



**WPI**

# Extracting Symbolic Models of Collective Behaviors

***Stephen Powers***

Joshua Smith

Carlo Pinciroli



<https://github.com/NESTLab/Extracting-Symbolic-Models-of-Collective-Behaviors>

# Why invent when you can steal?\*

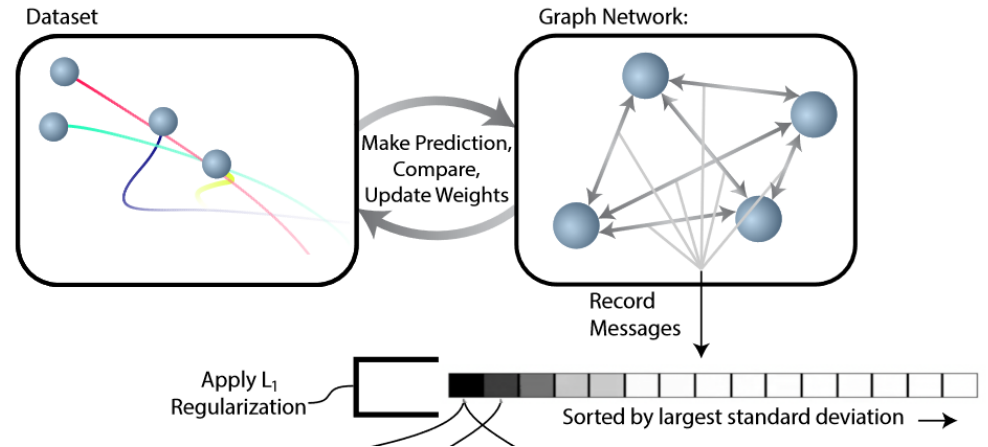
---

- Biomimicry
- Two options for me:
  - Become a biologist
  - Make friends with a biologist
- When in doubt, throw an AI at the problem
  - NNs are popular, use them?
  - Evolutionary Algorithms?

\* the presenter does not endorse stealing in any way from any sentient being.... the morality regarding stealing from non-sentient beings is a topic for very different presentation

