

Visualizing Genetic Programming Ancestries

McPhee¹, Casale², Finzel, Helmuth, & Spector

¹Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA

²Design Center Inc.
St. Paul, Minnesota, USA

21 July 2016
VizGEC

Why?

Say something clever about why we wanted to visualize these runs.

Outline

- 1 Background
- 2 Graph Structure
- 3 What did I learn?
- 4 Conclusions

Outline

1 Background

- Evolutionary Computation
- Run Configuration
- Database Configuration

2 Graph Structure

3 What did I learn?

4 Conclusions

Evolutionary Computation Versus Biological Evolution

Run Structure

In a run, we have many generations of individuals or programs.

- 1,000 individuals per generation
- Linear Plush Genome [?]
- Lexicase selection [?]
- Replace-Space-With-Newline
 - Two tasks
 - So two error values per test case
- 100 test cases, so 200 error values stored in vector

Database Content

- uuid
- uuid of parent(s)
- genetic operators
- plush genome
- error vector
- number of children
- number of selections
- Damerau-Levenshtein Distance

Outline

1 Background

2 Graph Structure

- Nodes
- Edges

3 What did I learn?

4 Conclusions

Node Basics

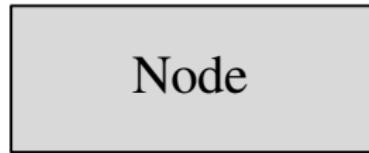
Nodes represent an individual or program.

They know:

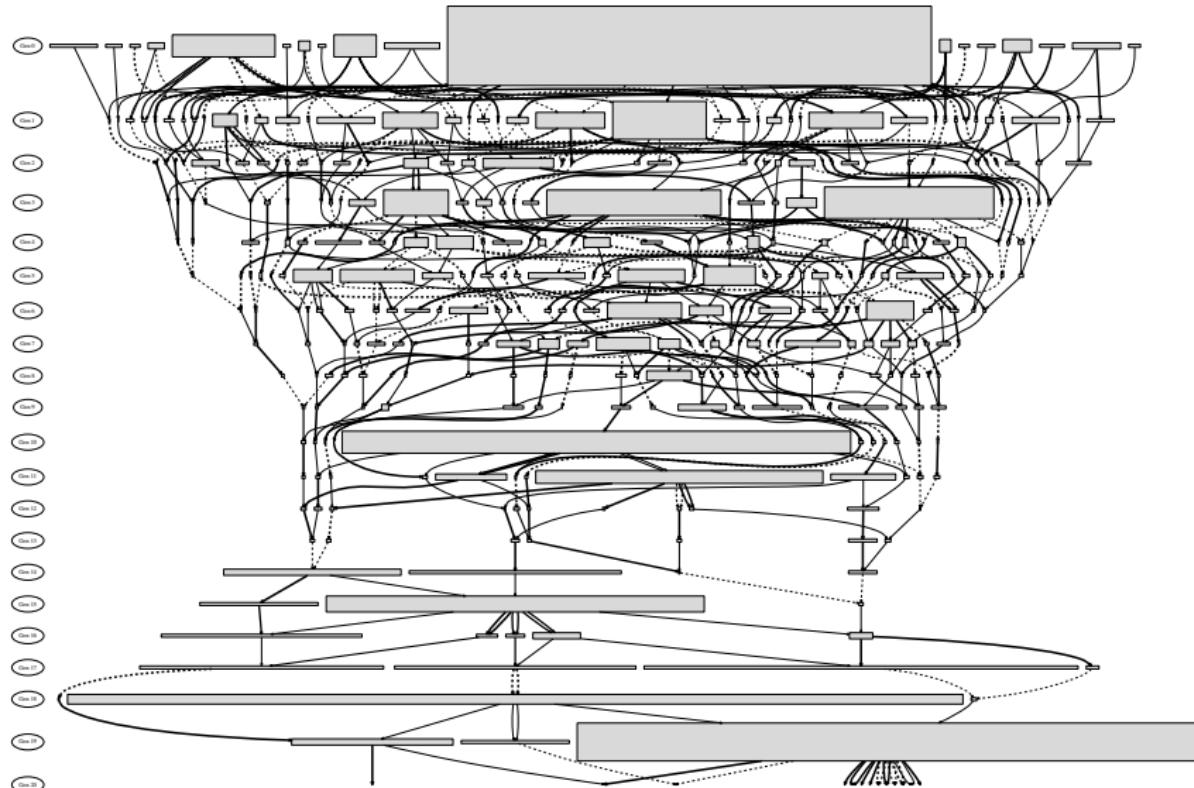
- Number of selections: Width
- Number of children: Height

