Overlapping community detection in complex networks with memetic algorithms

by

Ademir Cristiano Gabardo

Ph.D. Thesis

submitted in partial fulfilment

of the requirements for the Degree of

Doctor of Philosophy

Supervisor: Professor Regina Berretta

Co-supervisor: Professor Pablo Moscato

THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

The University of Newcastle

Faculty of Engineering and Built Environment

School of Electrical Engineering and Computing

July, 2018

Overlapping community detection in complex networks with memetic algorithms

Statement of Originality

I hereby certify that the work embodied in the thesis is my own work, conducted under normal supervision.

The thesis contains no material which has been accepted, or is being examined, for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.

Ademir Cristiano Gabardo
31st July 2018

Acknowledgments

First, I would like to express my sincere gratitude to my advisers, Professor Regina Berretta and Professor Pablo Moscato, for their valuable knowledge and continuous support of my PhD study. Their guidance helped me throughout my research and while writing this thesis. I could not have imagined having better advisers for my PhD.

This research was supported by the Brazilian National Council for Scientific and Technological Development (CNPQ).[1]

A research thesis may have a single author. However, this thesis is unquestionably not a journey of a single person there is so much to consider and there are so many to thank. These brief pages are not enough to express the enormous gratitude I have for all those somehow affected by my decision to pursue a PhD. With love and passion, I acknowledge my wife, ♡Juliana Gabardo♡, who supported me during this long journey. Without her love, I certainly would not have survived the challenging paths of academic research.

I also acknowledge my mother, Maria Francisca Gabardo, for her enormous courage in crossing the oceans and coming to Australia to bring me her heartwarming mother's love, and my father, Jair Gabardo, for supporting my decision and encouraging me to move forward. With deep respect, I acknowledge my father-in-law and my mother-in-law, who promptly supported our decision to move abroad.

I acknowledge my son, Adrian Gabardo, for embarking on this great adventure by my side and endure the challenges of moving overseas.

I am certain that the most valuable treasure in life are the good friends we meet along the way, and acknowledge my friends from the Priority Research Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine and Life and Economy Applications of Data Science research groups: Amer Abu Zaher, Amir

[1] CNPq - `http://cnpq.br`

Salehipour, Claudio Sanhueza Lobos, Francia Jimenez, Heloisa Milioli, Jake Fitzsimmons, Inna Tischenko, Leila Moslemi Naeni, Luke Mathieson, Lukasz Olech, Marta A. Jimenez, Nader Mahmoudi, Natalie de Vries, Nasimul Noman and Shannon Fenn. I thank them for the friendly environment, the countless coffee breaks and the support.

Although distant, but always present, I acknowledge my friend and mentor, Professor Heitor Silvério Lopes, who always had a word of support and encouragement.

I also extend my deep respect and esteem to the other friends have also been part of this journey, especially Luiz Fernando Nunes, Lauro Cesar Galvão, Chidambaram Chidambaram and Osvalcir Peters.

Ademir Cristiano Gabardo
The University of Newcastle
July 2018

Contents

List of Tables

List of Figures

List of Algorithms

Abstract

Intricate relationships exist among the billions of individuals who form our society. The interactions that occur between thousands of genes within our cells comprise our biology. Millions of financial institutions perform billions of transactions daily, creating complex networks of entities. These are just a few examples of the many complex systems surrounding us.

Network science comprehends the collection, management, analysis, interpretation and presentation of complex systems that employ networks. Collections of interconnected items that are usually represented by a graph showing a set of nodes joined by edges.

Complex networks often present community structures where nodes preferentially link to one another. Examples of community structures include groups of friends in society, groups of co-functioning genes in gene networks and groups of similar products in co-purchasing networks, among many others. Detecting the community structure in networks offers important information about the organisation and functioning of such groups. For many phenomena represented by networks, communities can be overlapping, with nodes participating in multiple communities. For example, a person participates in several social organisations; a gene is related to different biological functions; a product can be sold in different markets.

Revealing the community structure in complex networks is no trivial task and can lead to a non-deterministic polynomial-time hardness (NP-hard) computational problem. In this thesis, we approach the overlapping community detection problem using memetic algorithms, metaheuristics that employ a population-based search and local-search inspired by Darwinian principles of natural evolution and Dawkins's notion of a meme defined as a unit of cultural evolution. We detail the construction of two different memetic algorithms, present computational results, compare our methods with other state-of-the-art metaheuristics and present applications of our methods as case studies.