Biological Network Analysis

Mark B Gerstein
Yale

slides at
Lectures.GersteinLab.org

(See Last Slide for References & More Info.)
Networks occupy a midway point in terms of level of understanding

1D: Complete Genetic Partslist

~2D: Bio-molecular Network Wiring Diagram

3D and 4D: Detailed structural understanding of cellular machinery (e.g. ribosome in different functional states)

Networks as a universal language
Combining networks forms an ideal way of integrating diverse information

Metabolic pathway

Transcriptional regulatory network

Physical protein-protein Interaction

Co-expression Relationship

Genetic interaction (synthetic lethal)

Signaling pathways

Part of the TCA cycle
• Why Networks?

• Network Comparisons
  (reg. net. in many organisms)
  – in rel. to social hierarchy
  – scaling in rel. to partnerships
  – Computer OS Comparisons

• Network Dynamics Across Environments
  (prokaryote metab. pathways)
  – Metabolic Pathways
  – Entry pts. (Mem. Proteins)
Network Comparison #1
Comparing the Yeast Regulatory Network to a Governmental Hierarchy
Social Hierarchy

THE GOVERNMENT OF THE UNITED STATES

LEGISLATIVE BRANCH
THE CONGRESS
SENATE
HOUSE
ARCHITECT OF THE CAPITOL
UNITED STATES BOTANIC GARDEN
GENERAL ACCOUNTING OFFICE
GOVERNMENT PRINTING OFFICE
LIBRARY OF CONGRESS
CONGRESSIONAL BUDGET OFFICE

EXECUTIVE BRANCH
WHITE HOUSE OFFICE
OFFICE OF THE VICE PRESIDENT
OFFICE OF NATIONAL DRUG CONTROL POLICY
COUNCIL OF ECONOMIC ADVISORS
COUNCIL ON ECONOMICATIONAL POLICY DEVELOPMENT
COUNCIL ON ECONOMIC AND TECHNOLOGY POLICY
OFFICE OF ADMINISTRATION

JUDICIAL BRANCH
SUPREME COURT OF THE UNITED STATES
UNITED STATES COURTS OF APPEALS
UNITED STATES DISTRICT COURTS
TERRITORIAL COURTS
UNITED STATES COURT OF INTERNATIONAL TRADE
UNITED STATES COURT OF FEDERAL CLAIMS
UNITED STATES COURT OF APPEALS FOR THE ARMED FORCES
UNITED STATES TAX COURT
UNITED STATES COURT OF APPEALS FOR VETERANS CLAIMS
ADMINISTRATIVE OFFICE OF THE UNITED STATES COURTS
FEDERAL JUDICIAL CENTER
UNITED STATES SENTENCING COMMISSION

INDUSTRIAL BRANCH
AFRICAN DEVELOPMENT FOUNDATION
CENTRAL INTELLIGENCE AGENCY
COMMERCIAL SERVICE TRADING COMMISSION
CONSUMER PROTECTION SAFETY DEPARTMENT
CORPORATION FOR NATIONAL AND COMMUNITY SERVICE
DEFENSE NUCLEAR FACILITIES SAFETY BOARD
ENVIRONMENTAL PROTECTION AGENCY
EQUAL EMPLOYMENT OPPORTUNITY COMMISSION
EXPERT IMPORT BANK OF THE U.S.
FARM CREDIT ADMINISTRATION
FEDERAL COMMUNICATIONS COMMISSION
FEDERAL DEPOSIT INSURANCE CORPORATION
FEDERAL ELECTRIC COMMISSION
FEDERAL HOUSING FINANCE BOARD

