

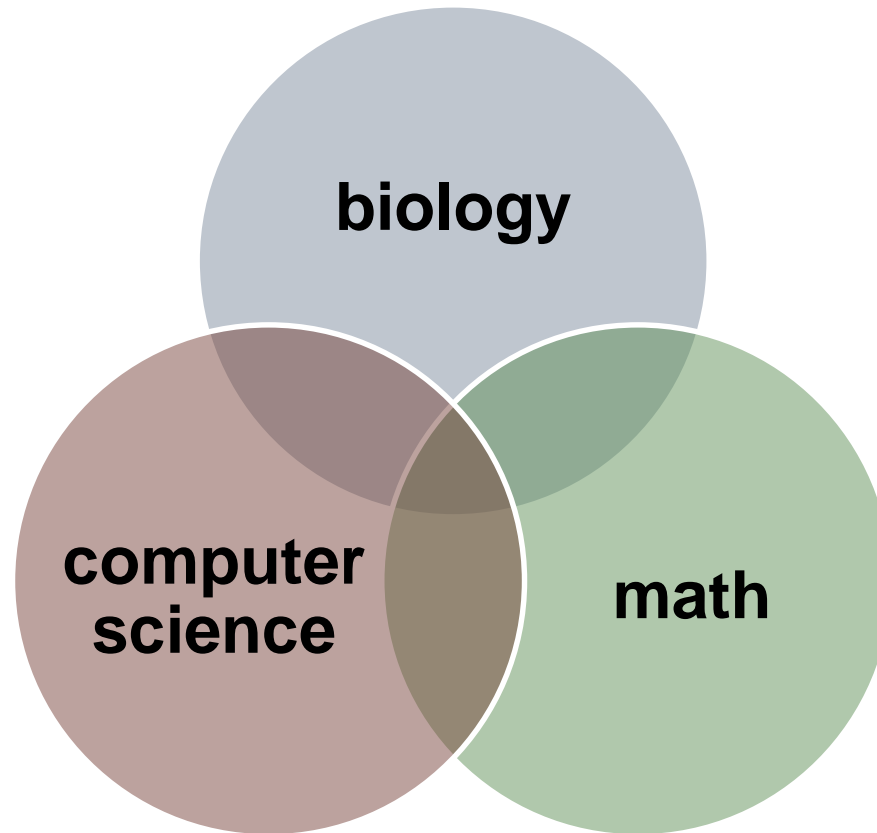
A Framework for Models and Modeling to Unify Mathematicians and Biologists and Improve Student Learning

Kam D. Dahlquist, Ph.D.
Department of Biology
Loyola Marymount University

**Breaking the Boundaries
in STEM Education Research**
April 7, 2017

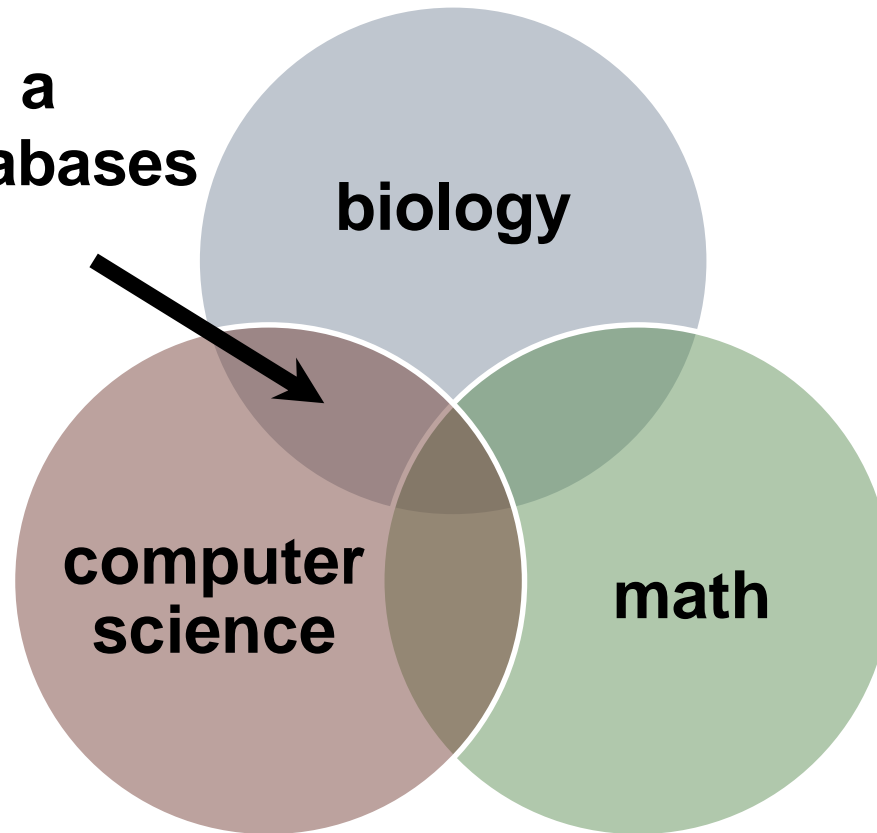
LMU|LA
Frank R. Seaver College
of Science and Engineering

Building Bridges between Biology, Computer Science, and Math in Research and Teaching



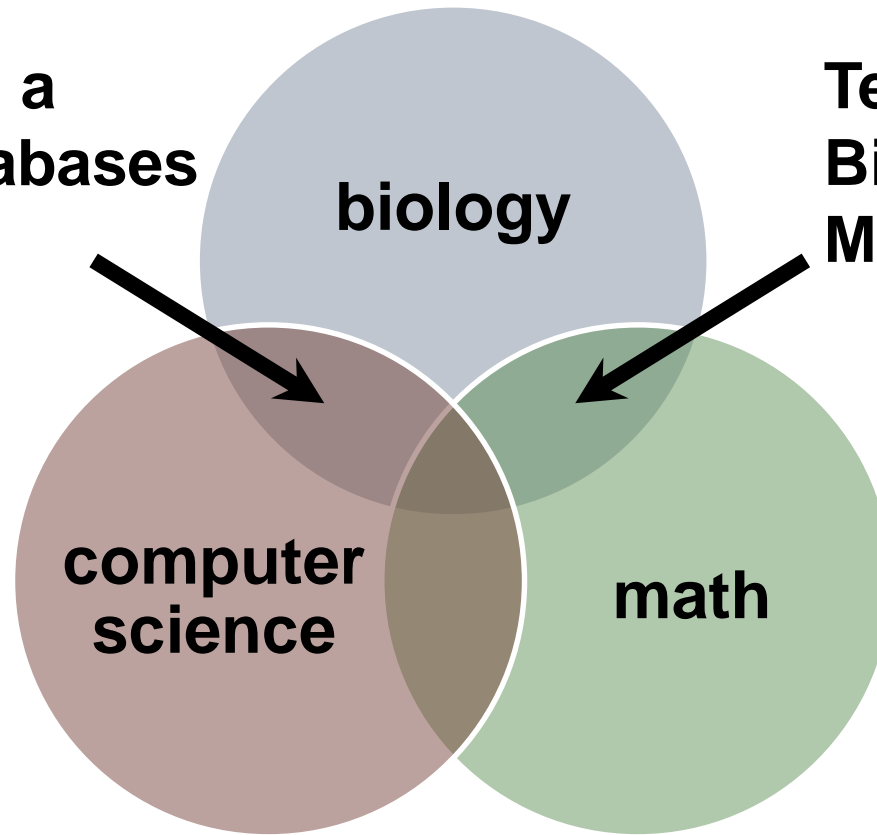
Building Bridges between Biology, Computer Science, and Math in Research and Teaching

**Team-teaching a
Biological Databases
course**



Building Bridges between Biology, Computer Science, and Math in Research and Teaching

**Team-teaching a
Biological Databases
course**

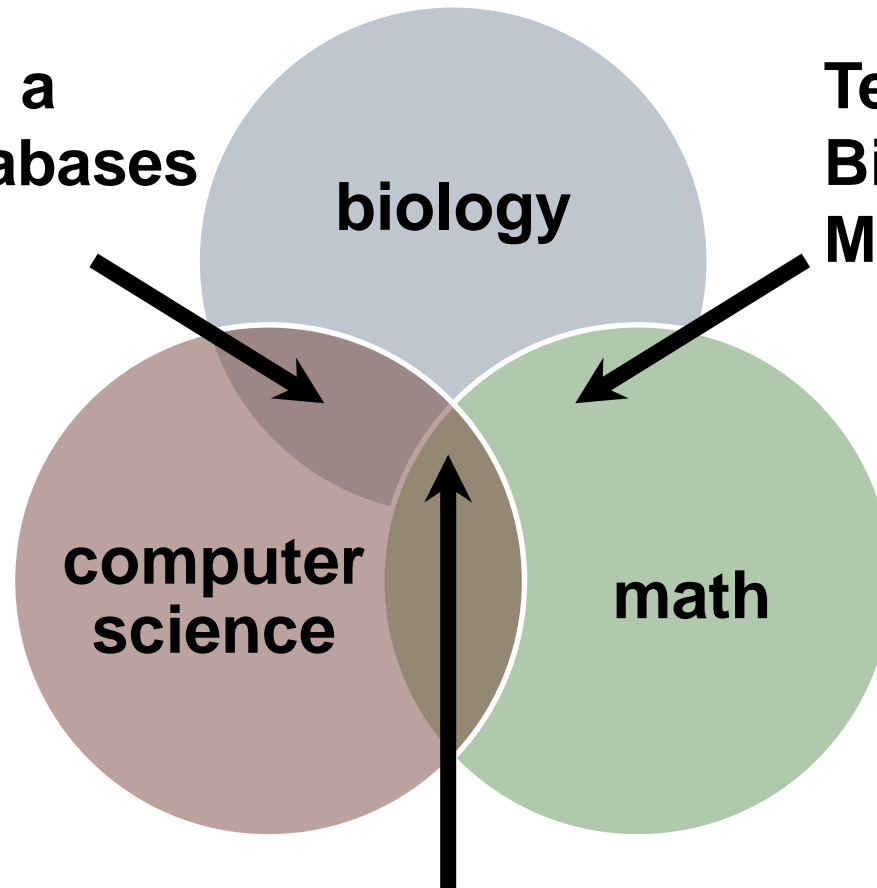


**Team-teaching a
Biomathematical
Modeling course**

Building Bridges between Biology, Computer Science, and Math in Research and Teaching

