Essentiality and Damage in Metabolic Networks

Laboratório de Bioinformática e Biologia Computacional
Graduate School in Computer Science
Universidade do Vale do Rio dos Sinos (UNISINOS)
Brazil

José Carlos Merino Mombach
Financial support: FAPERGS AND CNPq

Unisinos

Southern Genome Investigation Project

- Paraná, Santa Catarina and Rio Grande do Sul
- Pneumonia caused by *Mycoplasma hyopneumoniae*
- Other genomes
- Method to find potential enzyme targets for drugs

Laboratory Team

- Dr. Ney Lemke - Physics
- Dr. Cláudia Barcellos - Biochemistry
- Dr. Fabiana Herédia – Molecular biology

- Adriana dos Reis
- Guilherme Bedin
- Jean Schmith
- Rejane Apolo Ferreira

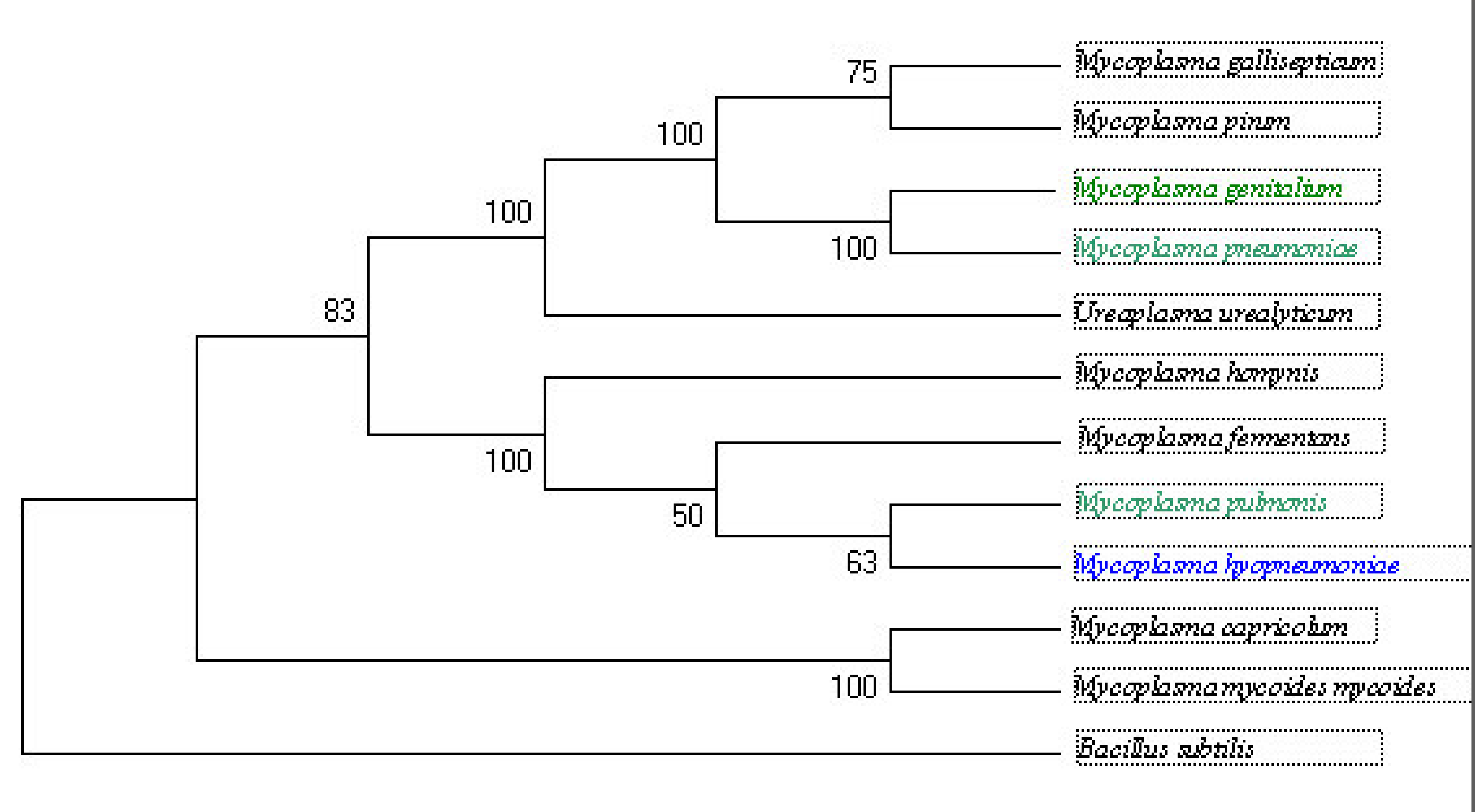
Genome Lengths

Mycoplasmas ~ 1.000.000 bp - 800 genes - 90%
E. coli - 4.600.000 - 4.500 - 89%
Yeast - 12.000.000 - 6.285 - 71%

C. elegans - 100.000.000 - 16.000 - 20%
Drosophila - 120.000.000 - 14.000 - 15,7%
Human - 3.000.000.000 - 40 a 50.000 - 3%

