Corbin Hopper COMP 561

Network Evolution Along Tractable Lines

I. Introduction

Biological networks are known to be computationally complex systems, but their way of untangling complexity is an ongoing study. An evolutionary model is defined to evolve a population of arbitrary networks based on their tractability on an NP-Hard problem. A previously defined network evolution problem (NEP) reduces to a knapsack optimization problem. NEP was previously reduced to the knapsack optimization problem and shown to be fundamentally NP-Hard. It is clearly grounded as both a biologically realistic scenario and a computational optimization problem. This was used to show that the topology of biological networks develops easier instances of NP-Hard problems.3

The evolutionary algorithm is built around generating easier instances of knapsack problems. Change in fitness over time is evaluated under varying environmental pressure, mutation frequency, and crossover percent. Lower mutation frequency is shown to more reliably balance the different aspects of tractability, but eventually reverts to only one property. A breadth first crossover operation is implemented, but is ineffective in its current form. A revised definition of tractability that is normalized over time, scale-invariant mutation frequency, and crossover with population agreement of node identity are proposed as improvements to the model.

II. The Network Evolution Problem (NEP)

NEP formally describes parameters that are used in the evolution of tractability. From the biological view, the network represents a gene regulatory network where nodes are genes and edges are simplified interactions (promotional or inhibitory). The omnipresent environment is termed the oracle, who gives ‘advice’ to the network in terms of which genes should be conserved or deleted. The difficulty resides in the interconnected nature of the network. For instance, conserving one gene also preferences its neighbors that it promotes, but these might be deemed unworthy by the oracle.

The problem of which genes to conserve is formally reduced to the knapsack optimization problem, where genes are included or not without redundancy. In this light, each gene is an object. The sum of the neighbors it influences in accordance with the oracle advice is its benefit, while the sum of adjacent nodes it affects contrary to the oracle advice is its damages. Pressure is the percent of nodes for which the oracle has a preference. Higher pressure tends to be less tractable as more adjacent nodes may become entwined. Meanwhile, tolerance is the number of influenced neighboring nodes contrary to the oracle advice allowed in the solution. 3

III. Biological Network Tractability

The knapsack optimization problem is NP-Hard, but not all knapsack instances are created equal. Biological networks have been shown to produce more tractable instances.Scale-free networks representative of biological systems tend to differentiate their nodes into densely connected hubs and sparsely linked outliers. These two properties generate simpler solutions to NP-Hard problems that can be accessed by mutations. Dense hubs increase the ability to garner many benefits by conserving or deleting a single node via a point-wise mutation. This is referred to as effective total benefits (ETB). Meanwhile, sparse peripheral nodes pose more obvious solutions. With only a few adjacent nodes, they often supply only benefits or only cause damages to a given knapsack instance. Hence they are a priori included in (green) or excluded from (red) the knapsack, again with a single point mutation. These are referred to as red-green nodes, as opposed to grey nodes that confer a mix of benefits and damages. One interpretation of this property is the red-green to grey node ratio (RGGR).1

Mutations are straightforward. ETB and RGGR are defined in such away that a single point mutation is capable of confers a large (ETB) or clear (RGGR) benefit. However, crossover on topologically based problems has been historically difficult. Early neuroevolution algorithms faced a similar issue: arbitrary crossovers could wreck the structure of the solution. One solution such as NEAT, is to record the history of evolution, such as added nodes and edges, and align solution sequences during crossover. Hence the population moves towards a consensus for each node. Moreover, NEAT solidifies premature crossover innovations by protecting newly crossed individuals by a speciation mechanism.4 Another solution proposes that coherent building blocks of the problem at hand should be conserved as independent components.2 For NEP, a node’s benefits and damages are based on interactions one edge away. Hence a breadth-first search is a natural description of the topology based on the problem definition. Breadth-first search crossover might then conserve tractable components.

Crossover on a highly interconnected network is, by nature, destructive. If complex systems are design to keep networks untangled, they may also be less prone to the trauma of crossover. The small-world nature of scale-free networks means that certain groups of nodes are tightly clustered, but few edges link disparate clusters. This could mean that crossover is less destructive to the topology of biological networks, since modules are coherent units that are easily spliced from the rest of the network.

IV. The Model

Evolution of tractability is defined in the context of several notable concepts, which correspond to a high level interpretation of model functions.

**Environmental pressure -** During evolution NEP is used to generate pressure in the form of knapsack instances based on a given network. The benefits and damages of each node are returned and whether or not they are included in the knapsack solution. A multi-set of the number of benefits of each node included in the knapsack solution is formed. ETB is then calculated from the sum of the non-redundant set of the multi-set. RGGR is calculated as the number of nodes with only benefits or only damages divided by nodes with mixed edges.

**Fitness of response -** Both ETB and RGGR describe different aspects of tractability and are incorporated into a network’s fitness. Ideally they would be equally weighted. Since their range is not known beforehand and changes as networks grow, each individual network is normalized relative to the rest of the population. ETB and RGGR are separately shifted to fall within [0,1] by subtracting the population’s minimum and dividing by the population’s range (in that category). As a result fitness only describes a network’s tractability in relation to other individuals at the same generation. As a result, quality of a parameter set is evaluated on its change in ETB and RGGR over time. This also allows for a separate analysis of the two components and an understanding of their balance (or lack of).