← Number of Selections →

↑ Number of Children ↓

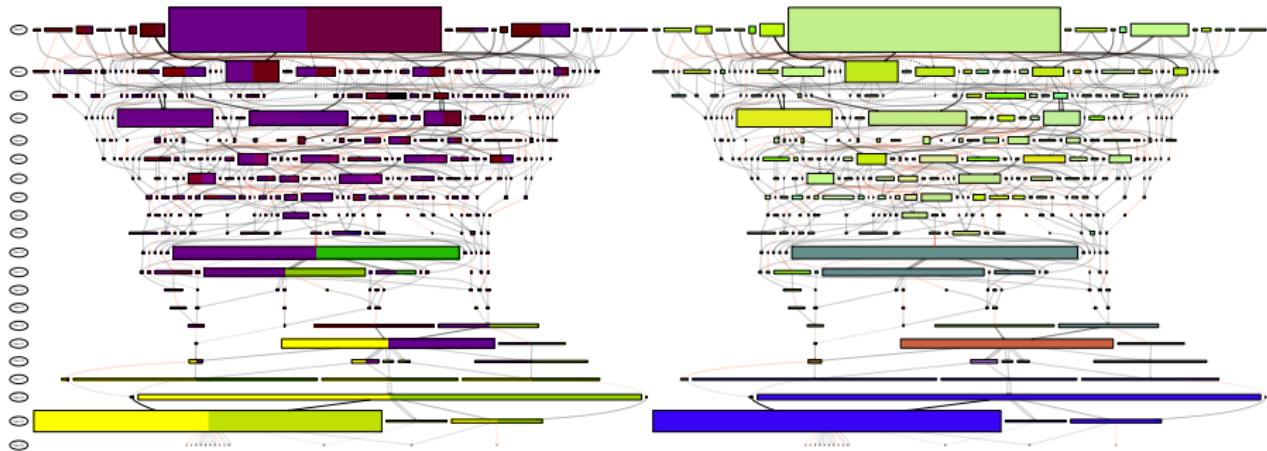


Run 0: No color

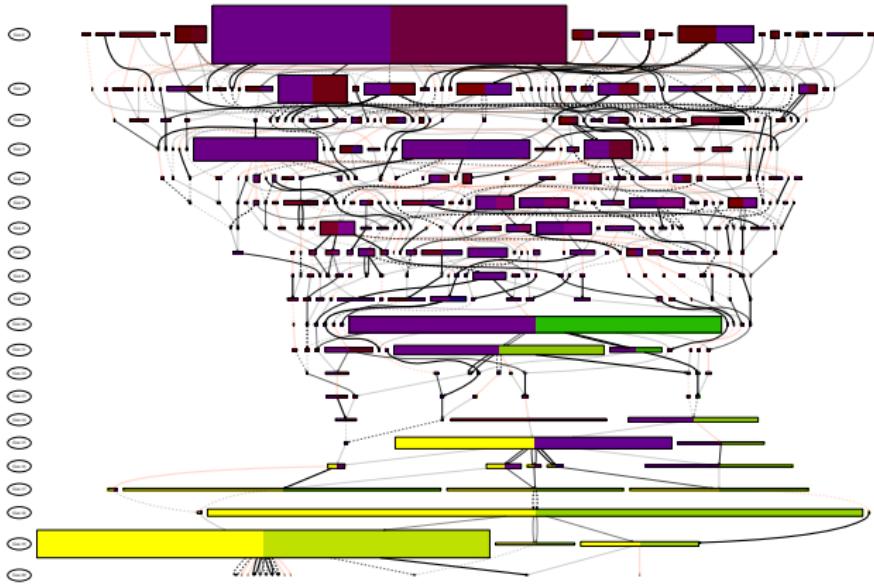


Node Coloring

There are two different techniques we use:
Error Based & RBM

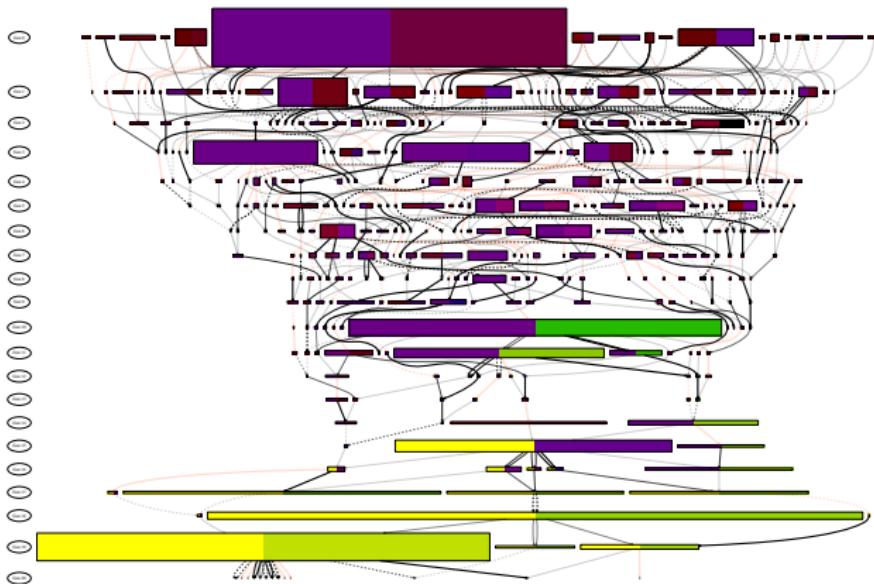


Error Based Node Coloring