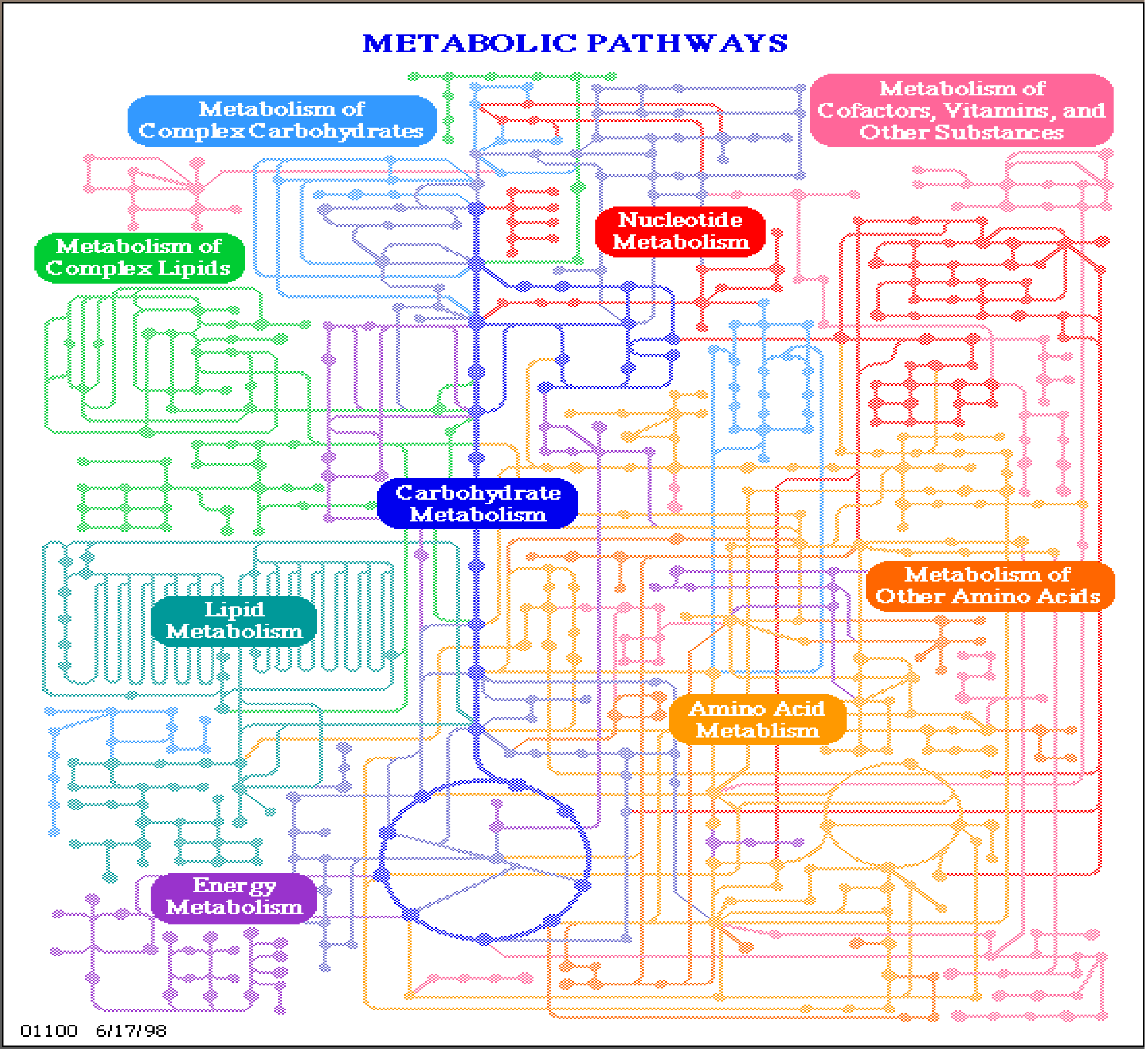
DNA → gene → protein
annotation

Mycoplasma Phylogeny



Metabolic Networks

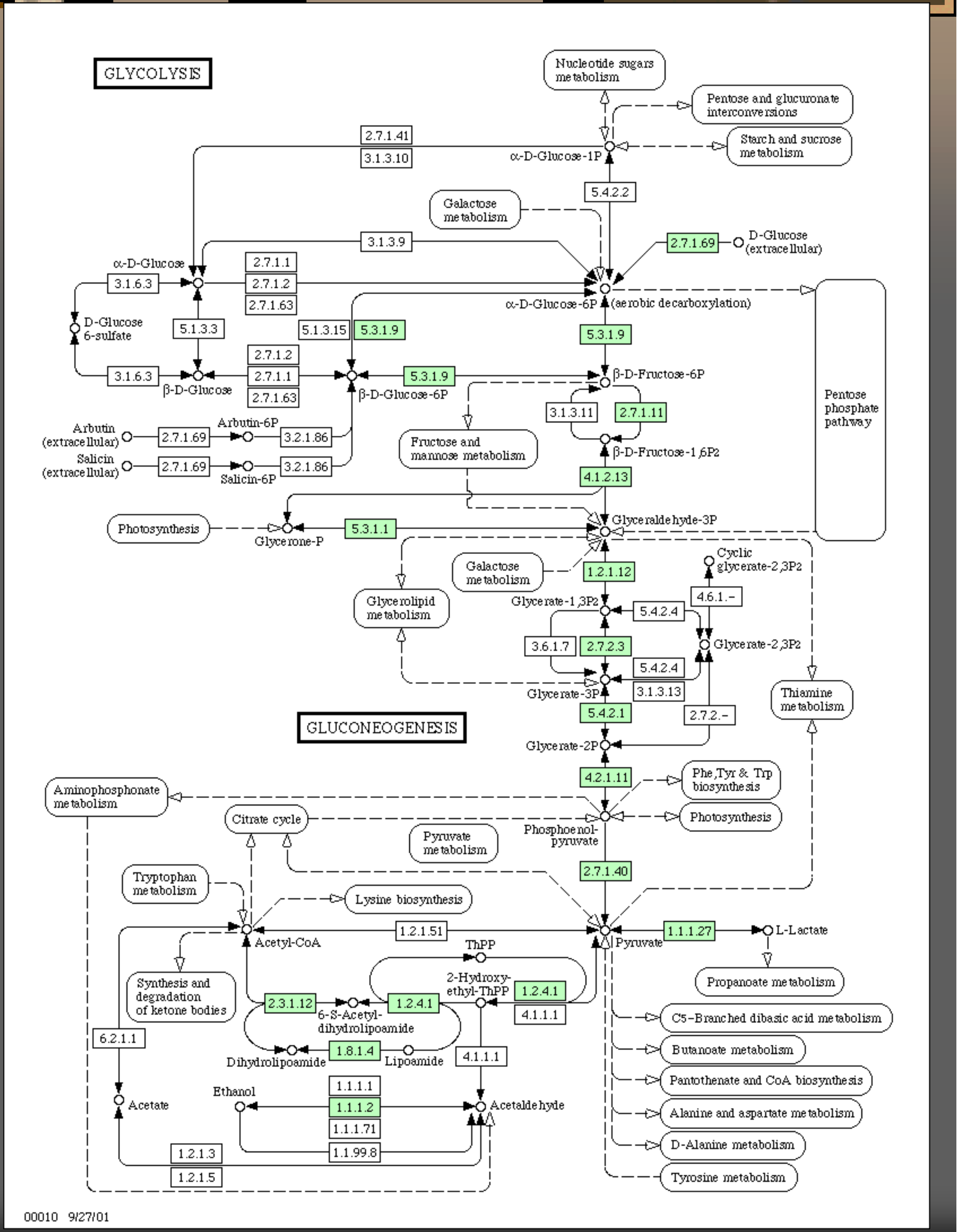
- Molecular circuitry of a cell



KEGG database