# Combining the two

- GNNs
  - Learn
  - Isolate relationships

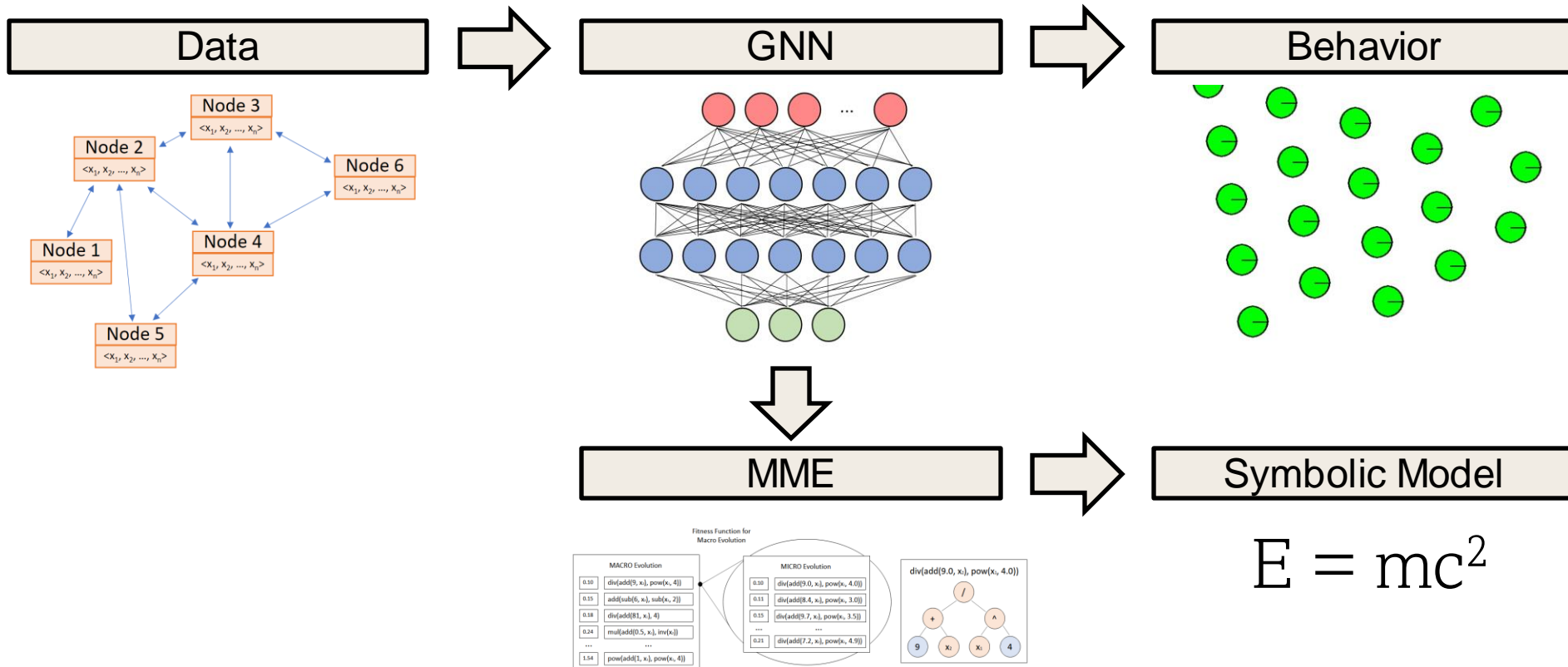


Discovering symbolic models from deep learning with inductive biases. Cranmer et al. [1]

- Evolutionary Algorithms
  - Symbolic models (able to be studied and manipulated)

[1] Cranmer, M.D., Sanchez-Gonzalez, A., Battaglia, P.W., Xu, R., Cranmer, K., Spergel, D.N., Ho, S.: Discovering symbolic models from deep learning with inductive biases.

