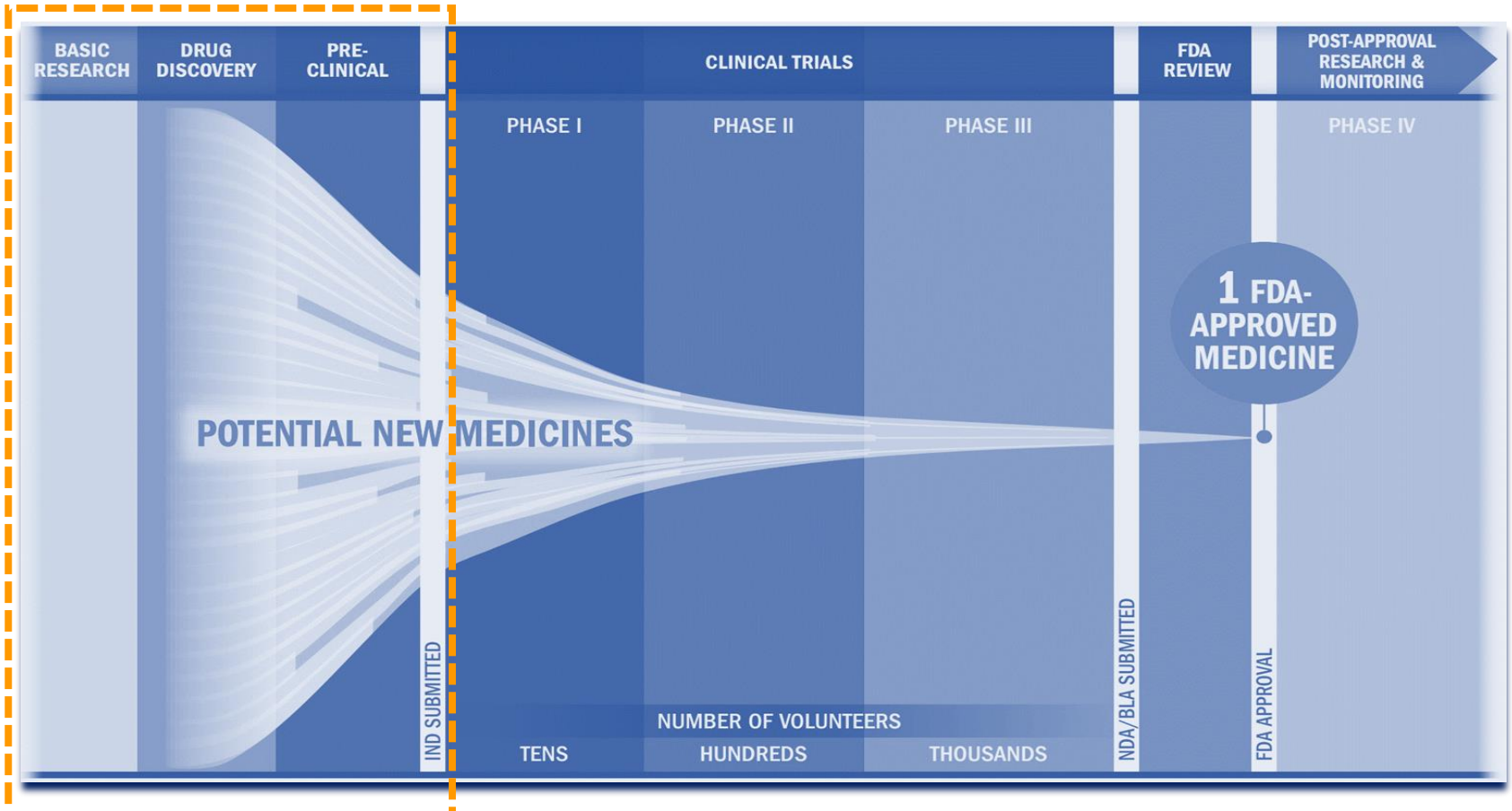
Deep Reinforcement Learning for Pre-Clinical Drug Development