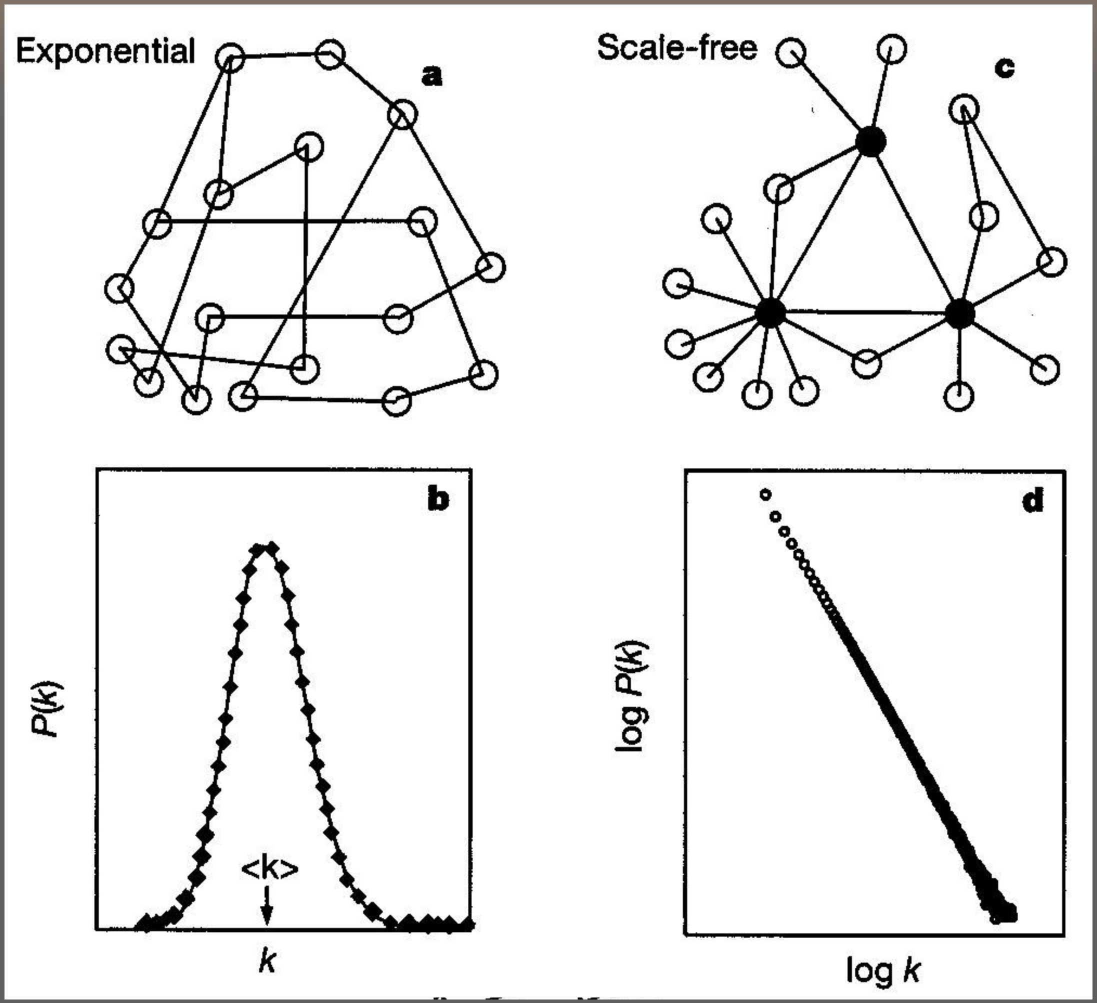
(Kyoto Encyclopedia of Genes And Genomes)

Directed graph



Metabolic Graphs

- Modeling metabolism is complex
- Graphs capture causality



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

(Poisson)