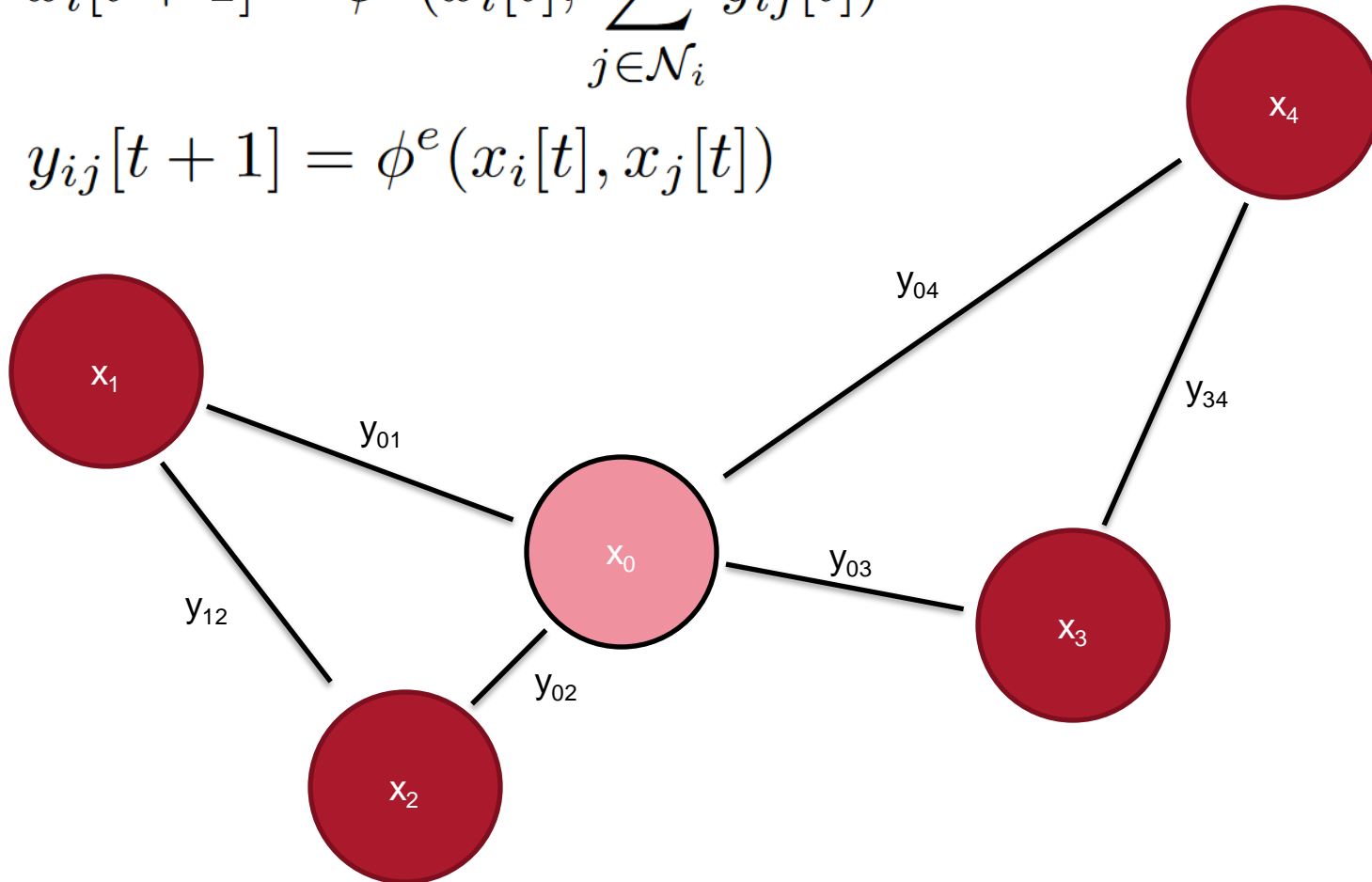
# Our system



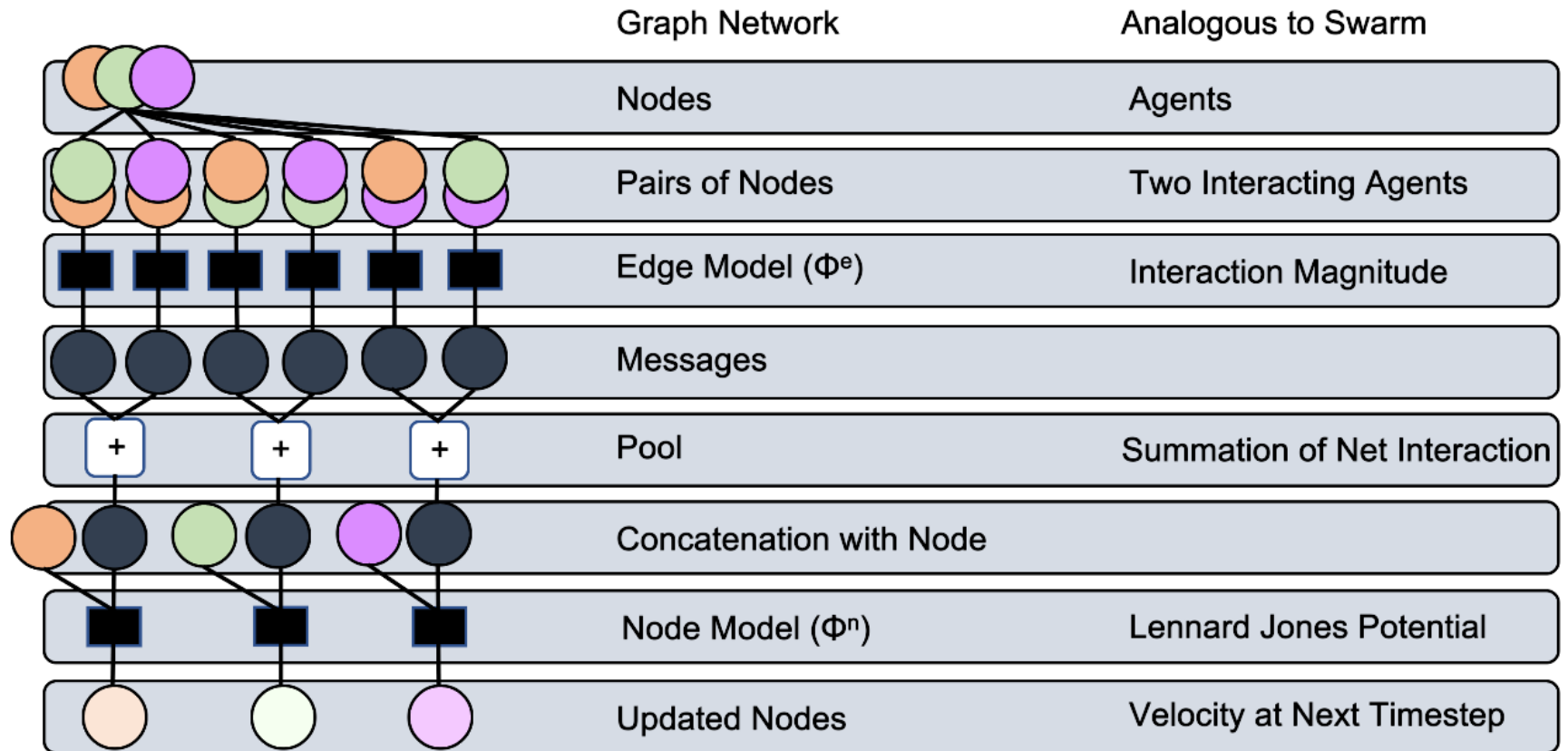
# GNN

$$x_i[t + 1] = \phi^n(x_i[t], \sum_{j \in \mathcal{N}_i} y_{ij}[t])$$