**Team-teaching a
Biological Databases
course**

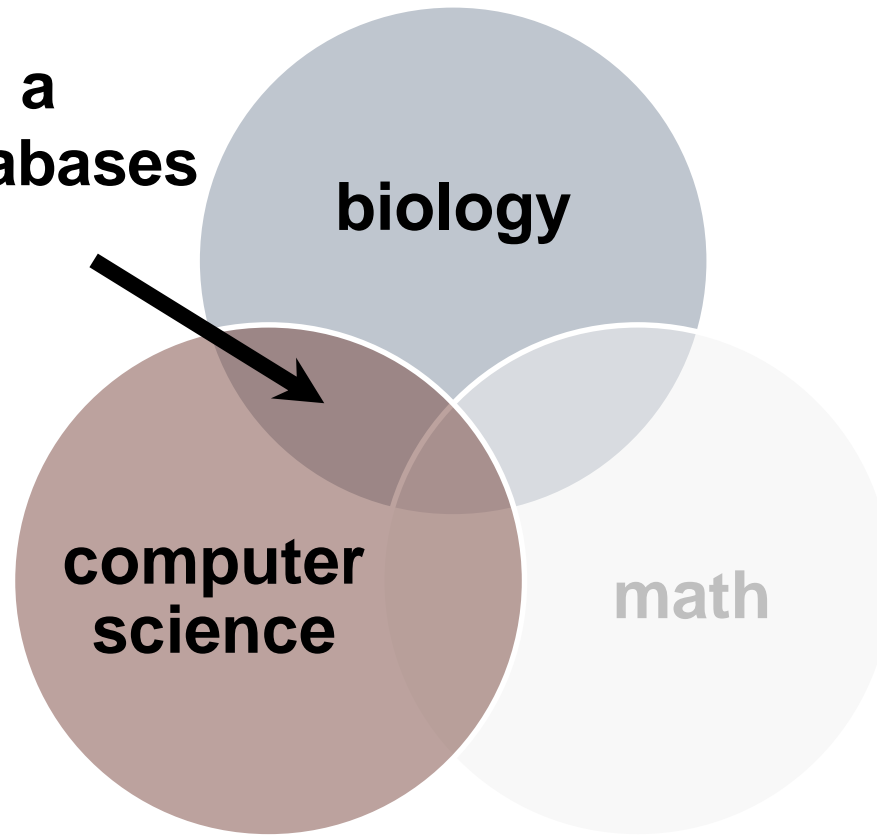
**Team-teaching a
Biomathematical
Modeling course**



**Bringing the three cultures together
for systems biology research with
undergraduates**

Building Bridges between Biology, Computer Science, and Math in Research and Teaching

**Team-teaching a
Biological Databases
course:
Open Source
Pedagogy**



Disconnect between Undergraduate Computer Science Training and Expectations/Skill Sets Required in Industry

Traditional computer science curriculum	“Real world” expectations (industry or research)	Best practices for software development
Students work alone	Members work together as teams	Project management
“Toy” programs and algorithms	Large modular projects	Up-front project design
Throwaway code after grading	Code maintained over an extended period of time	Program & process documentation; quality control; data standards

Official Open Source Definition

(<http://opensource.org>)

Free redistribution

**No discrimination against
fields of endeavor**

Source code

Distribution of license

Derived works

**License must not be
specific to a product**

**Integrity of the author's
source code**

**License must not
restrict other software**

**No discrimination against
persons or groups**

**License must be
technology-neutral**

Open Source Values Mirror STEM Curricular Reform

Open Source Values	Active Learning/ Bioinformatics Pedagogy	Open Source Practices & Tools
Source code is available, modifiable, and long-lived	Authentic problem to solve with realistic complexity	Central code repository; version control; provenance of code
Accountability to a developer and user community	Participatory and collaborative work; peer review	Task and bug trackers; continuous integration; test- driven workflows
Responsibilities accompany rights	Responsibility and ownership of the learning process	Documentation: in- line, user manual, Wiki

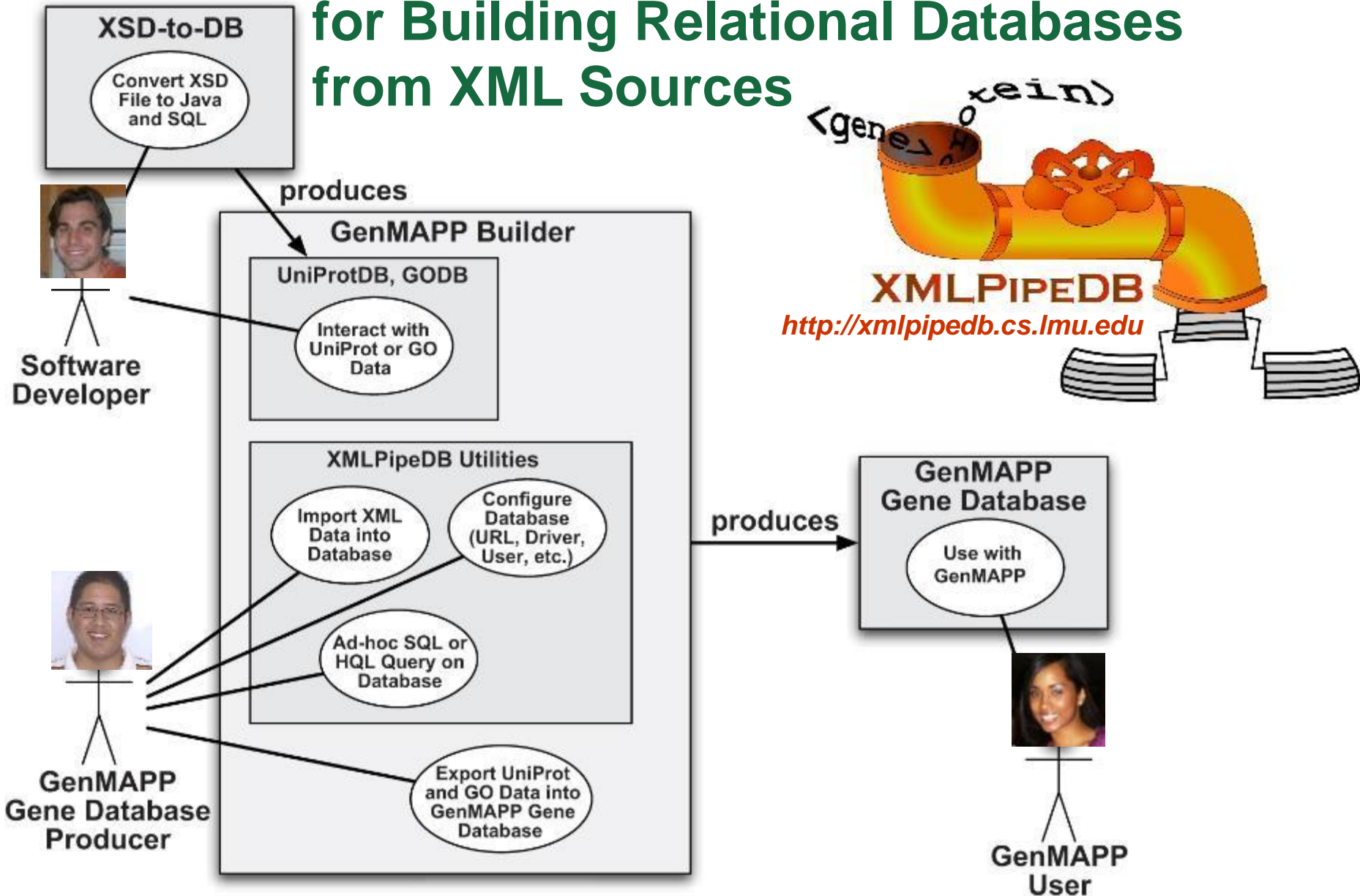
BIOL/CMSI 367: Biological Databases

- Team-taught by a biologist and a computer scientist
- Enrollment capped at 16 aims for equal numbers of biology and computer science majors
- Three-part framework:
 - Building blocks: genetic code and manipulating text
 - Going deeper: gene expression data and relational databases)
 - Integrating for research (gene database project)
- All coursework managed on a wiki
https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/Main_Page
- Research project to create a GenMAPP-compatible Gene Database for a new species with XMLPipeDB and use it to analyze a publicly available microarray dataset

