Analysis of Ancestry in Genetic Programming with a Graph Database

David Donatucci
M. Kirbie Dramdahl

1 May 2014
CSci4453: Databases

The Big Picture

- Genetic programming demonstrated to be effective for a variety of applications.
- Difficult to determine how this process works.
- Databases allow examination of the internal interactions of a run.
- Graph databases more efficient at this task than relational databases.
- This knowledge may be used to improve genetic programming algorithms.

Outline

Outline

Genetic Programming Overview

- Genetic Programming is based upon biological principles.
- Individuals form a population.
- Transformations
  - Crossover (XO)
  - Mutation
  - Reproduction
  - Elitism
- Transformations occur over a specified number of generations.
- Individuals are rated by their fitness.

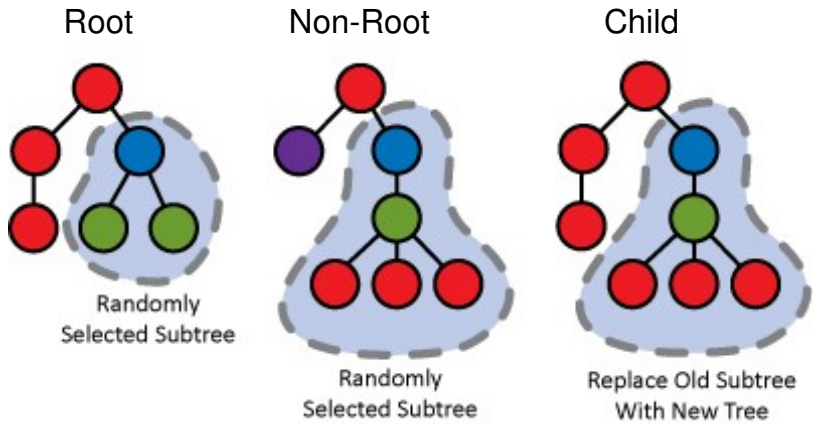
Transformations

Crossover sexual reproduction (root and non-root)

