liang Z, Xu M, Teng M, Niu L. *NetAlign: a web-based tool for comparison of protein interaction networks.*. Bioinformatics. 2006 Sep 1;22(17):2175-7. Epub 2006 Jun 9.

- MaWISh

KoyutuÌĺrk,M. et al. *Pairwise alignment of protein interaction networks*. J. Comput. Biol., 13, 182âĂŞ199.

PPI Network Aligners (multiple)

- IsoRankN (global)
- Graemlin (local)
- NetworkBLAST-M (local)

Alignment applications – Edge conservation

The edge correctness ratio of a mapping $\mu : G \to G'$ is the ratio of the edges that are preserved by μ, and it is defined by

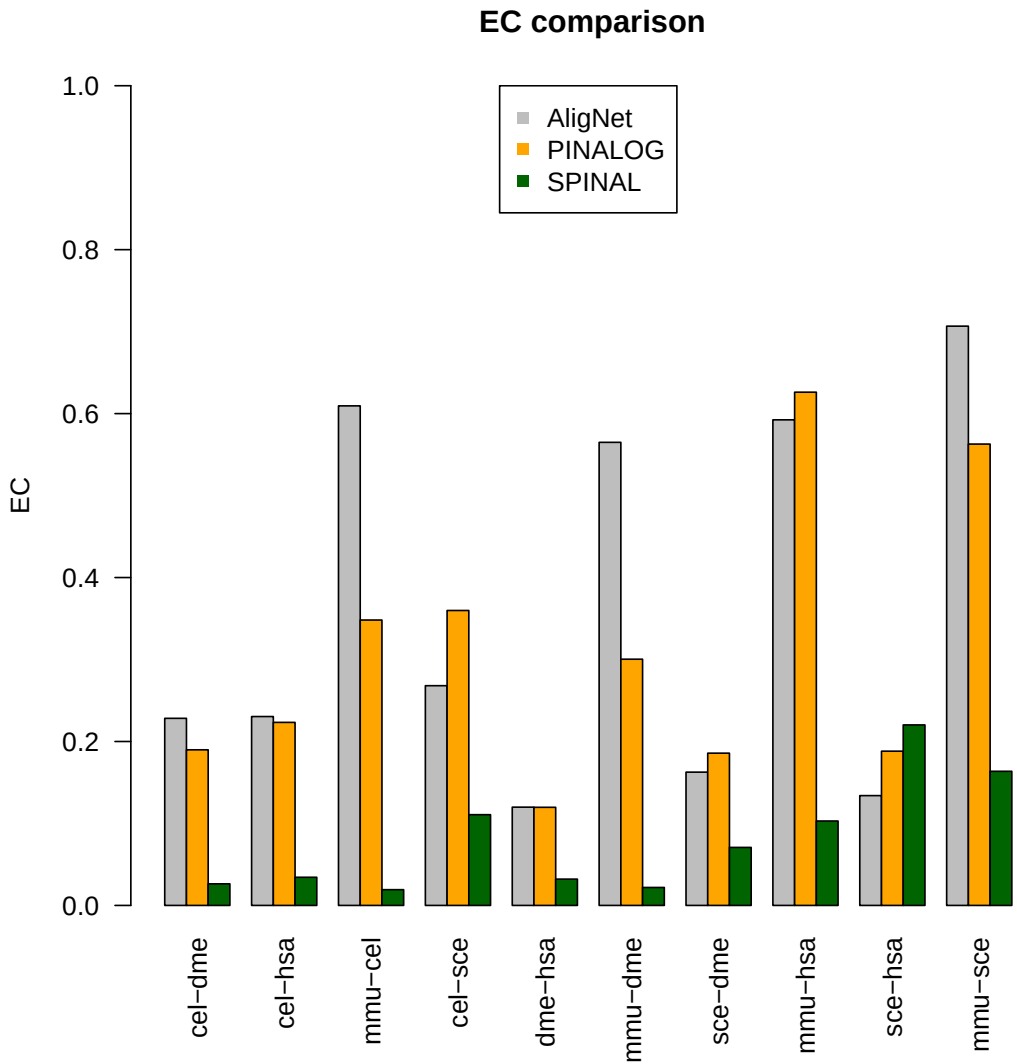
$$EC(\mu) = \frac{\left|\left\{\{u, v\} \in E \mid \{\mu(u), \mu(v)\} \in E'\right\}\right|}{|E|}.$$

Main applications

- Inferring diseases/noisy data
- Prediction of Protein-Protein interactions

Alkan F, Erten C *SiPAN:simultaneous prediction and alignment of protein.protein interaction networks*. Bioinformatics. 31(14):2356–2363 (2015)