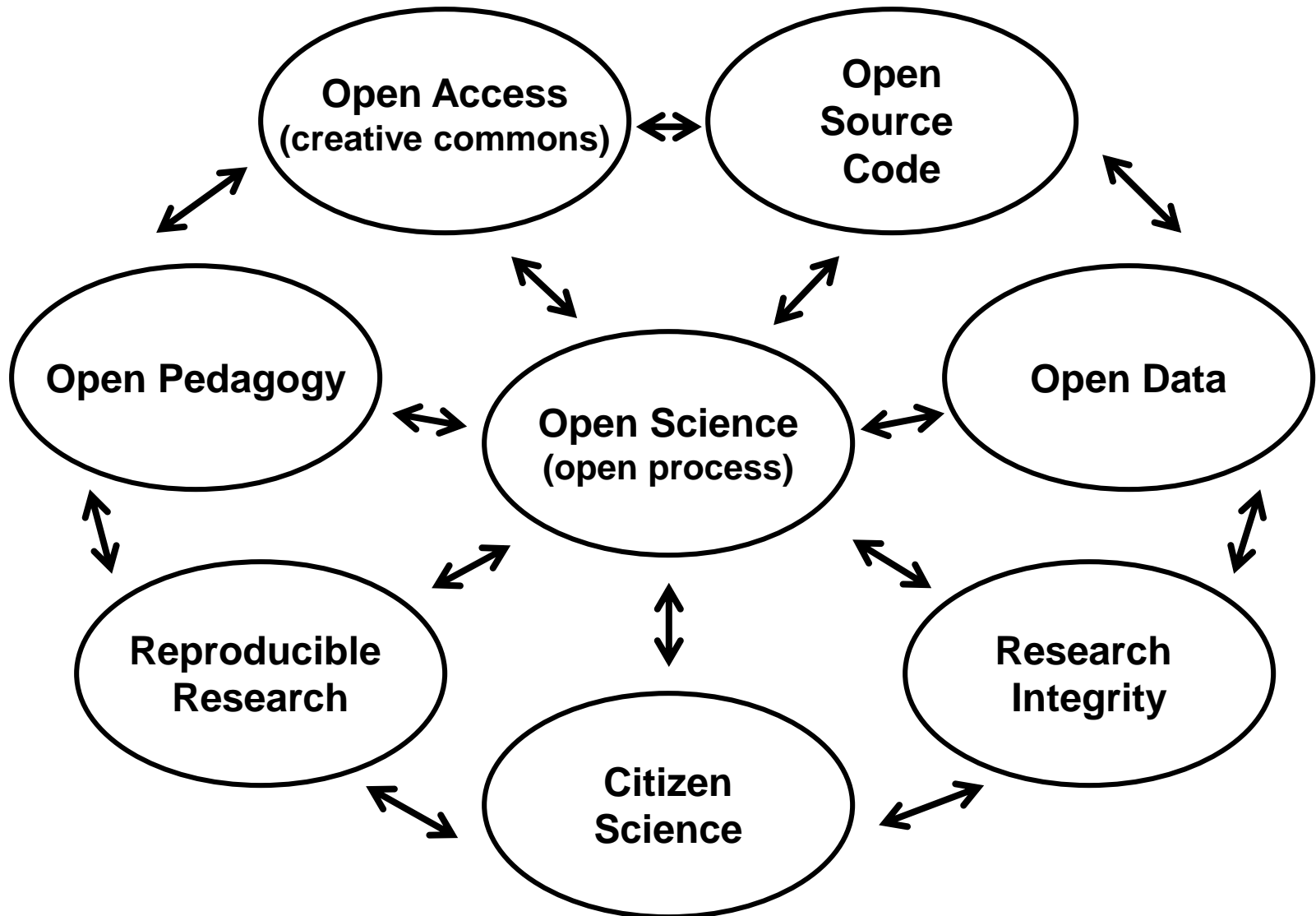
Students on Each Project Team Were Assigned Defined Roles

- **Project Manager**: sets the due dates for each milestone, manages the team wiki pages, makes sure that individuals are fulfilling their roles and performing the tasks on time, final responsibility for submitting project deliverables
- **Coder**: set-up of the development and testing environment, checking out code, making modifications, and checking code back in
- **Quality Control**: checks exported Gene Database for data integrity, verifying that the correct number of IDs from the XML source data are carried through a complete import/export cycle using pattern matching script and SQL queries
- **GenMAPP User**: locates a publicly available DNA microarray dataset, processes data (normalization, statistics), and analyzes data using GenMAPP and MAPPFinder

XMLPipeDB: an Open Source Tool Chain for Building Relational Databases from XML Sources

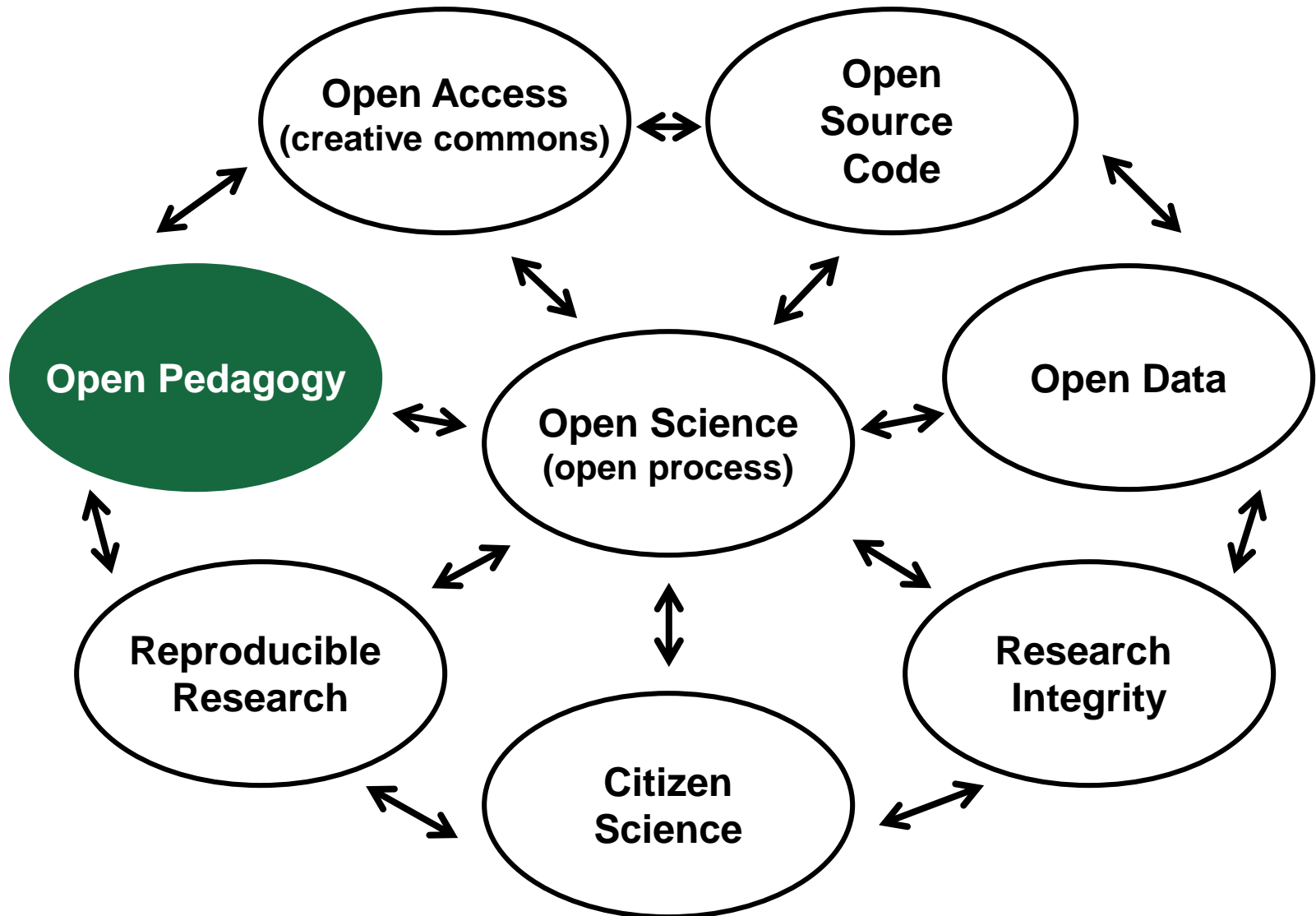


Open Science Ecosystem



With thanks to John Jungck

Open Science Ecosystem



Students Benefit from Open Source and Open Data



BioQUEST Curriculum Consortium
30+ years and going strong!
<http://www.bioquest.org>

