Using Novelty-biased GA to Sample Diversity in Graphs Satisfying Constraints

Peter Overbury peteroverbury@gmail.com Luc Berthouze L.Berthouze@sussex.ac.uk

Centre for Computational Neuroscience and Robotics Department of Informatics University of Sussex Falmer BN1 9QH, UK

ABSTRACT

Generating networks that satisfy a given set of constraints can be very challenging, especially when the metrics being controlled for are not very prescriptive and the networks could potentially exhibit very different higher-order structure within those constraints. Network-generating algorithms typically produce fairly contrived networks and lack mechanisms by which to systematically sample the space of network solutions. In this paper, we explore the potential of a multi-objective novelty-biased GA to provide a viable alternative to these algorithms. We believe our results provide the first proof of principle that (i) it is possible to use GAs to generate graphs satisfying set levels of key classical graph theoretic properties and (ii) it is possible to generate diverse solutions within these constraints.

Categories and Subject Descriptors

G.2.2 [Mathematics of Computing]: Graph Theory— Graph algorithms

Keywords

Network generation; degree distribution; clustering; higherorder structure; novelty search; graph diversity; coevolutionary dynamic; multi-objective optimisation

1. INTRODUCTION

Generating networks with set levels of classical network characteristics is a non-trivial task and available techniques typically are only capable of creating networks that satisfy a limited number of classical network characteristics such as global clustering, degree distribution, and, more recently, motif distribution for a limited set of such motifs. In terms of how realistic and usable the generated networks are, these algorithms have two significant limitations. First, they are only concerned with fitting the constraints, as opposed to

GECCO '15 July 11-15, 2015, Madrid, Spain © 2015 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-3488-4/15/07.

DOI: http://dx.doi.org/10.1145/2739482.2764637

generating diverse networks that fit the constraints. Second, there is increasing evidence that controlling for classical indicators such as degree distribution and global clustering does little by way of accounting for important structural differences including local clustering and higher-order structure. There is therefore a need for a more open ended or flexible method for generating networks, one which will not only satisfy a set of criteria but also provide the means to sample the diversity of the space of solutions. As a population-based method of searching the feature space, genetic algorithms (GA) are in principle well suited to this multi-objective problem. Here, we report preliminary work exploring the viability of using GAs to (a) generate networks that satisfy various requirements and (b) sample the diversity of the space of possible network structures under those requirements.

2. METHODS

An extended version [2] of this paper in which both methods and results are described in full is available at: http: //arxiv.org/abs/1503.06342.

We encoded the networks (assumed to be undirected and unweighted) through their adjacency matrix, thus making the extraction of their theoretic characteristics trivial. All measures described in the paper were implemented using the Brain Connectivity Toolbox [4] in the Matlab environment.

2.1 Fitness

There were two components to the fitness of an individual: CF, its compliance with the constraints (namely, a specified degree distribution and a set global clustering coefficient) and NF, its novelty in relation to valid networks previously evolved by the GA (measured in terms of 7 structural metrics). For simplicity, we used a weighted sum (or priori) approach [1], with the respective weighting determined through experiments. The NF fitness was subtracted from the CF fitness in order to reward diversity.

2.2 Population updating

To mitigate the effects of increasing network size, we used a steady state GA. Elitism was introduced as follows. When valid networks appeared, they were removed (one individual per generation at most) from the population and placed in a separately maintained pool of solutions. It was against those solutions that the NF fitness was calculated. Thus, the distribution of solutions in the pool could be thought of as a coarse approximation of the prevalence of structure types in the space of solutions.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).



Figure 1: Three networks from the pool of solutions obtained from one run for Poisson networks with 20 nodes and global clustering C=0.2.

2.3 Selection and mutation

The population was initialised by randomly assigning 1 or 0 to each element of the individual's genes. The fitness of each network in the population was calculated and then updated each time a mutation was applied to an individual. The mutation rate was set as the percentage of *potential* links of a randomly picked node whose values were flipped with probability 0.5. To avoid that similar networks keep being selected after their inclusion in the pool, rank-based selection in descending order was used.

3. **RESULTS**

We investigated the performance of our approach when sampling target configurations along two dimensions: network size (6-200 nodes) and topological complexity (homogeneous and heterogeneous networks with global clustering from 0.2 to 0.8). The degree distribution component of the CF fitness was calculated on the basis of the R^2 value obtained when fitting the evolved degree distribution with the expected one. Full results are available in [2]. In all cases, we found that it was possible to evolve valid networks and that there was diversity in the pool of solutions. Figure 1 illustrates one scenario of Poisson networks (see caption).

Effectiveness of our approach in controlling diversity in the pool of valid networks was assessed by comparing 3 conditions. In condition NF:off, the NF value was held to 0 throughout evolution. In NF:on, it was computed using all 7 structural metrics. Finally, in NF:Loc Clus, only the range of local clustering was used to determine NF. As shown by Figure 2, there is greater diversity with NF:on across all distributions. The increase in diversity is not uniform, with normally distributed networks showing the largest effect.

Diversity in the produced networks was also assessed in terms of network characteristics not used in the calculation of the NF fitness. Specifically, we determined the prevalence of 5 order-4 motifs (fully connected squares, squares with one diagonal, empty squares, stars and open quadruples) using the motif-counting algorithm in [3]. As shown by Figure 3, substantial diversity in motif composition was observed despite the preserved global clustering.

4. CONCLUSIONS

Being able to generate diverse networks exhibiting given network theoretic characteristics would be very useful. Indeed, an essential tool for inference is the availability of



Figure 2: Diversity in k-regular (K=5, C=0.6), Poisson (C=0.2) and normal (C=0.6) 12-node networks when three novelty scenarios are considered: no novelty (blue), default novelty (red), novelty calculated as the range of local clustering (green).



Figure 3: Prevalence of order-4 motifs in a sample of valid networks found for 12-node normallydistributed networks with C=0.6. Motifs are: fully connected square (c4), square with one diagonal (d4), square (e4), star (s4), open chain (u4).

proper null models. However, the current state of the art is quite limited, with available mathematical methods only capable of dealing with a few coarse indicators (typically degree distribution and global clustering). Even then, there is currently no mechanism by which to explore the diversity of the space of solutions for a given set of constraints. With this paper, we set on exploring the possibility that GAs provide a viable alternative. We described a novelty-biased methodology which we tested in a range of scenarios (different degree distributions and global clustering coefficients). Our results suggest that GAs are indeed a potentially viable approach to sampling the space of solutions of networks satisfying network theoretic characteristics. As far as we know, our paper is the first proof of concept in this area.

5. **REFERENCES**

- A. Konak, D. W. Coit, and A. E. Smith. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, 91(9):992–1007, 2006.
- [2] P. Overbury and L. Berthouze. Using novelty-biased GA to sample diversity in graphs satisfying constraints. arXiv e-prints, 2015, 1503.06342.
- [3] M. Ritchie, L. Berthouze, T. House, and I. Z. Kiss. Higher-order structure and epidemic dynamics in clustered networks. *Journal of Theoretical Biology*, 348:21–32, 2014.
- [4] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.