FEDERAL LABOR RELATIONS AUTHORITY
FEDERAL MARITIME COMMISSION
FEDERAL MEDIATION AND CONCILIATION SERVICE
FEDERAL MINERAL SAFETY AND HEALTH-REVIEW COMMISSION
FEDERAL RESERVE SYSTEM
FEDERAL RETIREMENT TRUST INVESTMENT BOARD
FEDERAL TRADE COMMISSION
GENERAL SERVICES ADMINISTRATION
INTER-AMERICAN FOUNDATION
MERIT SYSTEMS PROTECTION BOARD
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
NATIONAL ARCHIVES AND RECORDS ADMINISTRATION
NATIONAL CAPITAL PLANNING COMMISSION
NATIONAL CREDIT UNION ADMINISTRATION
NATIONAL FOUNDATION ON THE ARTS AND THE HUMANITIES
NATIONAL LABOR RELATIONS BOARD
NATIONAL MEDIATION BOARD
NATIONAL RAILROAD PASSENGER CORPORATION (AMTRAK)
NATIONAL SCIENCE FOUNDATION
NATIONAL TRANSPORTATION SAFETY BOARD
NUCLEAR REGULATORY COMMISSION
OCCUPATIONAL SAFETY AND HEALTH REVIEW COMMISSION
OFFICE OF GOVERNMENT ETHICS
OFFICE OF PERSONNEL MANAGEMENT
OFFICE OF SPECIAL COUNSEL
OVERSEAS PRIVATE INVESTMENT CORPORATION
PEACE CORPS
PENSION BENEFIT GUARANTY CORPORATION
POSTAL RATE COMMISSION
RAILROAD RETIREMENT BOARD
SECURITIES AND EXCHANGE COMMISSION
SELECTIVE SERVICE SYSTEM
SMALL BUSINESS ADMINISTRATION
SOCIAL SECURITY ADMINISTRATION
TENNESSEE VALLEY AUTHORITY
TRADE AND DEVELOPMENT AGENCY
U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT
U.S. COMMISSION ON CIVIL RIGHTS
U.S. INTERNATIONAL TRADE COMMISSION
U.S. POSTAL SERVICE
Determination of "Level" in Regulatory Network Hierarchy with Breadth-first Search

I. Example network with all 4 motifs

II. Finding terminal nodes (Red)

III. Finding mid-level nodes (Green)

IV. Finding top-most nodes (Blue)

Level 1

[Yu et al., PNAS (2006)]
Regulatory Networks have similar hierarchical structures

[S. cerevisiae]

1

2

3

4

[E. coli]

[Yu et al., Proc Natl Acad Sci U S A (2006)]
Yeast Regulatory Hierarchy: the Middle-managers Rule

A. Regulatory hierarchy in *S. cerevisiae*

[Yu et al., PNAS (2006)]
Yeast Network Similar in Structure to Government Hierarchy with Respect to Middle-managers

B. Governmental hierarchy of a representative city (Macao)
Characteristics of Regulatory Hierarchy: Middle Managers are Information Flow Bottlenecks

Average betweenness at each level

Yu et al., PNAS (2006)
Network Comparison #2
Broadening the comparison to different types of hierarchies & different types of biological networks
Different kinds of Hierarchies

- **Autocratic**
  - Well-defined levels and a clear chain of command
  - A military hierarchy

- **Democratic**
  - Without well-defined levels & with more co-regulatory partnerships
  - A club or a scientific collaboration network

- **Intermediate**
  - High degree of co-regulation and can be organized into hierarchies
  - A law firm

<table>
<thead>
<tr>
<th></th>
<th>Autocratic</th>
<th>Democratic</th>
<th>Intermediate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betweenness (▲)</td>
<td>1.03</td>
<td>3.6</td>
<td>3.3</td>
</tr>
<tr>
<td>Betweenness (△)</td>
<td>4.1</td>
<td>1.08</td>
<td>3.4</td>
</tr>
<tr>
<td>Var. Betw. (triangles)</td>
<td>2.1</td>
<td>0.58</td>
<td>1.74</td>
</tr>
<tr>
<td>Var. Betw. (all)</td>
<td>2.9</td>
<td>1.4</td>
<td>1.9</td>
</tr>
<tr>
<td>$D_{Net-collab}$</td>
<td>0</td>
<td>0.91</td>
<td>0.71</td>
</tr>
</tbody>
</table>