Fellow: Nathan Russell

Advisors: Jian Peng (CS), Marty Burke (Chemistry)

Big Picture



1.15 Billion

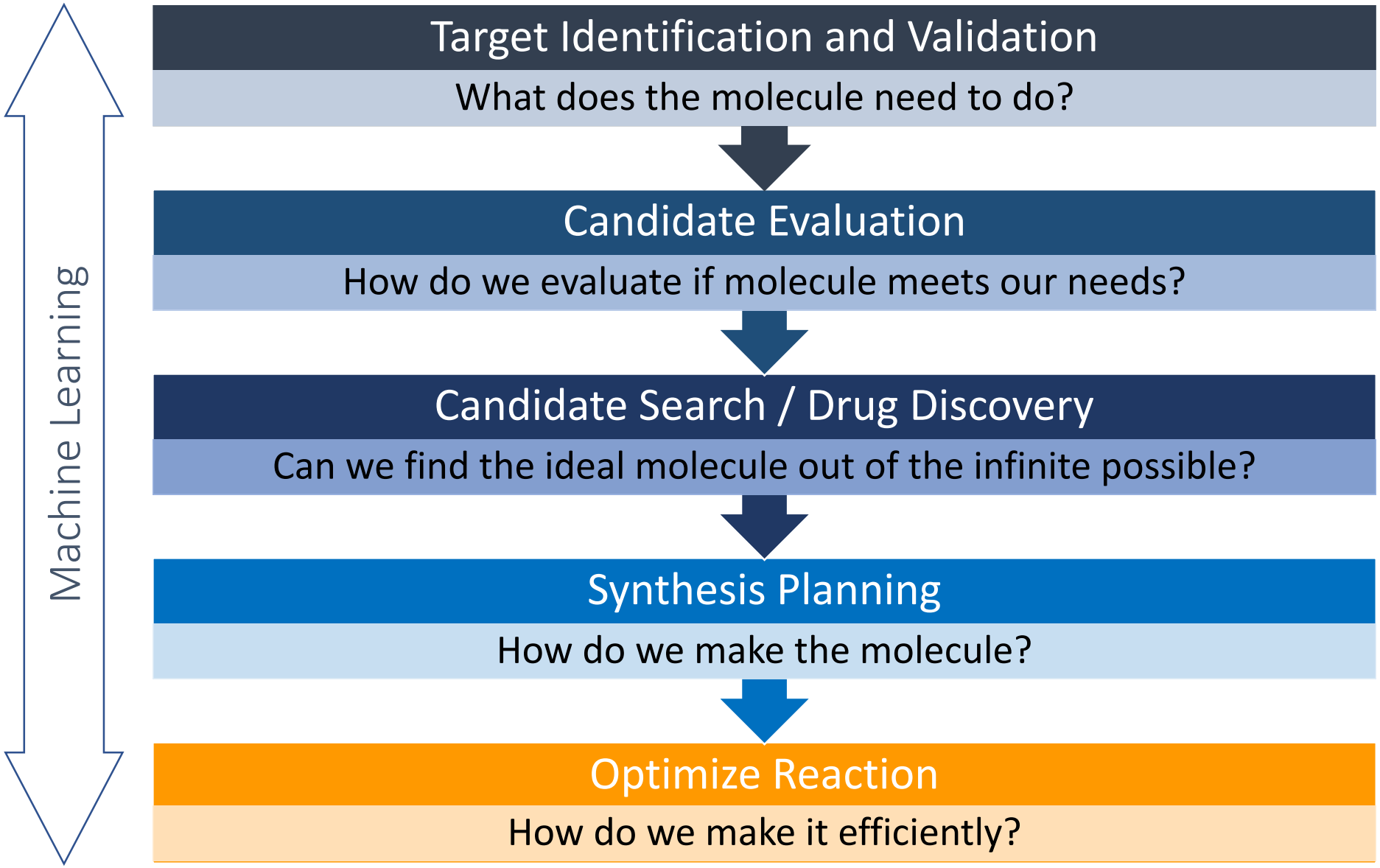
Direct Impact

1.53 Billion

Indirect Impact

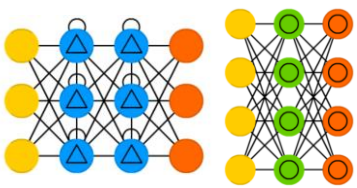
2017 $
Capitalized Costs
According to
Tufts CSDD 2014