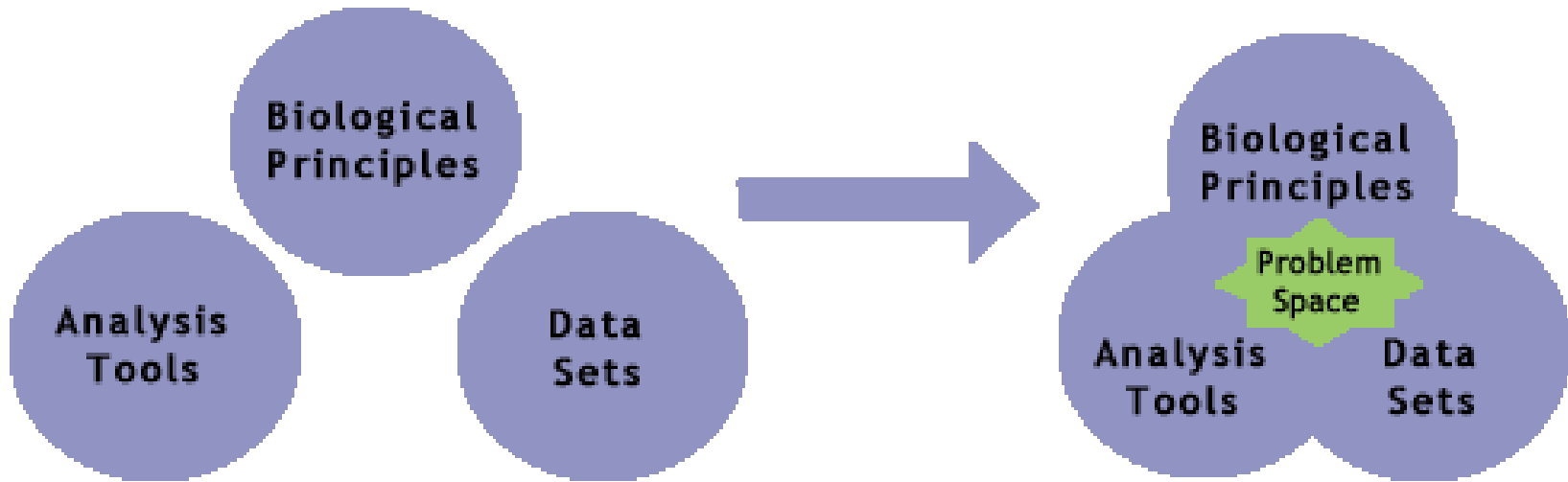


The Genome Consortium for Active Teaching
NextGen Sequencing Group
<http://gcat-seek.weebly.com/>



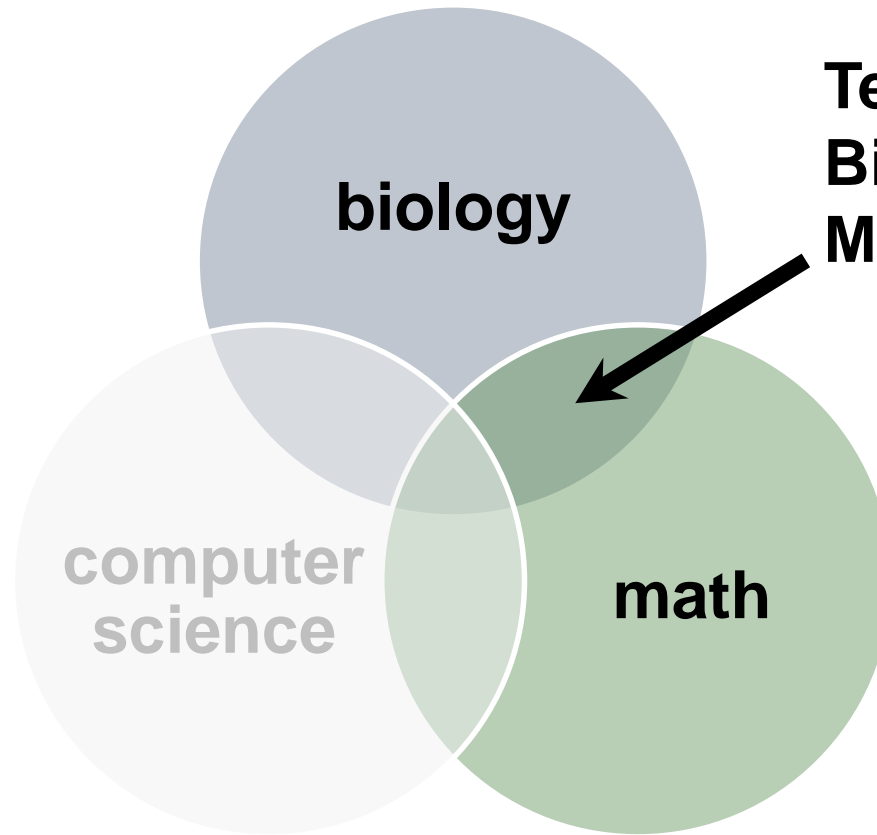
HHMI Science Education Alliance
Phage Hunters Advancing Genomics
and Evolutionary Science Program
<http://seaphages.org/>

Problem Spaces Are Flexible, Open-Ended, and Dynamic



- A brief introduction to a biological problem with references and links to background materials
- A collection of “open research questions” that might be addressed
- Examples of curricular materials used by other faculty
- Data, analysis tools, and other research resources
- Example of student work that explore the problem space

Building Bridges between Biology, Computer Science, and Math in Research and Teaching



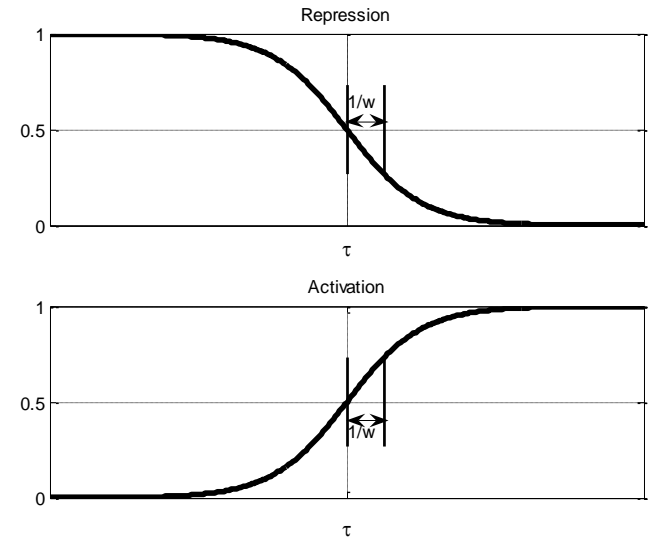
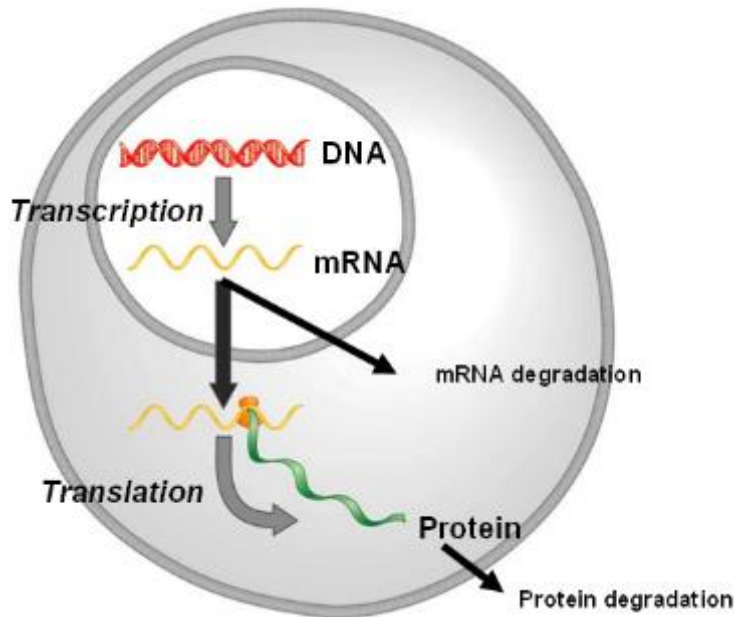
**Team-teaching a
Biomathematical
Modeling course:
Defining a New
Modeling
Framework**

BIOL/MATH 388: Biomathematical Modeling

- **Team-taught by a biologist and a mathematician**
- **Enrollment capped at 16 aims for equal numbers of biology and mathematics majors**
- **Three modeling projects in MATLAB:**
 - **Growth models (batch/logistic and chemostat)**
 - **Nitrogen metabolism in yeast**
 - **Systems Biology: cold shock gene regulatory network in yeast**
- **All coursework managed on a wiki**
<http://www.openwetware.org/wiki/BIOL398-05/S17>
- **Final research project based on NSF-funded collaborative work that produced the GRNmap modeling software**

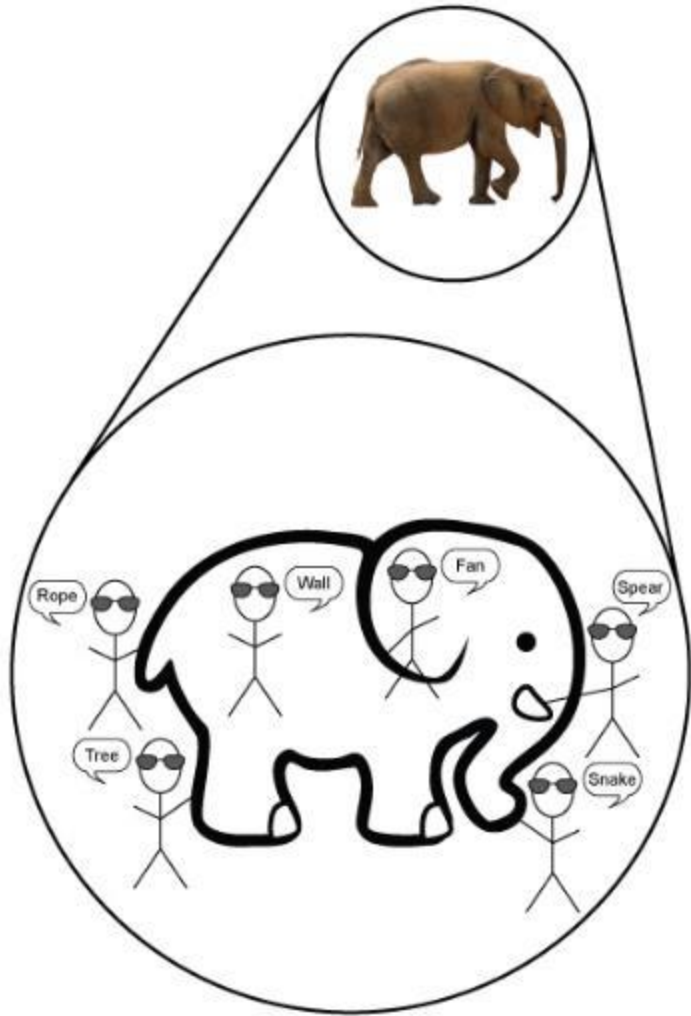
GRNmap: Gene Regulatory Network Modeling and Parameter Estimation

$$\frac{dx_i(t)}{dt} = \frac{P_i}{1 + \exp\left(-\left(\sum_j (w_{ij}x_j(t)) - b_i\right)\right)} - d_i x_i(t)$$



- The change in expression of one gene is *production – degradation*.
- Weight parameter, w , gives the direction (activation or repression) and magnitude of regulatory relationship.

