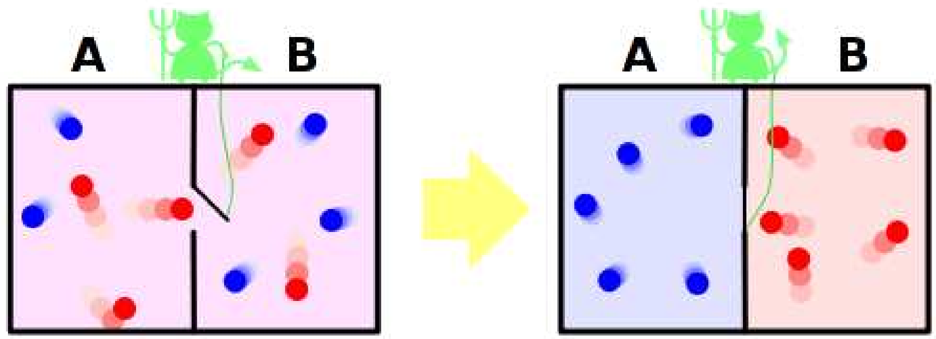
Role of Information in Physical Sciences

Some Examples from Life Sciences

Information and the Sciences – a Brief History

- Connections between information and the real world have intruigued researchers for centuries.

- Maxwell's thought experiment (circa 1867) in relation to the Second Law of Thermodynamics provides an elegant nexus.

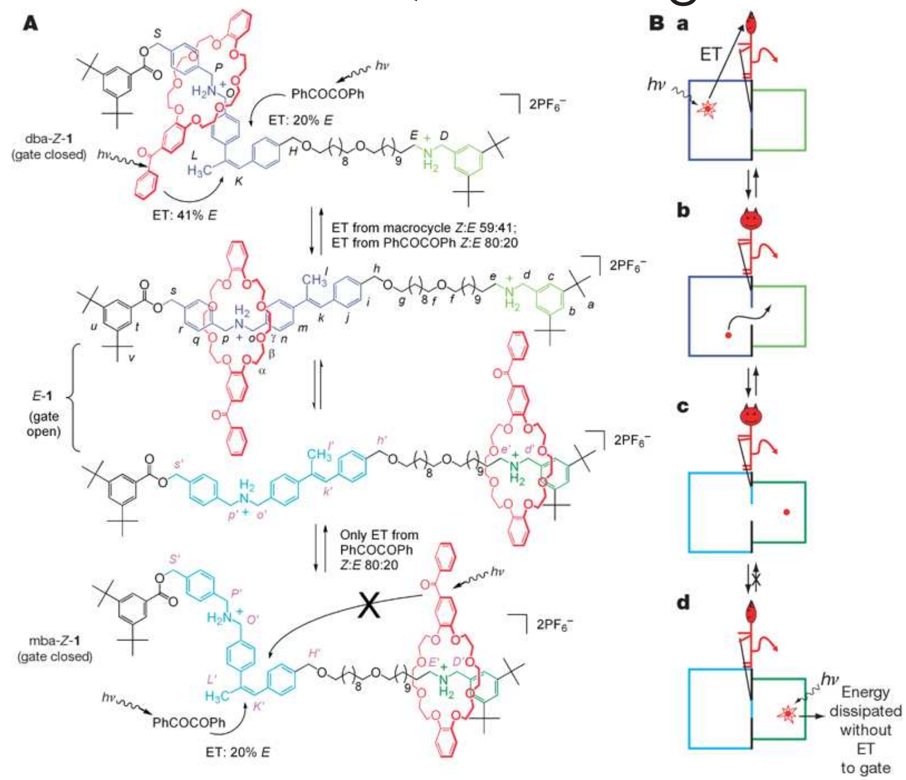


Maxwell's demon was one of the first observations relating energy to information.

- Szilard and Brillouin subsequently used this to equate one bit of information with $K_B T \ln 2$ joules of energy.

