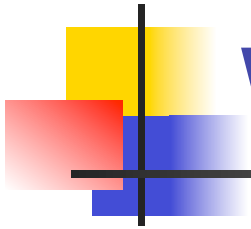
Modeling biomolecular networks: from metabolism and its regulation to protein-protein interactions

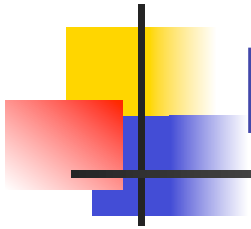
Sergei Maslov

Brookhaven National Laboratory



What is a complex system?

- Complex systems have many interacting components (10^{11} neurons, 10^{4} types of proteins, 10^{6} routers, 10^{9} web pages)
- All components are different from each other
- Systems traditionally studied by physics also have many interacting components (10^{23} electrons in a superconductor)
- But they are all the same!



Networks in complex systems

- Since components are different, first question: which pairs directly interact?
- The answer can be visualized as a network
- Network is the backbone of the underlying complex system
- In my talk I will first propose a model of evolution of the backbone (Part 1) and then put dynamics on it (Part 2)

Part 1: Parkinson’s law in biology

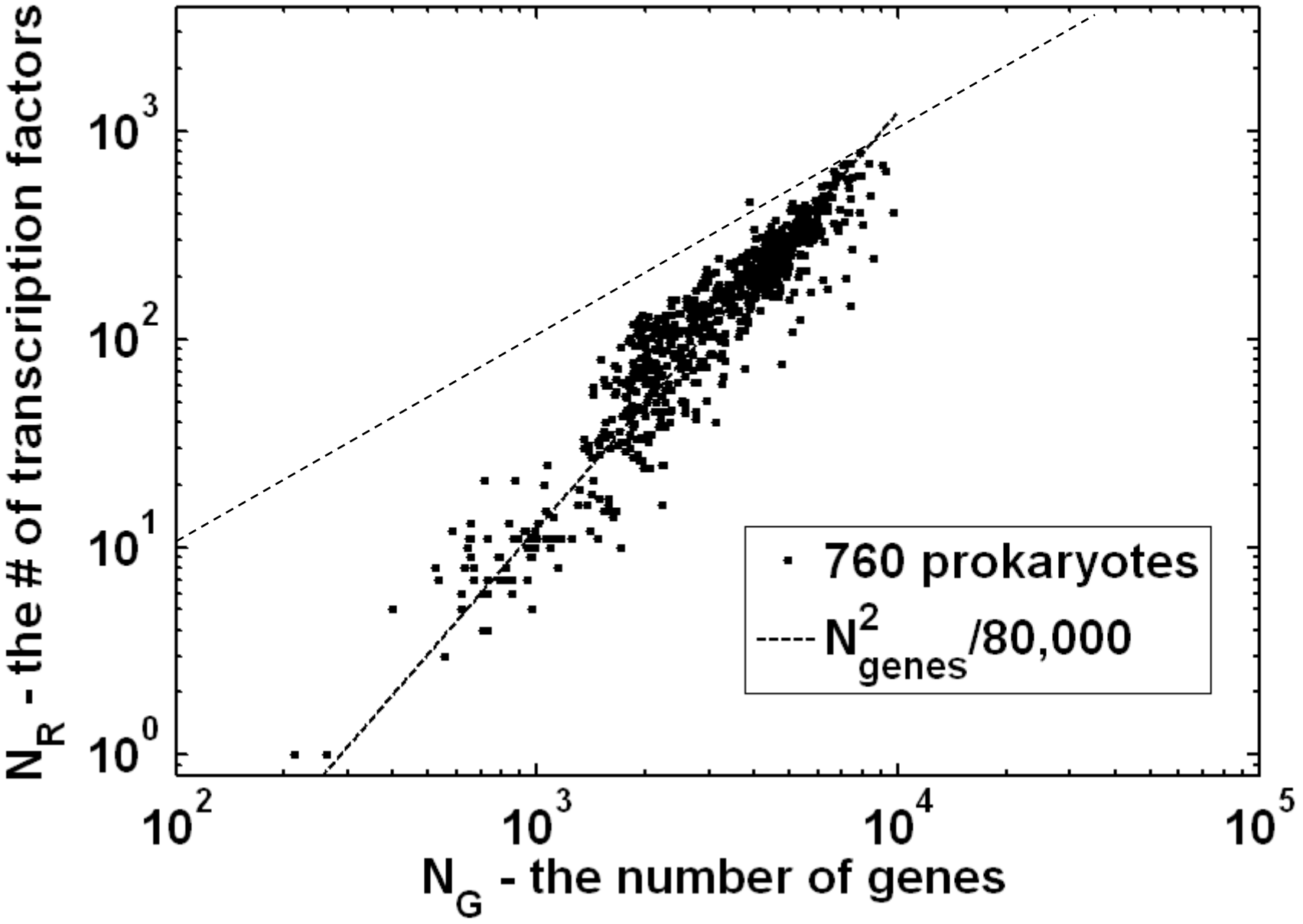
The Economist

Parkinson's Law

The report of the Royal Commission on the Civil Service was published on Thursday afternoon. Time has not permitted any comment in this week's issue of The Economist on the contents of the Report. But the startling discovery enunciated by a correspondent in the following article is certainly relevant to what should have been in it.