- Decreases $\sigma^2$ of ‘stress’
- Increases ‘collaboration’

[Bhardwaj et al., PNAS (2010), in press]
Collaborative Nature of the Nodes

Degree of Collaboration $D_{collab}^i$

More Collaborative: Democratic
More Autonomous: Autocratic

[Bhardwaj et al., PNAS (2010), in press]
Higher species are more show more collaborative nodes (more democratic)

[Bhardwaj et al., PNAS (2010), in press]
Collaborative Nature of the Levels

Level $L$

\[ D^i_{collab} = 0.1 \ 0.3 \ 0.2 \ 0.4 \]

\[ D^L_{Level\_collab} = \frac{0.1 + 0.3 + 0.2 + 0.4}{4} = 0.25 \]

[Bhardwaj et al., PNAS (2010), in press]
Collaboration Between Levels

\[ D_{\text{betw-level-collab}}^{L,M} = \frac{\sum \sum G_A \cap G_B}{|L| \cdot |M|} \]

[Bhardwaj et al., PNAS (2010), in press]
Middle Managers Interact the Most in Efficient Corporate Settings


[Bhardwaj et al., PNAS (2010), in press]
Co-regulation Instantiates a Multi-Input Motif

[Bhardwaj et al., PNAS (2010), in press]
Network Comparisons #3
Relating the size of co-regulation in partnership networks with the scale of the regulated
Co-regulation Partnerships

- Readily seen in many commonplace social contexts.
- An academic institution (say a high school), multiple teachers supervise the same set of students and have partnership interactions amongst themselves.

[Bhardwaj et al., PLoS Comp Biol (2010), in press]
Building and Analysis of Networks

Edge placed if two regulators co-regulate

<table>
<thead>
<tr>
<th>Network type</th>
<th>Species</th>
<th>Number of regulators</th>
<th>Number of targets</th>
<th>Number of interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription</td>
<td>E. coli</td>
<td>160</td>
<td>1,420</td>
<td>3,123</td>
</tr>
<tr>
<td>Transcription</td>
<td>Yeast</td>
<td>157</td>
<td>4,410</td>
<td>12,873</td>
</tr>
<tr>
<td>Transcription</td>
<td>Mouse</td>
<td>144</td>
<td>1,092</td>
<td>2,403</td>
</tr>
<tr>
<td>Transcription</td>
<td>Rat</td>
<td>91</td>
<td>461</td>
<td>1,092</td>
</tr>
<tr>
<td>Transcription</td>
<td>Human</td>
<td>156</td>
<td>3,032</td>
<td>6,896</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>Yeast</td>
<td>87</td>
<td>1,337</td>
<td>4,083</td>
</tr>
<tr>
<td>Modification</td>
<td>Human</td>
<td>518</td>
<td>1,218</td>
<td>2,782</td>
</tr>
</tbody>
</table>

[Bhardwaj et al., PLoS Comp Biol (2010), in press]
Scaling of Regulatory Partnerships with Targets
Scaling across many networks

(a) *E. coli* transcription

\[ y = 1.85 + 1.02x, R^2 = 0.81 \]

\[ y = 41(1 - e^{-0.046x}), R^2 = 0.84 \]

Linear in *E. coli* (Due to operons)

(b) Yeast transcription

\[ y = 126(1 - e^{-0.03x}), R^2 = 0.95 \]

(c) Rat transcription

\[ y = 52(1 - e^{-0.05x}), R^2 = 0.89 \]

\[ y = 49(1 - e^{-0.05x}), R^2 = 0.72 \]

Exponential Saturation in others

(d) Mouse transcription

\[ y = 73(1 - e^{-0.03x}), R^2 = 0.91 \]

\[ y = 21(1 - e^{-0.02x}), R^2 = 0.81 \]

(e) Human transcription

\[ y = 91(1 - e^{-0.04x}), R^2 = 0.90 \]

\[ y = 72(1 - e^{-0.02x}), R^2 = 0.92 \]