Information and the Sciences – a Brief History

Maxwell's Demon Realized (David Leigh, Nature, 2007)



Irradiation of rotaxane 1 at 350 nm in CD_3OD at 298 K interconverts the three diastereomers of 1 and, in the presence of benzil, drives the ring distribution away from the thermodynamic minimum, increasing the free energy of the molecular system without ever changing the binding strengths of the macrocycle or ammonium binding sites.

Information and the Sciences – a Brief History

- As the science of information developed, so did interest in correspondence between information theoretic and scientific measures.

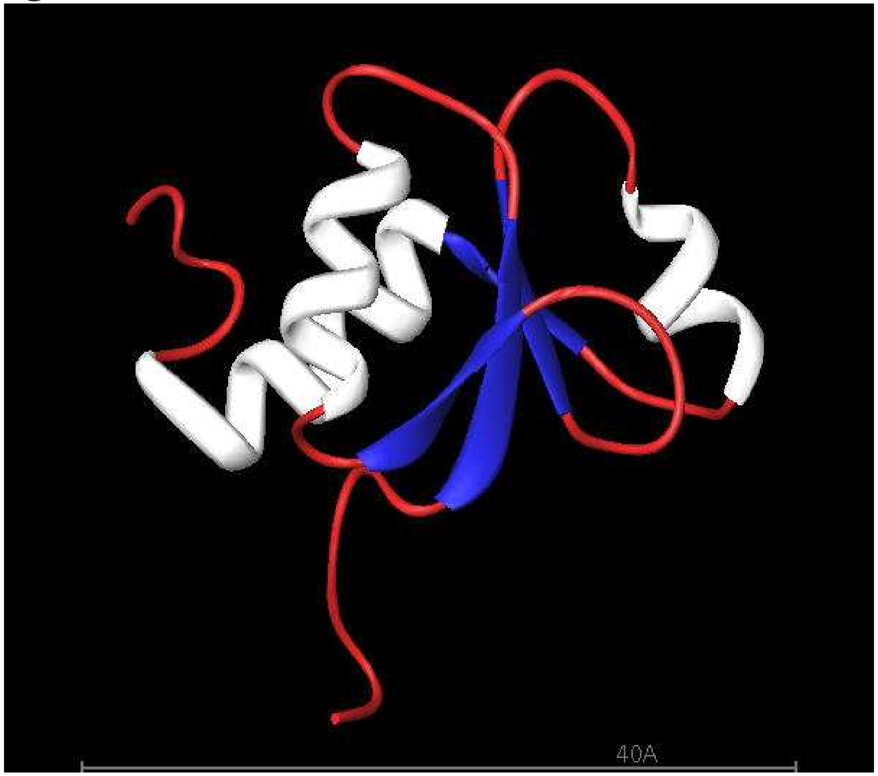
- Among the more intuitive and well studied is the correspondence between thermodynamic (Boltzmann-Gibbs) entropy and information theoretic (Shannon-Hartley) entropy.

- Drawing on these, two questions arise:
  - Can we draw on information theoretic formalisms to address foundational questions in scientific disciplines?
  - Can we draw on physical principles to address basic questions in computing?

- This interplay between information and science is the focus of the proposed institute.

Role of Information in Scientific Discovery: Examples

What is the "informative" component of XRCC1, BRCA-1, and H. sapiens DNA ligase III?



BRCT domain, from T.Thermophilus DNA Ligase.