**Breeding** – Once fitness is evaluated from multiple knapsack instances, only the 4 fittest networks in the population survive out of a population of 40 individuals. New individuals are either created as exact replicates of the survivors or are a cross between two survivors. The number of individuals born from crossover instead of exact duplicates of survivors is a variable parameter.

**Mutation** – Four types of mutation are possible, all with equal probability. A random edge may be added, removed, connected to or from a different node, or inverted in its sign (promotional or inhibitory).

**Crossover** – A random number of nodes from each parent is selected, such that the sum of nodes equals the original parent size plus or minus one. These are generated via a breadth first search from a random starting node. The two components are then joined by a single random edge.

V. Results

**Pressure and Tolerance** – first pressure and tolerance are explored with selective breeding, but no mutation or crossover. High pressure is set to 100%, where the oracle has an advice on all nodes, while low pressure is 50%. High tolerance is set to 20, while a low tolerance is set to 2. These are roughly appropriate for networks whose size remain under 120 nodes, since tolerance relates to the local neighborhood of a node. Interacting with 20 neighbors contrary to the oracle advice is only possible for highly connected nodes, whereas 2 advisory interactions is feasible even for relatively sparsely connected nodes.

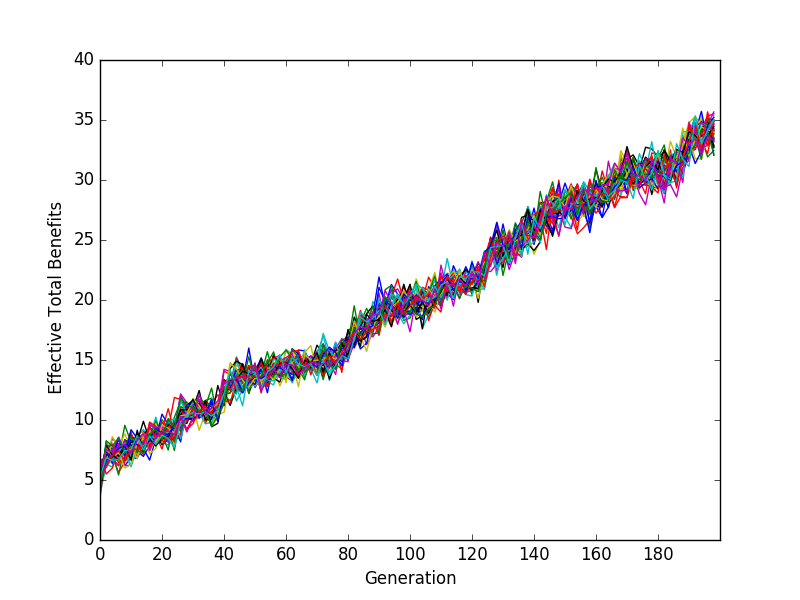
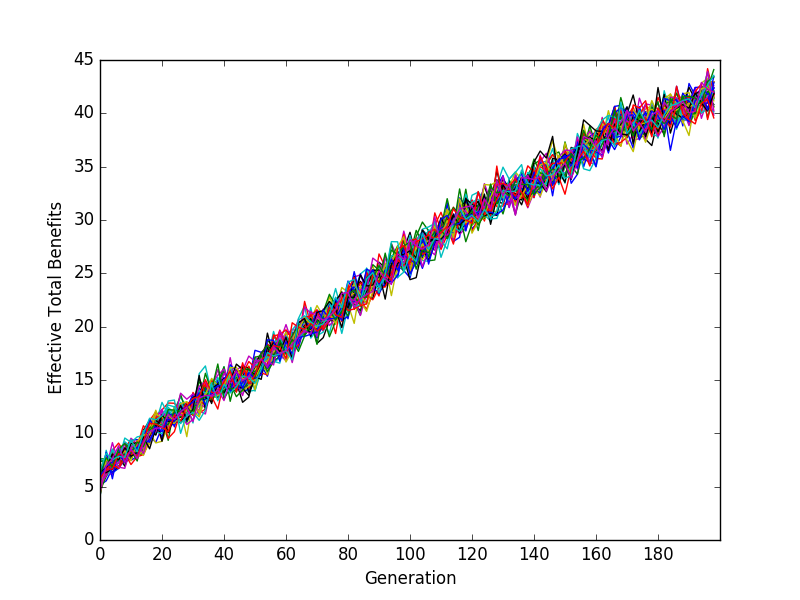
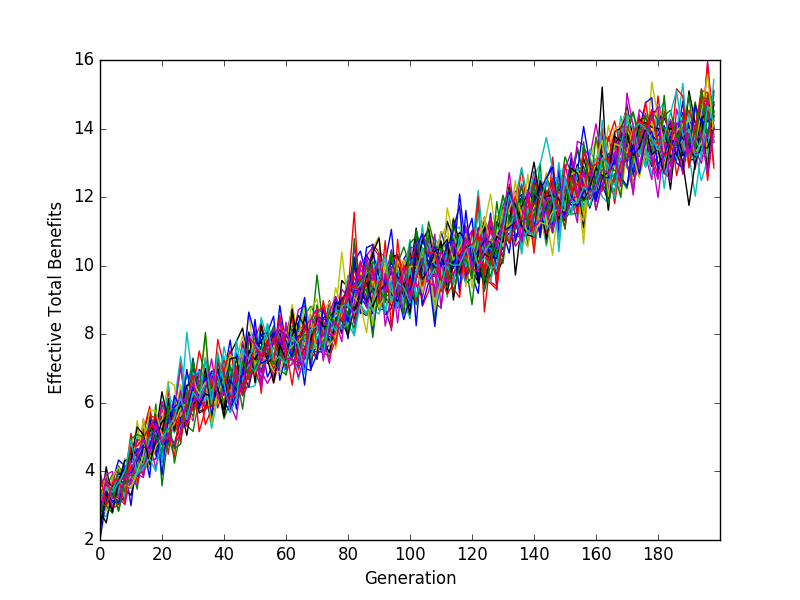
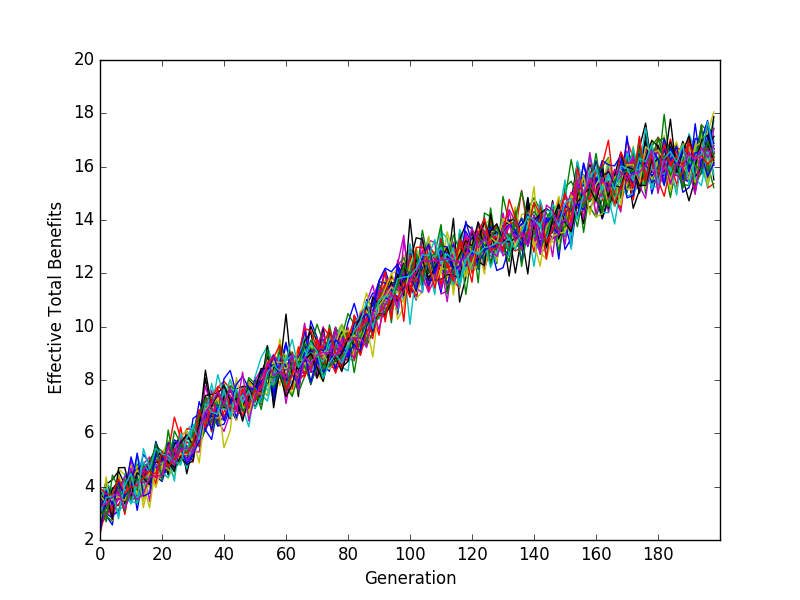
The model starts with networks of 20 nodes and evolves for 200 generations, where a new node is added every other generation. Figure 1. displays the results of the different pressure and tolerance conditions on a population, where each line represents one individual. High pressure exhibits a faster change in fitness and decreases noise for both ETB and RGGR. The increase is about two-fold, which is roughly proportional to the change in pressure. This makes intuitive sense, since pressure corresponds to how selective for fitness the environment is. High tolerance marginally decelerates ETB, possibly by allowing for more interactions against the oracle advice and hence reducing selection. Meanwhile, low tolerance has slightly faster RGGR growth under high pressure conditions. This is likely because a low tolerance preferences nodes that are not highly connected and hence unlikely to accrue damages.