Nov 19th 1955 | From *The Economist* print edition

Stover *et al*., Nature (2000) van Nimwegen, TIG (2003)



Let's play with this scaling law

- $N_R = N_G^2/80{,}000$ --> $\Delta N_R = \Delta N_G \, 2N_G/80{,}000$
- When a new regulated function is added $\Delta N_R = +1$, $\Delta N_G / \Delta N_R = 40{,}000/N_G$
  - ~40 new genes per function for $N_G = 1000$
  - ~4 new genes (1 regulator + 3 non-regulatory genes) for the largest bacterial genomes with $N_G \sim 10{,}000$
- One needs to explain why $\Delta N_G / \Delta N_R$ systematically decreases with genome size as $1/N_G$

"Home Depot" or toolbox model

PNAS

Toolbox model of evolution of prokaryotic metabolic networks and their regulation

Sergei Maslov[a,1], Sandeep Krishna[b], Tin Yau Pang[a,c], and Kim Sneppen[b]

[a]Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, NY 11973; [b]Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark; and [c]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3800

Edited by David J. Lipman, National Institutes of Health, Bethesda, MD, and approved April 16, 2009 (received for review March 23, 2009)

It has been reported that the number of transcription factors encoded in prokaryotic genomes scales approximately quadratically with their total number of genes. We propose a conceptual explanation of this finding and illustrate it using a simple model in which metabolic and regulatory networks of prokaryotes are

A simple evolutionary model explains both these empirical observations in a unified framework based on modular functional design of prokaryotic metabolic networks and their regulation.

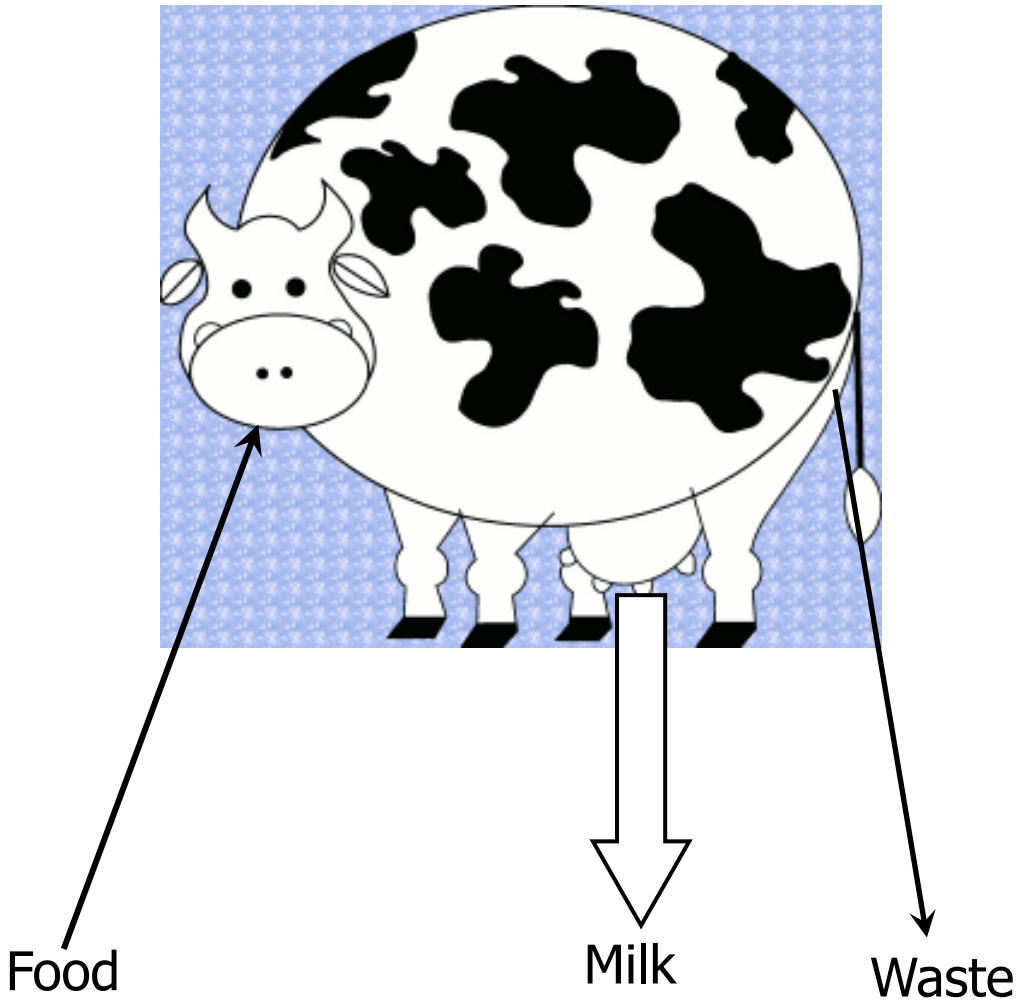
Toolbox View of Metabolic Networks

Disclaimer: authors of this study (unfortunately) received no financial support from Home Depot, Inc. Homebase, LTD or Obi, GMBH