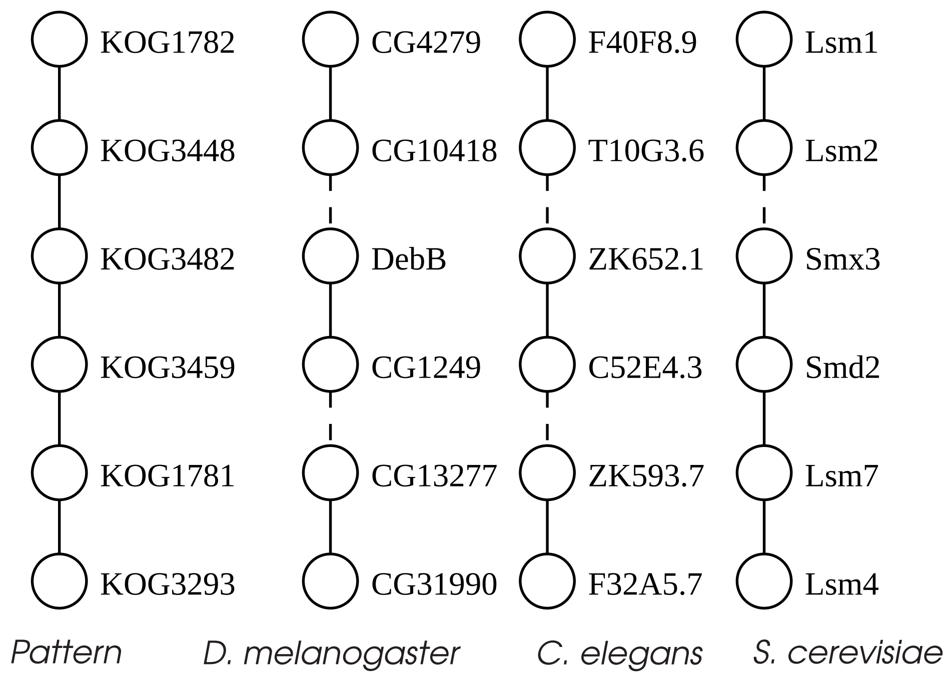
These are inferred using a variety of information correlation and extraction techniques (sequence analysis, Markov models), and experimentally validated.

Role of Information in Scientific Discovery: Examples

What is the "infomative" component of a given set of PPI networks?

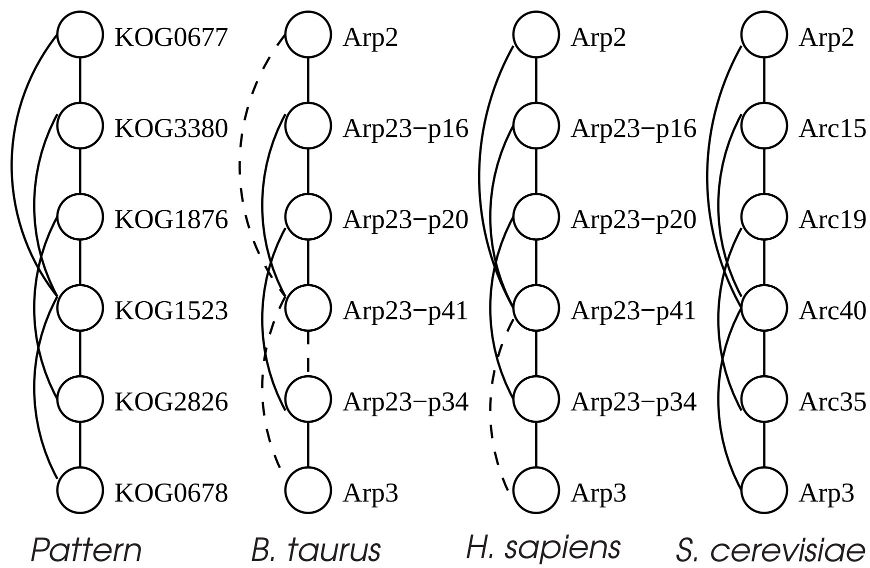
- PPI networks for 9 eukaryotic organisms derived from BIND and DIP
  - *A. thaliania, O. sativa, S. cerevisiae, C. elegans, D. melanogaster, H. sapiens, B. taurus, M. musculus, R. norvegicus*
  - # of proteins ranges from 288 (*Arabidopsis*) to 8577 (*fruit fly*)
  - # of interactions ranges from 340 (*rice*) to 28829 (*fruit fly*)
- Ortholog contraction
  - Group proteins according to existing COG ortholog clusters
  - Merge Homologene groups into COG clusters
  - Cluster remaining proteins via BLASTCLUST
  - Ortholog-contracted *fruit fly* network contains 11088 interactions between 2849 ortholog groups
- MULE is available at http://www.cs.purdue.edu/pdsl/