(f) Yeast phosphorylation

\[ y = 75(1 - e^{-0.087x}), R^2 = 0.81 \]

\[ y = 66(1 - e^{-0.057x}), R^2 = 0.79 \]

[Bhardwaj et al., PLoS Comp Biol (2010), in press]
Comparison to Social Networks: Partnership networks effectively saturate with increasingly complex output

For each central node c

\[
y = 635(1 - e^{-0.14x}), R^2 = 0.74
\]

\[
y = 694(1 - e^{-0.034x}), R^2 = 0.82
\]
A Simple Theoretical Model

On average, each regulator has $n$ targets
$f_i$: the number of partners
$k_i$: the number of targets

Integrating this, we get:

\[ f = m(1 - e^{-n/Nk}) = a(1 - e^{-bk}) \]

[Bhardwaj et al., PLoS Comp Biol (2010), in press]
Network Comparison #4: Comparing the structure and evolution of biological regulatory networks and software call graphs
**E. Coli Transcriptional regulatory network vs Linux kernel call graph**

<table>
<thead>
<tr>
<th>Basic properties of systems</th>
<th>E. coli transcriptional regulatory network</th>
<th>Linux call graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>Genes (TFs &amp; targets)</td>
<td>Functions (subroutines)</td>
</tr>
<tr>
<td>Edges</td>
<td>Transcriptional regulation</td>
<td>Function calls</td>
</tr>
<tr>
<td>External constraints</td>
<td>Natural environment</td>
<td>Hardware architecture, customer requirements</td>
</tr>
<tr>
<td>Origin of evolutionary changes</td>
<td>Random mutation &amp; natural selection</td>
<td>Designers’ fine tuning</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>E. coli transcriptional regulatory network</th>
<th>Linux call graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>1378</td>
<td>12391</td>
</tr>
<tr>
<td>Number of persistent nodes</td>
<td>72* (5%)</td>
<td>5120 (41%)</td>
</tr>
<tr>
<td>Number of edges</td>
<td>2967</td>
<td>33553</td>
</tr>
<tr>
<td>Number of modules</td>
<td>64</td>
<td>3665</td>
</tr>
<tr>
<td>Number of comparative references</td>
<td>200 bacterial genomes</td>
<td>24 versions of kernels</td>
</tr>
<tr>
<td>Years of evolution</td>
<td>Billions years</td>
<td>20 years</td>
</tr>
</tbody>
</table>

[Yan et al., PNAS (2010), in press]
E. coli transcriptional regulatory network

master regulator

middle manager

workhorse

Linux call graph

[Yan et al., PNAS (2010), in press]
Comparison: hierarchical organization

<table>
<thead>
<tr>
<th>Role</th>
<th>E. coli regulatory network</th>
<th>Linux call graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>master regulator</td>
<td>4.6</td>
<td>29.6</td>
</tr>
<tr>
<td>middle manager</td>
<td>5.1</td>
<td>58.2</td>
</tr>
<tr>
<td>workhorse</td>
<td>90.2</td>
<td>12.3</td>
</tr>
</tbody>
</table>

Degree distribution
Roles of hubs

Pyramidal vs Top-heavy

[Yan et al., PNAS (2010), in press]
Comparison: organization of modules

Modules are labeled by master regulators: TFs, high-level starting functions.

TRN: modules overlap little, components are less generic: “ompF”

<table>
<thead>
<tr>
<th></th>
<th>E. Coli TRN</th>
<th>Linux call graph</th>
</tr>
</thead>
<tbody>
<tr>
<td># of modules</td>
<td>64</td>
<td>3665</td>
</tr>
<tr>
<td>Average overlap</td>
<td>4.3%</td>
<td>80.7%</td>
</tr>
<tr>
<td>Maximum node reuse</td>
<td>15.6%</td>
<td>87.5%</td>
</tr>
<tr>
<td>Average node reuse</td>
<td>3.5%</td>
<td>8.4%</td>
</tr>
</tbody>
</table>