Low Pressure, High Tolerance

High Pressure, High Tolerance

High Pressure, Low Tolerance

Low Pressure, Low Tolerance



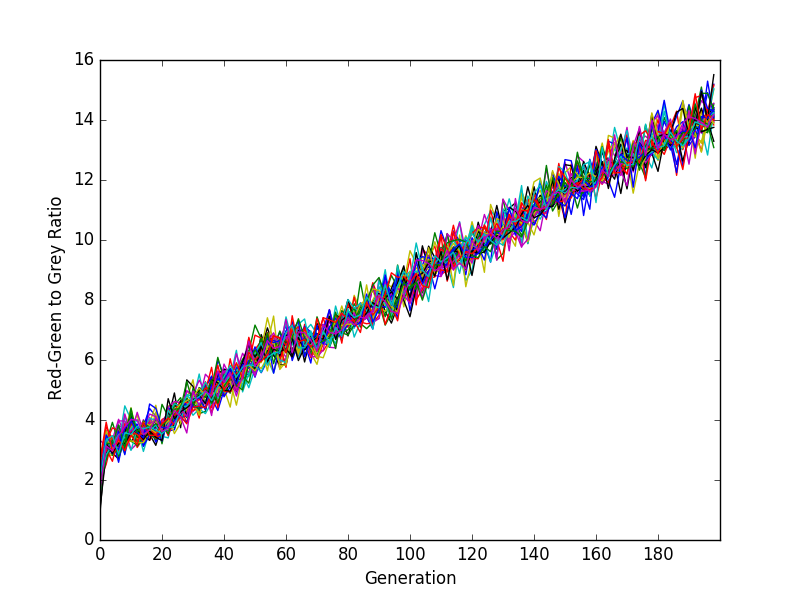
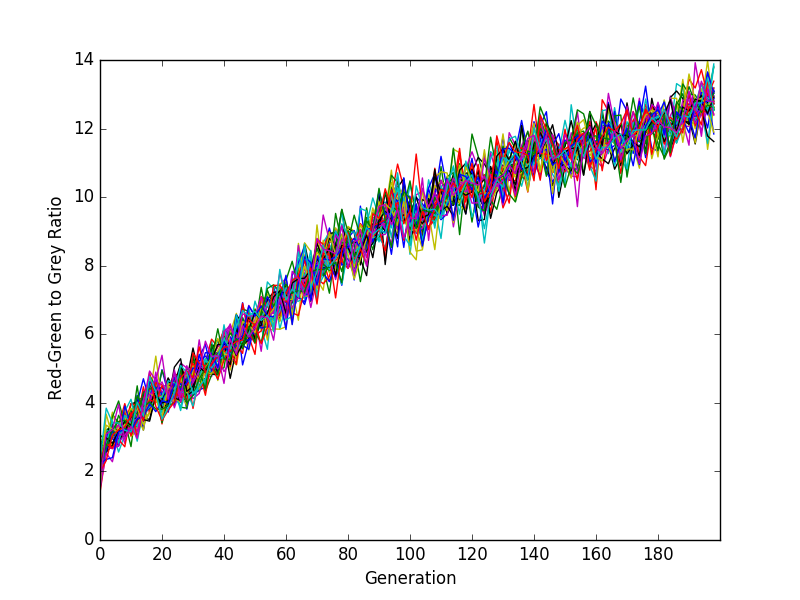