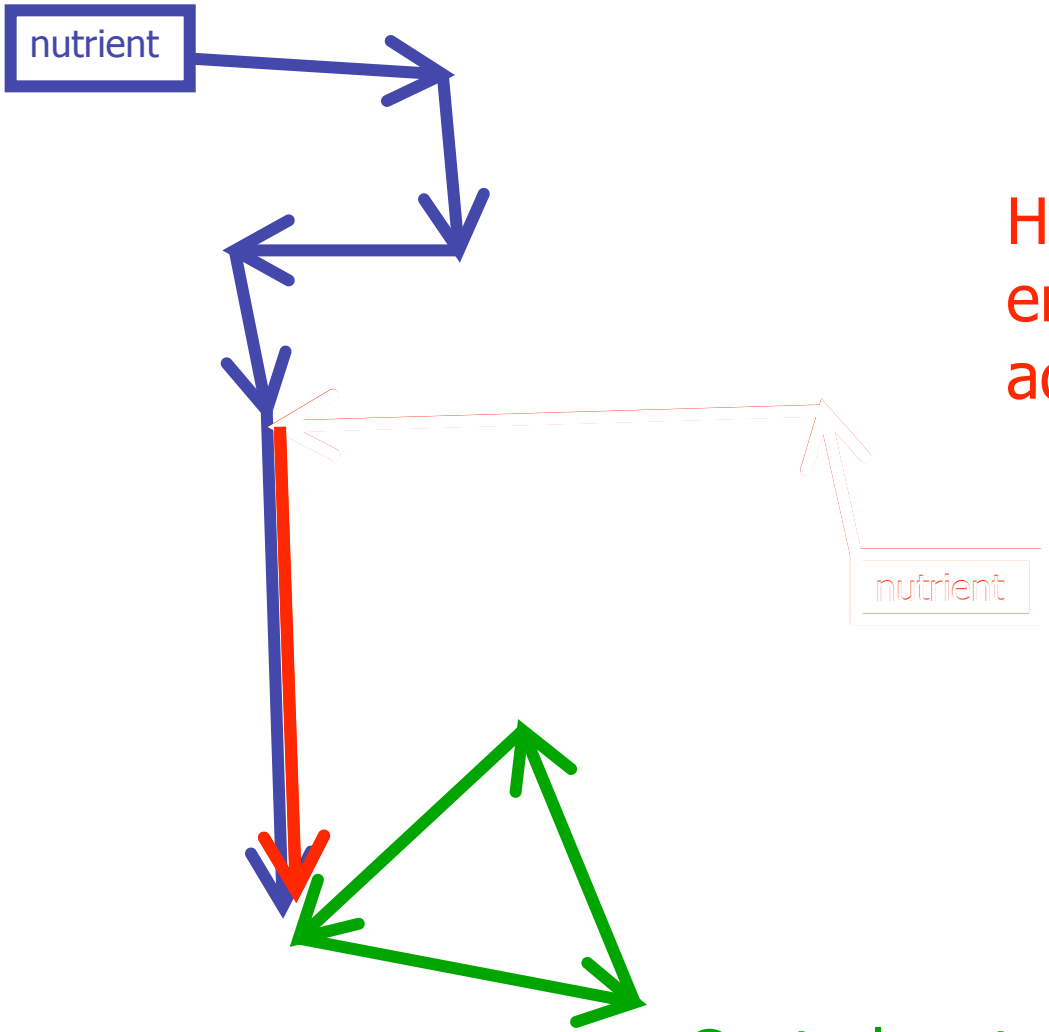
"Home Depot" argument

- Inspired by personal experience as a new homeowner
- Tools are bought to accomplish functional tasks e.g. fix a leaking faucet
- Duplicate tools are returned to "Home Depot"
- As your toolbox grows you need to get fewer and fewer new tools to accomplish a new task
- Bacteria also have tools encoded by non-regulatory "workhorse" genes (e.g. for metabolic enzymes)
- Entire pathways (collections of tools) are acquired from other bacteria by Horizontal Gene Transfer
- Regulators control these pathways (we assume one regulator per task/pathway)
- Redundant genes are promptly deleted (in prokaryotes)
- As the genome gets larger you need fewer new genes per new regulated function – FASTER THAN LINEAR SCALING

Spherical cow model of metabolic networks



Pathways could be also removed



Horizontal gene transfer: entire pathways could be added in one step

Central metabolism → anabolic pathways → biomass

• New functions are added by Horizontal Gene Transfer of entire pathways (collections of tools)

• They come from the universal network of size N_{univ} composed of all reactions in all organisms (bacterial answer to "Home depot")

• The current size of the toolbox (~# of genes ~ # of enzymes ~ # of metabolites): N_G