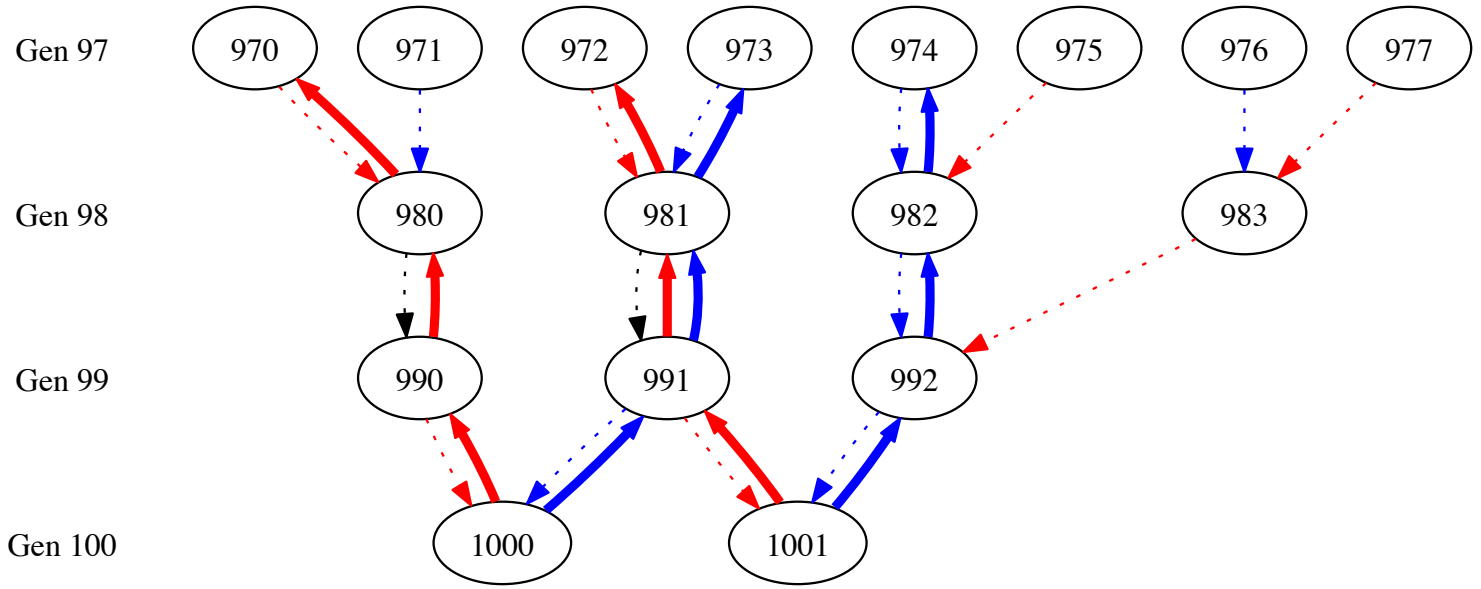
Mutation subtrees altered

Reproduction asexual reproduction

Elitism reproduction based on fitness



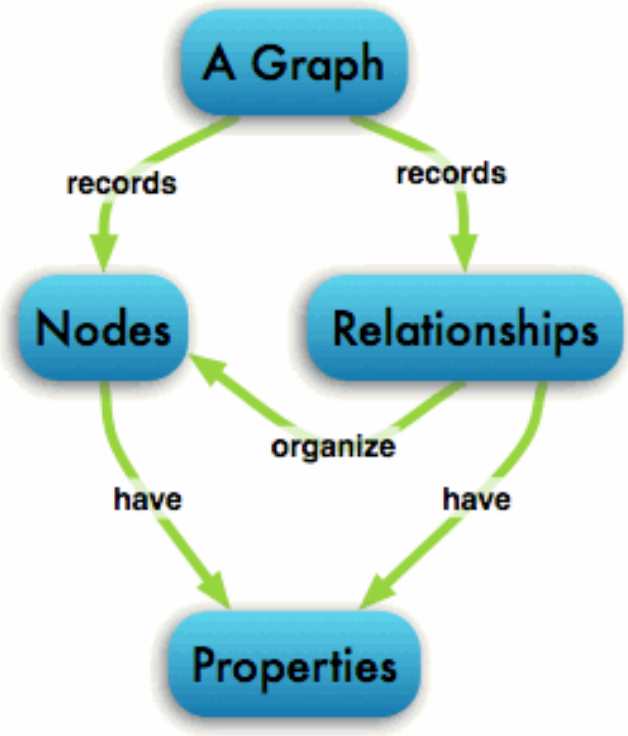
geneticprogramming.us



Outline

Neo4j

- information is stored using a graph
- nodes and relationships
- efficient recursive queries compared with traditional databases