Overlap(M2, M3) = \frac{|M2 \cap M3|}{|M2 \cup M3|} = \frac{2}{11}

Call graph: modules overlap, Functions are highly reused (generic): “printk”

[Yan et al., PNAS (2010), in press]
Comparison of persistent components

• Persistent genes (preserve among different genomes) vs persistent functions (preserve among different releases)

• Building of the hierarchy:
  ◊ TRN: Bottom up. Regulatory changes are the main driving forces of evolution
  ◊ Call graph: top down

[Yan et al., PNAS (2010), in press]
Evolutionary rate of persistent functions

Persistent genes evolve slowly

Two classes of persistent functions

Higher reuse

[Yan et al., PNAS (2010), in press]
**Why and so what?**

The difference can be explained by the nature of hubs evolution: tinkering vs design

- **Independent modules:**
  - robust
  - costly: the system needs a variety of tools for different tasks

- **Overlap modules (reuse):**
  - Less robust:
    - Breakdown of a generic component is harmful to the whole system
    - Fragile in the sense any change in a module may require compensating changes in a generic function
  - cost effective: components can be used by need to be fine-tuned

[Spearman correlation
r=0.25
P<10^{-75}

[Yan et al., PNAS (2010), in press]
Network Dynamics Across Environments: Metabolic Pathways

How do molecular networks change across environments? What pathways are used more? Used as a biosensor?
What is Metagenomics?

Traditional Genomics

1. Select organism and culture
2. Extract DNA and sequence
3. Assemble and annotate

Estimated that less than 1% of microbes can be cultured

Metagenomics

1. Collect sample from environment
2. Extract DNA and sequence
3. Assemble and annotate

Lose information about which gene belongs to which microbe
Sorcerer II Global Ocean Survey

Sorcerer II journey August 2003 - January 2006
Sample approximately every 200 miles

Rusch, et al., *PLOS Biology* 2007
Sorcerer II Global Ocean Survey

**Metadata**
- GPS coordinates, Sample Depth, Water Depth, Salinity, Temperature, Chlorophyll Content

**Metagenomic Sequence**
- 6.25 GB of data
- 0.1–0.8 µm size fraction (bacteria)
- 6.3 billion base pairs (7.7 million reads)
- Reads were assembled and genes annotated
- 1 million CPU hours to process

**Additional Metadata** via GPS coordinates

**Metabolic Pathways**

**Membrane Protein Families**

Rusch, et al., *PLOS Biology* 2007
Extracting Environmental Data from Other Sources

Nutrient Features Extracted:
- Phosphate
- Silicate
- Nitrate
- Apparent Oxygen Utilization
- Dissolved Oxygen

Sample Depth: 1 meter
Water Depth: 32 meters
Chlorophyll: 4.0 ug/kg
Salinity: 31 psu
Temperature: 11 C
Location: 41°5'28"N, 71°36'8"W

Annual Phosphate [umol/l] at the surface

World Ocean Atlas 2005
NOAA/NODC
40% of Oceans are Impacted by Humans

* Resolution is 1 km square
* Value of a activity at a particular location is determined by the type of ecosystem present:
  \[
  \text{Impact} = \sum \text{Features} \times \text{Ecosystem} \times \text{impact weight}
  \]

**Anthropogenic Features Extracted:**
- Ultraviolet radiation
- Shipping
- Pollution
- Climate Change
- Ocean Acidification

Mapping Raw Metagenomic Reads to a Matrix of Families or Pathways for each Site

Patel et. al., Genome Research 2010
Expressing data as matrices indexed by site, env. var., and pathway usage

[Rusch et. al., (2007) PLOS Biology; Gianoulis et al., PNAS (in press, 2009)]
Simple Relationships: Pairwise Correlations

Environmental Features

Chlorophyll      Temp

Pathways

Cobalamin Biosynthesis
Photosystem II
Photosystem I
Carbon Fixation (Dark rx)
Glutamine Degradation