• Probability to join the existing pathway: $p_{join} = N_G / N_{univ}$

• $L_{pathway} = 1/p_{join} = N_{univ}/N_G$

• Assume one regulator per function/pathway: $\Delta N_G / \Delta N_R = L_{pathway} = N_{univ}/N_G$

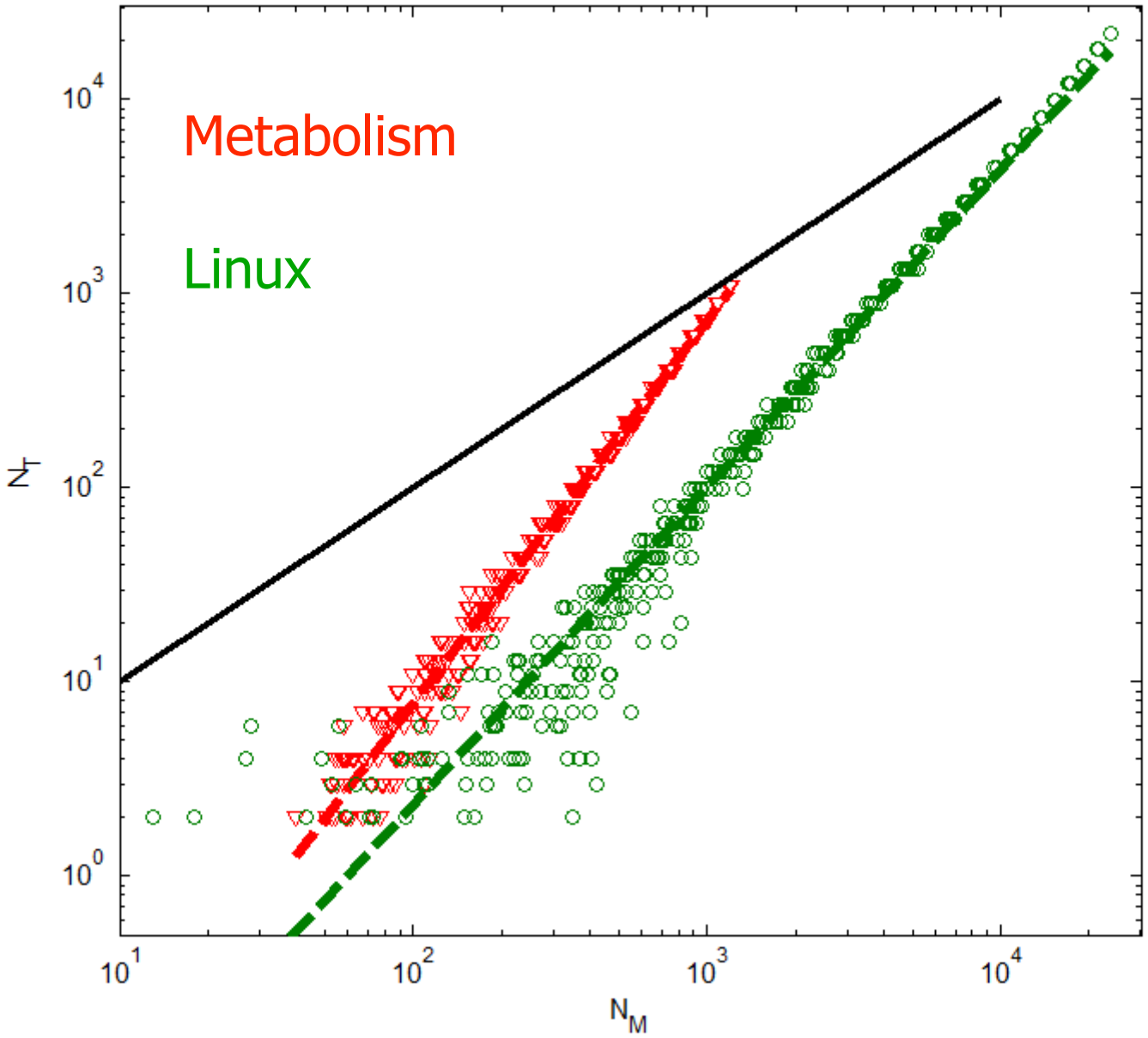
• Quadratic law: $N_R = N_G^2 / 2N_{univ}$

Different universal networks give the same result

- Random branched network: analytically solved to give $N_R \sim N_{met}^{2}$
- Union of all metabolic reactions in the KEGG database: numerically solved to give $N_R \sim N_{met}^{2.0 +/- 0.3}$