Alignment Testing: Edge correctness ratio



Alignment applications – Protein assignment

The functional coherence value, or *GO consistency*, of a mapping $\mu : G \to G'$ is defined as

$$FC(\mu) = \frac{\sum_{u \in V} FS(u, \mu(u))}{|V|}$$

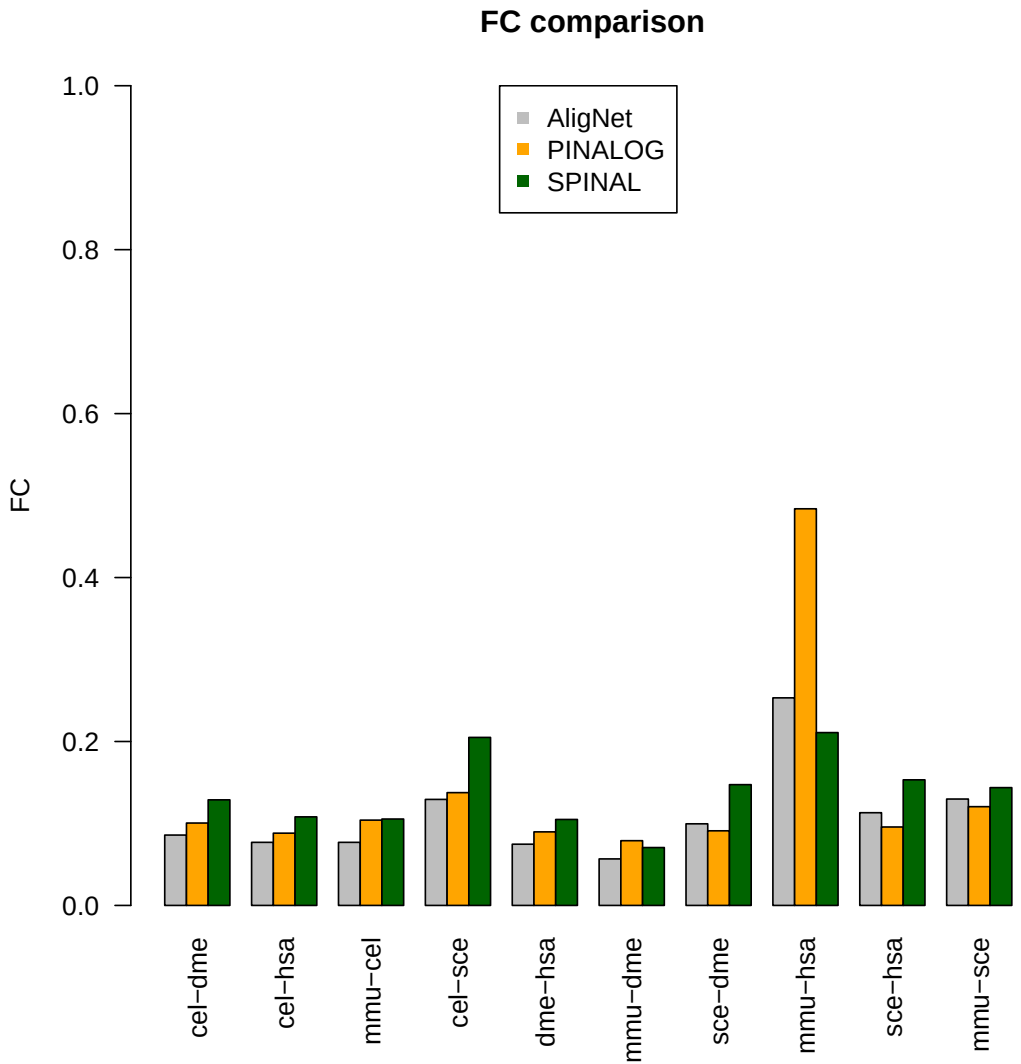
where, on its turn, the functional similarity score FS is defined by

$$FS(u, u') = \frac{|GO(u) \cap GO(u')|}{|GO(u) \cup GO(u')|},$$

with $GO(u)$ and $GO(u')$ the sets of GO annotations of the proteins u and u', respectively.

Application: Protein Function Prediction

Alignment Testing: functional coherence value



Multiple alignment applications

- Discovering Motifs
- Meaninful protein complexes
- Conserved Interactions
- Evolutionary relations

Metabolic networks...

Metabolic networks (pathways) aligners

- The right balance between network topology and biological information, is one of the most difficult and key points in every metabolic network alignment algorithm.

The global pairwise alignment is usually obtained by defining

- node score similarity based on network topology and sequence similarity
- final node assignment maximizing an overall score

Metabolic Network Aligners

- SubMap Final alignment obtained by first defining a one-to many mapping of subnetworks

Ferhat A, Manolis K, Tamer K. *SubMAP: Aligning Metabolic Pathways with Subnetwork Mappings.* J Comput Biol. 2011 Mar; 18(3): 219âĂŞ235. doi: 10.1089/cmb.2010.0280.

- MP-Align Final alignment obtained by first aligning "all" paths in every network.

Alberich et al. *MP-Align: alignment of metabolic pathways*. BMC Systems Biology 2014 8:58 DOI: 10.1186/1752-0509-8-58

Alberich R, Castro JA, Llabrés M, Palmer-Rodríguez P. *Metabolomics analysis: Finding out metabolic building blocks*. PLOS ONE 2017 https://doi.org/10.1371/journal.pone.0177031

- IsoRankN (adapted)

Cheng-Yu M et al. *Reconstruction of phyletic trees by global alignment of multiple metabolic networks*. BMC Bioinformatics201314(Suppl 2):S12 DOI: 10.1186/1471-2105-14-S2-S12

To do...

- Protein function prediction
- Protein interaction prediction
- Metabolic network simulation
- Modelling the "metabolomics" of a set of organisms like
  - lung, gut, soil microbiota

Thank you!


LOONEY TUNES
"That's all Folks!"