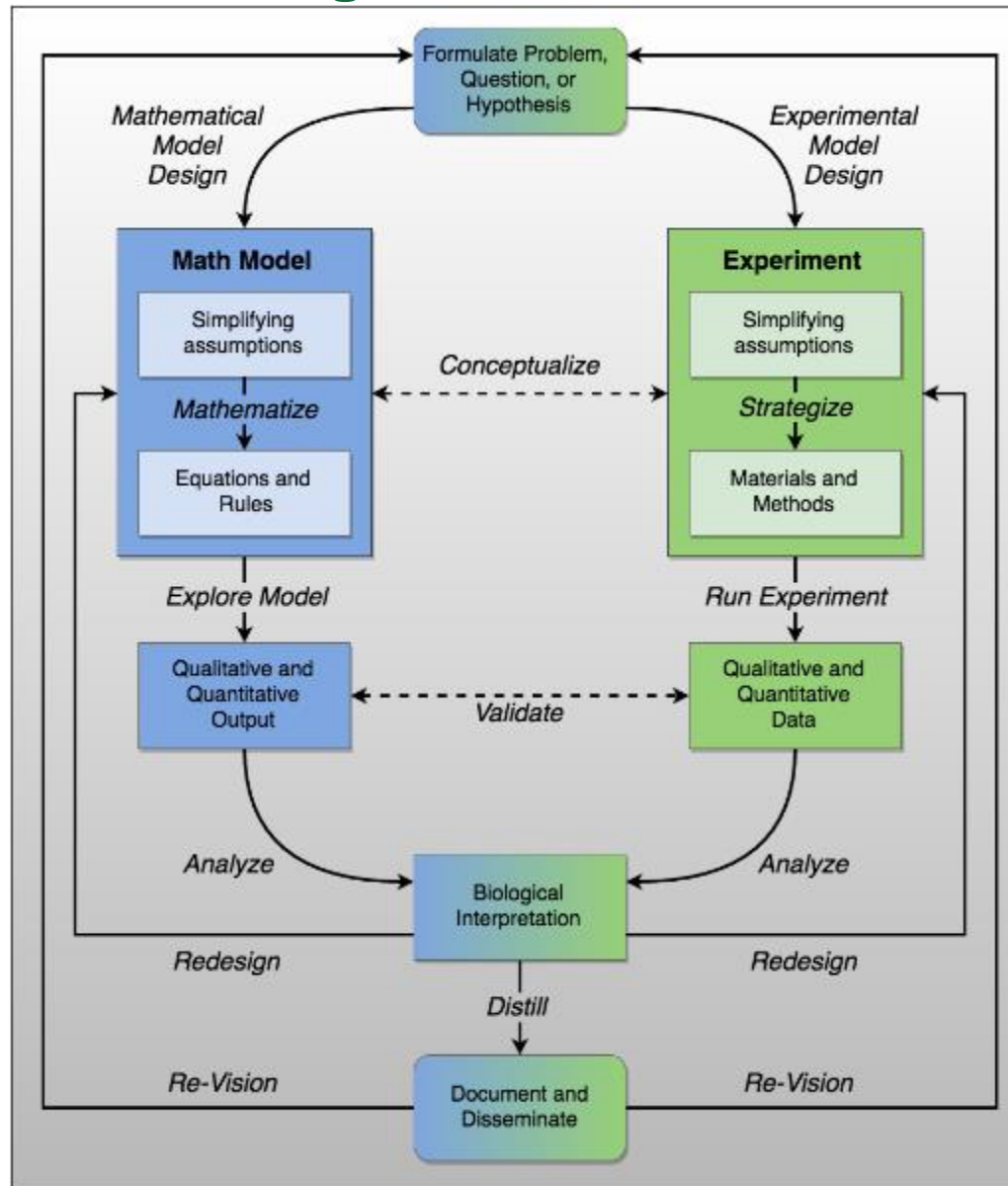
A *Model* is a Caricature: a Simplified Representation of Relationships or Processes in the Real World



Models can serve several purposes:

- Models are used to communicate ideas between scientists.
- Simple, unrealistic models help scientists explore complex systems.
- Models can lead to the development of conceptual frameworks and causal explanations (i.e., understanding)
- Models can make accurate predictions.

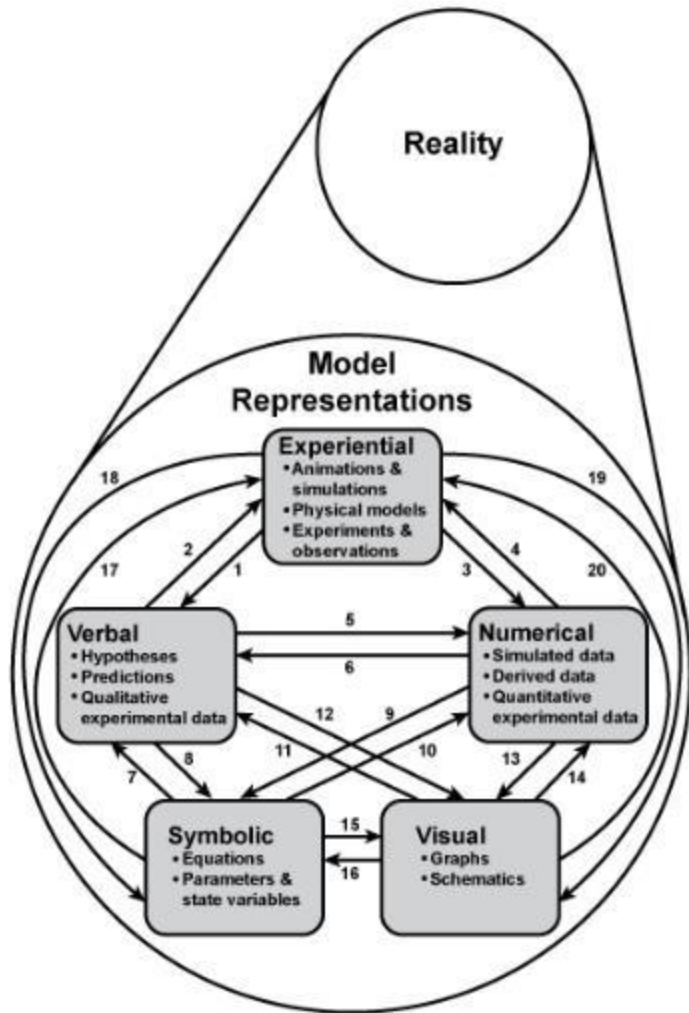
The Process of Modeling Mirrors the Process of Science



But Really, the *Process of Modeling* (and Science!) is Messier than That

We redefine *modeling* as individual activities such as:

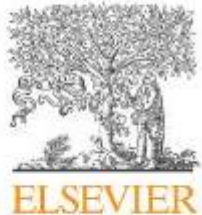
1. Moving from observations of reality to a model representation, either as an initial step in developing a model or as part of a model revision;
2. Moving from one model representation to another representation of the same model;
3. Comparing models to each other (model selection) or to reality (model validation).



We aim to be both more expansive and more inclusive of what counts as modeling.

Model Documentation Reminds Researchers to Make Our Process Explicit

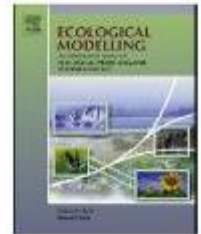
Ecological Modelling 280 (2014) 129–139



Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel



Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE



Volker Grimm^{a,b,c,*}, Jacqueline Augusiak^d, Andreas Focks^d, Béatrice M. Frank^e,
Faten Gabsi^f, Alice S.A. Johnston^g, Chun Liu^{g,h}, Benjamin T. Martin^{a,i}, Mattia Meli^j,
Viktoriia Radchuk^{c,e}, Pernille Thorbek^h, Steven F. Railsback^k

