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)

[Fleischmann et al., Science, 269 :496]  
[Chiu et al. Trends in Cell Biol, 16:144]
Networks as a universal language

- Internet [Burch & Cheswick]
- Food Web
- Electronic Circuit
- Disease Spread [Krebs]
- Protein Interactions [Barabasi]
- Neural Network [Cajal]
- Social Network

Networks are observed in various contexts:
- Internet
- Food Web
- Electronic Circuit
- Disease Spread
- Protein Interactions
- Neural Network
- Social Network

Books and references:
- Linked: The New Science of Networks by Albert-László Barabási
- Disease Spread
- Protein Interactions
- Neural Network
- Social Network

Images: Internet diagram, Food Web, Electronic Circuit, Disease Spread, Protein Interactions, Neural Network, Social Network

Do not reproduce without permission

(c) '09

3 - Lectures.GersteinLab.org
Using the position in networks to describe function

Guilt by association

Finding the causal regulator (the "Blame Game")

[NY Times, 2-Oct-05, 9-Dec-08]
Combining networks forms an ideal way of integrating diverse information

- Metabolic pathway
- Transcriptional regulatory network
- Physical protein-protein Interaction
- Co-expression Relationship

Part of the TCA cycle

Genetic interaction (synthetic lethal)
Signaling pathways
• Why Networks?

• Network Comparisons
  – Computer OS comparisons

• Integrating Networks with Molecular Structure & Motions (Dynasin)

• Network Dynamics Across Environments
  – Metabolic Pathways
  – Entry pts. (Mem. Proteins)
  – Generalizing Approach: Cross-patterns
Network Comparison: 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

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

Degree distribution
Roles of hubs

out-degree hubs e.g. “crp”

[Yan et al., PNAS (2010), in press]

% in E. coli regulatory network | % in Linux call graph
---|---
master regulator | 4.6 | 29.6
middle manager | 5.1 | 58.2
workhorse | 90.2 | 12.3

Comparison: hierarchical organization

<table>
<thead>
<tr>
<th>in-degree hubs</th>
<th>out-degree hubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. “printk”</td>
<td>e.g. “crp”</td>
</tr>
</tbody>
</table>

Degree distribution
Roles of hubs

out-degree hubs e.g. “crp”
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) = \( \frac{|M2 \cap M3|}{|M2 \cup M3|} = \frac{2}{11} \)

[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

[Yan et al., PNAS (2010), in press]
METHODOLOGY: MAP SNP AND CNV DATA ONTO ENSEMBL GENES, AND THEN MAP ENSEMBL GENES TO THE KNOWN INTERACTOME

Hapmap/Perlegen

SNPs

Interactome

~30000 interactions from HPRD and Y2H screens

Map to ENSEMBL genes

ENSEG000XXX: rsSNP00XXX
CNV_XXX
DN/DS XXXX
Recombination rate

Map to proteins in the interaction network

Database of Genomic Variants

CNVs + SDs

Result

• Dataset of network position / parameters (e.g. degree centrality or betweenness centrality) in relationship to SNPs, CNV’s, recombination rates and positive selection tests

POSITIVE SELECTION LARGELY TAKES PLACE AT THE NETWORK PERIPHERY

Positive selection in the human interactome

- High likelihood of positive selection
- Lower likelihood of positive selection
- Not under positive selection
- No data about positive selection

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

[Yan et al., PNAS (2010), in press]
Integrating Structural Interaction Network with Motion Dynamics in Human and Yeast
MOTIVATION

Network perspective:

\[ B_1 - A - B_3 \]

Structural biology perspective:

\[ \text{Cdk/cyclin complex} \]

Part of the RNA-pol complex

There remains a rich source of knowledge unmined by network theorists!

UTILIZING PROTEIN CRYSTAL STRUCTURES, WE CAN DISTINGUISH THE DIFFERENT BINDING INTERFACES

Map all interactions to available homologous structures of interfaces

PDB

Distinguish overlapping from non-overlapping interfaces

Simultaneously possible interactions: Multi-interface hub

Mutually exclusive interactions: Singlish-interface hub

SHORT DIGRESSION: THIS ALLOWS US TO DISTINGUISH SYSTEMATICALLY BETWEEN SIMULTANEOUSLY POSSIBLE AND MUTUALLY EXCLUSIVE INTERACTIONS

Simultaneously possible interactions

[Multiple Interface Hub]

Mutually exclusive interactions

[Singlish-Interface Hub]

Structural Interaction Network
(v2.0, ‘11, available for yeast, human, E coli)

- Non-hub
- Multi-interface hub
- Singlish-interface hub

Simultaneously possible interaction: Permanent
Mutually exclusive interaction: Transient

~8.5K Human & ~1.5K yeast edges

[Bhardwaj et al. ('11) Prot Sci]
dynasin.gersteinlab.org
sin.gersteinlab.org
THERE IS A PROBLEM WITH SCALE-FREENESS AND REALLY BIG HUBS IN INTERACTION NETWORKS

A really big hub (>200 Interactions)

Gedankenexperiment

How many maximum neighbors can a protein have?