Pre-Clinical Tasks

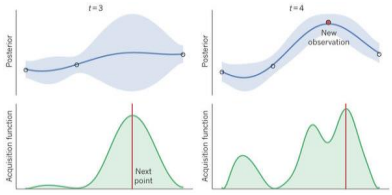


Resources & Contributions

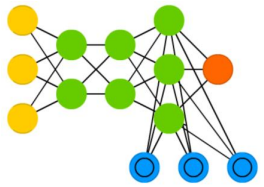
Contributions



Generative Multitask Molecular Network

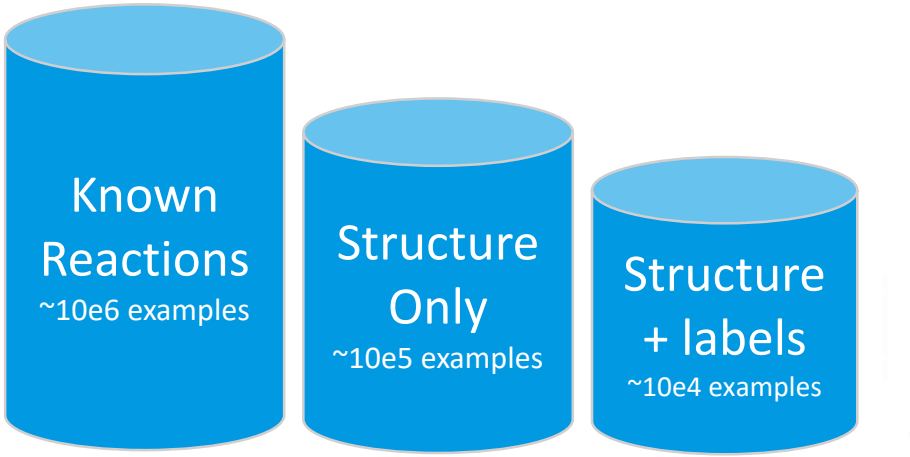


Model Assisted Bayesian Optimization



Graph Translation Policy Network

Resources



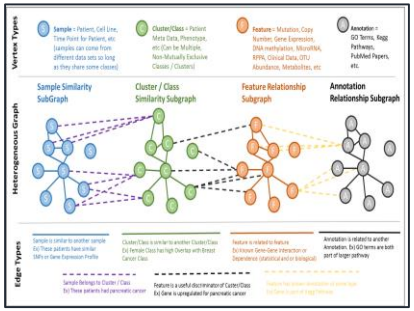
Automated Lab Evaluation

Target Identification and Validation (What I did 16-17)

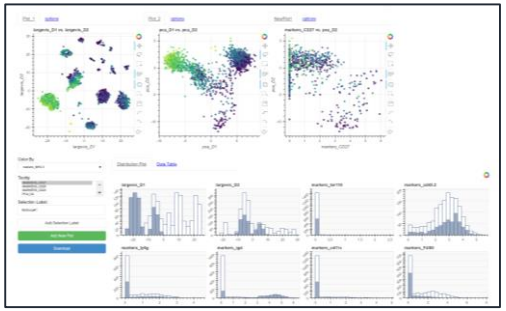
Common Tasks

- Gene / Protein Expression Studies
- Biomarker Discovery
- Biochemical and Cellular Pathway Analysis
- Cell Analysis