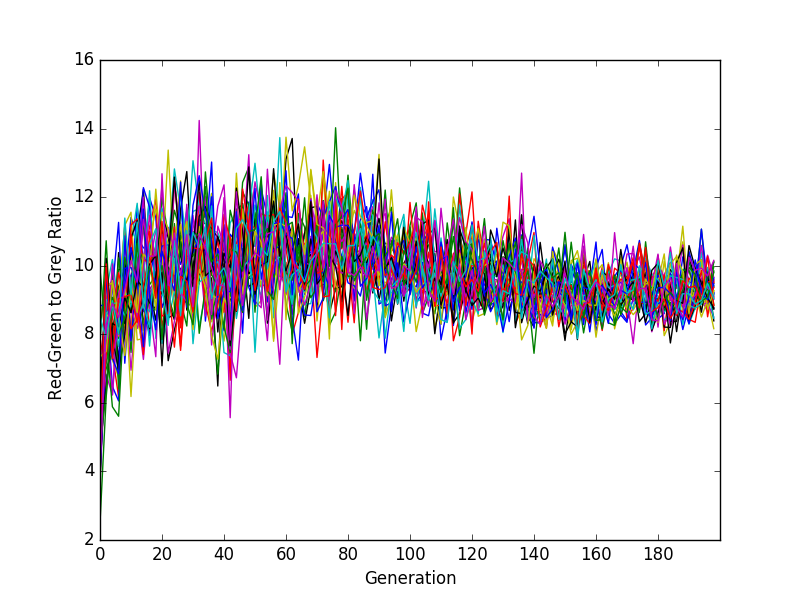
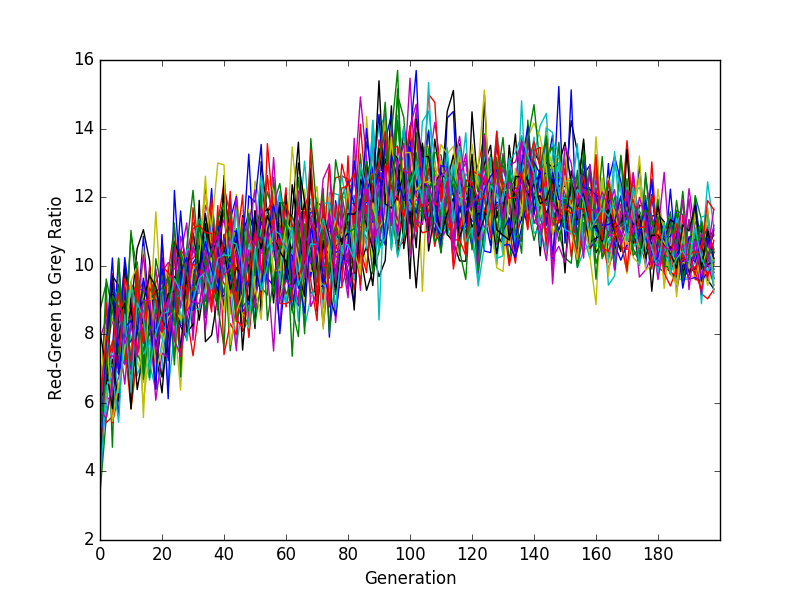
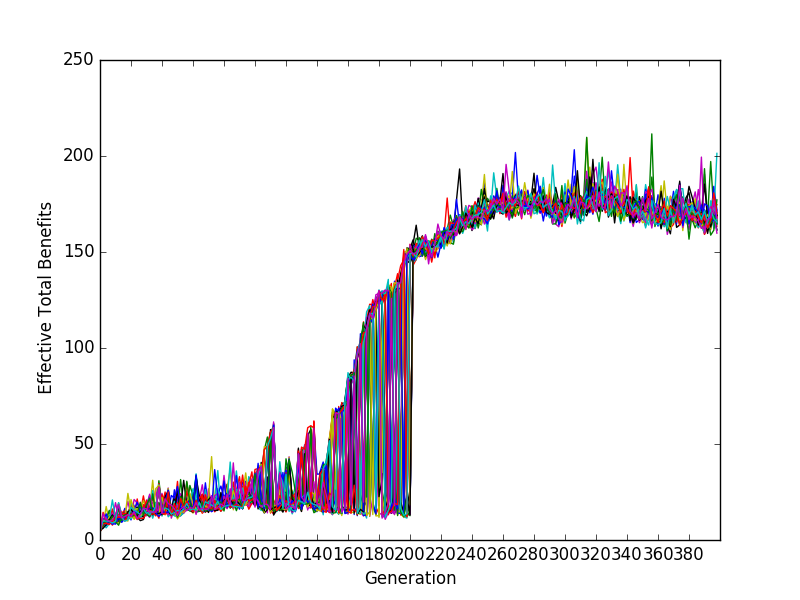
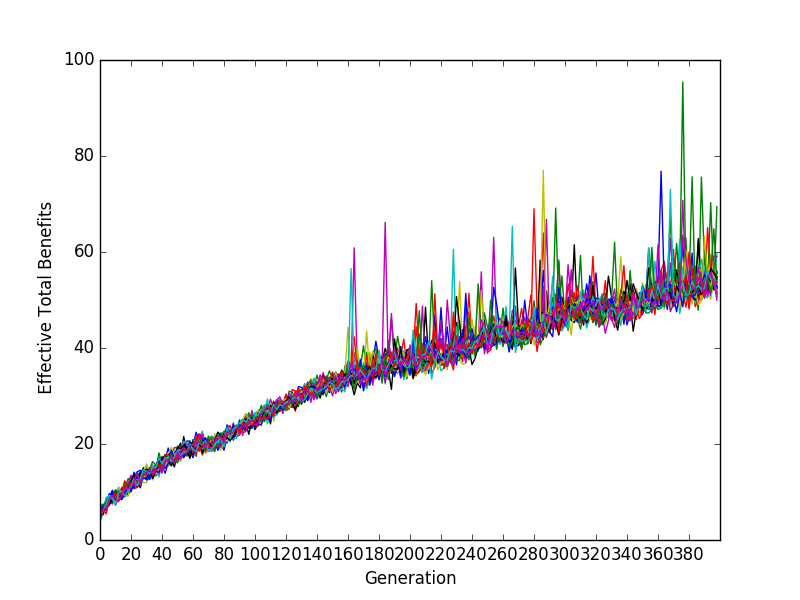