Wouldn’t it be great to be able to see the different binding interfaces?

• There appears to be an obvious discrepancy between >200 and 12.

Conclusion

• Clearly, a protein is very unlikely to have >200 simultaneous interactors.
• Some of the >200 are most likely false positives
• Some others are going to be mutually exclusive interactors (i.e. binding to the same interface).

Source: DIP, Institut fuer Festkoerperchemie (Univ. Tuebingen)
THERE DO NOT APPEAR TO BE THE KINDS OF REALLY BIG HUBS AS SEEN BEFORE – IS THE TOPOLOGY STILL SCALE-FREE?

Degree distribution

Properties

- With the maximum number of interactions at 13, there are no “really big hubs” in this network.
- Note that in other high-confidence datasets (or similar size), there are still proteins with a much higher degree.
- The degree distribution appears to top out much earlier and less scale free than that of other networks.

IT’S REALLY ONLY THE MULTI-INTERFACE HUBS THAT ARE SIGNIFICANTLY MORE LIKELY TO BE ESSENTIAL

Protein Motion

- Proteins show motion due to various reasons: binding of molecules (ions, lipids, peptides, proteins), change in pH or temp
How does Motion Affect Interaction?

[P1]

`Compatible` motions

[ `Conflicting` motion

[Bhardwaj et al. (’11) Prot Sci] dynasin.gersteinlab.org
An example

Protein Kinase CK2 β

Blue: alternate conformation

Protein Kinase CK2 α

[Bhardwaj et al. ('11) Prot Sci] dynasin.gersteinlab.org
Overview

[Bhardwaj et al. ('11) Prot Sci]

dynasin.gersteinlab.org

PPI
Map structural information on to networks

Structurally classify hubs and edges

SIN
Map alternate conformations onto network

Classify compatibility of edges from motions

DynaSIN
Hubs show more motion than Non-hubs

(A) Human

P-value < 2.2e-16

Non-hubs

Hubs

(B) Yeast

P-value < 2.2e-16

Non-hubs

Hubs

[Bhardwaj et al. ('11) Prot Sci] dynasin.gersteinlab.org
Multi-interface hubs show more motion than Singlish-interface ones

**Human**

- Singlish
- Multi

P-value < 2.2e-16

**Yeast**

- Singlish
- Multi

P-value < 2.2e-16

[Bhardwaj et al. ('11) Prot Sci] dynasin.gersteinlab.org
Contrasting Singlish- vs Multiple-interface hubs

Singlish: less unique interfaces, more Transient

Multiple: more unique interfaces, more Permanent

[Bhardwaj et al. ('11) Prot Sci] dynasin.gersteinlab.org
More Unique Interfaces result in Higher RMSD

(A) Range of RMSD (Ang)

Number of unique interfaces

P<2E-16  P<2E-16  P<1E-6  P<4E-3

(B) Range of RMSD (Ang)

Number of unique interfaces

P<2E-16  P<2E-16

[Bhardwaj et al. ('11) Prot Sci] dynasin.gersteinlab.org
Permanent Interactions more likely associated with Conflicting Motions

Fraction of conflicting motions

\[
\text{Fraction of conflicting motions} = \frac{\text{# of conflicting motions}}{\text{# of total motions}}
\]

Permanent/Transient?

(A) 0.45

Fraction of conflicting motions

Permanent Transient

0.40

Human

0.35

(B)

Fraction of conflicting motions

0.8

Permanent Transient

0.6

0.4

0.2

0.0

Yeast

[Refs: [Bhardwaj et al. (11) Prot Sci] dynasin.gersteinlab.org]
Rationalization: Permanent vs Transient Interactions

[Bhardwaj et al. (‘11) Prot Sci] dynasin.gersteinlab.org
Real examples: permanent interaction

Importin subunit beta-1 with Snurportin-1 (1.6 Å (7.9 Å)) and with GTP-binding nuclear protein RAN (2.3 Å).

Mitotic spindle assembly checkpoint protein Mad2a with itself (dimer) (6.7 Å) and with MAD1-Like 1 (4.4 Å).