Tools made to Support those Tasks

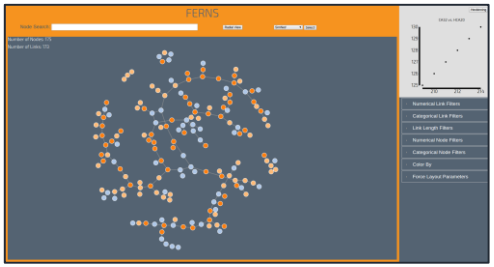


Heterogeneous Network Embedding

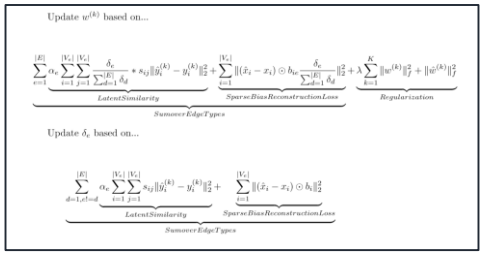


Scalable Manifold Embedding and Viz

Interpretable Pattern Discovery



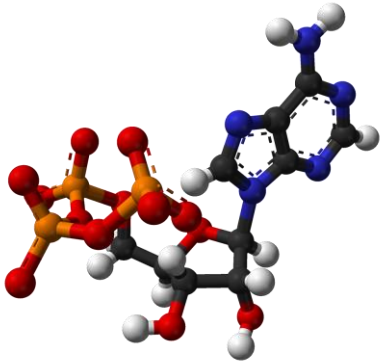
Real Time Filtering and Querying of Biological Networks



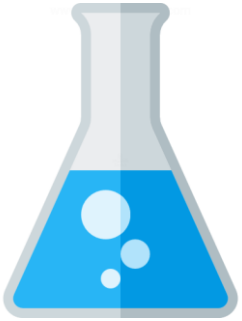
Deep Graph Embedding

Insilco Candidate Evaluation

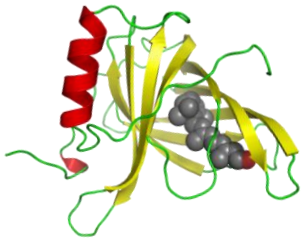
Molecule



Experimentation

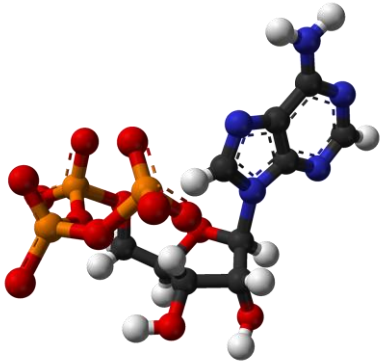


Ground Truth

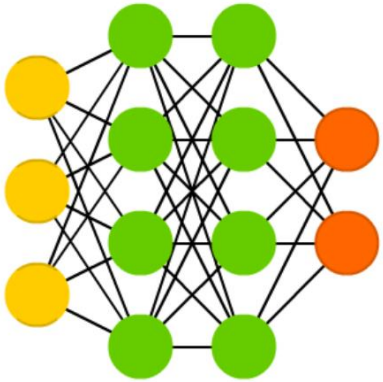


Binds to Protein
(Yes / No)