Figure 1. ETB and RGGR resulting from pressure at 50% (low) and 100% (high), and with tolerance at 2 (low) and 20 (high). Individuals are selected by fitness at each generation, but no mutation or crossover. Each line represents an individual in the population. Note that the axis marks are not consistent between images.

**Mutation** – The mutation rate represents the number of mutations per node in the network. A low mutation rate of 0.01 and a high mutation rate of 0.1 are considered. The model starts with networks of 20 nodes and evolves for 400 generations, where a new node is added every other generation. A pressure of 100% and tolerance of 10 were used. Figure 3. depicts the outcome. A high mutation rate has a faster change in fitness, but does so by only capitalizing on ETB, while entirely disregarding RGGR. Slower mutation means more gradual change in fitness, leading to a lower total fitness. Initially the slower mutation manages to balance ETB and RGGR, but abruptly sacrifices RGGR at generation 160.

High Mutation

Low Mutation



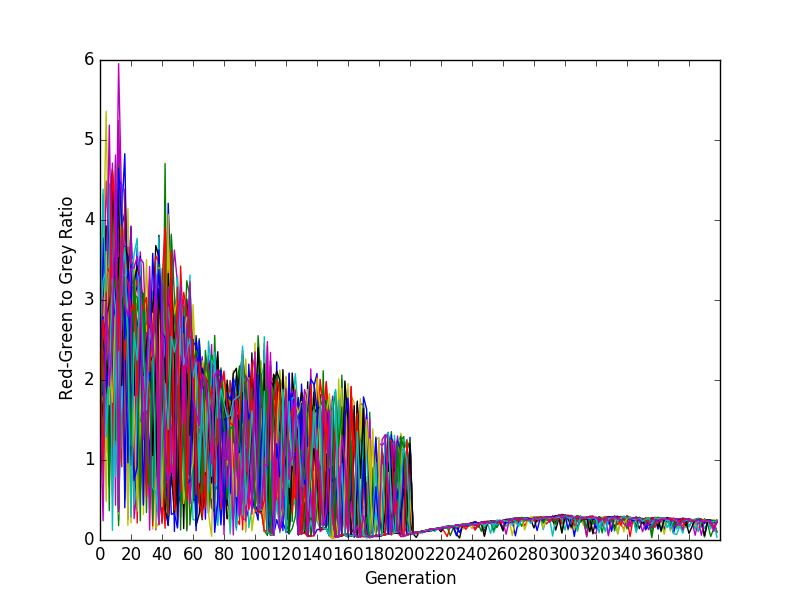
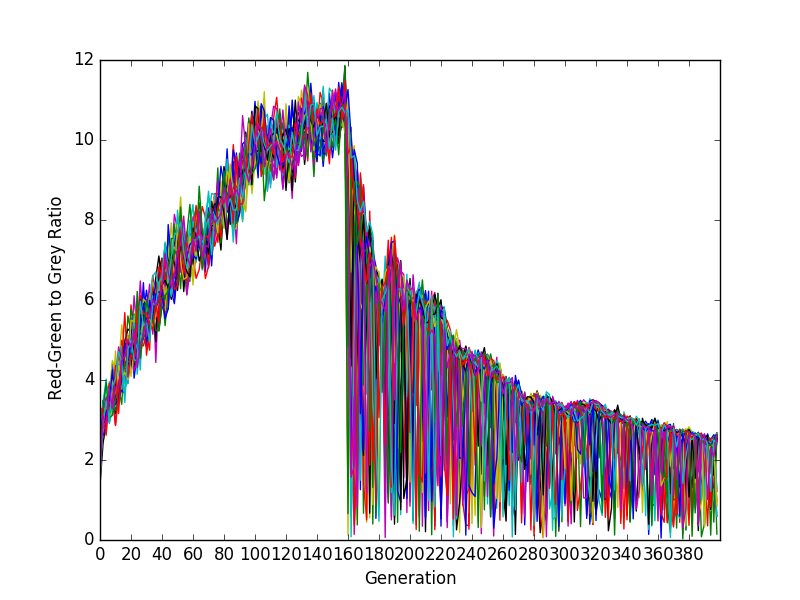


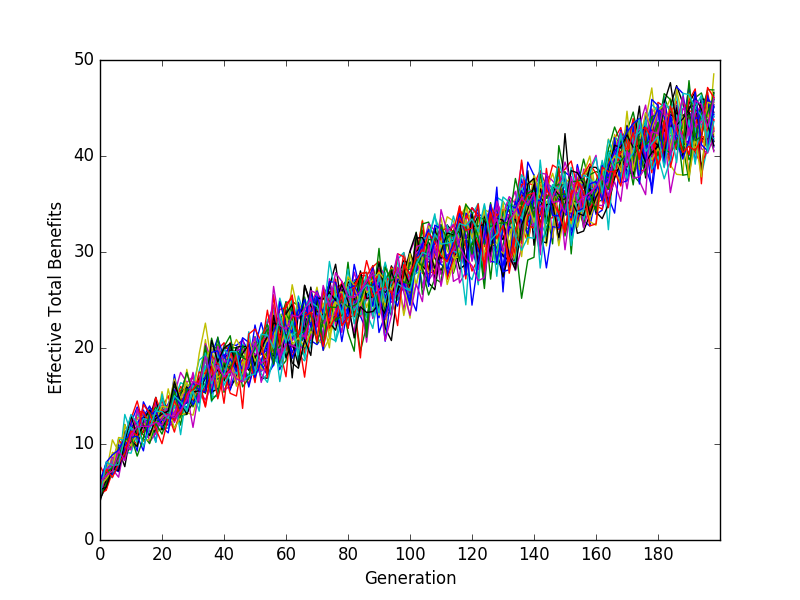
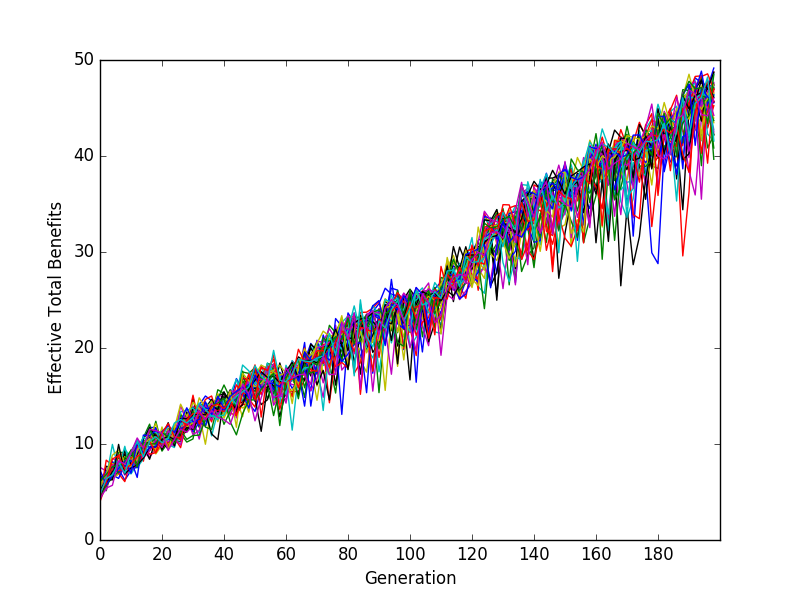
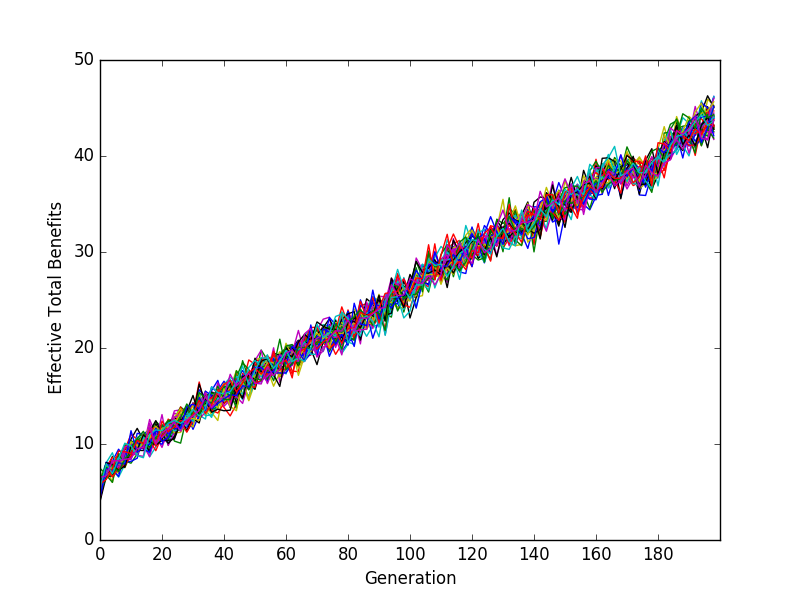
Figure 2. ETB and RGGR from 0.01 (low) and 0.1 (high) mutation rates. Individuals are mutated and selected by fitness at each generation, but are not bred with crossover. Each line represents an individual in the population.