Metadata

Temp   NaCl   Depth

Sites
B1    B2
15°C   27.2  10 m
23°C   36.6  5 m
Canonical Correlation Analysis: Simultaneous weighting

Lifestyle Index = a \( \text{km run/week} \) + b \( \text{Weight} \) + c

Fit Index = a \( \text{Blood pressure} \) + b \( \text{Weight} \) + c

# km run/week | Weight
--- | ---
Lifestyle Index | Fit Index
Canonical Correlation Analysis: Simultaneous weighting

<table>
<thead>
<tr>
<th>Lifestyle Index</th>
<th>Fit Index</th>
</tr>
</thead>
<tbody>
<tr>
<td># km run/week</td>
<td># km run/week</td>
</tr>
<tr>
<td>Weight</td>
<td>Weight</td>
</tr>
</tbody>
</table>

Lifestyle Index = a + b + c

Fit Index = a + b + c

Environmental Features
- Temp
- Chlorophyll
- etc

Metabolic Pathways/Protein Families
- Photosynthesis
- Lipid Metabolism
- etc
The goal of this technique is to interpret cross-variance matrices. We do this by defining a change of basis.

Gianoulis et al., PNAS 2009
Strength of Pathway co-variation with environment

Environmentally invariant  Environmentally variant

Gianoulis et al., PNAS 2009
Conclusion #1: energy conversion strategy, temp and depth

Gianoulis et al., PNAS 2009
Conclusion #2: Outer Membrane components vary with the environment

Membrane proteins interact with the environment, transporting available nutrients, sensing environmental signals, and responding to changes

Gianoulis et al., *PNAS* 2009
Patel et al. *Genome Research* 2010
Network Dynamics Across Environments: Membrane Proteins (Pathway Entry Points)
Membrane Proteins: Sensing and Responding the Environment

- 2.3 million predicted membrane proteins
- 1.2 million unique
- 850,000 mapped to 151 membrane protein COGs

107 variant membrane protein families
44 invariant membrane protein families
20% have NO KNOWN FUNCTION

Patel and Gianoulis et al., (in press) Genome Research
Membrane Proteins co-vary more than Metabolic Pathways

Median absolute structural Correlation Coefficient

Membrane Proteins  = 0.3

Metabolic Pathways  = 0.17

Patel and Gianoulis et al., (in press) Genome Research
CCA Limitations

Both the strength and the directionality of relationships between nodes is difficult to decipher in this format.
Protein Families and Environmental Features Network (PEN)

Distance: Dot product between 1st and 2nd Dimension of CCA

\[ a \cdot b = |a||b|\cos\theta \]

Patel et. al., Genome Research 2010
Protein Families and Environmental Features Network (PEN)

“Bi-modules”: groups of environmental features and membrane proteins families that are associated

UV, dissolved oxygen, apparent oxygen utilization, sample depth, and water depth are not in the network
Bi-module 1: Phosphate/Phosphate Transporters

Low Phosphate, high affinity phosphate transporters which are induced during phosphate limitation

High Phosphate, low affinity inorganic phosphate ion transporter which are constitutively expressed

Patel et al., Genome Research 2010
Negative relationship between areas of high ocean-based pollution and shipping and transporters involved in the uptake of iron

Pollution and Shipping may be a proxy for iron concentrations

Patel et. al., Genome Research 2010
Biosensors: 4 logs in 4 years
Beyond Canaries in a Coal Mine

$1000 Human ~ $1 E. coli
$100 Human ~ $.10 E. coli

1995: gel: $3G
2005: capil: $50M
2009: Lig: $1-5K
Pol: $10K-50K
2011: $100

Seq bp/$


Microbial Community

Carr and Church, Nat Biotech 2009

jgi.doe.gov

Lectures,GersteinLab.org™
• Why Networks?