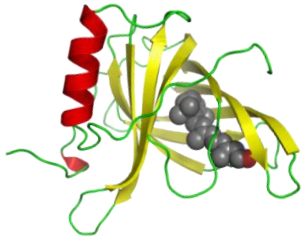
Molecule



Neural Network

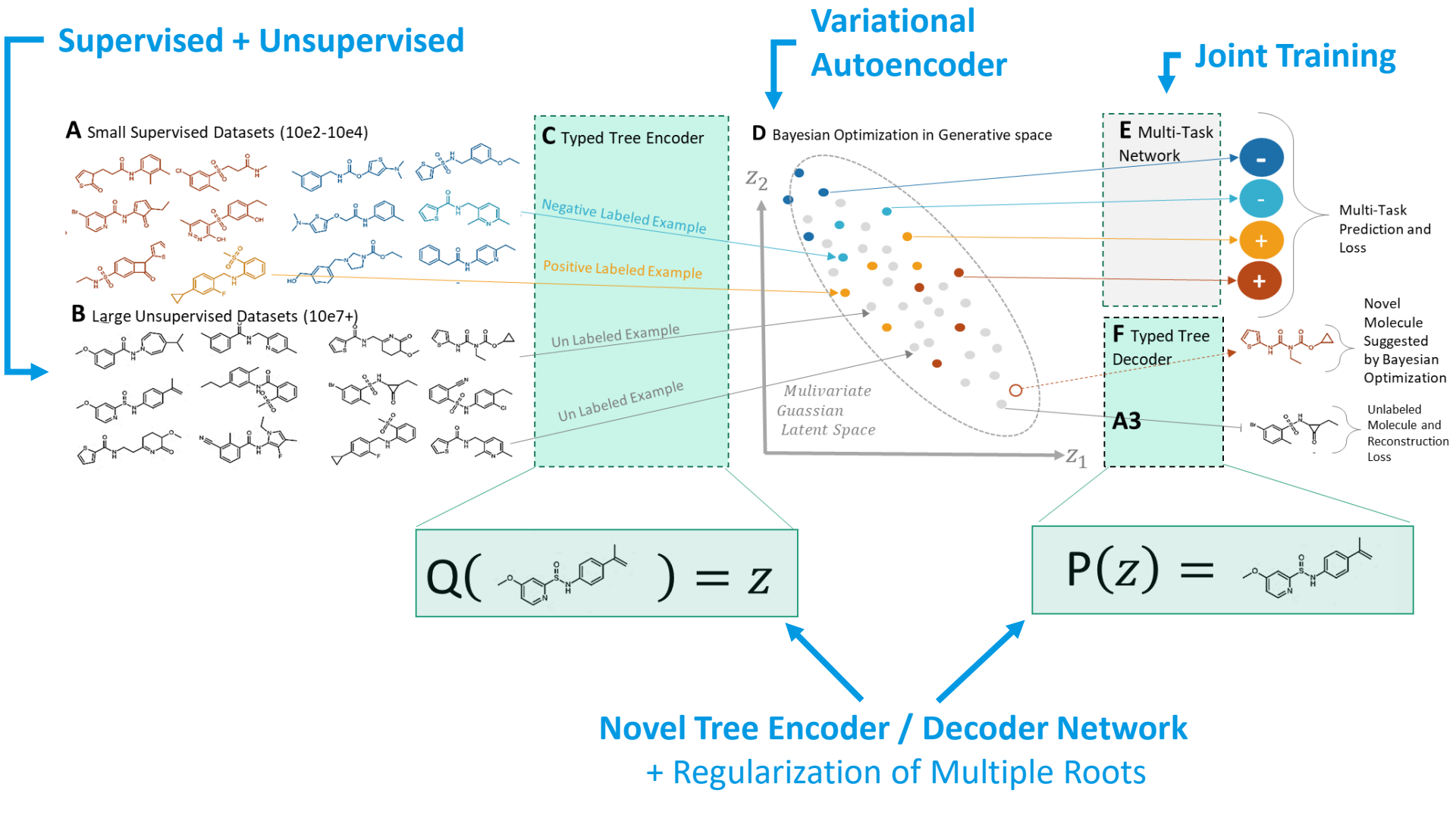


Prediction



Probability of
Binding (0,1)

Generative Multitask Molecular Network (GMMN)