**Crossover** – The breadth-first search based crossed over is tested with a low of 50% and high of all new individuals generated via crossover. The model starts with networks of 20 nodes and evolves for 200 generations, where a new node is added every other generation. A pressure of 100% and tolerance of 10 were used. Other pressure and tolerance combinations were explored, but do not improve crossover performance. Figure 2. reveals that the breadth-first crossover function has no noticeable benefit and may be counterproductive. ETB appears to remain fairly constant, although the thicker population bands indicate a noisier system. Meanwhile, RGGR accelerates slower with a high crossover, but yields similar acceleration using low or no crossover. As with ETB, significantly more noise is observed. Mutation and crossover were not explored together due to the ineffective nature of the crossover operation.

High Crossover

Low Crossover

No Crossover



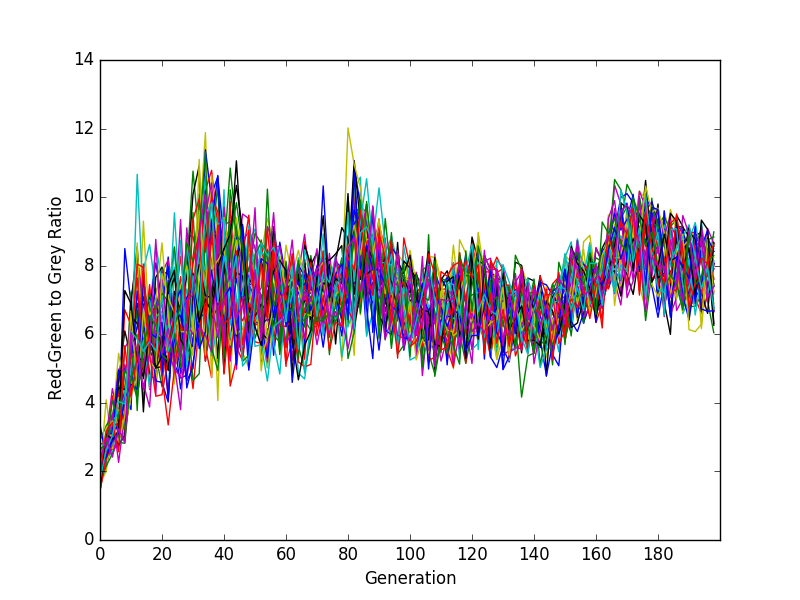
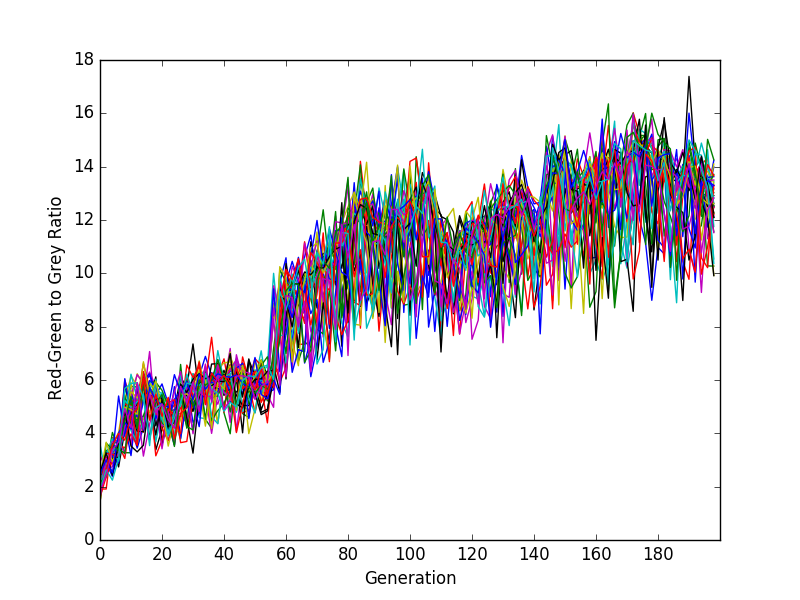
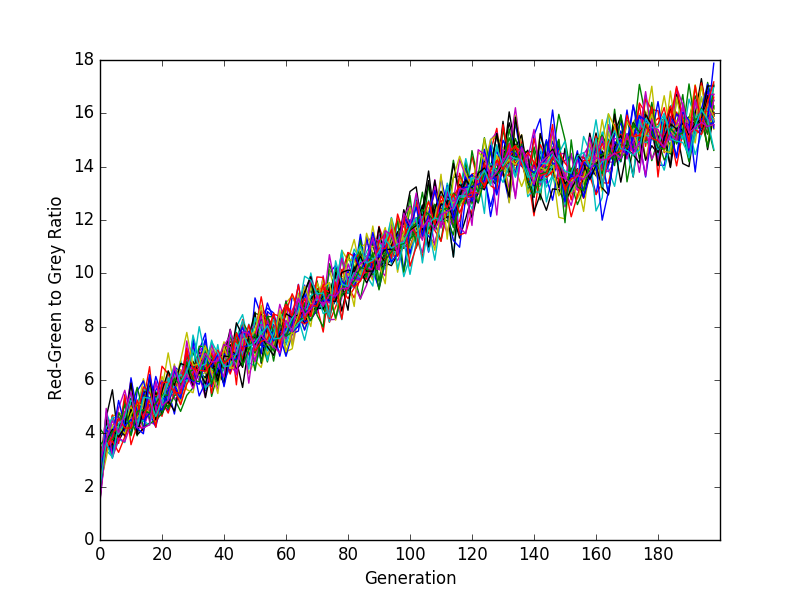