$$y_{ij}[t + 1] = \phi^e(x_i[t], x_j[t])$$



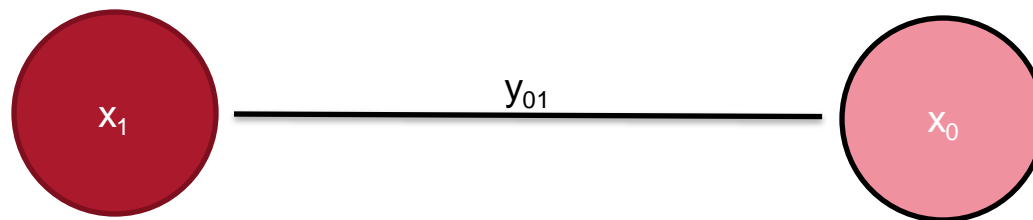
# How to use the GNN



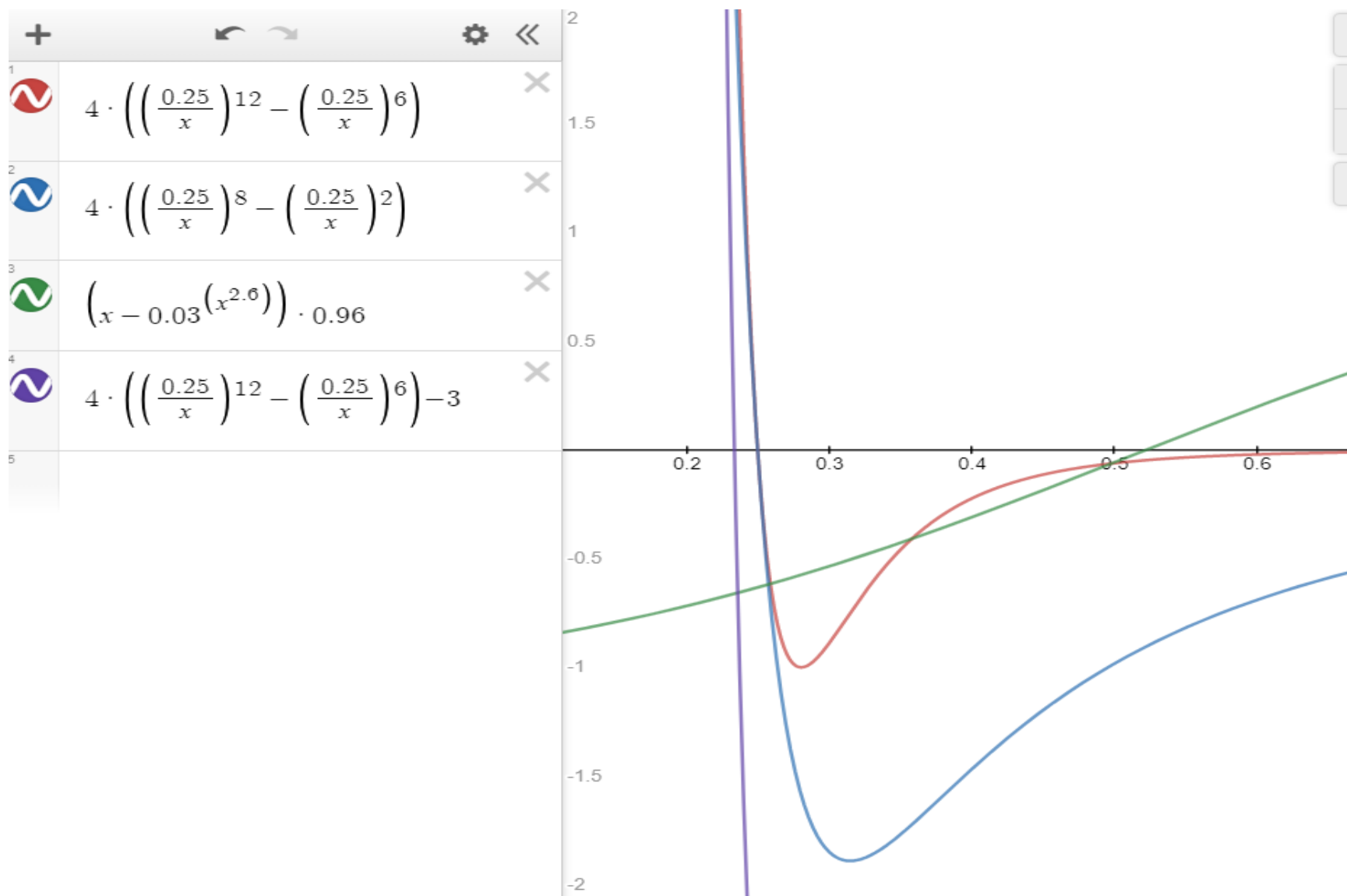