Graph Translation Policy Network with External Memory

Machine Translation

English ▹ 中文

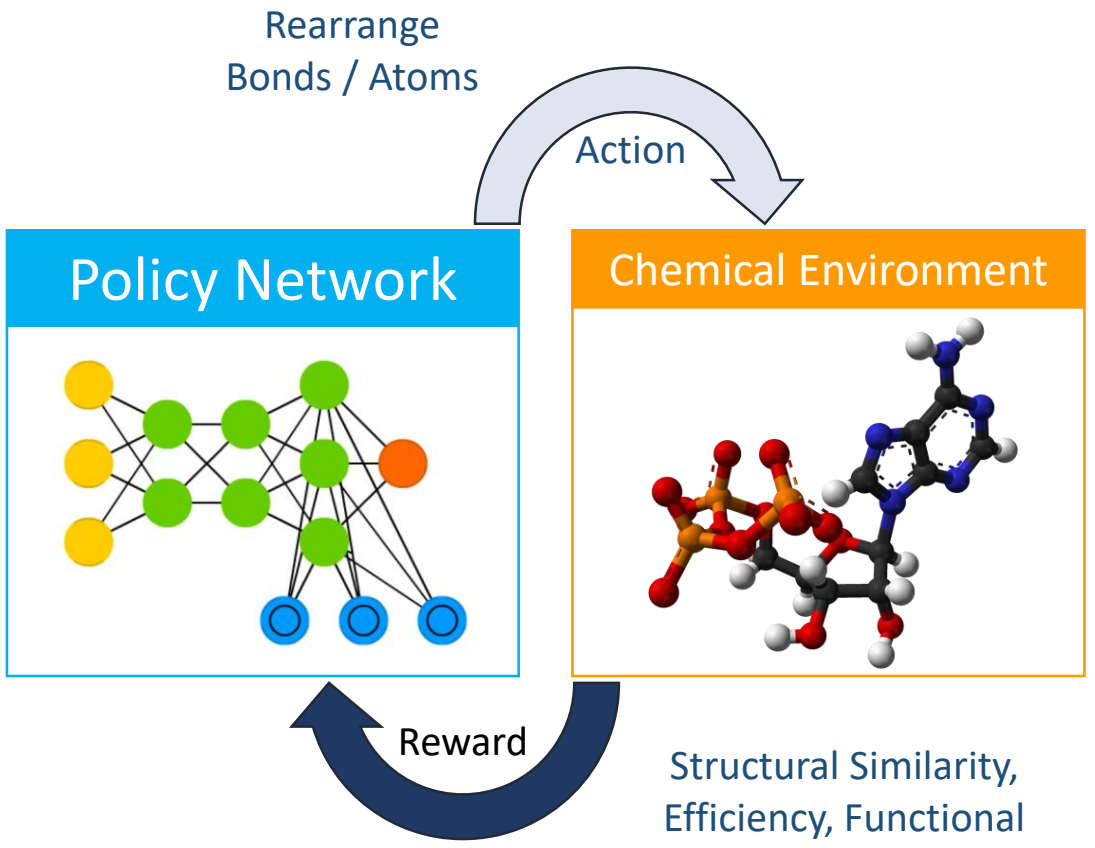
Reactants ▹ Products

Folded ▹ Unfolded

Good ▹ Great Molecule

- ✓ No Compression Bottleneck
- ✓ Arbitrary Input / output
- ✓ Reward Signal better than character level cross entropy