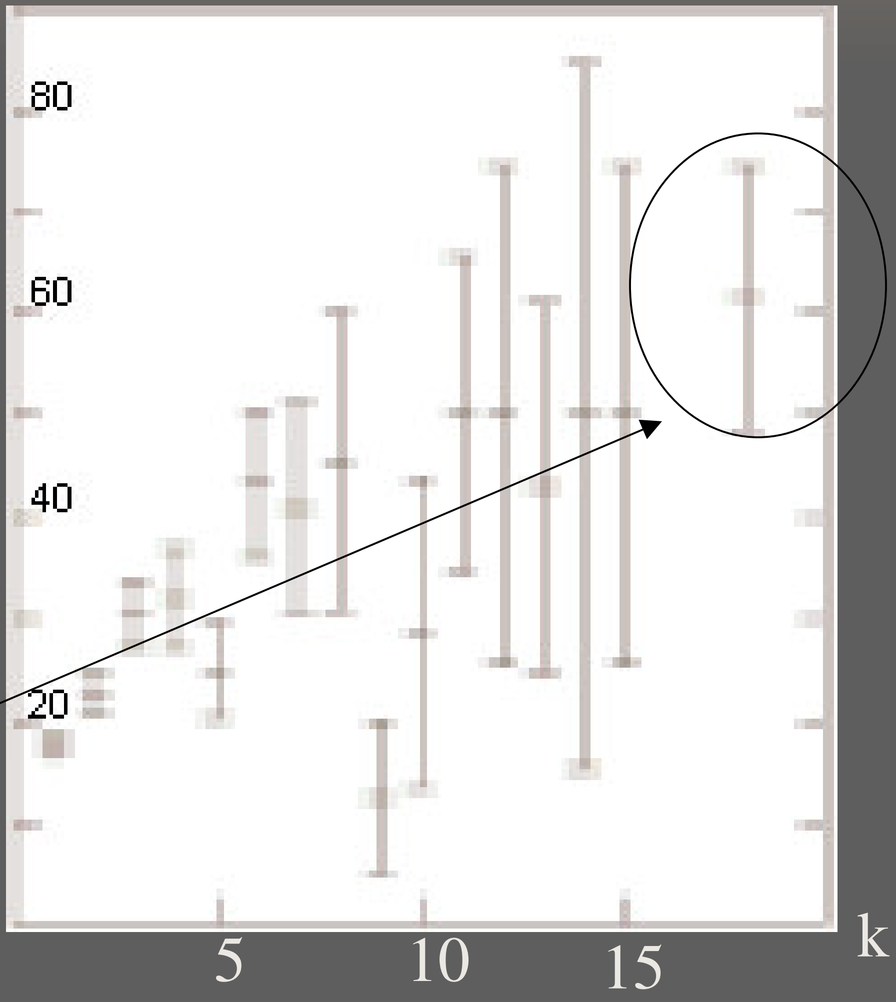
$P(k) \sim k^{-\gamma}$

Jeong et al., *Nature* v.407, 651, 2000.
Strogatz *Nature* (2001)

Interação Física entre Proteínas

1870 proteínas
2240 interações

Percentagem de Proteínas essenciais do *S. cerevisiae*



k<6 (93%)
21% são essenciais

k>15 (0.7%)
62% são essenciais

H. Jeong et al., *Nature* v.411, 41, 2001.

Relevance of Network Topology ?

- Can we get useful information?

- Network anatomy is related to function!
  (Strogatz)

Importance and topology ?

importance ←→ topology

metabolites ←→ connectivity (H_2O, ...)

phys. interac.
proteins ←→ connectivity

enzymes ←→ ? (damage)

Drug targets!!

Essentiality and Damage *E. coli*

- Can we determine important enzymes from the graph?