[Bhardwaj et al. ('11) Prot Sci] dynasin.gersteinlab.org
Real examples: transient interaction

- IGG receptor FCRN large subunit P51
- HFE
- T-cell surface glycoprotein CD1d
- T-cell surface glycoprotein CD1a
- HLA class I histocompatibility antigen α
- HLA class I histocompatibility antigen, CW-4

β-2-microglobulin

MHC Class I histocompatibility antigen

Distances:
- 0.6 Å
- 0.7 Å
- 0.8 Å
- 0.5 Å

[Bhardwaj et al. (11) Prot Sci] dynasin.gersteinlab.org
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

Select organism and culture

Extract DNA and sequence

Assemble and annotate

Estimated that less than 1% of microbes can be cultured

Metagenomics

Collect sample from environment

Extract DNA and sequence

Assemble and annotate

Lose information about which gene belongs to which microbe

Contig 1

Contig 2

. . .

atgctcgatctcg
atcgatctcgctg
atgccgatctaa
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

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

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

Additional Metadata via GPS coordinates

Membrane Protein Families

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

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

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

[ Gianoulis et al., PNAS (in press, 2009) ]
Canonical Correlation Analysis: Simultaneous weighting

Lifestyle Index = $a \cdot \text{km run/week} + \text{Weight} + b \cdot \text{Fit Index} + c$

Fit Index = $a \cdot \text{Running} + b \cdot \text{Bike} + c$
Canonical Correlation Analysis: Simultaneous weighting

Lifestyle Index = \[ a \times \text{# km run/week} + b \times \text{Weight} + c \]

Fit Index = \[ a \times \text{Environmental Features} + b \times \text{Metabolic Pathways/Protein Families} + c \]

Environmental Features
- Temp
- Chlorophyll
- etc

Metabolic Pathways/Protein Families
- Photosynthesis
- Lipid Metabolism
- etc
CCA: Finding Variables with Large Projections in "Correlation Circle"

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

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
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., (‘10) Genome Research
CCA Limitations

Both the strength and the directionality of relationships between nodes is difficult to decipher in this format.

Patel and Gianoulis et al., (‘10) Genome Research
Protein Families and Environmental Features Network (PEN)

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

$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta$

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
Bi-module 2: Iron Transporters/Pollution/Shipping

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

Carr and Church, Nat Biotech 2009
Cross Patterns:  
What is stacking? What are its limitations?  
Can we identify “cross-patterns” between differently indexed systems data?  
What formalism can express this new type of relationship?
PCA \Rightarrow CCA \Rightarrow \text{Cross Patterns} \Rightarrow \text{CRIT}

[Gianoulis et al. ('11) GenomeBiology]
PCA => CCA => Cross Patterns => CRIT

[Pathways] => [Env. Features] => [Commercial Impacts]

[Housing, Transportation, Healthcare...]

[Gianoulis et al. (’11) GenomeBiology]
PCA => CCA => Cross Patterns => CRIT

[Sites]

[Transfer of L1]

[Transfer of L2]

[Gene Properties (Charge, Size...)]

[Gianoulis et al. ('11) GenomeBiology]
PCA => CCA => Cross Patterns => CRIT

[Image: Diagram showing the relationships between PCA, CCA, Cross Patterns, and CRIT with specific mentions of DRUGS and PROTEINS with associated drug properties and protein properties.]
Cross Pattern Identification Technique (CRIT)

A. Traditional Single Gene Feature Integration (Stacking)

Gene-Centric Integration

Gene Pair-Centric Integration

B. Gene-Pair Features Integration (Networks)

C. Indirect Integration - Indexing on Sites

D. Indirect Integration - Transitivity

Drug-Features → Drug → Target → Target-Features

Drug Feature-Target Feature Cross-Patterns

M1, R1
M2, R3
M3, R2, R3

Metabolites
Tissues
Environmental Stress . . .
Connector Matrix

[Image of Connector Matrix]

[Image of TF charge, TF MW, site GC sites]

[Image of GenomeBiology article reference]

[Image of ORF distribution]

[Gianoulis et al. ('11) GenomeBiology] CRIT.gersteinlab.org
CRIT Algorithm

**Labeler:** Transfers label on columns of previous dataset to rows of new dataset.

---

[Gianoulis et al. (’11) GenomeBiology] CRIT.gersteinlab.org
CRIT Algorithm

**Labeler:** Transfers label on columns of previous dataset to rows of new dataset.

**Slicer:** Partitions rows into dark and light green slices.

[Gianoulis et al. (‘11) GenomeBiology] CRIT.gersteinlab.org
**CRIT Algorithm**

**Labeler:** Transfers label on columns of previous dataset to rows of new dataset.

**Slicer:** Partitions rows into dark and light green slices.

**Discriminator:** Returns a label for the columns based on whether the slices (from the rows) are sig different.

\[ L^n = f^n(M^n, L^{n-1}) \]