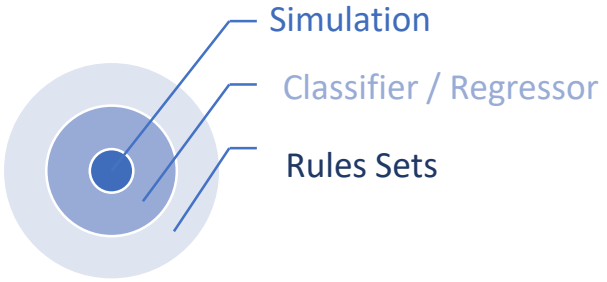
Deep Reinforcement Learning



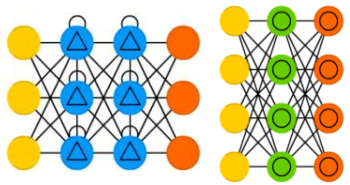
Candidate Search: 3 ways to discover

Pre 2016

Multi-Stage Virtual Screening



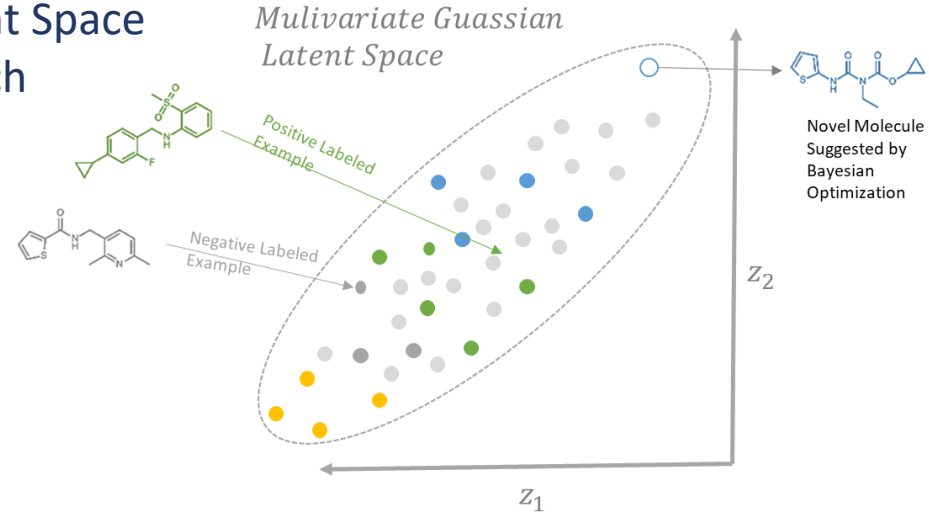
Improves



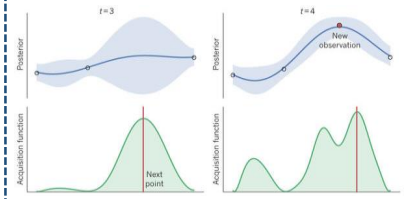
Generative Multitask Molecular Network

2016

Latent Space Search



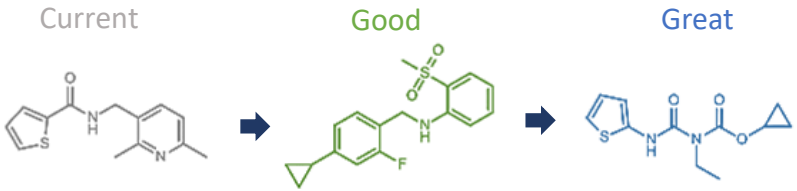
Improves



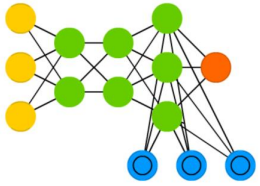
Model Assisted Bayesian Optimization

NEW

Search as Recursive Translation



Enables for the 1st Time

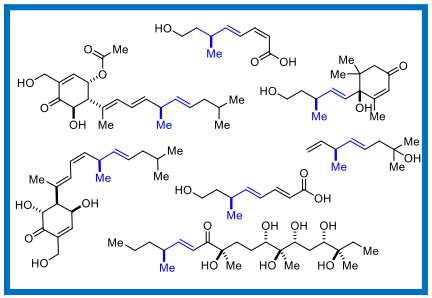


Graph Translation Policy Network

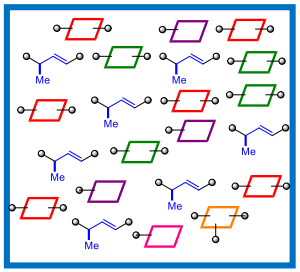
Synthesis Planning

Automation Friendly Natural Product Synthesis Library*

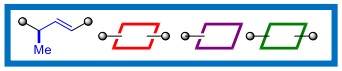
- New Heuristic and Combinatorial Optimization
- New Subgraph Isomorphism Clustering algorithms