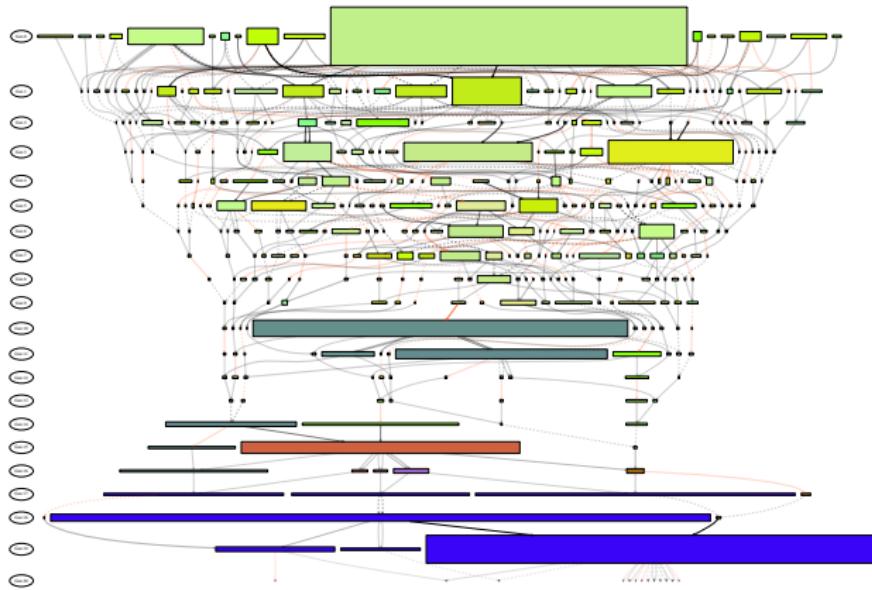
Two part problem, and based on an individual's total error.

Error Based Node Coloring



Additional shading for very high errors.

Restricted Boltzmann Machines Node Coloring



Dimensionality Reduction Problem, 200 tests \rightarrow 3 values.