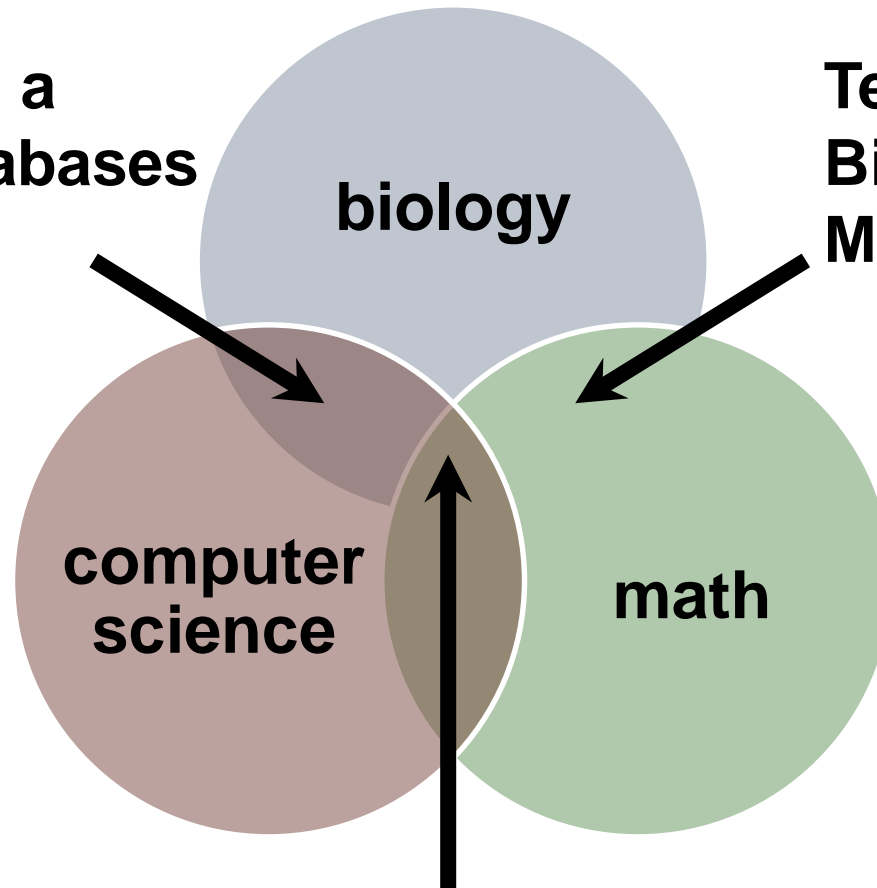
And this is exactly what is needed when teaching modeling to students:

- **Make the process explicit**
- **Help them reflect and connect**
- **Give it context again and again**

Building Bridges between Biology, Computer Science, and Math in Research and Teaching

**Team-teaching a
Biological Databases
course**

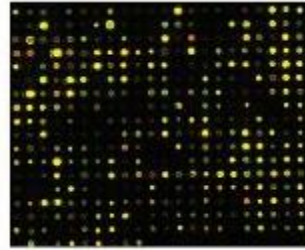
**Team-teaching a
Biomathematical
Modeling course**



**Bringing the three cultures together
for systems biology research with
undergraduates**

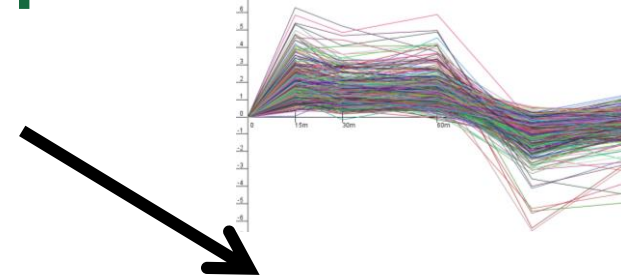
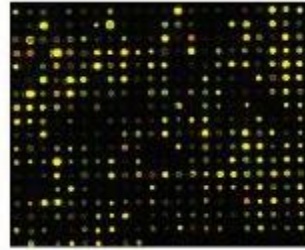
Systems Biology Approach to Understanding the Regulation of the Cold Shock Response in Yeast

Cold shock microarray
data from wt and TF
deletion strains



Systems Biology Approach to Understanding the Regulation of the Cold Shock Response in Yeast

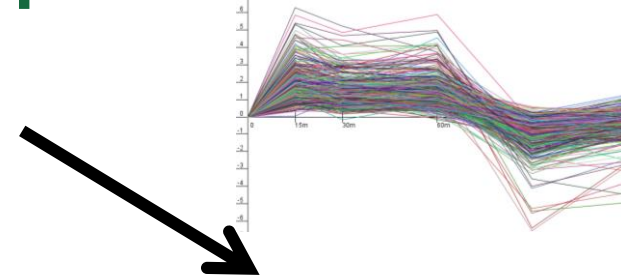
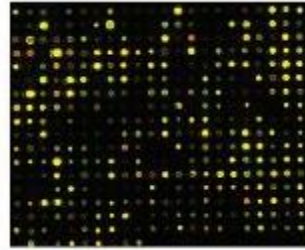
Cold shock microarray
data from wt and TF
deletion strains



Normalization,
statistical analysis,
clustering

Systems Biology Approach to Understanding the Regulation of the Cold Shock Response in Yeast

Cold shock microarray data from wt and TF deletion strains

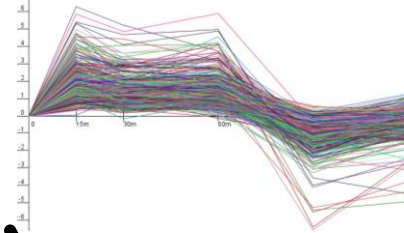
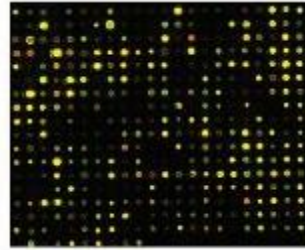


Normalization,
statistical analysis,
clustering

Derivation of gene
regulatory networks
from YEASTRACT

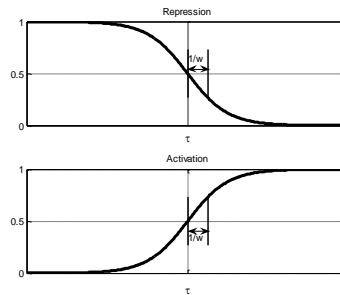
Systems Biology Approach to Understanding the Regulation of the Cold Shock Response in Yeast

Cold shock microarray data from wt and TF deletion strains



Normalization,
statistical analysis,
clustering

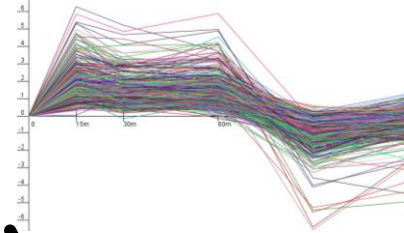
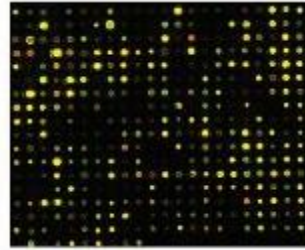
Derivation of gene
regulatory networks
from YEASTRACT



Dynamical systems
modeling using
GRNmap

Systems Biology Approach to Understanding the Regulation of the Cold Shock Response in Yeast

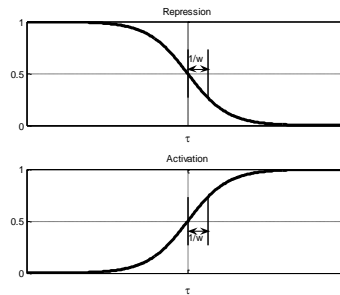
Cold shock microarray data from wt and TF deletion strains



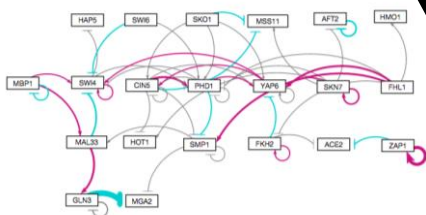
Normalization,
statistical analysis,
clustering

Derivation of gene
regulatory networks
from YEASTRACT

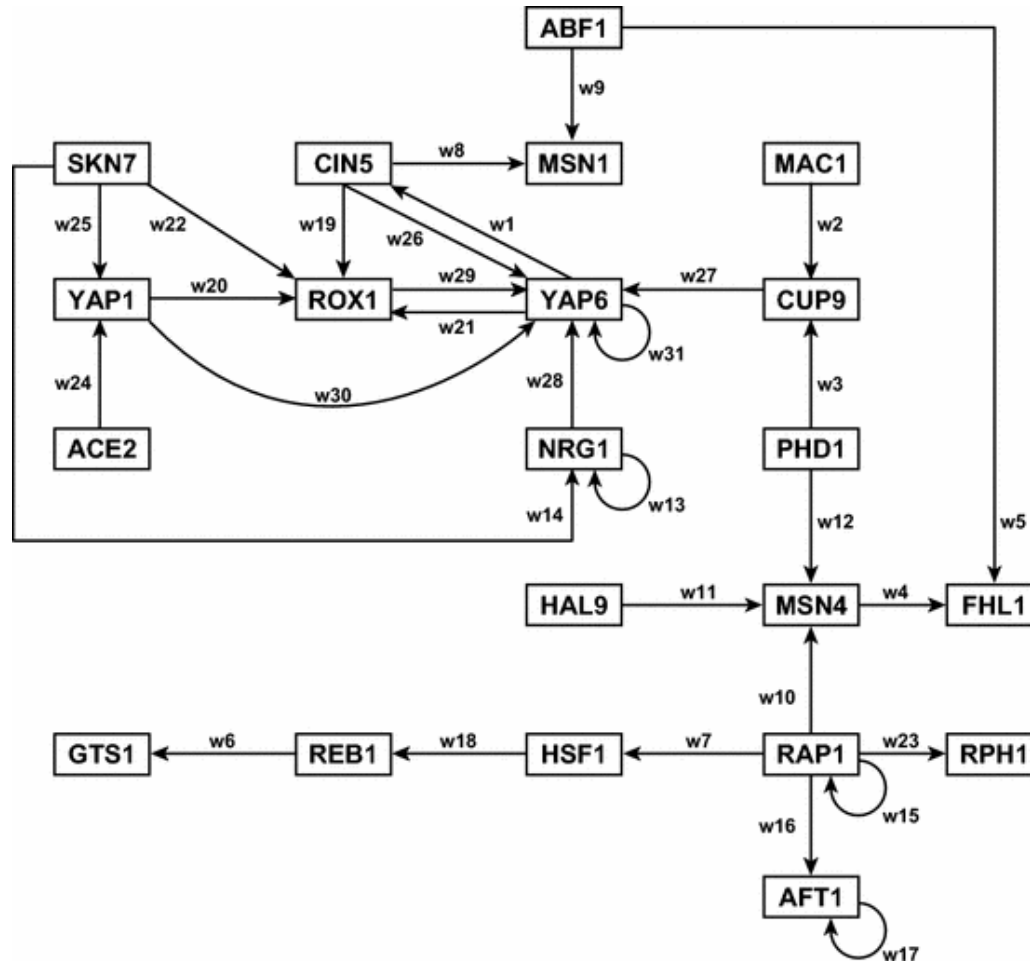
Visualization of
modeling results
using GRNsight



Dynamical systems
modeling using
GRNmap



This Gene Regulatory Network (GRN) Diagram Took Several Hours to Create by Hand in Adobe Illustrator



GRNsight Rapidly Generates GRN Graphs Using Our Customizations to the Open Source D3 Library



GRNsight

Web app and service for visualizing models of gene regulatory networks.

