

Outline

Sections

Problem and Solutions

- 1. P: Data problem in systems biology and metabolic engineering S: Torchcell, Neo4j, Biocypher
- 2. P: DCell visible neural network (Fitness) S: One Hot set net (Fitness)
- 3. P: DCell (interactions) S: DiffPool (interactions) - Compare ontology to learned clustering. Interpretability.
- 4. P: Multimodal models with expression data and morphology (multimodal learning) S: morphologic state prediction benefit from fitness and expression data?
- 5. P: Generative strain design (generative modeling) S: Solve the combinatorics problem when constructing multiplex mutants.

Models

- Set Net
- DCell
- Self-Attention Pool
- DiffPool
- cVAE

Dmf Fitness

model	nodes features (\mathcal{N})	edge features (\mathcal{E})	mae	mdae	pearson	spearman r^2	$\mu(1.0-1.1)$	$\sigma(1.0-1.1)$	$\mu(1.1-1.2)$	$\sigma(1.1-1.2)$
DCell	One Hot Genes	-								
Set Net	One Hot Genes	-								
Set Net	Codon Frequency	-								
Set Net	Fungal-UTR-Transformer Embeddings	-								
Set Net	Nucleotide-Transformer Embeddings	-								
Set Net	Prot T5	-								
Self-Attention Pool	One Hot Genes	PPI								
Self-Attention Pool	Codon Frequency	PPI								
Self-Attention Pool	Fungal-UTR-Transformer Embeddings	PPI								
Self-Attention Pool	Nucleotide-Transformer Embeddings	PPI								

model	nodes features (\mathcal{N})	edge features (\mathcal{E})	mae	mdae	pearson	spearman r^2	$\mu(1.0-1.1)$	$\sigma(1.0-1.1)$	$\mu(1.1-1.2)$	$\sigma(1.1-1.2)$
Self-Attention Pool	Prot T5	PPI								
DiffPool	One Hot Genes	PPI								
DiffPool	Codon Frequency	PPI								
DiffPool	Fungal-UTR-Transformer Embeddings	PPI								
DiffPool	Nucleotide-Transformer Embeddings	PPI								
DiffPool	Prot T5	PPI								

Features

Types: dna (sequence), interactions (edges)

- Graphs
 - PPI
 - GGI
 - Reg
- Ontology
 - GO
- Node Features
 - Median protein abundance
 - Median mRNA
 - Chromosome position
- Node Embeddings
 - Fungal UTR
 - Nucleotide Transformer
 - Codon frequency
 - ProtT5
 - One hot