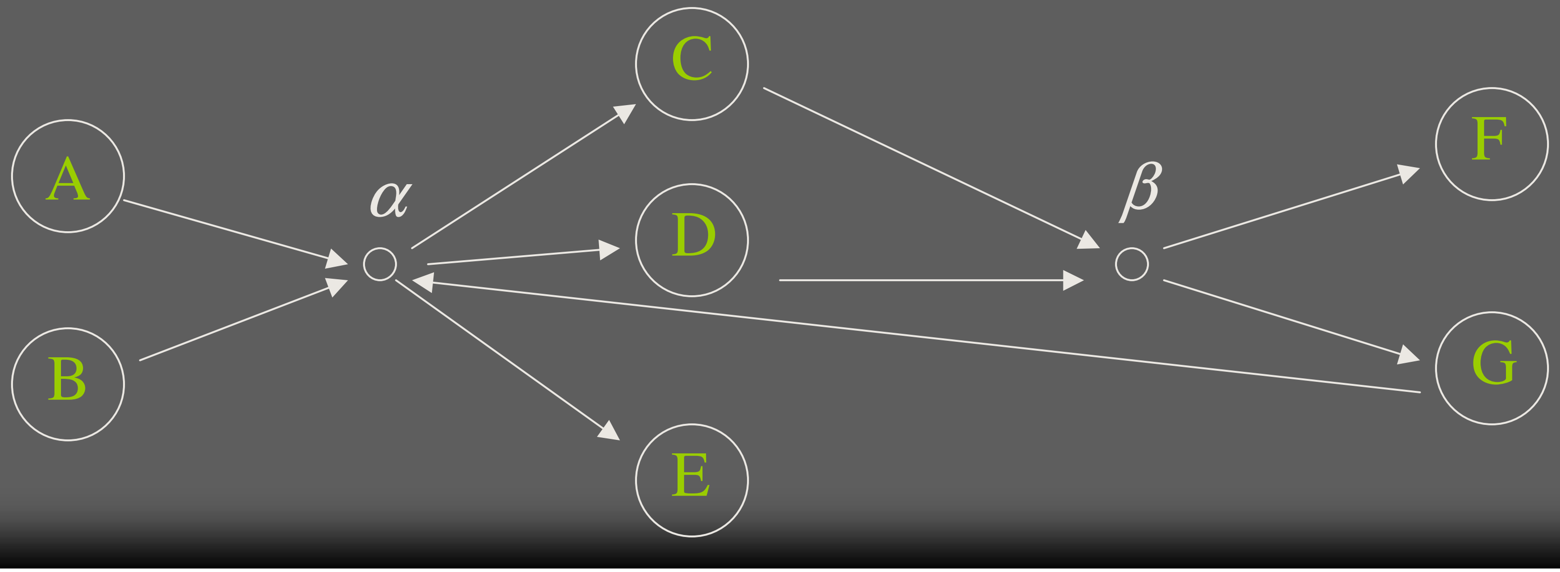
Importance = lethality

PEC Profiling of *E. coli* chromosome

New Graph Representation

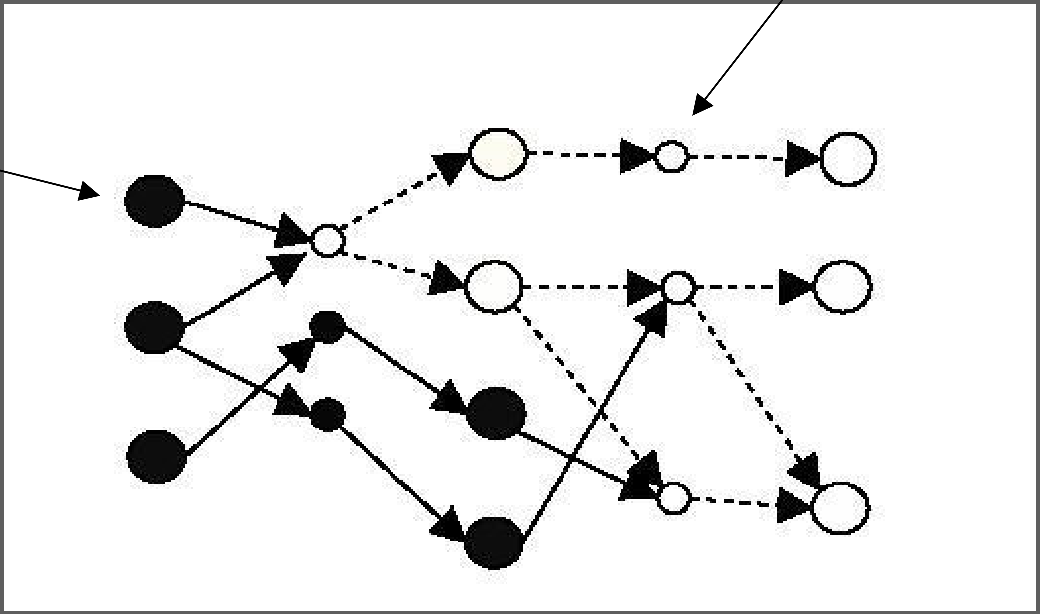
$$A + B + G \xrightarrow{\alpha} C + D + E$$

$$C + D \xrightarrow{\beta} F + G$$



Damage (*d*)

- Download the set of annotated EC's
- Number of metabolites (nodes) removed from the network