[Gianoulis et al. (‘11) GenomeBiology] CRIT.gersteinlab.org
**CRIT Algorithm**

**Labeler:** Transfers label on columns of previous dataset to rows of new dataset.

**Slicer:** Partitions rows into dark and light green slices.

**Discriminator:** Returns a label for the columns based on whether the slices (from the rows) are sig different.

\[ L^n = f^n(M^n, L^{n-1}) \]

Repeat...

TF Example

Gianoulis et al. Genome Biology 2011
Connector Matrix: Chemogenomics

[Gianoulis et al. (‘11) GenomeBiology] CRIT.gersteinlab.org (Skip?)
Calculating Drug Properties

Hillenmeyer et al. (2008) Science

[François Gianoulis et al. ('11) GenomeBiology] CRIT.gersteinlab.org (Skip?)
Chemogenomics: Identifying Cross Patterns

[Gianoulis et al. ('11) GenomeBiology] CRIT.gersteinlab.org (Skip?)
Cross Patterns

[a] 6 Drug Properties
281 DRUGS

Connector
1194 PROTEINS

[b] DRUG PROPERTIES
Molecular Weight (MW)
Charge
# of Aromatic Bonds (AB)
# of Aromatic Rings (AR)
Hydrophilicity
MlogP

PROTEIN PROPERTIES
Physicochemical & Composition
Localization
GO Process
GO Function
Network Stats
Environmental Stress

[Gianoulis et al. ('11) GenomeBiology] CRIT.gersteinlab.org  (Skip?)
Conclusion: Prop. of Drugs Mirror Direct Prop. of Protein

102 Charge “sensitive” proteins enriched for membrane proteins.

Charge  Growth Defect

[Gianoulis et al. ('11) GenomeBiology] CRIT.gersteinlab.org (Skip?)
• Why Networks?

• Integrating Networks with Molecular Structure & Motions (Dynasin)

• Network Dynamics Across Environments
  ◊ Metabolic Pathways
  ◊ Entry pts. (Mem. Proteins)
  ◊ Generalizing Approach: Cross-patterns

• Network Comparisons
  ◊ Computer OS comparisons
<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>Organization of modules</strong></td>
<td>Structure</td>
<td>Pyramidal</td>
</tr>
<tr>
<td>Characteristic hubs</td>
<td>Upper-level TFs with high out-degree</td>
<td>Generic workhorse functions with high in-degree</td>
</tr>
<tr>
<td>Downstream modules as labeled by</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Node reuse</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Overlap between modules</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Persistent nodes</strong></td>
<td>Characteristics</td>
<td>Specialized (non-generic) workhorses</td>
</tr>
<tr>
<td>Location in hierarchy</td>
<td>Mostly bottom</td>
<td>Mostly top</td>
</tr>
<tr>
<td>Evolutionary rate</td>
<td>Mostly conservative (e.g. dnaA)</td>
<td>Conservative (e.g. strlen) &amp; adaptive (e.g. mempool_alloc)</td>
</tr>
<tr>
<td><strong>Design principles</strong></td>
<td>Building of hierarchy</td>
<td>Bottom up</td>
</tr>
<tr>
<td>Optimal solution favors</td>
<td>Robustness</td>
<td>Cost-effectiveness (reuse of components)</td>
</tr>
</tbody>
</table>
Dynasin Summary

• Molecular interaction networks can be integrated with structure and motions, giving more realistic connectivity patterns

• Hub, particularly, multi-interface ones are more mobile

• Permanent interactions (which tend to occur in multi-interface hubs) more often involve conflicting motions than transient ones
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
CRIT Summary

• To leverage new data we need to move beyond gene-centric integration.
• We developed a formalism to test for and express cross patterns.
• We applied this method to identify cross patterns between TF properties and those of their target genes (and also between properties of drugs and their gene targets).
N Bhardwaj
K-K Yan
P Patel
T Gianoulis
D Clarke
P Kim

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

Job opportunities currently for postdocs & students

Networks.GersteinLab.org
Default Theme

• Default Outline Level 1
  – Level 2
More Information on this Talk

SUBJECT: Networks

DESCRIPTION:

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.