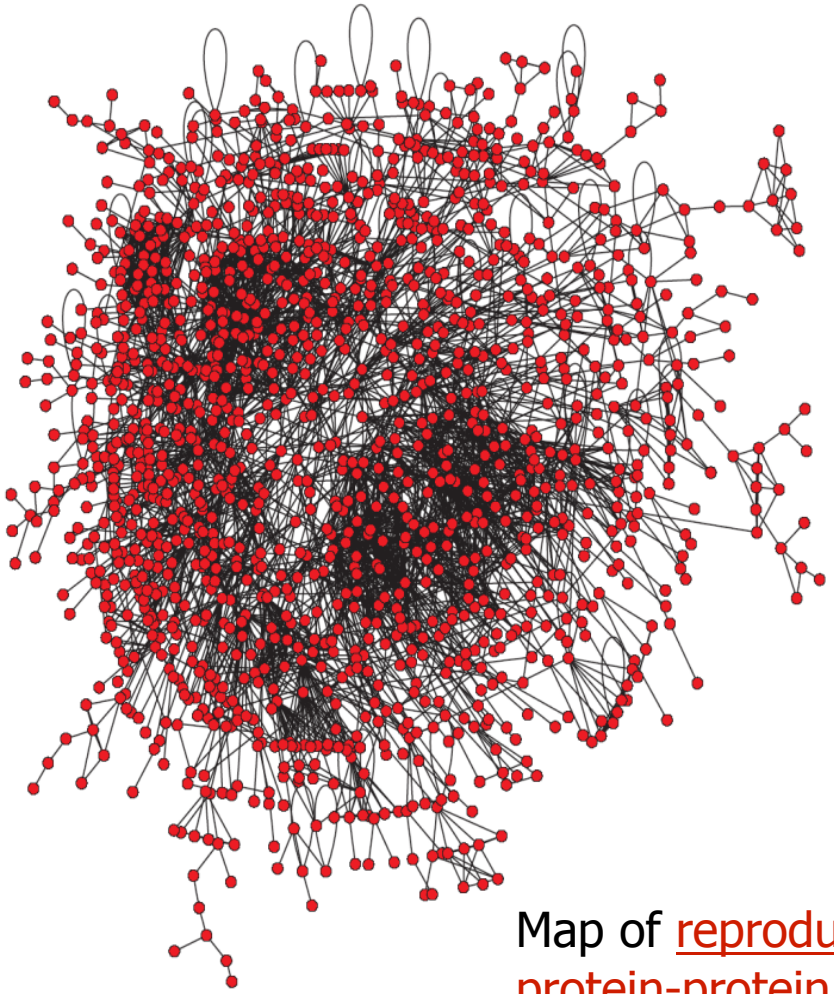
"Home Depot" model is not limited to biology

- Adding new units becomes easier as system grows: ECONOMY OF SCALE
- Quadratic scaling is expected in any multi-level system formed by mutually-dependent units
- Software packages installed in Linux
  $N_{\text{usable functions}} \sim N_{\text{installed packages}}^{1.7+/-0.3}$
- Expected: in networks of interdependent technological innovations (e.g. patents using other patents), supply networks of companies


Metabolism
Linux
10^4
10^3
10^2
10^1
10^0
N_T
10^1
10^2
10^3
10^4
N_M

Part 2: Mass Action Equilibrium in protein binding networks

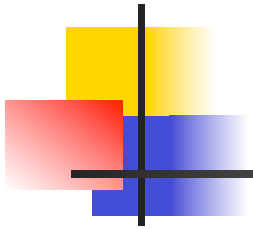
Small world of protein-protein interaction networks



Map of reproducible (>2 publications) protein-protein interactions in yeast

- >80% of proteins are all connected in one giant cluster of PPI network
- Small-world effect median network distance – 6 steps

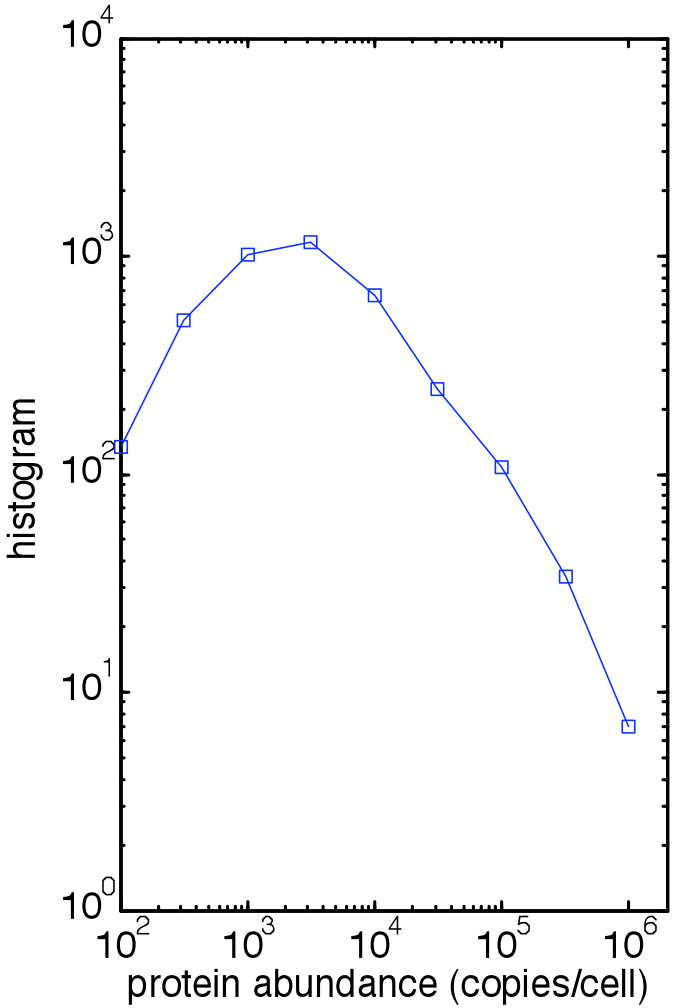
Why small-world property might cause problems