Conserved Protein Interaction Patterns



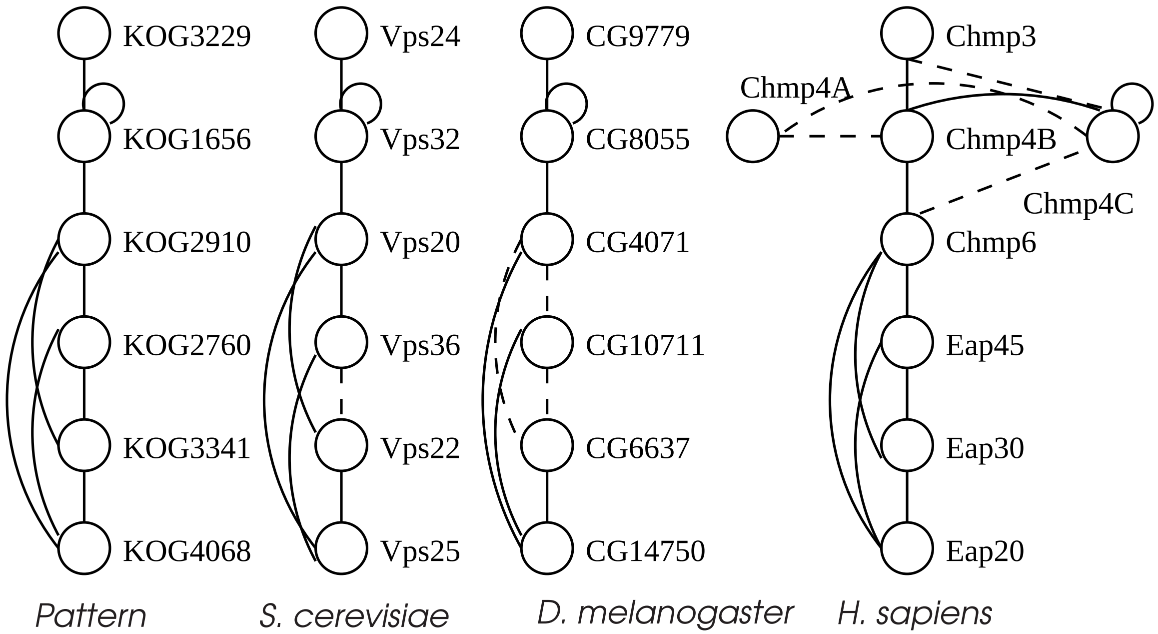
Small nuclear ribonucleoprotein complex ($p < 2e - 43$)

Conserved Protein Interaction Patterns



Actin-related protein Arp2/3 complex ($p < 9e-11$)

Conserved Protein Interaction Patterns



Endosomal sorting ($p < 1e - 78$)

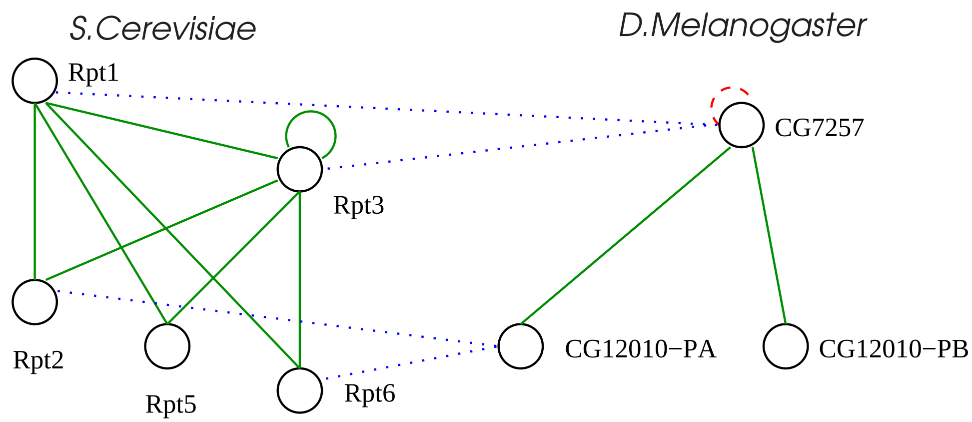
Role of Information in Scientific Discovery: Examples

What is the "informative" component shared by two given PPI networks (Yeast and Fruit Fly)?