A Joint Project of the
LMU Bioinformatics and
Biomathematics Groups

LMU|LA
Loyola Marymount
University

- Home
- About
- News
- Developers
- Documentation
- Wiki
- Publications
- People
- Contact
- Links
- Beta
- Privacy

Last modified: 8/11/16
1:55 PM

GRNsight v1.18.1

Welcome to GRNsight. Please select a file to upload.

File Edit Format Help Demo 21-genes_31-edges_Schade-data_input.xlsx

21 nodes
31 edges

Link Distance (1 - 1000): 500

Charge (-2000 - 0): -1000

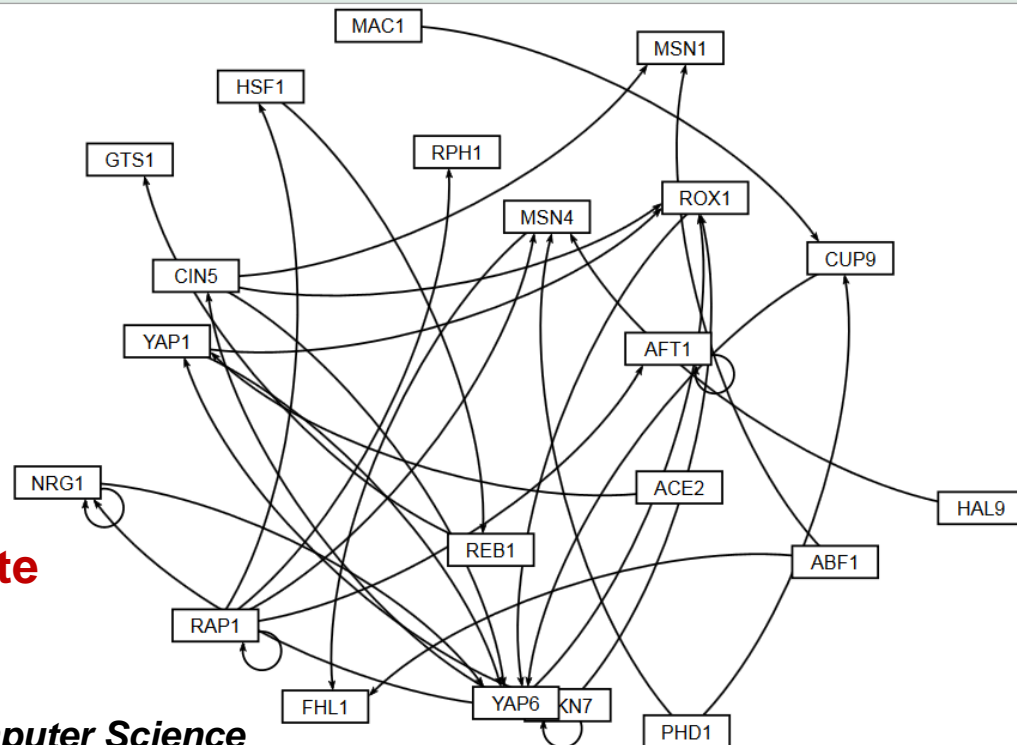
Charge Distance (0 - 2000): 1000

Gravity (0 - 1): 0.10

☐ Lock Force Graph Parameters

Reset Force Graph Parameters

Undo Reset



10 milliseconds to generate

Dahlquist et al. (2016) *PeerJ Computer Science*

GRNsight Rapidly Generates GRN Graphs Using Our Customizations to the Open Source D3 Library



GRNsight

Web app and service for visualizing models of gene regulatory networks.

A Joint Project of the
LMU Bioinformatics and
Biomathematics Groups

LMU|LA
Loyola Marymount
University

- Home
- About
- News
- Developers
- Documentation
- Wiki
- Publications
- People
- Contact
- Links
- Beta
- Privacy

Last modified: 8/11/16
1:55 PM

GRNsight v1.18.1

Welcome to GRNsight. Please select a file to upload.