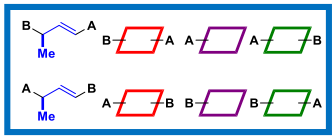
Linear natural products



FRAGMENTS from all allowed retrosyntheses



Smallest set of redundant FRAGMENTS that cover most natural product chemical space

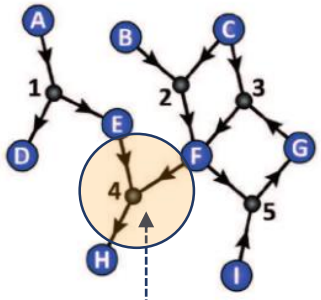


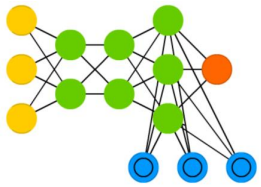
Double the FRAGMENTS to arrive at blocks

Generalized Synthetic Planning

Reactions

1) A → D + E
2) B + C → F
3) C + G → F
4) E + F → H
5) F + I → G





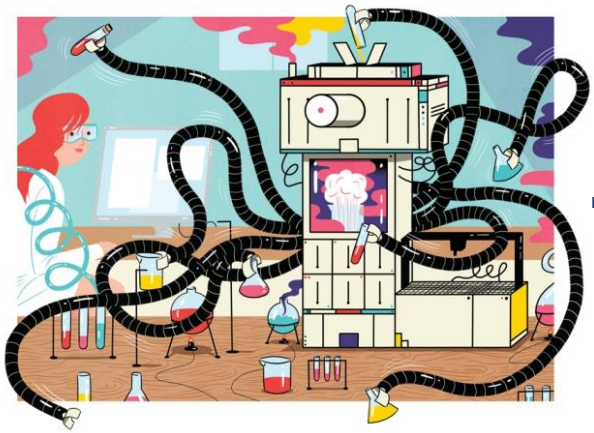
Graph Translation Policy Network

- MCTS based Retrosynthetic Planning can only use existing reactions
- New molecules will require new reactions and the GTPN can be used as a conditional generative model to propose new reactions given end points

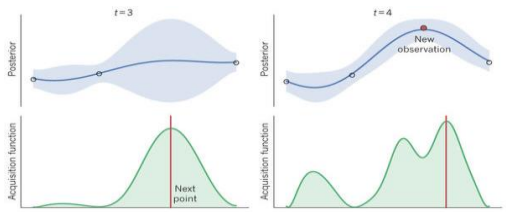
* Martin Burke, Andrea Palazzo, & Claire Simons are leading this endeavor

Optimize Synthesis

Lab Automation



Sequential Decision Making Agent

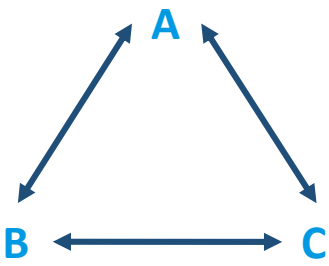
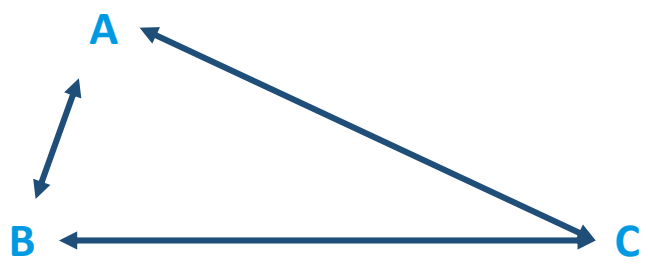


Model Assisted Parallel Bayesian Optimization

I. Pretrained model jointly optimizes over and learns latent distribution

II. Parallel Bayesian framework enables batch style lab automation

III. Learns within & between experiment(s)

	High Dimensional Sparse Binary Representation	Low Dimensional Dense Real Representation
Metric Space Properties		
Scaling N = # of Molecules D = Latent Dimensionality	N^2 Scaling	$N \; x \; D$ Scaling