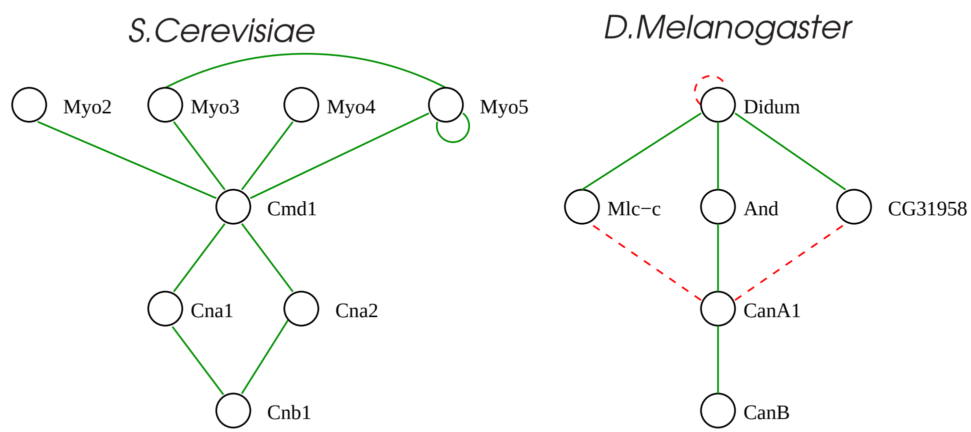
Rank	Score	z-score	# Proteins	# Matches	# Mismatches	# Dups.
1	15.97	6.6	18 (16, 5)	28	6	(4, 0)
	protein amino acid phosphorylation (69%)					
	JAK-STAT cascade (40%)					
2	13.93	3.7	13 (8, 7)	25	7	(3, 1)
	endocytosis (50%) / calcium-mediated signaling (50%)					
5	8.22	13.5	9 (5, 3)	19	11	(1, 0)
	invasive growth (sensu Saccharomyces) (100%)					
	oxygen and reactive oxygen species metabolism (33%)					
6	8.05	7.6	8 (5, 3)	12	2	(0, 1)
	ubiquitin-dependent protein catabolism (100%)					
	mitosis (67%)					
21	4.36	6.2	9 (5, 4)	18	13	(0, 5)
	cytokinesis (100%, 50%)					
30	3.76	39.6	6 (3, 5)	5	1	(0, 6)
	DNA replication initiation (100%, 80%)					

Subnets Conserved in Yeast and Fruit Fly

Proteosome regulatory particle subnet



Calcium-dependent stress-activated signaling pathway



Role of Information in Scientific Discovery: Examples

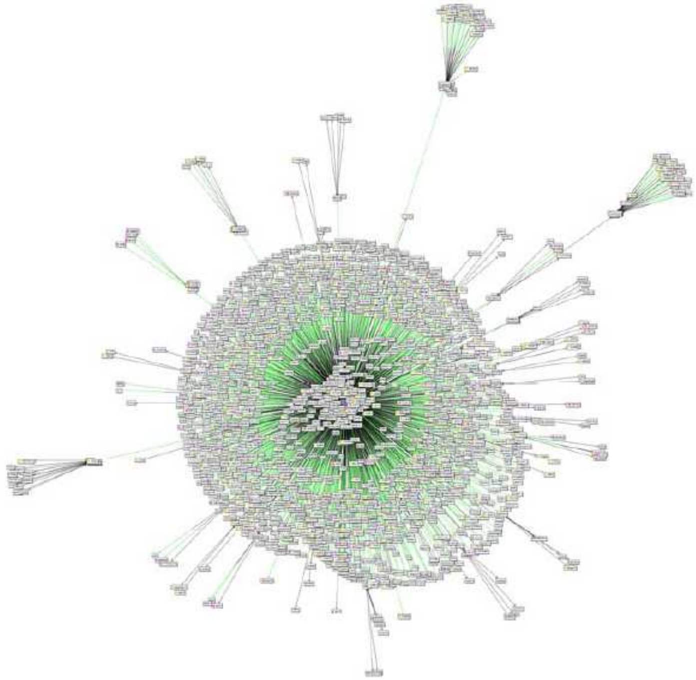
What are statistically significant functional pathways in gene regulatory networks?