Genetic Operators

Edges represent how an individual was created.

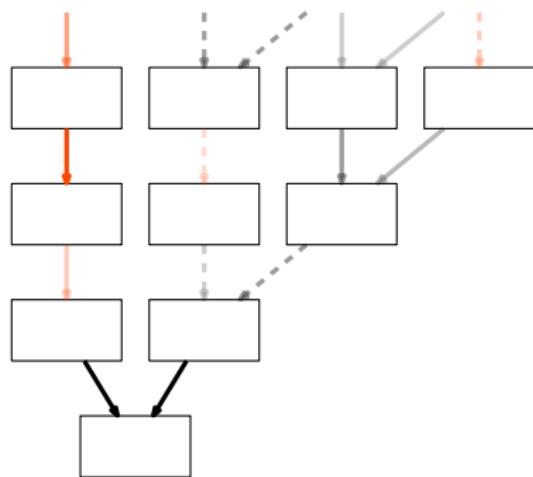
There are two main techniques: crossover and mutation.

- Alternation
- Uniform-Mutation
- Alternation & Uniform-Mutation
- Uniform-Close-Mutation

Edge Coloring & Style

Edges Coloring:

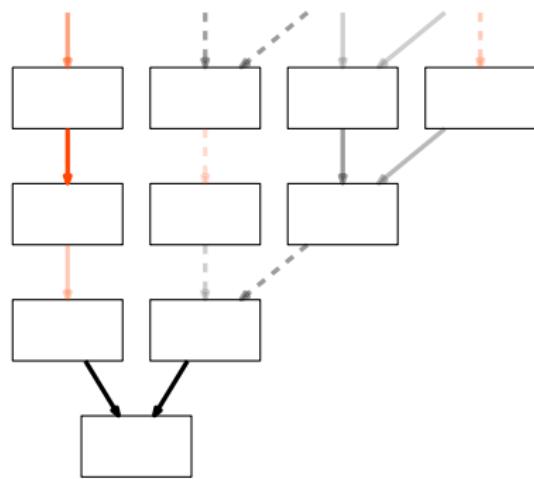
- Alternation is black and solid
- Uniform-Mutation is orange and solid
- Alternation & Uniform-Mutation black and dashed
- Uniform-Close-Mutation is orange and dashed



Edge Coloring & Style

Edge Opacity:

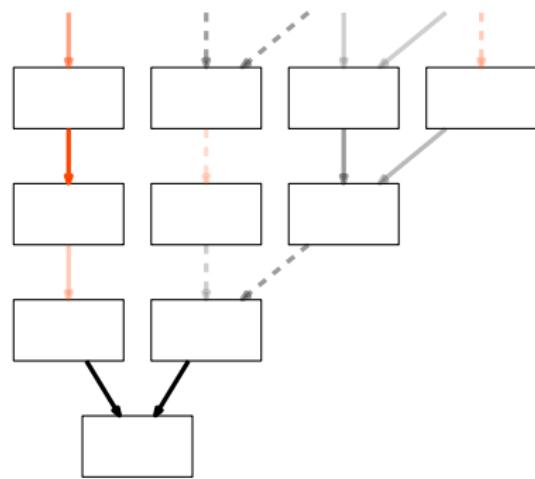
- Transparent: Few Children
- Opaque: Many Children