- Are small world networks more robust?
- Internet – more connected is better
- Small world binding networks could indiscriminately spread perturbations
  - Systematic changes: large deterministic changes in concentrations of a small number of proteins SM, I. Ispolatov, PNAS and NJP (2007)
  - Noise: small changes in concentrations of a large number of proteins K.-K. Yan, D. Walker, SM, PRL (2008)

My "spherical cow" assumptions

- Protein concentrations C_i of all yeast proteins (under the rich growth medium conditions) and subcellular localizations are experimentally known (group of Weissman @ UCSF)
- Consider only reproducible independently confirmed protein-protein interactions for non-catalytic binding (kinase-substrate pairs~5%)
- The network: ~4000 heterodimers and ~100 multi-protein complexes (we assume no cooperative binding in complexes) connecting ~1700 proteins
- Know the relevant average of dissociation constants K_{ij} ~10nM. Turned out their distribution around this average DOES NOT MATTER MUCH!!!
- Use "evolutionary motivated" binding strength: $K_{ij}=\max(C_i, C_j)/\text{const}$, which is sufficient to bind considerable fraction of twoproteins in a heterodimer



Law of Mass Action (LMA)

- $dD_{AB}/dt = r^{(on)}_{AB} F_A F_B - r^{(off)}_{AB} D_{AB}$
- In the equilibrium:

  $D_{AB}=F_A F_B/K_{AB}$; $C_A= F_A+D_{AB}$; $C_B= F_B +D_{AB}$

  or $F_A = C_A/(1+ F_B/K_{AB})$ and $F_B = C_B/(1+ F_A/K_{AB})$

- In a network:
  A system of ~2000 nonlinear equations for F_i that can be solved only numerically

$$F_i = \frac{C_i}{1 + \sum_{jnni.} FK_{jij}}$$

Propagation of perturbations: the *in silico* study

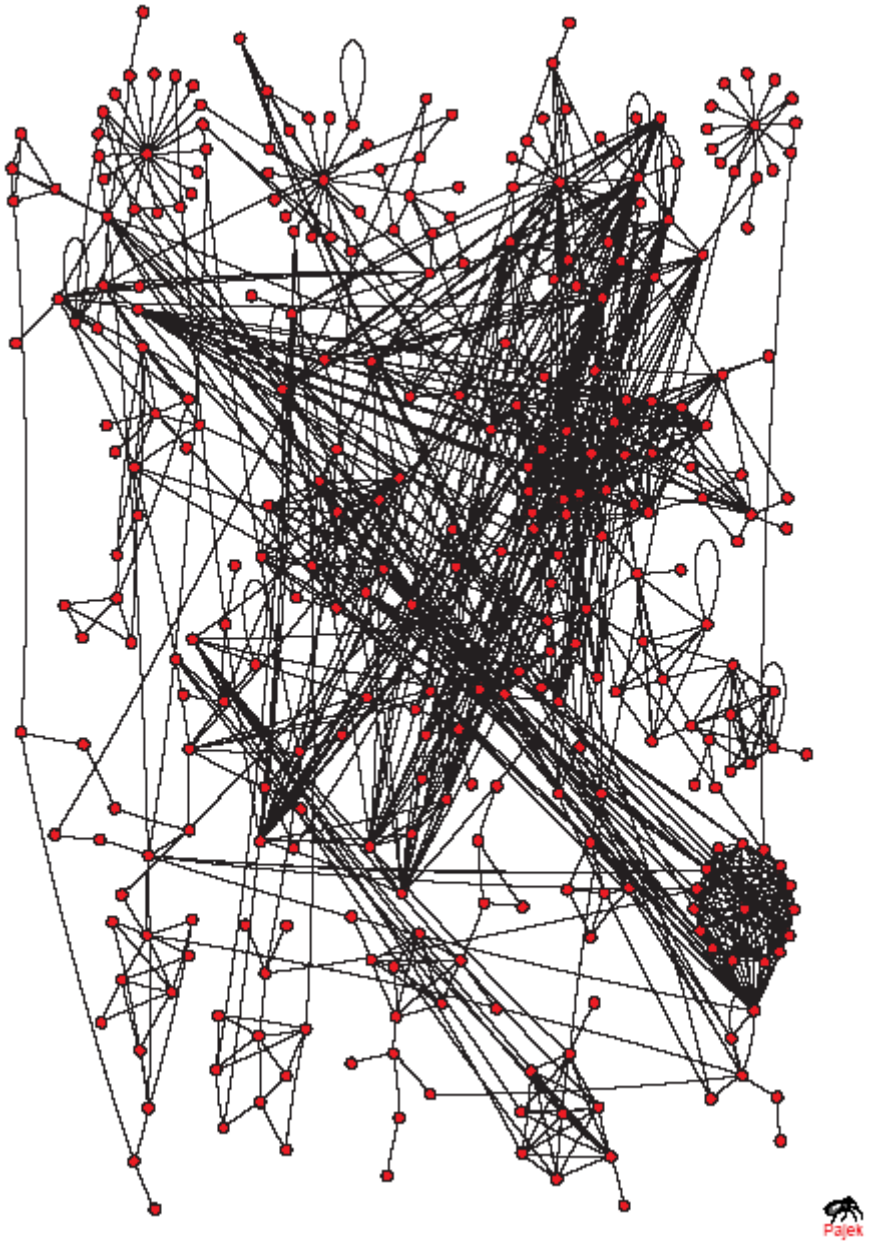
- Calculate the unperturbed mass action equilibrium
- Simulate a twofold increase of the concentration $C_A \rightarrow 2C_A$ of just one type of protein and recalculate equilibrium free concentrations F_i of all other proteins
- Look for cascading perturbations:
  A→ B → C → D with sign-alternation:
  A (↑ up), B (↓ down), C (↑ up), D (↓ down)