# Result

---

$$x_i[t + 1] = \phi^n(x_i[t], \sum_{j \in \mathcal{N}_i} y_{ij}[t])$$



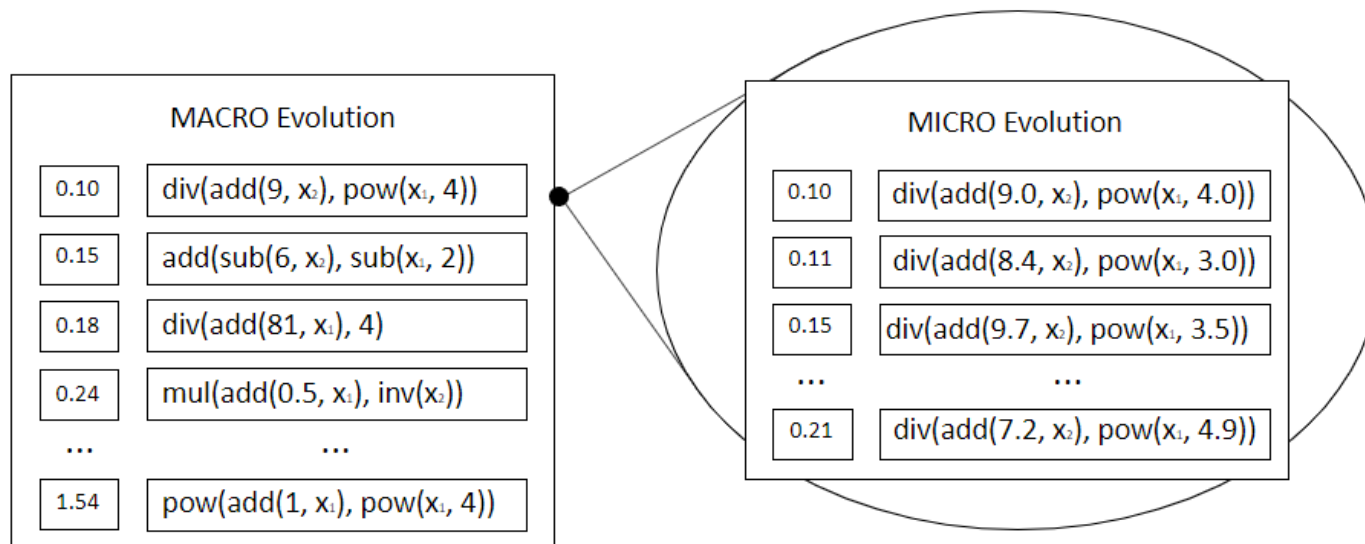
# Problem: Which one is the closest?



