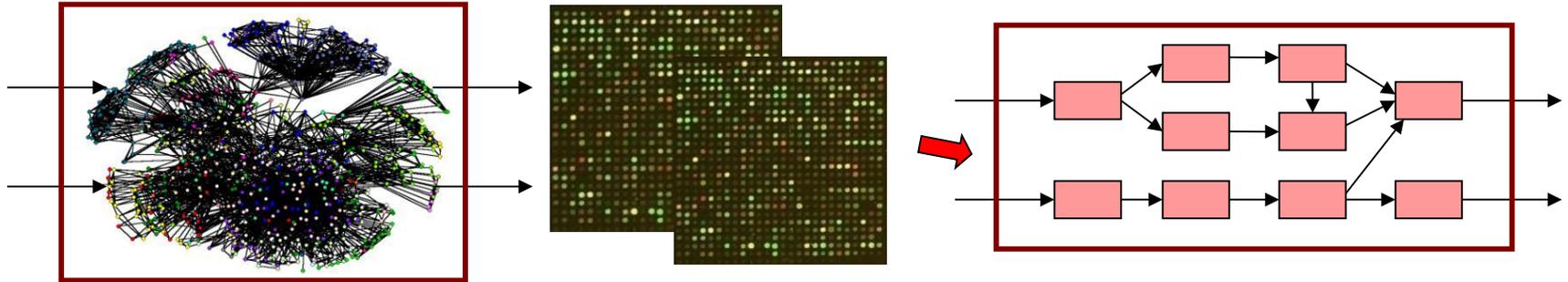


Modular Network Models



Impossible to
parameterize

Computationally
tractable

Biological systems are too complex for mechanistic models

Systems level modeling requires simpler abstractions

Define system in terms of independent *functional modules*

Modules interact with system only through their inputs and outputs

Input-output relationships and module connectivity determine system behavior

Focus on information flow, overall cause-effect relationships, rather than molecular mechanisms

1. Biological data sets are very high dimensional, often necessitating high throughput methods:
 - $\sim 10^4$ genes, $\sim 10^5$ methylation probe sites, etc.
 - When variables outnumber replicates there are difficulties in statistical inference
2. Specifying the machine learning task:
 - Unsupervised analysis where the structure of the graph is analyzed:
 - Visualizing the network in 2-D or 3-D embeddings
 - Clustering the network into modules based on some definition of a group/community of nodes
 - Supervised analysis where the network connectivity is tuned to produce a desired response variable:
 - Neural network classifiers/regressors
 - Fitting parameters of regulatory networks to best reproduce observed data from reverse phase protein assays
3. Picking the modeling approach:
 - Discriminative approaches often cast the problem in the language of convex optimization, and simply find solutions which minimize whatever cost function the biologist specifies. Robust, scale well, NOT interpretable.
 - Generative approaches attempt to learn a probability distribution which describes the observed data. Prone to over-fitting, learning many parameters can hinder scalability, but highly interpretable since one learns the best distribution which would have generated the observations.

rationale for this slide

Computational methods are becoming increasingly integrated into the study of biological systems. The mathematician / computer scientist needs to devote considerable time to understanding the experimental setup and data preparation before he/she can assist the biologist. I rarely hear the converse stated among my peers, that the experimentalists must be aware of the downstream modeling approaches BEFORE committing to a specific experimental design. We are now beginning to appreciate the multi-scale network complexity of biological systems, and it is imperative that the next generation of biologists be able to conceptualize problems in a computational framework.