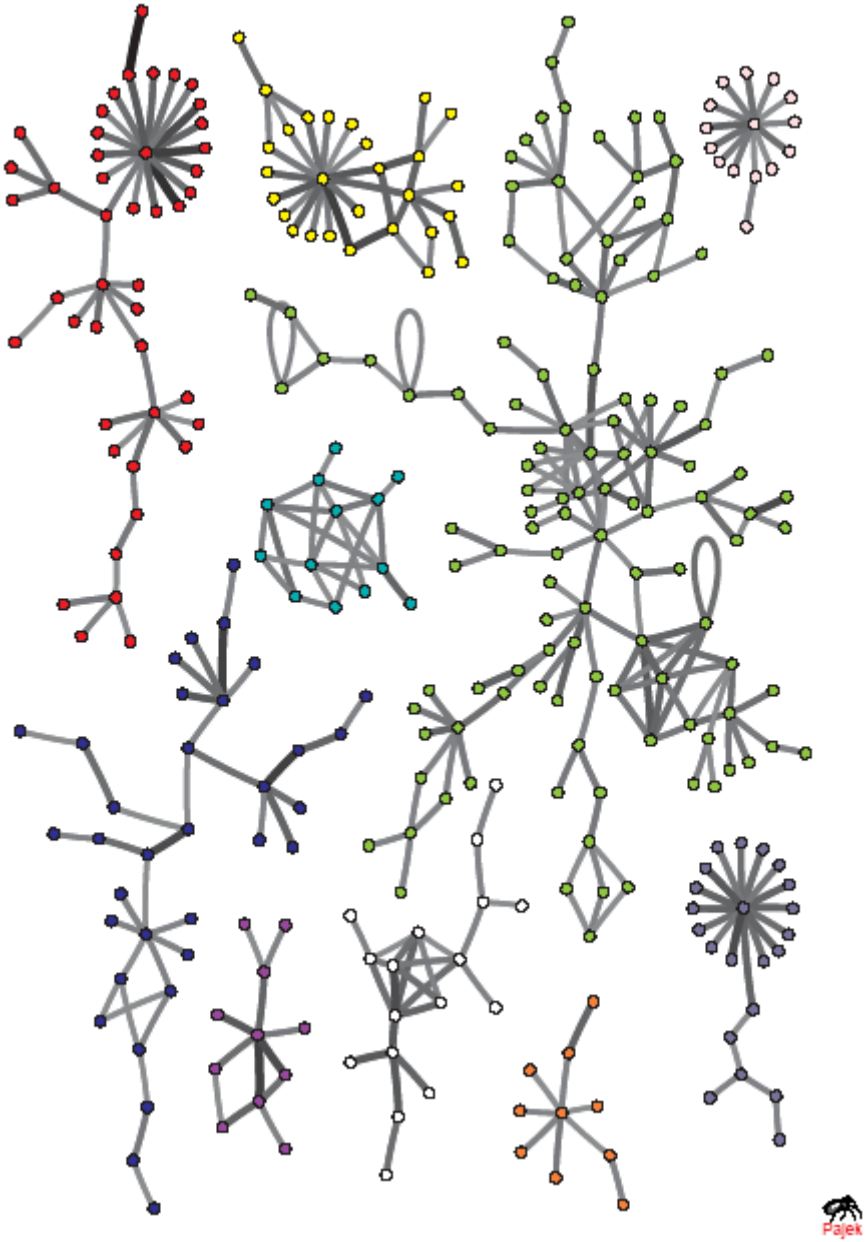
Propagation of large concentration changes in reversible protein-binding networks