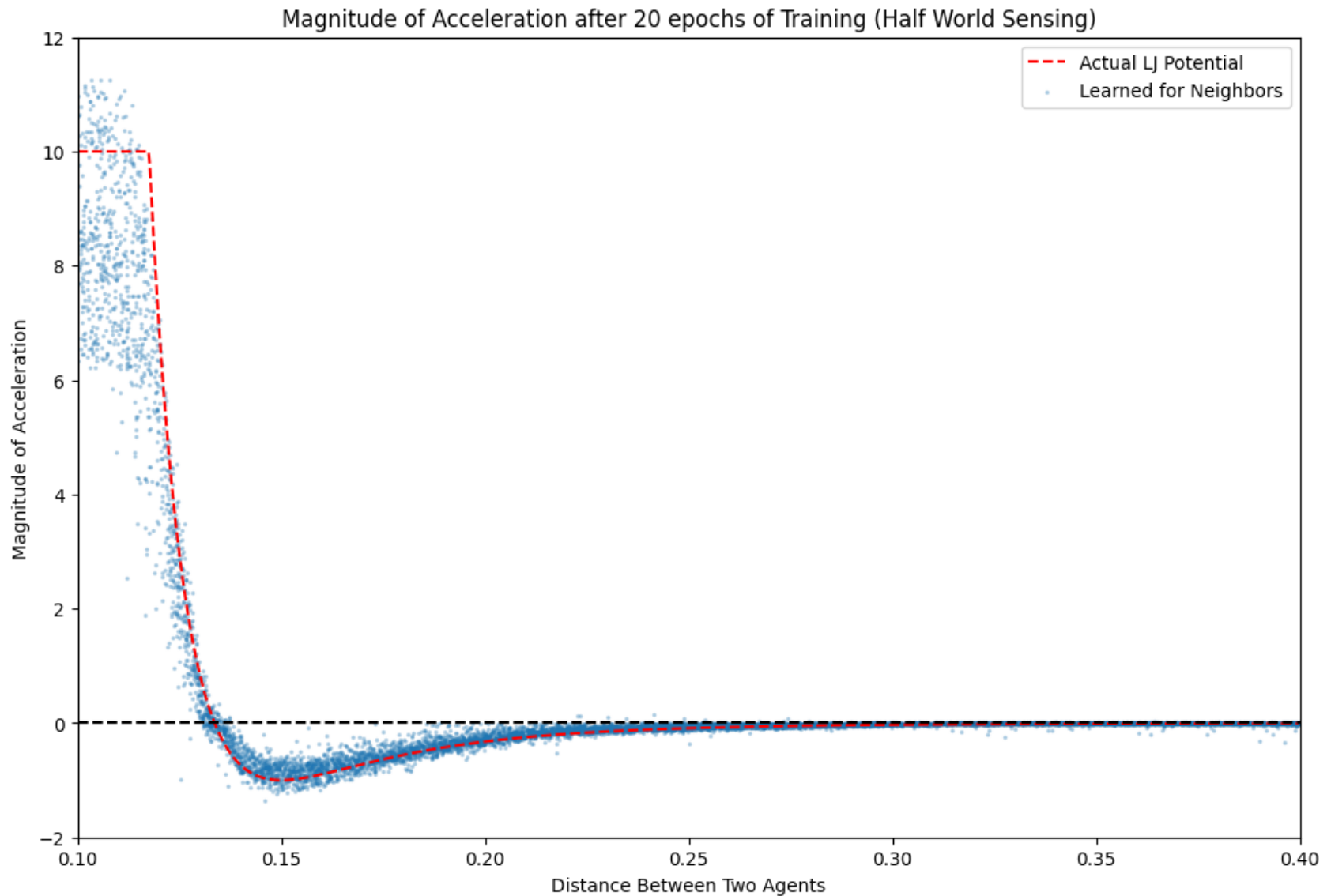


# Solution

- Embedding an EA inside an EA
  - Two optimizations:
    - Structure vs Params
  - Rank via MSE
    - Top survivors go through micro-evolution
      - More MSFI YAY!