Figure 3. ETB and RGGR from crossover with 0% (none), 50% (low), and 100% (high) new individuals bred via crossover. Individuals are selected by fitness at each generation, but are not mutated. Each line represents an individual in the population.

VI. Discussion

An evolutionary model of computational tractability is evaluated when subject to different amounts of pressure, tolerance, mutation rates, and crossover. Unsurprisingly, high pressure scales with a faster change in fitness and less noise. Low tolerance may promote RGGR growth by discouraging neighbor interactions. These observations guide further testing of mutation and crossover and their impact on the rate of change in fitness.

ETB and RGGR represent opposing forces of differentiation. ETB is inclined toward higher connectivity, while RGGR selects for sparsely connected nodes. Mutual success requires that the two operate on different subsets of nodes. A low mutation rate is shown to balance the two for a time, but eventually abandons RGGR. This occurs at generation 160, which is precisely the point where the net reaches size 100 (starting at size 20 and growing every other generation). Since the mutation rate is discrete, at network size of 100 is the transition between 9 mutations a generation and 10. Nodes that were previously contributing to RGGR are abruptly converted into hubs to capitalize on ETB. This suggests that mutation should not scale with the size of the network, as it should not be capable of converting a node’s identity in a single generation.

Meanwhile, the crossover operation did not yield positive results. It is possible that it could improve given some refinement. One explanation is that new crosses faced an initial drop in fitness, for instance in connecting the two parent components, and did not have enough time to rectify the problem before being out-competed. One possibility is to follow the NEAT approach and label each node such that a node in an individual has a corresponding node in the other individuals. The population could then speciate to protect new crossover innovations. This would also reinforce node-specific differentiation, where the population reaches a consensus as to which nodes should remain sparsely connect and which should evolve as hubs.

ETB and RGGR are normalized relative to the rest of the population to evaluate fitness. This results in evolution towards the feature with the most immediate opportunity, which roughly translates to the one with more variance within the population. When evolving too aggressively, such as a high mutation frequency, the model tends to prefer ETB. This implies that ETB is more immediately accessible, but not that it produces a more tractable network in the end. ETB and RGGR might be better normalized in comparison to their entire history, rather than within the current generation.

VII. References

1 Atiia, Ali and Walidspühl, Jérôme, “Computational Intractability Can Explain the Topology of Biological Networks.” (unpublished)

2 García-Pedrajas, Nicolás, Ortiz-Boyer, Domingo, and Hervás-Martínez, César. 2006. An alternative approach for neural network evolution with a genetic algorithm: Crossover by combinatorial optimization. *Neural Netw.* 19, 4 (May 2006), 514-528. DOI=http://dx.doi.org/10.1016/j.neunet.2005.08.014

3 Shamrani, Mohammed, François, Major, and Walidspühl, Jérôme, “Evolution by Computational Selection.” ARXIV, May 2015. Web.

4 Stanley, Kenneth O., and Risto Miikkulainen. “Evolving Neural Networks through Augmenting Topologies.” *Evolutionary Computation* 10.2 (2002): 99-127. Web.