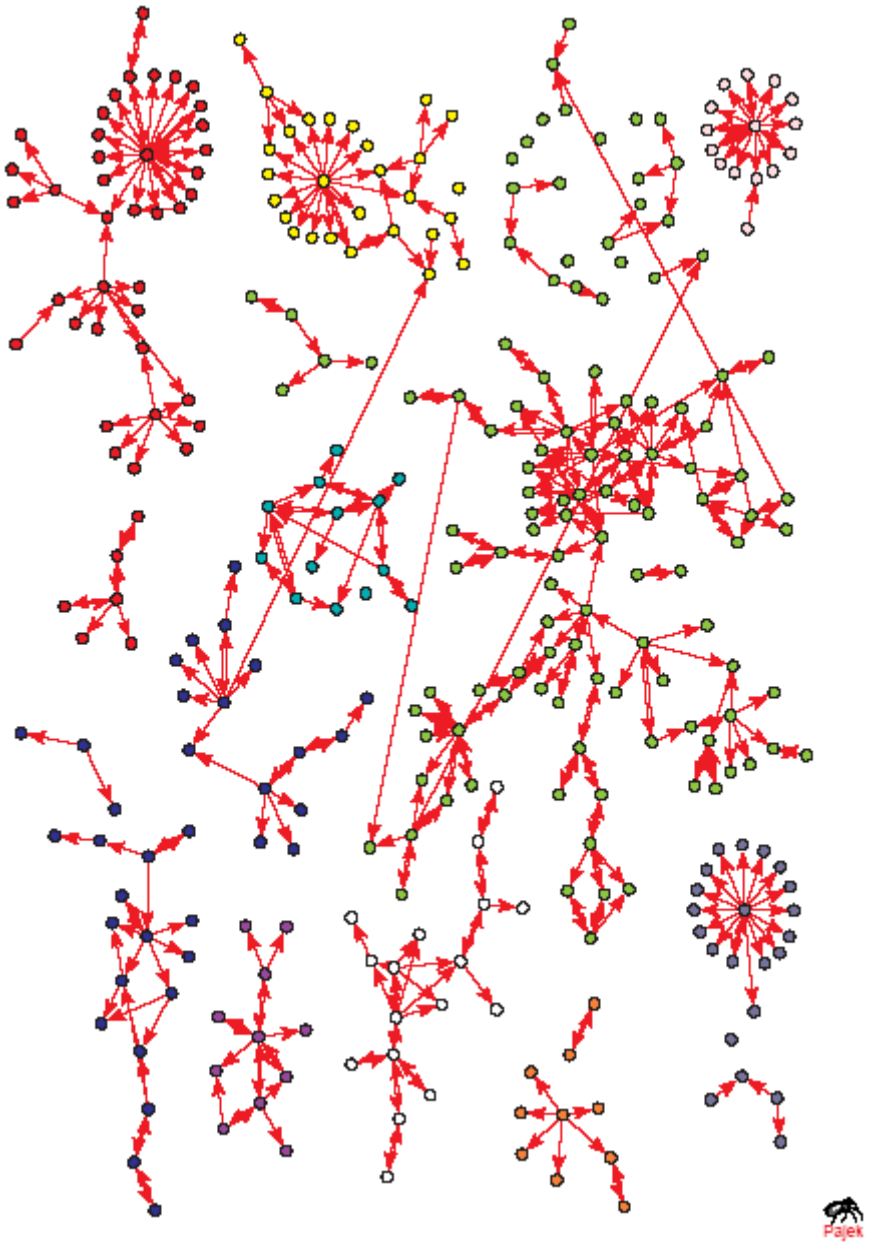
Sergei Maslov*† and I. Ispolatov†‡

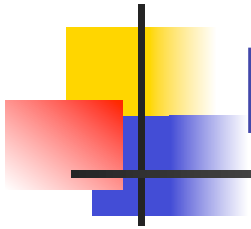
*Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, NY 11973; and ‡Ariadne Genomics, Inc., 9430 Key West Avenue, Suite 113, Rockville, MD 20850

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved June 21, 2007 (received for review March 28, 2007)


Pajek


Pajek


Pajek



Mapping to resistor network

- Conductivities σ_{ij} – heterodimer concentrations D_{ij}
- Losses to the ground σ_{iG} – free (unbound) concentrations F_i
- Perturbations spread along linear chains
  loosely conducting to neighbors and ground
- Mapping is exact for bi-partite networks → odd-length loops dampen perturbations

S.Maslov, K. Sneppen, I. Ispolatov, New J. Phys, (2007)

Collaborators and support

- Kim Sneppen (Center for Models of Life, NBI, Denmark)
- Sandeep Krishna (Center for Models of Life, NBI, Denmark)
- Tin Yau Pang (Stony Brook U and BNL)

- Jaroslav Ispolatov (Ariadne Genomics Inc)
- Koon-Kiu Yan (BNL → Yale U)
- Dylan Walker (BNL → NYU)

- DOE Division of Material Science, DE-AC02-98CH10886
- NIH/NIGMS, R01 GM068954
- DOE Grant on Systems Biology Knowledgebase

I have 1-2 postdoc positions to work on toolbox model.
If interested talk to me

Thank you!