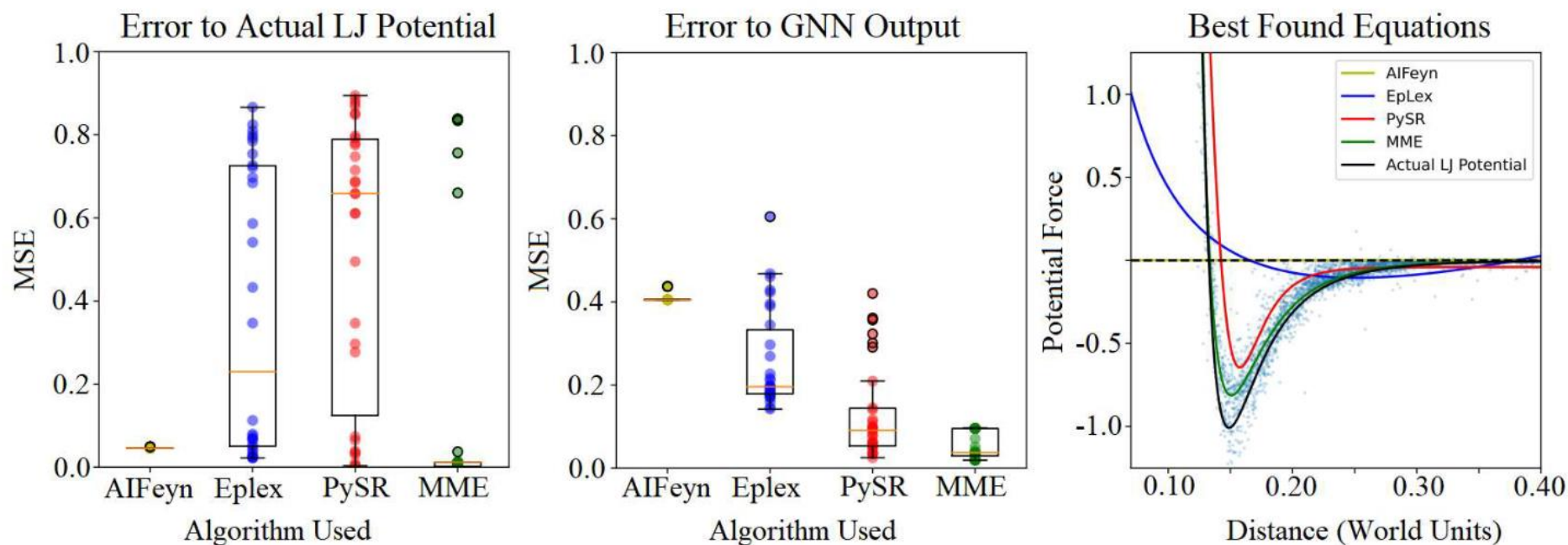


# GNN Output



# Symbolic Regression Results

- Each run 30 times



# MME Output

---

- Actual Equation (simplified)

$$\left( \frac{1.2e - 10}{x^{12}} \right) - \left( \frac{2.2e - 5}{x^6} \right)$$

- MME Found Equation

$$\left( \frac{8e - 9}{x^{10.07}} \right) - \left( \frac{9.8e - 6}{x^{6.54}} \right)$$

# Competition Output

---

$$\left(\frac{1.2e - 10}{x^{12}}\right) - \left(\frac{2.2e - 5}{x^6}\right)$$

- EpLex

$$2.38x - 1.3 + \left(\frac{0.15}{x}\right)$$

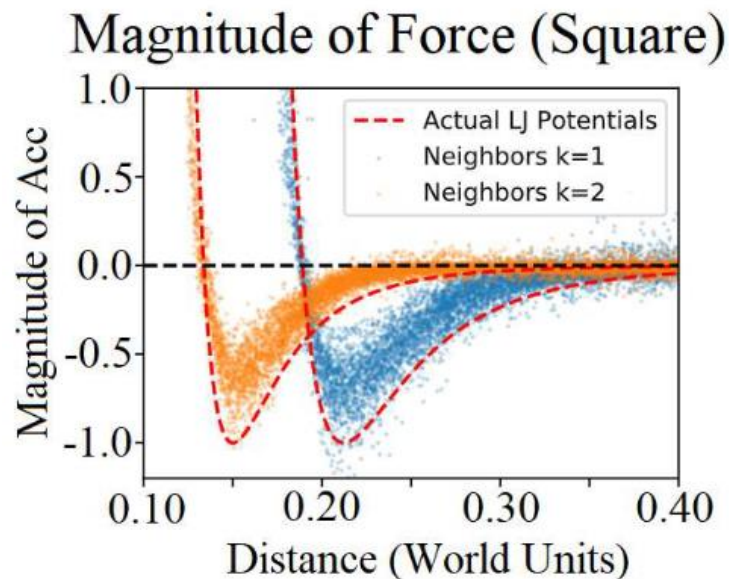
- AIFeynman

$$\left(5.6e - 5 \left(x + \frac{x^{0.5}}{-(x + 2)}\right)\right)^{0.5}$$

- PySR

$$\frac{-\left(0.42 - \frac{0.06}{x}\right)}{(x - 0.02)(8.14e8x^{11.66} - x + 0.272)} - 0.04$$

# Square Lattice Results



- Actual Equations

$$\text{Square (Kin)} \quad \left( \frac{7.84e-9}{x^{12}} \right) - \left( \frac{1.7e-4}{x^6} \right)$$

$$\text{Square (Non-Kin)} \quad \left( \frac{1.2e-10}{x^{12}} \right) - \left( \frac{2.2e-5}{x^6} \right)$$

- Found Equations

$$\text{Square (Kin)} \quad \left( \frac{6.8e-7}{x^{9.75}} \right) - \left( \frac{1.9e-5}{x^{7.75}} \right)$$

$$\text{Square (Non-Kin)} \quad \left( \frac{1.58e-9}{x^{10.67}} \right) - \left( \frac{4e-6}{x^{6.79}} \right)$$

# BOID Results

---

- Target Equation

$$2 \frac{\vec{x}}{||\vec{x}||} - \left( \frac{75}{||\vec{x}||} \right) \frac{\vec{x}}{||\vec{x}||} + 3\dot{\vec{x}}$$

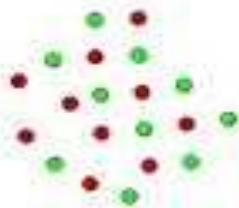
- MME Found Equation

$$0.59 \frac{\vec{x}}{||\vec{x}||} - \left( \frac{1}{||\vec{x}||} \right) \frac{\vec{x}}{||\vec{x}||} + \left( \frac{1}{||\dot{\vec{x}}||} \right) \dot{\vec{x}}$$