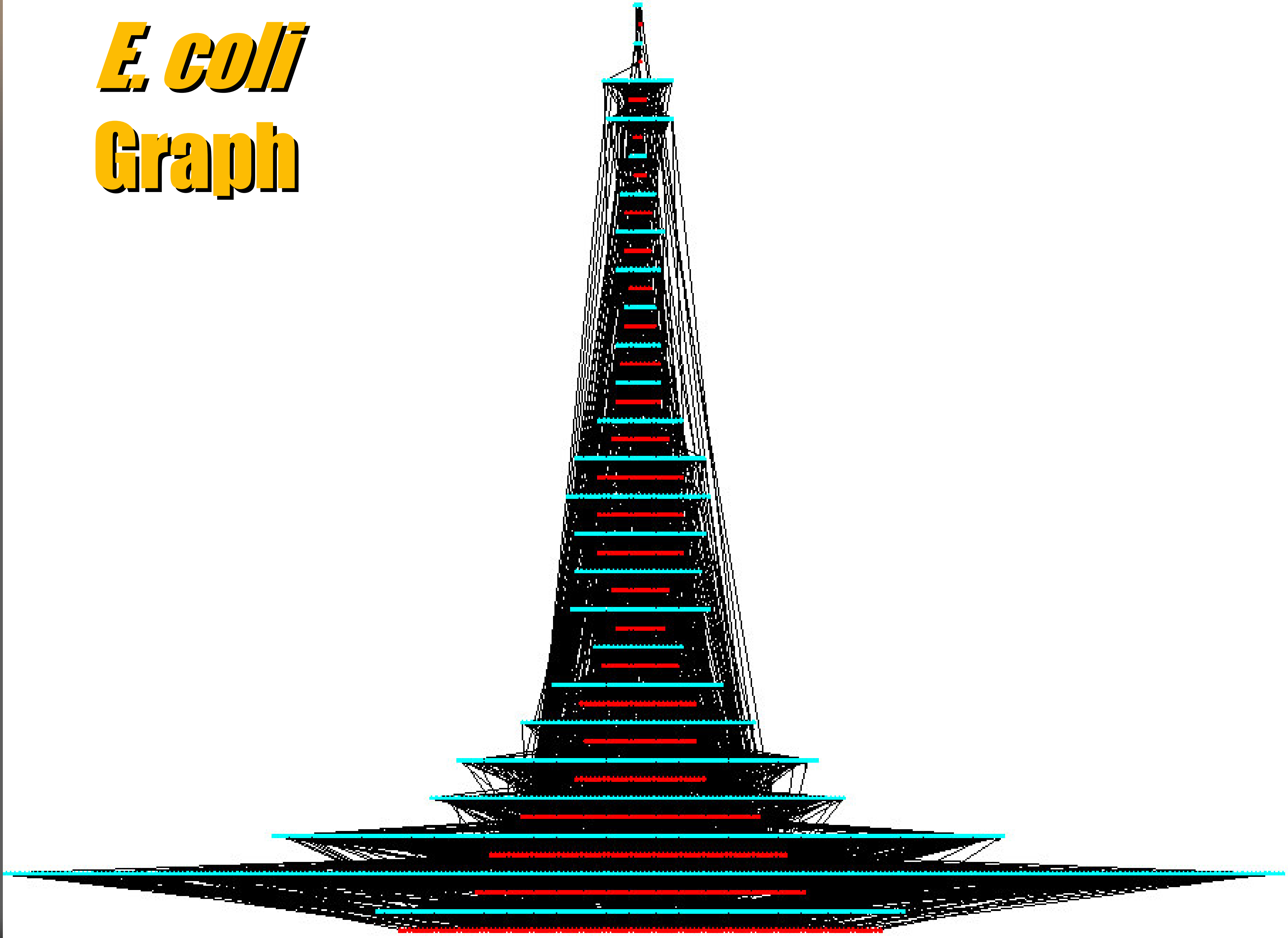


chemical reactions

metabolites

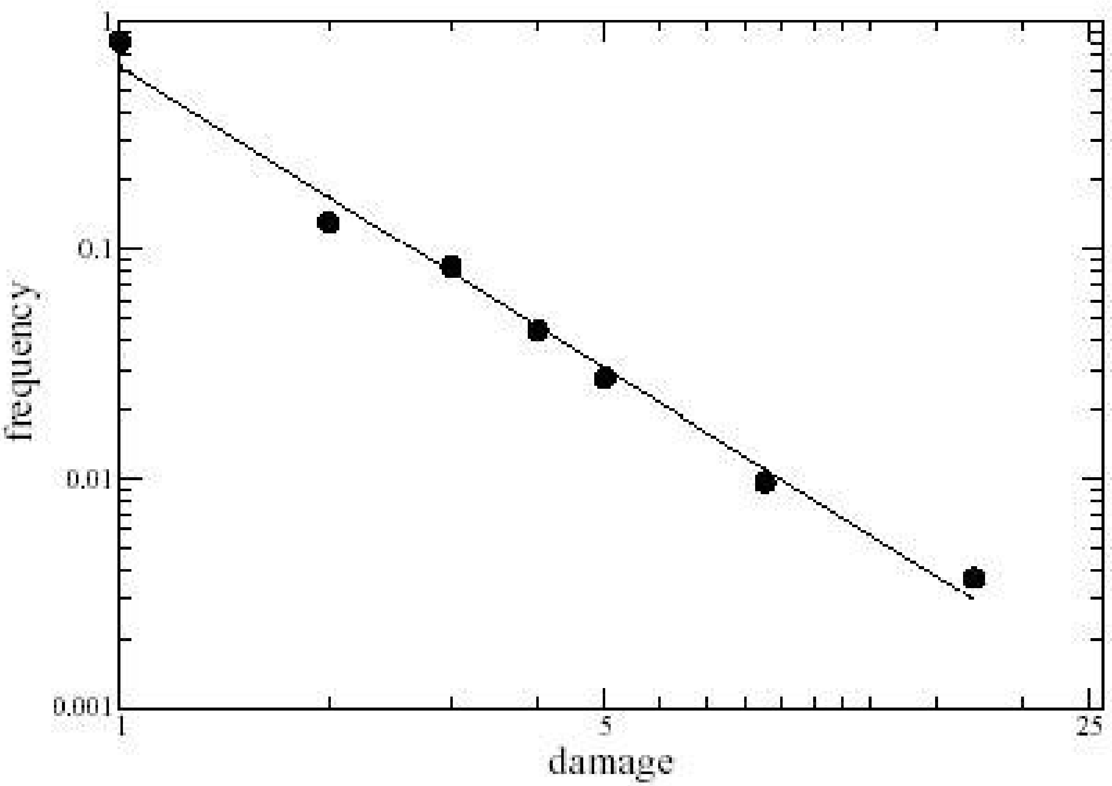
bipartite graph

E. coli Graph



Damage Frequency

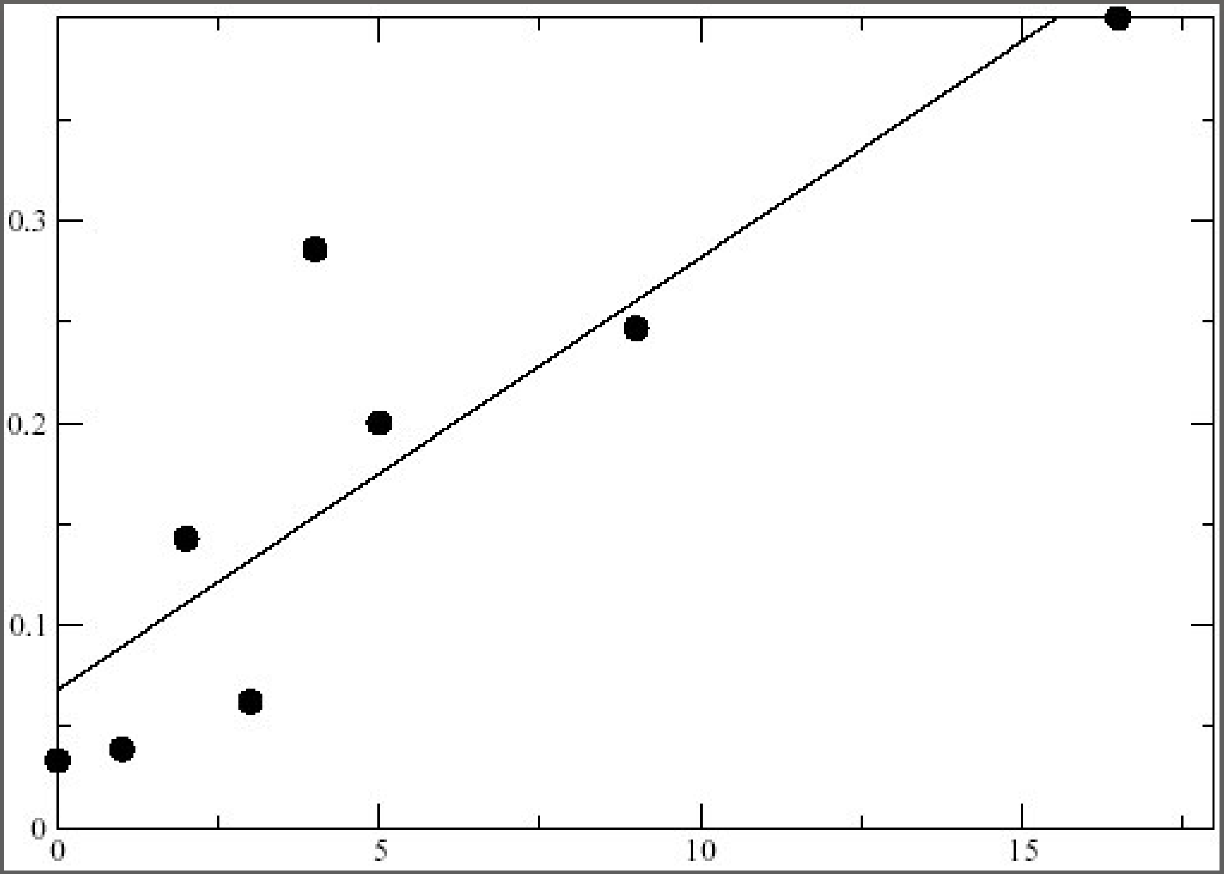