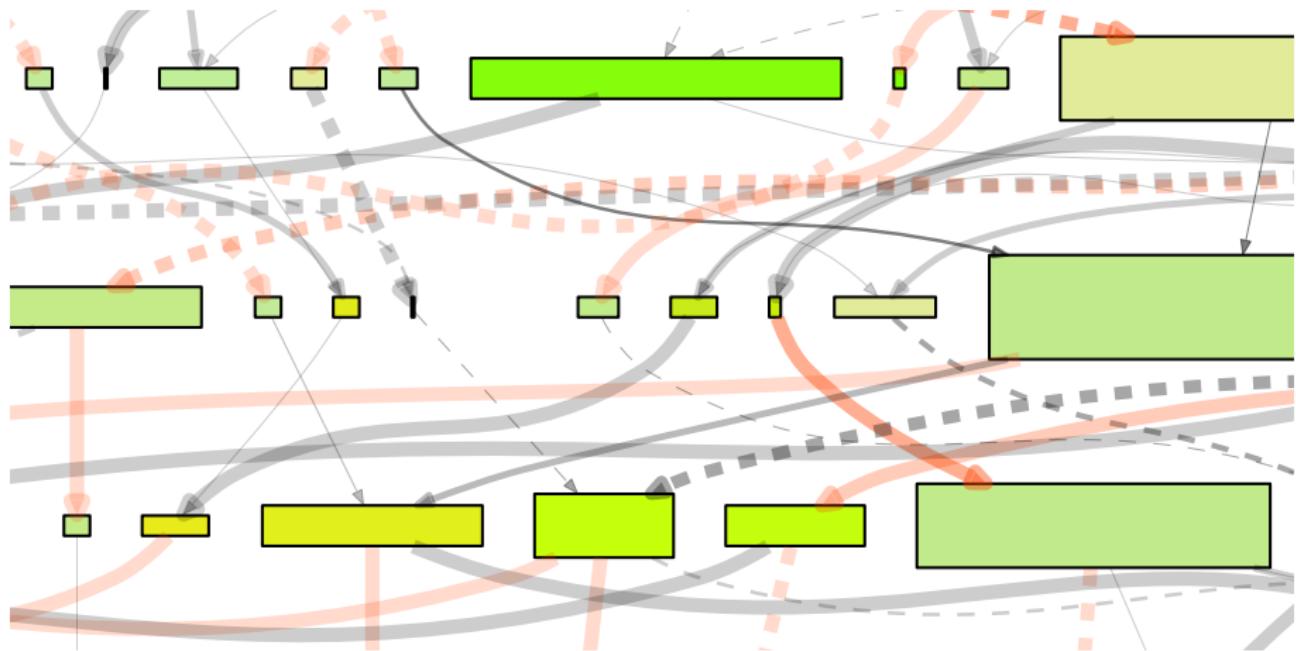
Edge Coloring & Style

Edge Thickness:

- Thin: Low DL-Distance
- Wide: High DL-Distance

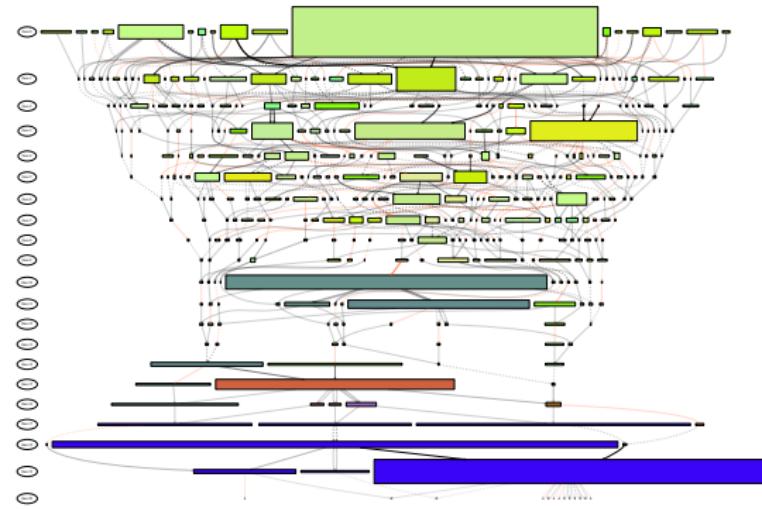


Edge Coloring & Style Example



Full Examples in RBM Coloring

- Run 0: Ends in 20 Generations
- Run 1: Ends in 129 Generations
 - Notice waist with large nodes