Neo4j `http://goo.gl/nzRWSV`

Outline

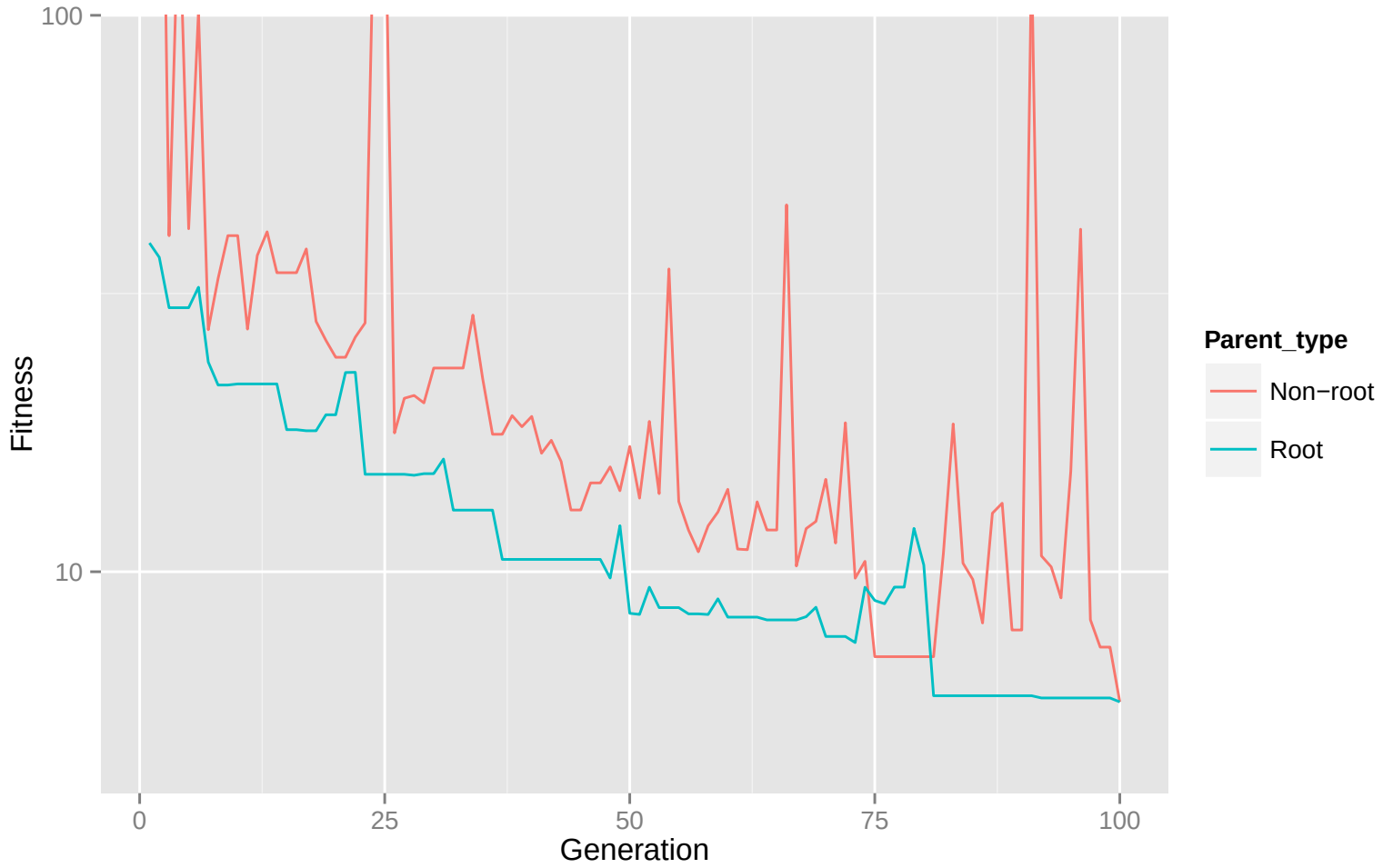
Run Configurations

Target Function	sin(x)
Variables	x (range 0.0 to 6.2, incremented by steps of 0.1)
Constants	range between -5.0 and 5.0
Operations	addition (+), subtraction (-), multiplication (*), protected division (/)
Generation Number	100
Population Size Per Gen	1,000 and 10,000
Transform Percentages	crossover (90%), mutation (1%), reproduction (9%)
Elitism	best 1%
Fitness	absolute error between target function and individual function

Outline

Fitness Over Time

What does the fitness of the "winning" individual's ancestry line look like over time?

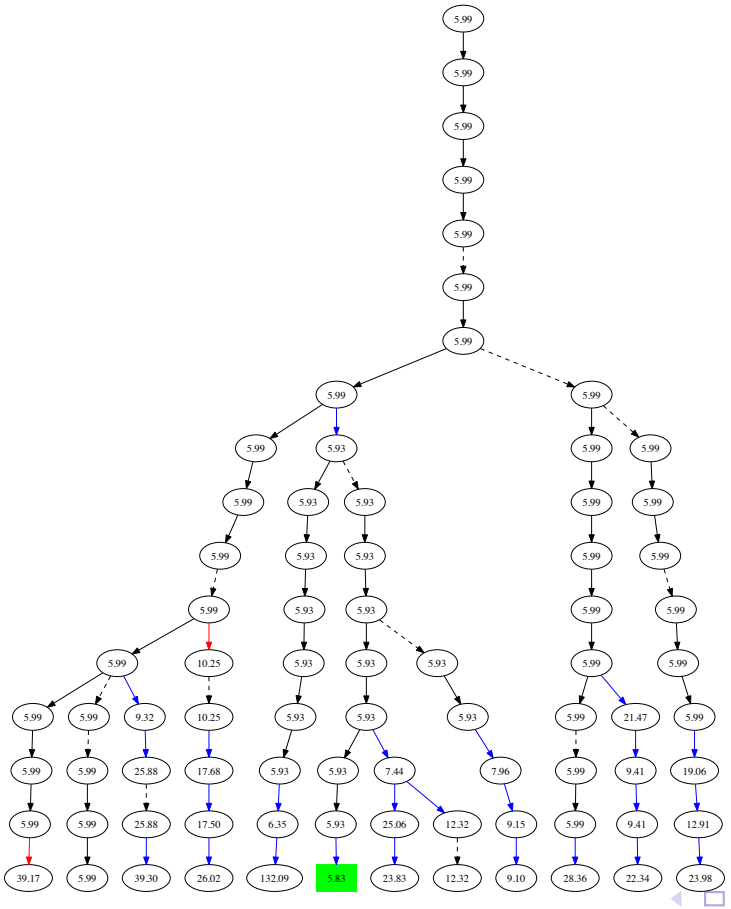


Percentage of Improved Transformations

How often does mutation and crossover improve fitness?

Common Ancestor

Does a group of individuals have a common root parent ancestor and how many initial generation individuals have descendants in the final generation?



Outline

Conclusions

- We can gather internal data efficiently.
- Provides more in depth information than statistical summaries.
- Support for hypotheses.

Future Work

- Trying different setup configurations.
- Enforcing the root parent to have better fitness in XO.
- Dynamically change parameters.

Thanks!

Questions?