λ = 2.04



d < 5: 91%   d > = 5: 9%

Essential Enzymes vs. *d*

Fraction



damage

Correlation

F-test

P-value = 0.0228

Another Significance Test

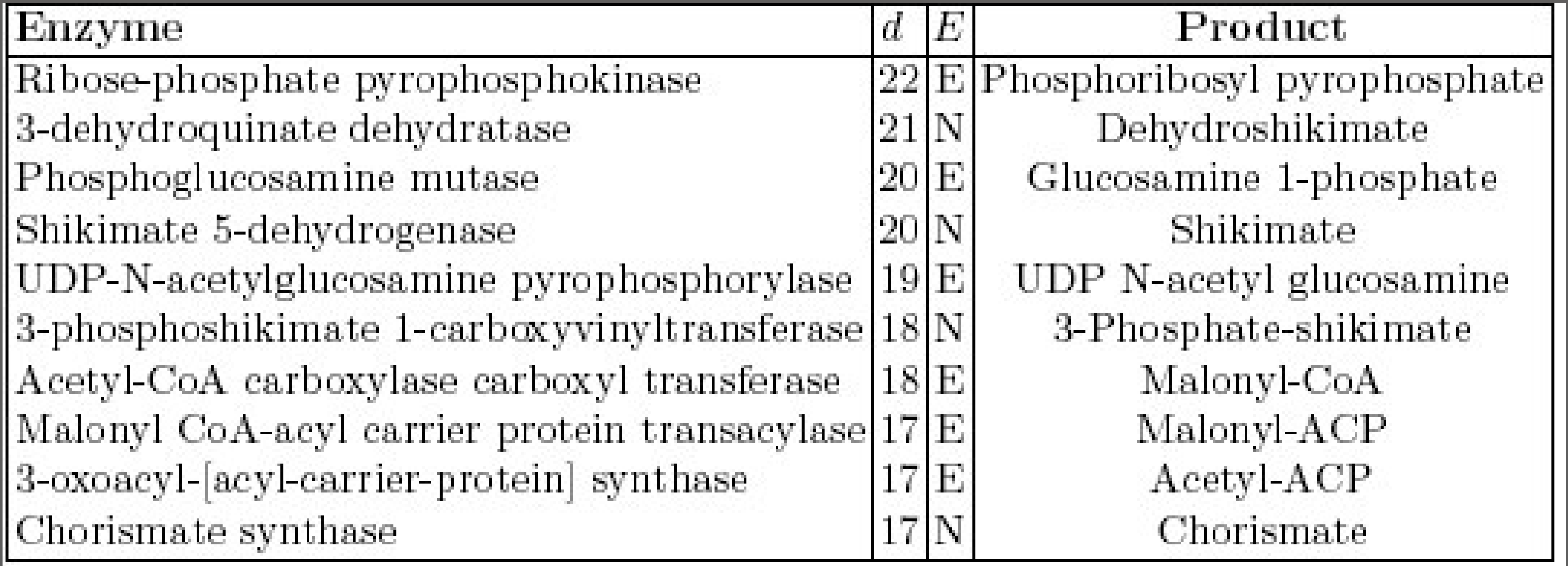
Two groups:

$d < 5$: 91% → 50% of not essentials

$d \geqslant 5$: 9% → 50% of essentials

$P\,(\,9\% \,;\, 50\% \text{ of essentials}\,) = 10^{-7}$

10 Enzymes Highest Damage



Enzyme	d	E	Product
Ribose-phosphate pyrophosphokinase	22	E	Phosphoribosyl pyrophosphate
3-dehydroquinate dehydratase	21	N	Dehydroshikimate
Phosphoglucosamine mutase	20	E	Glucosamine 1-phosphate
Shikimate 5-dehydrogenase	20	N	Shikimate
UDP-N-acetylglucosamine pyrophosphorylase	19	E	UDP N-acetyl glucosamine
3-phosphoshikimate 1-carboxyvinyltransferase	18	N	3-Phosphate-shikimate
Acetyl-CoA carboxylase carboxyl transferase	18	E	Malonyl-CoA
Malonyl CoA-acyl carrier protein transacylase	17	E	Malonyl-ACP
3-oxoacyl-[acyl-carrier-protein] synthase	17	E	Acetyl-ACP
Chorismate synthase	17	N	Chorismate

High damage, not essential => present in culture medium (?)

low damage, essential => not metabolism

Analysis

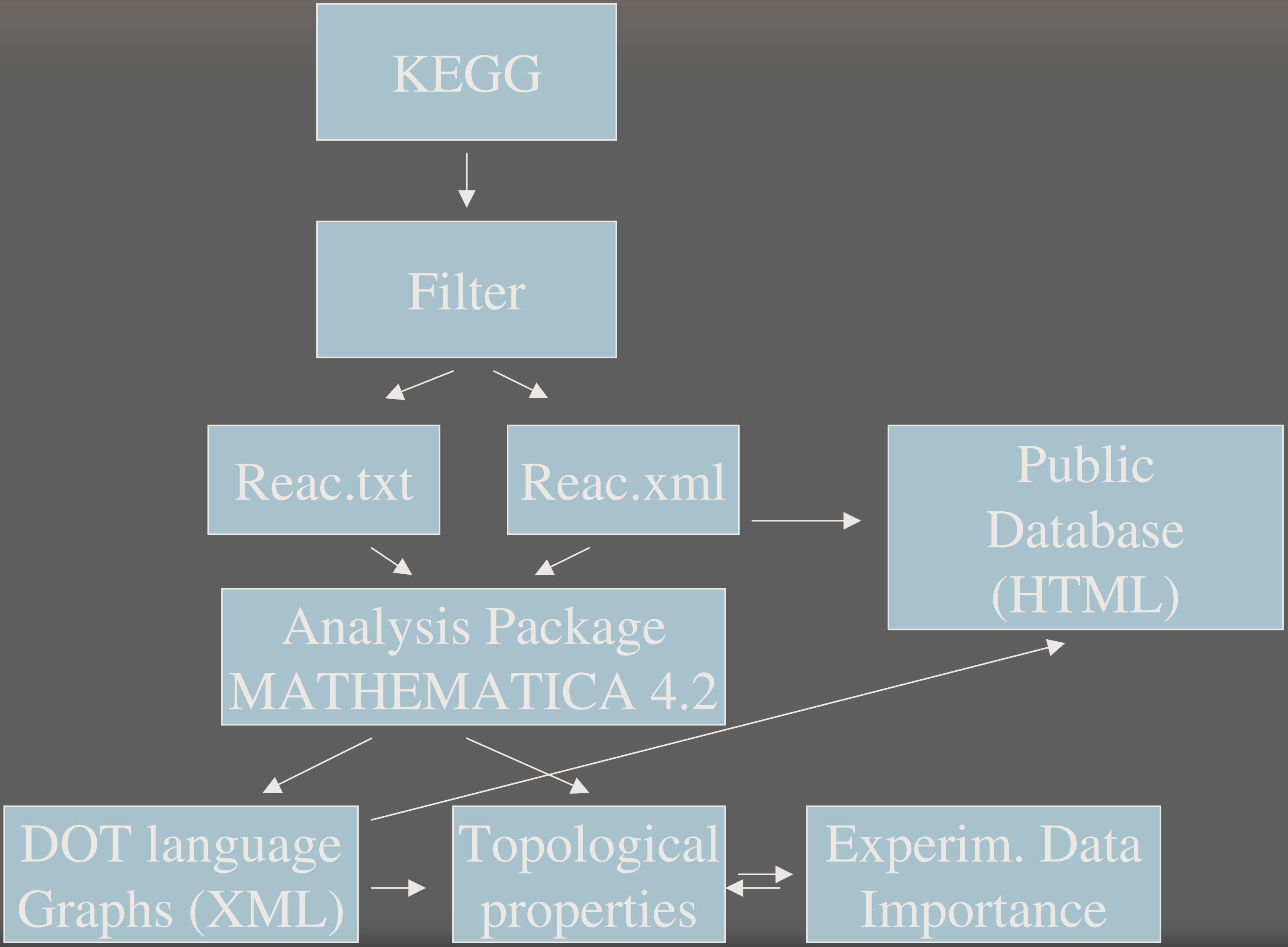
- In general → connect different maps
- Highest damaging enzyme → PRPP initial product for 4 different maps
- 5 enzymes → biosynthesis of cell wall (atractive targets)

Conclusions

- It is possible to “filter” important enzymes
- Application to mycoplasmas and other organisms
- Metabolic engineering
- Application to networks of interacting proteins

Submitted to *Bioinformatics*

Environment



KEGG

Filter

Reac.txt

Reac.xml

Public Database (HTML)

Analysis Package MATHEMATICA 4.2

DOT language Graphs (XML)

Topological properties

Experim. Data Importance

Extensions

Regulation

Not all maps are expressed at the same time!

(microarrays)