# Visual Results

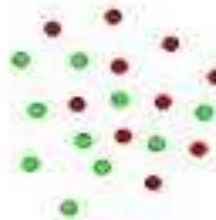


Original Simulation

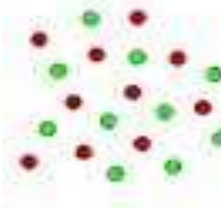


$$\begin{pmatrix} \frac{7.84e-9}{x^{12}} \\ \frac{1.2e-10}{x^{12}} \end{pmatrix} - \begin{pmatrix} \frac{1.7e-4}{x^6} \\ \frac{2.2e-5}{x^6} \end{pmatrix}$$

GNN



Symbolic Model



$$\begin{pmatrix} \frac{6.5e-7}{x^{9.35}} \\ \frac{1.58e-9}{x^{10.47}} \end{pmatrix} - \begin{pmatrix} \frac{1.3e-5}{x^{7.75}} \\ \frac{4e-6}{x^{4.75}} \end{pmatrix}$$





# Conclusion

---

- Wanted to find a way to model natural swarms
  - Not knowing original equations
  - Add in our potential knowledge of the situation
- Two stage system
  - GNN isolates relationships
  - MME performs symbolic regression
- Tested on Hex lattice, square lattice, and boids
  - Results match original equations more closely than other symbolic regression algorithms

# Questions?

---

[1] Cranmer, M.D., Sanchez-Gonzalez, A., Battaglia, P.W., Xu, R., Cranmer, K., Spergel, D.N., Ho, S.: Discovering symbolic models from deep learning with inductive biases. CoRR abs/2006.11287 (2020), <https://arxiv.org/abs/2006.11287>

# Holy cow, MME must take forever.

---

- Kinda no, but sorta yes.
  - It does NOT scale well relative to amount of data
    - Linked to the complexity of the target equation
      - Can be reduced by priors
  - Planning to submit extended version of work with MME analysis

# Cool story, where's the real data?

- FISH!

The screenshot shows the Oregon State University ScholarsArchive@OSU website. The header is orange with the OSU logo and 'Oregon State University' text. Below the header is a black bar with 'ScholarsArchive@OSU'. A navigation bar contains links for Home, About, Help, and Contact, along with a search bar and a 'Go' button. The main content area shows the 'Fish Schooling Data Subset' dataset page. It includes a breadcrumb trail 'Home / Fish Schooling Data Subset', a 'Dataset' icon, and a title 'Fish Schooling Data Subset' with 'Public' and 'Deposited' tags. A 'Citeable URL' is provided: <https://ir.library.oregonstate.edu/concern/datasets/zk51vq07c>. The 'Descriptions' section includes a table with 'Attribute Name' and 'Values'. The 'Creator' row lists: Katz, Yael; Tunström, Kolbjørn; Ioannou, Christos C.; Huepe, Cristián; and Couzin, Iain D. The 'Abstract' row describes the data as a JSON format file containing position, velocity, and fish identifier data for 300 golden shiners in a shallow rectangular water tank, with 5000 individual frames and a rate of 30 frames/s.

Oregon State University  
ScholarsArchive@OSU

English Login

Home About Help Contact

Enter search terms or select 'Go' to browse

Go

Advanced Search

Home / Fish Schooling Data Subset

Dataset

Fish Schooling Data Subset Public Deposited Analytics

No preview available  
[Download the file](#)

Citations: ▼

f t

Copy link

Citeable URL: <https://ir.library.oregonstate.edu/concern/datasets/zk51vq07c>

Descriptions

Attribute Name	Values
Creator	Katz, Yael Tunström, Kolbjørn Ioannou, Christos C. Huepe, Cristián Couzin, Iain D.
Abstract	The data is a JSON format file containing the position, velocity, and fish identifier data for 300 golden shiners in a shallow (depth of 4.5 to 5 cm) rectangular water tank (2.1 by 1.2 meters). There are 5000 individual frames (samples of position and velocity) corresponding to video taken at a rate of 30 frames/s and analysed to extract individual fish's trajectories. The fields px, py, vx, vy

<https://ir.library.oregonstate.edu/concern/datasets/zk51vq07c>

# Sneak Peak!

---



REAL fish data:  
~300 fish



GNN generated fish:  
~50 fish

# Turing Learning. It's nifty.

---

- Yes. It is. Look out for our next papers!