Outline

1 Background

2 Graph Structure

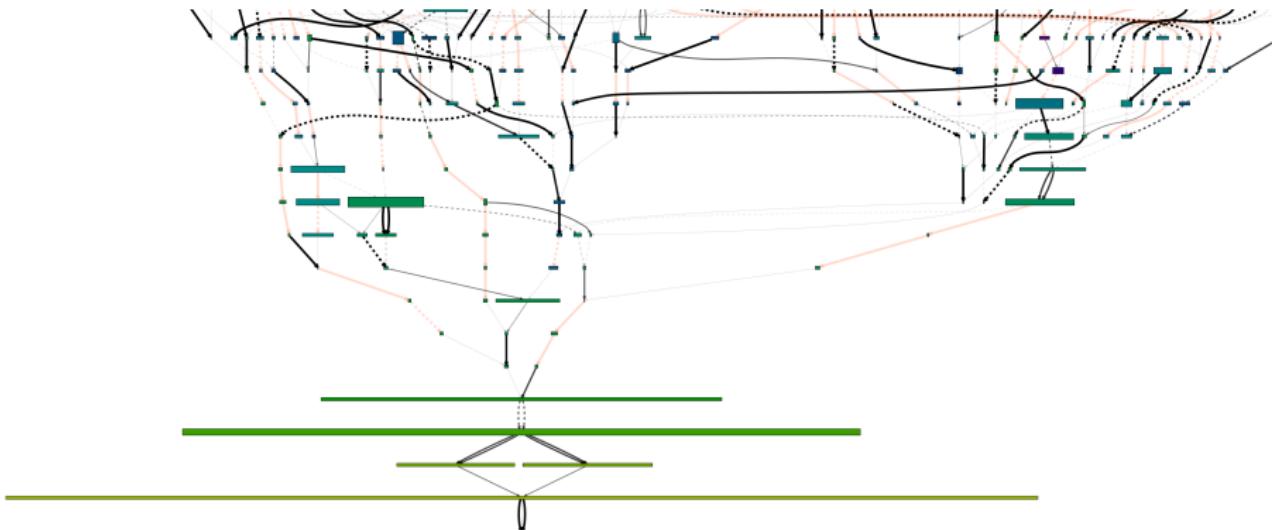
3 What did I learn?

- Hyperselection
- Filtering Ancestry Trees

4 Conclusions

Hyperselection

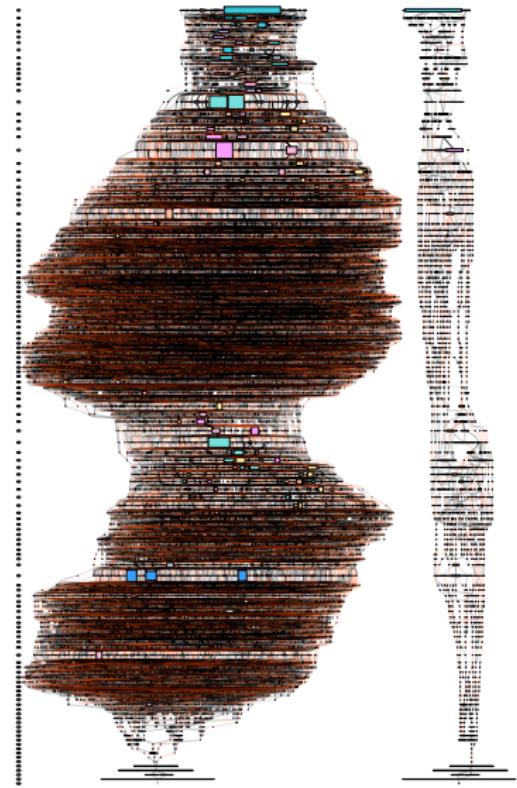
One focus of this research was to examine *hyperselection*. This occurs when an individual receives a noticeably high number selections in a population.



In our graphs, we can see this as a very wide node.