• Network Comparisons
  (reg. net. in many organisms)
  – in rel. to social hierarchy
  – scaling in rel. to partnerships
  – Computer OS Comparisons

• Network Dynamics Across Environments
  (prokaryote metab. pathways)
  – Metabolic Pathways
  – Entry pts. (Mem. Proteins)
Conclusions: Comparison of Social and Regulatory Hierarchies

- Middle managers dominate, sitting at info. flow bottlenecks
- Democratic v Autocratic
- Collaborative (locally democratic) fraction of networks increases with organism complexity
- Middle managers most collaborative
- Most interaction occur between 2 middle managers (as seen in efficient corporate hierarchies)
- Number of collaborative partners saturates even while scale of targets governed increases. (Also seen in social networks.)
<table>
<thead>
<tr>
<th>Hierarchical organization</th>
<th>E. coli transcriptional regulatory network</th>
<th>Linux call graph</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structure</strong></td>
<td>Pyramidal</td>
<td>Top-heavy</td>
</tr>
<tr>
<td><strong>Characteristic hubs</strong></td>
<td>Upper-level TFs with high out-degree</td>
<td>Generic workhorse functions with high in-degree</td>
</tr>
</tbody>
</table>

| Organization of modules   |                                         |                 |
|---------------------------|                                         |                 |
| **Downstream modules as labeled by** | Master TFs responsible for sensing environmental signals | High-level starting functions which initiate execution for specific tasks |
| **Node reuse**            | Low                                      | High            |
| **Overlap between modules** | Low                                     | High            |

| Persistent nodes          |                                         |                 |
|---------------------------|                                         |                 |
| **Characteristics**       | Specialized (non-generic) workhorses     | Generic or reusable functions |
| **Location in hierarchy** | Mostly bottom                            | Mostly top      |
| **Evolutionary rate**     | Mostly conservative (e.g. dnaA)          | Conservative (e.g. strlen) & adaptive (e.g. mempool_alloc) |

| Design principles         |                                         |                 |
|---------------------------|                                         |                 |
| **Building of hierarchy** | Bottom up                               | Top down        |
| **Optimal solution favors** | Robustness                             | Cost-effectiveness (reuse of components) |
Conclusions: Network Dynamics Across Environments

• Developed approach to connect quantitative features of environment to usage of pathways & families
  – CCA + PEN
• Applied to available aquatic datasets, identified footprints predictive of environment (potentially useful as biosensor)
• Integration of geospatial data can highlight unexpected trends as anthropogenic factors seem to be reflected in microbial function

• Specific Conclusions
  – Strong correlation exists between a community’s energy conversion strategies & env. parameters (e.g. temperature & chlorophyll)
  – Relation between Fe and P transporters & amt. of chemical in environment
    • For Fe illustrates impact of pollution & shipping
Acknowledgements

N Bhardwaj
K-K Yan
P Patel
T Gianoulis
H Yu

A Paccanaro
K Yip
R Bjornson
G Fang
Y Xia
J Korbel
J Raes
P Bork
D Engelman
M Snyder

Networks.GersteinLab.org
Default Theme

- Default Outline Level 1
  - Level 2
More Information on this Talk

SUBJECT: Networks

DESCRIPTION:
Computer Sci. Dept, Royal Holloway, University of London, London, 2010.07.02, 14:00-15:00; [I:ROYALH] (Medium-length networks talk, takes ~50 min.)

NOTES:
This PPT should work on mac & PC. Paper references in the talk were mostly from Papers.GersteinLab.org.

PERMISSIONS: This Presentation is copyright Mark Gerstein, Yale University, 2010. Please read permissions statement at http://www.gersteinlab.org/misc/permissions.html. Feel free to use images in the talk with PROPER acknowledgement (via citation to relevant papers or link to gersteinlab.org).

PHOTOS & IMAGES. For thoughts on the source and permissions of many of the photos and clipped images in this presentation see http://streams.gerstein.info. In particular, many of the images have particular EXIF tags, such as kwpotppt, that can be easily queried from flickr, viz: http://www.flickr.com/photos/mbgmbg/tags/kwpotppt.