File Edit Format Help Demo 21-genes_31-edges_Schade-data_input.xlsx

21 nodes
31 edges

Link Distance (1 - 1000): 500

Charge (-2000 - 0): -1000

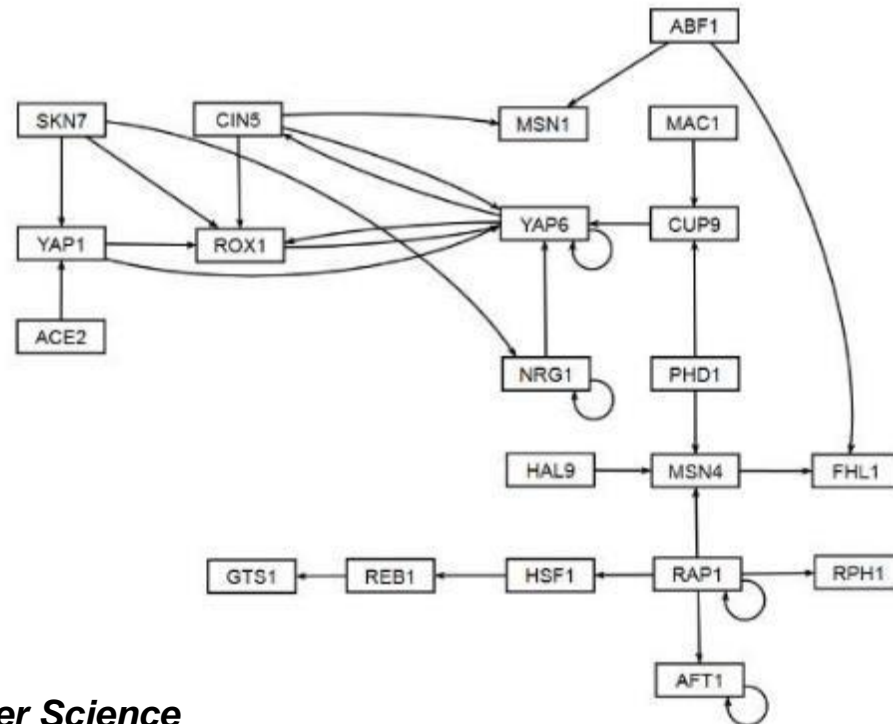
Charge Distance (0 - 2000): 1000

Gravity (0 - 1): 0.10

☐ Lock Force Graph Parameters

Reset Force Graph Parameters

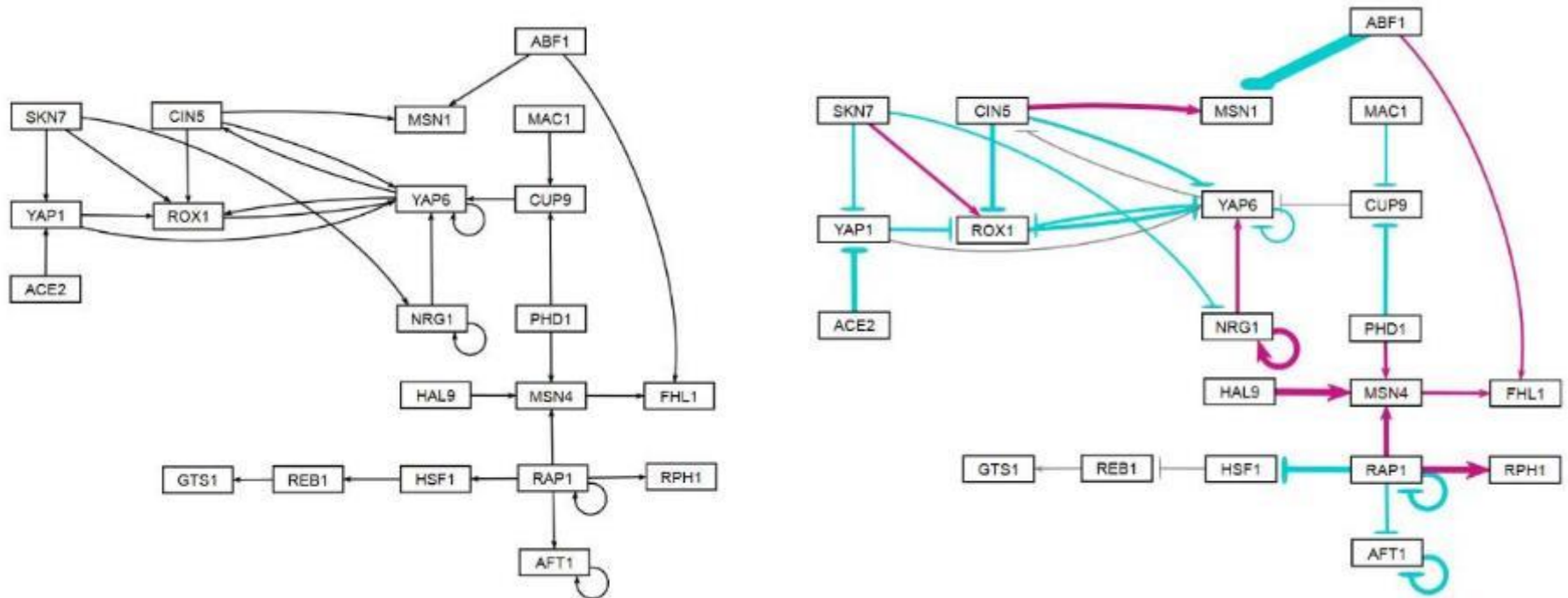
Undo Reset



**10 milliseconds to generate,
5 minutes to arrange**

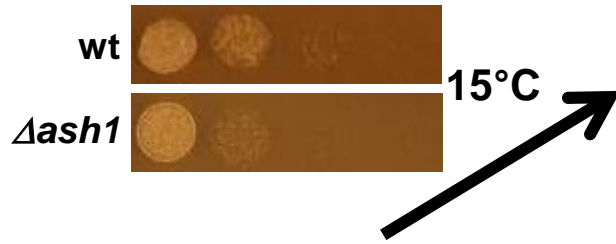
Dahlquist et al. (2016) *PeerJ Computer Science*

More Importantly, Weighted Graphs Reveal Patterns in the Data

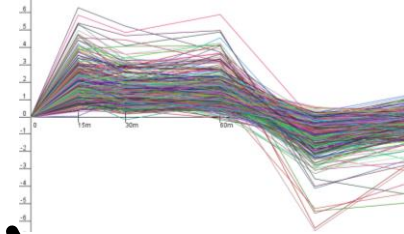
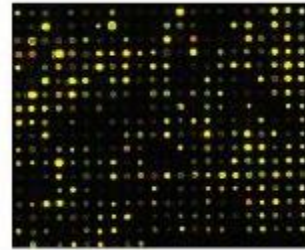


- There are two subgraphs in the network.
- Target genes are regulated by a balance of activation and repression.
- The same transcription factors both activate and repress expression.
- Regulatory chains weaken after two levels.
- The most complex motifs are actually simpler than they appear.

Systems Biology Approach to Understanding the Regulation of the Cold Shock Response in Yeast



Cold shock microarray data from wt and TF deletion strains

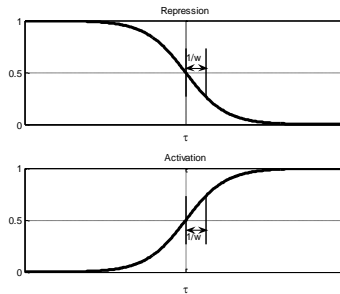


Interpretation,
new questions,
new experiments

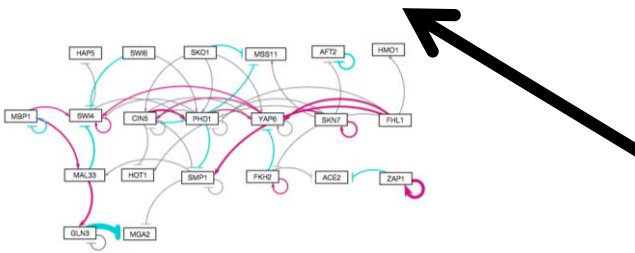
Normalization,
statistical analysis,
clustering

Visualization of
modeling results
using GRNsight

Derivation of gene
regulatory networks
from YEASTRACT



Dynamical systems
modeling using
GRNmap



A Biologist Bridging the Cultures of Computer Science and Mathematics

Open Source Pedagogy

- **Pairing bioinformatics tool users with tool developers works well for interdisciplinary work**
- **Emphasis on roles, process, and best practices**
- **Products of student research returned to the scientific community**
- **Instructors were already experienced with the bioinformatics domain**

Framework for Modeling

- **Bringing open source practices to a mathematics course was more challenging**
- **Needed to make the activities in the modeling process more explicit to students (and the biologist!)**
- **Instructors made substantial investment to learn each other's domains**

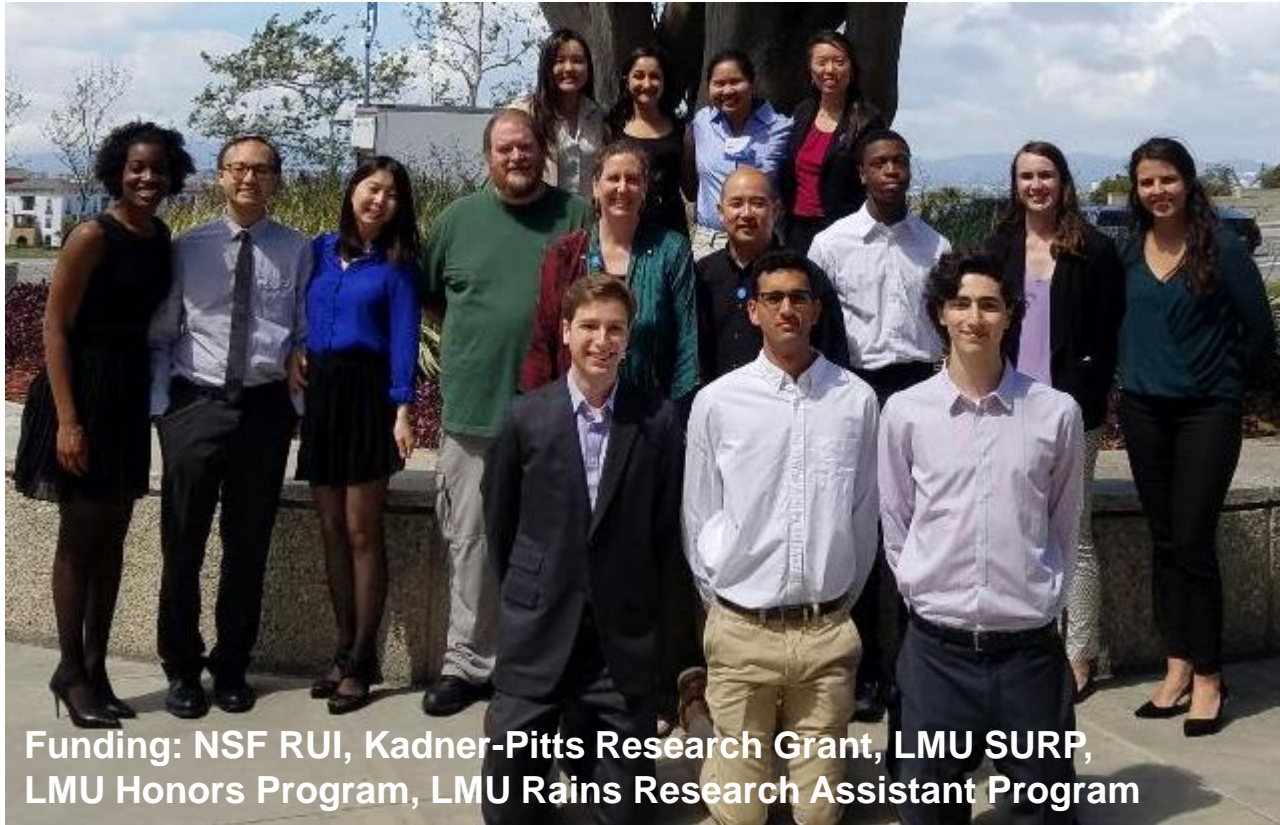
How to Start Interdisciplinary Conversations

- **Have a shared goal**
- **Start small**
- **Be willing to be a student as well as a teacher**
- **Pay attention to language**
 - **Disciplinary jargon disconnects**
 - **The same word means different things to different disciplines**
 - **Different words mean the same thing to different disciplines**
 - **Bring students into the conversation**
- **Be in it for the long term**

They Are Not Really “Digital Natives”

- **We have observed again and again that otherwise bright students are stymied by basic computer functions.**
 - Understanding where their files are (directories and paths)
 - Understanding what their files are (filetypes and extensions)
 - In general, what to do when you can’t just “drag and drop”
- **I’m interested in the intersection of computer literacy, data literacy, and information literacy skills.**
- **These also need to be taught in the context of authentic problem solving.**

GRNsight: <http://dondi.github.io/GRNsight/>
GRNmap: <http://kdahlquist.github.io/GRNmap/>
kdahlquist@lmu.edu



Funding: NSF RUI, Kadner-Pitts Research Grant, LMU SURP,
LMU Honors Program, LMU Rains Research Assistant Program

Pictured

Eileen Choe
Anindita Varshneya
Trixie Ann Roque
Monica Hong
Natalie Williams
Justin Kyle Torres
Jen Shin

Ben Fitzpatrick

Dondi Dionisio

Kam Dahlquist

Eddie Azinge
Maggie O'Neil
Kristen Horstmann
Brandon Klein
Mihir Samdarshi
Eddie Bachoura

Not pictured

Nicole Anguiano
Katherine Schecker
Nika Vafadari

https://qubeshub.org/groups/nimbios_wg_teachingquantbio
**Melissa Aikens, Hanna Callendar, Joe Dauer, Sam Donovan, Carrie Diaz Eaton,
Greg Goins, Kristen Jenkins, John Jungck, Drew Lamar, Glenn Ledder,
Bob Mayes, Richard Schugart**