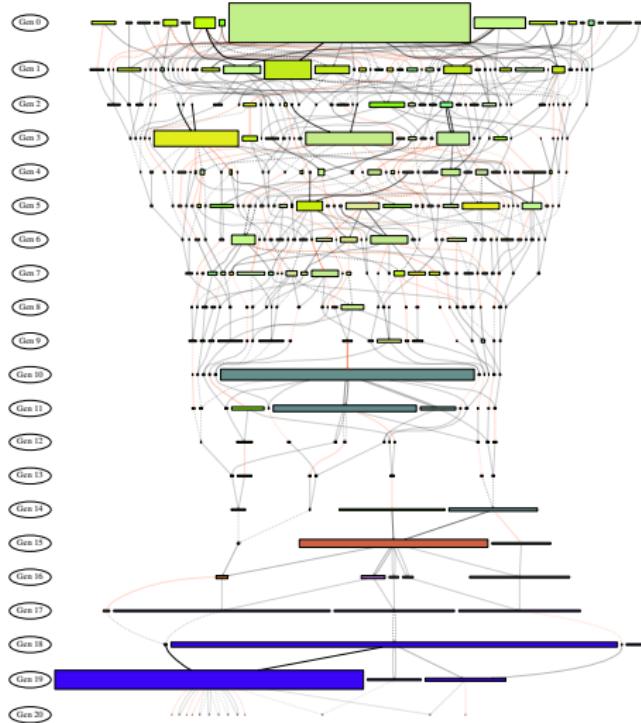
Filtering Ancestry Trees

- This is still a lot of data!
 - 22,435 Individuals
 - 1,597 Individuals
- We filter out parents who did not contribute much genetic information.
- Numbers on edges indicate similarity.
- The smaller the number is, the more similar they are.

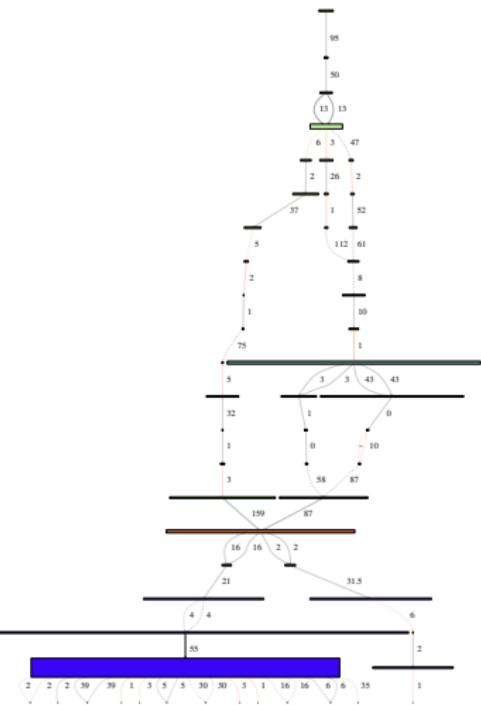


Another Filtered Example

394 Individuals



54 Individuals



Outline

1 Background

2 Graph Structure

3 What did I learn?

4 Conclusions

Conclusions

In Our Research We've:

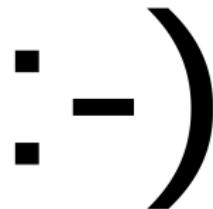
- Turned our complex runs into organized ancestry trees.
- Modified aspects of both nodes and edges to visualize analytics.
- Learned about hyperselection and filtering ancestry trees.

The Next Step:

- Presenting related work in May at GPTP in Ann Arbor, Michigan
- In July we'll be attending GECCO in Denver, Colorado

Thanks!

Thank you for your time & attention!



Special Thanks to Nic McPhee!

References

-  B. Burlacu, M. Affenzeller, M. Kommenda, S. Winkler, and G. Kronberger.
Visualization of genetic lineages and inheritance information in genetic programming.
In Christian Blum, *et al*, editors, *GECCO '13*, pages 1351–1358, New York, NY, USA, 2013.
-  T. Helmuth, N. F. McPhee, and L. Spector.
The Impact of Hyperselection on Lexicase Selection.
GECCO '16, Denver, CO, USA 2016.
-  N. F. McPhee, M. Casale, M. Finzel, T. Helmuth, and L. Spector.
Visualizing genetic programming ancestries using graph databases.
GECCO '16, Denver, CO, USA 2016.
-  N. F. McPhee, M. Finzel, M. Casale, T. Helmuth, and L. Spector.
A detailed analysis of a PushGP run.
(Forthcoming) *GPTP '16*, Ann Arbor, MI, USA 2016.