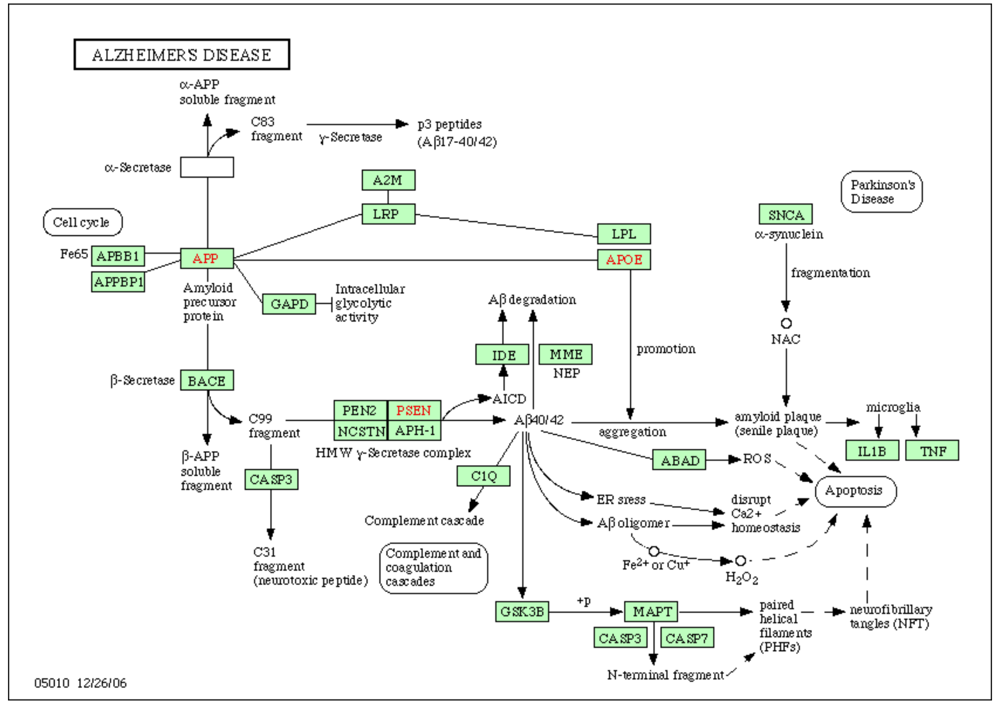
NARADA DEMO

Frequency	p-value	Pathway
276	5E-94	metabolic process ⊣ flagellum biogenesis → transport
136	3.1E-71	regulation of translation ⊣ DNA recombination → transport
38	4.9E-47	response to stimulus ⊣ transcription → cell motility
36	6.6E-35	flagellum biogenesis → ciliary or flagellar motility
56	1.4E-24	regulation of translation ⊣ transcription → carboxylic acid metabolism
178	8.3E-21	signal transduction ⊣ transcription → transport
14	8.6E-20	phosphate transport → transcription → phosphonate transport
16	2E-16	SOS response ⊣ regulation of transcription ⊣ DNA repair
501	1.2E-13	regulation of transcription, DNA-dependent → transport
12	3.6E-10	proteolysis ⊣ regulation of transcription ⊣ response to external stimulus
15	3.8E-7	nitrate assimilation ⊣ cytochrome complex assembly
10	1.4E-6	cell morphogenesis ⊣ protein secretion
178	3.8E-4	transcription → carbohydrate metabolic process

Information Sciences in Life Sciences, Ongoing Work

Characterizing phenotype and disease in networks.





Detecting "informative" parts of networks is an essential aspect of understanding disease and remediation (Alzheimers).